Introduction to interval computations

Götz Alefeld
Fakultät für Mathematik
Universität Karlsruhe
Karlsruhe, Federal Republic of Germany

Jürgen Herzberger
Fakultät für Mathematik
Universität Oldenburg
Oldenburg, Federal Republic of Germany

Translated by Jon Rokne
Department of Computer Science
The University of Calgary
Calgary, Alberta, Canada

ACADEMIC PRESS
A subsidiary of Harcourt Brace Jovanovich, Publishers New York, London, Paris, San Diego, San Francisco, São Paolo, Sydney, Tokyo, Toronto, 1983

Г. Алефельд
Ю. Херцбергер

Введение в интервальные вычисления

Перевод с английского
Г. Е. Минца и А. Г. Яковлева

под редакцией
Ю. В. Матиясевича

МОСКВА «МИР»
1987

ББК 22.19
А48
УДК 519.62

Алефельд Г., Херцбергер Ю.

А48 Введение в интервальные вычисления: Пер. с англ. — М.: Мир, 1987. — 360 с., ил.

Книга известных специалистов (ФРГ), посвященная новому направлению вычислительной математики — интервальному анализу, ставящему своей целью не только получение приближенного ответа для решаемой задачи, но и одновременное автоматическое вычисление оценки возможной погрешности. В книге рассматриваются алгоритмы решения основных задач линейной алгебры, нахождения вещественных и комплексных корней многочленов, решения систем нелинейных уравнений. Для этих алгоритмов оцениваются точность и трудоемкость; для части алгоритмов приведена реализация на Алголе 60.

Для студентов, аспирантов, инженеров и научных сотрудников, имеющих дело с вычислениями на ЭВМ.

А $\frac{1702070000-368}{041\,(01)-87}$ 27-87 ч. 1

ББК 22.19

НИЗ640

БИБЛИОТЕКА
Красноярского
политехнического
института

Редакция литературы по математическим наукам

© 1983, by Academic Press, Inc.
© перевод на русский язык, с дополнениями, «Мир», 1987

ПРЕДИСЛОВИЕ РЕДАКТОРА ПЕРЕВОДА

Эта книга относится к активно развивающемуся направлению вычислительной математики, получившему название «интервальный анализ» или даже «интервальная математика». Интерес к этому направлению обусловлен, в первую очередь, возрастающим применением ЭВМ для всевозможных расчетов. Если эти расчеты проводятся по традиционной схеме, то часто очень трудно, а иногда просто невозможно дать математически строгий ответ на естественный вопрос о соотношении числа, напечатанного машиной, и истинного значения вычисляемой величины. Интервальный анализ дает возможность получить такой ответ ценой увеличения времени счета.

Основная идея интервального анализа чрезвычайно проста. Вещественное число представляется в памяти ЭВМ не одним, а двумя машинными числами — оценкой снизу и оценкой сверху, образующими интервальное число. Арифметические операции над этими числами выполняются так, что если $[a_1, a_2] = [b_1, b_2] \circ [c_1, c_2]$, $b \in [b_1, b_2]$, $c \in [c_1, c_2]$, то $b \circ c \in [a_1, a_2]$, где $\circ \in \{+, -, \times, /\}$. Таким образом, интервальный анализ дает возможность автоматически учитывать погрешности в задании исходных данных и погрешности, вызываемые машинными округлениями. Это создает основу для аккуратного учета погрешностей, вызываемых используемым приближенным методом вычислений.

Первой монографией по интервальному анализу была вышедшая в 1966 г. книга Р. Е. Мура, которая во многом способствовала становлению этого направления. В настоящее время общее число опубликованных работ в этой области приближается к 2 тысячам. Такое обилие публикаций объясняется, в частности, тем, что практическая реализация описанной выше основной идеи интервального анализа сталкивается с большими трудностями. Например, оказалось, что алгоритм Гаусса может стать неприменимым к системе линейных уравнений с интервальными коэффициентами из-за возникающих делений на интервалы, содержащие нуль. В других случаях, когда традиционный численный метод переносился на интервальные числа, в результате вычислений получались интервалы, гарантированно содержащие истинное значение, но столь широкие, что найденные

двусторонние оценки были практически бесполезными. Выяснилось, что для успешного применения интервального анализа нужно пересмотреть весь арсенал численных методов.

Такой пересмотр и делается в данной книге. При этом авторы не ограничиваются только описанием различных алгоритмов, но и проводят их сравнение как по достигаемой точности, так и по вычислительной сложности. Большая часть книги посвящена задачам линейной алгебры, что неудивительно ввиду той базисной роли, которую играет линейная алгебра в численных методах. Вместе с тем вне рамок книги остались многие другие важные для приложений разделы математики, в которых интервальный анализ успешно применяется, например, обыкновенные дифференциальные уравнения. Имеющиеся по интервальной математике монографии и сборники собраны одним из переводчиков книги А. Г. Яковлевым в составленный специально для русского издания список (см. Книги по интервальной математике).

Интервальный анализ давно вышел за рамки чисто теоретического исследования и достаточно широко применяется на практике с помощью соответствующего программного обеспечения. В приложениях B и C приведены программы для интервальных вычислений на Алголе 60. В предисловии к немецкому изданию авторы обосновывают свой выбор этого языка. За прошедшее с тех пор время, с одной стороны, появились реализации интервальной арифметики на более мощных языках. С другой стороны, был разработан стандарт ANSI/IEEE на машинное представление чисел и правила выполнения операций над ними. Этот стандарт уже реализован программно на некоторых ЭВМ, а также воплощен аппаратно в новых микропроцессорах. Это в значительной степени облегчает программную реализацию интервальных вычислений, которые становятся ненамного более медленными, чем традиционные. Более подробные сведения о машинной реализации интервальных вычислений можно найти в Добавлении, также составленном А. Г. Яковлевым.

Следует сделать несколько замечаний по терминологии, которая не устоялась ни в мире, ни на русском языке, что надо иметь в виду при чтении работ других авторов. На интервальные числа можно смотреть двояко: как на способ задания вещественных чисел, которые мы знаем лишь с некоторой погрешностью, и как на самостоятельные объекты. Это различие почти нигде не ощущается, пожалуй, его единственное проявление — это определение равенства интервалов. При первом подходе равенство $[a_1, a_2] = [b_1, b_2]$ выполняется тогда и только тогда, когда $a_1 = a_2 = b_1 = b_2$, при втором, принятом в этой книге, оно справедливо тогда и только тогда, когда $a_1 = b_1$ и $a_2 = b_2$.

Переводчики вслед за авторами использовали слово «локализация» в двух смыслах, называя, например, локализацией

нуля функции как интервал, его содержащий, так и процесс нахождения такого интервала. С другой стороны, для перевода также многозначного термина interval evaluation переводчики использовали в соответствии со смыслом два разных выражения: интервальное оценивание и интервальная оценивающая функция. Следуя оригиналу, переводчики допускали такие вольности речи, как «неподвижная точка уравнения», «монотонность включения» и т. п.

Интервальный анализ в настоящее время активно развивается во многих странах. Регулярно проводятся конференции по интервальному анализу, в том числе международные. В Институте прикладной математики Фрейбургского университета (ФРГ) создана и непрерывно пополняется открытая для всех «Интервальная библиотека», содержащая почти все работы в этой области. Там же в периодическом издании Freiburger Intervall-Berichte по мере накопления материала публикуются библиографические списки, в последнем из которых (Garloff J. Interval Mathematics. A bibliography//Freiburger Intervall-Berichte. — 1985. — N6. — P.1 — 222) содержится более полутора тысяч ссылок. К сожалению, в нашей стране интервальный анализ — важное для приложений направление вычислительной математики — развивается и используется в гораздо меньшей степени, чем это имеет место за рубежом. Настоящая книга должна способствовать привлечению молодых сил в эту перспективную область.

Хотелось бы еще выразить благодарность авторам, приславшим список дополнительной литературы к русскому изданию, который поможет читателю составить наиболее полное представление о современном состоянии данной области.

Работа по переводу книги была распределена следующим образом: предисловия, гл. 1—6 и приложения перевел А. Г. Яковлев, гл. 7—22 перевел Г. Е. Минц.

Ю. В. Матиясевич

Йозефу Хайнхольду, который указал нам этот путь.

ПРЕДИСЛОВИЕ К АНГЛИЙСКОМУ ИЗДАНИЮ

Настоящая книга является пересмотренным и расширенным вариантом оригинального немецкого издания. Расположение материала и его содержание существенно не изменились. Все утверждения в предисловии к немецкому изданию, касающиеся соглашений о формулах, теоремах, леммах и определениях, остаются в силе, так же как и замечания, относящиеся к расположению и отбору материала.

Почти во всех разделах английского издания имеются небольшие изменения, перекомпоновки и некоторое количество дополнительного материала. Больше всего нового материала добавлено в гл. 3, 4, 7, 12, 14, 15, 18 и 19 (оценивание с помощью интервальной арифметики, интервальные вычисления с округлениями, симметрический короткошаговый метод, применимость алгоритма Гаусса, решение нелинейных уравнений и систем уравнений).

Как и в немецком издании, мы не могли дать здесь исчерпывающего обзора приложений интервальных вычислений, которые теперь стали еще более многочисленными. Такой обзор чрезмерно увеличил бы объем книги. Вместо этого мы сосредоточились на главных результатах и средствах, предоставляемых интервальной арифметикой. Интересующегося читателя мы отсылаем к новой расширенной библиографии в английском издании.

Английская рукопись была отпечатана г-жой Б. Траянович из Берлина и г-жой Ф. Найт из Калгари. Доктора Х. Корнелиус, Л. Платцёдер, Х. Швандт и Г. Майер прочли различные варианты рукописи.

Мы благодарны профессору В. Райнбольдту и издательству «Академик Пресс» за публикацию этой книги в серии «Computer Science and Applied Mathematics». Мы особенно благодарны нашему коллеге профессору Джону Рокну из Калгари, который весьма тщательно и точно перевел эту книгу; перевод субсидировался Университетом г. Калгари (фонд Killam Resident Fellowship).

Карлсруэ и Ольденбург — *Г. Алефельд и Ю. Херцбергер*

ПРЕДИСЛОВИЕ К НЕМЕЦКОМУ ИЗДАНИЮ

Эта книга создана на основе семинаров и лекций, которые мы помогали проводить, а также вели сами в университете г. Карлсруэ. Его целью является как систематическое введение в методы интервального анализа, так и единообразное изложение интервальных аналитических методов, получивших развитие в последние годы. Особое внимание уделено рассмотрению методов решения алгебраических задач вычислительной математики с использованием интервального анализа. Книга, таким образом, не претендует на полное представление и описание всего того, что было сделано за последнее время в области интервального анализа, и не дает полного обзора всех его возможных приложений.

В соответствии с целью книги первые главы содержат полное описание элементарных свойств четырех основных операций над вещественными интервалами и составленных из них выражений. При этом задача состоит в выработке системы правил для работы с интервалами, достаточной для дальнейших приложений. Эта система строится всего лишь на нескольких понятиях, таких как абсолютная величина, ширина и расстояние. Там, где это возможно, понятия, введенные для вещественных векторных пространств, обобщаются на случай интервальных. Поскольку в интервальном анализе встречаются пространства с нелинейной структурой, свойства операций в интервальных пространствах, вообще говоря, отличается от таких свойств в линейных пространствах.

Из аналогичных соображений мы исходим при описании двух реализаций интервальной арифметики комплексных чисел. Это так называемые прямоугольная и круговая комплексные интервальные арифметики. Ограничивая интервальные операции, можно показать сходство этих трех арифметик и использовать их в приложениях. Кроме того, читателю будут полезны имеющиеся в книге правила, общие для векторных пространств, а литература, указанная в ссылках, позволит ему ознакомиться с интервальными операциями и другими понятиями, необходимыми для специальных целей. В дальнейшем, однако, эти понятия в настоящей книге затронуты не будут.

Средства интервальной арифметики для работы с вещественными и комплексными числами, векторами и матрицами, составленными из таких чисел, применяются при решении нескольких избранных задач. Все они могут быть охарактеризованы как задачи, связанные с вычислением и уточнением множества решений. Их можно разделить на два класса: с «неточными» и «точными» исходными данными. Типичная задача из первого класса состоит в решении уравнения, данные для которого могут меняться в некотором интервале, путем вычисления наилучшего объемлющего множества. Примером может служить система линейных уравнений с коэффициентами, меняющимися в некоторых интервалах. В классе задач с точными исходными данными интервальный анализ используется в методах, которые порождают последовательность границ, сходящуюся к решению при сравнительно слабых предположениях.

Оглядываясь назад, можно сказать, что отправным пунктом в использовании интервального анализа было желание математиков-вычислителей сделать возможным выполнение алгоритмов на цифровой машине с полностью автоматическим учетом погрешностей округления и, тем самым, автоматически вычислять строгие границы погрешности. Так возник подход, порой дающий о себе знать и поныне, который состоит в замене вещественных значений в алгоритмах, разработанных для поля вещественных чисел, на интервалы с последующим выполнением этих алгоритмов в машинной интервальной арифметике; однако он редко позволяет получить практически ценные границы ошибки. Эффекты, возникающие при его применении, должным образом не изучены, а интерес к нему проявлен лишь в немногих работах. Исследование подобного рода наиболее полно представлено в [268].

Все рассмотренные в этой книге задачи относятся к упомянутым выше классам с «неточными» или «точными» данными. Построение объемлющих множеств, производимое обычно с помощью итерационной процедуры, выполняется таким образом, что сохраняются основные свойства решения, такие как монотонность последовательности включений или сходимость. Когда вычисления выполняются в подходящей машинной интервальной арифметике, соответствующие свойства также сохраняются. Рассмотрение машинной интервальной арифметики сводится тем самым к описанию требований к удобным реализациям интервальных операций в компьютере. В приложении В приведены написанные на Алголе интервальные процедуры; они использованы в алгоритмах из приложения С. Мы выбрали такую форму реализации машинной интервальной арифметики, поскольку на большинстве компьютеров имеются трансляторы с Алгола, с помощью которых могут быть выполнены эти вычисления. Поэтому

затраты на программирование интервальных операций как процедур невелики, что можно увидеть на примерах программ из приложения С. Языки программирования высокого уровня, допускающие введение нестандартных типов данных и операций, здесь не рассматриваются. Было бы желательно, однако, в будущем ввести в такие языки машинные интервалы и машинные интервальные операции. Это уже и теперь можно сделать, к примеру, в Алголе 68 или в Симуле, поскольку они являются расширяемыми языками. Однако эти языки не стали общеупотребительными. До тех пор, пока не появятся распространенные языки с подобными возможностями, предлагаемый метод останется, по-видимому, разумным компромиссом для сохранения общедоступности программ.

После этого короткого объяснения выбранного подхода и обзора содержания книги представим ключевые пункты отдельных разделов.

Книга включает 22 главы и три приложения. Большинство глав заканчивается замечаниями к излагаемому материалу и ссылками на результаты тематически связанных работ.

В гл. 1 простым способом вводятся вещественные интервальные операции и даются наиболее важные соотношения между вещественными интервалами. В гл. 2 рассмотрены метрические свойства. В гл. 3 объясняется понятие интервального оценивания вещественной функции и доказываются некоторые качественные утверждения. Эта глава не дает исчерпывающего описания темы по причинам, изложенным в замечаниях к ней. Тем не менее, она содержит все необходимые в дальнейшем свойства интервального оценивания. В гл. 4 вводится и обсуждается реализация вещественных интервальных вычислений на цифровых вычислительных машинах. Приведены условия построения машинной интервальной арифметики, гарантирующие сохранение большинства важнейших свойств вещественной интервальной арифметики при замене ее машинной. В гл. 5 вводятся два возможных подхода к аналогам интервалов на комплексной плоскости, а их метрические свойства изучаются в гл. 6.

Глава 7 содержит первые приложения. Приведены различные методы для уточнения области, в которой находится решение вещественного («точного») уравнения; доказаны утверждения о сходимости и порядке сходимости.

Две следующие главы содержат методы одновременного вычисления всех корней многочлена с точными коэффициентами. Наибольший интерес представляет тот факт, что предложенный в гл. 8 метод для многочленов, все корни которых вещественны и лежат в попарно непересекающихся интервалах, всегда сходится к этим корням по крайней мере квадратично. В гл. 9 приведены методы для общего случая многочлена с комплексными

корнями. Комплексные корни последовательно заключаются в круговые области. Эти методы локально сходятся и имеют сходимость третьего порядка. Глава 10 знакомит с арифметикой интервальных матриц и их метрическими свойствами. В гл. 11 доказываются некоторые утверждения, касающиеся нахождения неподвижной точки системы уравнений с интервальными коэффициентами. Для таких неподвижных точек также дана соответствующая интерпретация. Особый интерес в этой главе представляют возможности, открывающиеся для вычисления константы Липшица для систем уравнений с интервальными коэффициентами.

Главы 12—14 касаются специальных задач, возникающих при решении систем линейных уравнений в форме, поддающейся итерированию. В гл. 12 и 13 даны обобщения полношагового и короткошагового методов, а также метода релаксации, использующие системы линейных уравнений с интервальными коэффициентами. Определены необходимые и достаточные условия сходимости. Замечательно, что полношаговый и короткошаговый методы либо оба сходятся, либо оба расходятся. Метод релаксации не дает каких-либо преимуществ. Кроме того, в гл. 13 показано, что методы решения систем линейных уравнений, использующие теорию монотонно разложимых операторов, являются специальными случаями методов, разработанных с помощью интервального анализа.

В гл. 14 обсуждаются некоторые модификации полношагового и короткошагового методов и предлагается оптимальный метод.

Главы 15—17 касаются, в основном, адаптации алгоритма Гаусса для вычисления множества, включающего решение системы уравнений, имеющих интервальные коэффициенты. В гл. 15 описаны два класса интервальных матриц, для которых алгоритм Гаусса всегда выдает объемлющее множество. В гл. 16 намечен один из первых подходов к «решению» произвольной системы линейных уравнений и сделаны некоторые качественные утверждения. В гл. 17 мы обсуждаем возможность уточнения объемлющих множеств для систем уравнений с интервальными коэффициентами.

В гл. 18 представлены методы со сходимостью произвольно высоких порядков для итерационной локализации матрицы, обратной к вещественной, а также приведены утверждения о сходимости этих методов.

В гл. 19—22 рассматриваются решения «точных» систем нелинейных уравнений. В этих главах интервальный анализ используется (аналогично гл. 8 и 9) исключительного для того, чтобы гарантировать сходимость различных модификаций методов, перенесенных из традиционного численного анализа. При-

менение аппарата интервального анализа дает возможность доказать весьма полезные утверждения о сходимости, которые без привлечения этого аппарата могут быть доказаны лишь в локальной постановке или при наличии ограничивающих условий, например, выпуклости системы. В гл. 19 предлагаются некоторые модификации метода Ньютона для решения нелинейных систем уравнений путем итерационной локализации. Там мы используем алгоритм Гаусса для решения системы линейных уравнений с интервальными коэффициентами, упомянутой в гл. 15. Глава 20 содержит методы, не требующие решения линейных систем с интервальными коэффициентами, но имеющие тем не менее сходимость по крайней мере второго порядка. В гл. 21 предлагаются специальные методы для нелинейных систем, подходящие, например, для систем, возникающих при дискретизации нелинейных краевых задач. Здесь наиболее наглядно продемонстрировано, что разумное использование некоторых возможностей интервального анализа может приводить к разработке весьма полезных методов. В гл. 22 излагаются методы, которые применяются к системам линейных уравнений с интервальными коэффициентами, возникающими при использовании метода Ньютона для решения нелинейных систем. Множество решений такой линейной системы уравнений с интервальными коэффициентами содержит решение нелинейной системы, и она приближенно решается за один или несколько шагов полношагового или короткошагового метода. Такие методы, как это хорошо известно, предназначены для решения больших нелинейных систем уравнений, которые получаются, например, при дискретизации нелинейных дифференциальных уравнений в частных производных. Оказалось, что можно доказать сходимость этих методов для достаточно общего случая.

В приложении А мы даем определение порядка сходимости итерационного метода, записанного в интервальной арифметике. Это определение используется в других частях текста в формулировках утверждений о порядке сходимости. В приложении В мы приводим пример реализации машинной интервальной арифметики на Алголе 60. Приложение С содержит три алгол-процедуры для ранее обсужденных алгоритмов.

В конце каждой главы даны ссылки на статьи-первоисточники, а также добавочные ссылки на статьи, освещающие сходные проблемы. В конце книги приведен список литературы, не претендующий, однако на полноту. Он составлен из ссылок на обширную подборку публикаций, которые собирались и систематизировались в течение нескольких лет как материал для нашей собственной работы, лекций и семинаров. Тем не менее мы уверены, что он содержит ссылки на важнейшие статьи по интервальной математике, а также на другие статьи, связанные

с тематикой книги. Сюда относятся, в частности, публикации, касающиеся вычислительных аспектов решения непрерывных задач путем их дискретизации с учетом погрешностей дискретизации, особенно, когда мы имеем дело с задачей Коши или краевыми задачами. Если бы в книгу были включены результаты этих статей, то, несомненно, пришлось бы выйти за предполагавшийся объем книги.

Отсылки в тексте даются в квадратных скобках. Определения, теоремы, леммы и следствия, а также формулы имеют сквозную последовательную нумерацию в тексте каждой главы, что позволяет ссылаться на них указанием главы и порядкового номера. Таким образом, запись (6.8) обозначает формулу (8) из гл. 6. Если мы ссылаемся на формулу из этой же самой главы, то номер главы опускается. Такие же соглашения установлены для теорем, лемм, следствий и определений. Доказательства теорем, лемм и следствий заканчиваются знаком ■.

В большинстве глав, в которых рассматриваются численные методы, содержатся числовые примеры. Для всех них, за исключением лишь некоторых, программы были составлены П. Т. Шпеком на компьютере X8 в вычислительном центре университета г. Карлсруэ. Мы благодарны ему за его помощь.

Значительная часть материала книги была рассмотрена на семинаре по прикладной математике, проводимом проф. У. Кулишем и проф. Й. Вайсингером в Университете г. Карлсруэ. Мы благодарим их за это, а также всех участников семинара за многочисленные обсуждения. Эти обсуждения часто способствовали упрощению изложения и прояснению некоторых понятий. Добавим, что мы многое извлекли из цикла лекций проф. Кулиша.

Д-р Кристиан Ульрих из Института прикладной математики при Университете г. Карлсруэ прочел рукопись целиком. С его помощью было устранено много мелких ошибок.

Второй автор, со своей стороны, хотел бы поблагодарить компанию Volkswagenwerk за финансовую поддержку при завершении работ по некоторым частям книги.

Кроме того, нам хотелось бы также поблагодарить г-жу Лиобу Шинделе за весьма тщательную перепечатку рукописи.

Но главная наша благодарность — профессору Кулишу, который ввел нас в рассматриваемый здесь круг проблем и энергично содействовал совершенствованию этой книги.

Карлсруэ и Беркли *Г. Алефельд и Ю. Херцбергер*

Глава 1
ВЕЩЕСТВЕННАЯ ИНТЕРВАЛЬНАЯ АРИФМЕТИКА

В этой и последующих главах поле вещественных чисел будет обозначаться через $\mathbb{R}$, а строчные буквы a, b, c, ..., x, y, z будут использоваться для обозначения его элементов. Подмножество A множества $\mathbb{R}$, такое что

$$A = [a_1, a_2] = \{t \mid a_1 \leqslant t \leqslant a_2,\ a_1,\ a_2 \in \mathbb{R}\},$$

будет называться замкнутым вещественным интервалом, или просто интервалом, если это не сможет вызвать недоразумения. Впоследствии в некоторых случаях, чтобы избежать путаницы, мы будем обозначать границы интервала A через

$$i(A) = a_1 \quad \text{и} \quad s(A) = a_2.$$

Множество всех замкнутых вещественных интервалов обозначим через $I(\mathbb{R})$, а прописные буквы A, B, C, ..., X, Y, Z зарезервируем для обозначения его элементов. Всякое вещественное число x из $\mathbb{R}$ может считаться особым элементом из $I(\mathbb{R})$, имеющим вид $[x, x]$; чаще всего мы будем называть его точечным интервалом.

Определение 1. Два интервала $A = [a_1, a_2]$ и $B = [b_1, b_2]$ называются равными (записывается: $A = B$), если они равны в теоретико-множественном смысле. ■

Из этого определения непосредственно следует, что

$$A = B \Leftrightarrow a_1 = b_1, \quad a_2 = b_2.$$

Отношение равенства между двумя элементами из $I(\mathbb{R})$ рефлексивно, симметрично и транзитивно.

Теперь мы можем обобщить арифметику вещественных чисел, введя операции над элементами из $I(\mathbb{R})$.

Определение 2. Пусть $* \in \{+, -, \cdot, :\}$ — бинарная операция на множестве вещественных чисел. Если $A, B \in I(\mathbb{R})$, то

$$A * B = \{z = a * b \mid a \in A,\ b \in B\} \tag{1}$$

определяет бинарную операцию на $I(\mathbb{R})$. ■

В определении предполагается, что в случае деления $0 \notin B$, и в дальнейшем это явно указываться не будет. Заметим также, что символы операций на множествах $I(\mathbb{R})$ и $\mathbb{R}$ совпадают. Это не должно вызывать затруднений, поскольку из контекста всегда ясно, к чему применяется операция: к вещественным числам или интервалам.

Результат операции над интервалами $A = [a_1, a_2]$ и $B = [b_1, b_2]$ может быть получен явно с помощью формул

$$(2)\quad \begin{cases} A + B = [a_1 + b_1,\ a_2 + b_2], \\ A - A = [a_1 - b_2,\ a_2 - b_1] = A + [-1,\ -1] \cdot B, \\ A \cdot B = [\min\{a_1b_1,\ a_1b_2,\ a_2b_1,\ a_2b_2\}, \\ \qquad\qquad \max\{a_1b_1,\ a_1b_2,\ a_2b_1,\ a_2b_2\}], \\ A : B = [a_1a_2] \cdot [1/b_2,\ 1/b_1]. \end{cases}$$

Их обоснованием служит тот факт, что $z = f(x, y) = x * y$, где $* \in \{+, -, \cdot, :\}$ — непрерывная функция на компактном множестве. Следовательно, $f(x, y)$ принимает как наименьшее и наибольшее значения, так и все прочие значения между ними. Таким образом, $A * B$ — также замкнутый вещественный интервал. Теперь понятно, что (2) — это формулы для вычисления наименьшего и наибольшего значений $f(x, y)$. Из сказанного следует замкнутость множества $I(\mathbb{R})$ относительно введенных таким образом операций, а также изоморфизм между вещественными числами $x, y, \ldots$ и интервалами $[x, x], [y, y], \ldots$. Поэтому всюду далее операция $[x, x] * A$, в которой участвуют точечный интервал $[x, x]$ и произвольный интервал A, будет записываться в упрощенной форме $x * A$. Кроме того, мы часто будем опускать знак умножения.

Набор операций вида (1) может быть дополнен другими традиционными, в основном унарными операциями над интервалами.

Определение 3. Если $r(x)$ — непрерывная унарная операция на $\mathbb{R}$, то

$$r(X) = [\min_{x \in X} r(x),\ \max_{x \in X} r(x)]$$

определяет соответствующую ей операцию на $I(\mathbb{R})$. ■

Примерами таких унарных операций могут служить

$$X^k (k \in \mathbb{R}),\quad e^X,\quad \ln X,\quad \sin X,\quad \cos X \text{ и т. д.}$$

Соберем теперь вместе наиболее важные свойства операций на $I(\mathbb{R})$.

Теорема 4. *Пусть* $A, B, C \in I(\mathbb{R})$. *Тогда*

(3) $$A + B = B + A, \quad A \cdot B = B \cdot A \ (\textit{коммутативность});$$

(4) $$(A + B) + C = A + (B + C),$$
$$(A \cdot B) \cdot C = A \cdot (B \cdot C) \ (\textit{ассоциативность});$$

(5) $X = [0, 0]$ *и* $Y = [1, 1]$ — *единственные нейтральные элементы соответственно сложения и умножения, т. е.*

$$A = X + A = A + X \quad \textit{для всех} \quad A \in I(\mathbb{R}) \Leftrightarrow X = [0, 0],$$
$$A = Y \cdot A = A \cdot Y \quad \textit{для всех} \quad A \in I(\mathbb{R}) \Leftrightarrow Y = [1, 1];$$

(6) $I(\mathbb{R})$ *не имеет делителей нуля;*

(7) *произвольный элемент* $A = [a_1, a_2] \in I(\mathbb{R})$, *у которого* $a_1 \neq a_2$, *не имеет обратного ни по сложению, ни по умножению. Тем не менее,*

$$0 \in A - A \quad \textit{и} \quad 1 \in A : A;$$

$$A(B + C) \subseteq AB + AC \ (\textit{субдистрибутивность}),$$

(8) $$a(B + C) = aB + aC, \quad \textit{где} \quad a \in \mathbb{R},$$

$$A(B + C) = AB + AC, \quad \textit{где} \quad bc \geqslant 0 \quad \textit{для всех} \quad b \in B \quad \textit{и} \quad c \in C.$$

Доказательство. (3): Пусть $* \in \{+, \cdot\}$. Тогда

$$A * B = \{z = a * b \mid a \in A, \ b \in B\}$$
$$= \{z = b * a \mid b \in B, \ a \in A\} = B * A.$$

(4): Пусть $* \in \{+, \cdot\}$. Тогда

$$(A * B) * C = \{z = y * c \mid y \in A * B, \ c \in C\}$$
$$= \{z = (a * b) * c \mid a \in A, \ b \in B, \ c \in C\}$$
$$= \{z = a * (b * c) \mid a \in A, \ b \in B, \ c \in C\}$$
$$= \{z = a * x \mid a \in A, \ x \in B * C\} = A * (B * C).$$

(5): Необходимость доказывается тривиально. Если N и $\bar{N}$ — два нейтральных элемента сложения, то

$$N + \bar{N} = \bar{N} \quad \text{и} \quad \bar{N} + N = N.$$

Из свойства коммутативности (3) следует, что $N = \bar{N}$. Единственность $Y = [1, 1]$, нейтрального элемента умножения, может быть показана подобным же образом.

(6): Пусть $A \cdot B = 0$, т. е.

$$A \cdot B = \{z = a \cdot b \mid a \in A, \ b \in B\} = [0, 0].$$

Из этого следует, что по крайней мере один из интервалов A и B, принадлежащих $I(\mathbb{R})$, должен быть равен $[0, 0]$.

(7): Утверждения, которые нужно доказать, эквивалентны следующим:

$$A - B = [0, 0] \Rightarrow A = [a, a] = B,$$
$$A \cdot B = [1, 1] \Rightarrow A = [a, a], \quad B = [1/a, 1/a].$$

Пусть

$$A - B = \{z = a - b \mid a \in A, b \in B\} = [0, 0].$$

Отсюда следует, что $z = a - b = 0$ для всех $a \in A$, $b \in B$. Фиксируя b из B, получаем, что $a = b$ для всех a из A, т. е. $A = [b, b]$. Соответственно можно заключить, что $B = [a, a]$ и, следовательно, $a = b$. Второе утверждение доказывается подобным же образом.

Так как для a из A

$$0 = a - a \in \{z = x - y \mid x \in A, y \in A\},$$

то очевидно, что $0 \in A - A$. Аналогично, $1 \in A : A$, где $0 \notin A$.

(8):
$$A(B + C) = \{z = a(b + c) \mid a \in A, b \in B, c \in C\}$$
$$\subseteq \{y = ab + \tilde{a}c \mid a, \tilde{a} \in A, b \in B, c \in C\}$$
$$= AB + AC.$$

Для того чтобы показать невыполнение равенства в общем случае, приведем один пример:

$$A = [0, 1], \quad B = [1, 1], \quad C = [-1, -1],$$
$$A(B + C) = [0, 0] \subset [-1, 1] = AB + AC.$$

Далее имеем

$$a(B + C) = \{z = a(b + c) \mid b \in B, c \in C\}$$
$$= \{z = ab + ac \mid b \in B, c \in C\}$$
$$= \{x = ab \mid b \in B\} + \{y = ac \mid c \in C\}$$
$$= aB + aC.$$

Доказывая последнее равенство, будем считать b_1 и c_1 неотрицательными, что не приведет к потере общности. Если $a_1 \geqslant 0$, то

$$A(B + C) = [a_1(b_1 + c_1), a_2(b_2 + c_2)]$$

и

$$AB + AC = [a_1b_1, a_2b_2] + [a_1c_1, a_2c_2] = [a_1(b_1 + c_1), a_2(b_2 + c_2)],$$

т. е. для этого случая утверждение доказано.

Случай $a_2 \leqslant 0$ может быть сведен к $a_1 \geqslant 0$ путем замены A на $-A$. Если $a_1 a_2 < 0$, то получаем

$$A(B+C) = [a_1(b_2+c_2),\ a_2(b_2+c_2)],$$

а также

$$AB + AC = [a_1 b_2,\ a_2 b_2] + [a_1 c_2,\ a_2 c_2] = [a_1(b_2+c_2),\ a_2(b_2+c_2)],$$

что доказывает утверждение (8) и для этого случая. ■

Полное описание случаев, в которых выполняется дистрибутивный закон, дали Ратшек [254] и Спаниоль [304].

Теперь мы хотим остановиться на вопросе разрешимости уравнения

$$AX = B,$$

где $A \neq [0, 0]$ и $X \in I(\mathbb{R})$. Для того чтобы ответить на этот вопрос, введем вспомогательную функцию χ:

$$\chi(A) = \begin{cases} a_1/a_2, & \text{если } |a_1| \leqslant |a_2|, \\ a_2/a_1 & \text{в остальных случаях} \end{cases}$$

(эту функцию предложил Ратшек [256]).

Справедливо следующее утверждение: уравнение $AX = B$ разрешимо относительно X из $I(\mathbb{R})$ тогда и только тогда, когда

$$\chi(A) \geqslant \chi(B).$$

Решение не единственно лишь в случае

$$\chi(A) = \chi(B) \leqslant 0.$$

Доказательство этого утверждения можно найти у Ратшека [256] (для сравнения см. статью Берти [49]).

Проиллюстрируем приведенное утверждение примером. Пусть дано уравнение

$$[1, 2]\,X = [-1, 3].$$

Равенство выполняется лишь при $X = [-1/2, 3/2]$, поскольку

$$\chi[1, 2] = \frac{1}{2} > \chi[-1, 3] = -\frac{1}{3}.$$

С другой стороны, если рассмотреть множество решений всех уравнений вида

$$ax = b,$$

у которых

$$a \in [1, 2], \quad b \in [-1, 3],$$

то получим

$$\{x = b/a \mid a \in [1, 2],\ b \in [-1, 3]\} = [-1, 3]/[1, 2] = [-1, 3] \supset X.$$

Это множество решений существенно отличается от интервала X, удовлетворяющего равенству $AX = B$. По этой причине

мы не называем X решением уравнения $AX = B$, а предпочитаем говорить об «алгебраическом» решении.

Вообще, можно доказать следующее утверждение. Пусть уравнению $AX = B$, где $0 \notin A$, удовлетворяет некоторое X из $I(\mathbb{R})$. Тогда

$$X \subseteq B : A.$$

Действительно,

$$x \in X \Rightarrow \text{существуют } a \in A,\ b \in B, \text{ для которых}$$
$$ax = b \Rightarrow x = b/a \in B : A.$$

Заметим также, что равенство $AX = B$ может быть выполнено, даже если $B : A$ не определено. Примером служит уравнение

$$\left[-\tfrac{1}{3},\ 1\right] X = [-1,\ 2],$$

для которого единственным решением является $X = [-1,\ 2]$, причем $\chi\left[-\tfrac{1}{3},\ 1\right] > \chi[-1,\ 2]$.

Основное свойство интервальных вычислений — монотонность включения. Следующая теорема разъясняет это свойство.

Теорема 5. *Пусть* $A^{(k)},\ B^{(k)} \in I(\mathbb{R}),\ k = 1,\ 2,$ *и предполагается, что*

$$A^{(k)} \subseteq B^{(k)}, \quad k = 1,\ 2.$$

Тогда для операции $*$ *из* $\{+,\ -,\ \cdot,\ :\}$ *имеем*

$$A^{(1)} * A^{(2)} \subseteq B^{(1)} * B^{(2)}. \tag{9}$$

Доказательство. Так как $A^{(k)} \subseteq B^{(k)},\ k = 1,\ 2$, то

$$A^{(1)} * A^{(2)} = \{z = x * y \mid x \in A^{(1)},\ y \in A^{(2)}\}$$
$$\subseteq \{w = u * v \mid u \in B^{(1)},\ v \in B^{(2)}\} = B^{(1)} * B^{(2)}. \blacksquare$$

Приведем частный случай теоремы 5.

Следствие 6. *Пусть* $A,\ B \in I(\mathbb{R})$ *и* $a \in A,\ b \in B$. *Тогда*

$$a * b \in A * B,$$

где $* \in \{+,\ -,\ \cdot,\ :\}$. $\blacksquare$

Унарные операции $r(X)$ из определения 3 обладают сходными свойствами:

$$X \subseteq Y \Rightarrow r(X) \subseteq r(Y),$$
$$x \in X \Rightarrow r(x) \in r(X). \tag{10}$$

Непосредственное обобщение этих соотношений на случай интервальных выражений дано в теореме 3.1.

Замечания. Это элементарное введение в вещественную интервальную арифметику соответствует описанию, данному Муром [215]. Большинство унарных операций из определения 3 легко задаются в виде функций от левой и правой границ интервального аргумента. К примеру, это можно без труда проделать для монотонных функций x^k и $\sqrt{x}$.

Четыре основные операции ($+$, $-$, $\cdot$ и $:$) на точечных множествах общего вида ввел Янг [339]. Им были получены и некоторые элементарные соотношения, например (3), (4) и (8).

Кулиш [175] исследовал, какие свойства операций, заданных на множестве M, переносятся на множество всех его подмножеств $P(M)$. Интервальные операции вида (1) получаются при этом как частный случай в числе прочих результатов.

Представление интервалов, которое применил Сунага в [316], соответствует круговым комплексным интервалам, описываемым нами в гл. 5. При этом способе записи пара чисел (a, r) обозначает интервал $[a-r, a+r]$. Данное представление было использовано им в [316] для явного описания и последующего применения интервальных операций вида (1). Заметим, однако, что операции вида (1) не являются аналогами операций из гл. 5, даже когда последние применяются к вещественным интервалам.

Ортольф в [244] отождествил интервалы $[a_1, a_2]$ с точками (a_1, a_2) из $\mathbb{R} \times \mathbb{R}$. На этой основе ему удалось построить определение операций над всеми элементами $\mathbb{R} \times \mathbb{R}$. При $a_1 \leqslant a_2$ его определение сводится к операциям вида (2). Подобным же образом над точками из $\mathbb{R} \times \mathbb{R}$ вводятся отрицание (аддитивная инверсия) и, если $0 \notin [a_1, a_2]$, обращение (мультипликативная инверсия).

Кахан предложил в [143] обобщение интервальных операций вида (2). Наряду с обычными вещественными числами аргументами обобщенных операций могут быть $+\infty$ и $-\infty$. В некоторых случаях результатом операции оказывается «интервал» Ω, включающий все вещественные числа. Кроме Ω допускаются также интервалы $[a_1, a_2[$, $]a_1, a_2[$, $]a_1, a_2]$, причем разрешены $a_1 = \pm\infty$ и $a_2 = \pm\infty$. Более того, a_2 может быть меньше, чем a_1 [1]). Для интерпретации такого представления интервалов используется ориентированная окружность, на которой располагаются вещественные числа. Введенные подобным образом интервалы могут содержать $\infty \equiv +\infty \equiv -\infty$, быть открытыми и полуоткрытыми. Их арифметика определяется в соответствии с (1).

---

[1]) Например, запись [3,2] заменяет выражение $[-\infty, 2] \cup [3, +\infty]$. Подобные объекты могут возникать в результате разрешенного в этой арифметике деления на интервал, содержащий нуль. — *Прим. перев.*

В общем виде интервальные вычисления в частично упорядоченных пространствах описал Апостолатос [31]. И на этот раз $I(\mathbb{R})$ возникает как частный случай.

Клауа разработал трехзначную теорию множеств [152]. Он вводит так называемые частичные множества и частичные кардинальные числа. Получающаяся в результате арифметика кардинальных чисел для конечного случая в точности соответствует интервальной. Таким образом, аналогом интервальной арифметики на $I(\mathbb{R})$ служат операции над трехзначными числами. Наряду с отношением $=$ из определения 1 применяется более слабое отношение $=_{\#}$, которое для $A=[a_1, a_2]$ и $B=[b_1, b_2]$ задается следующим образом:

$$A=_{\#} B \Leftrightarrow A \cap B \neq \varnothing \Leftrightarrow \max\{a_1, b_1\} \leqslant \min\{a_2, b_2\}.$$

Отношение $=_{\#}$ рефлексивно и симметрично; кроме того,

$$A=B \Rightarrow A=_{\#} B.$$

Это означает, что если $A \neq_{\#} B$, то для всех a и b, таких что $a \in A$ и $b \in B$, мы всегда имеем $a \neq b$. Соответственно

$$A \neq_{\#} B \Rightarrow A \neq B.$$

Имея в виду отношение $=_{\#}$, можно рассмотреть $I(\mathbb{R})$ как разновидность обобщенного поля и, например, доказать следующие свойства:

$X - X =_{\#} 0$ для $X \in I(\mathbb{R})$;

$AX =_{\#} B \Leftrightarrow X =_{\#} B : A$ для $A, B, X \in I(\mathbb{R})$, причем $A \neq_{\#} 0$;

$X(Y+Z) =_{\#} XY + XZ$ для $X, Y, Z \in I(\mathbb{R})$.

Такого рода свойства использовал в своей диссертации и статьях Ян [137—139].

Каухер предложил расширенное множество $\overline{I(\mathbb{R})}$, получив его как результат дополнения $I(\mathbb{R})$ так называемыми нерегулярными интервалами [148], т. е. интервалами отрицательной ширины. В этом случае точечные интервалы $[a, a]$ больше не являются минимальными элементами в смысле порядка, задаваемого отношением $\subseteq$. Все структуры $I(\mathbb{R})$ переносятся на $I(\mathbb{R}) \cup \overline{I(\mathbb{R})}$, и с помощью несобственных элементов p и $-p$ достигается замкнутость. Подобным образом можно определить деление на интервал $[a_1, a_2]$, у которого $a_1 \leqslant 0 \leqslant a_2$ и $a_1 \neq a_2$.

* * *

Литература: [31], [33], [34], [49], [52], [53], [143], [148], [152], [175], [215], [236], [244], [251], [253], [254], [256], [259], [304], [316], [339].

Глава 2
ДАЛЬНЕЙШИЕ ПОНЯТИЯ И СВОЙСТВА

Введем теперь понятие расстояния на множестве вещественных интервалов.

Определение 1. Расстояние $q(A, B)$ между двумя интервалами A и B, такими что $A = [a_1, a_2]$, $B = [b_1, b_2] \in I(\mathbb{R})$, определяется равенством

$$q(A, B) = \max\{|a_1 - b_1|, |a_2 - b_2|\}. \blacksquare$$

Легко показать, что отображение q задает на $I(\mathbb{R})$ метрику. Действительно, q обладает следующими свойствами (см. например, работу Хаусдорфа [117]):

$$q(A, B) \geqslant 0 \quad \text{и} \quad q(A, B) = 0 \Leftrightarrow A = B,$$

$$q(A, B) \leqslant q(A, C) + q(B, C) \text{ (неравенство треугольника).}$$

Выполнение неравенства треугольника проверяется следующим образом:

$$\begin{aligned} q(A, C) + q(B, C) &= \max\{|a_1 - c_1|, |a_2 - c_2|\} \\ &\quad + \max\{|b_1 - c_1|, |b_2 - c_2|\} \\ &\geqslant \max\{|a_1 - c_1| + |b_1 - c_1|, |a_2 - c_2| + |b_2 - c_2|\} \\ &\geqslant \max\{|a_1 - b_1|, |a_2 - b_2|\} = q(A, B). \end{aligned}$$

Если применить введенное таким способом расстояние к точечным интервалам, то оно сведется к обычному расстоянию между вещественными числами. Иначе говоря,

$$q([a, a], [b, b]) = |a - b|.$$

Предложенная здесь метрика является для $I(\mathbb{R})$ хаусдорфовой. Хаусдорфова метрика обобщает понятие расстояния между двумя точками в метрическом пространстве (у нас таким пространством является $\mathbb{R}$ с $q(x, y) = |x - y|$) на случай пространства всех компактных непустых подмножеств данного пространства. Если U и V — непустые компактные множества вещественных чисел, то хаусдорфово расстояние определяется как

$$q(U, V) = \max\{\sup_{v \in V} \inf_{u \in U} q(u, v), \sup_{u \in U} \inf_{v \in V} q(u, v)\}.$$

Существуют другие полезные определения хаусдорфовой метрики (см. еще раз [117]). Легко убедиться в том, что для вещественных интервалов A и B хаусдорфова метрика задается выражением из определения 1.

Вводя на множестве $I(\mathbb{R})$ метрику, мы делаем его топологическим пространством. При этом понятия сходимости и непре-

рывности могут использоваться обычным образом, как и в случае метрического пространства. В этой связи мы получаем, что последовательность интервалов $\{A^{(k)}\}_{k=0}^{\infty}$ сходится к интервалу $A = [a_1, a_2]$ тогда и только тогда, когда последовательность границ отдельных членов последовательности сходится к его соответствующим границам. Следовательно, мы можем записать

$$(1) \qquad \lim_{k\to\infty} A^{(k)} = A \Leftrightarrow \left(\lim_{k\to\infty} a_1^{(k)} = a_1 \text{ и } \lim_{k\to\infty} a_2^{(k)} = a_2\right).$$

Доказательство этого утверждения мы опускаем, так как его легко получить непосредственно из определения расстояния между двумя интервалами.

Введенная нами метрика используется в следующей теореме.

Теорема 2. *Метрическое пространство* $(I(\mathbb{R}), q)$ *с метрикой из определения* 1 *является замкнутым метрическим пространством.* ■

(Это означает, что любая интервальная последовательность Коши сходится к интервалу.)

В теореме 3 рассматривается характер сходимости широко используемого класса интервальных последовательностей.

Теорема 3. *Каждая последовательность интервалов* $\{A^{(k)}\}_{k=0}^{\infty}$, *для которой справедливо соотношение*

$$A^{(0)} \supseteq A^{(1)} \supseteq A^{(2)} \supseteq \ldots,$$

сходится к интервалу $A = \bigcap_{k=0}^{\infty} A^{(k)}$.

Доказательство. Пусть имеется последовательность границ, такая что

$$a_1^{(0)} \leqslant a_1^{(1)} \leqslant a_1^{(2)} \leqslant a_1^{(3)} \leqslant \ldots \leqslant a_2^{(3)} \leqslant a_2^{(2)} \leqslant a_2^{(1)} \leqslant a_2^{(0)}.$$

Тогда последовательность нижних границ интервалов из $\{A^{(k)}\}_{k=0}^{\infty}$, является монотонной неубывающей последовательностью вещественных чисел, ограниченной сверху величиной $a_2^{(0)}$. Эта последовательность сходится к вещественному числу a_1. Аналогично, монотонная невозрастающая последовательность вещественных чисел $\{a_2^{(k)}\}_{k=0}^{\infty}$ сходится к вещественному числу a_2, причем $a_1 \leqslant a_2$. Равенство $A = \bigcap_{k=0}^{\infty} A^{(k)}$ проверяется столь же простым способом. ■

Как видно из доказательства, каждая последовательность $\{A^{(k)}\}_{k=0}^{\infty}$, для которой

$$A^{(0)} \supseteq A^{(1)} \supseteq A^{(2)} \supseteq A^{(3)} \supseteq \ldots \supseteq B,$$

сходится к такому интервалу A, что $A \supseteq B$.

Для арифметических, а также других определенных выше операций справедлива

Теорема 4. *Введенные в гл.* 1 *операции сложения, вычитания, умножения и деления интервалов непрерывны.*

Доказательство. Мы приводим доказательство только для операции сложения. Пусть $\{A^{(k)}\}_{k=0}^{\infty}$ и $\{B^{(k)}\}_{k=0}^{\infty}$ — две последовательности интервалов, причем $\lim\limits_{k\to\infty} A^{(k)} = A$ и $\lim\limits_{k\to\infty} B^{(k)} = B$. Из (1) вытекает, что последовательность интервальных сумм $\{A^{(k)} + B^{(k)}\}_{k=0}^{\infty}$ имеет предел

$$\begin{aligned}\lim_{k\to\infty} (A^{(k)} + B^{(k)}) &= \lim_{k\to\infty} [a_1^{(k)} + b_1^{(k)},\ a_2^{(k)} + b_2^{(k)}] \\ &= [\lim_{k\to\infty} (a_1^{(k)} + b_1^{(k)}),\ \lim_{k\to\infty} (a_2^{(k)} + b_2^{(k)})] \\ &= [a_1 + b_1,\ a_2 + b_2] = A + B.\end{aligned}$$

Доказательство непрерывности остальных операций может быть проведено аналогичным способом. ■

Обобщением теоремы 4 служит (см. определение 1.3)

Следствие 5. *Пусть r — непрерывная функция и* $r(X) = [\min\limits_{x\in X} r(x), \max\limits_{x\in X} r(x)]$. *Тогда $r(X)$ — непрерывное интервальное выражение.* ■

Доказательство этого следствия основывается непосредственно на факте непрерывности функции r и поэтому здесь будет опущено. Следствие 5 гарантирует непрерывность выражений, подобных X^k, $\sin X$ и e^X.

Определение 6. Пусть $A = [a_1, a_2] \in I(\mathbb{R})$. Абсолютной величиной этого интервала будем называть величину

$$|A| = q(A, [0, 0]) = \max\{|a_1|, |a_2|\}. \blacksquare$$

Абсолютную величину интервала можно записать и в виде

$$|A| = \max_{a\in A} |a|. \tag{2}$$

Очевидно, что если $A, B \in I(\mathbb{R})$, то

$$A \subseteq B \Rightarrow |A| \leqslant |B|. \tag{3}$$

Докажем теперь некоторые свойства, связанные с метрикой на $I(\mathbb{R})$.

Теорема 7. *Пусть* $A=[a_1, a_2]$, $B=[b_1, b_2]$, $C=[c_1, c_2]$, $D=[d_1, d_2]\in I(\mathbb{R})$. *Тогда*

$$q(A+B,\ A+C)=q(B,\ C), \tag{4}$$

$$q(A+B,\ C+D)\leqslant q(A,\ C)+q(B,\ D), \tag{5}$$

$$q(aB,\ aC)=|a|q(B,\ C),\quad a\in\mathbb{R}, \tag{6}$$

$$q(AB,\ AC)\leqslant|A|q(B,\ C). \tag{7}$$

Доказательство. (4): Из определения метрики q следует, что

$$q(A+B,\ A+C)=\max\{|a_1+b_1-(a_1+c_1)|,\ |a_2+b_2-(a_2+c_2)|\}$$
$$=\max\{|b_1-c_1|,\ |b_2-c_2|\}=q(B,\ C).$$

(5): Из неравенства треугольника, предыдущего свойства (4) и симметричности q вытекает, что

$$q(A+B,\ C+D)\leqslant q(A+B,\ B+C)+q(C+D,\ B+C)$$
$$=q(A,\ C)+q(B,\ D).$$

(6): $q(aB,\ aC)=\max\{|ab_1-ac_1|,\ |ab_2-ac_2|\}=|a|q(B,\ C)$.

(7): Пусть $A=[a_1, a_2]$. Для краткости будем использовать обозначения $i(A)=a_1$ и $s(A)=a_2$. Тогда утверждение (7) можно записать в виде

$$\max\{|i(AB)-i(AC)|,\ |s(AB)-s(AC)|\}\leqslant|A|q(B,\ C).$$

Докажем, что

$$|i(AB)-i(AC)|\leqslant|A|q(B,\ C).$$

Неравенство

$$|s(AB)-s(AC)|\leqslant|A|q(B,\ C)$$

доказывается аналогично.

Перепишем предыдущее соотношение (6):

$$\max\{|i(aB)-i(aC)|,\ |s(aB)-s(aC)|\}=|a|q(B,\ C).$$

Теперь без потери общности можно предположить, что

$$i(AB)\geqslant i(AC).$$

(Случай $i(AB)<i(AC)$ рассматривается точно так же.)

Поскольку

$$AC=\{ac\,|\,a\in A,\ c\in C\},$$

существует такое a из A, что

$$i(AC)=i(aC).$$

Из свойства монотонности включения следует, что

$$aB \subseteq AB \text{ и } i(aB) \geqslant i(AB),$$

откуда видно, что

$$i(aB) - i(aC) \geqslant i(AB) - i(AC) \geqslant 0.$$

Итак,

$$\begin{aligned} |i(AB) - i(AC)| &= i(AB) - i(AC) \leqslant i(aB) - i(aC) \\ &= |i(aB) - i(aC)| \leqslant |a| q(B, C) \\ &\leqslant |A| q(B, C). \blacksquare \end{aligned}$$

Отождествляя $|A|$ с $q(A, 0)$, получаем следующие легко проверяемые свойства абсолютного значения:

$$\begin{aligned} &|A| \geqslant 0 \text{ и } |A| = 0 \Leftrightarrow A = [0, 0], \\ &|A + B| \leqslant |A| + |B|, \\ &|xA| = |x|\,|A| \text{ для } x \in \mathbb{R}, \\ &|AB| = |A|\,|B|. \end{aligned} \tag{8}$$

Вот доказательство последнего равенства:

$$\begin{aligned} |AB| &= \max_{c \in AB} |c| = \max_{a \in A,\ b \in B} |ab| = \max_{a \in A,\ b \in B} (|a|\,|b|) \\ &= \max_{a \in A} |a| \max_{b \in B} |b| = |A|\,|B|. \end{aligned}$$

Остальные соотношения доказываются подобным же образом.

Определение 8. Шириной интервала $A = [a, a_2]$ будем называть

$$d(A) = a_2 - a_1 \geqslant 0. \blacksquare$$

Множество точечных интервалов можно теперь описать как

$$\{A \in I(\mathbb{R}) \mid d(A) = 0\}.$$

Из определения 8 сразу же получаем свойства

$$A \subseteq B \Rightarrow d(A) \leqslant d(B), \tag{9}$$

$$d(A \pm B) = d(A) + d(B). \tag{10}$$

Утверждение (9) доказывается тривиально — достаточно определение 8 переписать в виде

$$d(A) = \max_{a,\ b \in A} |a - b|. \tag{11}$$

Проверим свойство (10) для операции сложения:

$$\begin{aligned} d(A + B) &= d([a_1 + b_1,\ a_2 + b_2]) \\ &= a_2 + b_2 - (a_1 + b_1) \\ &= a_2 - a_1 + b_2 - b_1 = d(A) + d(B). \end{aligned}$$

Вычитание проверяется точно так же. Кроме того, имеет место теорема.

Теорема 9. *Пусть A и B — вещественные интервалы из $I(\mathbb{R})$. Тогда*

$$d(AB) \leqslant d(A)|B| + |A|d(B), \tag{12}$$

$$d(AB) \geqslant \max\{|A|d(B), |B|d(A)\}, \tag{13}$$

$$d(aB) = |a|d(B),\ a \in \mathbb{R}, \tag{14}$$

$$d(A^n) \leqslant n|A|^{n-1} d(A),\ n = 1, 2, \ldots, \tag{15}$$

$$(A^n := A \cdot A \cdot \ldots \cdot A,\ n \text{ раз}),$$

$$d((X-x)^n) \leqslant 2(d(X))^n, \text{ где } x \in X,\ n = 1, 2, \ldots, \tag{16}$$

$$((X-x)^n := (X-x)(X-x)\ \ldots\ (X-x),\ n \text{ раз}).$$

Если $C \in I(\mathbb{R})$ *и* $0 \in C$, то

$$|C| \leqslant d(C) \leqslant 2|C|. \tag{17}$$

Доказательство. (12): Используя тождество (11), получаем

$$\begin{aligned} d(AB) &= \max_{a, a' \in A,\ b, b' \in B} |ab - a'b'| \\ &= \max_{a, a' \in A,\ b, b' \in B} |ab - ab' + ab' - a'b'| \\ &\leqslant \max_{a, a' \in A,\ b, b' \in B} \{|a(b-b')| + |(a-a')b'|\} \\ &\leqslant \max_{a \in A,\ b, b' \in B} |a||b-b'| + \max_{a, a' \in A,\ b' \in B} |a-a'||b'| \\ &= \left(\max_{a \in A} |a|\right)\left(\max_{b, b' \in B} |b-b'|\right) + \left(\max_{a, a' \in A} |a-a'|\right)\left(\max_{b' \in B} |b'|\right) \\ &= |A|d(B) + d(A)|B|. \end{aligned}$$

(13): Сначала докажем, что

$$\begin{aligned} d(AB) &= \max_{a, a' \in A,\ b, b' \in B} |ab - a'b'| \geqslant \max_{a \in A,\ b, b' \in B} |ab - ab'| \\ &= \max_{a \in A,\ b, b' \in B} |a||b-b'| = |A|d(B). \end{aligned}$$

Подобным образом можно показать, что

$$d(AB) \geqslant |B|d(A),$$

откуда сразу же вытекает (13).

(14): $$d(aB) = \max_{b, b' \in B} |ab - ab'| = \max_{b, b' \in B} \{|a||b-b'|\}$$

$$= |a| \max_{b, b' \in B} |b-b'| = |a|d(B).$$

(15): При $n = 1$ имеет место равенство. Если неравенство выполняется для некоторого $n \geqslant 1$, то, используя (12) и последнее соотношение из (8), имеем

$$\begin{aligned} d(A^{n+1}) = d(A^n A) &\leqslant d(A^n)|A| + |A|^n d(A) \\ &\leqslant n|A|^{n-1} d(A)|A| + |A|^n d(A) \\ &= (n+1)|A|^n d(A). \end{aligned}$$

(16): Поскольку $x \in X$, из (9) и свойства монотонности включения получаем

$$\begin{aligned} d((X-x)^n) &\leqslant d((X-X)^n) = d([-d(X),\ d(X)]^n) \\ &= d([-(d(X))^n,\ (d(X))^n]) = 2(d(X))^n. \end{aligned}$$

(17): Так как $0 \in C = [c_1, c_2]$, то $c_1 \leqslant 0 \leqslant c_2$, откуда имеем

$$d(C) = c_2 - c_1 = |c_2| + |c_1| \geqslant \max\{|c_1|, |c_2|\} = |C|.$$

Итак,

$$d(C) = |c_1| + |c_2| \leqslant 2\max\{|c_1|, |c_2|\} = 2|C|. \blacksquare$$

Теперь докажем следующую теорему.

Теорема 10. *Пусть A, $B \in I(\mathbb{R})$, причем A — симметричный интервал, т. е. $A = -A$. Тогда имеют место следующие свойства:*

$$AB = |B|A, \tag{18}$$

$$d(AB) = |B|d(A). \tag{19}$$

Если $b_1 \geqslant 0$ или $b_2 \leqslant 0$, то второе свойство выполняется и в случае, когда $0 \notin A$.

Доказательство. Предположим, что $A = -A$, или, что то же самое, $a_2 = a = -a_1$. Тогда

$$\begin{aligned} AB &= [\min\{ab_1, ab_2, -ab_1, -ab_2\}, \max\{ab_1, ab_2, -ab_1, -ab_2\}] \\ &= [a\min\{b_1, -b_1, b_2, -b_2\}, a\max\{b_1, -b_1, b_2, -b_2\}] \\ &= [a(-|B|), a|B|] = [-a, a]|B| = |B|A. \end{aligned}$$

Опираясь на равенство (14), получаем (19). Остальные случаи могут быть доказаны аналогичным образом. ■

Теорема 11. *Для интервалов A и B из $I(\mathbb{R})$ справедливы следующие свойства:*

$$d(A) = |A - A|, \tag{20}$$

$$A \subseteq B \Rightarrow \frac{1}{2}(d(B) - d(A)) \leqslant q(A, B) \leqslant d(B) - d(A). \tag{21}$$

Доказательство.

(20): $$d(A) = a_2 - a_1 = |A - A|.$$

(21): Пусть $A \subseteq B$. Тогда $b_1 \leqslant a_1 \leqslant a_2 \leqslant b_2$ и, следовательно,

$$q(A, B) = \max\{|a_1 - b_1|, |a_2 - b_2|\} = \max\{a_1 - b_1, b_2 - a_2\}$$
$$\leqslant b_2 - a_2 + a_1 - b_1 = b_2 - b_1 - (a_2 - a_1) = d(B) - d(A),$$

откуда

$$q(A, B) = \max\{a_1 - b_1, a_2 - b_2\} \geqslant \frac{1}{2}(a_1 - b_1 + b_2 - a_2)$$
$$= \frac{1}{2}(d(B) - d(A)). \blacksquare$$

Введем теперь на $I(\mathbb{R})$ еще одну бинарную операцию. Пусть $A, B \in I(\mathbb{R})$. Тогда отношение

(22) $$A \cap B = \{c \mid c \in A, c \in B\}$$

представляет собой теоретико-множественное пересечение двух интервалов. Результат этой операции принадлежит $I(\mathbb{R})$ тогда и только тогда, когда пересечение не пусто. В этом случае

(23) $$A \cap B = [\max\{a_1, b_1\}, \min\{a_2, b_2\}].$$

В приводимом ниже следствии собраны важные свойства операции пересечения.

Следствие 12. *Пусть* $A, B, C, D \in I(\mathbb{R})$. *Тогда*

(24) $A \subset C, \quad B \subseteq D \Rightarrow A \cap B \subseteq C \cap D$ (*монотонность включения*).

Пока операция пересечения не выводит из $I(\mathbb{R})$, *она непрерывна.*

Доказательство. Монотонность включения (24) вытекает из определения (22). Доказательство непрерывности может быть получено с помощью (23). ■

Замечания. Мур в [215] использовал хаусдорфову метрику на $I(\mathbb{R})$, соответствующую определению 1. Некоторые из правил для вычисления абсолютного значения $|A|$ и ширины $d(A)$ имеются в работах Мура [215] и Кулиша [175]. Важное в приложениях неравенство (7) впервые доказал Майер [195]; его можно также найти в [122].

Иногда абсолютное значение вводится следующим способом, основанным на определении 1.3:

$$\operatorname{abs}(A) = \left[\min_{a \in A} |a|, \max_{a \in A} |a|\right].$$

Поскольку такое определение редко применяется в приложениях, мы также не будем его использовать.

Согласно С. М. Румпу (частное сообщение), в выражении (16) можно обойтись без сомножителя 2, если уточнить соответствующую оценку. Для x, принадлежащего X, $X - x = [a, b]$, где $a \leqslant 0$, $b \geqslant 0$. Предположим, что $b \geqslant -a = |a|$ (если это не так, будем иметь дело с $x - X$). Тогда

$$(X - x)(X - x) = [ab, b^2],$$

и с помощью полной индукции получаем

$$(X - x)^n = [ab^{n-1}, b^n].$$

Следовательно,

$$d((X - x)^n) = b^n - ab^{n-1} = b^{n-1}(b - a).$$

Теперь $b - a = d(X - x) = d(X)$, и, поскольку $a \leqslant 0$, $b \geqslant 0$, имеем $b \leqslant d(X - x) = d(X)$. Итак,

$$d((X - x)^n) \leqslant d(X)^n.$$

* * *

Литература: [117], [122], [143], [175], [195], [197], [215], [236], [268], [318].

Глава 3
ИНТЕРВАЛЬНОЕ ОЦЕНИВАНИЕ И МНОЖЕСТВО ЗНАЧЕНИЙ В СЛУЧАЕ ВЕЩЕСТВЕННЫХ ФУНКЦИЙ

В этой главе мы обсуждаем непрерывные вещественные функции. Пусть f относится к их числу. Аналитическое выражение для $f = f(x)$ представляет собой запись вычислительной процедуры, выдающей значение функции f для произвольного аргумента x. Примем при этом, что все выражения, с которыми мы будем иметь дело, составлены из операций и операндов, число которых конечно; одновременно мы предполагаем, что если эти выражения вычисляются в интервальной арифметике, то составляющие их операции трактуются в соответствии с определениями 1.2 и 1.3. Выражение, содержащее константы $a^{(0)}$, ..., $a^{(m)}$, будет для наглядности записываться в виде $f(x; a^{(0)}, \ldots, a^{(m)})$. Чтобы упростить изложение, в дальнейшем мы всегда будем предполагать, что каждая из констант $a^{(k)}$, $0 \leqslant k \leqslant m$, встречается в аналитическом выражении функции только один раз. Этого всегда можно добиться, вводя новые константы, равные константам, встречающимся неоднократно.

Пример. Двумя аналитическими выражениями функции g являются

$$g^{(1)}(x;\ a) = \frac{ax}{1-x}, \quad x \neq 1, \quad x \neq 0,$$

и

$$g^{(2)}(x;\ a) = \frac{a}{1/x - 1}, \quad x \neq 1, \quad x \neq 0.$$

Запись

$$W(f,\ X;\ A^{(0)},\ \ldots,\ A^{(m)}) = \{f(x;\ a^{(0)},\ \ldots,\ a^{(m)}) \mid x \in X,\ a^{(k)} \in A^{(k)},\ 0 \leqslant k \leqslant m\} = \left[\min_{\substack{x \in X \\ a^{(k)} \in A^{(k)}, \\ 0 \leqslant k \leqslant m}} f(x;\ a^{(0)},\ \ldots,\ a^{(m)}),\ \max_{\substack{x \in X \\ a^{(k)} \in A^{(k)}, \\ 0 \leqslant k \leqslant m}} f(x;\ a^{(0)},\ \ldots,\ a^{(m)})\right]$$

будет в дальнейшем обозначать диапазон изменения функции f, причем предполагается, что x из X и $a^{(k)}$ из $A^{(k)}$, $0 \leqslant k \leqslant m$, не зависят друг от друга. Согласно этому определению, интервал $W(f,\ X;\ A^{(0)},\ \ldots,\ A^{(m)})$ будет одним и тем же при любом аналитическом выражении для f.

Пример. Возьмем g из предыдущего примера. Для

$$A = [0,\ 1] \text{ и } X = [2,\ 3]$$

получаем

$$W(g,\ [2,\ 3];\ [0,\ 1]) = \left\{\frac{ax}{1-x} \,\middle|\, 2 \leqslant x \leqslant 3,\ 0 \leqslant a \leqslant 1\right\} = [-2,\ 0].$$

Введем теперь понятие интервального оценивания вещественной функции f. Пусть для f имеется аналитическое выражение. Заменив в этом выражении все вещественные операнды и операции над ними на интервальные операнды и операции, получим выражение $f(X;\ A^{(0)},\ \ldots,\ A^{(m)})$. Если все операнды попадают в области, на которых заданы операции из определений 1.2 и 1.3, то $f(X;\ A^{(0)},\ \ldots,\ A^{(m)})$ называется *интервальной оценивающей функцией*, или, для краткости, *оценкой* f, а получение ее значения — *вычислением*, или *оцениванием*, f *в интервальной арифметике*.

Для функций, рассматриваемых нами, замена описанного типа возможна всегда. Константы $a^{(0)},\ \ldots,\ a^{(m)}$, как и переменная x, превращаются в интервалы. Очевидно, что результат оценивания функции f зависит от выбора для нее аналитического выражения. Впоследствии мы будем использовать этот факт. А сейчас приведем простой пример.

Пример. Пусть g — функция из предыдущих двух примеров. Для $A = [0, 1]$ и $X = [2, 3]$ получим две различные оценки:

$$g^{(1)}([2, 3]; [0, 1]) = \frac{[0, 1][2, 3]}{1 - [2, 3]} = [-3, 0],$$

$$g^{(2)}([2, 3]; [0, 1]) = \frac{[0, 1]}{1/[2, 3] - 1} = [-2, 0] \neq g^{(1)}([2, 3]; [0, 1]).$$

Введенные выше обозначения можно распространить на функции от нескольких переменных. В этом случае множеством значений $f(x^{(1)}, \ldots, x^{(n)}; a^{(0)}, \ldots, a^{(m)})$ при независимых $x^{(k)}$ из $X^{(k)}$, $1 \leqslant k \leqslant n$, и $a^{(j)}$ из $A^{(j)}$, $0 \leqslant j \leqslant m$, становится $W(f, X^{(1)}, \ldots, X^{(n)}; A^{(0)}, \ldots, A^{(m)})$. Подобным же образом обобщается понятие интервальной оценки — теперь она обозначается через $f(X^{(1)}, \ldots, X^{(n)}; A^{(0)}, \ldots, A^{(m)})$[1]).

Приведем теперь пример выражения, из которого не удается получить его всюду определенный интервальный аналог путем простой замены операций и операндов. Вещественная функция

$$f(x) = 1 \Big/ \left(x^2 + \frac{1}{2}\right)$$

определена для всех x из $\mathbb{R}$. Представим f в виде

$$\tilde{f}(x) = 1 \Big/ \left(x \cdot x + \frac{1}{2}\right).$$

Заменим теперь независимую переменную x на интервал $X = [-1, 1]$, содержащийся в области определения f. Замена всех операций на соответствующие им интервальные приводит к интервальному выражению

$$\tilde{f}([-1, 1]) = \frac{1}{[-1, 1][-1, 1] + \frac{1}{2}} = \frac{1}{[-1, 1] + \frac{1}{2}} = \frac{1}{\left[-\frac{1}{2}, \frac{3}{2}\right]},$$

которое не определено.

А сейчас ознакомимся с рядом свойств интервального оценивания. Два свойства, используемые в последующих утверждениях, легко выводятся из теоремы 1.5 и следствия 1.6.

Теорема 1. *Пусть f — непрерывная вещественная функция, $f(x^{(1)}, \ldots, x^{(n)}; a^{(0)}, \ldots, a^{(m)})$ — аналитическое выражение для f. Предположим также, что для интервалов $Y^{(1)}, \ldots, Y^{(n)}, B^{(0)}, \ldots, B^{(m)}$ имеется оценка $f(Y^{(1)}, \ldots, Y^{(n)}; B^{(0)}, \ldots, B^{(m)})$. Тогда*

а) *для всех*

$$X^{(k)} \subseteq Y^{(k)}, \quad A^{(j)} \subseteq B^{(j)}, \quad 1 \leqslant k \leqslant n, \quad 0 \leqslant j \leqslant m$$

---

[1]) Множество $W(f, X^{(1)}, \ldots, X^{(n)}; A^{(0)}, \ldots, A^{(m)})$ Мур [215, 451] и другие авторы называют объединенным интервальным расширением, а функцию $f(X^{(1)}, \ldots, X^{(n)}; A^{(0)}, \ldots, A^{(m)})$ — естественным интервальным расширением.— *Прим. перев.*

справедливо свойство включения

$$(1)\qquad W(f,\ X^{(1)},\ \ldots,\ X^{(n)};\ A^{(0)},\ \ldots,\ A^{(m)}) \subseteq f(X^{(1)},\ \ldots,\ X^{(n)};\ A^{(0)},\ \ldots,\ A^{(m)});$$

б) *для всех*

$X^{(k)} \subseteq Z^{(k)} \subseteq Y^{(k)}$, $A^{(j)} \subseteq C^{(j)} \subseteq B^{(j)}$, $1 \leqslant k \leqslant n$, $0 \leqslant j \leqslant m$

имеет место монотонность включения

$$(1')\qquad f(X^{(1)},\ \ldots,\ X^{(n)};\ A^{(0)},\ \ldots,\ A^{(m)}) \subseteq f(Z^{(1)},\ \ldots,\ Z^{(n)};\ C^{(0)},\ \ldots,\ C^{(m)}).\ \blacksquare$$

Пример. *Функция f задана выражением*

$$f(x;\ a) = a - x/(1+x),\quad x \neq -1.$$

Для

$$X = \left[-\tfrac{1}{2},\ 1\right],\quad Z = \left[-\tfrac{1}{2},\ 2\right],\quad A = C = [2,\ 3]$$

получаем

$$W\left(f, \left[-\tfrac{1}{2},\ 1\right];\ [2,\ 3]\right) = \left[\tfrac{3}{2},\ 4\right] \subset f\left(\left[-\tfrac{1}{2},\ 1\right];\ [2,\ 3]\right) = [0,\ 4],$$

$$f\left(\left[-\tfrac{1}{2},\ 1\right];\ [2,\ 3]\right) = [0,\ 4] \subset f\left(\left[-\tfrac{1}{2},\ 2\right];\ [2,\ 3]\right) = [-2,\ 4].$$

Свойство включения (1) позволяет соотнести множество значений функции с ее интервальной оценкой. Позднее в этой главе мы дадим формулы для их качественного сравнения.

Можно привести примеры, когда в (1) достигается равенство. Очевидно, что к их числу относится случай однократного вхождения каждой из величин $x^{(1)}, \ldots, x^{(n)}; a^{(0)}, \ldots, a^{(m)}$ в выражение $f(x^{(1)}, \ldots, x^{(n)};\ a^{(0)}, \ldots, a^{(m)})$. Прочие классы функций, обладающих этим свойством, описаны Апостолатосом и Кулишем в [33, 34].

Теорема 2. *Пусть p — многочлен от вещественной переменной x, определяемый выражением*

$$p(x;\ a^{(0)},\ \ldots,\ a^{(m)}) = (\ldots((a^{(m)}x + a^{(m-1)})^{n_{m-1}} + a^{(m-2)})^{n_{m-2}} + \ldots + a^{(1)})^{n_1} + a^{(0)},$$

где $n_\nu \geqslant 2$, $1 \leqslant \nu \leqslant m-1$. Если встречающиеся в этом выражении степени вычисляются по формуле

$$X^k = \left[\min_{x \in X} x^k,\ \max_{x \in X} x^k\right]$$

(см. определение 1.3), *то*

$$W(p,\ X;\ a^{(0)},\ \ldots,\ a^{(m)}) = p(X;\ a^{(0)},\ \ldots,\ a^{(m)}).$$

Доказательство. Для $m = 2$ истинность теоремы очевидна:

$$p(x;\ a^{(0)},\ a^{(1)},\ a^{(2)}) = (a^{(2)}x + a^{(1)})^{n_1} + a^{(0)}.$$

Остальную часть доказательства получаем полной индукцией. ■

Не всякий многочлен можно привести к форме, требуемой теоремой 2. Однако многочлен второй степени

$$p(x;\ b^{(0)},\ b^{(1)}) = x^2 + b_{(1)}x + b^{(0)}$$

может быть преобразован к виду

$$p(x;\ a^{(0)},\ a^{(1)}) = (x + a^{(1)})^2 + a^{(0)},$$

где

$$a^{(1)} = b^{(1)}/2, \quad a^{(0)} = b^{(0)} - (b^{(1)})^2/4.$$

Прочие подробности см. в [33, 34].

Наряду с носящей общий характер теоремой 1 и разобранными выше частными случаями, представляет также интерес качественное утверждение о приближении множества значений функции f с помощью ее интервальной оценки. В случае функции одной вещественной переменной справедлива

Теорема 3. *Пусть f — вещественная функция от вещественной переменной x, $f(x;\ a^{(0)},\ \ldots,\ a^{(m)})$ — ее аналитическое выражение. Обозначим через $\tilde{f}(x^{(1)},\ \ldots,\ x^{(n)};\ a^{(0)},\ \ldots,\ a^{(m)})$ выражение, полученное заменой в $f(x,\ a^{(0)},\ \ldots,\ a^{(m)})$ каждого вхождения x на новую переменную $x^{(k)}$, $1 \leqslant k \leqslant n$. Пусть определена оценивающая функция $f(Y;\ A^{(0)},\ \ldots,\ A^{(m)})$, где $Y,\ A^{(0)},\ \ldots,\ A^{(m)} \in I(\mathbb{R})$. Кроме того, предположим, что для каждой переменной $x^{(k)}$, $1 \leqslant k \leqslant n$, из интервала Y и произвольно выбранных $x^{(j)}$ из Y, $1 \leqslant j \leqslant n$, $j \neq k$, и $a^{(j)}$ из $A^{(j)}$, $0 \leqslant j \leqslant m$, выражение $f(x^{(1)},\ \ldots,\ x^{(n)};\ a^{(0)},\ \ldots,\ a^{(m)})$ удовлетворяет условию Липшица. В остальном обозначения имеют тот же смысл, что и в теореме 1. При этих предположениях для $X \subseteq Y$ имеем*

$$q(W(f,\ X;\ A^{(0)},\ \ldots,\ A^{(m)}),\ f(X;\ A^{(0)},\ \ldots,\ A^{(m)})) \leqslant \gamma d(X), \quad \gamma \geqslant 0. \tag{2}$$

Доказательство. Во-первых, заметим, что

$$\tilde{f}(x,\ \ldots,\ x;\ a^{(0)},\ \ldots,\ a^{(m)}) = f(x;\ a^{(0)},\ \ldots,\ a^{(m)}), \quad x \in Y.$$

Теперь мы можем получить интервальную оценку для f в виде

$$f(X;\ A^{(0)},\ \ldots,\ A^{(m)}) = W(\tilde{f},\ X,\ \ldots,\ X;\ A^{(0)},\ \ldots,\ A^{(m)}), \quad X \subseteq Y.$$

Остается показать, что

$$q(W(f,\ X;\ A^{(0)},\ \ldots,\ A^{(m)}),\ W(\tilde{f},\ X,\ \ldots,\ X;\ A^{(0)},\ \ldots,\ A^{(m)})) \leqslant \gamma d(X), \quad X \subseteq Y.$$

Если теперь для $X \subseteq Y$ записать

$$W(f, X; A^{(0)}, \ldots, A^{(m)}) = [f(u; a^{(0)}, \ldots, a^{(m)}), f(v; b^{(0)}, \ldots, b^{(m)})],$$
$$u, v \in X, \quad a^{(j)}, b^{(j)} \in A^{(j)}, \quad 0 \leqslant j \leqslant m,$$
$$W(\tilde{f}, X, \ldots, X; A^{(0)}, \ldots, A^{(m)})$$
$$= [\tilde{f}(x^{(1)}, \ldots, x^{(n)}; c^{(0)}, \ldots, c^{(m)}),$$
$$\tilde{f}(y^{(1)}, \ldots, y^{(n)}; e^{(0)}, \ldots, e^{(m)})],$$
$$x^{(k)}, y^{(k)} \in X, \quad 1 \leqslant k \leqslant n, \quad c^{(j)}, e^{(j)} \in A^{(j)}, \quad 0 \leqslant j \leqslant m,$$

и принять во внимание соотношение

$$W(f, X; A^{(0)}, \ldots, A^{(m)}) \subseteq W(\tilde{f}, X, \ldots, X; A^{(0)}, \ldots, A^{(m)}),$$

то получим

$$\begin{aligned}
&|f(u; a^{(0)}, \ldots, a^{(m)}) - \tilde{f}(x^{(1)}, \ldots, x^{(n)}; c^{(0)}, \ldots, c^{(m)})| \\
&= |f(u; a^{(0)}, \ldots, a^{(m)}) - \tilde{f}(x^{(1)}, \ldots, x^{(n)}; c^{(0)}, \ldots, c^{(m)})| \\
&\leqslant |f(u; c^{(0)}, \ldots, c^{(m)}) - \tilde{f}(x^{(1)}, \ldots, x^{(n)}; c^{(0)}, \ldots, c^{(m)})| \\
&= |\tilde{f}(u, \ldots, u; c^{(0)}, \ldots, c^{(m)}) - \tilde{f}(x^{(1)}, \ldots, x^{(n)}; c^{(0)}, \ldots, c^{(m)})| \\
&\leqslant \gamma \max_{1 \leqslant k \leqslant n} |u - x^{(k)}| \leqslant \gamma d(X).
\end{aligned}$$

Разность верхних границ может быть оценена аналогичным образом. Получение этих двух оценок доказывает утверждение теоремы. ■

Теорема 3, как видно из ее доказательства, легко обобщается на случай функции от нескольких переменных $x^{(1)}, \ldots, x^{(n)}$. Вместо $\gamma d(X)$ имеем величину

$$\sum_{k=1}^{n} \gamma^{(k)} d(X^{(k)}) \quad \left(\leqslant \gamma \max_{1 \leqslant k \leqslant n} d(X^{(k)})\right).$$

Следующий пример иллюстрирует тот факт, что степень близости множества значений функции f и ее интервальной оценки зависит от выбора аналитического выражения $f(x; a^{(0)}, \ldots, a^{(m)})$.

Пример. Пусть $f(x) = x - x^2$ и $X = [0, 1]$. Тогда

$$W(f, [0, 1]) = \{x - x^2 \mid 0 \leqslant x \leqslant 1\} = \left[0, \frac{1}{4}\right].$$

Различные аналитические выражения для f дают следующие результаты:

$$f^{(0)}(x) = x - x^2 \Rightarrow f^{(0)}([0,\ 1]) = [0,\ 1] - [0,\ 1] = [-1,\ 1],$$
$$f^{(1)}(x) = x(1 - x) \Rightarrow f^{(1)}([0,\ 1]) = [0,\ 1](1 - [0,\ 1]) = [0,\ 1],$$
$$f^{(2)}(x) = \frac{1}{4} - \left(x - \frac{1}{2}\right)\left(x - \frac{1}{2}\right) \Rightarrow$$
$$f^{(2)}([0,\ 1]) = \frac{1}{4} - \left([0,\ 1] - \frac{1}{2}\right)\left([0,\ 1] - \frac{1}{2}\right) = \left[0,\ \frac{1}{2}\right],$$
$$f^{(3)}(x) = \frac{1}{4} - \left(x - \frac{1}{2}\right)^2 \Rightarrow$$
$$f^{(3)}([0,\ 1]) = \frac{1}{4} - \left([0,\ 1] - \frac{1}{2}\right)^2 = \left[0,\ \frac{1}{4}\right] = W(f,\ [0,\ 1]).$$

Для некоторых классов аналитических выражений можно доказать более сильные утверждения, нежели то, которое приведено в теореме 3. К их числу относится так называемая центрированная форма записи функции (ср. Хансен [114], Шуба и Миллер [73], Миллер [207]). Центрированная форма представляет собой специальное выражение, предназначенное для оценки функции f на интервале X. Ограничим наше рассмотрение случаем одной вещественной переменной. Выберем в X произвольную точку z и представим $f(x)$ в виде

$$f(x) = f(z) + (x - z)h(x - z), \tag{3}$$

где сомножитель $h(x - z)$ зависит от новой переменной $\tilde{z}$, равной $x - z$. Будем называть (3) формой $f(x)$, центрированной относительно z. Применительно к многочленам центрированная форма есть не что иное, как обычное тейлоровское разложение $f(x)$ в окрестности точки z, записанное с сомножителем $x - z$, имеющимся у всех членов, отличных от постоянного.

Рациональная функция $f(x) = p(x)/q(x)$ может быть, согласно Ратшеку [481], приведена к центрированной форме следующим образом. Пусть n — максимум из степеней многочленов $p(x)$ и $q(x)$. Для z из X определим

$$\gamma_\nu := p^{(\nu)}(z) - f(z)q^{(\nu)}(z), \quad 1 \leqslant \nu \leqslant n.$$

Функция

$$h(y) = \sum_{\nu=1}^{n} \gamma_\nu \frac{y^{\nu-1}}{\nu!} \Big/ \sum_{\nu=0}^{s} q^{(\nu)}(z) \frac{y^\nu}{\nu!}$$

является решением функционального уравнения

$$f(x) = f(z) + (x - z)h(x - z).$$

Теорема 4. *Пусть f — вещественная функция от вещественного аргумента x и*

$$f(x) = f(z) + (x - z)\,h(x - z).$$

— аналитическое выражение для f в центрированной форме. Кроме того, пусть имеется выражение $\tilde{h}(x^{(1)} - z, \ldots, x^{(n)} - z)$, аналогичное соответствующему выражению из теоремы 3. *Допустим, что для некоторого Y из $I(\mathbb{R})$ существует интервальная оценка $f(Y)$ и для каждой своей переменной $\tilde{h}(x^{(1)} - z, \ldots, x^{(n)} - z)$ удовлетворяет условию Липшица, подобно тому как это было в теореме* 3. *Тогда для $X \subseteq Y$ выполняется соотношение*

$$q(W(f, X) f(x)) \leqslant c(d(X))^2, \quad c \geqslant 0. \tag{4}$$

Доказательство. Поскольку

$$\tilde{h}(x - z, \ldots, x - z) = h(x - z)$$

и

$$\tilde{f}(x^{(0)}, \ldots, x^{(n)}) = f(z) + (x^{(0)} - z)\,\tilde{h}(x^{(1)} - z, \ldots, x^{(n)} - z),$$

то

$$\begin{aligned}\tilde{f}(x, \ldots, x) &= f(z) + (x - z)\,\tilde{h}(x - z, \ldots, x - z)\\ &= f(z) + (x - z)\,h(x - z) = f(x).\end{aligned}$$

Теперь можно получить интервальную оценку для f, записанной в центрированной форме:

$$f(X) = W(\tilde{f}, X, \ldots, X).$$

Этот результат позволяет переписать (4) в виде

$$q(W(f, X), W(\tilde{f}, X, \ldots, X)) \leqslant c(d(X))^2, \quad c \geqslant 0.$$

Пусть

$$\begin{aligned}W(\tilde{f}, X, \ldots, X) = [&f(z) + (x^{(0)} - z)\,\tilde{h}(x^{(1)} - z, \ldots, x^{(n)} - z),\\ &f(z) + (y^{(0)} - z)\,\tilde{h}(y^{(1)} - z, \ldots, y^{(n)} - z)],\end{aligned}$$

$$x^{(k)}, y^{(k)} \in X, \quad 0 \leqslant k \leqslant n.$$

Заметим, что

$$W(f, X) \subseteq W(\tilde{f}, X, \ldots, X).$$

Из (2.21) вытекает

$$q(W(f, X), W(\tilde{f}, X, \ldots, X)) \leqslant d(W(\tilde{f}, X, \ldots, X)) - d(W(f, X)).$$

Теперь положим

$$\min_{x \in X} |h(x - z)| = |h(w - z)|.$$

Легко убедиться в истинности соотношения

$$f(z)+(X-z)h(w-z)\subseteq f(z)+\{(x-z)h(x-z)\mid x\in X\}=W(f,\ X),$$

если проанализировать два случая, связанных со знаком $h(w-z)$.

Далее, исходя из (2.9) и (2.14), получаем

$$d(W(f,\ X))\geqslant d((X-z)h(w-z))=d(X)\,|h(w-z)|,\quad w\in X.$$

Наконец,

$$\begin{aligned}
&q(W(f,\ X),\ W(\tilde f,\ X,\ \ldots,\ X))\\
&\leqslant(y^{(0)}-z)\,\tilde h(y^{(1)}-z,\ \ldots,\ y^{(n)}-z)\\
&\quad-(x^{(0)}-z)\,\tilde h(x^{(1)}-z,\ \ldots,\ x^{(n)}-z)-d(X)\,|h(w-z)|\\
&=(y^{(0)}-z)\,\tilde h(y^{(1)}-z,\ \ldots,\ y^{(n)}-z)\\
&\quad-(y^{(0)}-z)\,\tilde h(x^{(1)}-z,\ \ldots,\ x^{(n)}-z)\\
&\quad+(y^{(0)}-z)\,\tilde h(x^{(1)}-z,\ \ldots,\ x^{(n)}-z)\\
&\quad-(x^{(0)}-z)\,h(x^{(1)}-z,\ \ldots,\ x^{(n)}-z)-d(X)\,|h(w-z)|\\
&=(y^{(0)}-z)(\tilde h(y^{(1)}-z,\ \ldots,\ y^{(n)}-z)\\
&\quad-\tilde h(x^{(1)}-z,\ \ldots,\ x^{(n)}-z))\\
&\quad+(y^{(0)}-x^{(0)})\,\tilde h(x^{(1)}-z,\ \ldots,\ x^{(n)}-z)\\
&\quad-d(X)\,|\tilde h(w-z,\ \ldots,\ w-z)|\\
&\leqslant|y^{(0)}-z\|\tilde h(y^{(1)}-z,\ \ldots,\ y^{(n)}-z)\\
&\quad-\tilde h(x^{(1)}-z,\ \ldots,\ x^{(n)}-z)|+|y^{(0)}-x^{(0)}\|\\
&\quad\times|\tilde h(x^{(1)}-z,\ \ldots,\ x^{(n)}-z)|-d(X)\,|\tilde h(w-z,\ \ldots,\ w-z)|\\
&\leqslant d(X)(|\tilde h(y^{(1)}-z,\ \ldots,\ y^{(n)}-z)-\tilde h(x^{(1)}-z,\ \ldots,\ x^{(n)}-z)|\\
&\quad+\|\tilde h(x^{(1)}-z,\ \ldots,\ x^{(n)}-z)|-|\tilde h(w-z,\ \ldots,\ w-z)\|)\\
&\leqslant d(X)\Big(c^{(1)}\max_{1\leqslant k\leqslant n}|y^{(k)}-x^{(k)}|+c^{(2)}\max_{1\leqslant k\leqslant n}|x^{(k)}-w|\Big)\\
&\leqslant d(X)(c^{(1)}+c^{(2)})\,d(X)=c(d(X))^2.
\end{aligned}$$

Выполнение использованного в этом построении условия Липшица было оговорено для $\tilde h$ и как следствие для $[\tilde h]$.

Утверждение теоремы 4 также можно распространить на случай функции нескольких переменных. Обобщенное таким образом соотношение (4) было дано Хансеном в [114]; его же, но с применением другой техники получили Шуба и Миллер [73].

Как следствие из теоремы 3 возникает

Теорема 5. *Пусть* f *— вещественная функция от вещественного аргумента* x, $f(x)$ *— аналитическое выражение для* f. *Будем считать, что выполнены все предположения теоремы* 3. *Тогда для* $X \subseteq Y$ *имеет место неравенство*

$$d(f(X)) \leqslant c\,d(X), \qquad c \geqslant 0. \tag{5}$$

Доказательство. Опираясь на теорему 3 и соотношение (2.21), получаем

$$\begin{aligned} d(f(X)) &\leqslant 2q(f(X),\ W(f,\ X)) + d(W(f,\ X)) \\ &\leqslant 2c^{(1)}d(X) + d(W(f,\ X)), \qquad c^{(1)} \geqslant 0. \end{aligned}$$

Исходя из условия Липшица для функции f, можно записать неравенство

$$d(W(f,\ X)) = f(x) - f(y) \leqslant c^{(2)}\,|\,x - y\,|, \quad \text{где} \quad x,\ y \in X, \quad c^{(2)} \geqslant 0,$$

из которого следует

$$d(f(X)) \leqslant 2c^{(1)}d(X) + c^{(2)}d(X) = cd(X),$$

что и требовалось доказать. ■

Соответствующее обобщение на случай нескольких переменных выглядит так:

$$d(f(X^{(1)},\ X^{(2)},\ \ldots,\ X^{(n)})) \leqslant \sum_{k=1}^{n} c^{(k)}d(X^{(k)}) \leqslant c \max_{1 \leqslant k \leqslant n} d(X^{(k)}). \tag{5'}$$

Теперь мы хотим доказать теорему о вхождении множества $W(f, X)$ в другое множество, появляющееся в результате вычисления интервального выражения на основе теоремы о среднем значении.

Теорема 6. *Пусть* f *— вещественная функция от вещественного аргумента* x, *дифференцируемая на интервале* $X = [x_1, x_2]$, *и пусть* $f'(x)$ *— аналитическое выражение для* f' *такое, что интервальное выражение для* $f'(X)$ *определено. Тогда если для* f' *справедливы предположения теоремы* 5, *то*

$$\text{(а)} \qquad W(f,\ X) \subseteq f(y) + f'(X)(X - y),$$

$$\text{(б)} \qquad q(W(f,\ X),\ f(y) + f'(X)(X - y)) \leqslant \tilde{c}\,(d(X))^2,$$

где $y \in X$ *и константа* $\tilde{c} \geqslant 0$.

Доказательство. (а) Из теоремы о среднем, примененной к x и y из X, получаем

$$f(x) = f(y) + f'(y + \theta(x - y))(x - y), \qquad 0 < \theta < 1.$$

Из

$$y + \theta(x - y) \in y + [0,\ 1](X - y) = X$$

с учетом монотонности включения вытекает

$$f(x) \in f(y) + f'(X)(X - y),$$

что доказывает утверждение (а).

(б) Пусть

$$W(f, X) = [f(u), f(v)], \qquad u, v \in X.$$

Тогда из теоремы о среднем следует, что

$$d(W(f, X)) = f(v) - f(u) = |f(v) - f(u)|$$
$$\geqslant |f(x_1) - f(x_2)| = |f'(\xi)| d(X), \qquad \xi \in X.$$

Формулы (2.12), (2.3) и (2.20) дают

$$d(f'(X)(X - y)) \leqslant |f'(X)| d(X) + d(f'(X))|X - y|$$
$$\leqslant |f'(X)| d(X) + d(f'(X)) d(X).$$

Так как $f'(\xi) \in f'(X)$, то, принимая во внимание (2.21), получаем

$$q(f'(X), f'(\xi)) \leqslant d(f'(X)).$$

Теперь используем неравенство

$$|f'(X)| - |f'(\xi)| \leqslant q(f'(X), f'(\xi)),$$

которое следует из (2.4), (2.5) и определения 2.6. Применяя к $f'(X)$ соотношения (а), (2.21) и теорему 5, получаем требуемый результат:

$$q(W(f, X), f(y) + f'(X)(X - y))$$
$$\leqslant d(f(y) + f'(X)(X - y)) - d(W(f, X))$$
$$\leqslant d(f'(X)) d(X) + (|f'(X)| - |f'(\xi)|) d(X)$$
$$\leqslant d(f'(X)) d(X) + q(f'(X), f'(\xi)) d(X)$$
$$\leqslant 2c(d(X))^2 = \bar{c}(d(X))^2. \blacksquare$$

Из теоремы 6 вытекает качественный результат теоремы 4 для записанного в центрированной форме выражения

$$f(y) + f'(X)(X - y), \qquad y \in X.$$

Это важный факт, поскольку уже для многочленов получение центрированной формы требует применения полной схемы Горнера. Теорема 6 также может быть обобщена на случай нескольких переменных. Детали этого мы опускаем.

Рассмотрим теперь рациональную функцию $f(x) = p(x)/q(x)$, где $p(x) = \sum_{\nu=0}^{r} a_\nu x^\nu$ и $q(x) = \sum_{\nu=0}^{s} b_\nu x^\nu$. Связав $p(x)$ и $q(x)$ неко-

торыми условиями, можно указать выражения, для которых сохраняет силу свойство

$$q(W(f, X), f(X)) \leqslant cd(X)^2, \qquad c \geqslant 0.$$

и более простые, нежели центрированная форма или использованное в теореме 6 представление, основанное на теореме о среднем.

Пусть даны $c = m(x)$— середина интервала и тейлоровские разложения $p(x) = \sum_{\nu=0}^{r} a'_\nu (x-c)^\nu$ и $q(x) = \sum_{\nu=0}^{s} b'_\nu (x-c)^\nu$. Без потери общности допустим, что $b'_0 = 1$ и $0 \notin q(X)$, где $q(X) = 1 + \sum_{\nu=1}^{s} b'_\nu (X-c)^\nu$. Если теперь

$$\operatorname{sign}(a'_1)\operatorname{sign}(b'_1 \cdot a'_0) \leqslant 0 \tag{7}$$

и предполагается, что $p(x)$ и $q(x)$ удовлетворяют неравенствам

$$d(p(X)) \leqslant c_1 d(X),$$

$$d(q(X)) \leqslant c_2 d(X),$$

то для интервального выражения

$$f(X) = \sum_{\nu=0}^{r} a'_\nu (X-c)^\nu \Big/ \left(1 + \sum_{\nu=1}^{s} b'_\nu (X-c)^\nu\right)$$

выполнено свойство (6). Это утверждение справедливо для обоих вышеприведенных выражений независимо от того, вычисляются они с помощью степеней $X-c$ либо по схеме Горнера.

Если мы по-прежнему находимся в условиях предположения (7) и

$$0 \notin 1 + (X-c)q'(X),$$

то в (6) можно подставить выражение

$$f(X) = \frac{a'_0 + (X-c)p'(X)}{1 + (X-c)q'(X)}.$$

Сомножитель $p'(X)$ представляет собой интервальную оценку для первой производной функции $p(x)$, удовлетворяющую соотношению $d(p'(X)) \leqslant \alpha d(X)$. Аналогично, $q'(X)$— интервальная оценка для первой производной $q(x)$, удовлетворяющая неравенству $d(q'(X)) \leqslant \beta d(X)$.

Доказательство свойства (6) применительно к двум рассмотренным выше случаям можно найти у Алефельда и Рокна [350]; здесь мы его повторять не будем.

В гл. 7 мы рассматриваем методы локализации нулей, использующие включения на участках монотонности функции.

Ниже мы дадим ряд возможных включений, применимых к отношению разностей. Эти включения будут частично упорядочены. Оказывается, что оптимальное включение может быть описано просто и систематично и что вычисления с соответствующими итерациями могут быть выполнены с теми же вычислительными затратами, что и интервальное оценивание производной. Эти включения выводятся другими способами, значительно отличающимися от использовавшихся у Хансена [395], решавшего ту же задачу.

Включения для примеров, приведенных Хансеном в [395], в точности соответствуют оптимальным включениям для этих же примеров, полученных нашими методами. Далее в своих рассуждениях мы следуем работе Алефельда [349a].

Пусть имеется многочлен

$$p(x) = \sum_{\nu=0}^{n} a_\nu x^\nu.$$

Справедливость двух нижеследующих равенств очевидна:

$$(8)\qquad p(x) - p(y) = \sum_{i=0}^{n} a_i (x^i - y^i) = \left(\sum_{i=1}^{n} a_i \sum_{j=1}^{i} x^{i-j} y^{j-1} \right)(x - y)$$

$$= \left(\sum_{i=1}^{n} \left(\sum_{j=i}^{n} a_j y^{j-i} \right) x^{i-1} \right)(x - y),$$

$$(9)\qquad p(x) - p(y) = \sum_{i=0}^{n} a_i (x^i - y^i) = \left(\sum_{i=1}^{n} a_i \sum_{j=1}^{i} y^{i-j} x^{j-1} \right)(x - y)$$

$$= \left(\sum_{i=1}^{n} \left(\sum_{j=i}^{n} a_j x^{j-i} \right) y^{i-1} \right)(x - y).$$

Для фиксированного y и произвольного x из X с помощью (8) и свойства монотонности включения получаем [1])

$$\frac{p(x) - p(y)}{x - y} \in \left(\sum_{i=1}^{n} c_{i-1} X^{i-1} \right)_{\mathrm{H}} =: J_1 \subseteq J_2 := \sum_{i=1}^{n} c_{i-1} X^{i-1},$$

где

$$c_{i-1} = \sum_{j=i}^{n} a_j y^{j-i}, \qquad 1 \leqslant i \leqslant n.$$

Здесь и далее буква H обозначает, что выражение, которое ею помечено, вычисляется по схеме Горнера. В многочлене J_2 степени X^r вычисляются по правилу $X^0 := 1$, $X^r := X^{r-1} X$ при $r \geqslant 1$.

[1]) Далее в тексте знаки: := и =: авторы используют как для обозначения выражения, стоящего со стороны, противоположной двоеточию, так и в качестве знаков присваивания в программистском смысле. — *Прим. перев.*

Включение $J_1 \subseteq J_2$ следует из закона субдистрибутивности. Для вещественного числа y и интервалов A_j, $0 \leqslant j \leqslant n-1$, всегда

$$\sum_{i=1}^{n} A_{i-1} y^{i-1} = \left(\sum_{i=1}^{n} A_{i-1} y^{i-1} \right)_{\text{н}}.$$

При фиксированном y и произвольном x из X, $x \neq y$, используя субдистрибутивность, данное равенство и формулу (9), получаем

$$\frac{p(x) - p(y)}{x - y} \in \sum_{i=1}^{n} (C_{i-1})_{\text{н}} y^{i-1} = \left(\sum_{i=1}^{n} (C_{i-1})_{\text{н}} y^{i-1} \right)_{\text{н}} =: J_3$$

$$\subseteq J_4 := \sum_{i=1}^{n} C_{i-1} y^{i-1} = \left(\sum_{i=1}^{n} C_{i-1} y^{i-1} \right)_{\text{н}},$$

где

$$(C_{i-1})_{\text{н}} = \left(\sum_{j=i}^{n} a_j X^{j-i} \right)_{\text{н}}, \qquad 1 \leqslant i \leqslant n,$$

и

$$C_{i-1} = \sum_{j=i}^{n} a_j X^{j-i}, \qquad 1 \leqslant i \leqslant n.$$

Докажем теперь еще одну теорему.

Теорема 7. *Введенные выше выражения удовлетворяют соотношениям*

(a) $J_1 \subseteq J_2 \subseteq J_4$,

(b) $J_1 \subseteq J_3 \subseteq J_4$,

(c) $J_4 \subseteq p'(X) = \sum_{\nu=1}^{n} \nu a_\nu X^{\nu-1}$.

Доказательство. Для простоты ограничимся случаем многочлена четвертой степени ($n = 4$). Общий случай может быть рассмотрен аналогичным образом.

(a) и (c) Нам достаточно показать, что $J_2 \subseteq J_4 \subseteq p'(X)$. Из свойства монотонности включения и соотношения (1.8) получаем

$$J_2 = \sum_{i=1}^{n} c_{i-1} X^{i-1}$$

$$= (a_1 + a_2 y + a_3 y^2 + a_4 y^3) X^0 + (a_2 + a_3 y + a_4 y^2) X$$
$$+ (a_3 + a_4 y) X^2 + a_4 X^3$$

$$\subseteq a_1 + a_2 X + a_3 X^2 + a_4 X^3 + a_2 y + a_3 y X + a_4 y X^2$$
$$+ a_3 y^2 + a_4 y^2 X + a_4 y^3$$

$$= a_1 + a_2X + a_3X^2 + a_4X^3 + (a_2 + a_3X + a_4X^2)\,y$$
$$+ (a_3 + a_4X)\,y^2 + a_4y^3 = J_4$$
$$\subseteq a_1 + a_2X + a_3X^2 + a_4X^3 + a_2X + a_3X^2 + a_4X^3$$
$$+ a_3X^2 + a_4X^3 + a_4X^3 = p'(X).$$

(b) Достаточно показать, что $J_1 \subseteq J_3$.

$$J_1 = ((c_3X + c_2)X + c_1)X + c_0$$
$$= ((a_4X + (a_3 + a_4y))X + a_2 + a_3y + a_4y^2)X + a_1 + a_2y + a_3y^2 + a_4y^3$$
$$\subseteq ((a_4X + a_3)X + a_4yX + a_2 + a_3y + a_4y^2)X + a_1 + a_2y + a_3y^2 + a_4y^3$$
$$= (((a_4X + a_3)X + a_2) + a_4yX + a_3y + a_4y^2)X + a_1 + a_2y + a_3y^2 + a_4y^3$$
$$= (((a_4X + a_3)X + a_2) + (a_4X + a_3)\,y + a_4y^2)X + a_1 + a_2y + a_3y^2 + a_4y^3$$
$$\subseteq ((a_4X + a_3)X + a_2)X + (a_4X + a_3)\,yX + a_4y^2X + a_1 + a_2y + a_3y^2 + a_4y^3$$
$$= (((a_4X + a_3)X + a_2)X + a_1)\,y^0 + ((a_4X + a_3)X + a_2)\,y$$
$$+ (a_4X + a_3)\,y^2 + a_4y^3 = J_3.$$

Итак, теорема доказана. ■

Нельзя ответить в общем случае на вопрос о том, какое из выражений: J_2 или J_3 — дает лучшее включение. Возможно и $J_2 \subseteq J_3$, и $J_3 \subseteq J_2$. Пусть, например,

$$p(x) = x^3 - x^2, \quad X = [-1, 2], \quad y = 1.$$

Тогда имеем

$$J_2 = (a_1 + a_2y + a_3y^2)\,X^0 + (a_2 + a_3y)\,X + a_3X^2 = X^2 = [-2, 4]$$

и

$$J_3 = ((a_3X + a_2)X + a_1)\,y^0 + (a_3X + a_2)\,y + a_3y^2$$
$$= (X - 1)X + (X - 1) + 1 = [-5, 4].$$

Здесь $J_2 \subset J_3$.

Если, с другой стороны, $y = 0$ и, следовательно, $c_{i-1} = a_i$, $1 \leqslant i \leqslant n$, то получаем

$$J_2 = \sum_{i=1}^{n} a_iX^{i-1} \quad \text{и} \quad J_3 = \left(\sum_{i=1}^{n} a_iX^{i-1}\right)_{\text{н}}.$$

Теперь $J_3 \subseteq J_2$.

Рассмотрим снова пример с $p(x) = x^3 - x^2$ при $y = 0$ и $X = [0, 2]$. В этом случае

$$J_2 = X^2 - X = [-2, 4] \quad \text{и} \quad J_3 = (X - 1)X = [-2, 2],$$

откуда имеем $J_3 \subset J_2$.

Получение интервалов J_1 и J_2 с помощью теоремы 7 требует предварительного нахождения $c_{i-1} = \sum_{j=i}^{n} a_j y^{j-i}$, $1 \leqslant i \leqslant n$. Если вычисляется значение многочлена $p(x)$ в точке y, что встречается, например, в итерационных методах гл. 7, то нахождение c_{i-1} не требует выполнения каких-либо дополнительных арифметических операций. Значения c_{i-1} могут быть найдены в процессе вычисления $p(y)$. Пусть, как и ранее,

$$p(x) = \sum_{i=0}^{n} a_i x_i.$$

Воспользуемся схемой Горнера

$$p_n := a_n,$$

и для $i = n, n-1, \ldots, 1$

$$p_{i-1} := p_i y + a_{i-1},$$

откуда $p_0 = p(y)$. По определению

$$\begin{aligned} c_{n-1} &= a_n && (= p_n), \\ c_{n-2} &= a_n y + a_{n-1} && (= p_{n-1}), \\ &\vdots && \vdots \\ c_0 &= c_1 y + a_1 && (= p_1). \end{aligned}$$

Следовательно, $c_{i-1} = p_i$, $1 \leqslant i \leqslant n$.

Примеры. Первые пять примеров взяты из работы Алефельда [349a].

(a) $p(x) = x^4 - 1, \quad x = [0.5, 3.5], \quad y = 2.$

Имеем

$$J_1 = J_2 = J_3 = J_4 = [10.625, 89.375],$$
$$p'(X) = (p'(X))_{\text{н}} = [0.5, 171.5].$$

Оценка J_1 совпадает с той, которую получил на этих же данных Хансен [395].

(b) $p(x) = x^3 + 4x - 16, \quad X = [-1, 3], \quad y = 1.$

Получаем

$$J_1 = J_2 = J_3 = J_4 = [1, 17],$$
$$p'(X) = (p'(X))_{\text{н}} = [-5, 31],$$

что совпадает с результатом Хансена [395].

(c) $$p(x) = \sum_{i=0}^{n} a_i x^i, \quad 0 \in X, \quad y = 0.$$

В этом случае

$$c_0 = a_1, \quad c_1 = a_2, \ \ldots, \ c_{n-1} = a_n$$

и

$$J_1 = \left(\sum_{i=1}^{n} c_{i-1} X^{i-1}\right)_{\mathrm{H}} = \left(\sum_{i=1}^{n} a_i X^{i-1}\right)_{\mathrm{H}}.$$

(d) $$p(x) = x^3 - x^2, \quad X = [1, 3], \quad y = 2.$$

$$J_1 = J_2 = J_3 = [4, 14] \subset [2, 16] = J_4 \subset (p'(X))_{\mathrm{H}}$$

$$= [1, 21] \subset [-3, 25] = p'(X),$$

что опять совпадает со значением, вычисленным Хансеном [395]. Однако при $X = [-1, 2]$ и $y = 1$

$$J_1 = J_2 = [-2, 4], \quad J_3 = [-5, 4], \quad J_4 = [-5, 7],$$

$$(p'(X))_{\mathrm{H}} = [-10, 8], \quad p'(X) = [-10, 14].$$

(e) Пусть $x_0 \in X$ и $f \in C^{n+1}(X)$. Используя тейлоровское разложение, получаем

$$f(x) = p(x) + \varphi(x),$$

где

$$\varphi(x) = \int_{x_0}^{x} \frac{(x-t)^n}{n!} f^{(n+1)}(t)\, dt$$

и

$$p(x) = \sum_{k=0}^{n} \frac{(x-x_0)^k}{k!} f^{(k)}(x_0).$$

Функция φ дифференцируема и

$$\varphi'(x) = \int_{x_0}^{x} \frac{(x-t)^{n-1}}{(n-1)!} f^{(n+1)}(t)\, dt.$$

Интегральная теорема о среднем дает формулу

$$\varphi'(x) = f^{(n+1)}(\eta) \int_{x_0}^{x} \frac{(x-t)^{n-1}}{(n-1)!}\, dt = \frac{(x-x_0)^n}{n!} f^{(n+1)}(\eta),$$

где η лежит между x и x_0. Применяя к φ теорему о среднем, получаем

$$f(x) - f(y) = p(x) - p(y) + \varphi(x) - \varphi(y)$$

$$= \left\{\sum_{k=1}^{n} c_{k-1}(x-x_0)^{k-1} + \varphi'(\xi)\right\}(x-y),$$

где

$$c_{k-1} = \sum_{l=k}^{n} (y - x_0)^{l-k} \frac{f^{(k)}(x_0)}{k!}, \quad 1 \leqslant k \leqslant n,$$

и

$$\varphi'(\xi) = \frac{(\xi - x_0)^n}{n!} f^{(n+1)}(\eta),$$

причем ξ лежит между x и y, а η — между x_0 и ξ. Полагая $y = x_0$, имеем

$$c_0 = f'(x_0)/1!, \ \ldots, \ c_{n-1} = f^{(n)}(x_0)/n!.$$

Если для $(n+1)$-й производной существует вычислимое интервальное выражение, то для $y = x_0$

$$\frac{f(x) - f(y)}{x - y} \in \sum_{k=1}^{n} \frac{f^{(k)}(x_0)}{k!} (X - x_0)^{k-1} + f^{(n+1)}(X) \frac{(X - x_0)^n}{n!},$$

поскольку $\eta, \xi \in X$. Эта оценка снова совпадает с данной Хансеном [395].

(f) $p(x) = x^7 + 3x^6 - 4x^5 - 12x^4 - x^3 - 3x^2 + 4x + 12,$

$$X = [1.8, 3], \quad y = 2.$$

Теперь

$$J_1 = [173.2362, 2400], \qquad J_2 = [161.4762, 2411.76],$$

$$J_3 = [24.72, 2400], \qquad J_4 = [-870.2933, 3443.5296],$$

$$(p'(X))_{\text{H}} = [71.799808, 6520], \qquad p'(X) = [-2378.791292, 8970.592].$$

Сделанные утверждения могут быть распространены на многомерный случай (ср.: Алефельд [349а]).

Замечания. В этой главе было рассмотрено интервальное оценивание вещественных функций. Мы сознательно не говорили о произвольных отображениях из $I(\mathbb{R})$ в $I(\mathbb{R})$. Приложения в последующих главах требуют использования многих свойств, которые мы в силах доказать только для интервальных оценок. Если разрешить на $I(\mathbb{R})$ отображения более общего вида, то для каждого приложения необходимо описать множество условий. Следующий пример показывает, сколь велик класс отображений из $I(\mathbb{R})$ в $I(\mathbb{R})$. Единственное ограничение здесь состоит в том, что если область определения сузить до $\mathbb{R}$, то множество значений также будет принадлежать $\mathbb{R}$. Итак, пусть f — вещественная функция, $f(x)$ — ее аналитическое выражение, $f(X)$ — оценивающая функция для $f(x)$. Тогда для произвольной $\varphi(x)$ такой, что $\varphi(0) = 0$,

$$\Psi(X) = f(X) + \varphi(d(X))[-1, 1]$$

определяет отображение из $I(\mathbb{R})$ в $I(\mathbb{R})$. Очевидно, что $\Psi([x, x]) \in \mathbb{R}$. Если $\varphi(x) \geqslant 0$ при $x \geqslant 0$, то $W(f, X) \subseteq \Psi(X)$, и если $\varphi(x)$ монотонна и не убывает при $x \geqslant 0$, то $\Psi(X)$ обладает свойством монотонности включения в форме (1′). Этот пример показывает, что, соответствующим образом выбирая $\varphi(x)$, можно строить отображения Ψ, имеющие различные свойства. Если же потребовать от отображений из $I(\mathbb{R})$ в $I(\mathbb{R})$ выполнения всех свойств интервальных оценок, то не найдется никаких других полезных нам отображений, кроме в точности этих оценок.

Мы хотим также показать, как можно сократить доказательство теоремы 4, используя теорему 5 примерно так же, как и при доказательстве теоремы 6. Записанная в центрированной форме интервальная оценка

$$f(X) = f(z) + (X - z)\, h(X - z)$$

удовлетворяет соотношению

$$W(f, X) \subseteq f(X).$$

В соответствии с (2.21) получаем

$$q(W(f, X),\ f(X)) \leqslant d(f(X)) - d(W(f, X)).$$

Пусть теперь

$$\min_{x \in X} |h(x - z)| = |h(w - z)|.$$

Тогда

$$f(z) + (X - z)\, h(w - z) \subseteq f(z) + \{(x - z)\, h(x - z) \mid x \in X\} = W(f, X).$$

С учетом (2.9), (2.14) и приведенного выше включения имеем

$$d(W(f, X)) \geqslant d((X - z)\, h(w - z)) = d(X)\, |h(w - z)|, \quad w \in X.$$

Исходя из (2.10), (2.12), (2.3) и (2.20), получаем

$$\begin{aligned} d(f(X)) &= d(f(z) + (X - z)\, h(X - z)) = d((X - z)\, h(X - z)) \\ &\leqslant |X - z|\, d(h(X - z)) + d(X)\, |h(X - z)| \\ &\leqslant d(X)\, d(h(X - z)) + d(X)\, |h(X - z)|. \end{aligned}$$

Поскольку $h(w - z) \in h(X - z)$, из (2.21) следует, что

$$q(h(X - z),\ h(w - z)) \leqslant d(h(X - z)).$$

Определение 2.6 и соотношения (2.4), (2.5) дают неравенство

$$|h(X - z)| - |h(w - z)| \leqslant q(h(X - z),\ h(w - z)).$$

На основе приведенных выше неравенств получаем

$$\begin{aligned} q(W(f, X), f(X)) &\leqslant d(f(z)+(X-z)h(X-z))-d(W(f, X)) \\ &\leqslant d(X)d(h(X-z))+d(X)\,|h(X-z)|-d(X)\,|h(w-z)| \\ &= d(X)d(h(X-z))+(|h(X-z)|-|h(w-z)|)d(X) \\ &\leqslant d(X)d(h(X-z))+q(h(X-z),\ h(w-z))d(X) \\ &\leqslant d(X)\cdot 2\cdot d(h(X-z)). \end{aligned}$$

И наконец, после применения теоремы 5 к выражению $h(X-z)$ оказывается, что

$$q(W(f, X), f(X)) \leqslant d(X)\cdot 2\cdot \tilde{c}\cdot d(X) = c(d(X))^2.$$

При выполнении некоторых условий дифференцируемости для аналитических выражений можно определить общие условия, *при которых* соотношение (4) справедливо (см. статью Херцбергера [405]). При этом теоремы 4 и 6 получаются как частные случаи. У Херцбергера в [406] можно найти простую интерполяционную формулу, вычисляющую множество значений для семейства многочленов с заданными коэффициентами. В этой работе используется тот вытекающий из теоремы 2 факт, что интервальное оценивание позволяет точно вычислить множество значений функции, когда все переменные и параметры входят в аналитическое выражение лишь по одному разу.

В готовящейся к публикации книге Ратшека и Рокна [1]) *подробно* обсуждают задачу аппроксимации множества значений функции с помощью интервальных оценок. Корнелиус и Лонер [384a] предложили выражение $f(X)$, для которого $q(W(f, X), f(X)) \leqslant cd(X)^{s+1}$, $s \geqslant 1$. При $s = 1, 2, 3$ $f(X)$ вычисляется с помощью весьма простого алгоритма. Детали см. в книге Ратшека и Рокна [1]).

* * *

Литература: [33], [34], [73], [97], [114], [122], [123], [157], [195], [197], [207], [208], [220], [257], [258], [262], [349a], [350], [384a], [395], [405], [406], [481].

Глава 4
МАШИННАЯ ИНТЕРВАЛЬНАЯ АРИФМЕТИКА

Теперь мы остановимся на вопросах реализации интервальных операций на цифровой вычислительной машине. Общеиз-

[1]) В оригинале здесь стоит ссылка на книгу, озаглавленную «Centered forms», *но, по-видимому, она вышла под другим названием: см. ссылку* [17] — *Прим. перев.*

вестно, что в машине может быть представлено лишь конечное множество чисел. Чаще всего они записываются в полулогарифмической форме, а точнее — в форме с плавающей точкой:

$$x = m \cdot b^e.$$

Здесь m — мантисса, b — основание степени [1]), e — порядок. Как правило, для внутримашинного представления выбирается основание b, равное 2, а мантисса нормализуется, т. е. ее абсолютное значение помещается в интервал $[1/2, 1)$ [2]). Целое e принадлежит интервалу $[e_{\min}, e_{\max}]$.

Множество машинных чисел описанного типа обозначим через $\mathbb{R}_M$, и всюду далее будем предполагать, что оно симметрично относительно нуля, т. е. $\mathbb{R}_M = -\mathbb{R}_M$. Для аппроксимации вещественных чисел, лежащих в интервале $[\min\limits_{y \in \mathbb{R}_M} y;\ \max\limits_{y \in \mathbb{R}_M} y]$, можно с успехом использовать машинные числа $\{\tilde{x} \mid \tilde{x} \in \mathbb{R}_M\}$. Аппроксимация достигается применением отображения

$$(1) \qquad fl\colon \mathbb{R} \ni x \to \tilde{x} = fl(x) \in \mathbb{R}_M.$$

Это отображение называется округлением, если выполнено свойство

$$(2) \qquad x \leqslant y \Rightarrow fl(x) \leqslant fl(y) \quad \text{(монотонность)}.$$

Округление, которое отображает $\mathbb{R}_M$ в $\mathbb{R}_M$ так, что

$$(3) \qquad x \in \mathbb{R}_M \Rightarrow fl(x) = x,$$

называется оптимальным [3]).

Особый интерес представляют так называемые направленные округления. Если для округления $\downarrow$ справедлива импликация

$$(4) \qquad x \in \mathbb{R} \Rightarrow \downarrow x \leqslant x,$$

то говорят об округлении вниз. Аналогично,

$$(5) \qquad \uparrow x := -(\downarrow(-x)), \qquad x \leqslant \mathbb{R},$$

---

[1]) Далее в тексте делается достаточно реалистичное, хотя и не оговоренное явно предположение, что основание степени b равно основанию системы счисления, в которой записана мантисса m. — *Прим. перев.*

[2]) Вообще же говоря, *нормализованным* называется такое число с плавающей точкой, у которого старшая цифра мантиссы отлична от нуля. Интервал, в который попадает при этом $|m|$, зависит от того, между какими разрядами m подразумевается точка. Поэтому приведенный здесь интервал отражает лишь частный случай. — *Прим. перев.*

[3]) Приведенное определение не является стандартным. Обычно под оптимальным понимают такое округление, которое отображает округляемое число x в $\tilde{x}$, ближайшее в некотором смысле к x. — *Прим. перев.*

определяет округление вверх [1]). Техника выполнения этих округлений для различных способов представления чисел неоднократно освещалась в литературе (см., например, работы Кнута [155], Купки [185], Уилкинсона [329], Миранкера и Кулиша [443], Мура [447]). Алгебраические свойства округлений и свойства, связанные с порядком, описаны, в частности, Кулишем [178, 180, 181]. Подобно тому как вещественные числа приближаются с помощью машинных, можно вещественные интервалы приближать машинными интервалами. В этом случае интервал X из $I(\mathbb{R})$, для которого справедливо соотношение $X \subseteq [\min_{y \in \mathbb{R}_M} y, \max_{y \in \mathbb{R}_M} y]$, заменяется соответствующим машинным интервалом из множества

$$I(\mathbb{R}_M) = \{[x_1, x_2] \mid x_1, x_2 \in \mathbb{R}_M, x_1 \leqslant x_2\} \subset I(\mathbb{R}).$$

Для того чтобы основные свойства интервальных операций выполнялись и для их машинных аналогов, применяется округление интервалов (интервальное округление)

$$\updownarrow : I(\mathbb{R}) \ni X \to \updownarrow X \in I(\mathbb{R}_M),$$

причем

$$X \in I(\mathbb{R}) \Rightarrow X \subseteq \updownarrow X \tag{6}$$

и

$$X, Y \in I(\mathbb{R}), \quad X \subseteq Y \Rightarrow \updownarrow X \subseteq \updownarrow Y. \tag{7}$$

Если рассмотреть переход от интервала $X = [x_1, x_2]$ из $I(\mathbb{R})$ к его машинному представлению $\check{X} = [\check{x}_1, \check{x}_2]$, то окажется, что (7) означает, что необходимо осуществить этот переход путем округления каждой из границ X. Из (6) следует, что границы должны быть округлены направленно. Таким образом, округление интервала X состоит в нахождении $\updownarrow X$ по правилу

$$\updownarrow X = \updownarrow [x_1, x_2] = [\downarrow x_1, \uparrow x_2]. \tag{8}$$

Проведенное обсуждение показывает, что для того, чтобы округлить интервал, достаточно иметь $\downarrow$ — направленное округление вниз. С другой стороны, $\uparrow$ и $\downarrow$ не обязательно должны быть связаны соотношением (5).

Если над двумя машинными числами x и y из $\mathbb{R}_M$ производится машинная операция $*$, где $* \in \{+, -, \cdot, :\}$, то ее результатом оказывается новое число z из $\mathbb{R}_M$. Проигнорировав возможность выхода за пределы допустимого диапазона (перепол-

[1]) Возможно, удобнее было бы, как это и делают некоторые авторы, говорить не о направлении вниз и вверх, а о направлении влево и вправо, поскольку вещественную ось привычно считать горизонтальной, а не вертикальной. — *Прим. перев.*

нение и антипереполнение[1])), можно, используя соответствующее округление fl, представить z в виде

$$z = fl(x * y). \tag{9}$$

Таким же образом мы можем определить результат машинной операции над интервалами.

Определение 1. Пусть $A, B \in I(\mathbb{R}_M)$, $* \in \{+, -, \cdot, :\}$, $\updownarrow$ — интервальное округление. Тогда результат операции $*$, выполненной над A и B с применением $\updownarrow$ есть

$$C = \updownarrow(A * B) \in I(\mathbb{R}_M). \blacksquare \tag{10}$$

Теперь мы покажем, что основные свойства интервальной арифметики при использовании этого определения сохраняются.

Теорема 2. *Для машинных интервальных операций, задаваемых определением* 1, *справедливо следующее утверждение:*

$$A^{(k)}, B^{(k)} \in I(\mathbb{R}_M),\ * \in \{+, -, \cdot, :\},\ A^{(k)} \subseteq B^{(k)},\ k = 1, 2,$$
$$\Rightarrow C^{(1)} = \updownarrow(A^{(1)} * A^{(2)}) \subseteq C^{(2)} = \updownarrow(B^{(1)} * B^{(2)}). \blacksquare \tag{11}$$

Доказательство теоремы 2 следует непосредственно из свойства интервальных округлений (7).

Утверждение (11) отражает не что иное, как свойство монотонности включения (1.9) применительно к машинным интервальным операциям.

Очередная теорема представляет интерес с точки зрения оценки погрешностей округлений.

Теорема 3. *Пусть* $\updownarrow$ — *интервальное округление, сводящееся с помощью* (8) *к направленным округлениям* $\downarrow$ *и* $\uparrow$, *и пусть* $* \in \{+, -, \cdot, :\}$. *Тогда*

$$\begin{aligned} &A, B \in I(\mathbb{R}_M) \Rightarrow A * B \subseteq C = \updownarrow(A * B) \in I(\mathbb{R}_M), \\ &a \in A, b \in B \Rightarrow a * b \in C = \updownarrow(A * B) \in I(\mathbb{R}_M). \end{aligned} \tag{12}$$

Если имеется округление fl, *применение которого приводит к выполнению неравенства*

$$\downarrow a \leqslant fl(a) \leqslant \uparrow a, \quad a \in \mathbb{R},$$

то для x, y, z *из* $\mathbb{R}_M$ *справедливо*

$$z = fl(x * y) \in Z = \updownarrow([x, x] * [y, y]) \in I(\mathbb{R}_M). \blacksquare \tag{13}$$

---

[1]) Если допустимым считать диапазон $[\min\limits_{x \in \mathbb{R}_M} x,\ \max\limits_{x \in \mathbb{R}_M} x]$, то попадание z в один из интервалов $(\max\limits_{\substack{x \in \mathbb{R}_M \\ x < 0}} x, 0)$ или $(0, \min\limits_{\substack{x \in \mathbb{R}_M \\ x > 0}} x)$ (а это и называется антипереполнением) не приводит к выходу за пределы допустимого диапазона, хотя может повлечь за собой другие неприятные последствия. — *Прим. перев.*

Доказательство свойств (12) и (13) мы опускаем, поскольку оно элементарно и следует непосредственно из соответствующих определений.

Интервальное оценивание аналитического выражения функции, проведенное с использованием операций из определения 1, дает интервалы, объемлющие значения оценивающей функции. Среди этих интервалов находятся и оценки множества значений функции. Более того, при выполнении подобных вычислений сохраняется свойство монотонности включения.

На практике машинные интервальные операции реализуются с помощью соответствующих программно-аппаратных средств. Эти средства могут служить поддержкой языка программирования высокого уровня (см., например, публикацию Апостолатоса и др. [38], а также спецвыпуск журнала «SIGNUM Newsletter» [518]). Один из вариантов реализации — набор подпрограмм, написанных, скажем, на Алголе. Рассмотрим вкратце последнюю возможность. В большинстве случаев, в частности у Криста [72], такой набор содержит средство, с помощью которого выполняется округление $\downarrow$. Это средство может быть, например, оформлено в виде процедуры-функции LOW; через нее определяются стандартные операции интервальной арифметики — ADD, SUB, MUL и DIV, а также элементарные функции (см. Херцбергер [124], Дьюар [77]). На деталях реализации подобных подпрограмм мы остановимся в приложении В.

Посмотрим теперь на алгоритмы, описанные в терминах вещественных чисел. К их числу можно отнести, скажем, схему Горнера или метод Гаусса. Если такой алгоритм реализуется на компьютере, т. е. с использованием машинной арифметики, то даже исходные данные в общем случае не могут быть представлены точно. Возникающие при этом трудности преодолеваются применением машинной интервальной арифметики. Исходные данные просто заключаются в интервалы, имеющие своими границами машинные числа. Если теперь представить себе, что алгоритм будет выполняться без учета погрешностей округления, то, как показано в гл. 3, ширина реализующего интервала станет, вообще говоря, возрастать в большей степени, чем это обусловлено исходными данными [1]). При наличии погрешностей округления описанное свойство проявляется еще сильнее.

---

[1]) Причина этого эффекта в том, что определение интервальных операций (1.2) предполагает предысторию получения операндов неизвестной. Например, согласно (1.2), $[1,3]-[1,3]=[-2,2]$, хотя, если предыстория получения операндов одна и та же, результат можно было бы принять равным нулю. Различные способы учета предыстории излагаются, в частности, в работах [7, 38, 64, 91. 92] (см. литературу к Добавлению). — *Прим. перев.*

А сейчас мы хотим обсудить следующий вопрос: на какой рост точности результата можно рассчитывать, если перейти от алгоритма, использующего машинную интервальную арифметику с t_1 цифрами в мантиссе, к алгоритму, использующему t_2-значную арифметику, где $t_2 > t_1$? Предполагается, что при таком переходе диапазон возможных значений порядка остается неизменным. Таким образом, все числа, представимые с t_1 цифрами, столь же точно записываются с t_2 цифрами.

Пусть $x \in \mathbb{R}$, $x \neq 0$ и

$$x = \left(\sum_{\nu=-1}^{-\infty} a_\nu b^\nu\right) b^e, \quad 1 \leqslant a_{-1} \leqslant b-1, \quad 0 \leqslant a_\nu \leqslant b-1, \quad \nu \leqslant -2.$$

Чтобы гарантировать единственность представления x, мы предполагаем, что не существует ν_0 такого, что при $\nu \leqslant \nu_0$ все a_ν равны $b-1$. Будем также считать, что число x не представимо в t_1-значной системе с плавающей точкой. (Если бы последнее допущение отсутствовало, то следующее рассуждение было бы совершенно излишним.) Пусть, кроме того, интервальное округление (8) осуществляется через оптимальное округление границ. В соответствии с (8) при $x > 0$ получаем

$$\updownarrow x = \updownarrow [x, x] = [\downarrow x, \uparrow x],$$

где

$$\downarrow x = \left(\sum_{\nu=-1}^{-t_1} a_\nu b^\nu\right) b^e, \quad \uparrow x = \left(\sum_{\nu=-1}^{-t_1} a_\nu b^\nu\right) b^e + b^{-t_1+e}.$$

Очевидно, что ширина $\updownarrow x$ есть

$$d(\updownarrow x) = b^{-t_1+e}.$$

Точно такой же результат получается для ширины $\updownarrow x$ при $x < 0$. В дальнейшем зависимость результата от длины мантиссы будет отражаться с помощью записи $fl_1(x)$ (соответственно $fl_2(x)$). Следовательно, под fl мы будем понимать интервальное округление вещественного числа (а впоследствии и вещественного интервала). Предыдущее равенство теперь может быть переписано в виде

$$d(fl_1(x)) = b^{-t_1+e}.$$

Аналогично, для мантиссы длины $t_2 = t_1 + l$ получаем

$$d(fl_2(x)) \leqslant b^{-t_1+e-l}.$$

Неравенство выполняется как строгое в случае, когда x представляется точно с t_2-разрядной мантиссой.

Как следствие

$$d(fl_2(x)) \leqslant b^{-l} d(fl_1(x)). \tag{14}$$

Из предположений, сформулированных для интервальных округлений, следует, что для двух машинных интервалов A и B

$$\updownarrow(A * B) = fl_1(A * B) = [(1 - \varepsilon_1)(A * B)_1,\ (1 + \varepsilon_2)(A * B)_2]$$

(см. определение 1). С помощью $(A*B)_1$ и $(A*B)_2$ вычисляются границы точного значения результата, причем

$$-\varepsilon_1(A * B)_1 \leqslant 0, \qquad \varepsilon_2(A * B)_2 \geqslant 0$$

и

$$|\varepsilon_1|,\ |\varepsilon_2| \leqslant b^{1-t_1}.$$

Следовательно, можно записать

$$fl_1(A * B) = A * B + [-\varepsilon_1(A * B)_1,\ \varepsilon_2(A * B)_2]. \tag{15a}$$

Оценкой ширины результата служит

$$d(fl_1(A * B)) \leqslant d(A * B) + 2b^{1-t_1}|A * B|. \tag{15b}$$

Эта оценка показывает, что когда используется мантисса фиксированной длины, то рост ширины $d(fl_1(A * B))$ определяется величиной $|A * B|$.

Пусть мы знаем, что $\tilde{x}$ принадлежит интервалу X из $I(\mathbb{R})$. Естественно выбрать некоторое x из X в качестве приближенного значения $\tilde{x}$. Оценим абсолютную и относительную погрешность такого приближения:

$$|x - \tilde{x}| \leqslant d(X) =: \Delta(X), \tag{16}$$

и если $0 \notin X$, $x \neq 0$, то

$$\left|\frac{x - \tilde{x}}{\tilde{x}}\right| \leqslant \frac{d(X)}{\min\{|x|\ |\ x \in X\}} =: \rho(X). \tag{17}$$

Теорема 4. *Пусть A, B, C и D — машинные интервалы, причем*

$$A \subseteq C, \qquad B \subseteq D, \tag{18}$$

и

$$\begin{aligned} d(C) &\leqslant s_1, & d(D) &\leqslant s_2, \\ d(A) &\leqslant b^{-l}s_1, & d(B) &\leqslant b^{-l}s_2. \end{aligned} \tag{19}$$

Предположим, что $*$ — одна из арифметических операций над вещественными интервалами и $0 \notin fl_1(C * D)$. Тогда границы для $\Delta(fl_2(A*B))$ (соответственно $\rho(fl_2(A*B))$) оказываются в b^l раз меньше, чем границы для $\Delta(fl_1(C * D))$ (соответственно $\rho(fl_1(C * D))$).

Доказательство. Используя (15b), (2.10), (2.12), неравенство

$$d(1/X) \leqslant |1/X|^2 d(X) \qquad (0 \notin X),$$

а также первую строку из (19), сразу же получаем

$$d(fl_1(C * D)) \leqslant d(C * D) + 2b^{1-t_1}|C * D|$$
$$\leqslant \left\{\begin{array}{ll} s_1 + s_2, & * = +, - \\ |C|s_2 + s_1|D|, & * = \cdot \\ |C||1/D|^2 s_2 + |1/D|s_1, & * = : \end{array}\right\} + 2b^{1-t_1}|C * D|.$$

На основе (18) и (19) аналогичным способом доказывается, что

$$(20)\quad d(A * B) \leqslant b^{-l}\left\{\begin{array}{ll} s_1 + s_2, & * = +, - \\ |C|s_2 + s_1|D|, & * = \cdot \\ |C||1/D|^2 s_2 + |1/D|s_1, & * = : \end{array}\right\}.$$

Из (18) и теоремы 2 следует включение

$$fl_2(A * B) \subseteq fl_2(C * D) \subseteq fl_1(C * D).$$

Оно справедливо, поскольку мы предположили, что интервальное округление задано через оптимальные округления границ. Таким образом,

$$(21)\quad \min\{|x|\,|\,x \in fl_2(A * B)\} \geqslant \min\{|x|\,|\,x \in fl_1(C * D)\}.$$

Из (15b), (20) и неравенства $|A * B| \leqslant |C * D|$ вытекает

$$d(fl_2(A * B)) \leqslant d(A * B) + 2b^{1-t_1-l}|C * D|$$
$$\leqslant \left\{\begin{array}{ll} s_1 + s_2, & * = +, - \\ |C|s_2 + s_1|D|, & * = \cdot \\ |C||1/D|^2 s_2 + |1/D|s_1, & * = : \end{array}\right\} + 2b^{1-t_1-l}|C * D|.$$

Это доказывает утверждение теоремы для верхних границ абсолютной погрешности. Для верхних границ относительной погрешности требуемый результат получается непосредственно из (21).

Простое, но важное следствие из теоремы 4 содержит

Теорема 5. *Допустим, что справедливы все приведенные выше предположения, касающиеся машинной интервальной арифметики. Кроме того, имеется заданный в поле вещественных чисел алгоритм, который выполняется в машинной интервальной арифметике с мантиссой длины t_1. Если затем этот алгоритм выполнить в арифметике с мантиссой длины t_2, где $t_2 = t_1 + l$, $l > 0$, то границы абсолютной и относительной погрешностей уменьшатся в b^l раз. (Под алгоритмом здесь понимается однозначно определенная последовательность арифметических операций вместе с конкретными входными данными.)*

Доказательство. Из (14) следует, что в нашем случае интервальное округление входных данных удовлетворяет важному предположению (19) теоремы 4. Свойства интервальной арифметики обеспечивают справедливость (18). Окончательно доказательство получаем из теоремы 4 применением полной индукции. ■

Теорема 5 указывает способ получения результата с наперед заданной абсолютной или относительной точностью. Пусть, например, d_1 — наибольшая ширина, которую имеют результирующие интервалы, вычисленные с помощью t_1-значной мантиссы,

Таблица 1

Верхняя граница $\rho(X_i)$ относительной погрешности в алгоритме Гаусса

Число цифр в мантиссе, t	15	20	25	30	35
1	> 1[a]	0.11×10^{-3}	0.11×10^{-8}	0.11×10^{-13}	0.11×10^{-18}
2	0.34×10^{0}	0.29×10^{-5}	0.29×10^{-10}	0.29×10^{-15}	0.29×10^{-20}
3	0.18×10^{-1}	0.17×10^{-6}	0.17×10^{-11}	0.17×10^{-16}	0.17×10^{-21}
4	0.16×10^{-2}	0.16×10^{-7}	0.16×10^{-12}	0.16×10^{-17}	0.16×10^{-22}
5	0.26×10^{-3}	0.25×10^{-8}	0.25×10^{-13}	0.25×10^{-18}	0.25×10^{-23}
6	0.64×10^{-4}	0.64×10^{-9}	0.64×10^{-14}	0.64×10^{-19}	0.64×10^{-24}
7	0.58×10^{-4}	0.58×10^{-9}	0.58×10^{-14}	0.58×10^{-19}	0.58×10^{-24}

[a]) $\rho(X_1) > 1$ означает здесь, что интервал X_1 содержит 0.

а ε — требуемая абсолютная точность. Если $d_1 \leqslant \varepsilon$, то цель достигнута. В противном случае число цифр в мантиссе увеличивается на l, где l удовлетворяет неравенству

$$b^{-l} d_1 \leqslant \varepsilon.$$

(Такой выбор не гарантирует, что абсолютная погрешность уменьшится в b^l раз. В соответствии с теоремой 5 эта оценка верна лишь для верхней границы абсолютной погрешности.)

Факты, обсужденные и доказанные в теореме 5, были изучены Румпом [501]; он же проиллюстрировал их числовыми примерами. Один из этих примеров — решение системы уравнений, задаваемой матрицей Гильберта размерности 7×7, причем в правой части каждого уравнения стоит 1. Результаты применения к этой системе алгоритма Гаусса, реализованного в машинной интервальной арифметике с 15, 20, 25, 30 и 35 цифрами в мантиссе, воспроизведены в табл. 1. Следует иметь в виду, что в ней дана лишь верхняя граница $\rho(X_i)$ относительной погрешности для каждой компоненты вектора результата. Дополнительные подробности можно найти у Румпа [501]. Про-

стой пример, иллюстрирующий содержание теоремы 5, приведен у Мура [451, с. 60].

А теперь рассмотрим следующую задачу. Пусть имеются машинные интервалы (т. е. вещественные интервалы, границами которых служат машинные числа)

$$C_0,\ A_0,\ B_0,\ D_0,\ A_1,\ B_1,\ D_1,\ \ldots,\ A_{n-1},\ B_{n-1},\ D_{n-1},$$

а также машинное число a_n. Требуется вычислить выражение

$$R_n = (1/a_n)\{C_0 - A_0(B_0 - D_0) - A_1(B_1 - D_1) - \ldots - A_{n-1}(B_{n-1} - D_{n-1})\}.$$

Теоретически можно воспользоваться таким алгоритмом:

$$\begin{aligned} &S_0 := C_0, \\ (S)\qquad &S_i := S_{i-1} - A_{i-1}(B_{i-1} - D_{i-1}), \quad 1 \leqslant i \leqslant n, \\ &R_n := S_n/a_n. \end{aligned}$$

На практике, однако, этот алгоритм выполняется в виде

$$\begin{aligned} &\overline{S}_0 := S_0 := C_0, \\ (\overline{S})\qquad &\overline{S}_i := fl(\overline{S}_{i-1} - fl(A_{i-1} fl(B_{i-1} - D_{i-1}))), \quad 1 \leqslant i \leqslant n, \\ &\overline{R}_n := fl(\overline{S}_n/a_n). \end{aligned}$$

Начав с (15а), установим $\mathrm{eps} := \frac{1}{2} b^{1-t}$ и для произвольных интервалов A и B получим, что

$$(22)\qquad fl(A * B) \subseteq A * B + [-\varepsilon, \varepsilon]\, A * B,$$

где $\max\{|\varepsilon_1|, |\varepsilon_2|\} \leqslant \varepsilon$, $\varepsilon = 2\ \mathrm{eps}$.

Предположим на время, что

$$\overline{S}_0 = S_0 = C_0, \quad \overline{S}_1, \ldots, \overline{S}_{n-1}$$

уже вычислено. Тогда из (22) следует

$$\begin{aligned} &fl(B_{n-1} - D_{n-1}) \subseteq B_{n-1} - D_{n-1} + |B_{n-1} - D_{n-1}|[-\varepsilon, \varepsilon], \\ &\quad fl(A_{n-1} fl(B_{n-1} - D_{n-1})) \\ &\subseteq A_{n-1}(B_{n-1} - D_{n-1} + |B_{n-1} - D_{n-1}|[-\varepsilon, \varepsilon]) \\ &\quad + |A_{n-1}(B_{n-1} - D_{n-1} + |B_{n-1} - D_{n-1}|[-\varepsilon, \varepsilon])|[-\varepsilon, \varepsilon] \\ &\subseteq A_{n-1}(B_{n-1} - D_{n-1}) + |A_{n-1}||B_{n-1} - D_{n-1}|[-2\varepsilon - \varepsilon^2, 2\varepsilon + \varepsilon^2], \end{aligned}$$

а значит,

$$\begin{aligned}\overline{S}_n \subseteq\ & \overline{S}_{n-1} - A_{n-1}(B_{n-1} - D_{n-1}) \\ & - |A_{n-1}||B_{n-1} - D_{n-1}|[-2\varepsilon - \varepsilon^2,\ 2\varepsilon + \varepsilon^2] \\ & + |\overline{S}_{n-1} - A_{n-1}(B_{n-1} - D_{n-1}) \\ & - |A_{n-1}||B_{n-1} - D_{n-1}|[-2\varepsilon - \varepsilon^2,\ 2\varepsilon + \varepsilon^2]|[-\varepsilon,\ \varepsilon] \\ \subseteq\ & \overline{S}_{n-1} - A_{n-1}(B_{n-1} - D_{n-1}) + |\overline{S}_{n-1}|[-\varepsilon,\ \varepsilon] \\ & + |A_{n-1}||B_{n-1} - D_{n-1}|[-3\varepsilon - 3\varepsilon^2 - \varepsilon^3,\ 3\varepsilon + 3\varepsilon^2 + \varepsilon^3].\end{aligned} \tag{23}$$

При помощи математической индукции покажем, что

$$\begin{aligned}\overline{S}_n \subseteq\ & S_n + [-\varepsilon,\ \varepsilon]\sum_{i=0}^{n-1}|\overline{S}_i| + [-3\varepsilon - 3\varepsilon^2 - \varepsilon^3,\ 3\varepsilon + 3\varepsilon^2 + \varepsilon^3] \\ & \times \sum_{i=0}^{n-1}|A_i||B_i - D_i|.\end{aligned} \tag{24}$$

Для $n = 1$ из (23) с учетом того, что $S_0 = S_0 = C_0$, получаем

$$\begin{aligned}\overline{S}_1 \subseteq\ & \overline{S}_0 - A_0(B_0 - D_0) + |\overline{S}_0|[-\varepsilon,\ \varepsilon] \\ & + |A_0||B_0 - D_0|[-3\varepsilon - 3\varepsilon^2 - \varepsilon^3,\ 3\varepsilon + 3\varepsilon^2 + \varepsilon^3] \\ =\ & S_1 + [-\varepsilon,\ \varepsilon]|\overline{S}_0| + [-3\varepsilon - 3\varepsilon^2 - \varepsilon^3,\ 3\varepsilon + 3\varepsilon^2 + \varepsilon^3]|A_0||B_0 - D_0|,\end{aligned}$$

откуда видна справедливость нашего предположения при $n = 1$. Если для некоторого $n \geqslant 1$ выполнено (24), тогда замена n на $n + 1$ в (23), а также использование (S) дают

$$\begin{aligned}\overline{S}_{n+1} \subseteq\ & \overline{S}_n - A_n(B_n - D_n) + [-\varepsilon,\ \varepsilon]|\overline{S}_n| \\ & + [-3\varepsilon - 3\varepsilon^2 - \varepsilon^3,\ 3\varepsilon + 3\varepsilon^2 + \varepsilon^3]|A_n||B_n - D_n| \\ \subseteq\ & S_{n+1} + [-\varepsilon,\ \varepsilon]\sum_{i=0}^{n}|\overline{S}_i| + [-3\varepsilon - 3\varepsilon^2 - \varepsilon^3,\ 3\varepsilon + 3\varepsilon^2 + \varepsilon^3] \\ & \times \sum_{i=0}^{n}|A_i||B_i - D_i|.\end{aligned}$$

Последнее выражение тождественно (24), у которого n заменено на $n + 1$. Повторное применение (22) приводит к окончательному результату

$$\overline{R}_n \subseteq \overline{S}_n/a_n + (|\overline{S}_n|/|a_n|)[-\varepsilon,\ \varepsilon]. \tag{25}$$

Неравенства (24) и (25) будут использованы в гл. 19.

Замечания. Понятие округления в том виде, как оно встречается в этой главе, подробно рассмотрено Миранкером и Кулишем [443]. Неравенство (15b), явившееся исходным пунктом при

обсуждении влияния погрешностей округления, можно найти у Валлиша и Грюцманна [530]. Оценка (24) была доказана Алефельдом и Рокном [351a]; мы еще вернемся к ней в гл. 18. Замечание, сделанное Муром в [447, с. 70], позволяет заключить, что теорема 5 уже была им доказана в [445, с. 102].

* * *

Литература: [33], [34], [38], [55], [72], [77], [98], [123], [124], [155], [169], [175], [178], [180], [181], [185], [195], [264], [266], [329], [332], [334], [351a], [433], [443], [445], [447], [451], [501], [518], [530].

Глава 5
КОМПЛЕКСНАЯ ИНТЕРВАЛЬНАЯ АРИФМЕТИКА

Теперь мы хотим определить так называемую комплексную интервальную арифметику. Будет показано, что многие из свойств и результатов, полученных для вещественной интервальной арифметики, можно перенести на случай комплексной. Чтобы это проделать, мы должны определить множества комплексных чисел, которые будут использоваться в качестве комплексных интервалов. Имеются два разумных подхода, к рассмотрению которых мы сейчас и перейдем.

A. Прямоугольники в качестве комплексных интервалов

Определение 1. Пусть A_1 и A_2 — произвольные элементы из $I(\mathbb{R})$. Тогда множество комплексных чисел

$$A = \{a = a_1 + ia_2 \mid a_1 \in A_1,\ a_2 \in A_2\} \qquad (i = \sqrt{-1})$$

называется комплексным интервалом. ■

Определенные таким образом множества комплексных чисел могут быть изображены на комплексной плоскости в виде прямоугольников со сторонами, параллельными осям координат. Множество всех таких комплексных интервалов обозначим через $R(\mathbb{C})$, а прописные буквы A, B, C, ..., X, Y, Z будем использовать для обозначения его элементов. Всякое A из $R(\mathbb{C})$ можно записать в виде $A = A_1 + iA_2$, где A_1, $A_2 \in I(\mathbb{R})$.

Комплексное число $a = a_1 + ia_2$ можно рассматривать как точечный комплексный интервал:

$$A = [a_1, a_1] + i[a_2, a_2] \in R(\mathbb{C}).$$

а каждый элемент A_1 из $I(\mathbb{R})$ — как сумму $A = A_1 + i[0, 0] \in R(\mathbb{C})$, откуда видно, что $I(\mathbb{R}) \subset R(\mathbb{C})$.

Определение 2. Пусть $A = A_1 + iA_2$ и $B = B_1 + iB_2$ — два элемента из $R(\mathbb{C})$. Тогда A и B считаются равными (запись: $A = B$), если

$$A_1 = B_1 \quad \text{и} \quad A_2 = B_2$$

(см. также определение 1.1). ■

Определенное здесь отношение равенства рефлективно, симметрично и транзитивно.

Теперь мы обобщим арифметику комплексных чисел на случай $R(\mathbb{C})$.

Определение 3. Пусть $*$ из $\{+, -, \cdot, :\}$ — бинарная операция над элементами из $I(\mathbb{R})$ (как в определении 1.2). Тогда если

$$A = A_1 + iA_2, \quad B = B_1 + iB_2 \in R(\mathbb{C}),$$

то мы полагаем

$$A \pm B = A_1 \pm B_1 + i(A_2 \pm B_2),$$
$$A \cdot B = A_1B_1 - A_2B_2 + i(A_1B_2 + A_2B_1),$$
$$A : B = (A_1B_1 + A_2B_2) : (B_1^2 + B_2^2) + i(A_2B_1 - A_1B_2) : (B_1^2 + B_2^2). \blacksquare$$

Считается, что в случае деления $0 \notin B_1^2 + B_2^2$. При вычислении степеней $B_1^2 = B_1B_1$ и $B_2^2 = B_2B_2$ это требование может оказаться невыполненными, даже если $0 \notin B_1 + iB_2$ в соответствии с определением 1.2. Если при этом оказывается, что $0 \in B_1^2 + B_2^2$, то деление не определено.

Чтобы проиллюстрировать сказанное, рассмотрим следующий пример.

Пример. Пусть

$$B = [-1, 1] + i[1, 3].$$

Тогда

$$0 \in [0, 10] = [-1, 1] + [1, 9] = B_1B_1 + B_2B_2.$$

Поэтому мы оговариваем, что в определении 3, если производится деление двух элементов из $R(\mathbb{C})$, выражение $B_1^2 + B_2^2$ следует вычислять по правилу

$$B_1^2 + B_2^2 = \{b_1^2 | b_1 \in B_1\} + \{b_2^2 | b_2 \in B_2\}$$

(см. также определение 1.3).

Тогда в приведенном выше примере

$$B_1^2 + B_2^2 = [0, 1] + [1, 9] = [1, 10].$$

Теперь рассмотрим более внимательно свойства введенной выше комплексной интервальной арифметики.

Сразу же видно, что если $A, B \in R(\mathbb{C})$, то равенство

$$A \pm B = \{a \pm b \mid a \in A, b \in B\}$$

справедливо для сложения (соответственно вычитания) на множестве $R(\mathbb{C})$. Аналогичное равенство для умножения и деления, вообще говоря, не выполняется. Это можно увидеть из следующего простого примера.

Пример. Пусть

$$A = [2, 4] + i[0, 0], \quad B = [1, 1] + i[1, 1].$$

Из определения 3 получаем

$$AB = [2, 4] + i[2, 4].$$

С другой стороны,

$$\{ab \mid a \in A, b \in B\} = \{s(1+i) \mid s \in \mathbb{R}, 2 \leqslant s \leqslant 4\} \subset AB.$$

Справедлива, однако, следующая теорема.

Теорема 4. *Операции, введенные определением* 3, *удовлетворяют соотношению*

$$\{a * b \mid a \in A, b \in B\} \subseteq A * B.$$

Для сложения и вычитания включение может быть заменено на равенство. Для умножения

$$AB = \inf \{X \in R(\mathbb{C}) \mid \{ab \mid a \in A, b \in B\} \subseteq X\},$$

где точная нижняя грань берется в смысле частичного порядка на $R(\mathbb{C})$, определяемого теоретико-множественным включением.

Доказательство. Случай сложения и вычитания был рассмотрен ранее. Пусть теперь $a \in A$ и $b \in B$. Используя монотонность включения вещественных интервалов, для $a = a_1 + ia_2$ и $b = b_1 + ib_2$ имеем

$$\begin{aligned} ab &= a_1b_1 - a_2b_2 + i(a_1b_2 + a_2b_1) \\ &\in A_1B_1 - A_2B_2 + i(A_1B_2 + A_2B_1) = AB. \end{aligned}$$

Поскольку каждая переменная входит в выражение $a_1b_1 - a_2b_2$ лишь один раз, то получаем, что

$$\{a_1b_1 - a_2b_2 \mid a_k \in A_k, b_k \in B_k, k = 1, 2\} = A_1B_1 - A_2B_2.$$

По той же причине

$$\{a_1b_2 + a_2b_1 \mid a_k \in A_k, b_k \in B_k, k = 1, 2\} = A_1B_2 + A_2B_1.$$

Последние два соотношения показывают, что для каждого вещественного числа c_1 такого, что

$$c_1 = a_1 b_1 - a_2 b_2 \in A_1 B_1 - A_2 B_2,$$

где $a_k \in A_k$, $b_k \in B_k$, $k = 1, 2$, можно найти другое вещественное число c_2 такое, что

$c_2 = a_2 b_1 + a_1 b_2 \in A_2 B_1 + A_1 B_2$, где $a_k \in A_k$, $b_k \in B_k$, $k = 1, 2$ и $c_1 + ic_2 \in AB$. Это и требовалось доказать.

Соотношение

$$\{a : b \mid a \in A,\ b \in B\} \subseteq A : B$$

также следует из монотонности включения. ■

Утверждение теоремы 4, касающееся умножения, не допускает, вообще говоря, распространения на случай деления. Тем не менее можно получить «уточнение» (в смысле сужения объемлющего интервала), если определить, что

$$A : B = A \cdot \frac{1}{B},$$

а затем вычислять $1/B$ по формуле

$$1/B = \inf \{X \in R(\mathbb{C}) \mid \{1/b \mid b \in B\} \subseteq X\}.$$

Данная возможность была предложена Рокном и Ланкастером [275] в виде набора формул, требующих значительной вычислительной работы. В публикации этих же авторов [276] приведены подпрограммы на Фортране.

B. Круги в качестве комплексных интервалов

Определение 5. Пусть a из $\mathbb{C}$ — комплексное число, и пусть $r \geqslant 0$. Мы называем множество

$$Z = \{z \in \mathbb{C} \mid |z - a| \leqslant r\}$$

кругом или круговым интервалом (или просто комплексным интервалом, когда не опасаемся спутать его с прямоугольным интервалом). ■

Множество всех кругов обозначим через $K(\mathbb{C})$, а прописные буквы $A, B, C, \ldots, X, Y, Z$ используем для обозначения его элементов. Круг Z с центром a и радиусом r будем записывать в виде

$$Z = \langle a, r \rangle.$$

Комплексные числа можно рассматривать как специальные элементы из $K(\mathbb{C})$, имеющие вид $\langle a, 0 \rangle$. Ясно, что $\mathbb{C} \subset K(\mathbb{C})$.

Определение 6. Два круга $A = \langle a, r_1 \rangle$ и $B = \langle b, r_2 \rangle$ называются равными (обозначение: $A = B$), если они равны в теоретико-множественном смысле. В этом случае $a = b$ и $r_1 = r_2$. ■

Это отношение равенства также рефлексивно, симметрично и транзитивно.

Операции на $K(\mathbb{C})$ вводятся как обобщения операций над вещественными числами следующим образом.

Определение 7. Пусть $*$ из $\{+, -, \cdot, :\}$ — бинарная операция над комплексными числами. Тогда если $A = \langle a, r_1 \rangle$ и $B = \langle b, r_2 \rangle$, то

$$A \pm B = \langle a \pm b,\ r_1 + r_2 \rangle,$$

$$A \cdot B = \langle ab,\ |a| r_2 + |b| r_1 + r_1 r_2 \rangle,$$

$$\frac{1}{B} = \left\langle \frac{\bar{b}}{b\bar{b} - r_2^2},\ \frac{r_2}{b\bar{b} - r_2^2} \right\rangle \quad \text{при условии, что} \quad 0 \notin B,$$

$$A : B = A \cdot \frac{1}{B} \quad \text{при условии, что} \quad 0 \notin B.$$

Здесь $|a| = \sqrt{a_1^2 + a_2^2}$ обозначает евклидову норму комплексного числа $a = a + ia_2$, а $\bar{b} = b_1 + ib_2$ — сопряженное с $b = b_1 + ib_2$. ■

Очевидно, что для сложения и вычитания кругов выполнено равенство

$$A \pm B = \{a \pm b \mid a \in A,\ b \in B\}.$$

То же справедливо для операции обращения круга: если мы применим теорию конформных отображений к отображению вида $w = 1/z$ для не содержащего нуля круга, то получим другой круг. Иными словами,

$$1/B = \{1/b \mid b \in B\}.$$

Элементарными преобразованиями легко проверить формулы определения 7 для центра и радиуса области $\{1/b \mid b \in B\}$.

Для умножения (а следовательно, и для деления) двух элементов из $K(\mathbb{C})$ по правилам определения 7 верно, вообще говоря, только то, что

$$\{z_1 z_2 \mid z_1 \in A,\ z_2 \in B\} \subseteq AB.$$

Это вытекает из следующих неравенств:

$$\begin{aligned} |z_1 z_2 - ab| &= |a(z_2 - b) + b(z_1 - a) + (z_1 - a)(z_2 - b)| \\ &\leqslant |a|\,|z_2 - b| + |b|\,|z_1 - a| + |z_1 - a|\,|z_2 - b| \\ &\leqslant |a| r_2 + |b| r_1 + r_1 r_2. \end{aligned}$$

Крир [168] обсуждает вопрос об «уточнении» в смысле включения результата для операции умножения элементов из $K(\mathbb{C})$. Такое уточнение требует значительной вычислительной работы, которая среди прочего предполагает нахождение корней многочлена третьей степени. Теряется также свойство, состоящее в том, что произведение центров операндов совпадает с центром произведения.

Аналогично теореме 1.4, соберем теперь вместе наиболее важные свойства операций на $R(\mathbb{C})$ и $K(\mathbb{C})$. Если не оговорено противное, то $I(\mathbb{C})$ можно понимать и как обозначение множества $R(\mathbb{C})$ с операциями из определения 3, и как обозначение множества $K(\mathbb{C})$ с операциями из определения 7.

Теорема 8. *Пусть* $A, B, C \in I(\mathbb{C})$. *Тогда*

(1) $A + B = B + A, \quad AB = BA$ (*коммутативность*);

(2) $(A + B) + C = A + (B + C)$,
$(AB)C = A(BC)$ *для* A, B, C *из* $K(\mathbb{C})$ (*ассоциативность*);

(3) $[0, 0] + i[0, 0] \in R(\mathbb{C})$ (*соответственно* $\langle 0, 0\rangle \in K(\mathbb{C})$) *и*
$[1, 1] + i[0, 0] \in R(\mathbb{C})$ (*соответственно* $\langle 1, 0\rangle \in K(\mathbb{C})$)
— *определенные единственным образом нейтральные элементы сложения* (*нуль*) *и умножения* (*единица*);

(4) $I(\mathbb{C})$ *не имеет делителей нуля*;

(5) *элемент* Z *множества* $I(\mathbb{C})$ *имеет противоположный и обратный элементы, если и только если* $Z \in \mathbb{C}$ *и в случае мультипликативных операций* $Z \neq 0$.
Однако $0 \in A - A$ *и* $1 \in A : A$;

(6) $A(B + C) \subseteq AB + AC$ (*субдистрибутивность*),
$a(B + C) = aB + aC$ *для* $a \in \mathbb{C}$.

Доказательство. Доказательство этих утверждений следует из определений операций 3 и 7. В качестве примера докажем (6) для $K(\mathbb{C})$. Если $A = \langle a, r_1\rangle$, $B = \langle b, r_2\rangle$, $C = \langle c, r_3\rangle \in K(\mathbb{C})$, то

$$\begin{aligned} A(B + C) &= \langle a, r_1\rangle\langle b + c, r_2 + r_3\rangle \\ &= \langle a(b + c), |a|(r_2 + r_3) + |b + c| r_1 + r_1(r_2 + r_3)\rangle \\ &\subseteq \langle ab + ac, |a| r_2 + |a| r_3 + |b| r_1 + |c| r_1 + r_1 r_2 + r_1 r_3\rangle \\ &= \langle ab, |a| r_2 + |b| r_1 + r_1 r_2\rangle + \langle ac, |a| r_3 + |c| r_1 + r_1 r_3\rangle \\ &= AB + AC. \end{aligned}$$

В случае $A = \langle a, 0\rangle$, т. е. равенства нулю r_1, доказательство приводит к соотношению

$$a(B + C) = aB + aC. \quad \blacksquare$$

Мы хотели бы особо отметить, что ассоциативный закон (2) в общем случае не выполняется при перемножении элементов $R(\mathbb{C})$. Это можно увидеть из следующего примера.

Пример.

$$A = [2, 4] + i[0, 0], \quad B = [1, 1] + i[1, 1], \quad C = [1, 1] + i[1, 1],$$
$$(AB)C = ([2, 4] + i[2, 4])([1, 1] + i[1, 1]) = [-2, 2] + i[4, 8],$$
$$A(BC) = ([2, 4] + i[0, 0])([0, 0] + i[2, 2]) = [0, 0] + i[4, 8].$$

$I(\mathbb{C})$ обладает также монотонностью включения.

Теорема 9. *Пусть* $A^{(k)}, B^{(k)} \in I(\mathbb{C}), k = 1, 2,$ *таковы, что*

$$A^{(k)} \subseteq B^{(k)}, \quad k = 1, 2.$$

Тогда соотношение

$$A^{(1)} * A^{(2)} \subseteq B^{(1)} * B^{(2)}$$

выполняется для операций $*$ *из* $\{+, -, \cdot, :\}$.

Доказательство. Это верно для $R(\mathbb{C})$, поскольку монотонность включения выполняется для элементов $I(\mathbb{R})$ (см. теорему 1.5).

В случае сложения и вычитания элементов $K(\mathbb{C})$ имеем

$$A^{(1)} \pm A^{(2)} = \{z = x \pm y \mid x \in A^{(1)}, y \in A^{(2)}\}$$
$$\subseteq \{w = u \pm v \mid u \in B^{(1)}, v \in B^{(2)}\} = B^{(1)} \pm B^{(2)}.$$

Рассмотрим теперь умножение на множестве $K(\mathbb{C})$. Пусть

$$A^{(k)} = \langle a^{(k)}, r^{(k)} \rangle, \quad B^{(k)} = \langle b^{(k)}, s^{(k)} \rangle, \quad k = 1, 2.$$

Предположение $A^{(k)} \subseteq B^{(k)}$, $k = 1, 2$, эквивалентно тому, что

$$|a^{(k)} - b^{(k)}| \leqslant s^{(k)} - r^{(k)}, \quad k = 1, 2.$$

Далее,

$$A^{(1)}A^{(2)} = \langle a^{(1)}a^{(2)}, |a^{(1)}| r^{(2)} + |a^{(2)}| r^{(1)} + r^{(1)}r^{(2)} \rangle,$$
$$B^{(1)}B^{(2)} = \langle b^{(1)}b^{(2)}, |b^{(1)}| s^{(2)} + |b^{(2)}| s^{(1)} + s^{(1)}s^{(2)} \rangle.$$

Требуется доказать, что

$$|a^{(1)}a^{(2)} - b^{(1)}b^{(2)}| \leqslant |b^{(1)}| s^{(2)} + |b^{(2)}| s^{(1)} + s^{(1)}s^{(2)}$$
$$- (|a^{(1)}| r^{(2)} + |a^{(2)}| r_1 + r^{(1)}r^{(2)}).$$

Из неравенства треугольника получаем

$$-|b^{(2)}| \leqslant -|a^{(2)}| + |a^{(2)} - b^{(2)}|,$$
$$-|b^{(1)}| \leqslant -|a^{(1)}| + |a^{(1)} - b^{(1)}|,$$

а поскольку

$$|a^{(k)} - b^{(k)}| \leqslant s^{(k)} - r^{(k)}, \quad k = 1, 2,$$

имеем

$$-|b^{(2)}|r^{(1)} \leqslant -|a^{(2)}|r^{(1)} + r^{(1)}(s^{(2)} - r^{(2)})$$
$$= -|a^{(2)}|r^{(1)} + r^{(1)}s^{(2)} - r^{(1)}r^{(2)},$$
$$-|b^{(1)}|r^{(2)} \leqslant -|a^{(1)}|r^{(2)} + r^{(2)}(s^{(1)} - r^{(1)})$$
$$= -|a^{(1)}|r^{(2)} + r^{(2)}s^{(1)} - r^{(1)}r^{(2)}.$$

Отсюда

$$|a^{(1)}a^{(2)} - b^{(1)}b^{(2)}| \leqslant |b^{(2)}||a^{(1)} - b^{(1)}| + |b^{(1)}||a^{(2)} - b^{(2)}|$$
$$+ |a^{(1)} - b^{(1)}||a^{(2)} - b^{(2)}|$$
$$\leqslant |b^{(2)}|(s^{(1)} - r^{(1)}) + |b^{(1)}|(s^{(2)} - r^{(2)})$$
$$+ (s^{(1)} - r^{(1)})(s^{(2)} - r^{(2)})$$
$$\leqslant |b^{(2)}|s^{(1)} + |b^{(1)}|s^{(2)} + s^{(1)}s^{(2)}$$
$$- (|a^{(2)}|r^{(1)} + |a^{(1)}|r^{(2)} + r^{(1)}r^{(2)}),$$

что доказывает теорему для случая умножения.

Из того, что

$$1/A^{(2)} = \{z = 1/x \mid x \in A^{(2)}\} \subseteq \{w = 1/u \mid u \in B^{(2)}\} = 1/B^{(2)},$$

следует, что

$$A^{(1)} : A^{(2)} = A^{(1)} \cdot \frac{1}{A^{(2)}} \subseteq B^{(1)} \cdot \frac{1}{B^{(2)}} = B^{(1)} : B^{(2)}.$$

Теорема доказана. ■

Частным случаем теоремы 9 является

Следствие 10. *Пусть* $A, B \in I(\mathbb{C})$ *и* $a \in A$, $b \in B$. *Тогда*

$$a * b \in A * B,$$

где $* \in \{+, -, \cdot, :\}$. ■

Замечания. Круги в качестве комплексных интервалов впервые ввели в систематическое употребление Гаргантини и Энричи [95, 120]. Рассмотренная в этой главе арифметика кругов была предложена ими же в работе [95], где она была использована для одновременной локализации корней многочлена (см. гл. 9). Другие приложения арифметики кругов можно найти в гл. 12, 15 и 16. Важное свойство монотонности включения в том виде, как оно сформулировано в теореме 9, впервые доказано здесь. Как уже было отмечено, в [168] обсуждается проблема такого определения умножения кругов, которое приводит к получению меньших множеств. Арифметика на $R(\mathbb{C})$, сводящаяся при выполнении некоторых условий к вещественной, предложена Алефельдом в [4]. Свойства этой арифметики исследовали также Бош [56], Рокн и Ланкастер [275].

Как уже неоднократно указывалось, в основе почти всех приложений интервальной арифметики лежит монотонность включения (теорема 9). Умножение, предложенное Криером в [168], не обладает этим свойством, что было показано там же на примере [168, с. 76].

Приближенная реализация арифметики прямоугольников на цифровой вычислительной машине не вызывает проблем, поскольку операции на $R(\mathbb{C})$ сводятся к операциям на $I(\mathbb{R})$. В гл. 4 было показано, как сделать возможной реализацию приближенных операций над элементами $I(\mathbb{R})$ на ЦВМ без потери наиболее важных свойств арифметики, что позволяет проделать то же самое для $R(\mathbb{C})$.

В [305] обсуждается возможность приближенной машинной реализации арифметики на множестве $K(\mathbb{C})$.

* * *

Литература: [4], [56], [95], [120], [168], [275], [276].

Глава 6
МЕТРИКА, АБСОЛЮТНАЯ ВЕЛИЧИНА И ШИРИНА В $I(\mathbb{C})$

В этой главе q будет обозначать метрику на $I(\mathbb{R})$, задаваемую определением 2.1. Введем теперь метрику на $R(\mathbb{C})$.

Определение 1. Пусть $A = A_1 + iA_2$ и $B = B_1 + iB_2$ принадлежат $R(\mathbb{C})$. Тогда расстояние p между A и B определяется формулой

$$p(A, B) = q(A_1, B_1) + q(A_2, B_2). \blacksquare$$

Если p используется в пространстве $I(\mathbb{R})$, то оно принимает те же самые значения, что и q из определения 2.1. Поэтому в дальнейшем расстояние на $R(\mathbb{C})$ будет обозначаться через q, так что

$$q(A, B) = q(A_1, B_1) + q(A_2, B_2).$$

Поскольку q является метрикой на $I(\mathbb{R})$, легко доказать, что q останется ею и при переходе к $R(\mathbb{C})$. Введение на $R(\mathbb{C})$ метрики q делает его топологическим пространством. Если теперь обычным для метрических пространств способом определить сходимость, то окажется, что последовательность $\{A^{(k)}\}_{k=0}^{\infty}$, где $A^{(k)} = A_1^{(k)} + iA_2^{(k)} \in R(\mathbb{C})$, сходится к $A = A_1 + iA_2$ — эле-

менту $R(\mathbb{C})$ тогда и только тогда, когда

$$\lim_{k\to\infty} A_1^{(k)} = A_1 \quad \text{и} \quad \lim_{k\to\infty} A_2^{(k)} = A_2. \tag{1}$$

Из факта замкнутости метрического пространства $(I(\mathbb{R}), q)$ вытекает, что $R(\mathbb{C})$ с заданной на нем метрикой q также образует замкнутое метрическое пространство.

Определение 2. Пусть $A = A_1 + iA_2 \in R(\mathbb{C})$. Тогда

$$|A| = q(A, 0) = |A_1| + |A_2| = q(A_1, 0) + q(A_2, 0)$$

называется абсолютной величиной A. ■

Если, в частности, $A = [a_1, a_1] + i[a_2, a_2] = a_1 + ia_2 = a$, то

$$|A| = |a| = |a_1| + |a_2|. \tag{2}$$

Итак, абсолютная величина из определения 2 не совпадает с евклидовой нормой комплексного числа. В дальнейшем из контекста всегда будет ясно, какая из них используется в конкретном случае. Кроме того, заметим, что из (2) следует справедливость соотношения

$$|A| = \max_{a \in A} |a|.$$

Пусть d в соответствии с определением 2.8 служит обозначением ширины вещественного интервала. Тогда имеем

Определение 3. Если $A = A_1 + iA_2 \in R(\mathbb{C})$, то шириной A будем называть

$$d(A) = d(A_1) + d(A_2). \blacksquare$$

Теперь перейдем к множеству $K(\mathbb{C})$.

Определение 4. Пусть $A = \langle a, r_1\rangle$, $B = \langle b, r_2\rangle \in K(\mathbb{C})$. Тогда назовем

(a) $q(A, B) = |a - b| + |r_1 - r_2|$ расстоянием между A и B,
(b) $|A| = |a| + r_1$ абсолютной величиной A и
(c) $d(A) = 2r_1$ шириной A. ■

Для того чтобы определить расстояние между двумя круговыми интервалами на комплексной плоскости, в определении 4 используется евклидова метрика. Когда абсолютная величина кругового интервала применяется к обычным комплексным числам, она совпадает с евклидовой нормой. Обратим внимание, что и на этот раз выполняется соотношение

$$|A| = \max_{a \in A} |a|.$$

Если сходимость последовательности, состоящей из элементов $K(\mathbb{C})$, определена, как обычно, с помощью соответствующей метрики, то можно легко убедиться в замкнутости метрического пространства $(K(\mathbb{C}), q)$. Опираясь на такого рода определение, получаем, что

$$\lim_{k\to\infty} A^{(k)} = A \Leftrightarrow \lim_{k\to\infty} a^{(k)} = a, \quad \lim_{k\to\infty} r^{(k)} = r, \tag{3}$$

где $\{A^{(k)}\}_{k=0}^{\infty} = \{\langle a^{(k)}, r^{(k)}\rangle\}_{k=0}^{\infty}$, $A = \langle a, r\rangle$.

В следующей теореме собраны наиболее важные свойства метрики, абсолютной величины и ширины на множествах $R(\mathbb{C})$ и $K(\mathbb{C})$.

Теорема 5. *Пусть имеются A, B, C, D из $I(\mathbb{R})$. Тогда*

$$q(A+B, A+C) = q(B, C), \tag{4}$$

$$q(A+B, C+D) \leqslant q(A, C) + q(B, D), \tag{5}$$

$$q(aB, aC) \leqslant |a| q(B, C), \quad a \in \mathbb{C}. \tag{6}$$

Если B и C принадлежат $K(\mathbb{C})$, то в (6) всегда имеет место равенство.

$$q(AB, AC) \leqslant |A| q(B, C), \tag{7}$$

$$|A| \geqslant 0, \quad |A| = 0 \Leftrightarrow A = 0, \tag{8}$$

$$|A+B| \leqslant |A| + |B|, \tag{9}$$

$$|aB| \leqslant |a|\,|B|, \quad a \in \mathbb{C}. \tag{10}$$

Если $B \in K(\mathbb{C})$, то (10) превращается в равенство.

$$|AB| \leqslant |A|\,|B|, \tag{11}$$

$$d(aB) = |a| d(B), \quad a \in \mathbb{C}, \tag{12}$$

$$d(AB) \leqslant |A| d(B) + |B| d(A), \tag{13}$$

$$d(A) = |A - A|, \tag{14}$$

$$d(AB) \geqslant |A| d(B), \tag{15}$$

$$d(A \pm B) = d(A) + d(B), \tag{16}$$

$$A \subseteq B \Rightarrow \frac{1}{2}(d(B) - d(A)) \leqslant q(A, B) \leqslant d(B) - d(A). \tag{17}$$

Доказательство. Сначала докажем перечисленные свойства из $R(\mathbb{C})$. Справедливость (4)—(7) следует непосредственно из соответствующих свойств (2.4)—(2.7), сформулированных в теореме 2.7 применительно к вещественным интервалам.

Пусть

$$A = A_1 + iA_2, \quad B = B_1 + iB_2,$$
$$C = C_1 + iC_2, \quad D = D_1 + iD_2 \in R(\mathbb{C}).$$

(4): $r(A+B, A+C) = q(A_1+B_1+i(A_2+B_2), A_1+C_1+i(A_2+C_2))$
$= q(A_1+B_1, A_1+C_1) + q(A_2+B_2, A_2+C_2)$
$= q(B_1, C_1) + q(B_2, C_2) = q(B; C).$

(5): $q(A+B, C+D) = q(A_1+B_1, C_1+D_1) + q(A_2+B_2, C_2+D_2)$
$\leqslant q(A_1, C_1) + q(B_1, D_1) + q(A_2, C_2) + q(B_2, D_2)$
$= q(A, C) + q(B, D).$

(6), (7):

$q(AB, AC) = q(A_1B_1 - A_2B_2, A_1C_1 - A_2C_2)$
$+ q(A_1B_2 + A_2B_1, A_1C_2 + A_2C_1)$
$\leqslant |A_1| q(B_1, C_1) + |A_2| q(B_2, C_2) + |A_1| q(B_2, C_2) + |A_2| q(B_1, G_1)$
$= (|A_1| + |A_2|) q(B, C) = |A| q(B, C).$

Свойства (8)—(11) доказываются на основании определения $|A|$.

(8): $|A| = q(A, 0) = q(A_1, 0) + q(A_2, 0) = |A_1| + |A_2| \geqslant 0,$
$|A| = 0 \Leftrightarrow |A_1| = |A_2| = 0 \Leftrightarrow A = 0.$

(9): $|A+B| = q(A+B, 0) \leqslant q(A, 0) + q(B, 0) = |A| + |B|$

(в соответствии с (5)).

(10), (11):

$$|AB| = q(AB, 0) = q(AB, A \cdot 0) \leqslant |A| q(B, 0) = |A||B|$$

(из 6) и (7)).

(12): Пусть $a = a_1 + ia_2 \in \mathbb{C}$. Из определения 5.3 следует

$$aB = a_1B_1 - a_2B_2 + i(a_1B_2 + a_2B_1),$$

и с помощью (2) получаем

$d(aB) = d(a_1B_1 - a_2B_2) + d(a_1B_2 + a_2B_1)$
$= d(a_1B_1) + d(a_2B_2) + d(a_1B_2) + d(a_2B_1)$
$= |a_1| d(B_1) + |a_2| d(B_2) + |a_1| d(B_2) + |a_2| d(B_1)$
$= (|a_1| + |a_2|)(d(B_1) + d(B_2)) = |a| d(B).$

(13): $d(AB) = d(A_1B_1 - A_2B_2) + d(A_1B_2 + A_2B_1)$
$= d(A_1B_1) + d(A_2B_2) + d(A_1B_2) + d(A_2B_1)$
$\leqslant |A_1| d(B_1) + |B_1| d(A_1) + |A_2| d(B_2) + |B_2| d(A_2)$
$+ |A_1| d(B_2) + |B_2| d(A_1) + |A_2| d(B_1) + |B_1| d(A_2)$

$$= (|A_1| + |A_2|)(d(B_1) + d(B_2)) + (|B_1| + |B_2|)(d(A_1) + d(A_2))$$
$$= |A|d(B) + |B|d(A).$$

(14): $d(A) = d(A_1) + d(A_2) = |A_1 - A_1| + |A_2 - A_2| = |A - A|.$

(15): $d(AB) = d(A_1B_1 - A_2B_2) + d(A_1B_2 + A_2B_1)$
$$\leqslant |A_1|d(B_1) + |A_2|d(B_2) + |A_1|d(B_2) + |A_2|d(B_1)$$
$$= (|A_1| + |A_2|)(d(B_1) + d(B_2)) = |A|d(B).$$

(16): $d(A \pm B) = d(A_1 \pm B_1) + d(A_2 \pm B_2)$
$$= d(A_1) + d(A_2) + d(B_1) + d(B_2)$$
$$= d(A) + d(B).$$

(17): Это свойство является прямым следствием (2.21).

Теперь пусть

$$A = \langle a, r_1 \rangle, \quad B = \langle b, r_2 \rangle,$$
$$C = \langle c, r_3 \rangle, \quad D = \langle d, r_4 \rangle \in K(\mathbb{C}).$$

(4): $q(A + B, A + C) = |a + b - (a + c)| + |r_1 + r_2 - (r_1 + r_3)|$
$$= |b - c| + |r_2 - r_3| = q(B, C).$$

(5): $q(A + B, C + D) = |a + b - (c + d)| + |r_1 + r_2 - (r_3 + r_4)|$
$$\leqslant |a - c| + |r_1 - r_3| + |b - d| + |r_2 - r_4|$$
$$= q(A, C) + q(B, D).$$

(6): $q(aB, aC) = |ab - ac| + ||a|r_2 - |a|r_3|$
$$= |a|\{|b - c| + |r_2 - r_3|\} = |a|q(B, C).$$

(7): $q(AB, AC) = |ab - ac| + ||a|r_2 + |b|r_1 + r_1r_2$
$$- (|a|r_3 + |c|r_1 + r_1r_3)|$$
$$\leqslant |a||b - c| + |a||r_2 - r_3| + r_1||b| - |c|| + r_1|r_2 - r_3|$$
$$\leqslant (|a| + r_1)(|b - c| + |r_2 - r_3|) = |A|q(B, C).$$

(8): $|A| = |a| + r_1 \geqslant 0, \quad |A| = 0 \Leftrightarrow (a = 0, r_1 = 0).$

(9): $|A + B| = |a + b| + |r_1 + r_2| \leqslant |a| + r_1 + |b| + r_2 = |A| + |B|.$

(10): $|aB| = |ab| + |a|r_2 = |a||B|.$

(11): $|AB| = q(AB, 0) = q(AB, A \cdot 0) \leqslant |A|q(B, 0) = |A||B|$
(из (7)).

(12): $d(aB) = 2|a|r_2 = |a|d(B).$

(13): $d(AB) = 2\{|a|r_2 + |b|r_1 + r_1r_2\} = 2\{(|a| + r_1)r_2 + |b|r_1\}$
$$\leqslant 2\{(|a| + r_1)r_2 + (|b| + r_2)r_1\} = |A|d(B) + |B|d(A).$$

(14): $$d(A) = 2r_1 = |\langle 0,\ 2r_1\rangle| = |A - A|.$$

(15): $$\begin{aligned} d(AB) &= 2\{|a|r_2 + |b|r_1 + r_1r_2\} \\ &= 2\{(|a| + r_1)r_2 + |b|r_1\} \\ &\geqslant 2(|a| + r_1)r_2 = |A|d(B). \end{aligned}$$

(16): $$d(A \pm B) = d(\langle a \pm b,\ r_1 + r_2\rangle) = 2(r_1 + r_2) = d(A) + d(B).$$

(17): $A \subseteq B \Leftrightarrow |a - b| \leqslant r_2 - r_1$. Следовательно,

$$\begin{aligned} \frac{1}{2}(d(B) - d(A)) &= |r_2| - |r_1| \leqslant |r_2 - r_1| \leqslant |a - b| + |r_1 - r_2| \\ &= q(A,\ B) \leqslant r_2 - r_1 + |r_2 - r_1| = d(B) - d(A). \end{aligned}$$ ■

С теоремой 2.4 связана

Теорема 6. *Операции* $\{+, -, \cdot, :\}$, *заданные на* $R(\mathbb{C})$ *определением* 5.3, *а на* $K(\mathbb{C})$ *— определением* 5.7, *непрерывны.*

Доказательство. Пусть $\{A^{(k)}\}_{k=0}^{\infty}$, $\{B^{(k)}\}_{k=0}^{\infty}$ — последовательности, у которых

$$A^{(k)} = A_1^{(k)} + iA_2^{(k)}, \qquad B^{(k)} = B_1^{(k)} + iB_2^{(k)} \in R(\mathbb{C})$$

и

$$\lim_{k\to\infty} A^{(k)} = A = A_1 + iA_2, \qquad \lim_{k\to\infty} B^{(k)} = B = B_1 + iB_2.$$

Докажем непрерывность умножения:

$$\begin{aligned} \lim_{k\to\infty} A^{(k)}B^{(k)} &= \lim_{k\to\infty} \{A_1^{(k)}B_1^{(k)} - A_2^{(k)}B_2^{(k)} + i\left(A_1^{(k)}B_2^{(k)} + A_2^{(k)}B_1^{(k)}\right)\} \\ &= \lim_{k\to\infty} \left(A_1^{(k)}B_1^{(k)} - A_2^{(k)}B_2^{(k)}\right) + i \lim_{k\to\infty} \left(A_1^{(k)}B_2^{(k)} + A_2^{(k)}B_1^{(k)}\right) \\ &= A_1B_1 - A_2B_2 + i(A_1B_2 + A_2B_1) = AB, \end{aligned}$$

поскольку операции отделения вещественной и мнимой частей комплексного числа непрерывны на $I(\mathbb{C})$.

Подобное доказательство можно провести и для остальных операций на $R(\mathbb{C})$ и $K(\mathbb{C})$. ■

Аналогично (2.22) вводится еще одна бинарная операция на $R(\mathbb{C})$. Теоретико-множественное пересечение A и B из $R(\mathbb{C})$, задаваемое формулой

(18) $$A \cap B = \{c \mid c \in A,\ c \in B\},$$

называется пересечением A и B. Это пересечение принадлежит $R(\mathbb{C})$, если оно не пусто. При $A = A_1 + iA_2$ и $B = B_1 + iB_2$ имеем

(19) $$A \cap B = A_1 \cap B_1 + i(A_2 \cap B_2),$$

где $A_i \cap B_i$ находится по формуле (2.23).

Со следствием 2.12 связано

Следствие 7. *Пусть* $A, B, C, D \in\in R(\mathbb{C})$. *Тогда имеем*

(20) $A \subseteq C,\ B \subseteq D \Rightarrow A \cap B \subseteq C \cap D$ (*монотонность включения*).

При этом операция пересечения непрерывна, если ее результат остается в $R(\mathbb{C})$. ■

Данное следствие можно доказать с помощью следствия 2.12, если последовательно применить последнее к вещественной и мнимой частям.

Замечания. Большая часть приведенных в этой главе результатов взята из диссертации Алефельда [4]. Свойства метрики на $K(\mathbb{C})$ собраны здесь впервые, хотя отчасти они были неявно использованы Гаргантини и Энричи [95]. Эти свойства нам еще потребуются, в частности, в гл. 11—17.

* * *

Литература: [4], [95].

Глава 7
ЛОКАЛИЗАЦИЯ НУЛЕЙ ФУНКЦИЙ ОДНОЙ ВЕЩЕСТВЕННОЙ ПЕРЕМЕННОЙ

В этой главе мы рассмотрим методы локализации нулей вещественной функции f одной вещественной переменной x. Эти методы позволят нам найти множество интервалов наименьшей возможной ширины, таких что каждый интервал содержит один или несколько нулей функции f из заданного интервала $X^{(0)} \in\in I(\mathbb{R})$. Особый интерес представляет случай одного изолированного нуля в $X^{(0)}$. При разработке таких методов мы обратим особое внимание на два обстоятельства. С одной стороны, методы должны быть применимы к широким классам функций при легко проверяемых условиях. С другой стороны, должна быть гарантирована локализация нулей и в том случае, когда рассматриваемые методы реализуются на вычислительной машине, где вместо обычной интервальной арифметики возникает машинная интервальная арифметика, описанная в гл. 4. Поэтому такие методы радикально отличаются от методов для конкретных классов функций [282] и от других процедур общего назначения, описанных в [62] и [70].

Простые реализации таких методов задаются с помощью так называемых методов деления. Это — интервальные варианты метода двоичного поиска или других методов поиска. Мы кратко

опишем такую процедуру. Для нее требуется лишь существование интервального вычисления функции f (см. гл. 3) в интервале $X^{(0)}$. Чтобы улучшить локализацию нулей в $X^{(0)}$, мы делим $X^{(0)}$ пополам точкой

$$m(X^{(0)}) = \frac{1}{2}\left(x_1^{(0)} + x_2^{(0)}\right)$$

на интервалы $U^{(0)}$ и $V^{(0)}$, такие что

$$X^{(0)} = U^{(0)} \cup V^{(0)} = \left[x_1^{(0)},\ m(X^{(0)})\right] \cup \left[m(X^{(0)}),\ x_2^{(0)}\right].$$

Если $0 \in f(U^{(0)})$, то $U^{(0)}$ может содержать нуль функции f, и потому мы повторяем ту же процедуру для $U^{(0)}$. Если $0 \in f(V^{(0)})$, то мы аналогичным образом повторяем процедуру для $V^{(0)}$. Если же мы имеем $0 \notin f(U^{(0)})$ или $0 \notin f(V^{(0)})$, то игнорируем соответствующий подынтервал, так как ввиду (3.1) он не может содержать нуля функции f. Поэтому такой подынтервал исключается из дальнейших вычислений. Описанный итерационный процесс порождает последовательность подынтервалов, содержащихся в $X^{(0)}$ и «подозрительных на наличие нуля функции f». Ширина этих интервалов стремится к нулю, так как она уменьшается вдвое на каждом шаге. Ввиду (3.5) эти постепенно вычисляемые интервалы сходятся к нулям функции f в интервале $X^{(0)}$. См. [37] и [215].

Чтобы предотвратить слишком сильный рост количества «подозрительных» интервалов, мы можем ввести следующую модификацию. На каждом шаге мы исследуем либо только правую половину интервала, либо только левую. Если на некотором шаге мы имеем $0 \notin f(Y)$ для этого полуинтервала Y, то процедура повторяется снова, начиная с интервала $\left[x_1^{(0)}, y_1\right] \subset X^{(0)}$ (соответственно $\left[y_2, x_2^{(0)}\right] \subset X^{(0)}$). Таким образом, мы последовательно вычисляем отдельные нули функции f в порядке справа налево (соответственно слева направо) и не сталкиваемся с проблемой хранения большого количества «подозрительных» интервалов [11, 85].

A. Методы ньютоновского типа

В этом и следующем разделах мы исследуем интервальные модификации метода Ньютона. Для этого рассмотрим непрерывную функцию f, имеющую нуль в данном интервале $X^{(0)} = \left[x_1^{(0)}, x_2^{(0)}\right]$:

$$f(\xi) = 0$$

для некоторого $\xi \in X^{(0)}$. Пусть

(1) $$f\left(x_1^{(0)}\right) < 0 \quad \text{и} \quad f\left(x_2^{(0)}\right) > 0$$

в граничных точках интервала $X^{(0)}$. Пусть, далее, m_1, m_2 — границы разностных отношений

$$(2)\quad 0 < m_1 \leqslant \frac{f(x)-f(\xi)}{x-\xi} = \frac{f(x)}{x-\xi} \leqslant m_2 < \infty, \quad \xi \neq x \in X^{(0)}.$$

Эти границы определяют интервал $M = [m_1, m_2] \in I(\mathbb{R})$. (Аналогичные соображения справедливы, если предположить,

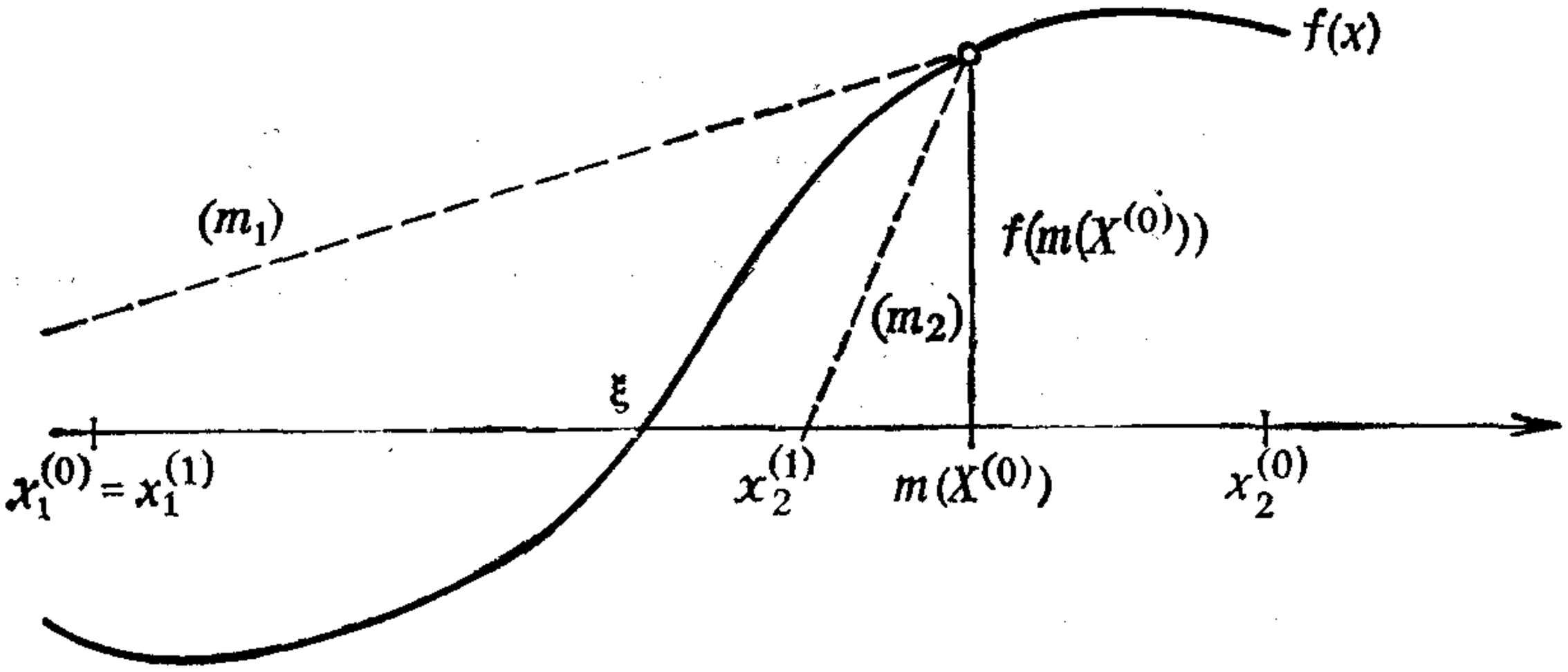


Рис. 1

что $f(x_1^{(0)}) > 0$, $f(x_2^{(0)}) < 0$ и $m_2 < 0$.) Очевидно, что при сделанных предположениях функция f не имеет других корней в $X^{(0)}$.

Начав с исходного локализующего интервала $X^{(0)} \ni \xi$, мы вычисляем итерационно новые интервалы $X^{(k)}$, $k \geqslant 1$, согласно следующей процедуре:

$$(3)\quad X^{(k+1)} = \{m(X^{(k)}) - f(m(X^{(k)}))/M\} \cap X^{(k)}, \quad k \geqslant 0,$$

где $m(X^{(k)}) \in X^{(k)}$.

Этот шаг изображен на рис. 1.

Не пользуясь интервальными операциями, можно записать итерации (3) в виде

$$(3')\quad \begin{aligned} x_1^{(k+1)} &= \begin{cases} \max\{x_1^{(k)}, m(X^{(k)}) - f(m(X^{(k)}))/m_1\}, & \text{если } f(m(X^{(k)})) \geqslant 0, \\ m(X^{(k)}) - f(m(X^{(k)}))/m_2, & \text{если } f(m(X^{(k)})) \leqslant 0, \end{cases} \\ x_2^{(k+1)} &= \begin{cases} m(X^{(k)}) - f(m(X^{(k)}))/m_2, & \text{если } f(m(X^{(k)})) \geqslant 0, \\ \min\{x_2^{(k)}, m(X^{(k)}) - f(m(X^{(k)}))/m_1\}, & \text{если } f(m(X^{(k)})) \leqslant 0. \end{cases} \end{aligned}$$

В обеих формулировках (3) и (3′)

$$m\colon\ I(\mathbb{R}) \to \mathbb{R}$$

обозначает некоторую процедуру выбора вещественного числа m из данного интервала. Часто используют его середину

$$(4)\qquad m(X) = \frac{1}{2}(x_1 + x_2).$$

Теперь мы перечислим важнейшие свойства последовательности итераций $\{X^{(k)}\}_{k=0}^{\infty}$.

Теорема 1. *Пусть f — непрерывная функция, ξ — нуль функции f в интервале $X^{(0)}$, причем имеют место* (1) *и* (2) *для интервала $M = [m_1, m_2]$, $m_1 > 0$. Тогда последовательность $\{X^{(k)}\}_{k=0}^{\infty}$, вычисленная по формулам* (3), *обладает следующими свойствами:*

5) $$\xi \in X^{(k)}, \quad k \geqslant 0,$$

(6) $$X^{(0)} \supset X^{(1)} \supset X^{(2)} \supset \ldots, \quad \text{где} \quad \lim_{k\to\infty} X^{(k)} = \xi,$$

либо эта последовательность стабилизируется через конечное число шагов на точке $[\xi, \xi]$;

(7) $$d(X^{(k+1)}) \leqslant (1 - m_1/m_2)\, d(X^{(k)}).$$

Доказательство. (5): Из (2) и следствия 1.6 получаем

$$\xi = m(X^{(0)}) - \frac{f(m(X^{(0)}))}{f(m(X^{(0)}))/(m(x^{(0)}) - \xi)}$$
$$\in \{m(X^{(0)}) - f(m(X^{(0)}))/M\} \cap X^{(0)} = X^{(1)}.$$

Для $k > 1$ применяем метод математической индукции. (6), (7): Предположим, что $f(m(X^{(k)})) > 0$. Теперь если имеет место $f(m(X^{(k)})) \geqslant (m(X^{(k)}) - x_1^{(k)})\, m_1$, то с помощью (3′) мы получаем

$$d(X^{(k+1)}) = x_2^{(k+1)} - x_1^{(k+1)} = m(X^{(k)}) - f(m(X^{(k)}))/m_2 - x_1^{(k)}$$
$$\leqslant (m(X^{(k)}) - x_1^{(k)}) - (m(X^{(k)}) - x_1^{(k)})\, m_1/m_2$$
$$= (m(X^{(k)}) - x_1^{(k)})\,(1 - m_1/m_2) \leqslant d(X^{(k)})\,(1 - m_1/m_2).$$

Если же имеет место $f(m(X^{(k)})) \leqslant (m(X^{(k)}) - x_1^{(k)})\, m_1$, то из (3′) получаем

$$d(X^{(k+1)}) = x_2^{(k+1)} - x_1^{(k+1)}$$
$$= m(X^{(k)} - f(m(X^{(k)}))/m_2 - m(X^{(k)}) + f(m(X^{(k)}))/m_1$$
$$= \frac{f(m(X^{(k)}))}{m_1}\,(1 - m_1/m_2) \leqslant (m(X^{(k)}) - x_1^{(k)})\,(1 - m_1/m_2)$$
$$\leqslant d(X^{(k)})(1 - m_1/m_2).$$

Случай $f(m(X^{(k)})) < 0$ доказывается аналогично. Если, однако, $f(m(X^{(k)})) = 0$, то $m(X^{(k)}) = \xi$, и потому $d(X^{(k+1)}) = 0$ и $X^{(k+1)} = \xi$, $i \geqslant 1$. Это доказывает (7). Ввиду $m_1 \leqslant m_2$

$$d(X^{(k+1)}) \leqslant \gamma^{k+1} d(X^{(0)}), \quad 0 \leqslant \gamma = (1 - m_1/m_2) < 1,$$

откуда следует

$$\lim_{k\to\infty} d(X^{(k+1)}) = 0.$$

Отсюда и из (5) следует $\lim\limits_{k\to\infty} X^{(k)} = \xi$, если только элементы последовательности не вырождаются в точку: $X^{(k_0+1)} = \xi$, $i \geqslant 1$, для некоторого k_0. Первая часть соотношения (6) следует из формул (3). ■

Таким образом, теорема 1 гарантирует, что в ее предположениях последовательные приближения $X^{(k)}$, $k \geqslant 0$, сходятся к нулю ξ функции f, причем каждый из этих интервалов содержит искомый нуль. Если же мы применяем (3), начав с интервала $X^{(0)}$, такого что $\xi \notin X^{(0)}$, то найдется индекс k_0, для которого пересечение в (3) пусто. Это легко доказать от противного, используя (7) и предположение, что пересечение непусто. Итерации вида (3) можно найти в [316] и [215]. Этот итерационный метод в общей постановке был глубоко исследован в [69]. Он связан также с итерацией монотонных функций для нахождения неподвижной точки (см., например, [103, гл. 5, разд. 2]). Аналогичные процедуры для многочленов были использованы Бауэром [46] еще в 1917 г.

Рассмотрим теперь два уточнения формул (3), возникающие при конкретном выборе точки m. Взяв в качестве m середину интервала, мы получаем следующее утверждение.

Следствие 2. *Если в предположениях теоремы* 1 *сделан выбор*

$$m(X^{(k)}) = \frac{1}{2}(x_1^{(k)} + x_2^{(k)}), \qquad k \geqslant 0,$$

то для последовательности приближений $\{X^{(k)}\}_{k=0}^{\infty}$ *верно неравенство*

$$d(X^{(k+1)}) \leqslant \frac{1}{2}(1 - m_1/m_2)\, d(X^{(k)}), \tag{8}$$

уточняющее (7).

Доказательство. В доказательстве соотношения (7) из теоремы 1 мы имеем при нашем конкретном выборе точки $m(X^{(k)})$, что

$$m(X^{(k)}) - x_1^{(k)} = \frac{1}{2} d(X^{(k)}).$$

Отсюда получаем (8). ■

Таким образом, при выборе середины интервала в качестве $m(X^{(k)})$ нам гарантировано уменьшение ширины локализующего интервала по крайней мере вдвое. Именно этот выбор m используется в методе, рассмотренном в [215].

Рассматривались и другие выборы $m(X^{(k)})$. Например, в [225] используется

$$m(X^{(k)}) = m(X^{(k-1)}) - f(m(X^{(k-1)}))/m_0 \quad \text{для} \quad m_0 \in M$$

и

$$m(X^{(k)}) \in \{x_1^{(k)}, x_2^{(k)}\},$$

если значение $m(X^{(k)})$ из предыдущей формулы не принадлежит интервалу $X^{(k)}$, $k \geqslant 1$.

Интервал M, содержащий разностное отношение (2), требуется и в теореме 1, и в следствии 2. Если функция f непрерывно дифференцируема и $f'(x) \neq 0$ для $x \in X^{(0)}$, то в силу теоремы о среднем можно положить

$$M = [\inf_{y \in X^{(0)}} f'(y), \sup_{y \in X^{(0)}} f'(y)].$$

В общем случае мы можем лишь локализовать этот интервал, например, путем интервального оценивания f', т. е. полагая

$$M = f'(X^{(0)}).$$

Условие $m_1 > 0$ может быть гарантировано в случае необходимости с помощью априорной оценки величины

$$\inf_{y \in X^{(0)}} f'(y).$$

B. Определение оптимального метода

В методе итераций (3), рассмотренном в разд. A, имеется определенная степень свободы при выборе $m(X^{(k)}) \in X^{(k)}$. В зависимости от выбора точек $m(X^{(k)})$ в интервалах $X^{(k)}$ мы получаем различные последовательности локализующих интервалов $\{X^{(k)}\}_{k=0}^{\infty}$. Эти последовательности в общем случае несравнимы почленно в смысле включения интервалов. Поэтому мы, очевидно, должны попытаться найти методы выбора $m(X^{(k)}) \in X^{(k)}$, порождающие последовательности $\{X^{(k)}\}_{k=0}^{\infty}$, в которых ширина отдельных элементов наименьшая возможная. Мы хотим уточнить это требование. Обозначим через $\varphi[X]$ класс функций f, обладающих следующими свойствами для данного интервала $X = [x_1, x_2]$:

(α) $f(x_1) < 0$ и $f(x_2) > 0$;

(β) для интервала $M = [m_1, m_2]$, такого что $m_1 > 0$, имеет место $m_1 \leqslant (f(x) - f(y))/(x - y) \leqslant m_2$ для $x \neq y$, $x, y \in X$.

Очевидно, что любая функция $f \in \varphi[X]$ имеет один и только один корень ξ в интервале X. Поэтому выполнены все условия применимости метода итераций (3) и справедливы все утверждения теоремы 1.

Процесс выбора подходящего $m(X^{(k)}) \in X^{(k)}$ мы разобьем на шаги по образцу [69]. Обозначим последовательные приближения (3) через $\{X^{(k)}\}_{k=0}^{\infty}$. Для вычисления нового приближения $X^{(k+1)}$ нам нужны величины $m(X^{(k)})$ и $f(m(X^{(k)}))$. Если мы зафиксируем величину $m(X^{(k)}) = x \in X^{(k)}$ из $X^{(k)}$, то $X^{(k+1)}$ будет зависеть только от $f(m(X^{(k)}))$. Но это значение функции f может меняться лишь между $y_1^{(k)}$ и $y_2^{(k)}$, ибо $f \in \varphi[X]$ и значения $f(m(X^{(i)}))$, $0 \leqslant i < k$, зафиксированы. Это позволяет нам определить наибольшую возможную ширину

$$\max \{d(X^{(k+1)}) \mid m(X^{(k)}) = x,\ y_1^{(k)} \leqslant f(m(X^{(k)})) \leqslant y_2^{(k)}\}.$$

Это «наихудший» случай для функции $f \in \varphi[X]$.

Теперь мы определим $\tilde{x} = m(X^{(k)}) \in X^{(k)}$ таким образом, чтобы минимизировать эту наибольшую ширину. Иными словами, вычисляется величина

$$\min_{x \in X^{(k)}} \{\max d(X^{(k+1)}) \mid m(X^{(k)}) = x,\ y_1^{(k)} \leqslant f(m(X^{(k)})) \leqslant y_2^{(k)}\}$$

и соответствующее значение $\tilde{x}$ выбирается в качестве $m(X^{(k)})$. Таким образом, определение $m(X^{(k)})$ производится путем минимизации «наихудшего» случая.

Теперь опишем эту процедуру подробно. Не умаляя общности, рассмотрим случай $f(m(X^{(k-1)})) > 0$. На рис. 2 отмечена область, куда могут попасть значения $f(m(X^{(k)}))$ в предположении, что $f \in \varphi[X]$, причем $f(m(X^{(k-1)})) > 0$.

Нижняя ограничивающая прямая (m_1) может отсутствовать на рис. 2, например, в случае, когда $f(m(X^{(i)})) > 0$, $0 \leqslant i < k-1$. Остальные значения $f(m(X^{(i)}))$, $0 \leqslant i < k-1$, не налагают никаких дополнительных ограничений на эту область.

Теперь мы оценим возможные значения $d(X^{(k+1)})$ при заданном $m(X^{(k)}) = x \in X^{(k)}$. Пусть сначала $f(m(X^{(k)})) \geqslant 0$. Для всех значений

$$0 \leqslant f(x) \leqslant (x - x_1^{(k)})\, m_1$$

мы получаем из (3′), что

$$d(X^{(k+1)}) = x - \frac{f(x)}{m_2} - x + \frac{f(x)}{m_1} = f(x)\left(\frac{1}{m_1} - \frac{1}{m_2}\right).$$

Аналогично, для всех значений

$$(x - x_1^{(k)})\, m_1 \leqslant f(x) \leqslant y_2^{(k)}$$

(3′) дает

$$d(X^{(k+1)}) = x - f(x)/m_2 - x_1^{(k)}.$$

(Отметим, что ввиду $x_1^{(k)} = \max\{x^{(k-1)}, \quad m(X^{(k-1)}) - f(m \times (X^{(k-1)}))/m_1\}$ мы всегда имеем $y_2^{(k)} \geqslant (x - x_1^{(k)})\, m_1$.) В первом

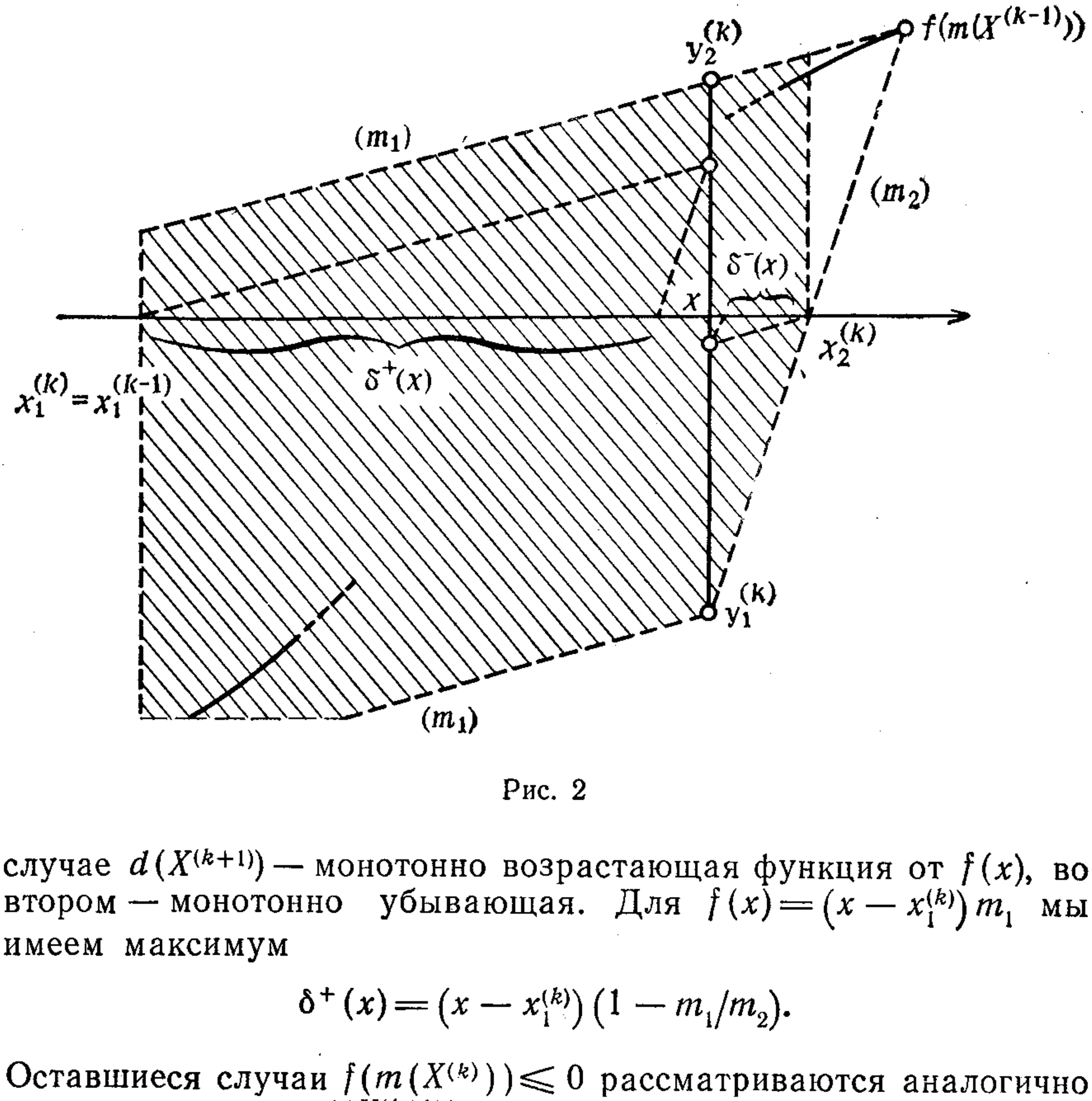


Рис. 2

случае $d(X^{(k+1)})$ — монотонно возрастающая функция от $f(x)$, во втором — монотонно убывающая. Для $f(x) = (x - x_1^{(k)})\, m_1$ мы имеем максимум

$$\delta^+(x) = (x - x_1^{(k)})\,(1 - m_1/m_2).$$

Оставшиеся случаи $f(m(X^{(k)})) \leqslant 0$ рассматриваются аналогично и дают максимум $d(X^{(k+1)})$, равный

$$\delta^-(x) = (x_2^{(k)} - x)\,(1 - m_1/m_2).$$

На рис. 2 показаны два варианта вычисления $X^{(k+1)}$, приводящие к максимальной ширине $\delta^+(x)$ (и соответственно $\delta^-(x)$).

Теперь мы определим минимум

$$\min_{x \in X^{(k)}} \max\{\delta^+(x), \ \delta^-(x)\}.$$

Выражения $\delta^+(x)$ и $\delta^-(x)$ удовлетворяют условию

$$\delta^+\left(\frac{1}{2}(x_1^{(k)}+x_2^{(k)})-t\right)=\delta^-\left(\frac{1}{2}(x_1^{(k)}+x_2^{(k)})+t\right)$$

для $0\leqslant|t|\leqslant\frac{1}{2}(x_2^{(k)}-x_1^{(k)})$. Поэтому минимум равен

$$d(X^{(k+1)})=\frac{1}{2}d(X^{(k)})(1-m_1/m_2)$$

и достигается в точке

$$\tilde{x}=\frac{1}{2}(x_1^{(k)}+x_2^{(k)}).$$

(Ср. со следствием 2.)

Распространим теперь принцип оптимизации, который мы применили к вычислению $X^{(k+1)}$, на определение величин $m(X^{(i)})$, $0\leqslant i<k$. Мы хотим определять значения $m(X^{(0)})=x^{(0)}$, ..., $m(X^{(k)})=x^k$ таким образом, чтобы при этом достигались величины

$$\min_{x^{(0)}\in X^{(0)}}\ \max_{y_1^{(0)}\leqslant f(x^{(0)})\leqslant y_2^{(0)}}\ \cdots\ \min_{x^{(k)}\in X^{(k)}}\ \max_{y_1^{(k)}\leqslant f(x^{(k)})\leqslant y_2^{(k)}} d(X^{(k+1)}).$$

Оказывается, что это просто, так как оптимальная величина $d(X^{(k+1)})$ пропорциональна величине $d(X^{(k)})$ при фиксированном $m(X^{(k-1)})$. Область допустимых значений функции $f(m(X^{(k-1)}))$ определяется лишь значением $f(m(X^{(k-2)}))$. Поэтому можно провести для $m(X^{(k-1)})$ те же рассуждения, что и для $m(X^{(k)})$, и получить оптимальное значение

$$m(X^{(k-1)})=\frac{1}{2}(x_1^{(k-1)}+x_2^{(k-1)}).$$

Соответствующим образом мы получаем результаты для

$$m(X^{(i)}),\quad i=k-2,\ k-3,\ \ldots,\ 0$$

в этом порядке.

Теорема 3. *Пусть метод итерации* (3) *применяется к функции* $f\in\varphi[X]$. *Если используется правило*

$$m(X^{(k)})=\frac{1}{2}(x_1^{(k)}+x_2^{(k)}),\quad 0\leqslant k\leqslant i,\quad i\geqslant 0,$$

то максимальная ширина $d(X^{(i+1)})$ *для функций* $f\in\varphi[X]$ *меньше, чем для любых других выборов точки* $m(X^{(k)})$. *Если* $f\in\varphi[X]$, *то мы имеем*

$$d(X^{(i+1)})\leqslant\frac{1}{2^{i+1}}(1-m_1/m_2)^{i+1}d(X^{(0)}).$$

Существует $g \in \varphi[X]$, для которой в последнем соотношении выполняется равенство. ■

Доказательство этой теоремы содержится в только что проведенном рассуждении. Следует отметить, что $g \in \varphi[X]$ можно выбрать в виде кусочно-линейной функции, проходящей через точки $(m(X^{(k)}), f(m(X^{(k)})))$, $0 \leqslant k \leqslant i$.

C. Квадратично сходящиеся методы

В методе (3) мы используем фиксированную пару m_1, m_2 границ для разностных отношений функции f. Эта процедура соответствует интервальному варианту упрощенного метода Ньютона. Если мы предположим, что f непрерывно дифференцируема и для производной f' имеется интервальная оценка $f'(X)$ (см. гл. 3), то мы можем определить интервальный вариант и для обычного метода Ньютона. Эта новая процедура получается, если мы модифицируем метод (3), заменяя интервал M на интервал

$$M^{(k)} = f'(X^{(k)}) \tag{9}$$

на k-м шаге итерации. Если известны априорные оценки

$$0 < l_1 \leqslant f'(x) \leqslant l_2. \quad x \in X^{(0)},$$

то можно гарантировать оценку $m_1 > 0$ и использовать выражение

$$M^{(k)} = [m_1^{(k)}, m_2^{(k)}] = f'(X^{(k)}) \cap L, \quad L = [l_1, l_2]. \tag{10}$$

Таким образом, мы получаем

$$X^{(k+1)} = \{m(X^{(k)}) - f(m(X^{(k)}))/M^{(k)}\} \cap X^{(k)}, \quad k \geqslant 0, \tag{11}$$

где $m(X^{(k)}) \in X^{(k)}$.

С помощью (11) порождается последовательность интервалов $\{X^{(k)}\}_{k=0}^{\infty}$, для которой можно доказать утверждение, аналогичное теореме 1.

Теорема 4. *Пусть f — непрерывно дифференцируемая функция и f' удовлетворяет в интервале $X^{(0)}$ условиям теоремы* 3.5. *Пусть, далее, для $X^{(0)}$ выполнено* (1), *нуль функции f в $X^{(0)}$ обозначен через ξ и интервалы $M^{(k)}$ определены формулами* (9) *или* (10). *Тогда последовательность $\{X^{(k)}\}_{k=0}^{\infty}$ либо удовлетворяет соотношениям*

$$\xi \in X^{(k)}, \ k \geqslant 0, \tag{5}$$

$$X^{(0)} \supset X^{(1)} \supset X^{(2)} \supset \ldots \quad \text{и} \quad \lim_{k \to \infty} X^{(k)} = \xi, \tag{6}$$

либо стабилизируется на значении $[\xi, \xi]$ *через конечное число шагов.*

$$(12)\qquad d\left(X^{(k+1)}\right) \leqslant \left(1 - m_1^{(k)}/m_2^{(k)}\right) d\left(X^{(k)}\right) \leqslant \beta \left(d\left(X^{(k)}\right)\right)^2, \quad \beta \geqslant 0,$$

т. е. R *— порядок метода итераций* (11) (*см. приложение* A) *удовлетворяет условию*

$$O_R((11), \xi) \geqslant 2.$$

Доказательство. Для $x \in X^{(k)}$ верно

$$\frac{f(x)}{x-\xi} = \frac{f(x)-f(\xi)}{x-\xi} = f'(\eta) \in M^{(k)}, \quad \eta = x + \theta(\xi - x), \quad 0 < \theta < 1.$$

Поэтому аналогичное утверждение для $M^{(k)}$ можно теперь доказать так же, как в теореме 1.

Остается установить (12). Как и в доказательстве теоремы 1, получаем

$$d\left(X^{(k+1)}\right) \leqslant \left(1 - \frac{m_1^{(k)}}{m_2^{(k)}}\right) d\left(X^{(k)}\right) = \frac{m_2^{(k)} - m_1^{(k)}}{m_2^{(k)}}\, d\left(X^{(k)}\right),$$

откуда с помощью 2.9 и теоремы 3.5 имеем

$$d\left(X^{(k+1)}\right) \leqslant \frac{d\left(M^{(k)}\right)}{m_1^{(0)}}\, d\left(X^{(k)}\right) \leqslant \frac{d\left(f'\left(X^{(k)}\right)\right)}{m_1^{(0)}}\, d\left(X^{(k)}\right)$$

$$\leqslant \left(c/m_1^{(0)}\right)\left(d\left(X^{(k)}\right)\right)^2, \quad c/m_1^{(0)} \geqslant 0. \ \blacksquare$$

Метод итераций (11) и теорема 4 соответствуют формулировкам из [215]. Сейчас мы модифицируем этот метод дальше. Заметим, что если $f(m(X^{(k)})) > 0$ (соответственно $f(m(X^{(k)})) < 0$, то искомый нуль ξ должен лежать в интервале $\left[x_1^{(k)}, m\left(X^{(k)}\right)\right]$ (соответственно $\left[m\left(X^{(k)}\right), x_2^{(k)}\right]$. Если $f\left(m\left(X^{(k)}\right)\right) = 0$, то $m\left(X^{(k)}\right) = \xi$, и процесс итерации заканчивается. Поэтому в (11) достаточно положить

$$M^{(k)} = f\left(Y^{(k)}\right) \cap L, \quad L \text{ взято из (10)},$$

где

$$(13)\qquad Y^{(k)} = \begin{cases} \left[x_1^{(k)}, m\left(X^{(k)}\right)\right], & \text{если} \quad f\left(m\left(X^{(k)}\right)\right) > 0, \\ \left[m\left(X^{(k)}\right), x_2^{(k)}\right], & \text{если} \quad f\left(m\left(X^{(k)}\right)\right) < 0, \\ X^{(k)} & \text{в противном случае.} \end{cases}$$

Тогда имеем $f'\left(Y^{(k)}\right) \subseteq f'\left(X^{(k)}\right)$ и $d\left(Y^{(k)}\right) \leqslant d\left(X^{(k)}\right)$, и, таким образом, легче выполнить условие $m_1^{(k)} > 0$. Теорема (4) справедлива и для метода (13).

По повсду выбора точек $m\left(X^{(k)}\right) \in X^{(k)}$ для метода (11) имеем утверждение, аналогичное следствию 2, и можем провести

те же рассуждения, что и разд. B (см. также [126]). Мы не будем углубляться в это.

Таблица 1

$X^{(k)}$	$Y^{(k)}$	$d(X^{(k)})/d(Y^{(k)})$
[1.00000000000, 1.153909281002]	[1.00000000000, 1.231579011696]	0.665
[1.074525733152, 1.075772270022]	[1.018539065305, 1.102153489956]	0.015
[1.07576435 5129, 1.075767749943]	[1.071809768336, 1.084762444669]	3×10^{-4}
[1.075766066086, 1.075766066088]	[1.075647094319, 1.075931180877]	6×10^{-9}
	[1.075766039501, 1.075766097327]	...
	[1.075766066085, 1.075766066090]	...
	[1.075766066085, 1.075766066088]	...

Теперь поясним интервальный метод Ньютона на числовых примерах.

Примеры. (α) Функция

$$f(x) = x^2\left(\frac{1}{3}x^2 + \sqrt{2}\sin x\right) - \sqrt{3}/19$$

имеет нуль ξ в интервале $X^{(0)} = [0.1, 1]$. Производную

$$f'(x) = x\left(\frac{4}{3}x^2 + \sqrt{2}(2\sin x + x\cos x)\right)$$

можно оценить в $X^{(0)}$:

$$l_1 = 0{,}0013 \leqslant f'(x) \leqslant l_2, \quad x \in X^{(0)}.$$

Локализующие интервалы

$X^{(k)}$, $k \geqslant 0$, по формулам (13),

$Y^{(k)}$, $k \geqslant 0$, по формулам (10)

вычислялись на машине в соответствии с (11) до тех пор, пока не переставали происходить изменения. Результаты приведены в табл. 1.

(β) *Многочлен*

$$p(x) = x(x^9 - 1) - 1$$

имеет единственный нуль ξ в интервале $X^{(0)} = [1, 1.5]$. Интервальное вычисление $p'(x)$ его производной

$$p'(x) = 10x^9 - 1$$

дает $0 \notin p'(x)$ для $X \subseteq X^{(0)}$. Итерированные локализующие интервалы

$$X^{(k)}, \quad k \geqslant 0, \quad \text{по формулам (13) с } L = p'(X^{(0)}),$$

$$Y^{(k)}, \quad k \geqslant 0, \quad \text{по формулам (9)}$$

вычислялись в соответствии с (11) с использованием (4). Получились значения, приведенные в табл. 2.

Таблица 2

$X^{(k)}$	$Y^{(k)}$	$d(X^{(k)})/d(Y^{(k)})$
[0.0999999999990, 0.43843885464 33]	[0.0999999999990, 0.5181776715881]	0.809
[0.3382030708107, 0.43843885464 33]	[0.3455588336928, 0.5181776715881]	0.581
[0.391505604 9954, 0.392448494 8316]	[0.3739864679691, 0.4075613703040]	0.028
[0.392378920 6719, 0.392379950 4692]	[0.3922481030413, 0.3925441306206]	0.004
[0.392379507 1350, 0.392379507 1378]	[0.3923794945039, 0.3923795211850]	0.001
	[0.392379507 1350, 0.392379507 1378]	—

На практике определяющим условием для итераций (11) является $m_1 > 0$. Мы показали, что можем добиться его выполнения с помощью (10), используя известную нижнюю оценку l_1 производной $f'(x)$ в интервале $X^{(0)}$. Если такая оценка l_1 неизвестна или если $0 \in f'(X^{(0)})$, то процедуру (11) нельзя даже начать. Поэтому перед началом итераций следует выполнить несколько шагов метода разбиения интервалов, описанного во введении к этой главе. Таким образом, мы найдем интервал $Y^{(0)} \subset X^{(0)}$, для которого верно $0 \notin f'(Y^{(0)})$.

Имеется еще одна модификация метода Ньютона, применимая даже в случае $0 \in f'(X^{(0)})$. Мы наметим сейчас этот метод, работающий даже при наличии нескольких нулей функции f в $X^{(0)}$. Если $0 \notin f'(X^{(0)})$, то этот метод совпадает с методом (11). Предположим поэтому, что $0 \in f(X^{(0)})$. Разобьем $X^{(0)}$ на подын-

тервалы

$$U^{(1)} = [x_1^{(0)},\ m(X^{(0)}) - |f(m(X^{(0)}))|/m_2^{(0)}],$$

$$V^{(1)} = [m(X^{(0)}) + |f(m(X^{(0)}))|/m_2^{(0)},\ x_2^{(0)}],$$

предполагая, что $f(m(X^{(0)})) \neq 0$. Все нули функции f в $X^{(0)}$ должны лежать также и в $U^{(1)} \cup V^{(1)}$. Действительно, нуль $\xi \in X^{(0)}$

Таблица 3

k	$X^{(k)}$
0	[1.8, 2.4]
1	[1.8, 2.07276180774 82]
2	[1.974290005 2812, 2.07276180778 42]
3	[1.994875714 7483, 2.00592154823 53]
4	[1.999988823 4200, 2.00001153900 70]
5	[1.999999999 894, 2.00000000010 7]
6	[2.0, 2.0]

Таблица 4

k	$X^{(k)}$
0	[1.8, 2.4]
1	[1.941953810 8826, 2.05669640504 88]
2	[1.999999997 5872, 2.00011129933 69]
3	[1.999999997 5872, 2.00000000295 95]
4	[2.0, 2.0]

должен удовлетворять неравенству

$$\left| \frac{f(m(X^{(0)}))}{\xi - m(X^{(0)})} \right| \leqslant m_2^{(0)},$$

откуда следует

$$|f(m(X^{(0)}))|/m_2^{(0)} \leqslant |\xi - m(X^{(0)})|$$

и

$$\xi \geqslant m(X^{(0)}) + \frac{|f(m(X^{(0)}))|}{m_2^{(0)}}$$

или

$$\xi \leqslant m(X^{(0)}) - \frac{|f(m(X^{(0)}))|}{m_2^{(0)}}$$

Из последнего неравенства, однако, следует, что $\xi \in U^{(1)} \cup V^{(1)}$. Кроме того, при условии $f(m(X^{(0)})) \neq 0$ имеет место

$$d(U^{(1)} \cup V^{(1)}) = d(X^{(0)}) - 2|f(m(X^{(0)}))|/m_2^{(0)} < d(X^{(0)}).$$

Теперь эту процедуру можно повторить для подынтервалов $U^{(1)}$ и $V^{(1)}$ и т. д. Суммарная ширина этих интервалов стремится к нулю. Если f имеет в $X^{(0)}$ только простые нули, то после некоторого шага итерации все они окажутся в непересекающихся

Таблица 5

k	0	1	2	3
$d_1^{(k)}/d_2^{(k)}$	1	2.37	492.35	3.7×10^6

подынтервалах. Далее, после некоторого шага k процедура превращается в итерацию вида (11). После этого либо подынтервалы стремятся к интервалу, содержащему нуль, либо в какой-то момент получается пустое пересечение.

Вместо того чтобы использовать $M^{(k)} := f'(X^{(k)})$ в (11), в соответствии с (3) мы можем для многочленов использовать интервалы J_1, J_2, J_3 и J_4 из теоремы 3.7 с $y := m(X^{(k)})$ и $X := X^{(k)}$ для локализации производной. Все утверждения теоремы 4 остаются справедливыми. Так как в теореме 3.7 было показано, что J_1 — оптимальная локализация, то для получения наилучшей локализации на каждом шаге разумно использовать именно этот интервал для локализации производной.

В этой связи рассмотрим следующий многочлен из [129].

Пример. Пусть

$$p(x) = x^7 + 3x^6 - 4x^5 - 12x^4 - x^3 - 3x^2 + 4x + 12.$$

Этот многочлен имеет корень ξ в $X^{(0)} = [1.8, 2.4]$. Используя формулы (11), мы находим локализующие интервалы для корня, вычисляя $M^{(k)} := p'(X^{(k)})$ по схеме Горнера. Таблица 3 содержит вычисленные интервалы.

Если $p'(X^{(k)})$ заменено на J_1, мы аналогичным образом получаем табл. 4. В табл. 5 мы приводим для каждого шага итерации частное $d_1^{(k)}/d_2^{(k)}$ от деления ширины первого итерированного интервала на ширину второго итерированного интервала. Вычисления в этом примере произведены на машине CDC 6500 (длина мантиссы — 48 бит) Вычислительного центра Берлинского технического университета.

D. Методы более высоких порядков

Теперь мы разработаем методы более высоких порядков для нахождения в интервале $X^{(0)} = [x_1^{(0)}, x_2^{(0)}]$ нуля ξ строго монотонно возрастающей или монотонно убывающей вещественной функции, обладающей непрерывными производными достаточно высокого порядка. Эти методы всегда сходятся. Идея описываемого построения принадлежит Эрманну [6]. С помощью этой идеи и методов интервального анализа можно разработать методы, для которых обязательно имеет место сходимость. Как и в разд. A, B, C, мы предположим, не умаляя общности, что

$$f(x_1^{(0)}) < 0 \quad \text{и} \quad f(x_2^{(0)}) > 0. \tag{1}$$

Пусть, далее, m_1 и m_2 снова обозначают границы разностных отношений

$$0 < m_1 \leqslant \frac{f(x) - f(\xi)}{x - \xi} = \frac{f(x)}{x - \xi} \leqslant m_2 < \infty, \tag{2}$$
$$\xi \neq x \in X^{(0)}.$$

Границы m_1 и m_2 определяют интервал $M = [m_1, m_2]$. Предполагаем, что функция f $(p+1)$-кратно непрерывно дифференцируема и для $F_i \in I(\mathbb{R})$, $2 \leqslant i \leqslant p+1$, имеет место

$$f^{(i)}(x) \in F_i, \quad x \in X^{(0)}. \tag{14}$$

Интервалы F_i могут быть найдены, например, с помощью интервального оценивания производных функций f на интервале $X^{(0)}$. Если интервальные выражения для производных не определены (например, из-за деления на X при $0 \in X$), то можно подразбить интервал $X^{(0)}$, а затем построить интервалы F_i, взяв объединение оценок, полученных для частичных интервалов.

Рассмотрим теперь следующий метод итераций:

$$\begin{cases} x^{(k)} = m(X^{(k)}) \in X^{(k)}, \\ X^{(k+1,\,0)} = \{x^{(k)} - f(x^{(k)})/M\} \cap X^{(k)}, \\ X^{(k+1,\,i)} = \left\{ x^{(k)} - \dfrac{1}{f'(x^{(k)})} \left[f(x^{(k)}) \right.\right. \\ \qquad + \displaystyle\sum_{\nu=2}^{i} \frac{f^{(\nu)}(x^{(k)})}{\nu!} (X^{(k+1,\,i-1)} - x^{(k)})^{\nu} \\ \qquad \left.\left. + \dfrac{1}{(i+1)!} F_{i+1} (X^{(k+1,\,i-1)} - x^{(k)})^{i+1} \right] \right\} \cap X^{(k+1,\,i-1)}, \\ \qquad\qquad 1 \leqslant i \leqslant p, \\ X^{(k+1)} = X^{(k+1,\,p)}, \qquad k \geqslant 0. \end{cases} \tag{15}$$

Как и в разд. А этой главы, $m(X)$ означает произвольный выбор вещественного числа в интервале X. Формулы, приведенные выше, требуют на каждом шаге вычисления значений $f(x^{(k)})$, $f'(x^{(k)})$, ..., $f^{(p)}(x^{(k)})$ и обладают следующими свойствами.

Теорема 5. *Пусть функция f имеет $(p+1)$ непрерывную производную, $p \geqslant 1$, и пусть для $X^{(0)}$ верно соотношение* (1). *Пусть ξ — нуль функции f в $X^{(0)}$, интервал $M = [m_1, m_2]$ определен соотношением* (2) *и имеет место* (14). *Тогда для итераций* (15) *верно, что*

$$\xi \in X^{(k)}, \quad k \geqslant 0, \tag{16}$$

и либо

$$X^{(0)} \supset X^{(1)} \supset X^{(2)} \supset \ldots \quad \text{и} \quad \lim_{k \to \infty} X^{(k)} = \xi, \tag{17}$$

либо эти итерации стабилизируются через конечное число шагов на точке $[\xi, \xi]$;

$$d(X^{(k+1)}) \leqslant \gamma (d(X^{(k)}))^{p+1} \tag{18}$$

для некоторого $\gamma \geqslant 0$. *Таким образом, согласно теореме 2 из приложения* A, *R-порядок последовательности* $\{X^{(k)}\}_{k=0}^{\infty}$ *не меньше, чем* $p+1$.

Доказательство. (16): Допустим, что $\xi \in X^{(k)}$ для некоторого $k \geqslant 0$. Это верно для $k = 0$ в силу условия теоремы. Как и в теореме 1, мы показываем, что

$$\xi \in X^{(k+1, 0)}.$$

Пусть $\xi \in X^{(k+1, i)}$ для некоторого $i \geqslant 0$. Это верно для $i = 0$. Теперь имеем

$$\xi - x^{(k)} \in X^{(k+1, i)} - x^{(k)}.$$

Из формулы Тейлора мы получаем

$$\begin{aligned} 0 = f(\xi) = f(x^{(k)}) &+ f'(x^{(k)})(\xi - x^{(k)}) \\ &+ \ldots + \frac{1}{(i+1)!} f^{(i+1)}(x^{(k)})(\xi - x^{(k)})^{i+1} \\ &+ \frac{1}{(i+2)!} f^{(i+2)}(\eta_{i+2})(\xi - x^{(k)})^{i+2} \end{aligned}$$

для некоторого η_{i+2}, лежащего между $x^{(k)}$ и ξ. Вместе с монотонностью это дает нам соотношение

$$\xi = x^{(k)} - \frac{1}{f'(x^{(k)})}\left[f(x^{(k)}) + \sum_{\nu=2}^{i+1} \frac{f^{(\nu)}(x^{(k)})}{\nu!}(\xi - x^{(k)})^{\nu} + \frac{f^{(i+2)}(\eta_{i+2})}{(i+2)!}(\xi - x^{(k)})^{i+2}\right]$$

$$\in \left\{x^{(k)} - \frac{1}{f'(x^{(k)})}\left[f(x^{(k)}) + \sum_{\nu=2}^{i+1} \frac{f^{(\nu)}(x^{(k)})}{\nu!}(X^{(k+1,\,i)} - x^{(k)})^{\nu} + \frac{F_{i+2}}{(i+2)!}(X^{(k+1,\,i)} - x^{(k)})^{i+2}\right]\right\} \cap X^{(k+1,\,i)} = X^{(k+1,\,i+1)}.$$

Поэтому мы имеем $\xi \in X^{(k+1,\,i)}$, $0 \leqslant i \leqslant p$, и $\xi \in X^{(k+1)} = X^{(k+1,\,p)}$.

(17): Тем же методом, что и в доказательстве теоремы 1, можно показать, что $X^{(k)} \supset X^{(k+1,\,0)}$, откуда следует $X^{(k)} \supset X^{(k+1)}$, $k \geqslant 0$, так как в формулах (15) мы берем пересечения. Затем, как и в теореме 1, можно показать, что

$$d(X^{(k+1,\,0)}) \leqslant (1 - m_1/m_2)\, d(X^{(k)}).$$

Так как в формулах (15) берутся пересечения, мы получаем

$$d(X^{(k+1)}) \leqslant (1 - m_1/m_2)\, d(X^{(k)}), \qquad k \geqslant 0.$$

Тем же методом, что и в теореме 1, мы получаем сходимость $\lim_{k\to\infty} X^{(k)} = \xi$. Остающаяся часть утверждения (17) доказывается так же, как в теореме 1.

18): Мы имеем $d(X^{(k+1,\,0)}) \leqslant d(X^{(k)})$, откуда

$$d(X^{(k+1,\,1)}) \leqslant d\left(x^{(k)} - \frac{1}{f'(x^{(k)})}\left(f(x^{(k)}) + \frac{1}{2}F_2(X^{(k+1,\,0)} - x^{(k)})^2\right)\right)$$

$$\leqslant \frac{1}{2}\, d\left(\frac{F_2}{f'(x^{(k)})}(X^{(k)} - X^{(k)})^2\right)$$

$$\leqslant \frac{1}{2}\, d\left(\frac{F_2}{M}[-(d(X^{(k)}))^2, (d(X^{(k)}))^2]\right)$$

$$= |F_2/M|(d(X^{(k)}))^2 = \gamma_1 (d(X^{(k)}))^2$$

с константой $\gamma_1 = |F_2/M|$, не зависящей от k.

Предположим, что для некоторого $i \geqslant 1$ мы имеем

$$d(X^{(k+1,\,i)}) \leqslant \gamma_i (d(X^{(k)}))^{i+1},$$

где γ_i не зависит от k. Это только что доказано для $i = 1$. Для $i > 1$ имеем из формул (15) с помощью правил из гл. 2 для вычисления ширины, что

$$d(X^{(k+1,\,i+1)}) \leqslant d\left(\sum_{\nu=2}^{i+1} \frac{f^{(\nu)}(x^{(k)})}{\nu!\, f'(x^{(k)})} (X^{(k+1,\,i)} - x^{(k)})^{\nu}\right.$$

$$\left. + \frac{1}{(i+2)!} \frac{F_{i+2}}{f'(x^{(k)})} (X^{(k+1,\,i)} - x^{(k)})^{i+2}\right)$$

$$\leqslant \sum_{\nu=2}^{i+1} \frac{1}{\nu!} \left| \frac{f^{(\nu)}(x^{(k)})}{f'(x^{(k)})} \right| d((X^{(k+1,\,i)} - x^{(k)})^{\nu})$$

$$+ \frac{1}{(i+2)!}\, d\left(\frac{F_{i+2}}{f'(x^{(k)})} (X^{(k+1,\,i)} - x^{(k)})^{i+2}\right)$$

$$\leqslant \sum_{\nu=2}^{i+1} \frac{1}{\nu!} \left| \frac{F_\nu}{M} \right| \nu\, |X^{(k+1,\,i)} - x^{(k)}|^{\nu-1}\, d(X^{(k+1,\,i)} - x^{(k)})$$

$$+ \frac{1}{(i+2)!}\, d\left(\frac{F_{i+2}}{f'(x^{(k)})} (X^{(k+1,\,i)} - x^{(k)})^{i+2}\right)$$

$$\leqslant \sum_{\nu=2}^{i+1} \frac{1}{(\nu-1)!} \left| \frac{F_\nu}{M} \right| |X^{(k)} - x^{(k)}|^{\nu-1}\, d(X^{(k+1,\,i)})$$

$$+ \frac{1}{(i+2)!}\, d\left(\frac{F_{i+2}}{M} (X^{(k)} - x^{(k)})^{i+2}\right)$$

$$\leqslant \sum_{\nu=2}^{i+1} \frac{1}{(\nu-1)!} \left| \frac{F_\nu}{M} \right| (d(X^{(k)}))^{\nu-1}\, \gamma_i (d(X^{(k)}))^{i+1}$$

$$+ \frac{1}{(i+2)!}\, d\left(\frac{F_{i+2}}{M} [-(d(X^{(k)}))^{i+2}, (d(X^{(k)}))^{i+2}]\right)$$

$$= (d(X^{(k)}))^{i+2} \sum_{\nu=2}^{i+1} \frac{1}{(\nu-1)!} \left| \frac{F_\nu}{M} \right| \gamma_i (d(X^{(k)}))^{\nu-2}$$

$$+ \frac{2}{(i+2)!} \left| \frac{F_{i+2}}{M} \right| (d(X^{(k)}))^{i+2}$$

$$\leqslant \left(\sum_{\nu=2}^{i+1} \frac{1}{(\nu-1)!} \left| \frac{F_\nu}{M} \right| \gamma_i (d(X^{(0)}))^{\nu-2}\right.$$

$$\left. + \frac{2}{(i+2)!} \left| \frac{F_{i+2}}{M} \right|\right) (d(X^{(k)}))^{i+2}$$

$$= \gamma_{i+1} (d(X^{(k)}))^{i+2}$$

с константой

$$\gamma_{i+1} = \gamma_i \sum_{\nu=2}^{i+1} \frac{1}{(\nu-1)!} \left| \frac{F_\nu}{M} \right| (d(X^{(0)}))^{\nu-2} + \frac{2}{(i+2)!} \left| \frac{F_{i+2}}{M} \right|,$$

не зависящей от k. Поэтому соотношение

$$d(X^{(k+1,\,i)}) \leqslant \gamma_i (d(X^{(k)}))^{i+1}$$

справедливо для $1 \leqslant i \leqslant p$.

Далее

$$d(X^{(k+1)}) = d(X^{(k+1,\,p)}) \leqslant \gamma_p (d(X^{(k)}))^{p+1}$$

с константой γ_p, не зависящей от k. Это доказывает формулу (18) для $\gamma = \gamma_p$, что и завершает доказательство теоремы. ■

Теперь мы хотим исследовать случай $p = 1$ несколько подробнее. Для $p = 1$ формулы (15) можно переписать в виде

$$\begin{cases} x^{(k)} = m(X^{(k)}) \in X^{(k)}, \\ X^{(k+1,\,0)} = \{x^{(k)} - f(x^{(k)})/M\} \cap X^{(k)}, \\ X^{(k+1,\,1)} = \left\{x^{(k)} - (1/f'(x^{(k)}))\left(f(x^{(k)}) + \frac{1}{2} F_2 (X^{(k+1,\,0)} - x^{(k)})^2\right)\right\} \cap X^{(k+1,\,0)}, \\ X^{(k+1)} = X^{(k+1,\,1)}, \quad k \geqslant 0. \end{cases}$$

Этот метод имеет те же свойства, что и метод Мура, приведенный в разд. С. Если не считать некоторых дополнительных арифметических операций, он менее трудоемок, чем метод Мура, так как значение функции и производной приходится вычислять в одной и той же точке $x^{(k)}$. Для метода Мура производную приходится вычислять с использованием интервала $X^{(k)}$. Это в общем случае требует больше затрат, чем вычисление значений в одной точке $x^{(k)}$. Если интервал F_2 вычисляется легко, то метод (15) с $p = 1$ предпочтительнее метода Мура. Эти результаты полностью справедливы лишь в теории, когда вычисления считаются точными. Если мы хотим гарантировать локализацию нулей при реализации метода на компьютере, то должны учесть влияние погрешностей округления. Это делается путем реализации всех действий в виде машинных интервальных операций. В частности, нам требуется вычислять $f'(x^{(k)})$, используя машинную интервальную арифметику. В этом случае метод (15) при $p = 1$ требует, если не считать нескольких арифметических операций, того же объема вычислений, что и метод Мура. Так как приходится вычислять еще и интервал F_2, следует, видимо, предпочесть метод Мура, если нужно учитывать погрешности округления.

Мы хотели бы еще упомянуть здесь о том, что в [160, разд. 9, теорема 3] рассмотрен следующий метод:

$$\begin{cases} x^{(k)} = m(X^{(k)}) \in X^{(k)}, \\ X^{(k+1)} = \left\{ x^{(k)} - (1/f'(x^{(k)}))(f(x^{(k)}) \right. \\ \left. + \frac{1}{2} f''(X^{(k)})(X^{(k)} - x^{(k)})^2 \right\} \cap X^{(k)}, \quad k \geqslant 0, \end{cases}$$

в предположении, что f дважды дифференцируема. Мы имеем $\xi \in X^{(k)}$, $k \geqslant 0$. Условия сходимости $\lim\limits_{k\to\infty} X^{(k)} = \xi$ не приводятся. Если метод сходится, то последовательность $d(X^{(k)})$ сходится к нулю квадратично, если $f'(\xi) \neq 0$. По сравнению с методом (15) для $p = 1$ мы должны на каждом шаге производить интервальное оценивание второй производной $f''(X^{(k)})$. Это уменьшает постоянную сходимость, но не улучшает порядок сходимости. (То же верно для метода (15) при $p = 1$, если на каждом шаге заменить постоянный интервал F_2 на $f''(X^{(k)})$.) Использование этого метода на практике, т. е. с учетом погрешностей округления при вычислении $f(x^{(k)})$ и $f'(x^{(k)})$, увеличивает объем вычислений еще примерно на треть. Так как сходимость не гарантирована, этот метод несколько менее привлекателен.

Применяя метод (15), мы должны выбрать конкретный порядок p. Отметим еще, что при определенных предположениях метод (15) оптимален при $p = 2$, т. е. он является методом третьего порядка. Подробности можно установить так же, как это сделано в [6].

E. Интерполяционные методы

Как и в предыдущем разделе мы рассматриваем сходящиеся методы высшего порядка. На этот раз в основу положен хорошо известный интерполяционный метод нахождения нулей функции. Теперь мы изменим его с помощью приемов из интервального анализа таким образом, чтобы всегда получать монотонную локализацию корней. Так же, как и в разд. D, нам нужны интервальные оценки старших производных функции f.

Различные методы определяются с помощью $(n+1)$-элементного множества неотрицательных параметров

$$m_0, m_1, \ldots, m_n.$$

Положим

$$r = \sum_{l=0}^{n} m_l$$

и допустим, что

$$m_0 m_n > 0,$$

откуда следует, что $r > 0$. Мы хотим найти нуль $\xi \in X^{(0)} = [x_1^{(0)}, x_2^{(0)}]$ функции f, которая предполагается дифференцируемой нужное число раз в $X^{(0)}$.

Теперь находятся интервалы H и K, такие что имеет место

$$f'(x) \in H,\ x \in X^{(0)}$$
$$f^{(r)}(x) \in K,\ x \in X^{(0)}. \quad \text{и} \quad 0 \notin H = [h_1, h_2],$$

Чтобы описать очередной $(k+1)$-й шаг итерации, допустим, что мы уже имеем $n+1$ попарно различных приближений к нулю ξ:

$$x^{(k)},\ x^{(k-1)},\ \ldots,\ x^{(k-n)} \text{ в } X^{(0)}$$

и что последний из найденных локализующих интервалов $X^{(k)}$ имеет вид

$$X^{(k)} = [x^{(k)} - \varepsilon^{(k)},\ x^{(k)} + \varepsilon^{(k)}]$$

для некоторого $\varepsilon^{(k)} \geqslant 0$. Допустим еще, что

$$\xi \neq x_1^{(0)} \text{ и } \xi \neq x_2^{(0)}$$

(см., например, [383]). После исполнения описываемых ниже шагов (S1)—(S5) определяется улучшенный локализующий интервал — новая аппроксимация $X^{(k+1)}$.

(S1) *Нахождение единственного интерполяционного многочлена Эрмита*

$$p_k(x) = p_{(m_0, m_1, \ldots, m_n)}(x;\ x^{(k)},\ \ldots,\ x^{(k-n)}),$$

удовлетворяющего интерполяционным условиям

$$p_k^{(j)}\left(x^{(k-i)}\right) = f^{(j)}\left(x^{(k-i)}\right), \quad 0 \leqslant i \leqslant n, \quad 0 \leqslant j \leqslant m_i - 1.$$

(*Мы полагаем* $f^{(0)} = f$, *и если* $m_i = 0$, *то условия в точке* $x^{(k-i)}$ *отсутствуют.*) *Определяется интервал* $Z^{(k)} \subset X^{(k)}$ *по формулам*

$$Z^{(k)} = \begin{cases} [x^{(k)} - \varepsilon^{(k)},\ x^{(k)}], & \text{если} \quad f(x^{(k)})\, h_1 > 0, \\ [x^{(k)},\ x^{(k)} + \varepsilon^{(k)}], & \text{если} \quad f(x^{(k)})\, h_1 < 0. \end{cases}$$

(S2) *Нахождение вещественного корня* $y^{(k)}$ *многочлена* $p_k(x)$ *в интервале*

$$[x^{(k)} - 2\varepsilon^{(k)},\ x^{(k)} + 2\varepsilon^{(k)}] \cap X^{(0)}.$$

Если такого корня нет, то переходим непосредственно к шагу (S5), *положив*

$$\tilde{X}^{(k+1)} := [\tilde{x}_1^{(k+1)},\ \tilde{x}_1^{(k+1)}] = Z^{(k)}.$$

(S3) *Вычисление интервала* $F^{(k)}$, *локализующего значение* $f(y^{(k)})$, *с помощью выражения*

$$F^{(k)} = \frac{K}{r!} \prod_{j=0}^{n} (y^{(k)} - x^{(k-j)})^{m_j}.$$

(S4) *Вычисление улучшенного локализующего интервала по формуле*

$$X^{(k+1)} = \{y^{(k)} - F^{(k)}/H\} \cap Z^{(k)}.$$

(S5) *Нахождение нового приближения*

$$x^{(k+1)} = (\tilde{x}_1^{(k+1)} + \tilde{x}_2^{(k+1)})/2,$$

нового значения

$$\varepsilon^{(k+1)} = (\tilde{x}_2^{(k+1)} - \tilde{x}_1^{(k+1)})/2$$

и нового интервала

$$X^{(k+1)} = [x^{(k+1)} - \varepsilon^{(k+1)},\ x^{(k+1)} + \varepsilon^{(k+1)}] = \tilde{X}^{(k+1)}.$$

(*Если оказывается, что некоторые из точек* $x^{(k+1)}$, $x^{(k)}$, ..., $x^{(k-n+1)}$ *совпали между собой, то можно взять* $x^{(k+1)} \in \tilde{X}^{(k+1)}$ *по формулам*

$$x^{(k+1)} = \begin{cases} \frac{1}{2}(\tilde{x}_1^{(k+1)} + \tilde{x}_2^{(k+1)}) + \tilde{\varepsilon}^{(k)}, & \text{если} \quad f(x^{(k)})\, h_1 > 0, \\ \frac{1}{2}(\tilde{x}_1^{(k+1)} + \tilde{x}_2^{(k+1)}) - \tilde{\varepsilon}^{(k)}, & \text{если} \quad f(x^{(k)})\, h_1 < 0, \end{cases}$$

с подходящим $\tilde{\varepsilon}^{(k)}$, *гарантирующим несовпадение этих точек и соотношение* $x^{(k+1)} \in \tilde{X}^{(k+1)}$. *Такой выбор возможен всегда, когда* $x^{(k-i)} \neq \xi$, $i = 0, 1, \ldots, n$. *Наконец, новое значение* $\varepsilon^{(k+1)}$ *выбирается по формуле*

$$\varepsilon^{(k+1)} = \max\{\tilde{x}_2^{(k+1)} - x^{(k+1)},\ x^{(k+1)} - \tilde{x}_1^{(k+1)}\},$$

что дает

$$X^{(k+1)} = [x^{(k+1)} - \varepsilon^{(k+1)},\ x^{(k+1)} + \varepsilon^{(k+1)}] \supset \tilde{X}^{(k+1)}.)$$

Следует отметить, что определение локализующих интервалов $\{X^{(k)}\}$ не использует перемен знака рассматриваемой функции. Это означает, что новый локализующий интервал будет вычислен всегда, даже если несколько последовательно вычисленных значений функции имеют один и тот же знак. Будет также показано, что шаги (S3) и (S4) могут быть пропущены лишь конечное число раз. Мы не требуем, чтобы локализующий интервал $X^{(i)}$ содержал какое-либо из предыдущих приближений, кроме $x^{(i)}$. Свойства определенного выше алгоритма собраны в следующей теореме.

Теорема 6. *Пусть f — вещественная функция, имеющая нуль ξ, для которого задан локализующий интервал*

$$X^{(0)} = \{x \mid\mid x^{(0)} - x \mid \leqslant \varepsilon^{(0)}\} \ni \xi,$$

такой что

$$\xi \neq x_1^{(0)}, \quad \xi \neq x_2^{(0)} \quad \left(X^{(0)} = \left[x_1^{(0)}, x_2^{(0)}\right]\right).$$

Пусть далее для ξ заданы попарно различные приближения

$$x^{(-n)}, \ x^{(-n+1)}, \ \ldots, \ x^{(0)} \in X^{(0)},$$

и производные функции f удовлетворяют условиям

$$f'(x) \in H, \text{ где } 0 \notin H, \ x \in X^{(0)}$$

и

$$f^{(r)}(x) \in K, \quad x \in X^{(0)}$$

для интервалов H, K и всех x в интервале $X^{(0)}$ при заданных неотрицательных параметрах

$$m_0, \ m_1 \ \ldots, \ m_n, \text{ где } r = \sum_{j=0}^{n} m_j, \quad m_0 m_n > 0.$$

Тогда для итераций, заданных шагами (S1—S5), *верны следующие утверждения:*

(19) $$\xi \in X^{(k)}, \quad k \geqslant 0,$$

(20) $$X^{(0)} \supset X^{(1)} \supset X^{(2)} \supset \ldots \text{ и } \lim_{k \to \infty} X^{(k)} = \xi$$

или после конечного числа шагов последовательность стабилизируется на точке $[\xi, \xi]$.

(21) *R-порядок итераций* (S1)—(S5) (*см. приложение* A) *равен* $O_R((\text{S1})-(\text{S5}), \xi) \geqslant s$, *где* s — *единственный положительный корень многочлена*

$$p(s) = s^{n+1} - \sum_{j=0}^{n} m_j s^{n-j}.$$

Доказательство. (19): В силу (S1) мы имеем $\xi \in Z^{(k)}$. Остаточный член интерполяционной формулы Эрмита дает

$$f(x) = p_k(x) + \frac{f^{(r)}(\eta)}{r!} \prod_{j=0}^{n} (x - x^{(k-j)})^{m_j},$$

где число η лежит в интервале, образованном точками

$$x, \ x^{(k)}, \ \ldots, \ x^{(k-n)},$$

лежащими в $X^{(k-n)}$. Так как $p_k(y^{(k)}) = 0$, получаем, используя свойство включения интервальной арифметики, что

$$f(y^{(k)}) \in \frac{K}{r!} \prod_{j=0}^{n} (y^{(k)} - x^{(k-j)})^{m_j} = F^{(k)}.$$

Так как (S4)— шаг упрощенного метода Ньютона, мы имеем $\xi \in \tilde{X}^{(k+1)}$, и потому $\xi \in X^{(k+1)}$.

(20): Непосредственно следует из определения шага (S1) вместе с (19) и определением шага (S5), где $X^{(k+1)}$ всегда можно выбрать таким образом, что

$$d(X^{(k+1)}) \leqslant c \cdot d(X^{(k)}),$$

где $(1/2) \leqslant c < 1$.

(21): Можно показать, что всегда имеется такой индекс k_1, что для всех шагов с номерами $i \geqslant k_1$ имеется вещественный корень в интервале $[x^{(k)} - 2\varepsilon^{(k)},\ x^{(k)} + 2\varepsilon^{(k)}]$. Доказательство этого утверждения основано на оценке, показывающей, что если $\varepsilon^{(k)}$ достаточно мало, то $p_k(x)$ меняет знак либо между $x^{(k)}$ и $x^{(k)} + 2\varepsilon^{(k)}$, либо между $x^{(k)}$ и $x^{(k)} - 2\varepsilon^{(k)}$. Оно аналогично доказательству из [320] и здесь не приводится. Таким образом, при оценке порядка сходимости мы всегда можем считать, что шаги (S3) и (S4) выполняются. Из (S4) получаем оценку ширины

$$d(\tilde{X}^{(k+1)}) \leqslant d(y^{(k)} - F^{(k)}/H) = d(F^{(k)}/H).$$

Используя (S3) и эту оценку, мы получаем

$$d(\tilde{X}^{(k+1)}) \leqslant \frac{1}{r!} \prod_{i=0}^{n} |y^{(k)} - x^{(k-i)}|^{m_j} d\left(\frac{K}{H}\right).$$

Так как имеет место соотношение

$$y^{(k)} \in [x^{(k-j)} - 2\varepsilon^{(k-j)},\ x^{(k-j)} + 2\varepsilon^{(k-j)}], \quad 0 \leqslant j \leqslant n,$$

мы получаем, наконец, что

$$d(\tilde{X}^{(k+1)}) \leqslant \tilde{c} \prod_{j=0}^{n} d(X^{(k-j)})^{m_j}$$

и

$$d(X^{(k+1)}) \leqslant c \prod_{j=0}^{n} d(X^{(k-j)})^{m_j}.$$

Окончательный результат мы получаем, применив теорему 3 из приложения А. ■

Обсудим теперь еще одно модифицированное семейство методов локализации. Оно получается путем модификации шага (S2) в случае, когда не существует нужного корня $y^{(k)}$. Чтобы

сделать возможным выполнение шага (S4), мы должны тогда определить $y^{(k)}$ иначе. Это делается так.

(S2′) *Нахождение вещественного корня* $y^{(k)}$ *многочлена* $p_k(x)$ *в интервале* $[x^{(k)} - 2\varepsilon^{(k)}, x^{(k)} + 2\varepsilon^{(k)}] \cap X^{(0)}$. *Если такого корня не существует, то полагаем*

$$y^{(k)} = \begin{cases} x^{(k)} - \varepsilon^{(k)}, & \text{если} \quad f(x^{(k)})\, h_1 > 0, \\ x^{(k)} + \varepsilon^{(k)}, & \text{если} \quad f(x^{(k)})\, h_1 < 0. \end{cases}$$

(S3′) *Вычисление локализующего интервала* $F^{(k)}$ *для величины* $f(y^{(k)})$ *с помощью выражения*

$$F^{(k)} = \begin{cases} \dfrac{K}{r!} \prod_{j=0}^{n} (y^{(k)} - x^{(k-j)})^{m_j}, & \text{если} \quad p_k(y^{(k)}) = 0, \\ p_k(y^{(k)}) + \dfrac{K}{r!} \prod_{j=0}^{n} (y^{(k)} - x^{(k-j)})^{m_j} & \text{в противном случае.} \end{cases}$$

Непосредственно очевидно, что теорема 6 истинна без каких-либо изменений и для итераций (S1), (S2′), (S3′), (S4), (S5).

Дальнейшие модификации, приведенные в [383], относятся к шагу (S1).

(S1′) *Построение единственного интерполяционного многочлена Эрмита*

$$p_k(x) = p_{(m_0, m_1, \ldots, m_n)}(x;\ x^{(k)}, \ldots, x^{(k-n)}),$$

удовлетворяющего условиям

$$p_k^{(j)}(x^{(k-i)}) = f^{(j)}(x^{(k-i)}), \quad 0 \leqslant i \leqslant n, \quad 0 \leqslant j \leqslant m_i - 1.$$

(*Мы полагаем* $f^{(0)} = f$ *и считаем, что при* $m_i = 0$ *условия в соответствующей точке отсутствуют.*) *Теперь мы вычисляем интервал* $Z^{(k)} \subset X^{(k)}$ *по формуле*

$$Z^{(k)} = \{x^{(k)} - f(x^{(k)})/H\} \cap X^{(k)}.$$

Очевидно, что теорема 6 верна без изменений также и для итераций (S1′), (S2′), (S3′), (S4), (S5).

Среди рассмотренных выше методов вычисления нулей содержался при $n = 1$ и $m_0 = m_1 = 1$ интервальный вариант метода секущих. В случае $n = 2$ и $m_0 = m_1 = m_2 = 1$ мы получаем интервальный вариант метода Мюллера.

Рассмотрим пример применения интервального метода секущих.

Пример. Рассмотрим функцию

$$f(x) = 2xe^{-5} + 1 - 2e^{-5x}$$

на интервале $X^{(0)} = [0, 1]$. Интервальный метод секущих в применении к этой формуле порождает приближения, приведенные в табл. 6.

Кроме описанного выше интервального метода секущих в [383] описан так называемый интервальный метод regula falsi (ложного основания). Этот метод предполагает примерно те же

Таблица 6

$X^{(k)}$
[0.000000000000, 1.000000000000]
[0.000000000000, 0.270316701135]
[0.1351583505675, 0.1417860581860]
[0.1380542457667, 0.1382588014849]
[0.1382505159257, 0.1382572086567]
[0.1382571542348, 0.1382572086567]
[0.1382571550288, 0.1382571550581]
[0.1382571550553, 0.1382571550581]

свойства функции f, что и интервальный метод секущих. Он также оказывается интерполяционным методом. В противоположность описанным ранее методам он использует разделенные разности Ньютона. Опишем его кратко.

Предположим, что функция f дважды непрерывно дифференцируема в интервале X и имеет в X единственный и притом простой нуль ξ. Далее, мы имеем интервалы H, K, удовлетворяющие условиям

$$f'(x) \in H, \quad x \in H, \quad \text{где} \quad 0 \notin H,$$
$$f''(x) \in K, \quad x \in X.$$

Интервальный *метод regula falsi*, сокращенно RF, описывается теперь так:

$$X^{(0)} = X, \; x^{(0)} = m(X^{(0)}) \quad \text{(середина интервала } X^{(0)}\text{)},$$
$$X^{(1)} = \{x^{(0)} - f(x^{(0)})/H\} \cap X^{(0)},$$
$$\mathrm{RF}\begin{cases} x^{(k)} = m(X^{(k)}) \quad \text{(середина интервала } X^{(k)}\text{)}, \\ Z^{(k+1)} = \{x^{(k)} - f(x^{(k)})/H\} \cap X^{(k)}, \\ X^{(k+1)} = \begin{cases} \left\{x^{(k)} - \dfrac{x^{(k)} - x^{(k-1)}}{f(x^{(k)}) - f(x^{(k-1)})}\left(f(x^{(k)}) + \dfrac{1}{2}K(Z^{(k+1)} - x^{(k)})\right.\right. \\ \left.\left. \times (Z^{(k+1)} - x^{(k-1)})\right\} \cap Z^{(k+1)}, & \text{если} \quad f(x^{(k)}) \neq 0, \\ Z^{(k+1)} & \text{в противном случае} \end{cases} \end{cases}$$

при $k \geqslant 1$.

Свойства алгоритма RF резюмированы в следующем утверждении.

Теорема 7. *Пусть дважды дифференцируемая функция f имеет простой нуль ξ в интервале X. Допустим далее, что выполнены условия*

$$f'(x) \in H, \quad x \in X \quad и \quad 0 \notin H,$$
$$f''(x) \in K, \quad x \in X.$$

Тогда последовательность $\{X^{(k)}\}$, вычисленная согласно процедуре RF, *либо удовлетворяет условиям*

$$\xi \in X^{(k)}, \quad k \geqslant 0, \tag{22}$$

$$X^{(0)} \supset X^{(1)} \supset X^{(2)} \supset \ldots \text{ и } \lim_{k \to \infty} X^{(k)} = \xi, \tag{23}$$

либо стабилизируется на точке $[\xi, \xi]$ через конечное число шагов.

Для некоторого $\gamma \geqslant 0$ имеет место соотношение

$$d(X^{(k+1)}) \leqslant \gamma d(X^{(k)}) d(X^{(k-1)}), \tag{24}$$

т. е. (см. приложение A)

$$O_R((\mathrm{RF}), \xi) \geqslant \frac{1}{2}(1 + \sqrt{5}).$$

Доказательство. (22): Доказывается с помощью математической индукции по k. Для $k = 1$ утверждение $\xi \in X^{(1)}$ очевидно. Допустим теперь, что для фиксированного k мы имеем $x^{(k)} = m(X^{(k)}) \neq \xi$ и $\xi \in X^{(k)}$. Тогда имеет место $\xi \in Z^{(k+1)}$ и $x_k \neq x_{k-1}$. Рассмотрение интерполяционной формулы Ньютона показывает, что

$$f(\xi) = f(x^{(k)}) + \frac{f(x^{(k)}) - f(x^{(k-1)})}{x^{(k)} - x^{(k-1)}}(\xi - x^{(k)})$$

$$+ \frac{1}{2} f''(\eta)(\xi - x^{(k)})(\xi - x^{(k-1)}),$$

где η находится в интервале, образованном точками $x^{(k)}$, $x^{(k-1)}$ и ξ. Из $f(\xi) = 0$ следует, что

$$\xi = x^{(k)} - \frac{x^{(k)} - x^{(k-1)}}{f(x^{(k)}) - f(x^{(k-1)})}\left(f(x^{(k)}) + \frac{f''(\eta)}{2}(\xi - x^{(k)})(\xi - x^{(k-1)})\right).$$

Из предположения $f''(\eta) \in K$ и того, что $\xi \in Z^{(k+1)}$, а также свойства включения для интервальной арифметики следует, что

$$\xi \in x^{(k)} - \frac{x^{(k)} - x^{(k-1)}}{f(x^{(k)}) - f(x^{(k-1)})}\left(f(x^{(k)}) + \frac{K}{2}(Z^{(k+1)} - x^{(k)})(Z^{(k+1)} - x^{(k-1)})\right),$$

откуда в силу $\xi \in Z^{(k+1)}$ мы получаем $\xi \in X^{(k+1)}$.

(23): Следует из построения $Z^{(k)}$ и того, что $X^{(k)} \subset Z^{(k)}$.
(24): Пусть $x^{(k)} \neq \xi$. Тогда мы получаем

$$d(X^{(k+1)}) \leqslant d\left(x^{(k)} - \frac{x^{(k)} - x^{(k-1)}}{f(x^{(k)}) - f(x^{(k-1)})}\left(f(x^{(k)}) + \frac{1}{2} K (Z^{(k+1)} - x^{(k)})(Z^{(k+1)} - x^{(k-1)})\right)\right)$$

$$= d\left(-\frac{x^{(k)} - x^{(k-1)}}{f(x^{(k)}) - f(x^{(k-1)})} \cdot \frac{1}{2} K (Z^{(k+1)} - x^{(k)})(Z^{(k+1)} - x^{(k-1)})\right).$$

Из

$$\frac{x^{(k)} - x^{(k-1)}}{f(x^{(k)}) - f(x^{(k-1)})} = (f'(\tau))^{-1} \in \frac{1}{H}, \quad \tau \in X$$

следует, что

$$d(X^{(k+1)}) \leqslant \frac{1}{2} d((K/H)(Z^{(k+1)} - x^{(k)})(Z^{(k+1)} - x^{(k-1)})).$$

Ввиду

$$X^{(k-1)} \supset X^{(k)} \supset Z^{(k+1)} \ni X^{(k+1)},$$

получаем, наконец,

$$d(X^{(k+1)}) \leqslant \frac{1}{2} d((K/H)(X^{(k)} - x^{(k)})(X^{(k-1)} - x^{(k-1)})).$$

Интервал $A \in I(\mathbb{R})$ с центром $m(A)$ удовлетворяет условию

$$A - m(A) = [-d(A)/2,\ d(A)/2],$$

откуда следует соотношение

$$(X^{(k)} - x^{(k)})(X^{(k-1)} - x^{(k-1)}) = [-d(X^{(k)})\, d(X^{(k-1)})/4,\ d(X^{(k)})\, d(X^{(k-1)})/4].$$

В применении к предыдущему неравенству оно дает

$$d(X^{(k+1)}) \leqslant \frac{1}{2} d((K/H)[-d(X^{(k)})\, d(X^{(k-1)})/4,\ d(X^{(k)})\, d(X^{(k-1)})/4]),$$

откуда получается

$$d(X^{(k+1)}) \leqslant \frac{1}{2} |K/H| \cdot 2 \cdot \frac{1}{4} d(X^{(k)})\, d(X^{(k-1)})$$

$$= \gamma\, d(X^{(k)})\, d(X^{(k-1)}),$$

где $\gamma = \frac{1}{4} |K/H|$. Из теоремы 3 (приложение A) мы получаем затем требуемое неравенство для R-порядка. ■

Используя для данной функции f разделенные разности более высокого порядка, мы можем построить новые методы, порядок сходимости которых тоже лежит между 1 и 2 и которые требуют лишь по одному новому значению функции на каждый шаг итерации. (Дальнейшее обсуждение этого вопроса см. в [383].)

Можно также *построить интервальные варианты методов из* [510], используя интервальный метод regula falsi (RF). Эти методы имеют порядок сходимости выше первого, хотя они используют значения только самой функции f. Мы опишем здесь такой метод, представляющий собой прямое обобщение интервального метода regula falsi.

Снова предполагается, что функция f имеет простой нуль ξ в интервале X, а интервалы H, K удовлетворяют условиям теоремы 7. Задан параметр p — целое число $\geqslant 1$. Теперь *параметрический метод* regula falsi, короче p-RF, формулируется следующим образом:

$$
p\text{-RF}\left\{\begin{array}{l}
X^{(0)} = X, \quad x^{(0)} = m(X^{(0)}), \\
X^{(1)} = \{x^{(0)} - f(x^{(0)})/H\} \cap X^{(0)}.
\end{array}\right.
$$

Для $k \geqslant 1$ вычисляем приближения по следующим формулам:

$$
p\text{-RF}\left\{\begin{array}{l}
x^{(k)} = m(X^{(k)}) \text{ (середина интервала } X^{(k)}), \\
X^{(k+1,\,0)} = \{x^{(k)} - f(x^{(k)})/H\} \cap X^{(k)}, \\
X^{(k+1,\,1)} = \begin{cases}
\left\{x^{(k)} - \dfrac{x^{(k)} - x^{(k-1)}}{f(x^{(k)}) - f(x^{(k-1)})}\left(f(x^{(k)})\right.\right. \\
\left.\left. + \dfrac{1}{2} K (X^{(k+1,\,0)} - x^{(k)})(X^{(k+1,\,0)} - x^{(k-1)})\right)\right\} \\
\cap X^{(k+1,\,0)}, \quad \text{если} \quad f(x^{(k)}) \neq 0, \\
X^{(k+1,\,0)} \qquad \text{в противном случае.}
\end{cases}
\end{array}\right.
$$

После этого производим вычисления для $i = 2, 3, \ldots, p$ (только для $p > 1$).

$$
p\text{-RF}\left\{\begin{array}{l}
z^{(i)} = m(X^{(k+1,\,i-1)}), \\
X^{(k+1,\,i)} = \begin{cases}
\left\{z^{(i)} - \dfrac{z^{(i)} - x^{(k)}}{f(z^{(i)}) - f(x^{(k)})}\left(f(z^{(i)})\right.\right. \\
\left.\left. + \dfrac{1}{2} K (X^{(k+1,\,i-1)} - z^{(i)})(X^{(k+1,\,i-1)} - x^{(k)})\right)\right\} \\
\cap X^{(k+1,\,i-1)}, \text{ если } f(x^{(k)}) \neq 0, \\
X^{(k+1,\,i-1)} \qquad \text{в противном случае,}
\end{cases} \\
X^{(k+1)} = X^{(k+1,\,p)}.
\end{array}\right.
$$

Метод p-RF обладает следующими свойствами.

Теорема 8. *Пусть функция f дважды непрерывно дифференцируема в интервале X и имеет там нуль ξ. Пусть выполнены условия*

$$
f'(x) \in H, \quad x \in H, \quad \text{где} \quad 0 \notin H,
$$
$$
f''(x) \in K, \quad x \in X.
$$

Тогда последовательность $\{X^{(k)}\}$, *вычисленная по формулам* p-RF, *удовлетворяет для* $p \geqslant 1$ *условиям*

$$\xi \in X^{(k)}, \quad k \geqslant 0, \tag{25}$$

$$X^{(0)} \supset X^{(1)} \supset X^{(2)} \supset \ldots, \text{ где } \lim_{k \to \infty} X^{(k)} = \xi, \tag{26}$$

или стабилизируется через конечное число шагов на точке $[\xi, \xi]$. (27) *Для некоторого* $\gamma \geqslant 0$ *справедлива оценка*

$$d(X^{(k+1)}) \leqslant \gamma d(X^{(k)})^p \, d(X^{(k-1)})$$

и

$$O_R((p\text{-RF}), \xi) \geqslant \frac{1}{2}(p + \sqrt{p^2+4}).$$

Доказательство. При $p = 1$ теорема 8 сводится к теореме 7, поэтому мы предполагаем далее, что $p \geqslant 2$.

Установим соотношение (25). Мы наметим доказательство методом математической индукции. Доказываемое утверждение очевидно для $k = 0, 1$. Мы допустим, что оно верно для фиксированного k и докажем его для $k+1$. Из нашего индукционного предположения следует, что

$$\xi \in X^{(k+1, 0)} \text{ и } X^{(k+1, 0)} = [\xi, \xi] \text{ для } x^{(k)} = \xi.$$

В случае когда $x^{(k)} = \xi$, мы имеем $f(x^{(k)}) = 0$, откуда следует, что

$$X^{(k+1, i)} = [\xi, \xi]$$

для $i = 2, 3, \ldots, p$. Поэтому имеем

$$\xi \in X^{(k+1)} = X^{(k+1, p)} = [\xi, \xi].$$

Если теперь $x^{(k)} \neq \xi$, то $f(x^{(k)}) \neq 0$ и получаем соотношение

$$0 = f(\xi) = f(x^{(k)}) + \frac{f(x^{(k)}) - f(x^{(k-1)})}{x^{(k)} - x^{(k-1)}}(\xi - x^{(k)}) + \frac{1}{2} f''(\eta)(\xi - x^{(k)})(\xi - x^{(k-1)}),$$

где $x^{(k)} \neq x^{(k-1)}$, $x^{(k-1)} \neq \xi$.

Тем же методом, что и в доказательстве теоремы 7, отсюда получаем, что

$$\xi \in x^{(k)} - \frac{x^{(k)} - x^{(k-1)}}{f(x^{(k)}) - f(x^{(k-1)})} \left(f(x^{(k)}) + \frac{1}{2} K (X^{(k+1, 0)} - x^{(k)}) \times (X^{(k+1, 0)} - x^{(k-1)}) \right)$$

и окончательно $\xi \in X^{(k+1, 1)}$. Остается показать, что

$$\xi \in X^{(k+1, i)}, \quad i = 2, 3, \ldots, p.$$

Для $x^{(k)} \neq \xi$ и $z^{(i)} \neq \xi$ снова показываем, используя остаточный член интерполяционной формулы Ньютона, что из $\xi \in X^{(k+1,\, i-1)}$ всегда следует $\xi \in X^{(k+1,\, i)}$.

Это снова очевидно для $x^{(k)} \neq \xi$ и $z^{(i)} = \xi$. Так как

$$\xi \in X^{(k+1,\, 1)},$$

мы получаем

$$\xi \in X^{(k+1,\, i)}, \quad i = 1,\ 2,\ \ldots,\ p,$$

а потому и

$$\xi \in X^{(k+1)} = X^{(k+1,\, p)}.$$

Соотношение (26) немедленно следует из формул, по которым вычисляется $X^{(k+1,\, 0)}$. Мы не приводим здесь элементарного обоснования этого факта.

(27): Не умаляя общности, мы предположим, что $x^{(k)} \neq \xi$. Тогда из определения процедуры p-RF немедленно следует, что

$$d(X^{(k+1,\, 0)}) < \frac{1}{2}\, d(X^{(k)}).$$

Аналогично тому, как доказано утверждение (24) в теореме (7), мы получаем, что

$$d(X^{(k+1,\, 1)}) \leqslant \frac{1}{4}\, |K/H|\, d(X^{(k)})\, d(X^{(k-1)})$$

и аналогичным образом

$$d(X^{(k+1,\, i)} \leqslant \frac{1}{4}\, |K/H|\, d(X^{(k+1,\, i-1)})\, d(X^{(k)}), \quad i = 2,\ 3,\ \ldots,\ p.$$

Эта простая рекурсия дает соотношение

$$d(X^{(k+1,\, i)}) \leqslant \beta_i\, d(X^{(k)})^i\, d(X^{(k-1)}),$$

$$\beta_i = \left(\frac{1}{4}\, |K/H|\right)^i, \quad i = 1,\ 2,\ \ldots,\ p.$$

Используя равенство $X^{(k+1)} = X^{(k+1,\, p)}$, получаем соотношение

$$d(X^{(k+1)}) \leqslant \beta_p\, d(X^{(k)})^p\, d(X^{(k-1)}),$$

из которого снова получаем требуемое неравенство для R-порядка с помощью теоремы 3 из приложения А. ■

Замечания. Были предложены и исследовались различные модификации методов (3) и (11). Вариант, приведенный в [407], был исследован в [268]. Процедура, аналогичная (3′), предложена в [162], где она была использована для уточнения оценок всех нулей в предписанном интервале $[x_1^{(0)},\ x_2^{(0)}]$. Уточнения нижней границы $x_1^{(k+1)}$ вычисляются отдельно от уточнений верхней границы $x_2^{(k+1)}$. Для этой процедуры нам требуются как значение функции $f(x_1^{(k)})$ (соответственно $f(x_2^{(k)})$), так и оценка $m = |M|$ величины $|f'(x)|$ для $x \in [x_1^{(k)},\ x_2^{(k)}]$. Тогда последовательность

нижних границ $x_1^{(0)}, x_1^{(1)}, \ldots$ сходится к наименьшему нулю, принадлежащему интервалу $[x_1^{0)}, x_2^{(0)}]$. Соответственно последовательность верхних границ сходится к наибольшему нулю функции $f(x)$ в $[x_1^{(0)}, x_2^{(0)}]$. Аналогичный метод для многочленов был разработан в [46].

В [44] использовано неявное представление метода (3) (соответственно (11)). Функция $f(x)$ локализуется интервальным выражением

$$I(x) = f(x^{(k)}) + (x - x^{(k)}) M^{(k)}, \quad x^{(k)} \in [x_1^{(k)}, x_2^{(k)}] = X^{(k)},$$

где $M^{(k)}$ определяется аналогично (9). Тогда интервалы $[x_1^{(k+1)}, x_2^{(k+1)}]$ вычисляются с помощью требования

$$[x_1^{(k+1)}, x_2^{(k+1)}] := \{x \mid x \in X^{(k)}, 0 \notin I(x)\} \cap X^{(k)}.$$

При этом возникают различные случаи в зависимости от того, верно ли, что $0 \in M^{(k)}$. Этот метод сходится при определенных условиях, и если имеется несколько нулей, то возникает несколько подпоследовательностей. Рекурсивная процедура для этого случая найдена в [45]. Этот метод, аналогичный методу, предложенному в явном виде в [6], был эскизно описан в конце разд. C. Этот метод в замкнутом виде можно также найти в диссертации Каухера [148]. Он был также недавно рассмотрен в [398] и применен к нахождению нулей производной дважды непрерывно дифференцируемой функции на интервале, что используется при нахождении глобального минимума функции.

Приведенные в теореме 1 (соответственно в теореме 4) результаты о методе (3) (соответственно о методе (11)) можно обобщить следующим образом. Предполагается, что для функции f на интервале $X^{(0)}$ имеется интервал M, такой что $0 \notin M$ и

$$\frac{f(x) - f(y)}{x - y} \in M \quad \text{для} \quad x, y \in X^{(0)}, \quad x \neq y. \tag{28}$$

Если

$$X^{(1)} = m(X^{(0)}) - f(m(X^{(0)}))/M \in X^{(0)},$$

то найдется $\xi \in X^{(1)}$, такое что $f(\xi) = 0$. Это будет в том случае, если мы предположим, не умаляя общности, что $f(m(X^{(0)})) > 0$, $m_1 > 0$, а затем $f(x_1^{(1)}) > 0$. Из этого предположения получается противоречие, если рассмотреть

$$x_1^{(1)} > m(X^{(0)}) - \frac{f(m(X^{(0)}))}{(f(m(X^{(0)})) - f(x_1^{(1)}))/(m(X^{(0)}) - x_1^{(1)})} \geqslant x_1^{(1)}.$$

Поэтому мы должны иметь $f(x_1^{(1)}) \leqslant 0$, откуда следует, что $X^{(0)}$, а тогда в силу теоремы 1 также и $X^{(1)}$ содержит нуль ξ. Такого рода теоремы существования найдены, например, в [143] и [306]. См. также лемму 11.10.

Мы хотим теперь, используя средства интервальной арифметики, дать короткое доказательство того, что уравнение $f(x) = 0$ имеет корень в интервале $X^{(0)} = [x^{(0)} - r,\ x^{(0)} + r]$. Предположим, что f дважды непрерывно дифференцируема и что

$$|f''(x)| \leqslant \gamma, \quad x \in X^{(0)}.$$

Допустим еще, что $f'(x^{(0)}) \neq 0$, и положим

$$\left|\frac{f(x^{(0)})}{f'(x^{(0)})}\right| = \eta, \quad \left|\frac{1}{f'(x^{(0)})}\right| = \beta.$$

Тогда для любого $y \in X^{(0)}$, $x^{(0)} \neq y$ и некоторого θ, лежащего между $x^{(0)}$ и y, имеем

$$\frac{f(x^{(0)}) - f(y)}{x^{(0)} - y} = f'(x^{(0)}) + \frac{1}{2} f''(\theta)(y - x^{(0)})$$

$$\in f'(x^{(0)}) + \frac{1}{2}\gamma[-r, r] =: M^{(0)}.$$

Если теперь $0 \notin M^{(0)}$ и

$$X^{(1)} = x^{(0)} - f(x^{(0)})/M^{(0)} \subset X^{(0)},$$

то мы показываем тем же методом, что и раньше, что имеется нуль $\xi \in X^{(1)} \subseteq X^{(0)}$. Требование $X^{(1)} \subseteq X^{(0)}$ выполнено тогда и только тогда, когда

$$\frac{1}{2}\beta\gamma r^2 - r + \eta \leqslant 0.$$

Это эквивалентно неравенству

$$\beta\gamma\eta \leqslant \frac{1}{2} \tag{29}$$

вместе с

$$(1 - \sqrt{1 - 2\beta\gamma\eta})/\beta\gamma \leqslant r \leqslant (1 + \sqrt{1 - 2\beta\gamma\eta})/\beta\gamma. \tag{30}$$

Неравенства (29) и (30) гарантируют выполнение условий теоремы Канторовича о существовании нуля в интервале $X^{(0)}$ (см. [250, теорема 22.4]).

Эти вопросы будут снова обсуждаться в гл. 19. Если $X^{(1)} \cap X^{(0)} = \varnothing$, то f не имеет нулей в $X^{(0)}$. Условие $X^{(1)} \cap X^{(0)} = \varnothing$ выполнено тогда и только тогда, когда

$$\frac{1}{2}\beta\gamma r^2 + r - \eta < 0,$$

т. е. при $\eta \neq 0$ тогда и только тогда, когда

$$0 \leqslant r < (-1 + \sqrt{1 + 2\beta\gamma\eta})/\beta\gamma.$$

Это утверждение может быть аналогичным образом использовано для доказательства теорем об исключении в банаховых пространствах. См. [347].

Теорема Канторовича использована в [274] для локализации корня функции f с учетом влияния ошибок округления. Ср. также с [273].

* * *

Литература: [4], [6], [11], [12], [29], [37], [44], [45], [46], [62], [69], [70], [74], [75], [79], [85], [88], [94], [103], [116], [119], [126], [128], [129], [134], [150], [160], [162], [215], [225], [245], [246], [250], [268], [273], [274], [282], [313], [314], [316], [320], [347], [382], [383], [384], [397], [398], [407], [510].

Глава 8

МЕТОДЫ ОДНОВРЕМЕННОЙ ЛОКАЛИЗАЦИИ ВЕЩЕСТВЕННЫХ КОРНЕЙ МНОГОЧЛЕНОВ

В этой главе мы рассмотрим интервальные методы ньютоновского типа для вычисления интервалов, локализующих все вещественные корни вещественного многочлена. Сначала рассматривается случай, когда все корни вещественны. Комплексные корни рассматриваются в гл. 9. Для случая, когда все корни вещественные и простые, мы строим короткошаговый метод, сходящийся быстрее, чем квадратично. В качестве приложения мы используем этот метод для вычисления всех собственных значений симметрической трехдиагональной матрицы.

Дан вещественный многочлен

$$p(x) = a^{(n)}x^n + a^{(n-1)}x^{n-1} + \ldots + a^{(0)}. \tag{1}$$

и мы предполагаем в дальнейшем, что

$$a^{(n)} = 1.$$

Далее предполагается, что этот многочлен имеет n вещественных корней $\xi^{(1)}, \xi^{(2)}, \ldots, \xi^{(n)}$, которые собраны в вектор $(\xi^{(i)})$, причем кратные корни выписаны столько раз, какова их кратность. Предполагается, что для всех корней известны локализующие интервалы

$$\xi^{(i)} \in X^{(0,\,i)} = \left[x_1^{(0,\,i)}, x_2^{(0,\,i)}\right], \quad 1 \leqslant j \leqslant n.$$

Предположим сначала, что все эти локализующие интервалы попарно не пересекаются, т. е.

$$X^{(0,\,j)} \cap X^{(0,\,k)} = \varnothing, \qquad 1 \leqslant j < k \leqslant n. \tag{2}$$

Многочлен $p(x)$ можно записать в виде

$$p(x) = \prod_{j=1}^{n} (x - \xi^{(j)})$$

или

$$p(x) = (x - \xi^{(i)}) \prod_{j=1,\ j \neq i}^{n} (x - \xi^{(j)}),$$

откуда следует

$$\xi^{(i)} = x - p(x) \Big/ \prod_{j=1,\ j \neq i}^{n} (x - \xi^{(j)}).$$

Если мы выберем $x = x^{(0,\,i)} \in X^{(0,\,i)}$, то получим, что

$$0 \notin \prod_{j=1,\ j \neq i}^{n} (x^{(0,\,i)} - X^{(0,\,j)}),$$

и из (1.9) следует, что

$$\xi^{(i)} \in X^{(1,\,i)} = \left\{ x^{(0,\,i)} - p(x^{(0,\,i)}) \Big/ \prod_{j=1,\ j \neq i}^{n} (x^{(0,\,i)} - X^{(0,\,j)}) \right\} \cap X^{(0,\,i)}.$$

Таким образом, интервальное выражение, стоящее в правой части этого равенства, задает новый локализующий интервал $X^{(1,\,i)}$, для которого верно

$$\xi^{(i)} \in X^{(1,\,i)} \subseteq X^{(0,\,i)}.$$

Это соотношение порождает следующую итерационную схему:

$$X^{(k+1,\,i)} = \left\{ x^{(k,\,i)} - p(x^{(k,\,i)}) \Big/ \prod_{j=1,\ j \neq i}^{n} (x^{(k,\,i)} - X^{(k,\,j)}) \right\} \cap X^{(k,\,i)}, \tag{3}$$

где

$$x^{(k,\,i)} \in X^{(k,\,i)}, \qquad 1 \leqslant i \leqslant n, \qquad k \geqslant 0.$$

Для интервального выражения в знаменателе вводится сокращение

$$Q^{(k,\,i)} = \prod_{j=1,\ j \neq i}^{n} (x^{(k,\,i)} - X^{(k,\,j)}).$$

Итерационная схема (3) задает полношаговый метод локализации корней многочлена $\xi^{(i)}$, $1 \leqslant i \leqslant n$. Если в $Q^{(k,\,i)}$ мы всегда используем последнее вычисленное значение локализующих интервалов, то получаем

$$R^{(k,\,i)} = \prod_{j=1}^{i-1} (x^{(k,\,i)} - X^{(k+1,\,j)}) \prod_{j=i+1}^{n} (x^{(k,\,i)} - X^{(k,\,j)}),$$

что приводит к соответствующей короткошаговой итерации. Теперь мы хотим провести для этой короткошаговой итерации такие же рассуждения, как для метода (7.13). Локализующий интервал $X^{(k+1,\,i)}$ урезается до $Y^{(k+1,\,i)}$ в зависимости от знаков выражений $p(x^{(k+1,\,i)})$ и $R^{(k,\,i)}$. Функция sign для интервалов определяется соотношением

$$\text{(4)}\qquad \operatorname{sign}(X)=\begin{cases} 1, & \text{если} \quad x_1>0,\\ -1, & \text{если} \quad x_2<0,\\ 0 & \text{в противном случае.}\end{cases}$$

Интервалы $Y^{(k+1,\,i)}$, также содержащие корни $\xi^{(i)}$, определяются тогда следующим образом:

$$Y^{(k+1,\,i)}=\begin{cases}\left[x_1^{(k+1,\,i)},\ x^{(k+1,\,i)}\right], & \text{если } \operatorname{sign}(R^{(k,\,i)})\operatorname{sign}(p(x^{(k+1,\,i)}))>0,\\ \left[x^{(k+1,\,i)},\ x_2^{(k+1,\,i)}\right], & \text{если } \operatorname{sign}(R^{(k,\,i)})\operatorname{sign}(p(x^{(k+1,\,i)}))<0,\\ X^{(k+1,\,i)} & \text{в противном случае.}\end{cases}$$

Заметим, что всегда имеет место соотношение

$$\operatorname{sign}(R^{(0,\,i)})=\operatorname{sign}(R^{(1,\,i)})=\ \ldots, \qquad 1\leqslant i\leqslant n.$$

Используя только что введенные новые локализующие интервалы, мы вычисляем теперь новое значение знаменателя с помощью выражения

$$S^{(k+1,\,i)}=\prod_{j=1}^{i-1}(x^{(k+1,\,i)}-Y^{(k+2,\,j)})\cdot\prod_{j=i+1}^{n}(x^{(k+1,\,i)}-Y^{(k+1,\,j)}).$$

Применяя его, мы приходим к следующему модифицированному короткошаговому методу:

$$\text{(Б)}\qquad \begin{cases} Y^{(0,\,i)}=X^{(0,\,i)},\ x^{(0,\,i)}\in X^{(0,\,i)},\\ X^{(k+1,\,i)}=\{x^{(k,\,i)}-p(x^{(k,\,i)})/S^{(k,\,i)}\}\cap X^{(k,\,i)}, \quad \text{где}\\ S^{(k,\,i)}=\prod_{j=1}^{i-1}(x^{(k,\,i)}-Y^{(k+1,\,j)})\prod_{j=i+1}^{n}(x^{(k,\,i)}-Y^{(k,\,j)}),\\ x^{(k+1,\,i)}\in X^{(k+1,\,i)},\\ Y^{(k+1,\,i)}=\begin{cases}\left[x_1^{(k+1,\,i)},\ x^{(k+1,\,i)}\right], & \text{если } \operatorname{sign}(S^{(k,\,i)})\operatorname{sign}(p(x^{(k+1,\,i)}))>0,\\ \left[x^{(k+1,\,i)},\ x_2^{(k+1,\,i)}\right], & \text{если } \operatorname{sign}(S^{(k,\,i)})\operatorname{sign}(p(x^{(k+1,\,i)}))<0,\\ X^{(k+1,\,i)} & \text{в противном случае.}\end{cases}\end{cases}$$

Оба метода (3), (5) можно считать интервальными вариантами известных методов одновременного нахождения корней вещественных многочленов [78, 81, 87, 149]. Преимущество интервальных вариантов этих методов состоит в том, что они не только дают локализующие интервалы для корней, но и всегда сходятся при сделанных выше допущениях. Это устанавливается в следующей теореме.

Теорема 1. *Пусть дан многочлен* (1), *имеющий n простых корней* $\xi^{(i)}$, $1 \leqslant i \leqslant n$, *причем известны локализующие интервалы* $X^{(0, i)} \ni \xi^{(i)}$, *для которых верно* (2). *Тогда последовательность приближений* $\{X^{(k, i)}\}_{k=0}^{\infty}$, $1 \leqslant i \leqslant n$, *вычисленная по формулам* (3) (*соответственно* (5)), *либо удовлетворяет условиям*

$$\xi^{(i)} \in X^{(k, i)}, \quad k \geqslant 0,$$

и

$$X^{(0, i)} \supset X^{(1, i)} \supset X^{(2, i)} \supset \ldots, \quad \text{где} \quad \lim_{k \to \infty} X^{(k, i)} = \xi^{(i)},$$

либо стабилизируется за конечное число шагов на точке $[\xi^{(i)}, \xi^{(i)}]$. ■

Теорема 1 получается тем же методом, что и соответствующее утверждение в теореме 7.1, так как каждое из интервальных выражений $Q^{(k, i)}$, $S^{(k, i)}$ обладает либо свойством (7.2), либо соответствующим свойством в интервале $X^{(k, i)}$ для $m_2 < 0$.

Выбирая в обоих методах

$$x^{(k, i)} = \frac{1}{2}\left(x_1^{(k, i)} + x_2^{(k, i)}\right)$$

и рассматривая структуру выражений (3) и (5), мы немедленно получаем, что ширина локализующего интервала для каждого из корней уменьшается на каждом шаге итерации по крайней мере вдвое.

Теорема 1 частично сохраняется и в случае, когда многочлен имеет кратные корни. Если мы выпишем эти кратные корни вместе

$$\xi^{(m)}, \ \xi^{(m+1)}, \ \ldots, \ \xi^{(n)},$$

то оба метода (3), (5) нужно изменить таким образом, чтобы вычисление локализующих интервалов производилось только для индексов $1 \leqslant i < m$. Если эти интервалы для простых корней перевычисляются на каждом шаге, а остальные интервалы остаются неизменными, то теорема 1 сохраняется для простых корней (см. [25]).

Можно обобщить метод (3) таким образом, что в теореме 1 удастся заменить предположение (2), касающееся локализующих интервалов $X^{(0, i)}$, $1 \leqslant i \leqslant n$, на более слабое условие. Для

этого следует полностью использовать возможность варьирования значения $x^{(k,\,i)} \in X^{(k,\,i)}$, а не применять систематическую процедуру выбора, где это значение полагают равным, например, среднему арифметическому границ интервала. Такое обобщение рассмотрено в [24].

Теперь мы рассмотрим подробнее поведение последовательности

$$\{d(X^{(k,\,i)})\}_{k=0}^{\infty}, \qquad 1 \leqslant i \leqslant n.$$

Для метода (3) мы получаем оценку

$$d(X^{(k+1,\,i)}) \leqslant d(\{x^{(k,\,i)} - p(x^{(k,\,i)})/Q^{(k,\,i)}\}) = d(p(x^{(k,\,i)})/Q^{(k,\,i)}) = |p(x^{(k,\,i)})|\, d(1/Q^{(k,\,i)})$$

с помощью (2.9), (2.10) и (2.14). Так как

$$|p(x^{(k,\,i)})| = |p(x^{(k,\,i)}) - p(\xi^{(i)})| = |(x^{(k,\,i)} - \xi^{(i)})\, p'(\tilde{\eta}^{(k,\,i)})| \leqslant d(X^{(k,\,i)})\, |p'(\tilde{\eta}^{(k,\,i)})| \leqslant d(X^{(k,\,i)})\, |p'(X^{(0,\,i)})|,$$

отсюда следует, что

$$d(X^{(k+1,\,i)}) \leqslant d(X^{(k,\,i)})\, |p'(X^{(0,\,i)})|\, d(1/Q^{(k,\,i)}).$$

Применяя теорему 3.5, мы получаем оценку

$$d(1/Q^{(k,\,i)}) \leqslant \gamma^{(k,\,i)} d(Q^{(k,\,i)})$$

и, так как

$$Q^{(k,\,i)} \subseteq \prod_{j=1,\ j \neq i}^{n} (X^{(0,\,i)} - X^{(0,\,j)}),$$

получаем далее

$$d\left(\frac{1}{Q^{(k,\,i)}}\right) \leqslant \gamma^{(i)} d(Q^{(k,\,i)}) = \gamma^{(i)} d\left(\prod_{j=1,\ j \neq i}^{n} (x^{(k,\,i)} - X^{(k,\,j)})\right)$$

с константами $\gamma^{(i)}$, зависящими только от $X^{(0,\,j)}$, $1 \leqslant j \leqslant n$. Многократно применяя (2.12), мы получаем далее

$$d\left(\frac{1}{Q^{(k,\,i)}}\right) \leqslant \gamma^{(i)} \sum_{j=1,\ j \neq i}^{n} \eta^{(i,\,j)} d(X^{(k,\,j)})$$

с подходящими константами $\eta^{(i,\,j)}$, зависящими только от $X^{(0,\,j)}$, $1 \leqslant j \leqslant n$, так как $X^{(k,\,j)} \subseteq X^{(0,\,j)}$ и применимы (2.3) и (2.9). Резюмируя все это, мы получаем следующее неравенство:

$$(6) \qquad d(X^{(k+1,\,i)}) \leqslant |p'(X^{(0,\,i)})|\, \gamma^{(i)} d(X^{(k,\,i)}) \sum_{j=1,\ j \neq i}^{n} \eta^{(i,\,j)} d(X^{(k,\,j)}),$$

$$1 \leqslant i \leqslant n.$$

То же самое рассуждение можно провести для метода (5); единственное дополнительное обстоятельство, которое нужно учесть — это соотношение $Y^{(k,\,i)} \subseteqq X^{(k,\,i)}$. Это приводит к неравенству

$$(7)\qquad d(X^{(k+1,\,i)}) \leqslant |p'(X^{(0,\,i)})|\gamma^{(i)} d(X^{(k,\,i)}) \left(\sum_{j=1}^{i-1} \eta^{(i,\,j)} d(X^{(k+1,\,j)}) + \sum_{j=i+1}^{n} \eta^{(i,\,j)} d(X^{(k,\,j)}) \right), \qquad 1 \leqslant i \leqslant n.$$

В следующем утверждении оценивается R-порядок методов (3) и (5).

Теорема 2. *В предположениях и обозначениях теоремы* 1 *для R-порядка методов* (3) *и* (5) *имеет место*

$$(8)\qquad Q_R((3),\,(\xi^{(i)})) \geqslant 2$$

и

$$(9)\qquad Q_R((5),\,(\xi^{(i)})) \geqslant 1 + \sigma^{(n)},$$

где $\sigma^{(n)} > 1$ — единственный положительный корень многочлена

$$\tilde{q}^{(n)}(y) = y^n - y - 1.$$

Доказательство. (8): Из (6) немедленно получаем, что

$$d(X^{(k+1,\,i)}) \leqslant |p'(X^{(0,\,i)})|\gamma^{(i)} \left(\sum_{j=1,\ j \neq i}^{n} \eta^{(i,\,j)} \right) (d^{(k)})^2$$

$$\leqslant \max_{1 \leqslant i \leqslant n} \left\{ |p'(X^{(0,\,i)})|\gamma^{(i)} \left(\sum_{j=1\ j \neq i}^{n} \eta^{(i,\,j)} \right) \right\} (d^{(k)})^2$$

$$\leqslant \gamma (d^{(k)})^2, \qquad 1 \leqslant i \leqslant n,$$

где

$$d^{(k)} = \max_{1 \leqslant i \leqslant n} \{d(X^{(k,\,i)})\}.$$

Отсюда следует, что

$$d^{(k+1)} = \max_{1 \leqslant i \leqslant n} \{d(X^{(k+1,\,i)})\} \leqslant \gamma (d^{(k)})^2.$$

Тогда из теоремы 2 приложения А следует, что

$$O_R((3),\,(\xi^{(i)})) \geqslant 2$$

(9) Доказательство соотношения (9) требует бóльших усилий. Пусть

$$\gamma = \max_{1 \leqslant i,\,j \leqslant n} \{\eta^{(i,\,j)} |p'(X^{(0,\,i)})|\gamma^{(i)}\}.$$

Тогда мы можем переписать (7) в виде

$$d(X^{(k+1,\ i)}) \leqslant \gamma d(X^{(k,\ i)}) \left(\sum_{j=1}^{i-1} d(X^{(k+1,\ j)}) + \sum_{j=i+1}^{n} d(X^{(k,\ j)}) \right).$$

Применяя подстановку

$$d(X^{(k,\ i)}) = \frac{1}{(n-1)\gamma} h^{(k,\ i)}, \qquad 1 \leqslant i \leqslant n, \qquad \hat{\varepsilon} = \frac{1}{n-1}$$

это соотношение можно далее переписать в виде

$$h^{(k+1,\ i)} \leqslant \hat{\varepsilon} h^{(k,\ i)} \left(\sum_{j=1}^{i-1} h^{(k+1,\ j)} + \sum_{j=i+1}^{n} h^{(k,\ j)} \right).$$

Предположим без потери общности, что

$$h^{(0,\ i)} \leqslant h < 1, \qquad 1 \leqslant i \leqslant n.$$

Тогда мы имеем

$$h^{(k+1,\ i)} \leqslant h^{u^{(k+1,\ i)}}, \qquad 1 \leqslant i \leqslant n, \qquad k \geqslant 0.$$

Вектор $u_{\mathrm{p}}^{(k+1)} = (u^{(k+1,\ 1)}, \ldots, u^{(k+1,\ n)})^{\mathrm{T}}$, компоненты которого — целые числа, может быть вычислен по правилу

$$u_{\mathrm{p}}^{(k+1)} = \mathscr{A}_{\mathrm{p}} u_{\mathrm{p}}^{(k)}, \qquad k \geqslant 0,$$

с помощью начального вектора $u_{\mathrm{p}}^{(0)} = (1, 1, \ldots, 1)^{\mathrm{T}}$. (Здесь векторы с вещественными компонентами, называемые точечными векторами, обозначаются буквами a_{p}, b_{p}, ... с индексами p (point), чтобы отличить их от интервальных векторов [1]). То же делается для «точечных матриц». Эти обозначения будут часто встречаться в гл. 10.)

Тогда матрица $\mathscr{A}_{\mathrm{p}}$ имеет вид

$$\mathscr{A}_{\mathrm{p}} = \begin{pmatrix} 1 & 1 & & & & & \\ & 1 & 1 & & 0 & & \\ & & 1 & 1 & & & \\ & & & \ddots & \ddots & & \\ & 0 & & & \ddots & & \\ & & & & & 1 & 1 \\ 1 & 1 & 0 & \cdots & & 0 & 1 \end{pmatrix}$$

Доказательство этого соотношения может быть проведено методом математической индукции, и мы не приводим его здесь. Матрица $\mathscr{A}_{\mathrm{p}}$ неотрицательна, и ее направленный граф (см. [325]) очевидным образом является сильно связным. Отсюда

---

[1]) В английском оригинале применяются обозначения $\dot{\mathscr{A}}$, $\dot{\mathscr{B}}$, ... и т. д. В русском издании вместо точки по техническим причинам мы используем нижний индекс p. — *Прим. ред.*

следует, что $\mathscr{A}_p$ неприводима. Тогда из теоремы Перрона — Фробениуса следует, что $\mathscr{A}_p$ имеет собственное число $\lambda^{(1)}$, равное ее спектральному радиусу $\rho(\mathscr{A}_p)$. По теореме 2.9 из [325] матрица $\mathscr{A}_p$ примитивна. Поэтому остальные ее собственные числа удовлетворяют соотношению

$$\lambda^{(1)} = \rho(\mathscr{A}_p) > |\lambda^{(2)}| \geqslant \ldots \geqslant |\lambda^{(n)}|.$$

Так как матрица $\mathscr{A}_p$ примитивна, мы имеем для некоторого натурального числа $k^{(0)}$, что

$$\mathscr{A}_p^k = (a_{ij}^{(k)}) > O_p, \qquad k \geqslant k^{(0)}.$$

Однако в [102] показано, что для матриц, обладающих двумя последними свойствами, мы имеем

$$\lim_{k \to \infty} (a_{ij}^{(k+1)} / a_{ij}^{(k)}) = \lambda^{(1)}.$$

Поэтому для данного $\varepsilon > 0$ верно, что

$$a_{ij}^{(k+1)} / a_{ij}^{(k)} \geqslant \rho(\mathscr{A}_p) - \varepsilon, \qquad k \geqslant k(\varepsilon) \geqslant k^{(0)},$$

или

$$a_{ij}^{(k+1)} \geqslant \alpha(\rho(\mathscr{A}_p) - \varepsilon), \qquad 1 \leqslant i, \quad j \leqslant n,$$

где

$$\alpha = \min_{1 \leqslant i,\ j \leqslant n} a_{ij}^{(k)} > 0.$$

Отсюда следует, что

$$a_{ij}^{(k+2)} \geqslant a_{ij}^{(k+1)} (\rho(\mathscr{A}_p) - \varepsilon) \geqslant \alpha(\rho(\mathscr{A}_p) - \varepsilon),$$

что дает

$$a_{ij}^{(k+r)} \geqslant \alpha(\rho(\mathscr{A}_p) - \varepsilon)^r, \qquad 1 \leqslant i, \quad j \leqslant n, \quad r \geqslant 0.$$

Используя правило вычисления векторов $u_p^{(k)}$, мы получаем тогда

$$u_p^{(k+r)} = \mathscr{A}_p^{k+r} u_p^{(0)} = \left(\sum_{i=1}^{n} a_{ij}^{(k+r)} \right) \geqslant (n\alpha(\rho(\mathscr{A}_p) - \varepsilon)^r)\, e_p,$$

где $e_p = (1, 1, \ldots, 1)^T$. Поэтому получаем

$$h^{(k+r,\ i)} \leqslant h^{u^{(k+r,\ i)}} \leqslant h^{n\alpha(\rho(\mathscr{A}_p) - \varepsilon)^r},$$

$$1 \leqslant i \leqslant n, \qquad r \geqslant 0, \qquad k \geqslant k(\varepsilon) \geqslant k^{(0)}.$$

Это легко переформулировать в виде

$$d(X^{(k+r,\ i)}) \leqslant (\hat{\varepsilon}/\gamma)\, h^{n\alpha(\rho(\mathscr{A}_p) - \varepsilon)^r}.$$

Пусть теперь

$$d^{(k)} = \max_{1 \leqslant i \leqslant n} \{d(X^{(k,\ i)})\}.$$

Тогда получаем

$$d^{(k+r)} \leqslant (\hat{\varepsilon}/\gamma)\, h^{\alpha n\,(\rho\,(\mathscr{A}_p)-\varepsilon)^r}.$$

Поэтому мы можем заключить, что R-фактор удовлетворяет соотношению

$$R_{\rho\,(\mathscr{A}_p)-\varepsilon}\{d^{(k)}\} = \limsup_{r\to\infty} (d^{(k+r)})^{[1/(\rho\,(\mathscr{A}_p)-\varepsilon)^r]}$$

$$\leqslant \limsup_{r\to\infty} \left(\frac{\hat{\varepsilon}}{\gamma}\, h^{\alpha n\,(\rho\,(\mathscr{A}_p)-\varepsilon)^r}\right)^{[1/(\rho\,(\mathscr{A}_p)-\varepsilon)^r]} = h^{\alpha n} < 1.$$

Отсюда следует, что

$$O_R((5), (\xi^{(i)})) \geqslant \rho(\mathscr{A}_p) - \varepsilon$$

для всех $\varepsilon > 0$, а потому

$$O_R((5), (\xi^{(i)})) \geqslant \rho(\mathscr{A}_p).$$

Рассмотрим теперь характеристический многочлен $q^{(n)}(\lambda)$ матрицы $\mathscr{A}_p$:

$$q^{(n)}(\lambda) = (\lambda - 1)^n - (\lambda - 1) - 1.$$

Полагая $\tau = \lambda - 1$, мы можем написать его в виде

$$\tilde{q}^{(n)}(\tau) = \tau^n - \tau - 1.$$

Так как

$$\tilde{q}^{(n)}(1) = -1 < 0 \quad \text{и} \quad \tilde{q}^{(n)}(2) = 2^n - 3 \geqslant 1 > 0$$

для $n \geqslant 2$, то по правилу Декарта многочлен $\tilde{q}^{(n)}(\tau)$ имеет ровно один положительный корень $\sigma^{(n)}$, для которого

$$1 < \sigma^{(n)} < 2.$$

Поэтому спектральный радиус матрицы $\mathscr{A}_p$ удовлетворяет соотношению

$$\rho(\mathscr{A}_p) = 1 + \sigma^{(n)} > 2,$$

откуда получается

$$O_R((5), (\xi^{(i)})) \geqslant 1 + \sigma^{(n)}. \blacksquare$$

Мы опишем теперь одно приложение метода (5). Пусть дана вещественная симметрическая матрица $\mathscr{A}'_p = (a_{ij})$ размерности $n \times n$. Нужно определить собственные числа этой матрицы, т. е. числа λ, для которых выполнено равенство

$$\mathscr{A}'_p x_p = \lambda x_p \quad \text{при} \quad x_p \neq o_p.$$

Чтобы сделать это, применяем конечное число раз ортогональные преобразования подобия

$$\tilde{\mathscr{A}}_p = \mathscr{U}_p^T \mathscr{A}_p \mathscr{U}_p,$$

преобразующие, вообще говоря, неразреженную матрицу $\mathscr{A}'_p$ в матрицу $\mathscr{A}_p$, имеющую вид

$$\mathscr{A}_p = \begin{pmatrix} a^{(1)} & b^{(1)} & & & 0 \\ b^{(1)} & a^{(2)} & b^{(2)} & & \\ & \ddots & \ddots & \ddots & \\ 0 & & & b^{(n-1)} & a^{(n)} \end{pmatrix}$$

Затем собственные числа матрицы $\mathscr{A}_p$ (а тем самым и матрицы $\mathscr{A}'_p$) вычисляются как корни характеристического многочлена

$$p(\lambda) = \det(\lambda \mathscr{T}_p - \mathscr{A}_p)$$

матрицы $\mathscr{A}_p$. Значение $p(\lambda)$ может быть найдено с помощью следующей рекуррентной формулы (см. [300]):

$$(10) \quad \begin{cases} f^{(0)}(\lambda) = 1, \quad f^{(1)}(\lambda) = \lambda - a^{(1)}, \\ f^{(k)}(\lambda) = (\lambda - a^{(k)}) f^{(k-1)}(\lambda) - (b^{(k-1)})^2 f^{(k-2)}(\lambda), \quad 2 \leqslant k \leqslant n, \\ p(\lambda) = f^{(n)}(\lambda). \end{cases}$$

Если теперь $\mathscr{A}_p$ имеет только простые собственные числа и для них найдены непересекающиеся локализующие интервалы (например, с помощью теоремы Гершгорина), то мы можем применить (5). Это иллюстрирует следующий пример.

Пример. (α) Рассмотрим матрицу

$$\mathscr{A}_p = \begin{pmatrix} 15 & 1 & & & & & & & \\ 1 & 10 & 1 & & & & & & \\ & 1 & 7 & 1 & & & & & \\ & & 1 & 4 & 1 & & & & \\ & & & 1 & 0 & 1 & & & \\ & & & & 1 & -4 & 1 & & \\ & & & & & 1 & -7 & 1 & \\ & & & & & & 1 & -10 & 1 \\ & & & & & & & 1 & -15 \end{pmatrix}$$

Применяя теорему Гершгорина, мы получаем для собственных чисел матрицы $\mathscr{A}_p$ следующие локализующие интервалы:

$$X^{(0,1)} = [+\,13.99999999995, +\,16.00000000005],$$
$$X^{(0,2)} = [+\,7.99999999974, +\,12.00000000005],$$
$$X^{(0,3)} = [+\,4.99999999981, +\,9.00000000015],$$
$$X^{(0,4)} = [+\,1.99999999992, +\,6.00000000022],$$
$$X^{(0,5)} = [-\,2.00000000008, +\,2.00000000008],$$
$$X^{(0,6)} = [-\,6.00000000022, -\,1.99999999992],$$
$$X^{(0,7)} = [-\,9.00000000015, -\,4.99999999981],$$
$$X^{(0,8)} = [-\,12.00000000005, -\,7.99999999974],$$
$$X^{(0,9)} = [-\,17.00000000005, -\,12.99999999995].$$

Исходя из этих интервалов, мы строим итерации по формулам (5) и получаем следующие результаты:

$$X^{(5,1)} = [+\,\underline{15.19709300868}, \quad +\,\underline{15.19709300872}],$$
$$X^{(4,2)} = [+\,\underline{10.13174515464}, \quad +\,\underline{10.13174515471}],$$
$$X^{(4,3)} = [+\,\underline{7.001927580904}, \quad +\,\underline{7.001927580960}],$$
$$X^{(4,4)} = [+\,\underline{3.920346203678}, \quad +\,\underline{3.920346203715}],$$
$$X^{(5,5)} = [-\,0.1096791595101 \times 10^{-10}, \quad +\,0.1096791595101 \times 10^{-10}],$$
$$X^{(4,6)} = [-\,\underline{3.920346203719}, \quad -\,\underline{3.920346203674}],$$
$$X^{(4,7)} = [-\,\underline{7.001927580969}, \quad -\,\underline{7.001927580895}],$$
$$X^{(3,8)} = [-\,\underline{10.13174515473}; \quad -\,\underline{10.13174515463}],$$
$$X^{(3,9)} = [-\,\underline{15.19709300876}, \quad -\,\underline{15.19709300866}].$$

Эти интервалы невозможно улучшить, используя имеющуюся программу (см. приложение C). Подчеркнуты знаки, совпадающие в верхней и нижней границах.

(β) Теперь рассмотрим матрицу

$$\mathscr{A}_p = \begin{pmatrix} 12 & 1 & & & 0 \\ 1 & 9 & 1 & & \\ & 1 & 6 & 1 & \\ & & 1 & 3 & 1 \\ 0 & & & 1 & 0 \end{pmatrix}$$

(ср. [521a, p. 375]). Снова применяя теорему Гершгорина, мы находим для собственных чисел матрицы $\mathscr{A}_p$ следующие лока-

лизующие интервалы:

$$X^{(0,1)} = [+\,10.99999999998,\ +\,13.00000000003],$$
$$X^{(0,2)} = [+\,6.99999999970,\ +\,11.00000000003],$$
$$X^{(0,3)} = [+\,3.99999999989,\ +\,8.00000000021],$$
$$X^{(0,4)} = [+\,0.99999999945,\ +\,5.00000000019],$$
$$X^{(0,5)} = [-\,1.00000000004,\ +\,1.00000000004].$$

Следующие уточненные интервалы были вычислены с помощью итерационного метода (5) (ср. с замечаниями, сделанными после теоремы 1) [1]):

$$X^{(1,1)} = [+\,12.11013986010,\ +\,12.55506993010],$$
$$X^{(1,2)} = [+\,9.006328989416,\ +\,9.048379503166],$$
$$X^{(1,3)} = [+\,5.99999999958,\ +\,6.00000000041],$$
$$X^{(1,4)} = [+\,2.979804773200,\ +\,2.987022580008],$$
$$X^{(1,5)} = [-\,0.3230758693540,\ -\,0.3162523763767].$$

$$X^{(2,1)} = [+\,12.31617201370,\ +\,12.31774922532],$$
$$X^{(2,2)} = [+\,9.016110401580,\ +\,9.016149094187],$$
$$X^{(2,3)} = [+\,5.99999999958,\ +\,6.00000000013],$$
$$X^{(2,4)} = [+\,2.983860239266,\ +\,2.983864788268],$$
$$X^{(2,5)} = [-\,0.3168759526293,\ -\,0.3168750526051],$$

$$X^{(3,1)} = [+\,12.31687595112,\ +\,12.31687595546],$$
$$X^{(3,2)} = [+\,9.016136303134,\ +\,9.016136303198],$$
$$X^{(3,3)} = X^{(2,3)}$$
$$X^{(3,4)} = [+\,2.983863696823,\ +\,2.983863696853],$$
$$X^{(3,5)} = [-\,0.3168759526293,\ -\,0.3168759526051],$$
$$X^{(4,1)} = [+\,12.31687595258,\ +\,12.31687595266],$$
$$X^{(4,2)} = [+\,9.016136303134,\ +\,9.016136303181],$$
$$X^{(4,3)} = X^{(3,3)}$$
$$X^{(4,4)} = X^{(3,4)}$$
$$X^{(4,5)} = [-\,0.3168759526284,\ -\,0.3168759526051].$$

---

[1]) Этот пример наглядно показывает, каким плохим критерием точности является количество «правильных» цифр — $X^{(1,3)}$ вычислено гораздо точнее,

* * *

Литература [1], [24], [25], [26], [57], [61], [78], [81], [87], [102], [129], [149], [246a], [285], [300], [325], [521].

Глава 9
МЕТОДЫ ОДНОВРЕМЕННОЙ ЛОКАЛИЗАЦИИ КОМПЛЕКСНЫХ КОРНЕЙ МНОГОЧЛЕНОВ

В этой главе мы обсудим метод одновременной локализации комплексных (в общем случае) корней многочленов, предложенный Гаргантини и Энричи [95]. Пусть задан многочлен

$$p(z) = a^{(n)}z^n + a^{(n-1)}z^{n-1} + \ldots + a^{(1)}z + a^{(0)}, \tag{1}$$

где $a^{(i)} \in \mathbb{C}$, $0 \leqslant i \leqslant n$, $n \geqslant 2$. Предположим далее, что заданы n интервалов

$$W^{(0,\,i)} = \langle z^{(0,\,i)},\ r^{(0,\,i)} \rangle \in K(\mathbb{C}),$$

для которых

$$\zeta^{(i)} \in W^{(0,\,i)}, \quad p(\zeta^{(i)}) = 0, \quad 1 \leqslant i \leqslant n, \tag{2}$$

$$W^{(0,\,i)} \cap W^{(0,\,j)} = \varnothing, \quad 1 \leqslant i < j \leqslant n. \tag{3}$$

Элемент $Z \in K(\mathbb{C})$ представляется в дальнейшем в виде $Z = \langle m(Z),\ r(Z) \rangle$.

Рассмотрим следующий метод итерации:

$$\begin{cases} z^{(k,\,i)} = m(W^{(k,\,i)}), \\ C^{(k,\,i)} = \sum\limits_{j=1,\ j \neq i}^{n} \dfrac{1}{z^{(k,\,i)} - W^{(k,\,j)}}, \\ q(z^{(k,\,i)}) = \dfrac{p'(z^{(k,\,i)})}{p(z^{(k,\,i)})} \quad \text{для} \quad p(z^{(k,\,i)}) \neq 0, \\ W^{(k+1,\,i)} = \langle z^{(k+1,\,i)},\ r^{(k+1,\,i)} \rangle = -\dfrac{1}{q(z^{(k,\,i)}) - C^{(k\ \ i)}}, \\ 1 \leqslant i \leqslant n, \quad k \geqslant 0, \end{cases} \tag{4}$$

и пусть

$$r^{(k)} = \max_{1 \leqslant i \leqslant n} \{r^{(k,\,i)}\}, \tag{5}$$

$$\rho^{(k)} = \min_{1 \leqslant i < j \leqslant n} \{\min\{|z| \mid z \in z^{(k,\,i)} - W^{(k,\,j)}\}\}. \tag{6}$$

чем $X^{(1,\,1)}$, $X^{(1,\,2)}$, $X^{(1,\,4)}$, $X^{(1,\,5)}$, хотя мы так и не знаем ни одной цифры в его десятичной записи. — *Прим. ред.*

Для $i \neq j$ из (3) следует, что

$$\min\{|z| \mid z \in z^{(0,i)} - W^{(0,i)}\} = |z^{(0,i)} - z^{(0,j)}| - r^{(0,j)} \geqslant \rho^{(0)}. \tag{7}$$

Определим еще величины $\eta^{(k)}$ соотношением

$$\rho^{(k)} = (n-1)\eta^{(k)}. \tag{8}$$

Тогда для итерационной схемы (4) верно следующее.

Теорема 1. *Пусть $p(z)$ есть многочлен (1) и его корни $\zeta^{(i)}$, $1 \leqslant i \leqslant n$, удовлетворяют условиям (2) и (3). В обозначениях (5), (6), (8) пусть*

$$6r^{(0)} \leqslant \eta^{(0)}. \tag{9}$$

Тогда

(a) *итерация (4) всегда осуществима, причем*

$$\zeta^{(i)} \in W^{(k,i)}, \quad 1 \leqslant i \leqslant n, \quad k \geqslant 0;$$

(b) *имеет место неравенство*

$$r^{(k+1)} \leqslant \frac{1}{\rho^{(0)}(\eta^{(0)} - 4r^{(0)})}(r^{(k)})^3 \leqslant \frac{1}{12(n-1)} r^{(k)}, \quad k \geqslant 0.$$

Замечание. Из (6) следует, что $\lim\limits_{k\to\infty} r^{(k)} = 0$. Поэтому в силу (a) получаем, что

$$\lim_{k\to\infty} W^{(k,i)} = \zeta^{(i)}, \quad 1 \leqslant i \leqslant n.$$

Если учесть определения 6.4c и 10.6a, то из (6) следует тогда с помощью теоремы 2 из приложения A, что R-порядок итераций (4) удовлетворяет условию $O_R((4), (\xi^{(i)})) \geqslant 3$.

Доказательство. (a): Из

$$|z^{(0,i)} - \zeta^{(i)}| \leqslant r^{(0,i)} \leqslant r^{(0)},$$

$$|z^{(0,i)} - \zeta^{(j)}| \geqslant |z^{(0,i)} - z^{(0,j)}| - |z^{(0,j)} - \zeta^{(j)}|$$

$$\geqslant |z^{(0,i)} - z^{(0,i)}| - r^{(0,i)} \geqslant \rho^{(0)}$$

следует, что

$$|q(z^{(0,i)})| = \left|\sum_{j=1}^{n} \frac{1}{z^{(0,i)} - \zeta^{(j)}}\right| \tag{10}$$

$$\geqslant \left|\frac{1}{z^{(0,i)} - \zeta^{(i)}}\right| - \sum_{j=1,\ j\neq i} \left|\frac{1}{z^{(0,i)} - \zeta^{(j)}}\right|$$

$$\geqslant \frac{1}{r^{(0)}} - \frac{1}{\eta^{(0)}} \quad \text{для} \quad z^{(0,i)} \neq \zeta^{(i)}.$$

Из

$$|z^{(0,i)} - z^{(0,j)}| - r^{(0,j)} \geqslant \rho^{(0)} > 0$$

мы имеем

$$0 \notin z^{(0,\,i)} - W^{(0,\,i)},$$

а также

$$\frac{1}{z^{(0,\,1)} - W^{(0,\,i)}} \subset \left\langle 0,\ \frac{1}{\rho^{(0)}} \right\rangle,$$

$$C^{(0,\,i)} = \sum_{j=1,\ j \neq i}^{n} \frac{1}{z^{(0,\,i)} - W^{(0,\,j)}} \subset \sum_{j=1,\ j \neq i}^{n} \left\langle 0,\ \frac{1}{\rho^{(0)}} \right\rangle = \left\langle 0,\ \frac{1}{\eta^{(0)}} \right\rangle,$$

$$q\,(z^{(0,\,i)}) - C^{(0,\,i)} \subset \langle q\,(z^{(0,\,i)}),\ 1/\eta^{(0)} \rangle. \tag{11}$$

Так как

$$|\,q\,(z^{(0,\,i)})\,| - 1/\eta^{(0)} \geqslant 1/r^{(0)} - 2/\eta^{(0)} > 0,$$

то ясно, что

$$0 \notin q\,(z^{(0,\,i)}) - C^{(0,\,i)},$$

и потому определены

$$W^{(1,\,i)}, \qquad 1 \leqslant i \leqslant n.$$

Ввиду

$$\frac{p'\,(z^{(0,\,i)})}{p\,(z^{(0,\,i)})} = \sum_{j=1}^{n} \frac{1}{z^{(0,\,i)} - \zeta^{(j)}}$$

из (2) и монотонности включения следует, что

$$\zeta^{(i)} = z^{(0,\,i)} - p\,(z^{(0,\,i)}) \Big/ \left[p'\,(z^{(0,\,i)}) - p\,(z^{(0,\,i)}) \sum_{j=1,\ j \neq i}^{n} \frac{1}{z^{(0,\,i)} - \zeta^{(j)}} \right]$$

$$\in z^{(0,\,i)} - \frac{1}{q\,(z^{(0,\,i)}) - C^{(0,\,i)}} = W^{(1,\,i)}, \qquad 1 \leqslant i \leqslant n.$$

Это доказывает (a) для $k = 1$.

(b): Из

$$|\,z^{(0,\,i)} - z^{(0,\,j)}\,|^2 - (r^{(0,\,j)})^2 \geqslant (\rho^{(0)} + r^{(0,\,j)})^2 - (r^{(0,\,j)})^2 \geqslant (\rho^{(0)})^2$$

мы получаем, что

$$r\left(\frac{1}{z^{(0,\,i)} - W^{(0,\,j)}}\right) = \frac{r^{(0,\,j)}}{|\,z^{(0,\,i)} - z^{(0,\,j)}\,|^2 - (r^{(0,\,j)})^2} \leqslant \frac{r^{(0)}}{(\rho^{(0)})^2},$$

а потому

$$r\,(C^{(0,\,i)}) \leqslant \frac{n-1}{\rho^{(0)}} \cdot \frac{r^{(0)}}{\rho^{(0)}} = \frac{r^{(0)}}{\eta^{(0)}\rho^{(0)}}.$$

Используя это неравенство и (11), получаем теперь

$$r\,(q\,(z^{(0,\,i)}) - C^{(0,\,i)}) = r\,(C^{(0,\,i)}),$$

$$|\,m\,(q\,(z^{(0,\,i)}) - C^{(0,\,i)})\,| \geqslant 1/r^{(0)} - 2/\eta^{(0)} + r\,(q\,(z^{(0,\,i)}) - C^{(0,\,i)})$$

$$= 1/r^{(0)} - 2/\eta^{(0)} + r\,(C^{(0,\,i)}),$$

а отсюда и неравенство

$$r(W^{(1,\,i)}) = r\left(\frac{1}{q(z^{(0,\,i)}) - C^{(0,\,i)}}\right)$$

$$= \frac{r(q(z^{(0,\,i)}) - C^{(0,\,i)})}{|m(q(z^{(0,\,i)}) - C^{(0,\,i)})|^2 - (r(q(z^{(0,\,i} - C^{(0,\,i)}))^2}$$

$$\leqslant \frac{(r^{(0)})^3}{\rho^{(0)}(\eta^{(0)} - 4r^{(0)})},$$

т. е.

$$r^{(1)} \leqslant \frac{(r^{(0)})^3}{\rho^{(0)}(\eta^{(0)} - 4r^{(0)})}. \tag{12}$$

Применяя (9), мы получаем из предыдущей оценки неравенство

$$r^{(1)} \leqslant \frac{1}{12(n-1)} r^{(0)}.$$

Пусть

$$\delta^{(0)} = \max_{1 \leqslant i \leqslant n} \{|z^{(0,\,i)} - z^{(1,\,i)}|\}.$$

Тогда, применяя (6), мы получаем

$$\rho^{(1)} \geqslant \rho^{(0)} - \delta^{(0)} - 2r^{(1)}. \tag{13}$$

Чтобы оценить $\delta^{(0)}$, мы используем (10), (11) и соотношение

$$z^{(1,\,i)} - z^{(0,\,i)} \in \frac{1}{q(z^{(0,\,i)}) - C^{(0,\,i)}},$$

чтобы получить

$$|z^{(1,\,i)} - z^{(0,\,i)}| \leqslant \left|\frac{1}{\langle q(z^{(0,\,i)}),\, 1/\eta^{(0)}\rangle}\right| = \frac{1}{|q(z^{(0,\,i)})| - 1/\eta^{(0)}} \leqslant \frac{r^{(0)}\eta^{(0)}}{\eta^{(0)} - 2r^{(0)}},$$

что дает, наконец,

$$\delta^{(0)} \leqslant \frac{r^{(0)}\eta^{(0)}}{\eta^{(0)} - 2r^{(0)}}. \tag{14}$$

Используя (12), (13) и (14), получаем из (9), что

$$\eta^{(1)} - 6r^{(1)} = \rho^{(1)}/(n-1) - 6r^{(1)} \tag{15}$$

$$\geqslant \eta^{(0)} - r^{(0)}\left(\frac{\eta^{(0)}}{\eta^{(0)} - 2r^{(0)}} + \frac{8(r^{(0)})^2}{\rho^{(0)}(\eta^{(0)} - 4r^{(0)})}\right)$$

$$\geqslant \eta^{(0)} - 3r^{(0)} \geqslant 0,$$

т. е.

$$\eta^{(1)} \geqslant 6r^{(1)}.$$

Отсюда можно так же, как и раньше, получить, что

$$r^{(2)} \leqslant \frac{1}{\rho^{(1)}(\eta^{(1)} - 4r^{(1)})}(r^{(1)})^3 \leqslant \frac{1}{12(n-1)} r^{(1)}.$$

Из (13) тем же способом, каким было получено (15), мы выводим, что

$$\eta^{(1)} - 4r^{(1)} \geqslant \eta^{(0)} - r^{(0)}\left(\frac{\eta^{(0)}}{\eta^{(0)} - 2r^{(0)}} + \frac{6(r^{(0)})^2}{\rho^{(0)}(\eta^{(0)} - 4r^{(0)})}\right) \geqslant \tag{16}$$

и

$$\eta^{(1)} \geqslant \eta^{(0)} - r^{(0)}\left(\frac{\eta^{(0)}}{\eta^{(0)} - 2r^{(0)}} + \frac{2(r^{(0)})^2}{\rho^{(0)}(\eta^{(0)} - 4r^{(0)})}\right) \geqslant 0. \tag{17}$$

Используя оба последних неравенства, мы получаем из (9)

$$\eta^{(1)}(\eta^{(1)} - 4r^{(1)}) \geqslant (\eta^{(0)})^2 - \eta^{(0)} r^{(0)}\left(\frac{2\eta^{(0)}}{\eta^{(0)} - 2r^{(0)}} + \frac{8(r^{(0)})^2}{\rho^{(0)}(\eta^{(0)} - 4r^{(0)})}\right)$$
$$\geqslant \eta^{(0)}(\eta^{(0)} - 4r^{(0)}),$$

а потому

$$r^{(2)} \leqslant \frac{1}{\rho^{(0)}(\eta^{(0)} - 4r^{(0)})}(r^{(1)})^3.$$

Оставшаяся часть доказательства получается методом математической индукции. ■

Теперь мы проиллюстрируем итерационный метод (4). Для этого рассмотрим задачу вычисления собственных чисел гессенберговой матрицы с помощью последовательности локализаций. Нужные при этом значения характеристического многочлена и его производных можно вычислить с помощью метода Хаймана (см. [521a, p. 353]). В качестве конкретного примера рассмотрим матрицу

$$\mathscr{H}_p = \begin{pmatrix} 12+16i & 1 & 0 & 0 \\ 0 & 9+12i & 1 & 0 \\ 0 & 0 & 6+8i & 1 \\ 1 & 0 & 0 & 3+4i \end{pmatrix},$$

где $i = \sqrt{-1}$. Применяя теорему Гершгорина, мы получаем, что каждый из кругов

$$W^{(0,1)} = \langle 12+16i,\ 1\rangle, \quad W^{(0,2)} = \langle 9+12i,\ 1\rangle,$$
$$W^{(0,3)} = \langle 6+8i,\ 1\rangle, \quad W^{(0,4)} = \langle 3+4i,\ 1\rangle$$

содержит в точности одно собственное число матрицы $\mathscr{H}_p$. С помощью (4) мы получаем уточненные локализующие множества $W^{(k,i)}$ для собственных чисел матрицы $\mathscr{H}_p$. Ниже в табл. 1 используется представление

$$W^{(k,i)} = \langle m(W^{(k,i)}),\ r(W^{(k,i)})\rangle,$$

где

$$m(W^{(k,i)}) = \mathrm{Re}(m(W^{(k,i)})) + i\,\mathrm{Im}(m(W^{(k,i)})).$$

Дополнительные числовые примеры можно найти в [95] и [305].

Замечания. Итерационный метод, исследованный в теореме 1, можно назвать полношаговым методом. Если на каждом шаге использовать только что уточненные значения приближений, то

Таблица 1

k	i	Re	Im	r
1	1	+ 11.99875131516	+ 15.99953080496	0.1001255×10^{-6}
	2	+ 9.003742419628	+ 12.00140833328	0.1494005×10^{-5}
	3	+ 5.996257580383	+ 7.998591666711	0.1493969×10^{-5}
	4	+ 3.001248654837	+ 4.000469195035	0.1000782×10^{-6}
2	1	+ 11.99875136181	+ 15.99953080159	0.1019500×10^{-9}
	2	+ 9.003742437190	+ 12.00140832752	$0.8760740 \times 10^{-10}$
	3	+ 5.996257562811	+ 7.998591672458	$0.3665239 \times 10^{-10}$
	4	+ 3.001248638204	+ 4.000469198423	$0.2555951 \times 10^{-10}$
3	1	+ 11.99875136181	+ 15.99953080159	0.1019496×10^{-9}
	2	+ 9.003742437190	+ 12.00140832752	$0.8760740 \times 10^{-10}$
	3	+ 5.996257562811	+ 7.998591672458	$0.3665353 \times 10^{-10}$
	4	+ 3.001248638204	+ 4.000469198423	$0.2556093 \times 10^{-10}$

получим метод, который можно назвать короткошаговым. Аналогично тому, как это сделано в [26], можно показать, что этот метод имеет более чем кубический порядок сходимости к нулю радиуса аппроксимирующего круга. Другой класс методов нахождения корней многочленов с помощью круговой арифметики был введен в [389].

* * *

Литература: [26], [95], [120], [305], [389], [521a].

Глава 10
ОПЕРАЦИИ НАД ИНТЕРВАЛЬНЫМИ МАТРИЦАМИ

Множество вещественных матриц размерности $m \times n$ обозначается через $M_{mn}(\mathbb{R})$, а множество комплексных матриц размерности $m \times n$ — через $M_{mn}(\mathbb{C})$. Элементы множества $M_{mn}(\mathbb{R})$, $M_{mn}(\mathbb{C})$ обозначаются через $\mathcal{A}_p$, $\mathcal{B}_p$, $\mathcal{C}_p$, ..., $\mathcal{X}_p$, $\mathcal{Y}_p$, $\mathcal{Z}_p$. Матрицы-столбцы, т. е. вещественные или комплексные векторы, обозначаются через a_p, b_p, e_p, ..., x_p, y_p, z_p. Множество вещественных n-мерных векторов обозначается через $V_n(\mathbb{R})$, мно-

жество комплексных векторов — через $V_n(\mathbb{C})$. Аналогичным образом мы обозначаем через $M_{mn}(I(\mathbb{R}))$ множество матриц, элементами которых являются вещественные интервалы, а через $M_{mn}(I(\mathbb{C}))$ — множество матриц, элементами которых являются комплексные интервалы; здесь $I(\mathbb{C})$ может обозначать как $R(\mathbb{C})$, так и $K(\mathbb{C})$. Элементы множества $M_{mn}(I(\mathbb{R}))$ (соответственно $M_{mn}(I(\mathbb{C}))$ обозначаются через $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$, ..., $\mathcal{X}$, $\mathcal{Y}$, $\mathcal{Z}$, и мы называем их вещественными (соответственно комплексными) интервальными матрицами. Интервальные матрицы-столбцы, т. е. вещественные или комплексные интервальные векторы, обозначаются через a, b, c, ..., x, y, z. Множество вещественно-интервальных векторов-столбцов обозначается через $V_n(I(\mathbb{R}))$, а множество комплексно-интервальных векторов-столбцов — через $V_n(I(\mathbb{C}))$. Интервальные векторы и матрицы записываются, как обычно, в виде $\mathcal{A}=(A_{ij})$ в случае матриц и $a=(A_i)$ в случае векторов. Интервальная матрица, все компоненты которой являются точечными интервалами, называется *точечной матрицей*. *Точечные векторы* определяются аналогично. Упомянем очевидные соотношения $M_{mn}(I(\mathbb{R}))\subset M_{mn}(R(\mathbb{C}))$ и $V_{mn}(I(\mathbb{R}))\subset V_n(R(\mathbb{C}))$.

Определение 1. Две интервальные матрицы $\mathcal{A}=(A_{ij})$ и $\mathcal{B}=(B_{ij})$ размерности $m\times n$ равны (это записывается, как обычно, в виде $\mathcal{A}=\mathcal{B}$), если равны их соответствующие компоненты. Иными словами, $\mathcal{A}=\mathcal{B}\leftrightarrow A_{ij}=B_{ij}$, $1\leqslant i\leqslant m$, $1\leqslant j\leqslant n$. ■

Сейчас мы введем частичный порядок на множестве интервальных матриц.

Определение 2. Пусть $\mathcal{A}=(A_{ij})$ и $\mathcal{B}=(B_{ij})$ — интервальные матрицы размерности $m\times n$. Тогда полагаем

$$\mathcal{A}\subseteq\mathcal{B}\leftrightarrow A_{ij}\subseteq B_{ij},\quad 1\leqslant i\leqslant m,\quad 1\leqslant j\leqslant m.\ \blacksquare$$

Отношение $\mathcal{A}\subset\mathcal{B}$ вводится аналогичным поэлементным определением. Если при этом $\mathcal{A}_p=(a_{ij})$ — точечная матрица, то мы пишем также $\mathcal{A}_p\in\mathcal{B}$. Каждую интервальную матрицу можно рассматривать как множество точечных матриц. Отношения $\subseteq$ и $\subset$ между множествами точечных матриц понимаются в обычном теоретико-множественном смысле.

Наша следующая цель — определить операции над интервальными матрицами, формально соответствующие операциям над точечными матрицами.

Определение 3. (а) Пусть $\mathcal{A}=(A_{ij})$, $\mathcal{B}=(B_{ij})$ — две интервальные матрицы размерности $m\times n$. Тогда соотношения

$$\mathcal{A}\pm\mathcal{B}:=(A_{ij}\pm B_{ij})$$

определяют соответственно сложение и вычитание интервальных матриц.

(b) Пусть $\mathscr{A} = (A_{ij})$ — интервальная матрица размерности $m \times r$ и $\mathscr{B} = (B_{ij})$ — интервальная матрица размерности $r \times n$. Тогда соотношение

$$\mathscr{A}\mathscr{B} := \left(\sum_{\nu=1}^{r} A_{i\nu} B_{\nu j} \right)$$

определяет умножение интервальных матриц. В частности, для интервальной матрицы $\mathscr{A} = (A_{ij})$ размерности $n \times r$ и интервального вектора $u = (U_i)$ размерности r мы имеем

$$\mathscr{A}u = \left(\sum_{\nu=1}^{r} A_{i\nu} U_\nu \right).$$

(c) Пусть $\mathscr{A} = (A_{ij})$ — интервальная матрица и X — интервал. Тогда полагаем

$$X\mathscr{A} = \mathscr{A}X := (XA_{ij}). \quad \blacksquare$$

В дальнейшем предполагается, что интервальные матрицы, участвующие в интервальной операции, имеют нужное для этой операции число строк и столбцов, и это обстоятельство не будет специально оговариваться. Далее предполагается, что интервальные операнды (т. е. аргументы операций) имеют подходящие элементы. Если, например, мы имеем $\mathscr{A} \in M_{mn}(K(\mathbb{C}))$, то произведение $\mathscr{A}\mathscr{B}$ определено, только если $\mathscr{B} \in M_{nr}(K(\mathbb{C}))$.

Операции над интервальными матрицами и векторами были формально введены в определения 3. Для вещественных интервальных операций мы имеем простое определение 1.2. Для интервальных матриц аналогичное определение невозможно, однако в общем случае имеет место

$$\{\mathscr{A}_{\mathrm{p}}\mathscr{B}_{\mathrm{p}} \mid \mathscr{A}_{\mathrm{p}} \in \mathscr{A},\ \mathscr{B}_{\mathrm{p}} \in \mathscr{B}\} \subseteq \{\mathscr{C}_{\mathrm{p}} \mid \mathscr{C}_{\mathrm{p}} \in \mathscr{A}\mathscr{B}\}.$$

Доказательство получается с помощью монотонности отношения включения для интервальных операций. Следующий пример показывает, что равенство не имеет места в общем случае. Пусть

$$\mathscr{A}_{\mathrm{p}} = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}, \qquad u_{\mathrm{p}} = \begin{pmatrix} [0, & 1] \\ [0, & 1] \end{pmatrix}$$

Тогда мы имеем

$$\mathscr{A}_{\mathrm{p}}u = \begin{pmatrix} [0, & 2] \\ [-1, & 1] \end{pmatrix}.$$

и если возьмем

$$x_{\mathrm{p}} = \begin{pmatrix} 2 \\ -1 \end{pmatrix} \in \mathscr{A}_{\mathrm{p}}u,$$

то видим, что не найдется $y_{\mathrm{p}} \in u$, такого что $\mathscr{A}_{\mathrm{p}}y_{\mathrm{p}} = x_{\mathrm{p}}$.

Пусть $\mathscr{A}$, $\mathscr{B} \in M_{mn}(I(\mathbb{R}))$ и $c_p \in V_n(\mathbb{R})$. Тогда справедливы соотношения

$$(1)\qquad \left\{\begin{array}{c} \{\mathscr{A}_p \pm \mathscr{B}_p \mid \mathscr{A}_p \in \mathscr{A},\ \mathscr{B}_p \in \mathscr{B}\} = \mathscr{A} \pm \mathscr{B}, \\ \{\mathscr{A}_p c_p \mid \mathscr{A}_p \in \mathscr{A}\} = \mathscr{A} c_p, \end{array}\right\}$$

которыми мы будем пользоваться в дальнейшем.

Множество интервальных матриц замкнуто относительно операций из определения 3. Множество вещественных или комплексных матриц изоморфно соответствующему множеству точечных матриц. Именно по этой причине в определении 3 использованы те же символы операций, что и для соответствующих вещественных и комплексных операций.

Теперь мы сформулируем некоторые полезные свойства введенных операций.

Теорема 4. *Пусть $\mathscr{A}$, $\mathscr{B}$, $\mathscr{C}$ — интервальные матрицы. Тогда*

$$(2)\qquad \mathscr{A} + \mathscr{B} = \mathscr{B} + \mathscr{A},$$

$$(3)\qquad \mathscr{A} + (\mathscr{B} + \mathscr{C}) = (\mathscr{A} + \mathscr{B}) + \mathscr{C},$$

$$(4)\qquad \mathscr{A} + \mathscr{O}_p = \mathscr{O}_p + \mathscr{A} = \mathscr{A},\ \textit{где}\ \mathscr{O}_p\ \textit{— нулевая матрица},$$

$$(5)\qquad \mathscr{A}\mathscr{I}_p = \mathscr{I}_p \mathscr{A} = \mathscr{A},\ \textit{где}\ \mathscr{I}_p\ \textit{— единичная матрица},$$

$$(6)\qquad \left\{\begin{array}{l} (\mathscr{A} + \mathscr{B})\mathscr{C} \subseteq \mathscr{A}\mathscr{C} + \mathscr{B}\mathscr{C} \\ \mathscr{C}(\mathscr{A} + \mathscr{B}) \subseteq \mathscr{C}\mathscr{A} + \mathscr{C}\mathscr{B} \end{array}\right\}\ (\textit{субдистрибутивность}),$$

$$(7)\qquad (\mathscr{A} + \mathscr{B})\mathscr{C}_p = \mathscr{A}\mathscr{C}_p + \mathscr{B}\mathscr{C}_p,$$

$$(8)\qquad \mathscr{C}_p(\mathscr{A} + \mathscr{B}) = \mathscr{C}_p\mathscr{A} + \mathscr{C}_p\mathscr{B},$$

$$(9)\qquad \left\{\begin{array}{c} \mathscr{A}(\mathscr{B}_p\mathscr{C}_p) \subseteq (\mathscr{A}\mathscr{B}_p)\mathscr{C}_p, \\ (\mathscr{A}_p\mathscr{B})\mathscr{C} \subseteq \mathscr{A}_p(\mathscr{B}\mathscr{C})\ \textit{для}\ \mathscr{C} = -\mathscr{C}, \\ \mathscr{A}_p(\mathscr{B}\mathscr{C}_p) = (\mathscr{A}_p\mathscr{B})\mathscr{C}_p, \\ \mathscr{A}(\mathscr{B}\mathscr{C}) = (\mathscr{A}\mathscr{B})\mathscr{C}\ \textit{для}\ \mathscr{A},\ \mathscr{B},\ \mathscr{C} \in M_{mn}(I(\mathbb{R})) \\ \textit{и}\ \mathscr{B} = -\mathscr{B},\ \mathscr{C} = -\mathscr{C}. \end{array}\right.$$

Доказательство. Соотношения (2)—(8) доказываются поэлементно с использованием формул из теорем 1.4 и 5.8. Докажем (9) для квадратных матриц. Из дистрибутивности (1.8) и формулы (5.6) получаем

$$\mathscr{A}(\mathscr{B}_p\mathscr{C}_p) = \sum_{j=1}^{n} A_{ij}\left(\sum_{l=1}^{n} b_{jl}c_{lk}\right)\Bigg) \subseteq \left(\sum_{j=1}^{n}\sum_{l=1}^{n} A_{ij}b_{jl}c_{lk}\right)$$

$$= \left(\sum_{l=1}^{n}\left(\sum_{j=1}^{n} A_{ij}b_{jl}\right)c_{lk}\right) = (\mathscr{A}\mathscr{B}_p)\mathscr{C}_p,$$

что доказывает первое из соотношений (9). Равенства

$$(\mathscr{A}_{\mathrm{p}}\mathscr{B})\mathscr{C} = \left(\sum_{l=1}^{n}\left(\sum_{k=1}^{n} a_{ik}B_{kl}\right)C_{lj}\right) = \left(\sum_{l=1}^{n}\left|\sum_{k=1}^{n} a_{ik}B_{kl}\right|C_{lj}\right)$$

$$\subseteq \left(\sum_{l=1}^{n}\left(\sum_{k=1}^{n} |a_{ik}||B_{kl}|\right)C_{lj}\right) \subseteq \left(\sum_{l=1}^{n}\left(\sum_{k=1}^{n} |a_{ik}||B_{kl}|C_{lj}\right)\right)$$

$$= \left(\sum_{k=1}^{n} |a_{ik}|\left(\sum_{l=1}^{n} |B_{kl}|C_{lj}\right)\right)$$

$$= \left(\sum_{k=1}^{n} a_{ik}\left(\sum_{l=1}^{n} B_{kl}C_{lj}\right)\right) = \mathscr{A}_{\mathrm{p}}(\mathscr{B}\mathscr{C}).$$

дают второе соотношение. Из равенств

$$(\mathscr{A}_{\mathrm{p}}\mathscr{B})\mathscr{C}_{\mathrm{p}} = \left(\sum_{l=1}^{n}\left(\sum_{k=1}^{n} a_{ik}B_{kl}\right)c_{lj}\right) = \left(\sum_{l=1}^{n}\left(\sum_{k=1}^{n} a_{ik}B_{kl}c_{lj}\right)\right)$$

$$= \left(\sum_{k=1}^{n} a_{ik}\left(\sum_{l=1}^{n} B_{kl}c_{lj}\right)\right)$$

мы получаем третье соотношение.

Последнее соотношение получается следующим образом с помощью третьей из формул (1.8):

$$\mathscr{A}(\mathscr{B}\mathscr{C}) = \left(\sum_{k=1}^{n} A_{ik}\left(\sum_{l=1}^{n} B_{kl}C_{lj}\right)\right)$$

$$= \left(\sum_{k=1}^{n} |A_{ik}|\left(\sum_{l=1}^{n} |B_{kl}|C_{lj}\right)\right)$$

$$= \left(\sum_{k=1}^{n}\left(\sum_{l=1}^{n} |A_{ik}||B_{kl}|C_{lj}\right)\right)$$

$$= \left(\sum_{l=1}^{n}\left(\sum_{k=1}^{n} |A_{ik}||B_{kl}|C_{lj}\right)\right)$$

$$= \left(\sum_{l=1}^{n}\left(\sum_{k=1}^{n} |A_{ik}||B_{kl}|\right)C_{lj}\right)$$

$$= \left(\sum_{l=1}^{n}\left(\sum_{k=1}^{n} |A_{ik}|B_{kl}\right)C_{lj}\right)$$

$$= \left(\sum_{l=1}^{n}\left(\sum_{k=1}^{n} A_{ik}B_{kl}\right)\right)C_{lj} = (\mathscr{A}\mathscr{B})\mathscr{C}.$$

В общем случае ассоциативный закон не имеет места для интервальных матриц. Это показывает следующий

Пример.

$$\begin{bmatrix} [-1, 1] & 1 \\ -1 & [0, 1] \end{bmatrix} \left\{ \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} -1 & 0 \\ 1 & -1 \end{bmatrix} \right\} = \begin{bmatrix} 1 & [-2, 0] \\ [0, 1] & [0, 1] \end{bmatrix},$$

$$\left\{ \begin{bmatrix} [-1, 1] & 1 \\ -1 & [0, 1] \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \right\} \begin{bmatrix} -1 & 0 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} [-1, 3] & [-2, 0] \\ [0, 1] & [0, 1] \end{bmatrix}.$$

Основное свойство монотонности включения справедливо и для интервальных матричных операций.

Теорема 5. *Пусть* $\mathscr{A}^{(k)}$, $\mathscr{B}^{(k)}$, $k = 1, 2$, — *интервальные матрицы. Далее, пусть* X, Y — *интервалы и*

$$\mathscr{A}^{(k)} \subseteq \mathscr{B}^{(k)}, \quad k = 1, 2 \quad \textit{и} \quad X \subseteq Y.$$

Тогда соотношения

$$(10) \qquad \begin{cases} \mathscr{A}^{(1)} * \mathscr{A}^{(2)} \subseteq \mathscr{B}^{(1)} * \mathscr{B}^{(2)}, \\ X\mathscr{A}^{(1)} \subseteq Y\mathscr{B}^{(1)} \end{cases}$$

имеют место для $* \in \{+, -, \cdot\}$. ■

Доказательство соотношений (10) проводится покомпонентно с использованием (1.9) и теоремы 5.9. Имеет место частный случай соотношений (10):

$$\mathscr{A}_{\mathrm{p}} \in \mathscr{A}, \quad \mathscr{B}_{\mathrm{p}} \in \mathscr{B} \Rightarrow \mathscr{A}_{\mathrm{p}} * \mathscr{B}_{\mathrm{p}} \in \mathscr{A} * \mathscr{B}, \ * \in \{+, -, \cdot\},$$

$$x \in X, \quad \mathscr{A}_{\mathrm{p}} \in \mathscr{A} \Rightarrow x\mathscr{A}_{\mathrm{p}} \in X\mathscr{A}.$$

Мы введем теперь понятия ширины и абсолютной величины интервальных матриц.

Определение 6. Пусть $\mathscr{A} = (A_{ij})$ — интервальная матрица. Тогда

(a) вещественная неотрицательная матрица

$$d(\mathscr{A}) := (d(A_{ij}))$$

называется шириной матрицы $\mathscr{A}$;

(b) вещественная неотрицательная матрица

$$|\mathscr{A}| := (|A_{ij}|)$$

называется матрицей абсолютных величин или абсолютной величиной матрицы $\mathscr{A}$. ■

Мы соберем теперь в одном месте некоторые свойства ширины и абсолютной величины интервальных матриц. Частичный порядок

$$\mathscr{X}_{\mathrm{p}} \leqslant \mathscr{Y}_{\mathrm{p}} \Leftrightarrow x_{ij} \leqslant y_{ij}, \quad 1 \leqslant i \leqslant m, \quad 1 \leqslant j \leqslant n$$

используется здесь для вещественных интервальных матриц $\mathscr{X}_{\mathrm{p}}$ и $\mathscr{Y}_{\mathrm{p}}$ размерности $m \times n$. Перечислим эти свойства.

$$\mathscr{A} \subseteq \mathscr{B} \Rightarrow d(\mathscr{A}) \leqslant d(\mathscr{B}), \tag{11}$$

$$d(\mathscr{A} \pm \mathscr{B}) = d(\mathscr{A}) + d(\mathscr{B}), \tag{12}$$

$$d(\mathscr{A}) = \sup_{\mathscr{A}'_{\mathrm{p}}, \mathscr{A}_{\mathrm{p}} \in \mathscr{A}} |\mathscr{A}_{\mathrm{p}} - \mathscr{A}'_{\mathrm{p}}|, \tag{13}$$

$$|\mathscr{A}| = \sup_{\mathscr{A}_{\mathrm{p}} \in \mathscr{A}} |\mathscr{A}_{\mathrm{p}}|, \tag{14}$$

$$\mathscr{A} \subseteq \mathscr{B} \Rightarrow |\mathscr{A}| \leqslant |\mathscr{B}|, \tag{15}$$

$$|\mathscr{A}| \geqslant \mathscr{O}_{\mathrm{p}} \quad \text{и} \quad |\mathscr{A}| = \mathscr{O}_{\mathrm{p}} \Leftrightarrow \mathscr{A} = \mathscr{O}_{\mathrm{p}},$$

$$|\mathscr{A} + \mathscr{B}| \leqslant |\mathscr{A}| + |\mathscr{B}|,$$

$$|x\mathscr{A}| = |\mathscr{A}x| = |x||\mathscr{A}|, \qquad x \in \mathbb{C}, \tag{16}$$

$$\mathscr{A} \in M_{mn}(I(\mathbb{R})) \quad \text{или} \quad \mathscr{A} \in M_{mn}(K(\mathbb{C})),$$

$$|\mathscr{A}\mathscr{B}| \leqslant |\mathscr{A}||\mathscr{B}|,$$

$$d(\mathscr{A}\mathscr{B}) \leqslant d(\mathscr{A})|\mathscr{B}| + |\mathscr{A}|d(\mathscr{B}), \tag{17}$$

$$d(\mathscr{A}\mathscr{B}) \geqslant |\mathscr{A}|d(\mathscr{B}), \; d(\mathscr{A}\mathscr{B}) \geqslant d(\mathscr{A})|\mathscr{B}| \tag{18}$$

$$\begin{cases} d(a\mathscr{B}) = |a|d(\mathscr{B}), \; a \in \mathbb{C}, \\ d(\mathscr{A}_{\mathrm{p}}\mathscr{B}) = |\mathscr{A}_{\mathrm{p}}|d(\mathscr{B}), \quad d(\mathscr{B}\mathscr{A}_{\mathrm{p}}) = d(\mathscr{B})|\mathscr{A}_{\mathrm{p}}|. \end{cases} \tag{19}$$

Для вещественных интервальных матриц $\mathscr{A}$, $\mathscr{B}$ мы имеем

$$\mathscr{O}_{\mathrm{p}}^{\#} \in \mathscr{A} \Rightarrow |\mathscr{A}| \leqslant d(\mathscr{A}) \leqslant 2|\mathscr{A}|, \tag{20}$$

$$\mathscr{A} = (-1)\mathscr{A} \Rightarrow \mathscr{A}\mathscr{B} = \mathscr{A}|\mathscr{B}|, \tag{21}$$

$$\mathscr{O}_{\mathrm{p}} \in \mathscr{A}, \quad 0 \notin \mathscr{B}_{ij} \quad \text{для} \quad \mathscr{B} = (B_{ij}) \Rightarrow d(\mathscr{A}\mathscr{B}) = d(\mathscr{A})|\mathscr{B}|. \tag{22}$$

Доказательство этих свойств проводится покомпонентно с использованием свойств $I(\mathbb{R})$ из гл. 2 и свойств $I(\mathbb{C})$ из гл. 6.

Мы заметим также, что соотношения (20)—(22) неверны для комплексных интервальных матриц. Рассмотрим, например, (21) для матрицы размерности 1×1 с элементами из $R(\mathbb{C})$, т. е. интервал $A = A_1 + iA_2 \in R(\mathbb{C})$. Утверждение $A = (-1)A$ эквивалентно равенствам

$$A_1 = (-1)A_1, \qquad A_2 = (-1)A_2$$

для $A = A_1 + iA_2$. Используя (2.18), мы получаем для $B = B_1 + iB_2$, что

$$\begin{aligned} AB &= (A_1B_1 - A_2B_2) + i(A_1B_2 + A_2B_1) \\ &= A_1|B_1| + A_2|B_2| + i(A_1|B_2| + A_2|B_1|). \end{aligned}$$

С другой стороны, мы имеем

$$A|B| = A_1(|B_1| + |B_2|) + iA_2(|B_1| + |B_2|).$$

Эти два интервала различны в общем случае, например при $B_1 = 0$. Так как соотношения (20)—(22) не потребуются нам для комплексных интервалов, мы не будем рассматривать случаев, в которых они справедливы.

Теперь мы введем понятие матрицы расстояний для пары интервальных матриц.

Определение 7. Пусть $\mathscr{A} = (A_{ij})$ и $\mathscr{B} = (B_{ij})$— интервальные матрицы. Тогда вещественная неотрицательная матрица

$$q(\mathscr{A}, \mathscr{B}) := (q(A_{ij}, B_{ij}))$$

называется матрицей расстояний или расстоянием между матрицами $\mathscr{A}$ и $\mathscr{B}$. ■

Соотношения

$$q(\mathscr{A}, \mathscr{B}) = \mathscr{O}_{\mathrm{p}} \Leftrightarrow \mathscr{A} = \mathscr{B},$$
$$q(\mathscr{A}, \mathscr{B}) \leqslant q(\mathscr{A}, \mathscr{C}) + q(\mathscr{B}, \mathscr{C})$$

очевидным образом справедливы для расстояний между интервальными матрицами вместе с соотношениями

$$q(\mathscr{A} + \mathscr{C}, \mathscr{B} + \mathscr{C}) = q(\mathscr{A}, \mathscr{B}), \tag{23}$$

$$q(\mathscr{A} + \mathscr{B}, \mathscr{C} + \mathscr{D}) = q(\mathscr{A}, \mathscr{C}) + q(\mathscr{B}, \mathscr{D}), \tag{24}$$

$$q(\mathscr{A}\mathscr{B}, \mathscr{A}\mathscr{C}) \leqslant |\mathscr{A}|\, q(\mathscr{B}, \mathscr{C}). \tag{25}$$

Доказательства последних свойств проводятся покомпонентно с использованием соответствующих свойств $I(\mathbb{R})$ (см. гл. 2) или $I(\mathbb{C})$ (см. гл. 6). С помощью понятия расстояния между интервальными матрицами, введенного в определении 7, можно, используя монотонную норму матриц $\|\cdot\|$, определить метрику на множестве интервальных матриц (см., например, [195]) как $\|q(\mathscr{A},\mathscr{B})\|$. Множество всех интервальных матриц размерности $m \times n$ можно также рассматривать как $m \cdot n$-кратное произведение полного метрического пространства $I(\mathbb{C})$ на себя. Известные теоремы топологии (см., например, [318]) показывают, что это произведение снова является полным метрическим пространством. Сходимость в произведении пространств эквивалентна сходимости отдельных компонент. Поэтому справедливы следующие утверждения.

(26) Сходимость последовательности $\{\mathscr{A}^{(k)}\}_{k=0}^{\infty}$ интервальных матриц размерности $m \times n$ к матрице $\mathscr{A}$, т. е. $\lim\limits_{k\to\infty} \mathscr{A}^{(k)} = \mathscr{A}$, эквивалентна

$$\lim_{k\to\infty} A_{ij}^{(k)} = A_{ij}, \quad 1 \leqslant i \leqslant m, \quad 1 \leqslant j \leqslant n.$$

Следствие 8. *Любая последовательность интервальных матриц* $\{\mathscr{A}^{(k)}\}_{k=0}^{\infty}$ *размерности* $m \times n$, *для которой имеет место*

$$\mathscr{A}^{(0)} \supseteq \mathscr{A}^{(1)} \supseteq \mathscr{A}^{(2)} \supseteq \ldots,$$

сходится к интервальной матрице $\mathscr{A} = (A_{ij})$, *где*

$$A_{ij} = \bigcap_{k=0}^{\infty} A_{ij}^{(k)}, \quad 1 \leqslant i \leqslant m, \quad 1 \leqslant j \leqslant n. \blacksquare$$

Это утверждение следует из (26) определения 2 и утверждения об интервалах, аналогичного следствию (8).

Следствие 9. *Операции, введенные в определении* 3, *непрерывны.* ■

Доказательство получается из того, что непрерывность операций на элементах влечет за собой непрерывность операций в целом. В силу определения 3 и теорем 2.6, 6.6 элементы результата операции непрерывно зависят от операндов.

Следующее соотношение справедливо ввиду (2.21) и (5.17).

$$(27) \quad \mathscr{X} \subseteq \mathscr{Y} \Rightarrow \frac{1}{2}(d(\mathscr{Y}) - d(\mathscr{X})) \leqslant q(\mathscr{X}, \mathscr{Y}) \leqslant d(\mathscr{Y}) - d(\mathscr{X}).$$

Теперь мы хотим ввести операцию пересечения на $M_{mn}(I(\mathbb{R}))$ и $M_{mn}(I(\mathbb{C}))$ таким же образом, как это было сделано в гл. 2 для элементов множества $I(\mathbb{R})$ и в гл. 6 для элементов множества $\mathbb{R}(\mathbb{C})$. Так как $M_{mn}(I(\mathbb{R})) \subset M_{mn}(R(\mathbb{C}))$, достаточно определить эту операцию на $M_{mn}(R(\mathbb{C}))$. Пусть $\mathscr{A} = (A_{ij})$, $\mathscr{B} = (B_{ij}) \in M_{mn}(R(\mathbb{C}))$. Тогда определяем пересечение $\mathscr{A}$ и $\mathscr{B}$ как теоретико-множественное пересечение

$$\mathscr{A} \cap \mathscr{B} = \{\mathscr{C}_p \mid \mathscr{C}_p \in \mathscr{A}, \mathscr{C}_p \in \mathscr{B}\}.$$

Пересечение двух интервальных матриц $\mathscr{A}$ и $\mathscr{B}$ принадлежит множеству $M_{mn}(R(\mathbb{C}))$ тогда и только тогда, когда это теоретико-множественное пересечение непусто. В этом случае мы имеем

$$\mathscr{A} \cap \mathscr{B} = (A_{ij} \cap B_{ij}),$$

где $A_{ij} \cap B_{ij}$, $1 \leqslant i \leqslant n$, $1 \leqslant j \leqslant m$, строится согласно (2.23) (соответственно (6.19)).

Аналогично следствиям 2.12 и 6.7 мы получаем

Следствие 10. *Пусть* $\mathscr{A}, \mathscr{B}, \mathscr{C}, \mathscr{D} \in M_{mn}(R(\mathbb{C}))$. *Тогда имеем*

$$\mathscr{A} \subseteq \mathscr{C}, \mathscr{B} \subseteq \mathscr{D} \Rightarrow \mathscr{A} \cap \mathscr{B} \subseteq \mathscr{C} \cap \mathscr{D} \quad (\textit{монотонность включения})$$

и пересечение, если оно не выводит за пределы $M_{mn}(R(\mathbb{C}))$, *является непрерывной операцией.* ■

Как и в случае следствий 2.12 и 6.7, утверждение следует из того, что непрерывность операций на элементах влечет за собой непрерывность операций в целом.

Чтобы обобщить понятие билинейности на операторы из $V_n(\mathbb{C}) \times V_n(\mathbb{C})$ в $V_n(\mathbb{C})$, мы рассмотрим теперь трехмерные массивы с интервальными элементами. Множество всех таких массивов обозначается через $M_{n^3}(I(\mathbb{C}))$. Имеем

$$\mathscr{B} = (B_{ijk}) \in M_{n^3}(I(\mathbb{C})),$$

где

$$B_{ijk} \in I(\mathbb{C}), \quad 1 \leqslant i, j, \quad k \leqslant n.$$

Равенство, включение и сложение определяются поэлементно, т. е. так же, как для интервальных матриц. Аналогично определяются ширина, расстояние и абсолютная величина. Например, определение

$$|\mathscr{B}| := (|\mathscr{B}_{ijk}|)$$

вводит билинейный оператор из $V_n(\mathbb{R}) \times V_n(\mathbb{R})$ в $V_n(\mathbb{R})$. Множество всех билинейных операторов из $V_n(\mathbb{R}) \times V_n(\mathbb{R})$ в $V_n(\mathbb{R})$ обозначается через $M_{n^3}(\mathbb{R})$.

Определение 11. Пусть $\mathscr{B} = (B_{ijk}) \in M_{n^3}(I(\mathbb{C}))$, $x = (X_i)$ $y = (Y_i) \in V_n(I(\mathbb{C}))$ и $\mathscr{A} \in M_{nn}(I(\mathbb{C}))$.

Тогда полагаем

(a) $$\mathscr{C} := \mathscr{B}x \in M_{nn}(I(\mathbb{C})), \quad \text{где}$$

$$C_{ij} = \sum_{k=1}^{n} B_{ijk} X_k, \quad 1 \leqslant i, j \leqslant n,$$

(b) $$z := (\mathscr{B}x)\, y \in V_n(I(\mathbb{C})), \text{ где}$$

$$Z_i = \sum_{j=1}^{n} C_{ij} Y_j = \sum_{j=1}^{n} \left(\sum_{k=1}^{n} B_{ijk} X_k \right) Y_j, \quad 1 \leqslant i < n.$$

Вместо $\mathscr{B}(x)\, y$ мы будем иногда писать $\mathscr{B}xy$.

(c) Далее полагаем

$$\mathscr{C} := \mathscr{A}\mathscr{B} \in M_{n^3}(I(\mathbb{C})),$$

где

$$C_{ijk} = \sum_{\nu=1}^{n} A_{i\nu} B_{\nu jk}, \quad 1 \leqslant i, j, \quad k \leqslant n. \blacksquare$$

Следующая теорема содержит некоторые соотношения, нужные в дальнейшем.

Теорема 12. *Пусть*

$$\mathscr{A}_{\mathrm{p}} = (a_{ij}) \in M_{nn}(\mathbb{C}), \quad \mathscr{B} = (B_{ijk}) \in M_{n^3}(I(\mathbb{C})),$$
$$x = (X_i), \quad y = (Y_i) \in V_n(I(\mathbb{C})).$$

Тогда

(a) $$(\mathscr{A}_{\mathrm{p}}\mathscr{B})\,xy \subseteq \mathscr{A}_{\mathrm{p}}(\mathscr{B}xy).$$

Если $x = -x$, то

(b) $$\mathscr{B}xx = \frac{1}{2}|\,\mathscr{B}\,|\,d(x)\,x$$

и

(c) $$d(\mathscr{B}xx) = \frac{1}{2}|\,\mathscr{B}\,|\,d(x)\,d(x).$$

Доказательство. (a) Пусть $\mathscr{C} = \mathscr{A}_{\mathrm{p}}\mathscr{B} = (C_{ijk})$ и $z = (Z_i) = \mathscr{B}xy$. Тогда с помощью определения 11 и субдистрибутивности мы получаем

$$(\mathscr{A}_{\mathrm{p}}\mathscr{B})\,xy = \left(\sum_{j=1}^{n}\left(\sum_{k=1}^{n} C_{ijk}X_k\right)Y_j\right) = \left(\sum_{j=1}^{n}\left(\sum_{k=1}^{n}\left(\sum_{\nu=1}^{n} a_{i\nu}B_{\nu jk}\right)X_k\right)Y_j\right)$$

$$\subseteq \left(\sum_{j=1}^{n}\left(\sum_{k=1}^{n}\left(\sum_{\nu=1}^{n} a_{i\nu}B_{\nu jk}X_k\right)\right)Y_j\right)$$

$$= \left(\sum_{j=1}^{n}\left(\sum_{\nu=1}^{n}\left(\sum_{k=1}^{n} a_{i\nu}B_{\nu jk}X_k\right)\right)Y_j\right)$$

$$= \left(\sum_{j=1}^{n}\left(\sum_{\nu=1}^{n} a_{i\nu}\left(\sum_{k=1}^{n} B_{\nu jk}X_k\right)\right)Y_j\right)$$

$$\subseteq \left(\sum_{j=1}^{n}\left(\sum_{\nu=1}^{n} a_{i\nu}\left(\sum_{k=1}^{n} B_{\nu jk}X_k\right)Y_j\right)\right)$$

$$= \left(\sum_{\nu=1}^{n}\left(\sum_{j=1}^{n} a_{i\nu}\left(\sum_{k=1}^{n} B_{\nu jk}X_k\right)Y_j\right)\right)$$

$$= \left(\sum_{\nu=1}^{n} a_{i\nu}\left(\sum_{j=1}^{n}\left(\sum_{k=1}^{n} B_{\nu jk}X_k\right)Y_j\right)\right)$$

$$= \left(\sum_{\nu=1}^{n} a_{i\nu}Z_\nu\right) = \mathscr{A}_{\mathrm{p}}(\mathscr{B}xy).$$

(b) Полагая $z = \mathscr{B}xx$ и используя симметричность по x, получаем в обозначениях определения 11 (b), что

$$Z_i = \sum_{j=1}^{n}\left(\sum_{k=1}^{n} B_{ijk}X_k\right)X_j = \sum_{j=1}^{n}\left|\sum_{k=1}^{n} B_{ijk}X_k\right|X_j$$

$$= \sum_{j=1}^{n}\left|\left(\sum_{k=1}^{n}|\,B_{ijk}\,|\,X_k\right) = \sum_{j=1}^{n}\left(\sum_{k=1}^{n}\frac{1}{2}|\,B_{ijk}\,|\,d(X_k)\right)X_j.\right.$$

Поэтому $\mathscr{B}xx = \frac{1}{2}|\,\mathscr{B}\,|\,d(x)\,x.$

(c) Используя соотношения из п. (b), получаем

$$d(Z_i) = \sum_{j=1}^{n} \left(\sum_{k=1}^{n} \tfrac{1}{2} |B_{ijk}| d(X_k) \right) d(X_j),$$

т. е.

$$d(\mathscr{B}xx) = \tfrac{1}{2} |\mathscr{B}| d(x) d(x).$$

Замечание. Операции над интервальными матрицами применялись в работе [36] еще без систематического исследования их свойств. Приведенные выше утверждения о трехмерных массивах с интервальными элементами были доказаны в [469a].

* * *

Литература: [36], [175], [195], [197], [215], [250], [318],[322], [469a].

Глава 11

ИТЕРАЦИОННАЯ ЛОКАЛИЗАЦИЯ НЕПОДВИЖНОЙ ТОЧКИ ДЛЯ СИСТЕМ НЕЛИНЕЙНЫХ УРАВНЕНИЙ

Рассмотрим теперь функцию $f(x_\mathrm{p})$ от векторной переменной $x_\mathrm{p} = (x_1, \ldots, x_n)^T \in V_n(\mathbb{C})$, принимающую значения в $\mathbb{C}$. Будем предполагать, что функция f построена при помощи основных арифметических операций, а также стандартных функций синус, косинус и т. д. Это означает, что ее можно вычислять и как интервальную функцию. Предположим еще, что функция f зависит от m параметров $a_1, a_2, \ldots, a_m \in \mathbb{C}$. Таким образом, мы можем записать f в виде

$$f(x_\mathrm{p}) = f(x_1, x_2, \ldots, x_n; a_1, a_2, \ldots, a_m).$$

Пусть теперь заданы n функции такого вида $f_i(x_\mathrm{p})$, $1 \leqslant i \leqslant n$. Тогда соотношение

$$y_\mathrm{p} = \mathscr{f}_\mathrm{p}(x_\mathrm{p}) = (f_i(x_\mathrm{p}))$$

определяет отображение из $V_n(\mathbb{C})$ в $V_n(\mathbb{C})$, а соотношение

$$y = \mathscr{f}_\mathrm{p}(x) = (f_i(x))$$

определяет отображение на множестве n-компонентных интервальных векторов

$$\mathscr{f}_\mathrm{p}\colon\ V_n(I(\mathbb{C})) \to V_n(I(\mathbb{C})), \quad \text{где} \quad \mathscr{f}_\mathrm{p}(x) = (f_i(x)).$$

Ниже мы собираемся использовать интервальную арифметику для локализации решений системы уравнений

$$\mathcal{f}_{\mathrm{p}}(x_{\mathrm{p}}) = o_{\mathrm{p}},$$

где

$$\mathcal{f}_{\mathrm{p}}(x_{\mathrm{p}}) = (f_i(x_{\mathrm{p}}))$$

и

$$f_i(x_{\mathrm{p}}) = f_i(x_1, \ldots, x_n; a_{i1}, \ldots, a_{im_i}) \qquad 1 \leqslant i \leqslant n,$$

в предположении, что параметры a_{ij} независимо изменяются в некоторых комплексных интервалах.

Много возможностей для решения этой задачи дает метод итераций. Заметим, что данное уравнение всегда может быть преобразовано к виду

$$x_{\mathrm{p}} = \mathcal{f}_{\mathrm{p}}(x_{\mathrm{p}}).$$

Вычисление правой части этого уравнения для произвольного интервального вектора $x^{(0)}$ дает интервальный вектор $x^{(1)}$. Продолжая в том же духе, мы получаем метод итераций

$$x^{(k+1)} = \mathcal{f}_{\mathrm{p}}(x^{(k)}), \qquad k \geqslant 0.$$

Немедленно возникают следующие вопросы. (a) Когда существует последовательность $\{x^{(k)}\}_{k=0}^{\infty}$? (b) Когда эта последовательность сходится? (c) Когда предел x^* единствен? (d) Какое отношение имеет предел x^* к решению сформулированной выше задачи?

Сначала мы докажем элементарную теорему о неподвижной точке, опирающуюся на монотонность интервального оценивания функций относительно включения (см. также гл. 3).

Теорема 1. *Пусть дано отображение*

$$\mathcal{f}_{\mathrm{p}}\colon V_n(\mathbb{C}) \to V_n(\mathbb{C}), \quad \text{где} \quad \mathcal{f}_{\mathrm{p}}(x_{\mathrm{p}}) = (f_i(x_{\mathrm{p}})),$$

причем функции $f_i(x_{\mathrm{p}})$ имеют указанный выше вид. Рассмотрим метод итераций в $V_n(I(\mathbb{C}))$, заданный соотношением

$$x^{(k+1)} = \mathcal{f}_{\mathrm{p}}(x^{(k)}), \qquad k \geqslant 0,$$

и удовлетворяющий условию

$$x^{(1)} \subseteq x^{(0)}.$$

Тогда имеет место следующее.

(1) *Последовательность результатов итерации $\{x^{(k)}\}_{k=0}^{\infty}$ сходится к пределу x, такому что $x = \mathcal{f}_{\mathrm{p}}(x)$.*
(2) *Любой вектор $x_{\mathrm{p}} \in x^{(0)}$, удовлетворяющий уравнению $x_{\mathrm{p}} = \mathcal{f}_{\mathrm{p}}(x_{\mathrm{p}})$, содержится в x, т. е. имеет место*

$$\{x_{\mathrm{p}} \mid x_{\mathrm{p}} \in x^{(0)},\ x_{\mathrm{p}} = \mathcal{f}_{\mathrm{p}}(x)\} \subseteq x$$

Доказательство. (1) По предположению имеем $x^{(1)} \subseteq x^{(0)}$. Так как интервальные вычисления монотонны относительно включения, мы получаем

$$x^{(2)} = f_{\mathrm{p}}(x^{(1)}) \subseteq f_{\mathrm{p}}(x^{(0)}) = x^{(1)} \subseteq x^{(0)}.$$

С помощью математической индукции можно показать, что

$$\ldots \subseteq x^{(3)} \subseteq x^{(2)} \subseteq x^{(1)} \subseteq x^{(0)}.$$

Из следствия 10.8 вытекает, что эта последовательность сходится к некоторому элементу $x \in V_n(I(\mathbb{C}))$. Из непрерывности интервальных оценок следует, что

$$x = \lim_{k \to \infty} x^{(k)} = \lim_{k \to \infty} f_{\mathrm{p}}(x^{(k)}) = f_{\mathrm{p}}(x).$$

(2). Пусть $x_{\mathrm{p}} \in x^{(0)}$ и $x_{\mathrm{p}} = f_{\mathrm{p}}(x_{\mathrm{p}})$.

Из монотонности интервальных операций относительно включения снова следует, что

$$x_{\mathrm{p}} = f_{\mathrm{p}}(x_{\mathrm{p}}) \in f_{\mathrm{p}}(x^{(0)}) = x^{(1)},$$

и по индукции получаем

$$x \in x^{(k)}, \qquad k \geqslant 0,$$

откуда следует, что $x_{\mathrm{p}} \in x$. ■

Условия теоремы 1 гарантируют существование неподвижной точки, но не ее единственность. Это показывает следующий пример.

Пример. Рассмотрим уравнение

$$X = X \cdot X \cdot X$$

в $R(\mathbb{C})$.Очевидно, что этому уравнению удовлетворяют следующие элементы $R(\mathbb{C})$:

$$X = [-1, 1], [1, 1], [-1, -1], [0, 1], [-1, 0], [0, 0], i[-1, 1].$$

Мы докажем теперь другую теорему о неподвижной точке, имеющую несколько иные условия и использующую несколько иной итерационный процесс.

Теорема 2. *Пусть задано отображение*

$$f_{\mathrm{p}}\colon V_n(\mathbb{C}) \to V_n(\mathbb{C}), \quad \textit{где} \quad f_{\mathrm{p}}(x_{\mathrm{p}}) = (f_i(x_{\mathrm{p}}))$$

с функциями $f_i(x_{\mathrm{p}})$ указанного выше вида.

Рассмотрим итерационный процесс

$$x^{(k+1)} = f_{\mathrm{p}}(x^{(k)}) \cap x^{(k)}, \qquad k \geqslant 0,$$

в $V_n(R(\mathbb{C}))$. Предположим, что существует элемент $\tilde{x}_{\mathrm{p}} \in x^{(0)}$, удовлетворяющий уравнению $\tilde{x}_{\mathrm{p}} = f_{\mathrm{p}}(\tilde{x}_{\mathrm{p}})$. Тогда верно следующее.

(3) *Последовательные приближения* $\{x^{(k)}\}_{k=0}^{\infty}$ *удовлетворяют условию* $\lim\limits_{k\to\infty} x^{(k)} = x$, *причем* $x = \mathcal{f}_{\mathrm{p}}(x) \bigcap x$.

(4) *Любой вектор* $x_{\mathrm{p}} \in x^{(0)}$, *удовлетворяющий уравнению* $x_{\mathrm{p}} = \mathcal{f}_{\mathrm{p}}(x_{\mathrm{p}})$, *содержится в* x, *т. е.*

$$\{x_{\mathrm{p}} \mid x_{\mathrm{p}} \in x^{(0)},\ x_{\mathrm{p}} = \mathcal{f}_{\mathrm{p}}(x_{\mathrm{p}})\} \subseteq x.$$

Доказательство. Рассуждением, аналогичным доказательству теоремы 1, из $\tilde{x}_{\mathrm{p}} \in x^{(0)}$ мы получаем соотношение $x_{\mathrm{p}} \in \mathcal{f}_{\mathrm{p}}(x^{(0)}) \bigcap x^{(0)} = x^{(1)}$. Применяя индукцию, получаем $x_{\mathrm{p}} \in x^{(k)}$, $k \geqslant 0$. Так как пересечение здесь всегда непусто, мы получаем последовательность интервалов

$$x^{(0)} \supseteq x^{(1)} \supseteq \ldots,$$

которая в силу следствия 10.8 сходится к некоторому пределу x. Ввиду непрерывности интервальных вычислений и пересечений мы имеем $x = \mathcal{f}_{\mathrm{p}}(x) \bigcap x$ для этого предела, а также $\hat{x}_{\mathrm{p}} \in x$ для всех $\hat{x}_{\mathrm{p}} = \mathcal{f}_{\mathrm{p}}(\hat{x}_{\mathrm{p}}) \in x^{(0)}$. ■

По поводу единственности неподвижной точки в теореме 2 можно сказать то же самое, что и в случае теоремы 1. Эти две элементарные теоремы о неподвижной точке подробно рассмотрены в работе [232]. Мы приведем теперь две теоремы о неподвижной точке, которые будут применены в дальнейшем к двум конкретным итерационным процедурам. В отличие от теорем 1 и 2, имеющих довольно общий характер, эти новые теоремы обеспечивают единственность неподвижной точки. Сначала введем одно понятие.

Определение 3. (См. [295].) Пусть

$$\mathcal{f}_{\mathrm{p}}:\ V_n(\mathbb{C}) \to V_n(\mathbb{C}), \qquad \mathcal{f}_{\mathrm{p}}(x_{\mathrm{p}}) = (f_i(x_{\mathrm{p}})),$$

— отображение указанного выше вида. $\mathcal{f}_{\mathrm{p}}$ называется $\mathcal{P}_{\mathrm{p}}$-сжатием, если существует неотрицательная матрица $\mathcal{P}_{\mathrm{p}}$, такая что

$$q(\mathcal{f}_{\mathrm{p}}(x), \mathcal{f}_{\mathrm{p}}(y)) \leqslant \mathcal{P}_{\mathrm{p}} q(x, y) \quad \text{для всех} \quad x, y \in V_n(I(\mathbb{C})),$$

где

$$\rho(\mathcal{P}_{\mathrm{p}}) < 1. \ \blacksquare$$

Здесь ρ обозначает спектральный радиус матрицы $\mathcal{P}_{\mathrm{p}}$, а q обозначает расстояние между двумя интервальными векторами, определенное в гл. 10. Докажем следующее утверждение.

Теорема 4. *Если* $\mathcal{f}_{\mathrm{p}}:\ V_n(\mathbb{C}) \to V_n(\mathbb{C})$ *есть* $\mathcal{P}_{\mathrm{p}}$*-сжатие, то уравнение* $x = \mathcal{f}_{\mathrm{p}}(x)$ *имеет единственную неподвижную точку* $x^* \in V_n(I(\mathbb{C}))$. *При этом для любого* $x^{(0)} \in V_n(I(\mathbb{C}))$ *итерации*

$$x^{(k+1)} = \mathcal{f}_{\mathrm{p}}(x^{(k)}), \qquad k \geqslant 0,$$

сходятся к x^*.

Доказательство. Из того, что $\rho(\mathscr{P}_p) < 1$ и $\mathscr{P}_p \geqslant \mathscr{O}_p$, следует существование матрицы $(\mathscr{I}_p - \mathscr{P}_p)^{-1}$ и соотношение

$$(\mathscr{I}_p - \mathscr{P}_p)^{-1} = \sum_{l=0}^{\infty} \mathscr{P}_p^l \geqslant \sum_{l=0}^{m-1} \mathscr{P}_p^l \geqslant \mathscr{O}_p.$$

Тогда мы получаем для любого k и $m \geqslant 1$

$$q\left(x^{(k+m)}, x^{(k)}\right) \leqslant \sum_{l=0}^{m-1} \mathscr{P}_p^l q\left(x^{(k+1)}, x^{(k)}\right) \leqslant (\mathscr{I}_p - \mathscr{P}_p)^{-1} \mathscr{P}^k q\left(x^1, x^{(0)}\right).$$

Так как $\lim\limits_{k\to\infty} \mathscr{P}_p^k = \mathscr{O}_p$, каждая компонента последовательности $\{x^{(k)}\}_{k=0}^{\infty}$, а значит, и сама эта последовательность удовлетворяют условию сходимости Коши. Так как пространство $V_n(I(\mathbb{C}))$ полно, а отображение $\mathscr{P}_p$ является сжатием и потому непрерывно, мы получаем

$$\lim_{k\to\infty} x^{(k)} = x^* \quad \text{и} \quad x^* = \mathscr{f}_p(x^*).$$

Единственность неподвижной точки следует из соотношений

$$q(x^*, y^*) = q(\mathscr{f}_p(x^*), \mathscr{f}_p(y^*)) \leqslant \mathscr{P}_p q(x^*, y^*)$$

и $(\mathscr{I}_p - \mathscr{P}_p)^{-1} \geqslant \mathscr{O}_p$. ■

Эта теорема — частный случай более общего результата, доказанного Шрёдером [295].

Вот еще одна теорема о неподвижной точке, которая будет использована в дальнейшем.

Теорема 5. *Пусть*

$$\mathscr{f}_p\colon V_n(\mathbb{C}) \to V_n(\mathbb{C})$$

и

$$g\colon V_n(\mathbb{C}) \times V_n(\mathbb{C}) \to V_n(\mathbb{C}),$$

где $\mathscr{f}_p$ *и* g_p *имеют описанный выше вид, причем*

$$g_p(x, x) = \mathscr{f}_p(x) \quad \textit{для всех} \quad x \in V_n(I(\mathbb{C})),$$

$$q(g_p(x, z), g(y, z)) \leqslant \mathscr{Q}_p q(x, y),$$

$$q(g(z, x), g(z, y)) \leqslant \mathscr{R}_p q(x, y) \quad \textit{для всех} \quad x, y, z \in V_n(I(\mathbb{C}))$$

и

$$\mathscr{Q}_p \geqslant \mathscr{O}_p, \quad \mathscr{R}_p \geqslant \mathscr{O}_p, \quad \rho(\mathscr{Q}_p) < 1, \quad \rho\left((\mathscr{I}_p - \mathscr{Q}_p)^{-1} \mathscr{R}_p\right) < 1.$$

Тогда уравнение $\mathscr{f}_p(x) = x$ *имеет единственную неподвижную точку* $x^* \in V_n(I(\mathbb{C}))$, *и для любого* $x^{(0)}$ *существует единственная последовательность* $\{x^{(k)}\}_{k=0}^{\infty}$, *удовлетворяющая уравнению*

$$x^{(k+1)} = g_p(x^{k+1}, x^k), \quad k \geqslant 0$$

При этом

$$\lim_{k\to\infty} x^{(k)} = x^*.$$

Доказательство. Матрицу $(\mathscr{I} - \mathscr{Q})^{-1}\mathscr{R}_p$ можно рассматривать как итерационную матрицу, соответствующую матрице $\mathscr{I}_p - \mathscr{Q}_p - \mathscr{R}_p$. Из $\mathscr{P}_p = \mathscr{Q}_p + \mathscr{R}_p \geqslant \mathscr{O}_p$ и того, что $\rho(\mathscr{P}_p) < 1$, следует $\rho((\mathscr{I}_p - \mathscr{Q}_p)^{-1}\mathscr{R}_p) < 1$ (см. также [35] и [325]). Отсюда с помощью (10.23) и (10.24) получаем для произвольных x, $y \in V_n(I_n(\mathbb{C}))$, что

$$\begin{aligned} q(\mathscr{f}_p(x), \mathscr{f}_p(y)) &\leqslant q(\mathscr{g}_p(x, x), \mathscr{g}_p(x, y)) + q(\mathscr{g}_p(x, y), \mathscr{g}_p(y, y)) \\ &\leqslant (\mathscr{R}_p + \mathscr{Q}_p) q(x, y) \end{aligned}$$

т. е. что $\mathscr{f}_p$ есть $\mathscr{P}_p$-сжатие. В силу предыдущей теоремы уравнение $\mathscr{f}_p(x) = x$ имеет единственную неподвижную точку $x^* \in V_n(I_n(\mathbb{C}))$. Из наших предположений следует, что отображение $\mathscr{g}(\cdot, z)$ для фиксированного $z \in V_n(I_n(\mathbb{C}))$ является $\mathscr{Q}_p$-сжатием. Применяя предыдущую теорему, получаем, что $\mathscr{g}(\cdot, x^{(k)})$ имеет единственную неподвижную точку $x^{(k+1)}$. Таким образом, существование последовательности $\{x^{(k)}\}_{k=0}^{\infty}$ установлено для произвольного $x^{(0)} \in V_n(I(\mathbb{C}))$. Из того, что

$$\begin{aligned} q(x^{(k+1)}, x^*) &\leqslant q(\mathscr{g}_p(x^{(k+1)}, x^{(k)}), \mathscr{g}(x^*, x^{(k)})) + q(\mathscr{g}(x^*, x^{(k)}), \mathscr{g}(x^*, x^*)) \\ &\leqslant \mathscr{Q}_p q(x^{(k+1)}, x^*) + \mathscr{R}_p q(x^{(k)}, x^*) \end{aligned}$$

или

$$\begin{aligned} q(x^{(k+1)}, x^*) &\leqslant (\mathscr{I}_p - \mathscr{Q}_p)^{-1}\mathscr{R}_p q(x^{(k)}, x^*) \leqslant \\ &\leqslant ((\mathscr{I}_p - \mathscr{Q}_p)^{-1}\mathscr{R})^{k+1} q(x^{(0)}, x^*), \end{aligned}$$

следует, что $\lim\limits_{k\to\infty} x^{(k)} = x^*$, так как $\rho((\mathscr{I}_p - \mathscr{Q}_p)^{-1}\mathscr{R}_p) < 1$. Это завершает доказательство теоремы. ■

Результат из теоремы 5 был сначала получен для отображений из $V_n(\mathbb{R})$ в $V_n(\mathbb{R})$ (см. [243]).

В связи с двумя последними теоремами мы хотим продемонстрировать соотношение между единственной неподвижной точкой и потенциальными решениями уравнения

$$x_p = \mathscr{f}_p(x_p), \qquad x_p \in V_n(\mathbb{C}).$$

Следствие 6. *Пусть задано отображение*

$$\mathscr{f}\colon\ V_n(\mathbb{C}) \to V_n(\mathbb{C}), \qquad \mathscr{f}_p(x_p) = (f_i(x_p)),$$

причем

$$f_i(x) = f_i(X_1, X_2, \ldots, X_n; A_{i1}, \ldots, A_{im_i}), \quad 1 \leqslant i \leqslant n.$$

Пусть выполнены условия одной из теорем 4, 5, *и пусть* x^* — *единственная неподвижная точка уравнения* $x = \mathscr{f}_p(x)$, *суще-*

ствование которой доказано там. Тогда

$$\{x_{\mathrm{p}} \mid x_{\mathrm{p}} = f_{\mathrm{p}}(x_{\mathrm{p}}),\quad a_{ij} \in A_{ij},\quad 1 \leqslant i \leqslant n,\quad 1 \leqslant j \leqslant m_i\} \subseteq x^*.$$

Доказательство. Рассмотрим уравнение

$$x_{\mathrm{p}} = f_{\mathrm{p}}(x_{\mathrm{p}}),\qquad x_{\mathrm{p}} \in V_n(\mathbb{C}),$$

при фиксированном выборе элементов $a_{ij} \in A_{ij}$, $1 \leqslant i \leqslant n$, $1 \leqslant j \leqslant m_i$, и допустим, что ему удовлетворяет элемент $x^*_{\mathrm{p}} \in V_n(\mathbb{C})$. Мы можем тогда начать итерации в теореме 4 или 5 со значения $x^{(0)} = x^*_{\mathrm{p}}$ и в пределе получим неподвижную точку x^*. Так же, как это было сделано в доказательстве п. (2) теоремы 1, мы можем использовать монотонность включения, чтобы показать, что всегда имеет место

$$x^*_{\mathrm{p}} \in x^{(k)},\qquad k \geqslant 0.$$

Отсюда следует $x^*_{\mathrm{p}} \in x^*$. ■

Теперь мы рассмотрим практическое нахождение констант Липшица для интервальных вычислений. Мы увидим, что константы Липшица для интервальных вычислений будут мажорировать константы Липшица для соответствующих точечных функций. Это означает, что каждая из систем $x_{\mathrm{p}} = f(x_{\mathrm{p}})$, рассмотренных в следствии 6, удовлетворяет условиям теорем 4 и 5 при ограничении на множество $V_n(\mathbb{C})$, а потому имеет единственное решение x^*_{p}.

С помощью найденных констант Липшица мы сможем проверить, выполнены ли в конкретных условиях предположения теоремы 4 или 5. Для простоты мы ограничимся пространством $V_n(I(\mathbb{R}))$. Аналогичные формулы для $V_n(I(\mathbb{C}))$ могут быть получены без труда. Для подготовки мы докажем сейчас одно свойство метрики q, которое следует из того, что она является метрикой Хаусдорфа (ср. [215]).

Лемма 7. *Пусть* $Y, Z \in I(\mathbb{R})$ *и* $\alpha \geqslant 0$. *Тогда*

$q(Y, Z) \leqslant \alpha \Leftrightarrow$ {*для любого* $y \in Y$ *существует* $z \in Z$ *со свойством* $|y - z| \leqslant \alpha$, *и для любого* $z \in Z$ *существует* $y \in Y$ *со свойством* $|z - y| \leqslant \alpha$}.

Доказательство. Если $Y = [y_1, y_2]$ и $Z = [z_1, z_2]$, то $q(Y, Z) = \max\{|y_1 - z_1|, |y_2 - z_2|\}$. Докажем сначала импликацию $\Rightarrow$. Пусть $q(Y, Z) \leqslant \alpha$ и зафиксирован $y \in Y$. Если $y \in Z$, то можно взять $z = y$, что дает первую половину правой части $\Rightarrow$. Если же $y \notin Z$, то при $z_1 > y$ мы получаем для $z = z_1$, что

$$\alpha \geqslant |z_1 - y_1| = z_1 - y_1 \geqslant z_1 - y = |z_1 - y|,$$

а в случае $z_2 < y$ получаем для $z = z_2$, что

$$\alpha \geqslant |z_2 - y_2| = y_2 - z_2 \geqslant y - z_2 = |y - z_2|.$$

Так как в этом рассуждении y и z равноправны, оно дает и вторую половину правой части $\Rightarrow$ для фиксированного $z \in Z$.

Докажем обратную импликацию $\Leftarrow$. Пусть сначала $y_1 \leqslant z_1$. Зафиксируем $y = y_1$. Тогда найдется $z \geqslant z_1$, такое что

$$\alpha \geqslant |z - y_1| = z - y_1 \geqslant z_1 - y_1 = |z_1 - y_1|.$$

Если же $z_1 < y_1$, то зафиксируем $z = z_1$. Тогда найдется $y \geqslant y_1$, для которого

$$\alpha \geqslant |y - z_1| = y - z_1 \geqslant y - z_1 = |y_1 - z_1|.$$

Таким образом, всегда получается $|y_1 - z_1| \leqslant \alpha$. Соотношение $|y_2 - z_2| \leqslant \alpha$ доказывается тем же методом. Это и дает $q(Y, Z) \leqslant \alpha$. ■

Теперь мы используем эту лемму, чтобы получить одно утверждение о константах Липшица для интервальных вычислений.

Теорема 8. *Пусть f — вещественная функция вещественной переменной x. Из выражения $f(x; a_1, \ldots, a_m)$, принадлежащего функции f (см. начало гл. 3), построим выражение $\tilde{f}(x_1, x_2, \ldots, x_n; a_1, \ldots, a_m)$, заменяя каждое вхождение переменной x на новую переменную x_i, $1 \leqslant i \leqslant n$.*

Допустим, что это новое выражение удовлетворяет условию Липшица

$$|\tilde{f}(x_1, \ldots, y_i, \ldots, x_n; a_1, \ldots, a_m) - \tilde{f}(x_1, \ldots, z_i, \ldots, x_n; a_1, \ldots, a_m)| \leqslant l_i |y_i - z_i|$$

при каждом i, $1 \leqslant i \leqslant n$ и фиксированных

$$x_k \in X, \quad 1 \leqslant k \leqslant i, \quad i \neq k \quad \text{и} \quad a_k \in A, \quad 1 \leqslant k \leqslant m.$$

Если существует интервальная оценка $f(X; A_1, \ldots, A_m)$, то для любых интервалов $Y \subseteq X$ и $Z \subseteq X$ выполнено следующее условие Липшица:

$$(5) \quad q(f(Y; A_1, \ldots, A_m), f(Z; A_1, \ldots, A_m)) \leqslant \left(\sum_{i=1}^{n} l_i\right) q(Y, Z).$$

Доказательство. Из способа построения выражения $\tilde{f}(x_1, \ldots, x_n; a_1, \ldots, a_m)$ следует (ср. с гл. 3), что

$$f(X; A_1, \ldots, A_m) = \tilde{f}(X, \ldots, X; A_1, \ldots, A_m) = W(\tilde{f}, X, \ldots, X; A_1, \ldots, A_m).$$

Если произвольным образом выбрано $u \in f(Y; A_1, \ldots, A_m)$, то мы имеем

$$u = \tilde{f}(y_1, \ldots, y_n; a_1, \ldots, a_m) \quad \text{для} \quad y_i \in Y, \quad 1 \leqslant i \leqslant n,$$
$$a_k \in A_k, \quad 1 \leqslant k \leqslant m.$$

Из леммы 7 следует, что для произвольного y_i существует $z_i \in Z$, для которого

$$|y_i - z_i| \leqslant q(Y, Z).$$

Рассматривая значение функции

$$\tilde{f}(z_1, \ldots, z_n; a_1, \ldots, a_m) \in f(Z; A_1, \ldots, A_m),$$

мы получаем с помощью сделанных предположений и многократного использования неравенства треугольника, что

$$\begin{aligned}
&|\tilde{f}(y_1, \ldots, y_n; a_1, \ldots, a_m) - \tilde{f}(z_1, \ldots, z_n; a_1, \ldots, a_m)| \\
&\leqslant |f(y_1, \ldots, y_n; a_1, \ldots, a_m) \\
&\quad - \tilde{f}(z_1, y_2, \ldots, y_n; a_1, \ldots, a_m)| + \ldots \\
&\quad + |\tilde{f}(z_1, \ldots, z_{n-1}, y_n; a_1, \ldots, a_m) \\
&\quad - \tilde{f}(z_1, \ldots, z_n; a_1, \ldots, a_m)| \\
&\leqslant \sum_{i=1}^{n} l_i |y_i - z_i| \leqslant \left(\sum_{i=1}^{n} l_i\right) q(Y, Z) = \alpha.
\end{aligned}$$

Аналогичное неравенство можно установить для произвольного $v \in f(Z, A_1, \ldots, A_m)$. Применяя лемму (7), получаем неравенство (5). ■

Условия теоремы 8 на практике почти всегда выполнены. Рассматриваемые здесь функциональные выражения составлены из основных арифметических операций и стандартных функций. Поэтому они почти всегда дифференцируемы на своей области определения.

В качестве приложения доказанной теоремы приведем несколько конкретных примеров.

Примеры. (a) $f(x; a) = ax$. Интервальная оценка этой функции для $A \in I(\mathbb{R})$ и произвольных $Y, Z \in I(\mathbb{R})$ удовлетворяет в силу (5) неравенству

$$q(AY, AZ) \leqslant |A| q(Y, Z).$$

(Ср. также с формулой (2.7).)

(b) $$f(x; a_0, \ldots, a_n) = \sum_{k=0}^{n} a_k x^k.$$

В силу (5) интервальная оценка этой функции удовлетворяет для $A_k \in I(\mathbb{R})$, $0 \leqslant k \leqslant n$, и произвольных $Y, Z \in I(\mathbb{R})$ неравенству

$$q\left(\sum_{k=0}^{n} A_k Y^k,\ \sum_{k=0}^{n} A_k Z^k\right) \leqslant \left(\sum_{k=1}^{n} k\,|A_k|\,|X|^{k-1}\right) q(Y, Z).$$

Здесь X — наименьший интервал, содержащий $Y \cup Z$. Полученное неравенство верно независимо от того, как вычисляется X^k — как произведение $X \cdot \ldots \cdot X$ или как одноместная операция согласно определению 1.3.

Читателю предоставляется в качестве простого упражнения получить аналогичную формулу для рациональных выражений.

(c) $$f(x;\ a) = x/2 - 2e^{ax}.$$

В силу (5) интервальная оценка этой функции удовлетворяет для $A \in I(\mathbb{R})$ и произвольных $Y, Z \in I(\mathbb{R})$ неравенству

$$q(Y/2 - 2e^{AY},\ Z/2 - 2e^{AZ}) \leqslant \left(\frac{1}{2} + 2\,|A|\,e^{|A|\,|X|}\right) q(Y, Z).$$

Здесь X снова обозначает наименьший интервал, содержащий $Y \cup Z$.

Теорема 8 без труда переносится на функции нескольких переменных. Таким образом, мы получаем формулы для интервальных вычислений отображений $\mathscr{f}_{\mathrm{p}}$, описанных в начале этой главы. Нужно только применить эти неравенства покомпонентно. Это без труда следует из уже доказанных результатов, и мы опускаем подробности. Такие формулы в свою очередь позволяют нам вычислять практически матрицу $\mathscr{P}_{\mathrm{p}}$, участвующую в определении 3. Мы приведем простой пример.

Пример. Дано n функций

$$f_i(x_1, \ldots, x_n;\ a_{i1}, \ldots, a_{in}) = \sum_{\nu=1}^{n} \sin(a_{i\nu} x_\nu), \quad 1 \leqslant i \leqslant n.$$

Интервальная оценка существует для произвольных интервалов $X_\nu \in I(\mathbb{R})$, $1 \leqslant \nu \leqslant n$ и $A_{ij} \in I(\mathbb{R})$, $1 \leqslant i, j \leqslant n$. Кроме того, каждая из данных функций удовлетворяет условиям теоремы 8 для данных интервалов A_{ij}, $1 \leqslant i, j \leqslant n$, а именно

$$\begin{aligned} &|f_i(x_1, \ldots, x_\nu, \ldots, x_n;\ a_{i1}, \ldots, a_{in}) \\ &\qquad - f_i(x_1, \ldots, y_\nu, \ldots, x_n;\ a_{i1}, \ldots, a_{in})| \\ &\leqslant |A_{i\nu}|\,|x_\nu - y_\nu|, \quad x_j \in X_j, \quad 1 \leqslant j \leqslant n, \quad j \neq \nu, \quad x_\nu, \quad y_\nu \in X_\nu. \end{aligned}$$

Если мы теперь применим теорему 8 к каждой компоненте, то получим

$$\mathcal{f}_{\mathrm{p}}(x)=(f_i(x))=\left(\sum_{\nu=1}^{n}\sin(A_{i\nu}X_{\nu})\right),$$

$$q(\mathcal{f}_{\mathrm{p}}(x),\ \mathcal{f}(y))\leqslant\mathcal{P}_{\mathrm{p}}q(x,\ y),\quad\text{где}\quad\mathcal{P}_{\mathrm{p}}=(|A_{ij}|).$$

Если теперь $\rho(\mathcal{P}_{\mathrm{p}})<1$, то, согласно теореме 4, уравнение

$$x=\mathcal{f}_{\mathrm{p}}(x)=(f_i(x))$$

имеет единственную неподвижную точку, которая может быть найдена методом итераций. Применяя теорему 5, можно показать, что короткошаговый метод также сходится.

В теореме 4 мы предположим для простоты, что отображение $\mathcal{f}_{\mathrm{p}}$ определено на всем пространстве $V_n(\mathbb{C})$, что $\mathcal{f}_{\mathrm{p}}$ можно интервально оценить для всех элементов множества $V_n(I(\mathbb{C}))$ и что $\mathcal{f}_{\mathrm{p}}$ есть $\mathcal{P}_{\mathrm{p}}$-сжатие. Простые примеры [в частности, пример (b)] показывают, однако, что матрица $\mathcal{P}_{\mathrm{p}}$ в определении 3 может зависеть от x и y и что условие $\rho(\mathcal{P}_{\mathrm{p}})<1$ может нарушаться для некоторых x, y В этом случае может помочь следующее утверждение.

Теорема 9. *Пусть* $\mathcal{f}_{\mathrm{p}}$: $\theta\subseteq V_n(\mathbb{C})\to V_n(\mathbb{C})$ *есть* $\mathcal{P}_{\mathrm{p}}$*-сжатие для всех* x, y *из замкнутого множества* $I(\theta)=\{z\in V_n(I(\mathbb{C}))\mid z\subseteq\theta\}$. *Если* $\mathcal{f}_{\mathrm{p}}(x)\in I(\theta)$ *для всех* $x\in I(\theta)$, *то уравнение* $x=\mathcal{f}_{\mathrm{p}}(x)$ *имеет единственную неподвижную точку* $x^*\in I(\theta)$, *причем итерации* $x^{(k+1)}=\mathcal{f}_{\mathrm{p}}(x^{(k)})$, $k\geqslant 0$, *сходятся к этой неподвижной точке* x^* при любом начальном векторе $x^{(0)}\in I(\theta)$. ■

Доказательство можно провести аналогично доказательству теоремы 4.

Заметим, что из неравенства

$$q(x^{(k+m)},\ x^{(k)})\leqslant\sum_{j=1}^{m}\mathcal{P}_{\mathrm{p}}^{j}q(x^{(k)},\ x^{(k-1)})\leqslant(\mathcal{I}_{\mathrm{p}}-\mathcal{P}_{\mathrm{p}})^{-1}\mathcal{P}_{\mathrm{p}}q(x^{(k)},\ x^{(k-1)}),$$

которое следует из теоремы 4, мы получаем оценку погрешности

$$q(x^{(k)},\ x^{*})\leqslant(\mathcal{I}_{\mathrm{p}}-\mathcal{P}_{\mathrm{p}})^{-1}\mathcal{P}_{\mathrm{p}}q(x^{(k)},\ x^{(k-1)})$$

для теорем 4 и 9, переходя к пределу при $m\to\infty$ в предыдущем неравенстве.

Сформулируем еще одну лемму, которая будет использована позднее.

Лемма. 10. *Пусть* $\mathcal{f}_{\mathrm{p}}$: $x\subseteq\theta\subseteq V_n(\mathbb{R})\to V_n(\mathbb{R})$ — *непрерывное отображение. Пусть отображение* $\mathcal{Y}_{\mathrm{p}}$: $x\subseteq\theta\subseteq V_n(\mathbb{R})\to V_n(\mathbb{R})$ *также непрерывно и для всех* $x_{\mathrm{p}}\in x$ *существует обращение*

$\mathcal{Y}_p(x_p)^{-1}$. *Если* $f_p(x_p) \in x$ *при всех* $x_p \in x$ *верно для отображения* f_p: $x \to V_n(\mathbb{R})$, *определенного формулой*

$$f_p(x_p) := x_p - \mathcal{Y}_p(x_p) f_p(x_p),$$

то f_p *имеет нуль в интервале* x. ■

Простое доказательство этого утверждения использует теорему Брауэра о неподвижной точке. Действительно, f_p отображает выпуклое компактное множество $\{x_p \mid x_p \in x\} \subset V_n(\mathbb{R})$ в себя и поэтому имеет неподвижную точку. Так как $\mathcal{Y}_p(x_p)$ несингулярно, отображение f_p имеет нуль в x.

Замечания. Лемма 10 сформулирована и доказана Алефельдом в [349]. Она обобщает утверждение для постоянного $\mathcal{Y}_p$, доказанное Муром в [448]. См. также [451].

Эта лемма важна, так как мы можем проверить выполнение ее условий с помощью конечного числа арифметических операций над интервальным вектором, найдя мажоранту для отображения $\mathcal{Y}_p(x_p)$ в виде интервального выражения, зависящего от интервального вектора x.

Подробности такой проверки описаны в следующих главах, в частности в гл. 19.

Прямая проверка условий теоремы Брауэра о неподвижной точке представляет собой, напротив, очень трудную задачу.

* * *

Литература: [35], [122], [215], [220], [232], [243], [295], [325] [349], [448], [451].

Глава 12
СИСТЕМЫ ЛИНЕЙНЫХ УРАВНЕНИЙ, ПОДДАЮЩИЕСЯ МЕТОДУ ИТЕРАЦИИ

Мы предполагаем, что рассматриваемая здесь система линейных уравнений уже имеет вид

$$x_p = \mathcal{A}_p x_p + b_p, \tag{1}$$

где $\mathcal{A}_p = (a_{ij})$ и $b_p = (b_i)$.

Пусть известно, что элементы a_{ij} матрицы $\mathcal{A}_p$ лежат в интервалах A_{ij}, а компоненты b_i вектора b_p лежат в интервалах B_i. Нас интересует множество решений, получающихся, когда входные данные изменяются в данных интервалах. Поэтому мы вводим интервальную матрицу $\mathcal{A} = (A_{ij})$ размерности $n \times n$, со-

держащую интервальные коэффициенты системы, и интервальный вектор $\mathcal{b} = (B_i)$, содержащий интервальные правые части. Рассмотрим теперь отображение

$$\mathcal{f}_{\mathrm{p}}\colon\ V_n(\mathbb{C}) \to V_n(\mathbb{C}),$$

определяемое формулой

$$\mathcal{f}_{\mathrm{p}}(x_{\mathrm{p}}) = \mathcal{A}_{\mathrm{p}} x_{\mathrm{p}} + \mathcal{b}_{\mathrm{p}}.$$

Прежде всего мы имеем следующее утверждение.

Теорема 1. *Метод последовательных приближений*

$$(x)^{(k+1)} = \mathcal{A}x^{(k)} + \mathcal{b}, \qquad k \geqslant 0, \tag{2}$$

сходится к единственной неподвижной точке x^* уравнения

$$x = \mathcal{A}x + \mathcal{b}$$

для любого $x^{(0)} \in V_n(I(\mathbb{C}))$ тогда и только тогда, когда $\rho(|\mathcal{A}|) < 1$.

Доказательство. Покажем, что $\mathcal{f}_{\mathrm{p}}$ есть $|\mathcal{A}|$-сжатие. Пусть $x,\ y \in V_n(I(\mathbb{C}))$. Применяя (10.23) и (10.25), получаем

$$q(\mathcal{f}_{\mathrm{p}}(x), \mathcal{f}_{\mathrm{p}}(y)) = q(\mathcal{A}x + \mathcal{b}, \mathcal{A}y + \mathcal{b}) \leqslant |\mathcal{A}|\, q(x, y).$$

Из теоремы 11.4 следует, что условие $\rho(|\mathcal{A}|) < 1$ достаточно для сходимости метода и единственности неподвижной точки. Для доказательства обратного утверждения допустим, что последовательные приближения $x^{(k+1)} = \mathcal{A}x^{(k)} + \mathcal{b}$, $k \geqslant 0$, сходятся для каждого $x^{(0)} \in V_n(I(\mathbb{C}))$ к неподвижной точке x^*. Мы должны показать, что $\rho(|\mathcal{A}|) < 1$. Из теоремы Перрона и Фробениуса (см. [325]) следует, что вещественная неотрицательная матрица $|\mathcal{A}|$ имеет неотрицательный собственный вектор, соответствующий собственному числу $\lambda = \rho(|\mathcal{A}|)$. Из сходимости приближений $x^{(k+1)} = \mathcal{A}x^{(k)} + \mathcal{b}$, $k \geqslant 0$, к x^* при любом начальном векторе $x^{(0)}$ следует, что последовательность $\{d(x^{(k)})\}_{k=0}^{\infty}$ сходится к $d(x^*)$. Выберем теперь $x^{(0)}$ таким образом, чтобы $d(x^{(0)})$ был собственным вектором, соответствующим собственному числу $\lambda = \rho(|\mathcal{A}|)$ матрицы $|\mathcal{A}|$, причем хотя бы одна компонента вектора $d(x^{(0)})$ была строго больше, чем соответствующая компонента вектора $d(x^*)$. Тогда из (2) следует в силу (10.12) и (10.18), что в предположении $\lambda = \rho(|\mathcal{A}|) \geqslant 1$ верно

$$\begin{aligned} d(x^{(1)}) &= d(\mathcal{A}x^{(0)} + \mathcal{b}) = d(\mathcal{A}x^{(0)}) + d(\mathcal{b}) \\ &\geqslant d(\mathcal{A}x^{(0)}) \geqslant |\mathcal{A}|\, d(x^{(0)}) = \lambda d(x^{(0)}) \geqslant d(x^{(0)}), \\ d(x^{(2)}) &\geqslant |\mathcal{A}|\, d(x^{(1)}) \geqslant \lambda |\mathcal{A}|\, d(x^{(0)}) = \lambda^2 d(x^{(0)}) \geqslant d(x^{(0)}). \end{aligned}$$

Для произвольного k получаем

$$d(x^{(k+1)}) \geqslant |\mathcal{A}|\, d(x^{(k)}) \geqslant \ldots \geqslant \lambda^{(k+1)} d(x^{(0)}) \geqslant d(x^{(0)}).$$

Переходя к пределу при $k \to \infty$, получаем

$$d(x^*) \geqslant d(x^{(0)}),$$

что противоречит выбору $x^{(0)}$. Поэтому верно $\rho(|\mathscr{A}|) < 1$. ■

Мы хотим теперь установить связь этой теоремы с задачей, сформулированной во введении к этой главе. (См. также следствие 11.6.)

Теорема 2. *Пусть $\mathscr{A}$ — интервальная матрица, такая что $\rho(|\mathscr{A}|) < 1$. Тогда для неподвижной точки x^* уравнения $x^* = \mathscr{A}x^* + b$ (которая существует и единственна в силу теоремы 1) верно соотношение*

$$\{y_p = (\mathscr{I}_p - \mathscr{A}_p)^{-1} b_p \mid \mathscr{A}_p \in \mathscr{A},\ b_p \in b\} \subseteq \{x_p \mid x_p \in x^*\}.$$

Если $\mathscr{A} = (A_{ij}) \in M_{nn}(I(\mathbb{R}))$, $b \in V_n(I(\mathbb{R}))$ и неравенство $i(A_{ij}) \geqslant 0$ справедливо для $A_{ij} = [i(A_{ij}), s(A_{ij})]$, то x^* оптимальна в следующем смысле. Не существует интервального вектора $x \in V_n(I(\mathbb{R}))$ такого, что $x \subseteq x^*$, $x \neq x^*$, но

$$\{y_p = (\mathscr{I}_p - \mathscr{A}_p)^{-1} b_p \mid \mathscr{A}_p \in \mathscr{A},\ b_p \in b\} \subseteq \{x_p \mid x_p \in x\}.$$

Доказательство. Мы покажем сначала, что система линейных уравнений

$$y_p = \mathscr{A}_p y_p + b_p$$

имеет решение

$$y_p = (\mathscr{I}_p - \mathscr{A}_p)^{-1} b_p$$

для $\mathscr{A}_p \in \mathscr{A}$ и $b_p \in b$. Так как верно $\mathscr{A}_p \in \mathscr{A}$, мы имеем $|\mathscr{A}_p| \leqslant |\mathscr{A}|$ и из теоремы Перрона и Фробениуса (см. [21]) получаем

$$\rho(\mathscr{A}_p) \leqslant \rho(|\mathscr{A}_p|) \leqslant \rho(|\mathscr{A}|) < 1.$$

Отсюда следует, что матрица $\mathscr{I}_p - \mathscr{A}_p$ несингулярна, что и требовалось.

Рассмотрим теперь последовательные приближения

$$x^{(k+1)} = \mathscr{A}x^{(k)} + b, \quad k \geqslant 0,$$

где $x^{(0)} = y_p = \mathscr{A}_p y_p + b_p$. Из монотонности включения следует, что

$$y_p = \mathscr{A}_p y_p + b_p \in \mathscr{A}x^{(0)} + b = x^{(1)},$$

и для произвольного k

$$y_p = \mathscr{A}_p y_p + b_p \in \mathscr{A}x^{(k)} + b = x^{(k+1)}.$$

Из $\rho(|\mathscr{A}|) < 1$ следует, что $\lim_{k\to\infty} x^{(k)} = x^*$, а потому и $y_p \in x^*$. Так как x^* не зависит от начального вектора, мы получаем первую часть теоремы.

Для доказательства второй части теоремы построим вектор $u_{\mathrm{p}} \in V_n(\mathbb{R})$ из n нижних границ компонент вектора x^*. Из верхних границ аналогичным образом строится вектор $v_{\mathrm{p}} \in V_n(\mathbb{R})$. Тогда из $x^* = \mathscr{A}x^* + b$ следует по правилам интервальной арифметики, что

$$u_{\mathrm{p}} = \mathscr{A}^*_{\mathrm{p}} u_{\mathrm{p}} + r_{\mathrm{p}} \quad \text{и} \quad v_{\mathrm{p}} = \mathscr{A}^{**}_{\mathrm{p}} v_{\mathrm{p}} + s_{\mathrm{p}},$$

где $u_{\mathrm{p}i} = (u_i)$, $v_{\mathrm{p}i} = (v_i)$ и

$$\mathscr{A}^*_{\mathrm{p}} = (a^*_{ij}), \quad a^*_{ij} = \begin{cases} i(A_{ij}), & u_j > 0, \\ s(A_{ij}), & u_j \leqslant 0, \end{cases}$$

$$\mathscr{A}^{**}_{\mathrm{p}} = (a^{**}_{ij}), \quad a^{**}_{ij} = \begin{cases} s(A_{ij}), & v_j < 0 \\ i(A_{ij}), & v_j \leqslant 0, \end{cases}$$

$$r_{\mathrm{p}} = (i(B_i)), \quad s_{\mathrm{p}} = (s(B_i)).$$

Из этих равенств следует, что u_{p} и v_{p} являются членами множества

$$\{y_{\mathrm{p}} = (\mathscr{I}_{\mathrm{p}} - \mathscr{A}_{\mathrm{p}})^{-1} b_{\mathrm{p}} \mid \mathscr{A}_{\mathrm{p}} \in \mathscr{A},\ b_{\mathrm{p}} \in b\},$$

что и завершает доказательство. ■

Метод итерации, рассмотренный в теореме 1, можно назвать полношаговым (T) по аналогии с соответствующим методом для «точечной системы уравнений». Аналогичный короткошаговый метод (S) получается разложением интервальной матрицы $\mathscr{A}$ в сумму

$$\mathscr{A} = \mathscr{L} + \mathscr{D} + \mathscr{U},$$

где $\mathscr{L}$ — строго нижняя треугольная матрица, $\mathscr{D}$ — диагональная матрица и $\mathscr{U}$ — строго верхняя треугольная матрица. Тогда короткошаговый итерационный метод определяется формулами

$$x^{(k+1)} = \mathscr{L}x^{(k+1)} + (\mathscr{D} + \mathscr{U})x^{(k)} + b, \quad k \geqslant 0. \tag{3}$$

Следующее утверждение касается сходимости этого короткошагового метода.

Теорема 3. *Итерационный метод*

$$x^{(k+1)} = \mathscr{L}x^{(k+1)} + (\mathscr{D} + \mathscr{U})x^{(k)} + b, \quad k \geqslant 0,$$

с произвольным начальным вектором $x^{(0)} \in V_n(I(\mathbb{C}))$ *сходится к единственной неподвижной точке* x^* *тогда и только тогда, когда*

$$\rho((\mathscr{I}_{\mathrm{p}} - |\mathscr{L}|)^{-1}(|\mathscr{D}| + |\mathscr{U}|)) < 1.$$

Доказательство. Мы собираемся применить теорему 11.5 и с этой целью полагаем

$$f\colon V_n(\mathbb{C}) \to V_n(\mathbb{C}),$$

где

$$f_{\mathrm{p}}(x_{\mathrm{p}}) = \mathscr{L}_{\mathrm{p}} x_{\mathrm{p}} + (\mathscr{D}_{\mathrm{p}} + \mathscr{U}_{\mathrm{p}}) x_{\mathrm{p}} + b_{\mathrm{p}}$$

и

$$g_{\mathrm{p}}\colon\ V_n(\mathbb{C}) \times V_n(\mathbb{C}) \to V_n(\mathbb{C}),$$

где

$$g_{\mathrm{p}}(x_{\mathrm{p}},\ y_{\mathrm{p}}) = \mathscr{L}_{\mathrm{p}} x_{\mathrm{p}} + (\mathscr{D}_{\mathrm{p}} + \mathscr{U}_{\mathrm{p}}) y_{\mathrm{p}} + b_{\mathrm{p}}.$$

Мы имеем тогда

$$g_{\mathrm{p}}(x,\ x) = f_{\mathrm{p}}(x_{\mathrm{p}}) \quad \textit{для всех} \quad x \in V_n(I(\mathbb{C})),$$

и из (10.23) и (10.25) следует, что

$$q(g_{\mathrm{p}}(x,\ z),\ g_{\mathrm{p}}(y,\ z)) = q(\mathscr{L} x + (\mathscr{D} + \mathscr{U}) z + b,\ \mathscr{L} y + (\mathscr{D} + \mathscr{U}) z + b) \leqslant |\mathscr{L}|\, q(x,\ y),$$

$$q(g_{\mathrm{p}}(z,\ x),\ g_{\mathrm{p}}(z,\ y)) = q(\mathscr{L} z + (\mathscr{D} + \mathscr{U}) x + b,\ \mathscr{L} z + (\mathscr{D} + \mathscr{U}) y + b) \leqslant (|\mathscr{D}| + |\mathscr{U}|)\, q(x,\ y)$$

для всех x, y_{p}, $z \in V_n(I(\mathbb{C}))$. Мы имеем $\rho(|\mathscr{L}|) = 0$, так как $|\mathscr{L}|$ — строго нижняя треугольная матрица. Поэтому, полагая $Q_{\mathrm{p}} := |\mathscr{L}|$, $\mathscr{R}_{\mathrm{p}} := |\mathscr{D}| + |\mathscr{U}|$, мы оказываемся в условиях теоремы 11.5, так что условие

$$\rho((\mathscr{I}_{\mathrm{p}} - |\mathscr{L}|)^{-1}(|\mathscr{D}| + |\mathscr{U}|)) < 1$$

достаточное. Доказательство необходимости этого условия для сходимости при любом начальном векторе проводится так же, как соответствующее доказательство для полношагового метода в теореме 1. ■

Теорема Штейна и Розенберга, а также ее обобщение (см. [325] и [35]) утверждают, что для $\mathscr{A} = \mathscr{L} + \mathscr{D} + \mathscr{U}$ верна эквивалентность $\rho(|\mathscr{A}|) < 1$ тогда и только тогда, когда

$$\rho((\mathscr{I}_{\mathrm{p}} - |\mathscr{L}|)^{-1}(|\mathscr{D}| + |\mathscr{U}|)) < 1.$$

Так как условия сходимости полношагового и короткошагового методов необходимы и достаточны, мы получаем следующее утверждение.

Теорема 4. *Полношаговый метод* (2) *сходится для любого начального значения* $x^{(0)} \in V_n(I(\mathbb{C}))$ *к единственной неподвижной точке тогда и только тогда, когда короткошаговый метод* (3) *сходится к единственной неподвижной точке для любого начального значения* $x^{(0)} \in V_n(I(\mathbb{C}))$. ■

Этот результат существенно отличается от соответствующего результата для точечных систем уравнений (см., например, [325]), где сходимость или расходимость полношагового метода

не обязательно означает сходимость или расходимость коротко-шагового метода.

Поскольку умножение интервальных матриц на интервальные векторы в общем случае не дистрибутивно, то даже для случая $\rho(|\mathscr{A}|)<1$ не очевидно, что неподвижная точка полношаговой итерации, удовлетворяющая равенству

$$x^* = \mathscr{A}x^* + b,$$

совпадает с неподвижной точкой короткошаговой итерации, удовлетворяющей равенству

$$\tilde{x}^* = \mathscr{L}\tilde{x}^* + (\mathscr{D}+\mathscr{U})\tilde{x}^* + b.$$

Однако, используя специальный вид матриц $\mathscr{L}$, $\mathscr{D}$ и $\mathscr{U}$, мы получаем с помощью определения действий над интервальными матрицами и векторами, что

$$\tilde{x}^* = \mathscr{L}\tilde{x}^* + (\mathscr{D}+\mathscr{U})\tilde{x}^* + b = (\mathscr{L}+\mathscr{D}+\mathscr{U})\tilde{x}^* + b = \mathscr{A}\tilde{x}^* + b.$$

Отсюда получается, что $x^* = \tilde{x}^*$, и мы приходим к следующему утверждению.

Следствие 5. *Если* $\rho(|\mathscr{A}|)<1$, *то полношаговый и короткошаговый методы сходятся к неподвижной точке* x^* *уравнения*

$$x = \mathscr{A}x + b. \ \blacksquare$$

Рассмотрим теперь симметрический короткошаговый метод (SS), в котором матрица $\mathscr{A}$ раскладывается в сумму

$$\mathscr{A} = \mathscr{L} + \mathscr{U},$$

где $\mathscr{L}$ — строго нижняя, а $\mathscr{U}$ — строго верхняя треугольные матрицы. Метод (SS) определяется соотношениями

$$\text{(SS)} \quad \begin{cases} x^{(k+1/2)} = \mathscr{L}x^{(k+1/2)} + \mathscr{U}x^{(k)} + b, \\ x^{(k+1)} = \mathscr{L}x^{(k+1/2)} + \mathscr{U}x^{(k+1)} + b, \quad k \geqslant 0. \end{cases}$$

Если не все диагональные элементы матрицы $\mathscr{A}$ обращаются в нуль, то вместо этого мы должны рассмотреть итерационный метод

$$\text{(SS}') \quad \begin{cases} x^{(k+1/2)} = \mathscr{L}x^{(k+1/2)} + \mathscr{D}x^{(k)} + \mathscr{U}x^{(k)} + b, \\ x^{(k+1)} = \mathscr{L}x^{(k+1/2)} + \mathscr{D}x^{(k+1/2)} + \mathscr{U}x^{(k+1)} + b, \quad k \geqslant 0. \end{cases}$$

Для этого итерационного метода можно доказать утверждения, аналогичные тем, которые будут доказаны ниже для метода (SS).

Для точечных матриц и точечных векторов метод (SS) был впервые рассмотрен в [345]. В [455] было указано, что в случае точечных систем уравнений метод (SS) можно организовать с

теми же затратами труда (если не считать первого шага итерации), что и для короткошагового метода. Мы рассмотрим этот вопрос в гл. 14, где будет описана модификация метода (SS), для которой верен тот же результат. Сходимость метода (SS) будет получена из следующего более общего результата.

Теорема 6. *Пусть интервальная матрица* $\mathcal{A} \in M_{nn}(I(\mathbb{C}))$ *разложена в сумму* $\mathcal{A} = \mathcal{M} + \mathcal{N}$ *двух интервальных матриц* $\mathcal{M}$ *и* $\mathcal{N}$, *для которых верно* $\rho(|\mathcal{M}|) < 1$ *и* $\rho(|\mathcal{N}|) < 1$. *Тогда для произвольного вектора* $b \in V_n(I(\mathbb{C}))$ *верно следующее:*

(a) *Для любого интервального вектора* $x^{(0)} \in V_n(I(\mathbb{C}))$ *существует последовательность* $\{x^{(k)}\}_{k=0}^{\infty}$, *которая удовлетворяет итерационным формулам*

$$\text{(V)} \quad \begin{cases} x^{(k+1/2)} = \mathcal{M}x^{(k+1/2)} + \mathcal{N}x^{(k)} + b, \\ x^{(k+1)} = \mathcal{M}x^{(k+1/2)} + \mathcal{N}x^{(k+1)} + b, \quad k = 0, 1, 2, \ldots. \end{cases}$$

(b) *Если* $\rho\left((\mathcal{I}_p - |\mathcal{N}|)^{-1}|\mathcal{M}|(\mathcal{I}_p - |\mathcal{M}|)^{-1}|\mathcal{N}|\right) < 1$, *то уравнение* $x = \mathcal{A}x + b$ *имеет единственную неподвижную точку* x^*. *Если сверх того*

$$\mathcal{A}x^* = (\mathcal{M} + \mathcal{N})x^* = \mathcal{M}x^* + \mathcal{N}x^*,$$

то последовательность, вычисленная по формулам (V), *сходится к* x^* *для любого начального вектора* $x^{(0)}$. (*Как мы уже видели, для интервальных матриц не выполнен дистрибутивный закон. См. гл.* 10 *формулы* (6).)

(c) *Обратно, если уравнение* $x = \mathcal{A}x + b$ *имеет единственную неподвижную точку* x^* *и последовательность* (V) *сходится к* x^* *для любого начального приближения* $x^{(0)}$, *то*

$$\mathcal{A}x^* = (\mathcal{M} + \mathcal{N})x^* = \mathcal{M}x^* + \mathcal{N}x^*$$

и

$$\rho\left((\mathcal{I}_p - |\mathcal{N}|)^{-1}|\mathcal{M}|(\mathcal{I}_p - |\mathcal{M}|)^{-1}|\mathcal{N}|\right) < 1.$$

Доказательство. (a) Для произвольного интервального вектора z из (10.23) и (10.25) следует, что для любых векторов x, y имеет место

$$q(\mathcal{M}x + \mathcal{N}z + b, \mathcal{M}y + \mathcal{N}z + b) = q(\mathcal{M}x, \mathcal{M}y) \leqslant |\mathcal{M}|q(x, y).$$

Ввиду $\rho(|\mathcal{M}|) < 1$ мы получаем по теореме 1, что для любого k уравнение

$$x^{(k+1/2)} = \mathcal{M}x^{(k+1/2)} + \mathcal{N}x^{(k)} + b$$

имеет единственную неподвижную точку $x^{(k+1/2)}$. Аналогично можно показать, что для любого k уравнение

$$x^{(k+1)} = \mathcal{M}x^{(k+1/2)} + \mathcal{N}x^{(k+1)} + b$$

имеет единственную неподвижную точку $x^{(k+1)}$. Тем самым доказаны существование и единственность последовательности $\{x^{(k)}\}_{k=0}^{\infty}$ при данном начальном векторе $x^{(0)}$.

(b) Сначала покажем, что $\rho(|\mathscr{A}|)<1$. Так как $\rho(|\mathscr{M}|)<1$, $\rho(|\mathscr{N}|)<1$, то обратные матрицы

$$(\mathscr{I}_p-|\mathscr{M}|)^{-1} \text{ и } (\mathscr{I}_p-|\mathscr{N}|)^{-1}$$

существуют и неотрицательны по теореме 3.8 из [325]. Поэтому вещественная матрица

$$(\mathscr{I}_p-|\mathscr{N}|)^{-1}|\mathscr{M}|(\mathscr{I}_p-|\mathscr{M}|)^{-1}|\mathscr{N}|=(\mathscr{I}-|\mathscr{N}|)^{-1}\times(\mathscr{I}_p-|\mathscr{M}|)^{-1}|\mathscr{M}||\mathscr{N}|$$

также неотрицательна. Применяя теорему 3.8 из [325], получаем

$$\mathscr{O}_p\leqslant(\mathscr{I}_p-(\mathscr{I}_p-|\mathscr{N}|)^{-1}(\mathscr{I}_p-|\mathscr{M}|)^{-1}(|\mathscr{M}||\mathscr{N}|)^{-1}$$
$$=(\mathscr{I}_p-(|\mathscr{M}|+|\mathscr{N}|))^{-1}(\mathscr{I}_p-|\mathscr{M}|)(\mathscr{I}_p-|\mathscr{N}|),$$

а используя

$$(\mathscr{I}_p-|\mathscr{M}|)^{-1}\geqslant\mathscr{O}_p,\quad(\mathscr{I}_p-|\mathscr{N}|)^{-1}\geqslant\mathscr{O}_p,$$

имеем, наконец,

$$(\mathscr{I}_p-(|\mathscr{M}|+|\mathscr{N}|))^{-1}\geqslant\mathscr{O}_p.$$

Отсюда по теореме 3.8 из [325] следует неравенство

$$\rho(|\mathscr{M}|+|\mathscr{N}|)<1.$$

Из соотношения

$$|\mathscr{A}|=|\mathscr{M}+\mathscr{N}|\leqslant|\mathscr{M}|+|\mathscr{N}|$$

(см. [199 гл. 10]) ввиду теоремы Перрона и Фробениуса (теорема 2.7 из [325]) следует, что

$$\rho(|\mathscr{A}|)\leqslant\rho(|\mathscr{M}|+|\mathscr{N}|)<1.$$

Поэтому ввиду теоремы 1 уравнение $x=\mathscr{A}x+b$ имеет единственную неподвижную точку x^*. Из равенства

$$x^*=\mathscr{A}x^*+b=\mathscr{M}x^*+\mathscr{N}x^*+b$$

с помощью (10.23)—(10.25) и (V) следует, что

$$q(x^{(k+1)},\ x^*)=q(\mathscr{M}x^{(k+1/2)}+\mathscr{N}x^{(k+1)}+b,\ \mathscr{M}x^*+\mathscr{N}x^*+b)$$
$$\leqslant q(\mathscr{M}x^{(k+1/2)}+\mathscr{N}x^{(k+1)},\ \mathscr{M}x^{(k+1/2)}+\mathscr{N}x^*)$$
$$+q(\mathscr{M}x^{(k+1/2)}+\mathscr{N}x^*,\ \mathscr{M}x^*+\mathscr{N}x^*)$$
$$\leqslant|\mathscr{N}|q(x^{(k+1)},\ x^*)+|\mathscr{M}|q(x^{(k+1/2)},\ x^*),$$

а так как $(\mathscr{I}_p - |\mathscr{N}|)^{-1} \geqslant Q_p$, это дает

$$q(x^{(k+1)}, x^*) \leqslant (\mathscr{I}_p - |\mathscr{N}|)^{-1} |\mathscr{M}| q(x^{(k+1/2)}, x^*).$$

Аналогичным образом получаем

$$q(x^{(k+1/2)}, x^*) \leqslant (\mathscr{I}_p - |\mathscr{M}|)^{-1} |\mathscr{N}| q(x^{(k)}, x^*);$$

откуда, наконец,

$$q(x^{(k+1)}, x^*) \leqslant (\mathscr{I}_p - |\mathscr{N}|)^{-1} |\mathscr{M}| (\mathscr{I}_p - |\mathscr{M}|)^{-1} |\mathscr{N}| q(x^k, x^*)$$
$$\leqslant \{(\mathscr{I}_p - |\mathscr{N}|)^{-1} |\mathscr{M}| (\mathscr{I}_p - |\mathscr{M}|)^{-1} |\mathscr{N}|\}^{k+1} q(x^{(0)}, x^*).$$

Из того, что спектральный радиус выражения в фигурных скобках меньше 1, получаем, что $\lim_{k\to\infty} x^{(k)} = x^*$.

(c): Пусть уравнение $x = \mathscr{A}x + b$ имеет единственную неподвижную точку x^*. Из неравенства

$$q(x^{(k+1/2)}, x^*) \leqslant (\mathscr{I}_p - |\mathscr{M}|)^{-1} |\mathscr{N}| q(x^{(k)}, x^*),$$

которое выводится так же, как в доказательстве п. (b), следует, что последовательность $\{x^{(k+1/2)}\}_{k=0}^{\infty}$ сходится к x^* для любого $x^{(0)}$. Отсюда и из верхнего равенства (V) следует при $k \to \infty$, что

$$x^* = \mathscr{M}x^* + \mathscr{N}x^* + b,$$

т. е.

$$o_p = q(x^*, x^*) = q(\mathscr{M}x^* + \mathscr{N}x^* + b, \mathscr{A}x^* + b) = q(\mathscr{M}x^* + \mathscr{N}x^*, \mathscr{A}x^*)$$

или

$$\mathscr{A}x^* = (\mathscr{M} + \mathscr{N})x^* = \mathscr{M}x^* + \mathscr{N}x^*.$$

Мы должны еще доказать, что

$$\rho((\mathscr{I}_p - |\mathscr{N}|)^{-1} |\mathscr{M}| (\mathscr{I}_p - |\mathscr{M}|)^{-1} |\mathscr{N}|) < 1.$$

Чтобы сделать это, поступаем так же, как в доказательстве теоремы 1.

Из теоремы Перрона и Фробениуса [325, теорема 2.7] мы знаем, что матрица

$$(\mathscr{I}_p - |\mathscr{N}|)^{-1} |\mathscr{M}| (\mathscr{I}_p - |\mathscr{M}|)^{-1} |\mathscr{N}|$$

имеет неотрицательный собственный вектор, соответствующий неотрицательному собственному числу

$$\tilde{\lambda} = \rho((\mathscr{I}_p - |\mathscr{N}|)^{-1} |\mathscr{M}| (\mathscr{I}_p - |\mathscr{M}|)^{-1} |\mathscr{N}|).$$

Теперь мы выбираем $x^{(0)}$ так, чтобы $d(x^{(0)})$ был собственным вектором, соответствующим собственному числу $\tilde{\lambda}$, и при этом хотя бы одна компонента вектора $d(x^{(0)})$ была больше, чем со-

ответствующая компонента вектора $d(x^*)$. Тогда из (V) с помощью (10.12) и (10.18) следует, что

$$d(x^{(k+1/2)}) \geqq |\mathscr{M}|\, d(x^{(k+1/2)}) + |\mathscr{N}|\, d(x^{(k)})$$

или

$$d(x^{(k+1/2)}) \geqq (\mathscr{I}_p - |\mathscr{M}|)^{-1} |\mathscr{N}|\, d(x^{(k)}),$$

а также

$$d(x^{(k+1/2)}) \geqq (\mathscr{I}_p - |\mathscr{M}|)^{-1} |\mathscr{N}|\, d(x^{(k)})$$

или

$$d(x^{(k+1)}) \geqq (\mathscr{I}_p - |\mathscr{N}|)^{-1} |\mathscr{M}|\, d(x^{(k+1/2)}).$$

Наконец, мы получаем

$$\begin{aligned} d(x^{(k+1)}) &\geqq (\mathscr{I}_p - |\mathscr{N}|)^{-1} |\mathscr{M}| (\mathscr{I}_p - |\mathscr{M}|)^{-1} |\mathscr{N}|\, d(x^{(k)}) \\ &\geqq \{(\mathscr{I}_p - |\mathscr{N}|)^{-1} |\mathscr{M}| (\mathscr{I}_p - |\mathscr{M}|)^{-1} |\mathscr{N}|\}^{(k+1)}\, d(x^{(0)}) \\ &= \tilde{\lambda}^{(k+1)}\, d(x^{(0)}). \end{aligned}$$

Из сходимости итераций (V) к x^* следует сходимость последовательности $\{d(x^{(k)})\}_{k=0}^{\infty}$ к $d(x^*)$. Предположение $\tilde{\lambda} \geqq 1$ приводит к неравенству

$$d(x^{(k+1)}) \geqq \tilde{\lambda}^{(k+1)}\, d(x^{(0)}) \geqq d(x^{(0)}), \quad k \geqq 0,$$

что в пределе при $k \to \infty$ дает $d(x^*) \geqq d(x^{(0)})$, а это противоречит выбору вектора $x^{(0)}$. Поэтому $\tilde{\lambda} < 1$, и теорема доказана. ■

Теперь возьмем в теореме 6

$$\mathscr{M} := \mathscr{L}, \quad \mathscr{N} := \mathscr{U}.$$

Тогда имеем $\rho(|\mathscr{M}|) = \rho(|\mathscr{N}|) = 0$. Ввиду специального выбора матриц $\mathscr{L}$ и $\mathscr{U}$ равенство

$$\mathscr{A}x = \mathscr{L}x + \mathscr{U}x$$

справедливо для всех интервальных векторов. Это дает приводимое ниже следствие теоремы 6.

Следствие 7. *Симметрический короткошаговый метод* (SS) *сходится к единственной неподвижной точке* x^* *уравнения* $x = \mathscr{A}x + b$ *для любого начального вектора* $x^{(0)}$ *тогда и только тогда, когда спектральный радиус матрицы*

$$(\mathscr{I}_p - |\mathscr{U}|)^{-1} |\mathscr{L}| (\mathscr{I}_p - |\mathscr{L}|)^{-1} |\mathscr{U}|$$

меньше единицы. ■

Мы уже установили в доказательстве теоремы 6(b), что неравенство

$$\rho((\mathscr{I}_p - |\mathscr{N}|)^{-1} |\mathscr{M}| (\mathscr{I}_p - |\mathscr{M}|)^{-1} |\mathscr{N}|) < 1$$

влечет за собой $\rho(|\mathscr{A}|)<1$. Теперь мы хотим показать, что в предположении

$$|\mathscr{A}|=|\mathscr{M}|+|\mathscr{N}|$$

верна и обратная импликация, если выполнены условия $\rho(|\mathscr{M}|)<1$ и $\rho(|\mathscr{N}|)<1$ из теоремы 6. Итак, предположим, что

$$\rho(|\mathscr{A}|)=\rho(|\mathscr{M}|+|\mathscr{N}|)<1.$$

В силу теоремы 3.8 из [325] матрица, обратная к $\mathscr{I}_p-|\mathscr{A}|$, существует, и мы имеем $(\mathscr{I}_p-|\mathscr{A}|)^{-1}\geqslant\mathscr{O}_p$. Рассмотрим следующее разложение:

$$\mathscr{I}_p-|\mathscr{A}|=(\mathscr{I}_p-|\mathscr{M}|)(\mathscr{I}_p-|\mathscr{N}|)-|\mathscr{M}||\mathscr{N}|.$$

Из $\rho(|\mathscr{M}|)<1$, $\rho(|\mathscr{N}|)<1$ следует, что обращения матриц $\mathscr{I}_p-|\mathscr{M}|$ и $\mathscr{I}_p-|\mathscr{N}|$ существуют и неотрицательны. Поэтому неотрицательно и произведение $(\mathscr{I}_p-|\mathscr{N}|)^{-1}\times(\mathscr{I}_p-|\mathscr{M}|)^{-1}$. Тем самым рассматриваемое разложение матрицы $\mathscr{I}_p-|\mathscr{A}|$ регулярно, так как матрица $|\mathscr{M}||\mathscr{N}|$ неотрицательна [325, определение 3.5]. Отсюда с помощью теоремы 3.13 из [325] мы получаем соотношение

$$\rho((\mathscr{I}_p-|\mathscr{N}|)^{-1}(\mathscr{I}_p-|\mathscr{M}|)^{-1}|\mathscr{M}||\mathscr{N}|)=\rho((\mathscr{I}_p-|\mathscr{N}|)^{-1}\times|\mathscr{M}|(\mathscr{I}_p-|\mathscr{M}|)^{-1}|\mathscr{N}|)<1.$$

Собирая все вместе и применяя теорему 6(b), получаем следующее утверждение.

Теорема 8. *Пусть интервальная матрица $\mathscr{A}$ разложена в сумму $\mathscr{A}=\mathscr{M}+\mathscr{N}$ двух интервальных матриц, для которых выполнено*

$$|\mathscr{A}|=|\mathscr{M}|+|\mathscr{N}| \text{ и } \rho(|\mathscr{M}|)<1,\ \rho(|\mathscr{N}|)<1.$$

Тогда неравенство

$$\rho((\mathscr{I}_p-|\mathscr{N}|)^{-1}|\mathscr{M}|(\mathscr{I}_p-|\mathscr{M}|)^{-1}|\mathscr{N}|)<1$$

эквивалентно неравенству

$$\rho(|\mathscr{A}|)<1. \ \blacksquare$$

Равенство $|\mathscr{A}|=|\mathscr{M}|+|\mathscr{N}|$ будет выполнено, например, в случае, когда $\mathscr{A}=(A_{ij})$ разложена в сумму $\mathscr{A}=\mathscr{M}+\mathscr{N}$ с $\mathscr{M}=(M_{ij})$, $\mathscr{N}=(N_{ij})$ таким образом, что для любых $1\leqslant i, j\leqslant n$ по крайней мере одна из компонент M_{ij} и N_{ij} равна нулю.

Это последнее условие выполнено, например, для разложения, с которого начинается описание симметрического коротко-шагового метода (SS). Теорема 8 немедленно дает приводимое ниже утверждение.

Следствие 9. *Симметрический короткошаговый метод сходится к единственной неподвижной точке* x^* *уравнения* $x = \mathcal{A}x + b$ *для любого начального вектора* $x^{(0)} \in V_n(I(\mathbb{C}))$ *тогда и только тогда, когда* $\rho(|\mathcal{A}|) < 1$, *т. е. когда полношаговый метод (а потому и короткошаговый метод) сходится к* x^* *для любого* $x^{(0)} \in V_n(I(\mathbb{C}))$.

Теперь мы рассмотрим скорость, с которой сходится к x^* последовательность $\{x^{(k)}\}_{k=0}^{\infty}$ интервальных векторов, порожденных итерационной процедурой

$$x^{(k+1)} = \mathcal{f}_p(x^{(k)}), \quad k \geqslant 0. \tag{4}$$

Определение 10. *Пусть* $x^* = \mathcal{f}_p(x^*)$ *и пусть* $\mathfrak{S}$ *— множество всех последовательностей* $\{x^{(k)}\}_{k=0}^{\infty}$, *вычисленных по формуле* (4) *и удовлетворяющих условию* $\lim_{k\to\infty} x^{(k)} = x^*$. *Тогда величина*

$$\alpha = \sup\left\{\lim_{k\to\infty} \sup \| q(x^{(k)}, x^*\|^{1/k} \{x^{(k)}\}_{k=0}^{\infty} \in \mathfrak{S}\right\}$$

называется асимптотическим фактором сходимости итерации (4) *к точке* x^*. ■

Пусть $\{x^{(k)}\}_{k=0}^{\infty}$ — последовательность, сходящаяся к x^*. Положим

$$\beta = \lim_{k\to\infty} \sup \| q(x^{(k)}, x^*)\|^{1/k}.$$

Так как $\lim_{k\to\infty} \| q(x^{(k)}, x^*)\| = 0$, мы получаем, что $0 \leqslant \beta < 1$, а значит, и $0 \leqslant \alpha \leqslant 1$. Из определения β следует, что для любого $\varepsilon > 0$ найдется k_0, такое что

$$\| q(x^{(k)}, x^*)\| \leqslant (\beta + \varepsilon)^k, \quad k \geqslant k_0. \tag{5}$$

Если $\beta < 1$, то можно выбрать $\varepsilon > 0$ таким образом, что $\beta + \varepsilon < 1$. Тогда неравенство (5) показывает, что последовательность $\| q(x^{(k)}, x^*)\|$ асимптотически сходится к нулю не хуже, чем геометрическая прогрессия со знаменателем $\beta + \varepsilon$. Точная верхняя грань по всем последовательностям из $\mathfrak{S}$ характеризует асимптотически наихудший выбор вектора $x^{(0)}$. Определение 10 — это непосредственное обобщение определения фактора асимптотической сходимости для последовательностей точечных векторов (см., например, [240] или [242]).

Для дальнейшего важно, что α не зависит от нормы. Чтобы убедиться в этом, достаточно показать, что β не зависит от нормы. Пусть $\|\cdot\|$ и $\|\cdot\|'$ — две нормы на векторах. Из теоремы об эквивалентности норм следует, что существуют вещественные числа $d \geqslant c > 0$, такие, что $c\| x_p\| \leqslant \| x_p\|' \leqslant d\| x_p\|$ для любых

точечных векторов. Отсюда мы получаем

$$\limsup_{k\to\infty} \| q(x^{(k)}, x^*)\|^{1/k} \leqslant \lim_{k\to\infty} \left(\frac{1}{c}\right)^{1/k} \limsup_{k\to\infty} \| q(x^{(k)}, x^*)\|'^{1/k}$$

$$\leqslant \lim_{k\to\infty} \left(\frac{d}{c}\right)^{1/k} \limsup_{k\to\infty} \|q(x^{(k)}, x^*)\|^{1/k}$$

$$= \limsup_{k\to\infty} \| q(x^{(k)}, x^*)\|^{1/k}.$$

Перед тем как применить определение 10 к методам, рассмотренным в этой главе, мы докажем одну теорему о точечных матрицах.

Рассмотрим положительный точечный вектор $h_{\mathrm{p}} = (h_i)$, $h_i > 0$ и диагональную точечную матрицу $\mathscr{H}_{\mathrm{p}} = \mathrm{diag}(1/h_i)$. Тогда неравенство

$$\| x_{\mathrm{p}}\| = \max_{1\leqslant i\leqslant n} (|x_i|/h_i)$$

определяет монотонную норму на векторах (короче, векторную норму), т .е. из

$$|x_{\mathrm{p}}| \leqslant |y_{\mathrm{p}}|$$

следует, что

$$\| x_{\mathrm{p}}\| \leqslant \| y_{\mathrm{p}}\|$$

(в частности, из $o_{\mathrm{p}} \leqslant x_{\mathrm{p}} \leqslant y_{\mathrm{p}}$ следует, что справедливо неравенство $\| x_{\mathrm{p}}\| \leqslant \| y_{\mathrm{p}}\|$).

Полагая

$$\| \mathscr{A}_{\mathrm{p}}\| = \sup_{\| x_{\mathrm{p}}\|=1} \| \mathscr{A}_{\mathrm{p}} x_{\mathrm{p}}\|,$$

мы получаем норму на матрицах, подчиненную векторной норме $\| x_{\mathrm{p}}\|$ (см. [74, с. 135]), причем

$$\| \mathscr{A}_{\mathrm{p}}\| = \max_{1\leqslant i\leqslant n} \left(\frac{1}{h_i} \sum_{j=1}^{n} |a_{ij}| h_j\right).$$

Теперь мы докажем следующее утверждение.

Теорема 11. *Для любой точечной матрицы* $\mathscr{A}_{\mathrm{p}} \geqslant \mathscr{O}_{\mathrm{p}}$ *и любого* $\varepsilon > 0$ *существует монотонная векторная норма* $\| \cdot \|$, *такая что имеет место*

$$\| \mathscr{A}_{\mathrm{p}}\| = \sup_{\| x_{\mathrm{p}}\|=1} \| \mathscr{A}_{\mathrm{p}} x_p\| \leqslant \rho(\mathscr{A}_{\mathrm{p}}) + \varepsilon.$$

Доказательство. Пусть матрица $\mathscr{A}_{\mathrm{p}} \geqslant \mathscr{O}_{\mathrm{p}}$ неприводима (см. [325]). Тогда $\mathscr{A}_{\mathrm{p}}$ имеет положительный собственный вектор $c_{\mathrm{p}} = (c_i)$, $c > 0$, соответствующий собственному числу $\lambda = \rho(\mathscr{A}_{\mathrm{p}})$

(см. [325, теорема 2.1]). Из равенства $\mathscr{A}_{\mathrm{p}} c_{\mathrm{p}} = \lambda c_{\mathrm{p}}$ следует, что

$$\rho(\mathscr{A}_{\mathrm{p}}) = \lambda = \frac{1}{c_i} \sum_{j=1}^{n} a_{ij} c_j = \| \mathscr{A}_p \| = \sup_{\| x_{\mathrm{p}} \| = 1} \| \mathscr{A}_{\mathrm{p}} x_{\mathrm{p}} \|,$$

где

$$\| x_{\mathrm{p}} \| = \max_{1 \leqslant i \leqslant n} \frac{| x_i |}{c_i},$$

т. е.

$$\| \mathscr{A}_{\mathrm{p}} \| \leqslant \rho(\mathscr{A}_{\mathrm{p}}) + \varepsilon.$$

Если $\mathscr{A}_{\mathrm{p}} \geqslant \mathscr{O}_{\mathrm{p}}$ приводима, то определим неприводимую матрицу $\tilde{\mathscr{A}}_{\mathrm{p}} \geqslant \mathscr{O}_{\mathrm{p}}$ равенствами

$$\tilde{\mathscr{A}}_p = (\tilde{a}_{ij}), \quad \text{где } \tilde{a}_{ij} = \begin{cases} a_{ij}, & \text{если } a_{ij} > 0 \\ a > 0, & \text{если } a_{ij} = 0. \end{cases}$$

Очевидно, что $\tilde{\mathscr{A}}_{\mathrm{p}} \geqslant \mathscr{A}_{\mathrm{p}} \geqslant \mathscr{O}_{\mathrm{p}}$. Если теперь $\tilde{c}_{\mathrm{p}} = (\tilde{c}_i)$ — положительный собственный вектор матрицы $\tilde{\mathscr{A}}_{\mathrm{p}}$, соответствующий собственному числу $\lambda = \rho(\tilde{\mathscr{A}}_{\mathrm{p}})$, то из неравенства $|\lambda| \leqslant \| \tilde{\mathscr{A}}_{\mathrm{p}} \|$ (которое верно для любой матричной нормы) следует, что

$$\rho(\tilde{\mathscr{A}}_{\mathrm{p}}) = \frac{1}{\tilde{c}_i} \sum_{j=1}^{n} \tilde{a}_{ij} \tilde{c}_j = \| \tilde{\mathscr{A}}_{\mathrm{p}} \| \geqslant \max_{1 \leqslant i \leqslant n} \frac{1}{c_i} \sum_{j=1}^{n} a_{ij} \tilde{c}_j = \| \mathscr{A}_{\mathrm{p}} \| \geqslant \rho(\mathscr{A}_{\mathrm{p}}),$$

где $\| x_{\mathrm{p}} \| = \max_{1 \leqslant i \leqslant n} | x_i | / \tilde{c}_i$ и $\| \mathscr{A}_{\mathrm{p}} \| = \sup_{\| x_{\mathrm{p}} \| = 1} \| \mathscr{A}_{\mathrm{p}} x_{\mathrm{p}} \|$. Так как спектральный радиус $\rho(\mathscr{A}_{\mathrm{p}})$ — непрерывная функция от элементов матрицы $\mathscr{A}_{\mathrm{p}}$ (см., например, [240]), мы получаем, что для каждого $\varepsilon > 0$ существует $\delta = \delta(\varepsilon) > 0$, такое что

$$\rho(\tilde{\mathscr{A}}_{\mathrm{p}}) - \rho(\mathscr{A}_{\mathrm{p}}) \leqslant \varepsilon$$

имеет место для всех $a \leqslant \delta(\varepsilon)$. Так как $\rho(\tilde{\mathscr{A}}_{\mathrm{p}}) \geqslant \| \mathscr{A}_{\mathrm{p}} \|$, мы получаем отсюда, что

$$\| \mathscr{A}_{\mathrm{p}} \| - \rho(\mathscr{A}_{\mathrm{p}}) \leqslant (\tilde{\mathscr{A}}_{\mathrm{p}}) - \rho(\mathscr{A}_p) \leqslant \varepsilon$$

или

$$\| \mathscr{A}_{\mathrm{p}} \| \leqslant \rho(\mathscr{A}_{\mathrm{p}}) + \varepsilon. \blacksquare$$

После этой подготовки мы легко докажем следующее утверждение.

Теорема 12. *Пусть дано уравнение*

$$x = \mathscr{A} x + b,$$

где b — интервальный вектор и $\mathscr{A}$ — интервальная матрица, для которой $\rho(|\mathscr{A}|) < 1$. Тогда асимптотический фактор сходимости α_{T} для полношагового метода удовлетворяет неравенству

$$\alpha_{\mathrm{T}} \leqslant \rho(|\mathscr{A}|),$$

фактор α_s *для короткошагового метода удовлетворяет неравенству*

$$\alpha_S \leqslant \rho((\mathscr{I}_p - |\mathscr{L}|)^{-1}(|\mathscr{D}| + |\mathscr{U}|)),$$

а фактор α_{SS} *для симметрического короткошагового метода удовлетворяет неравенству*

$$\alpha_{SS} \leqslant \rho((\mathscr{I}_p - |\mathscr{U}|)^{-1}|\mathscr{L}|(\mathscr{I}_p - |\mathscr{L}|)^{-1}|\mathscr{U}|).$$

Доказательство. Проведем доказательство для полношагового метода. Если $\rho(|\mathscr{A}|) < 1$, то, согласно теореме 1, полношаговый метод сходится для любого начального вектора $x^{(0)}$ к единственной неподвижной точке x^* уравнения $x = \mathscr{A}x + b$. Выбрав $x^{(0)}$ произвольным образом, мы получаем из свойств расстояния q, что

$$q(x^{(1)}, x^*) \leqslant |\mathscr{A}| q(x^{(0)}, x^*),$$
$$q(x^{(2)}, x^*) \leqslant |\mathscr{A}|^2 q(x^{(0)}, x^*),$$

и для произвольного $k \geqslant 1$

$$q(x^{(k)}, x^*) \leqslant |\mathscr{A}|^k q(x^{(0)}, x^*).$$

Используя монотонную векторную норму, которая существует по теореме 11, и подчиненную ей матричную норму, имеем

$$\| q(x^{(k)}, x^*) \| \leqslant (\rho(|\mathscr{A}|) + \varepsilon^k \| q(x^{(0)}, x^*) \|,$$

т. е

$$\limsup_{k \to \infty} \| q(x^{(k)}, x^*) \|^{1/k} \leqslant \rho(|\mathscr{A}|) + \varepsilon.$$

Так как $\varepsilon > 0$ было произвольным и α_T не зависит от нормы, мы получаем неравенство $\alpha_T \leqslant \rho(|\mathscr{A}|)$, т. е. утверждение теоремы для этого случая. Остальные случаи рассматриваются аналогично. ■

Неизвестно, можно ли поставить знак равенства в наших оценках для α_T, α_S, α_{SS}. Чтобы доказать, что это верно, скажем, для полношагового метода, нужно было бы указать начальный вектор $x^{(0)}$, для которого имеет место

$$\limsup_{k \to \infty} \| q(x^{(k)}, x^*) \| = \rho(|\mathscr{A}|).$$

Это можно довольно легко сделать в конкретных случаях, но доказательства для общего случая нет.

Замечания. Непосредственное применение интервального анализа к итерационному решению систем уравнений было впервые рассмотрено в [36]. Там было показано, что полношаговый метод (2) для вещественной интервальной матрицы $\mathscr{A}$ и вещественного интервального вектора b сходится к единственной непо-

движной точке, если $\|\,|\mathscr{A}|\,\| < 1$. Необходимые и достаточные условия из теорем 1 и 3 были найдены в [195] для вещественных интервальных матриц и в [4] для интервальных матриц с элементами из $R(\mathbb{C})$. Результаты теорем 1 и 3 для интервальных матриц с элементами из $K(\mathbb{C})$ новые. Теорему 11, которая нужна для доказательства теоремы 12, можно найти в [195].

Результаты для симметрического короткошагового метода (SS) были установлены в [348]. Результаты, аналогичные второй части теоремы 2, можно найти в [361]; см. также [36]. Заметим, не вдаваясь в подробности, что более общую задачу о неподвижной точке

$$\mathscr{X} = \mathscr{A}\mathscr{X} + \mathscr{B}, \quad \mathscr{A}, \mathscr{B}, \mathscr{X} \in M_{nn}(I(\mathbb{C}))$$

можно исследовать тем же методом, который был использован в этой главе (см. [36]). Например, если $\rho(\mathscr{A}) < 1$, то для итерации

$$\mathscr{X}^{(k+1)} = \mathscr{A}\mathscr{X}^{(k)} + \mathscr{B}, \quad k \geqslant 0,$$

имеется единственная неподвижная точка $\mathscr{X}^*$ при любой начальной матрице $\mathscr{X}^{(0)} \in M_{nn}(I(\mathbb{C}))$. Мы имеем в этом случае

$$\{\mathscr{Y}_{\mathrm{p}} = (\mathscr{I}_{\mathrm{p}} - \mathscr{A}_{\mathrm{p}})^{-1}\mathscr{B}_{\mathrm{p}} \,|\, \mathscr{A}_{\mathrm{p}} \in \mathscr{A},\ \mathscr{B}_{\mathrm{p}} \in \mathscr{B}\} \subseteq \mathscr{X}^*.$$

Мы хотели бы также сделать некоторые замечания по поводу условия $\rho(|\mathscr{A}|) < 1$, которое в силу теоремы 1 необходимо и достаточно для сходимости полношагового метода. В частном случае, когда $\mathscr{A}$ — точечная матрица и b — точечный вектор, мы получаем, что последовательные приближения

$$x^{(k+1)} = \mathscr{A}_{\mathrm{p}}x^{(k)} + b_p, \quad k \geqslant 0, \tag{6}$$

сходятся к решению

$$x_{\mathrm{p}} = (\mathscr{I}_{\mathrm{p}} - \mathscr{A}_{\mathrm{p}})^{-1} b_{\mathrm{p}}$$

для любого интервального вектора $x^{(0)}$ тогда и только тогда, когда $\rho(|\mathscr{A}_{\mathrm{p}}|) < 1$. С другой стороны, последовательные приближения

$$x_{\mathrm{p}}^{(k+1)} = \mathscr{A}_{\mathrm{p}}x_{\mathrm{p}}^{(k)} + b_{\mathrm{p}}, \quad k \geqslant 0, \tag{7}$$

сходятся для всех $x_{\mathrm{p}}^{(0)} \in V_n(\mathbb{C})$ к

$$x_{\mathrm{p}} = (\mathscr{I}_{\mathrm{p}} - \mathscr{A}_{\mathrm{p}})^{-1} b,$$

если $\rho(\mathscr{A}_{\mathrm{p}}) < 1$ (см. [325]).

Если мы проводим итерацию (6) в $V_n(K(\mathbb{C}))$, то в обозначениях $\mathscr{A}_{\mathrm{p}} = (a_{rs})$, $a_{rs} = a_{rs}^{(1)} + ia_{rs}^{(2)}$, $1 \leqslant r,\ s \leqslant n$, мы имеем в виду определения 10.6b, 6.4b:

$$|\mathscr{A}_{\mathrm{p}}| = |\mathscr{A}_{\mathrm{p}}|_2 := \left(\sqrt{(a_{rs}^{(1)})^2 + (a_{rs}^{(2)})^2}\right).$$

Если, с другой стороны, мы проводим итерацию (6) в $V_n(R(\mathbb{C}))$, то ввиду определений 10.6b, 6.2 имеем

$$|\mathscr{A}_p| = |\mathscr{A}_p|_1 := (|a_{rs}^{(1)}| + |a_{rs}^{(2)}|).$$

В силу неравенства

$$\sqrt{a_1^2 + a_2^2} \leqslant |a_1| + |a_2|$$

имеем

$$|\mathscr{A}_p|_2 \leqslant |\mathscr{A}_p|_1.$$

Из

$$\mathscr{O}_p \leqslant |\mathscr{A}_p|_2 \leqslant |\mathscr{A}_p|_1$$

и теоремы 2.8 из [325] следует, что

$$\rho(\mathscr{A}_p) \leqslant \rho(|\mathscr{A}_p|_2) \leqslant \rho(|\mathscr{A}_p|_1). \tag{8}$$

Последнее неравенство показывает, что (6) требует в общем случае более сильных предположений, чем (7). Это происходит потому, что необходимое и достаточное условие для (6) гарантирует сходимость для любых интервальных векторов, а множество всех точечных векторов, которые допускаются в (7) в качестве начальных, является собственным подмножеством множества всех интервальных векторов. Этот вопрос рассматривается в [195].

Вторая часть неравенства (8) показывает, что при реализации итерации (6) в $V_n(K(\mathbb{C}))$ от $\mathscr{A}_p$ требуется выполнение не менее сильных условий, чем в случае $V_n(R(\mathbb{C}))$. Это верно и для систем уравнений, коэффициенты которых — невырожденные интервалы. С этой точки зрения арифметические операции в $K(\mathbb{C})$ имеют некоторые преимущества перед операциями в $R(\mathbb{C})$.

* * *

Литература: [4], [7], [8], [35], [36], [74], [122], [175], [195], [197], [198], [199], [200], [201], [240], [242], [325], [345], [348], [361], [455].

Глава 13
МЕТОДЫ РЕЛАКСАЦИИ

Мы уже рассмотрели полношаговый, короткошаговый и симметрический короткошаговый методы. Существует много других методов решения линейных систем точечных уравнений вида

$$x_p = \mathscr{A}_p x_p + b_p,$$

для которых можно уменьшить асимптотический фактор сходимости путем введения одного или нескольких параметров (см. [325]). Бóльшую часть этих приемов можно перенести на итерационные методы, использующие интервальные векторы. В качестве примера мы рассмотрим метод релаксации для коротко-шагового случая.

Как и в короткошаговом методе, разложим матрицу $\mathscr{A}$ в сумму $\mathscr{A} = \mathscr{L} + \mathscr{D} + \mathscr{U}$, где $\mathscr{L}$ — нижняя строго треугольная матрица, $\mathscr{U}$ — верхняя строго треугольная матрица и $\mathscr{D}$ — диагональная матрица. Затем строим последовательные приближения

$$\tilde{X}_i^{(k+1)} = \sum_{j=1}^{i-1} A_{ij} X_j^{(k+1)} + \sum_{j=i}^{n} A_{ij} X_j^{(k)} + B_i,$$

$$X_i^{(k+1)} = (1 - \omega) X_i^{(k)} + \omega \tilde{X}_i^{(k+1)}, \quad 1 \leqslant i \leqslant n, \quad k \geqslant 0,$$

начиная с произвольного интервального вектора $x^{(0)}$. С помощью векторных обозначений эти формулы можно записать в виде

$$x^{(k+1)} = (1 - \omega) x^{(k)} + \omega \{\mathscr{L} x^{(k+1)} + (\mathscr{D} + \mathscr{U}) x^{(k)} + b\}, \quad k \geqslant 0,$$

где $\omega > 0$ — параметр. Так же, как это делалось для полношагового или короткошагового метода, можно показать, что

$$\rho((\mathscr{I}_p - \omega |\mathscr{L}|)^{-1} \{|1 - \omega| \mathscr{I}_p + \omega (|\mathscr{D}| + |\mathscr{U}|)\}) < 1$$

является необходимым и достаточным условием сходимости метода к единственной неподвижной точке для произвольного начального вектора.

Можно далее показать, что при $\rho(|\mathscr{A}|) < 1$ это условие выполнено для всех значений ω, которые удовлетворяют неравенству

$$0 < \omega < 2/(1 + \rho(|\mathscr{A}|))$$

(см., например, [4]). Если полношаговый метод сходится и ω удовлетворяет приведенному неравенству, то метод релаксации тоже сходится к неподвижной точке $\tilde{x}^*$, удовлетворяющей уравнению

$$\tilde{x}^* = (1 - \omega) \tilde{x}^* + \omega (\mathscr{A} \tilde{x}^* + b).$$

Эта неподвижная точка, вообще говоря, отлична от неподвижной точки x^* уравнения $x = \mathscr{A} x + b$. Чтобы показать это, заметим сначала, что для всех вещественных чисел a, b и невырожденных интервалов Z (т. е. таких Z, что $d(Z) > 0$) при $ab > 0$ выполнено соотношение

$$(a + b) Z = aZ + bZ.$$

Пусть $0 < \omega < 1$. Тогда имеем

$$(1 - \omega) x^* + \omega (\mathscr{A} x^* + b) = (1 - \omega) x^* + \omega x^* = (1 - \omega + \omega) x^* = x^*;$$

т. е. $x^* = \tilde{x}^*$, так как неподвижная точка $\tilde{x}^*$, вычисленная методом релаксации, единственна.

Если же, напротив, мы имеем $\omega > 1$, то

$$\begin{aligned} \mathscr{A}x^* + b = x^* &= (1 - \omega + \omega)\, x^* \subseteq (1 - \omega)\, x^* + \omega x^* \\ &= (1 - \omega)\, x^* + \omega(\mathscr{A}x^* + b) \\ &= (1 - \omega)\, x^* + \omega\{\mathscr{L}x^* + (\mathscr{D} + \mathscr{U})\, x^* + b\}. \end{aligned}$$

Если мы возьмем начальное приближение $\tilde{x}^{(0)} = x^*$ для метода релаксации, то из этого включения и монотонности включения следует, что

$$x^* \subseteq (1 - \omega)\, \tilde{x}^{(0)} + \omega\{\mathscr{L}\tilde{x}^{(1)} + (\mathscr{D} + \mathscr{U})\, \tilde{x}^{(0)} + b\} =: \tilde{x}^{(1)}.$$

С помощью математической индукции можно показать, что $x^* \subseteq \tilde{x}^{(k)}$, $k \geqslant 0$, т. е. $x^* \subseteq \tilde{x}^*$. Из простых примеров видно, что включение здесь собственное. Поэтому при $\omega > 1$ мы должны учитывать, что применение метода релаксации «увеличивает» неподвижную точку уравнения $x^* = \mathscr{A}x^* + b$. Это нежелательный эффект, так как задача состоит в нахождении интервального вектора, который содержит множество

$$\{y_p = (\mathscr{I}_p - \mathscr{A}_p)^{-1} b_p \mid \mathscr{A}_p \in \mathscr{A},\ b_p \in b\}$$

и дает достаточно хорошую локализацию.

В частном случае, когда $\mathscr{A}$ — точечная матрица и b — точечный вектор, мы всегда имеем $x^* = x^*$. Множество точечных векторов — это подмножество множества интервальных векторов. Если мы выберем начальный вектор точечным, то все последовательные приближения и неподвижная точка также будут точечными векторами. Поэтому мы имеем для полношагового метода равенство

$$x_p^* = \mathscr{A}_p x_p^* + b_p,$$

а для метода релаксации — равенство

$$\tilde{x}_p^* = (1 - \omega)\, \tilde{x}_p^* + \omega\,(\mathscr{A}_p \tilde{x}_p^* + b_p),$$

откуда следует, что $\tilde{x}_p^* = x_p^*$. Таким образом, для случая точечной матрицы и точечного вектора оба метода сходятся к решению

$$x_p^* = (\mathscr{I}_p - \mathscr{A}_p)^{-1} b_p$$

уравнения $x_p = \mathscr{A}_p x_p + b_p$.

Мы хотим теперь исследовать этот случай несколько подробнее. Последовательность $\{d(x^{(k)})\}_{k=0}^{\infty}$, полученная вычислением ширины из последовательности $\{x^{(k)}\}_{k=0}^{\infty}$, сходится к нулевому

вектору, так как $\lim_{k\to\infty} x^{(k)} = x_{\mathrm{p}}^* = (\mathscr{I}_{\mathrm{p}} - \mathscr{A}_{\mathrm{p}})^{-1} b_{\mathrm{p}}$. Поэтому кажется, что естественно характеризовать скорость сходимости величиной $\tilde{\alpha}$, вводимой следующим определением.

Определение 1. Пусть $x_{\mathrm{p}}^* = \mathscr{f}_{\mathrm{p}}(x_{\mathrm{p}}^*)$, и пусть $\mathfrak{S}$ обозначает множество всех последовательностей $\{x^{(k)}\}_{k=0}^{\infty}$, которые получаются применением итерационного метода

$$x^{(k+1)} = \mathscr{f}_{\mathrm{p}}(x^{(k)}), \quad k \geqslant 0,$$

и для которых $\lim_{k\to\infty} x^{(k)} = x_{\mathrm{p}}^*$. Тогда величина

$$\tilde{\alpha} = \sup\left\{\limsup_{k\to\infty} \| d(x_{\mathrm{p}}^{(k)}) \|^{1/k} \mid \{x^{(k)}\}_{k=0}^{\infty} \in \mathfrak{S}\right\}$$

называется асимптотическим фактором сходимости этого итерационного метода в точке x_{p}^*. ■

По аналогии с величиной α (определение 12.10) мы можем сказать, что $\tilde{\alpha}$ характеризует асимптотически наихудший случай при произвольном выборе $x^{(0)}$. Точно так же, как для α, можно показать, что $\tilde{\alpha}$ не зависит от используемой нормы.

Докажем теперь следующее утверждение.

Теорема 2. *Пусть задано уравнение*

$$x_{\mathrm{p}} = \mathscr{A}_{\mathrm{p}} x_{\mathrm{p}} + b_{\mathrm{p}}$$

для точечной матрицы $\mathscr{A}_{\mathrm{p}}$, *такой что* $\rho(|\mathscr{A}_{\mathrm{p}}|) < 1$, *и точечного вектора* b_{p}. *Тогда асимптотический фактор сходимости* $\tilde{\alpha}_{\mathrm{T}}$ *для полношагового метода* (12.2) *удовлетворяет равенству*

$$\tilde{\alpha}_{\mathrm{T}} = \rho(|\mathscr{A}_{\mathrm{p}}|),$$

а асимптотический фактор сходимости $\tilde{\alpha}_{\mathrm{R}}$ *для метода релаксации — равенству*

$$\tilde{\alpha}_{\mathrm{R}} = \rho((\mathscr{I}_{\mathrm{p}} - \omega|\mathscr{L}_{\mathrm{p}}|)^{-1}\{|1-\omega|\mathscr{I}_{\mathrm{p}} + \omega(|\mathscr{D}_{\mathrm{p}}| + |\mathscr{U}_{\mathrm{p}}|)\})$$

для

$$0 < \omega < 2/(1 + \rho(|\mathscr{A}_{\mathrm{p}}|)).$$

Доказательство. Проведем рассуждение для метода релаксации. Начав с произвольного интервального вектора $x^{(0)}$, мы с помощью (10.12) и (10.19) получаем из нашей итерационной формулы

$$x^{(k+1)} = (1-\omega)x^{(k)} + \omega\{\mathscr{L}_{\mathrm{p}} x^{(k+1)} + (\mathscr{D}_{\mathrm{p}} + \mathscr{U}_{\mathrm{p}})x^{(k)} + b_{\mathrm{p}}\}, \quad k \geqslant 0$$

что

$$d(x^{(k+1)}) = |1-\omega| d(x^{(k)}) + \omega|\mathscr{L}_{\mathrm{p}}| d(x^{(k+1)}) + \omega(|\mathscr{D}_{\mathrm{p}}| + |\mathscr{U}_{\mathrm{p}}|) d(x^{(k)})$$

или

$$d(x^{(k+1)}) = (\mathscr{T}_p - \omega|\mathscr{L}_p|)^{-1}\{|1-\omega|\mathscr{T}_p + \omega(|\mathscr{D}_p| + |\mathscr{U}_p|)\}\, d(x^{(k)})$$
$$= \{(\mathscr{T}_p - \omega|\mathscr{L}_p|)^{-1}(|1-\omega|\mathscr{T}_p + \omega(|\mathscr{D}_p| + |\mathscr{U}_p|))\}^{k+1} d(x^{(0)}).$$

Отсюда непосредственно получаем, что

$$\tilde{\alpha}_R \leqslant \rho((\mathscr{T}_p - \omega|\mathscr{L}_p|)^{-1}\{|1-\omega|\mathscr{T}_p + \omega(|\mathscr{D}_p| + |\mathscr{U}_p|)\}).$$

Если теперь выбрать конкретный вектор $x^{(0)}$ так, что $d(x^{(0)})$ — собственный вектор неотрицательной матрицы $(\mathscr{T}_p - \omega|\mathscr{L}_p|)^{-1} \times \{|1-\omega|\mathscr{T}_p + \omega(|\mathscr{D}_p| + |\mathscr{U}_p|)\}$, соответствующий собственному числу λ, равному спектральному радиусу этой матрицы, то из уравнения для $d(x^{(k+1)})$ следует, что

$$d(x^{(k+1)}) = \lambda^{k+1} d(x^{(0)}),$$

откуда получается нужное утверждение. Доказательство для полношагового метода можно провести аналогично. ■

Только что доказанная теорема позволяет сформулировать утверждение об асимптотически наискорейшем (в смысле определения 1) методе.

Теорема 3. *В условиях теоремы 2*

$$\min\{\tilde{\alpha}_R | 0 < \omega < 2/(1+\rho(|\mathscr{A}_p|))\} = \tilde{\alpha}_R|_{\omega=1}$$
$$= \rho((\mathscr{T}_p - |\mathscr{L}_p|)^{-1}(|\mathscr{D}_p| + |\mathscr{U}_p|)) = \tilde{\alpha}_S,$$

а также

$$\tilde{\alpha}_S \leqslant \tilde{\alpha}_T.$$

Доказательство. Рассмотрим вещественную точечную матрицу

$$\mathscr{P}_p = ((1 - |1-\omega|)/\omega)\,\mathscr{T}_p - |\mathscr{A}_p|$$

и ее разложение $\mathscr{P}_p = \mathscr{M}_{p\omega} - \mathscr{N}_{p\omega}$, где

$$\mathscr{M}_{p\omega} = (1/\omega)(\mathscr{T}_p - \omega|\mathscr{L}_p|), \quad \mathscr{N}_{p\omega} = (1/\omega)\{|1-\omega|\mathscr{T}_p + \omega(|\mathscr{D}_p| + |\mathscr{U}_p|)\}.$$

Это разложение регулярно (см. [325]), так как $\mathscr{M}_{p\omega}^{-1}$ существует и $\mathscr{M}_{p\omega}^{-1} \geqslant \mathscr{O}_p$, $\mathscr{N}_{p\omega} \geqslant \mathscr{O}_p$. Если ω удовлетворяет неравенству $0 < \omega < 2/(1+\rho|\mathscr{A}_p|)$, то $\mathscr{P}_p^{-1} \geqslant \mathscr{O}_p$, и в силу неравенства $\mathscr{N}_{p\omega} \geqslant \mathscr{N}_{p1}$ мы получаем с помощью очевидного обобщения теоремы 3.15 из [325], что

$$\rho(\mathscr{M}_{p1}^{-1}\mathscr{N}_{p1}) = \rho((\mathscr{T}_p - |\mathscr{L}_p|)^{-1}(|\mathscr{D}_p| + |\mathscr{U}_p|)) \leqslant \rho(\mathscr{M}_{p\omega}^{-1}\mathscr{N}_{p\omega}) < 1.$$

Этим доказана первая часть теоремы. Ее вторая часть следует из теоремы Штейна и Розенберга и ее обобщения (см.

[325] и [35]), утверждающего, что из $\rho(|\mathscr{A}_p|)<1$ следует

$$\rho((\mathscr{I}_p-|\mathscr{L}_p|)^{-1}(|\mathscr{D}_p|+|\mathscr{U}_p|))\leqslant\rho(|\mathscr{A}_p|). \blacksquare$$

В предыдущей теореме утверждается, что в случае системы точечных уравнений невозможно асимптотически (в смысле определения 1) ускорить сходимость короткошагового метода, используя метод релаксации. Кроме того, короткошаговый метод сходится не медленнее, чем полношаговый.

Теперь рассмотрим связь между описанным выше итерационным методом в пространстве интервальных векторов и принципом локализации решений, который получен иным способом (см. [74, с. 275 и далее], [325] и [296]).

Рассмотрим систему линейных уравнений

$$x_p=\mathscr{A}_p x_p+b_p,$$

где $\mathscr{A}_p$ — вещественная точечная матрица и b_p — вещественный точечный вектор. Мы вводим естественный (т. е. покомпонентный) порядок на точечных матрицах и точечных векторах. Точечная матрица $\mathscr{P}_p$ называется изотонной (соответственно антитонной), если из $x_p\geqslant o_p$ следует $\mathscr{P}_p x_p\geqslant o_p$ (соответственно $\leqslant o_p$). Теперь матрица $\mathscr{A}_p$ раскладывается в сумму изотонной и антитонной точечных матриц:

$$\mathscr{A}_p=\mathscr{T}_{p1}+\mathscr{T}_{p2}.$$

Начиная с пары точечных векторов $v_p^{(0)}$ и $w_p^{(0)}$, для которых верно $v_p^{(0)}\leqslant w_p^{(0)}$, наш итерационный метод вычисляет две последовательности $\{v_p^{(k)}\}_{k=0}^{\infty}$ и $\{w_p^{(k)}\}_{k=0}^{\infty}$ по формулам

$$\text{(1)}\qquad \begin{cases} v_p^{(k+1)}=\mathscr{T}_{p1}v_p^{(k)}+\mathscr{T}_{p2}w_p^{(k)}+b_p,\\ w_p^{(k+1)}=\mathscr{T}_{p1}w_p^{(k)}+\mathscr{T}_{p2}v_p^{(k)}+b_p, \quad k\geqslant 0.\end{cases}$$

Если теперь $v_p^{(0)}\leqslant v_p^{(1)}\leqslant w_p^{(1)}\leqslant w_p^{(0)}$, то с помощью математической индукции можно показать, что

$$v_p^{(0)}\leqslant v_p^{(1)}\leqslant\ldots\leqslant v_p^{(k)}\leqslant v_p^{(k+1)}\leqslant w_p^{(k+1)}\leqslant w_p^{(k)}\leqslant\ldots\leqslant w_p^{(1)}\leqslant w_p^{(0)}.$$

Поэтому последовательности $\{v_p^{(k)}\}_{k=0}^{\infty}$ и $\{w_p^{(k)}\}_{k=0}^{\infty}$ сходятся, и простые рассуждения показывают, что решение x_p^* уравнения $x_p=\mathscr{A}_p x_p+b_p$ существует и находится между граничными точками. Если $\rho(|\mathscr{A}_p|)=\rho(\mathscr{T}_{p1}-\mathscr{T}_{p2})<1$, то $v_p^*=w_p^*=x_p^*$.

Рассматривая вместе с методом итераций для $\{v_p^{(k)}\}_{k=0}^{\infty}$ и $\{w_p^{(k)}\}_{k=0}^{\infty}$ также итерационный метод

$$\text{(2)}\qquad x^{(k+1)}=\mathscr{A}_p x^{(k)}+b_p, \quad k\geqslant 0,$$

с интервалом $x^{(0)} = ([v_i^{(0)}, w_i^{(0)}])$ в качестве начального приближения и учитывая правила умножения интервальных матриц на интервальные векторы, мы сразу видим, что границы компонент последовательности интервальных векторов $\{x^{(k)}\}_{k=0}^{\infty}$ совпадают с компонентами последовательностей $\{v_p^{(k)}\}_{k=0}^{\infty}$ и $\{w_p^{(k)}\}_{k=0}^{\infty}$.

Сказанное только что об итерационном методе (2) показывает, что в методе (1) можно избавиться от предположения $v_p^{(0)} \leqslant v_p^{(1)} \leqslant w_p^{(1)} \leqslant w_p^{(0)}$, если выполнено условие $v_p^{(0)} \leqslant x_p^* \leqslant w_p^{(0)}$, которое гарантирует локализацию решения x_p^*. Последовательности $\{v_p^{(k)}\}_{k=0}^{\infty}$ и $\{w_p^{(k)}\}_{k=0}^{\infty}$ в этом случае уже *не обязательно* сходятся монотонно. Монотонность можно восстановить, если брать пересечение после каждого шага (см. гл. 14).

Замечания. Утверждения, аналогичные тем, которые мы доказали для метода релаксации, справедливы и для симметрического метода релаксации (SR)

$$x^{(k+1/2)} = (1-\omega)\,x^{(k)} + \omega\{\mathscr{L}x^{(k+1/2)} + \mathscr{U}x^{(k)} + b\},$$

$$x^{(k+1)} = (1-\omega)\,x^{(k+1/2)} + \omega\{\mathscr{L}x^{(k+1/2)} + \mathscr{U}x^{(k+1)} + b\}, \quad k \geqslant 0,$$

который при $\omega = 1$ сводится к симметрическому короткошаговому методу (SS), описанному в гл. 12. Необходимым и достаточным условием сходимости этого метода к единственной неподвижной точке при произвольном начальном интервальном векторе является

$$\rho((\mathscr{I}_p - \omega|\mathscr{U}|)^{-1}(|1-\omega|\mathscr{I}_p + \omega|\mathscr{L}|)(\mathscr{I}_p - \omega|\mathscr{L}|)^{-1} \times (|1-\omega|\mathscr{I}_p + \omega|\mathscr{U}|)) < 1.$$

Если $\rho(|\mathscr{A}|) < 1$, то предыдущее условие выполняется при

$$0 < \omega < 2/(1 + \rho(|\mathscr{A}|)).$$

Доказательство можно провести так же, как для метода релаксации.

Если $\mathscr{A}$ — точечная матрица, то по аналогии с рассуждением из теоремы 2 можно показать, что асимптотический фактор сходимости, введенный в определении 1, равен

$$\tilde{\alpha}_{SR} = \rho((\mathscr{I}_p - \omega|\mathscr{U}_p|)^{-1}(|1-\omega|\mathscr{I}_p + \omega|\mathscr{L}_p|) \times (\mathscr{I}_p - \omega|\mathscr{L}_p|)^{-1}(|1-\omega|\mathscr{I}_p + \omega|\mathscr{U}_p|)),$$

и по аналогии с теоремой 3 имеем

$$\tilde{\alpha}_{SS} \leqslant \tilde{\alpha}_{SR}$$

для $0 < \omega < 2/(1 + \rho(|\mathscr{A}_p|))$.

Покажем, что имеет место $\tilde{\alpha}_{ss} \leqslant \tilde{\alpha}_s$. Матрица

$$(\mathscr{I}_p - |\mathscr{L}_p|)^{-1} |\mathscr{U}_p|$$

неотрицательна и всегда приводима, так как ее первый столбец состоит из нулей. Если добавить положительную матрицу $\Delta\mathscr{U}_p$ (вообще говоря, не являющуюся верхней треугольной), то можно сделать матрицу $(\mathscr{I}_p - |\mathscr{L}_p|)^{-1}(|\mathscr{U}_p| + \Delta\mathscr{U}_p)$ неприводимой.

Используя теорему Перрона и Фробениуса (см. [325]), мы получаем

$$(\mathscr{I}_p - |\mathscr{L}_p|)^{-1}(|\mathscr{U}_p| + \Delta\mathscr{U}_p)\, x_p = \lambda x_p,$$

где λ — спектральный радиус матрицы из левой части и вектор x_p имеет только положительные компоненты. Простое преобразование (корректное при $0 < \lambda < 1$ и достаточно малой $\Delta\mathscr{U}_p$) дает

$$x_p = (\mathscr{I}_p - (1/\lambda)(|\mathscr{U}_p| + \Delta\mathscr{U}_p))^{-1} |\mathscr{L}_p|\, x_p$$
$$\geqslant (\mathscr{I}_p - (|\mathscr{U}_p| + \Delta\mathscr{U}_p))^{-1} |\mathscr{L}_p|\, x_p.$$

Окончательно получаем

$$(\mathscr{I}_p - (|\mathscr{U}_p| + \Delta\mathscr{U}_p))^{-1} |\mathscr{L}_p| (\mathscr{I}_p - |\mathscr{L}_p|)^{-1} (|\mathscr{U}_p| + \Delta\mathscr{U}_p)\, x_p \leqslant \lambda x_p.$$

Это неравенство верно и при $\lambda = 0$. Применяя теперь теорему 1.11 из [366], получаем, что спектральный радиус матрицы из левой части не превосходит λ. Так как это верно для всех матриц $\Delta\mathscr{U}_p$, делающих матрицу $(\mathscr{I}_p - |\mathscr{L}_p|)^{-1}(|\mathscr{U}_p| + \Delta\mathscr{U}_p)$ неприводимой, то мы получаем нужное утверждение, устремляя $\Delta\mathscr{U}_p$ к нулю, так как собственные числа непрерывно зависят от элементов матрицы.

* * *

Литература: [3], [4], [35], [74], [201], [296], [325], [366].

Глава 14

ОПТИМАЛЬНОСТЬ СИММЕТРИЧЕСКОГО КОРОТКОШАГОВОГО МЕТОДА СО ВЗЯТИЕМ ПЕРЕСЕЧЕНИЯ НА КАЖДОМ ШАГЕ

В этой главе предполагается, что все интервальные матрицы взяты из пространства $M_{nn}(R(\mathbb{C}))$, а все интервальные векторы — из $V_n(R(\mathbb{C}))$.

Мы собираемся теперь исследовать некоторые модификации полношагового, короткошагового и симметрического короткоша-

гового методов. Если полношаговый метод

$$x^{(k+1)} = \mathscr{A}x^{(k)} + b$$

имеет начальный вектор $x^{(0)}$ для которого $x^* \subseteq x^{(0)}$, то из монотонности включения следует, что

$$x^* = \mathscr{A}x^* + b \subseteq \mathscr{A}x^{(0)} + b = x^{(1)}.$$

Это показывает, что и $x^{(1)}$ содержит неподвижную точку и вектор $x^{(0)}$, а значит, и их пересечение $x^{(0)} \cap x^{(1)}$. Поэтому естественно продолжать итерацию, используя это новое включение. Это приводит к итерационной процедуре

$$x^{(k+1)} = \{\mathscr{A}x^{(k)} + b\} \cap x^{(k)}, \quad k \geqslant 0,$$

которую мы будем называть полношаговым методом со взятием пересечения на каждом шаге (TI).

Если провести те же рассмотрения для короткошагового метода, то получится итерационная процедура

$$x^{(k+1)} = \{\mathscr{L}x^{(k+1)} + (\mathscr{D} + \mathscr{U})x^{(k)} + b\} \cap x^{(k)}, \quad k \geqslant 0,$$

которую мы назовем короткошаговым методом со взятием пересечения на каждом шаге (SI).

В случае короткошагового метода имеется еще одна возможность:

$$X_i^{(k+1)} = \left\{\sum_{j=1}^{i-1} A_{ij}X_j^{(k+1)} + \sum_{j=i}^{n} A_{ij}X_j^{(k)} + B_i\right\} \cap X_i^{(k)}, \quad 1 \leqslant i \leqslant n, \quad k \geqslant 0.$$

После того как вычислена первая компонента, образуется пересечение со старым приближением, дающее новое приближение. Это новое приближение используется для вычисления нового приближения для второй компоненты и т. д. Эта модификация называется короткошаговым методом со взятием пересечения после каждой компоненты (SIC).

Наконец, для случая, когда все диагональные элементы матрицы $\mathscr{A}$ обращаются в нуль, рассмотрим (снова в предположении $x^* \subseteq x^{(0)}$) итерационную процедуру

$$X_i^{(k+1/2)} = \left\{\sum_{j=1}^{i-1} A_{ij}X_j^{(k+1/2)} + \sum_{j=i+1}^{n} A_{ij}X_j^{(k)} + B_i\right\} \cap X_i^{(k)}, \quad 1 \leqslant i \leqslant n,$$

$$X_i^{(k+1)} = \left\{\sum_{j=1}^{i-1} A_{ij}X_j^{(k+1/2)} + \sum_{j=i+1}^{n} A_{ij}X_j^{(k+1)} + B_i\right\} \cap X_i^{(k+1/2)},$$

$$1 \leqslant i \leqslant n, \quad k \geqslant 0.$$

Мы будем называть эту процедуру симметрическим короткошаговым методом со взятием пересечения после каждой компоненты (SSIC).

Метод (SSIC) можно выполнять таким образом, что он будет требовать на каждом шаге (за исключением самого первого) того же количества интервальных умножений, что и метод (SIC). В обоих случаях требуется $n^2 - n$ умножений (в предположении, что диагональные элементы обращаются в 0). См. также [455]. При этом используются следующие соображения. Допустим, что для некоторого $k > 0$ известны суммы

$$\sum_{j=i+1}^{n} A_{ij} X_j^{(k)}, \quad 1 \leqslant i \leqslant n-1.$$

Вычисление векторов $x^{(k+1/2)}$ по методу (SSIC) требует $\frac{1}{2}(n^2 - n)$ умножений. Если запоминается $n - 1$ сумма

$$\sum_{j=1}^{i-1} A_{ij} X_j^{(k+1/2)}, \quad 2 \leqslant i \leqslant n,$$

то вычисление $x^{(k+1)}$ из $x^{(k+1/2)}$ требует еще $\frac{1}{2}(n^2 - n)$ умножений. Таким образом, всего для вычисления приближения $x^{(k+1)}$, исходя из $x^{(k)}$ нам требуется $n^2 - n$ умножений, как и в методе (SIC). Если при вычислении $x^{(k+1)}$ из $x^{(k)}$ запоминаются суммы

$$\sum_{j=i+1}^{n} A_{ij} X_j^{(k+1)},$$

то все эти рассуждения проходят для индекса, увеличенного на 1.

Следующее утверждение показывает, что итерационные методы (TI), (SI), (SIC) и (SSIC) сходятся к неподвижной точке x^*.

Теорема 1. *Пусть $\mathcal{A}$ — интервальная матрица и $\rho(|\mathcal{A}|) < 1$. Если x^* — неподвижная точка уравнения $x = \mathcal{A}x + b$ и $x^{(0)} \supseteq x^*$, то итерационные методы* (TI), (SI), (SIC) *и* (SSIC) *сходятся к x^*.*

Доказательство. Докажем теорему для метода (TI). Так как мы берем пересечения, полученная последовательность приближений $\{x^{(k)}\}_{k=0}^{\infty}$ удовлетворяет условию

$$x^{(0)} \supseteq x^{(1)} \supseteq \ldots \supseteq x^{(k)} \supseteq x^{(k+1)} \supseteq \ldots .$$

В силу следствия 10.8 эта последовательность сходится к некоторому пределу $\tilde{x}^*$, и этот предел удовлетворяет условию $x^* \subseteq \tilde{x}^*$. Операция взятия пересечения также непрерывна (если пересечение непусто), поэтому при $k \to \infty$ мы имеем

$$\tilde{x}^* = \{\mathcal{A}\tilde{x}^* + b\} \cap \tilde{x}^*,$$

откуда следует, что

$$\mathcal{A}\tilde{x}^* + b \supseteq \tilde{x}^*.$$

Мы рассматриваем полношаговый метод

$$y^{(k+1)} = \mathcal{A}y^{(k)} + b, \qquad k \geqslant 0,$$

где $y^{(0)} = \tilde{x}^*$. Из сказанного следует, что

$$y^{(1)} = \mathcal{A}y^{(0)} + b = \mathcal{A}\tilde{x}^* + b \supseteq \tilde{x}^* \supseteq x^*,$$
$$y^{(2)} = \mathcal{A}y^{(1)} + b \supseteq \mathcal{A}\tilde{x}^* + b \supseteq \tilde{x}^* \supseteq x^*$$

и вообще

$$y^{(k+1)} = \mathcal{A}y^{(k)} + b \supseteq \mathcal{A}\tilde{x}^* + b \supseteq \tilde{x}^* \supseteq x^*.$$

Последовательность $\{y^{(k)}\}_{k=0}^{\infty}$, вычисленная с помощью рассматриваемого итерационного метода, сходится к x^* в силу теоремы 12.1. Последнее из доказанных включений дает при $k \to \infty$ соотношение

$$x^* \supseteq \tilde{x}^* \supseteq x^*,$$

т. е. $\tilde{x}^* = x^*$. Для остальных методов доказательство аналогично. ■

Теперь мы сравним методы (T), (S), (TI), (SI), (SIC) и (SSIC) по скорости сходимости для случая, когда итерации начинаются с интервального вектора, содержащего неподвижную точку x^*.

В первой теореме метод (T) сравнивается с (TI), а метод (S) — с (SI).

Теорема 2. *Пусть последовательности* $\{x^{(k)}\}_{k=0}^{\infty}$ *и* $\{\tilde{x}^{(k)}\}_{k=0}^{\infty}$ *вычислены согласно итерационным методам* (T) *и* (TI) *в предположении, что* $x^{(0)} \supseteq \tilde{x}^{(0)} \supseteq x^*$. *Тогда имеет место*

$$x^{(k)} \supseteq \tilde{x}^{(k)} \supseteq x^* \quad \text{для всех} \quad k \geqslant 0.$$

Такое же утверждение справедливо для последовательностей, вычисленных согласно итерационным методам (S) *и* (SI).

Доказательство. Мы докажем теорему для методов (T) и (TI). Соотношение $\tilde{x}^{(k)} \supseteq x^*$, $k \geqslant 0$, уже было доказано в предположении $\tilde{x}^{(0)} \supseteq x^*$ при получении формул для итерационного метода (TI). Допустим, что для некоторого $k \geqslant 0$ уже доказано, что

$$x^{(k)} \supseteq \tilde{x}^{(k)}.$$

Для $k=0$ это верно ввиду нашего исходного допущения. Используя монотонность включения, мы получаем

$$x^{(k+1)}=\mathscr{A}x^{(k)}+b\supseteq\mathscr{A}\tilde{x}^{(k)}+b\supseteq\{\mathscr{A}\tilde{x}^{(k)}+b\}\cap\tilde{x}^{(k)}=\tilde{x}^{(k+1)},$$

что завершает доказательство по индукции.

Доказательство для методов (S) и (SI) можно провести аналогичным образом. ■

Теорема 3. *Пусть последовательности* $\{x^{(k)}\}_{k=0}^{\infty}$ *и* $\{\tilde{x}^{(k)}\}_{k=0}^{\infty}$ *вычислены согласно итерационным методам* (TI) *и* (SIC) *в предположении, что* $x^{(0)}\supseteq\tilde{x}^{(0)}\supseteq x^*$. *Тогда имеет место*

$$x^{(k)}\supseteq\tilde{x}^{(k)}\supseteq x^* \quad \textit{для всех} \quad k\geqslant 0.$$

Такое же утверждение справедливо для последовательностей, вычисленных согласно итерационным методам (S) *и* (SIC).

Доказательство. Нам нужно доказать лишь соотношение $x^{(k)}\supseteq\tilde{x}^{(k)}$. Мы проведем доказательство для последовательностей, вычисленных согласно итерационным методам (TI) и (SIC). Допустим, что для некоторого $k\geqslant 0$ верно

$$x^{(k)}\supseteq\tilde{x}^{(k)}.$$

Для $k=0$ это верно в силу нашего исходного допущения. Тогда в обозначениях $x^{(k)}=(X_i^{(k)})$, $\tilde{x}^{(k)}=(\tilde{X}_i^{(k)})$, $\mathscr{A}=(A_{ij})$, $b=(B_i)$ мы имеем

$$X_1^{(k+1)}=\left(\sum_{j=1}^{n}A_{1j}X_j^{(k)}+B_1\right)\cap X_1^{(k)},$$

$$\tilde{X}_1^{(k+1)}=\left(\sum_{j=1}^{n}A_{1j}\tilde{X}_j^{(k)}+B_1\right)\cap\tilde{X}_1^{(k)}.$$

Из $\tilde{X}_i^{(k)}\subseteq X_i^{(k)}$, $1\leqslant i\leqslant n$, и монотонности включения следует, что

$$\sum_{j=1}^{n}A_{1j}\tilde{X}_j^{(k)}+B_1\subseteq\sum_{j=1}^{n}A_{1j}X_j^{(k)}+B_1,$$

а потому

$$\tilde{X}_1^{(k+1)}\subseteq X_1^{(k+1)}.$$

Ввиду $\tilde{X}_1^{(k+1)}\subseteq\tilde{X}_1^{(k)}\subseteq X_1^{(k)}$ отсюда следует, что

$$A_{21}\tilde{X}_1^{(k+1)}+\sum_{j=2}^{n}A_{2j}\tilde{X}_j^{(k)}+B_2\subseteq\sum_{j=1}^{n}A_{2j}X_j^{(k)}+B_2,$$

а ввиду

$$X_2^{(k+1)} = \left(\sum_{j=1}^{n} A_{2j} X_j^{(k)} + B_2 \right) \cap X_2^{(k)},$$

$$\tilde{X}_2^{(k+1)} = \left(A_{21} \tilde{X}_1^{(k+1)} + \sum_{j=2}^{n} A_{2j} \tilde{X}_j^{(k)} + B_2 \right) \cap \tilde{X}_2^{(k)}$$

получаем

$$\tilde{X}_2^{(k+1)} \subseteq X_2^{(k+1)}.$$

Таким же образом мы показываем, что $\tilde{X}_i^{(k+1)} \subseteq X_i^{(k+1)}$, $3 \leqslant i \leqslant n$, т. е. $\tilde{x}^{(k+1)} \subseteq x^{(k+1)}$, что завершает доказательство по индукции. Доказательство для последовательностей, вычисленных согласно итерационным методам (SI) и (SIC), можно провести аналогичным образом. ■

Теорема 4. *Пусть последовательности* $\{z^{(k)}\}_{k=0}^{\infty}$ *и* $\{x^{(k)}\}_{k=0}^{\infty}$ *вычислены согласно итерационным методам* (SIC) *и* (SSIC) *в предположении* $z^{(0)} \supseteq x^{(0)} \supseteq x^*$. *Тогда имеет место*

$$z^{(k)} \supseteq x^{(k)} \supseteq x^* \quad \textit{для всех} \quad k \geqslant 0.$$

Доказательство. Допустим, что для некоторого $k \geqslant 0$ верно $z^{(k)} \supseteq x^{(k)} \supseteq x^*$. Для $k = 0$ это верно в силу нашего исходного допущения. Из формул для (SSIC) и первой формулы для (SIC) получаем с помощью математической индукции по индексам компонент, что

$$z^{(k+1)} \supseteq x^{(k+1/2)} \supseteq x^*.$$

С помощью второй формулы для (SSIC) получаем, еще раз применяя индукцию по индексам компонент, что

$$x^{(k+1/2)} \supseteq x^{(k+1)} \supseteq x^*.$$

Сочетание этих включений дает

$$z^{(k+1)} \supseteq x^{(k+1)} \supseteq x^*. \quad \blacksquare$$

Используя доказанные выше утверждения, мы можем теперь указать оптимальный метод. Пусть (M) обозначает любой метод из множества

$$\{(\mathrm{T}), (\mathrm{S}), (\mathrm{TI}), (\mathrm{SI}), (\mathrm{SIC}), (\mathrm{SSIC})\}.$$

Мы допускаем в качестве начального вектора любой интервальный вектор $x^{(0)}$, такой что $x^* \subseteq x^{(0)}$, где x^* — неподвижная точка, т. е. решение уравнения $x = \mathcal{A}x + b$. Введем частичный порядок на рассматриваемом множестве итерационных методов, полагая $(\mathrm{M}) \leqslant (\tilde{\mathrm{M}})$, если $x^{(k)} \subseteq \tilde{x}^{(k)}$ для всех $k \geqslant 0$. Здесь

$\{x^{(k)}\}_{k=0}^{\infty}$ и $\{\tilde{x}^{(k)}\}_{k=0}^{\infty}$ обозначают последовательности, вычисленные согласно методам (M) и ($\tilde{\text{M}}$). Из теоремы 2 имеем

$$(\text{TI}) \leqslant (\text{T}) \quad \text{и} \quad (\text{SI}) \leqslant (\text{S}).$$

Аналогично из теоремы 3 имеем

$$(\text{SIC}) \leqslant (\text{TI}) \quad \text{и} \quad (\text{SIC}) \leqslant (\text{SI}).$$

Наконец, из теоремы 4 имеем, что

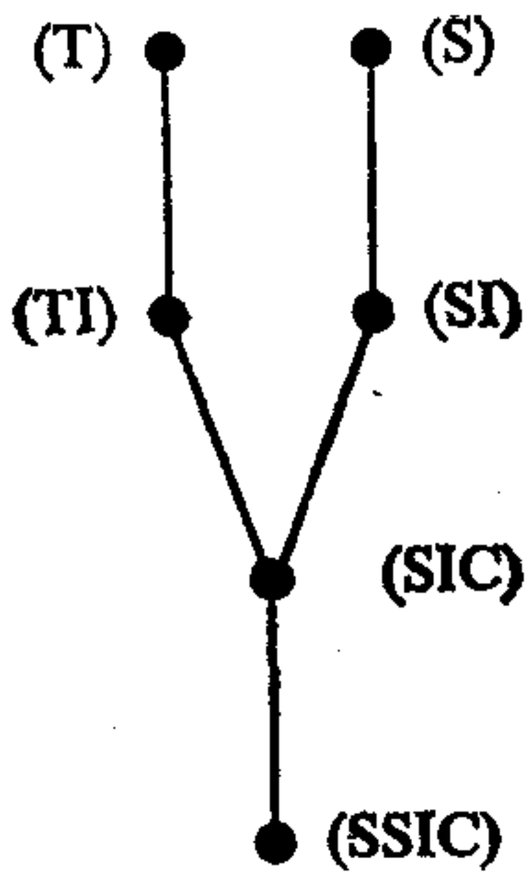
$$(\text{SSIC}) \leqslant (\text{M})$$

для любого из рассматриваемых итерационных методов (M).

Объединяем эти результаты в следующее утверждение.

Теорема 5. *Пусть $\mathscr{A}$ — интервальная матрица, такая что $\rho(|\mathscr{A}|) < 1$, и b — интервальный вектор. Если начать вычисление по одному из методов* (T), (S), (TI), (SI), (SIC) *и* (SSIC) *с вектора $x^{(0)}$, удовлетворяющего соотношению $x^{(0)} \supseteq x^* = \mathscr{A}x^* + b$, то метод* (SSIC) *всегда даст наименьшую (в смысле включения) локализацию последовательностей для x^*.* ■

Следующая диаграмма наглядно выражает содержание теоремы 5:



Чтобы проиллюстрировать теорему 5, мы просчитали различные примеры на машине UNIVAC 1108 (где мантисса содержит 8 десятичных цифр) вычислительного центра Университета в Карлсруэ.

Для каждого примера мы приводим начальный вектор $x^{(0)} = (X_i^{(0)})$ и число итераций k^*, после которого процедура стабилизируется. Примеры показывают, что метод (SIC) требует примерно на 25 % больше шагов итерации, чем метод (SSIC).

В первых двух примерах и исходные данные, и неподвижные точки — невырожденные интервалы. В этом случае приводим также и вектор $x^{(k^*)}$. В остальных примерах приводим только наибольшую ширину компоненты вектора $x^{(k)}$, т. е. величину

$d^{(k)} = \max\limits_{1 \leqslant i \leqslant n} \{d(X_i^{(k)})\}$. Все примеры были приведены к виду $x = \mathscr{A}x + b$ таким образом, что диагональные элементы матрицы $\mathscr{A}$ равны нулю.

I. **Пример.** [3, пример 1, стр. 356].

$$x^{(0)} = \begin{pmatrix} [0.9, & 1.2] \\ [0.4, & 0.7] \\ [0, & 0.2] \\ [-0.4, & -0.1] \end{pmatrix},$$

$$(\mathrm{SIC}): k^* = 37, \quad (\mathrm{SSIC}): k^* = 31,$$

$$x^{(k^*)} = \begin{pmatrix} [1.0328601, & 1.0597579] \\ [0.55075440, & 0.57481398] \\ [0.099483623, & 0.12251379] \\ [-0.24354582, & -0.21269841] \end{pmatrix}.$$

II. **Пример.** [3, пример 1, стр. 357].

$$x^{(0)} = \begin{pmatrix} [0.8, & 1.0] \\ [0.65, & 0.85] \\ [0.55. & 0.7] \end{pmatrix}, \qquad x^{(k^*)} = \begin{pmatrix} [0.89636817, & 0.89647991] \\ [0.76505755, & 0.76520225] \\ [0.61424734, & 0.61452184] \end{pmatrix},$$

$$(\mathrm{SIC}): k^* = 18, \quad (\mathrm{SSIC}): k^* = 15.$$

III. **Пример.** [512, пример 1, стр. 332].

$$x^{(0)} = (X_i^{(0)}), \quad \text{где} \quad X_i^{(0)} = [0, 1], \quad i = 1(1)8.$$

$$(\mathrm{SIC}): k^* = 27, \quad (\mathrm{SSIC}): k^* = 22.$$

$d^{(k)}$	k				
	0	5	10	15	20
(SIC)	1	4.0×10^{-2}	8.4×10^{-4}	1.8×10^{-5}	3.8×10^{-7}
(SSIC)	1	7.0×10^{-3}	5.4×10^{-5}	4.1×10^{-7}	3.7×10^{-9}

IV. **Пример.** [512, пример 2, стр. 333].

$$x^{(0)} = (X_i^{(0)}), \quad \text{где} \quad X_i^{(0)} = [0, 0.5], \quad 1 \leqslant i \leqslant 4.$$

$$(\mathrm{SIC}): k^* = 56, \quad (\mathrm{SSIC}): k^* = 44.$$

$d^{(k)}$	k					
	0	5	10	20	30	40
(SIC)	5.0×10^{-1}	1.1×10^{-1}	2.9×10^{-2}	6.7×10^{-4}	2.7×10^{-5}	1.5×10^{-6}
(SSIC)	5.0×10^{-1}	3.7×10^{-2}	4.7×10^{-3}	7.4×10^{-5}	1.2×10^{-6}	1.8×10^{-8}

V. Пример. [508, пример A, стр. 85].

$x^{(0)} = (X_i^{(0)})$, где $X_i^{(0)} = [-0.036016,\ 0.674056]$, $1 \leqslant i \leqslant 3$.

$$(\mathrm{SIC}): k^* = 52, \quad (\mathrm{SSIC}): k^* = 42.$$

$d^{(k)}$	k				
	0	5	10	20	40
(SIC)	7.1×10^{-1}	1.2×10^{-1}	2.0×10^{-2}	5.3×10^{-4}	3.7×10^{-7}
(SSIC)	7.1×10^{-1}	5.5×10^{-2}	5.7×10^{-3}	6.2×10^{-5}	3.7×10^{-9}

VI. **Пример.** Возьмем в примере V

$$X_i^{(0)} = [0.059459,\ 0.643243], \quad 1 \leqslant i \leqslant 3.$$

$$(\mathrm{SIC}): k^* = 51, \quad (\mathrm{SSIC}): k^* = 41.$$

$d^{(k)}$	k				
	0	5	10	20	40
(SIC)	5.8×10^{-1}	1.0×10^{-1}	1.7×10^{-2}	4.3×10^{-4}	3.0×10^{-7}
(SSIC)	5.8×10^{-1}	4.5×10^{-2}	4.7×10^{-3}	5.1×10^{-5}	3.7×10^{-9}

Использование локализующих множеств важно при реализации итерационных вычислений на ЭВМ. Если в этом случае мы начинаем вычисления с вектора, представимого в машине и содержащего неподвижную точку, т. е. решение уравнения $x^* = \mathscr{A}x^* + b$, то все следующие приближения снова содержат эту неподвижную точку. Так как вычисление новых приближений искажается погрешностями округления, мы можем в действительности в какой-то момент «потерять» неподвижную точку: некоторый вновь вычисленный интервал уже не будет содержать ее. Если все операции выполняются в машинной интервальной арифметике (см. гл. 4), то свойство содержать неподвижную точку не будет потеряно. Если мы применяем метод, где берутся пересечения, то получается последовательность

$$\tilde{x}^{(0)} \supseteq \tilde{x}^{(1)} \supseteq \ldots \tilde{x}^{(k-1)} \supseteq \tilde{x}^{(k)} = \tilde{x}^{(k+1)} = \ldots.$$

Последовательность приближений, вычисленных на машине, стабилизируется, начиная с некоторого номера k^*. Это следует из того факта, что на цифровой машине представимо лишь конечное количество вещественных чисел.

Сейчас мы покажем, что для методов (TI), (SI), (SIC) и (SSIC) после конечного числа шагов не нужно брать пересече-

ний. Мы сформулируем и докажем такую теорему для метода итераций (TI).

Теорема 6. *Пусть* $\mathscr{A}$ *— интервальная матрица, для которой* $\rho(|\mathscr{A}|)<1$. *Пусть вектор* $x^{(0)}=(X_i^{(0)})=([i(X_i^{(0)}),\ s(X_i^{(0)})])$ *выбран таким образом, что для неподвижной точки* $x^*=([i(X_i^*),\ s(X_i^*)])$, *т. е. решения уравнения* $x=\mathscr{A}x+b$ *выполнено включение* $x^{(0)}\supset x^*$, *которое вводится соотношениями* $i(X_i^{(0)})<i(X_i^*)\leqslant s(X_i^*)<s(X_i^{(0)})$, $i=1,\ 2,\ \ldots,\ n$. *Тогда существует* $\tilde{k}\geqslant 0$, *такое что при всех* $k\geqslant\tilde{k}$ *для итерационного метода*

$$x^{(k+1)}=(\mathscr{A}x^{(k)}+b)\cap x^{(k)},\qquad k\geqslant 0,$$

выполнено равенство

$$x^{(k+1)}=(\mathscr{A}x^{(k)}+b)\cap x^{(k)}=\mathscr{A}x^{(k)}+b,$$

т. е. верно включение

$$\mathscr{A}x^{(k)}+b\subseteq x^{(k)}.$$

Доказательство. Мы ограничимся случаем, когда все элементы матрицы $\mathscr{A}$ и векторов b, $x^{(0)}$ принадлежат $I(\mathbb{R})$. Случай, когда разрешены элементы из $R(\mathbb{C})$, может быть рассмотрен аналогичным образом. Сначала мы покажем, что из включения $x^{(0)}\supset x^*$ следует, что не может выполняться соотношение

$$x^{(0)}\subseteq\mathscr{A}x^{(0)}+b.$$

Действительно, если бы оно выполнялось, то из формул

$$z^{(k+1)}=\mathscr{A}z^{(k)}+b,\qquad k\geqslant 0,$$

определяющих наш метод итераций, следовало бы при $z^{(0)}=x^{(0)}$, что

$$z^{(1)}=\mathscr{A}z^{(0)}+b=\mathscr{A}x^{(0)}+b\supseteq x^{(0)}\supset x^*,$$
$$z^{(2)}=\mathscr{A}z^{(1)}+b\supseteq\mathscr{A}x^{(0)}+b\supseteq x^{(0)}\supset x^*$$

и вообще

$$z^{(k+1)}=\mathscr{A}z^{(k)}+b\supseteq\mathscr{A}x^{(0)}+b\supseteq x^{(0)}\supset x^*,\qquad k\geqslant 0.$$

Так как $\rho(|\mathscr{A}|)<1$, мы имели бы тогда $\lim\limits_{k\to\infty}z^{(k)}=z^*$, где $z^*=\mathscr{A}z^*+b$. Из последнего соотношения следует, что $z^*\supseteq\supseteq x^{(0)}\supset x^*$. Это противоречит единственности неподвижной точки, т. е. решения уравнения $x=\mathscr{A}x+b$.

Теперь полагаем

$$x^{(k)}=(X_i^{(k)}),\qquad\mathscr{A}=(A_{ij}),\qquad b=(B_i),\qquad y^{(k)}=(Y_i^{(k)}),$$

где

$$Y_i^{(k+1)} = \sum_{j=1}^{n} A_{ij} X_j^{(k)} + B_i, \qquad k \geqslant 0, \qquad 1 \leqslant i \leqslant n.$$

Из только что установленного факта следует, что найдется номер i, $1 \leqslant i \leqslant n$, такой что имеет место в точности одна из следующих двух возможностей:

(a) $Y_i^{(1)} \subset X_i^{(0)}$, т. е. $X_i^{(1)} = Y_i^{(1)} \cap X_i^{(0)} = Y_i^{(1)}$;

(b) $Y_i^{(1)} \not\subset X_i^{(0)}$ и $X_i^{(1)} = Y_i^{(1)} \cap X_i^{(0)} \subset X_i^{(0)}$.

В случае (a) мы получаем ввиду $x^{(0)} \supseteq x^{(1)}$ и монотонности включения, что

$$Y_i^{(2)} = \sum_{j=1}^{n} A_{ij} X_j^{(1)} + B_i \subseteq \sum_{j=1}^{n} A_{ij} X_j^{(0)} + B_i = Y_i^{(1)} = X_i^{(1)},$$

$$X_i^{(2)} = Y_i^{(2)} \cap X_i^{(1)} = Y_i^{(2)},$$

а в общем случае ввиду $x^{(k)} \supseteq x^{(k+1)}$ методом математической индукции получаем

$$Y_i^{(k+1)} = \sum_{j=1}^{n} A_{ij} X_j^{(k)} + B_i \subseteq \sum_{j=1}^{n} A_{ij} X_j^{(k-1)} + B_i = Y_i^{(k)} = X_i^{(k)},$$

$$X_i^{(k+1)} = Y_i^{(k+1)} \cap X_i^{(k)} = Y_i^{(k+1)}.$$

(b). Полагая $i(A) = a_1, s(A) = a_2$ для $A = [a_1, a_2] \in I(\mathbb{R})$, мы можем, не умаляя общности, считать, что

$$i(Y_i^{(1)}) < i(X_i^{(0)}) \leqslant s(Y_i^{(1)}) < s(X_i^{(0)}),$$

т. е.

$$X_i^{(1)} = [i(X_i^{(0)}),\ s(Y_i^{(1)})].$$

(Возможен еще случай

$$i(X_i^{(0)}) < i(Y_i^{(1)}) \leqslant s(X_i^{(0)}) < s(Y_i^{(1)}),$$

т. е.

$$X_i^{(1)} = [i(Y_i^{(1)}),\ s(X_i^{(0)})],$$

но он рассматривается аналогично.) Так как $x^{(0)} \supseteq x^{(1)}$, мы имеем

$$Y_i^{(2)} = \sum_{j=1}^{n} A_{ij} X_j^{(1)} + B_i \subseteq \sum_{j=1}^{n} A_{ij} X_j^{(0)} + B_i = Y_i^{(1)},$$

т. е.

$$i\ (Y_i^{(1)}) \leqslant i(Y_i^{(2)}),\ s(Y_i^{(2)}) \leqslant s(Y_i^{(1)}) = s(X_i^{(1)}),$$

а также

$$X_i^{(2)} = [\max\{i(Y_i^{(2)}),\ i(X_i^{(0)})\},\ s(Y_i^{(2)})].$$

Так как $x^{(k)} \supseteq x^{(k+1)}$, то можно показать методом математической индукции, что

$$i\left(Y_i^{(k)}\right) \leqslant i\left(Y_i^{(k+1)}\right),\ s\left(Y_i^{(k+1)}\right) \leqslant s\left(Y_i^{(k)}\right) = s\left(X_i^{(k)}\right), \qquad k \geqslant 1,$$

т. е.

$$X_i^{(k+1)} = \left[\max\left\{i\left(Y_i^{(k+1)}\right),\ i\left(X_i^{(0)}\right)\right\},\ s\left(Y_i^{(k+1)}\right)\right].$$

По предположению мы имеем $i(X^{(0)}) < i\left(X_i^*\right)$, а по теореме 1 $\lim\limits_{k\to\infty} i\left(X_i^{(k)}\right) = i\left(X_i^*\right)$. Поэтому найдется $k_0 \geqslant 1$, такое что имеет место

$$\max\left\{i\left(Y_i^{(k_0+1)}\right),\ i\left(X_i^{(0)}\right)\right\} = i\left(Y_i^{(k_0+1)}\right)$$

или

$$X_i^{(k_0+1)} = \left[i\left(Y_i^{(k_0+1)}\right),\ s\left(Y_i^{(k_0+1)}\right)\right],$$

т. е.

$$X_i^{(k_0+1)} = Y_i^{(k_0+1)} \cap X_i^{(k_0)} = Y_i^{(k_0+1)}.$$

Метод математической индукции позволяет теперь установить, что

$$X_i^{(k_0+\nu)} = Y_i^{(k_0+\nu)} \cap X_i^{(k_0+\nu-1)} = Y_i^{(k_0+\nu)}, \qquad \nu \geqslant 1,$$

так как $x^{(k)} \supseteq x^{(k+1)}$. Ввиду соотношений $\lim\limits_{k\to\infty} x^{(k)} = x^*$ и $x^{(0)} \supset x$ получаем, что для любого i, $1 \leqslant i \leqslant n$, не удовлетворяющего ни одному из условий (a), (b), хотя бы одно из этих условий выполнится после нескольких следующих шагов итерации

Наконец, получаем

$$x^{(k+1)} = (\mathscr{A}x^{(k)} + b) \cap x^{(k)} = \mathscr{A}x^{(k)} + b, \qquad k \geqslant \tilde{k} \geqslant 0. \ \blacksquare$$

Замечания. Бóльшая часть результатов этой главы доказана в статьях, указанных в списке литературы. Соответствующие результаты приводятся также в статьях, где рассматриваются методы релаксации (см., в частности, [4] и [201]). Аналогичные результаты для систем так называемых полулинейных уравнений приведены в [10]. Теорема 5 была первоначально доказана в [348].

Теоремы этой главы без всяких изменений переносятся на соответствующие итерационные методы нахождения неподвижной точки, т. е. решения нелинейного уравнения

$$x = \mathscr{f}_{\mathrm{p}}(x),$$

где

$$\mathscr{f}_{\mathrm{p}}(x_{\mathrm{p}}) = \left(f_i\left(x_1, \ldots, x_n; a_{i1}, \ldots, a_{im_i}\right)\right)$$

есть $\mathscr{P}_{\mathrm{p}}$-сжатие. Самое существенное для доказательства этих теорем свойство — монотонность включения.

Аналогичные теоремы для комплексных точечных систем уравнений с использованием круговой арифметики и подходя-

щей модификацией процесса взятия пересечений были приведены в [247].

* * *

Литература: [3], [4], [10], [195], [197], [198], [199], [200], [201], [247], [348], [455], [508], [512].

Глава 15
О ПРИМЕНИМОСТИ МЕТОДА ГАУССА К СИСТЕМАМ УРАВНЕНИЙ С ИНТЕРВАЛЬНЫМИ КОЭФФИЦИЕНТАМИ

Пусть $\mathscr{A}$ — интервальная матрица, $\mathscr{b}$ — интервальный вектор. Будем предполагать, что обращение $\mathscr{A}_p^{-1}$ существует для всех $\mathscr{A}_p \in \mathscr{A}$. Мы хотим найти множество

$$\mathfrak{L} = \{x_p \mid \mathscr{A}_p x_p = \mathscr{b}_p,\ \mathscr{A}_p \in \mathscr{A},\ \mathscr{b}_p \in \mathscr{b}\}.$$

Это множество в общем случае не имеет простого описания. Поэтому мы ограничимся его локализацией с помощью интервального вектора. Очевидный способ нахождения такого интервального вектора — применение непосредственного обобщения метода Гаусса на системы с интервальными коэффициентами. Иными словами, пусть нам дана таблица коэффициентов

$$\begin{array}{cccc} A_{11} & \dots & A_{1n} & B_1 \\ \vdots & & \vdots & \vdots \\ A_{n1} & \dots & A_{nn} & B_n. \end{array}$$

Применяя формулы

$$\begin{aligned} A'_{1j} &= A_{1j}, && 1 \leqslant j \leqslant n, && B'_1 = B_1, \\ A'_{ij} &= A_{ij} - A_{1j}(A_{i1}/A_{11}), && 2 \leqslant i,\ j \leqslant n, \\ B'_i &= B_i - B_1(A_{i1}/A_{11}), && 2 \leqslant i \leqslant n, \\ A'_{i1} &= 0, && 2 \leqslant i \leqslant n, \end{aligned}$$

в предположении $0 \notin A_{11}$, мы вычисляем новую таблицу коэффициентов

$$\begin{array}{cccc} A'_{11} & A'_{12} & \dots & A'_{1n} & B'_1 \\ 0 & A'_{22} & \dots & A'_{2n} & B'_2 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & A'_{n2} & \dots & A'_{nn} & B'_n. \end{array}$$

Покажем теперь, что имеет место

$$\{x_\text{р} \mid \mathscr{A}_\text{р} x_\text{р} = b_\text{р},\ \mathscr{A}_\text{р} \in \mathscr{A},\ b_\text{р} \in b\} \subseteq \{y_\text{р} \mid \mathscr{A}'_\text{р} y_\text{р} = b'_\text{р},$$
$$\mathscr{A}'_\text{р} \in \mathscr{A}',\ b'_\text{р} \in b'\}.$$

Допустим, что $\mathscr{A}_\text{р} = \mathscr{A}$ и $b_\text{р} \in b$, и рассмотрим систему линейных уравнений

$$\mathscr{A}_\text{р} x_p = b_\text{р}.$$

Мы строим матрицу $\mathscr{A}'_\text{р} = (a'_{ij})$ и вектор $b'_\text{р} = (b'_i)$, где

$$a'_{1j} = a_{1j}, \qquad 1 \leqslant j \leqslant n, \qquad b'_1 = b_1,$$

и

$$a'_{ij} = a_{ij} - a_{1i}\,(a_{j1}/a_{11}), \qquad 2 \leqslant i,\ j \leqslant n,$$
$$b'_i = b_i - b_1\,(a_{i1}/a_{11}), \qquad 2 \leqslant i \leqslant n,$$
$$a'_{i1} = 0, \qquad 2 \leqslant i \leqslant n.$$

Из теории линейных уравнений хорошо известно, что система уравнений $\mathscr{A}'_\text{р} y_\text{р} = b'_\text{р}$ имеет те же решения, что и $\mathscr{A}_\text{р} x_\text{р} = b_\text{р}$. Из монотонности включения следует $\mathscr{A}'_\text{р} \in \mathscr{A}'$ и $b'_\text{р} \in b'$, что и доказывает наше утверждение.

Если описанный выше шаг проведен $n - 1$ раз, то исходная таблица коэффициентов превращается в верхнюю треугольную матрицу

$$\begin{matrix} \tilde{A}_{11} & \tilde{A}_{12} & \dots & \tilde{A}_{1n} & \tilde{B}_1 \\ & \tilde{A}_{22} & & . & . \\ & & & . & . \\ & & & . & . \\ & & & \tilde{A}_{nn} & \tilde{B}_n \end{matrix},$$

для которой имеет место

$$\{x_\text{р} \mid \mathscr{A}_\text{р} x_\text{р} = b_\text{р},\ \mathscr{A}_\text{р} \in \mathscr{A},\ b_\text{р} \in b\}$$
$$\subseteq \{\tilde{x}_\text{р} \mid \tilde{\mathscr{A}}_\text{р} \tilde{x}_\text{р} = \tilde{b}_\text{р}, \tilde{\mathscr{A}}_\text{р} \in \tilde{\mathscr{A}},\ \tilde{b}_\text{р} \in \tilde{b}\}.$$

Используя формулы

$$X_n = \tilde{B}_n / \tilde{A}_{nn},$$

$$X_i = \left(\tilde{B}_i - \sum_{j=i+1}^{n} \tilde{A}_{ij} X_j\right) \Big/ \tilde{A}_{ii}, \qquad 1 \leqslant i \leqslant n-1,$$

мы получаем тогда интервальный вектор $x = (X_i)$, удовлетворяющий условию

$$\{x_\text{р} \mid \mathscr{A}_\text{р} x_\text{р} = b_\text{р},\ \mathscr{A}_\text{р} \in \mathscr{A},\ b_\text{р} \in b\} \subseteq x.$$

В частности, если $\mathscr{A}_{\mathrm{p}} = (a_{ij})$— невырожденная точечная матрица, то метод Гаусса применим, когда в правой части стоит произвольный интервальный вектор. При этом в процессе исключения по Гауссу может потребоваться перестановка столбцов. Это эквивалентно умножению матрицы $\mathscr{A}_{\mathrm{p}}$ слева на матрицу перестановки перед началом процесса исключения.

Теперь мы определим отображение

$$g_{\mathrm{p}}\colon\ M_{nn}(\mathbb{C}) \times V_n(\mathbb{C}) \to V_n(\mathbb{C})$$

для невырожденной матрицы $\mathscr{A}_{\mathrm{p}}$ и точечного вектора b_{p}. Это отображение представляет собой применение метода Гаусса к системе линейных уравнений

$$\mathscr{A}_{\mathrm{p}} x_{\mathrm{p}} = b_{\mathrm{p}},$$

дающее результат

$$x_{\mathrm{p}} = g_{\mathrm{p}}(\mathscr{A}_{\mathrm{p}},\ b_{\mathrm{p}}).$$

Отображение g_{p} единственно, но, как обычно, для g_{p} имеются различные выражения. Например, мы имеем $\mathscr{A}_{\mathrm{p}}^{-1} b_{\mathrm{p}} = g_{\mathrm{p}}(\mathscr{A}_{\mathrm{p}},\ b_{\mathrm{p}})$. Кроме того, метод Гаусса дает различные выражения для g_{p} в зависимости от выбора главных элементов.

Следующие свойства не зависят от выбора главных элементов. Интервальное выражение для g_{p} обозначается через $g_{\mathrm{p}}(\mathscr{A},\ b)$. Поэтому интервальный вектор x, получаемый после выполнения описанного выше метода Гаусса, можно задать равенством $x = g_{\mathrm{p}}(\mathscr{A},\ b)$.

Имеем следующие свойства:

(1) $$\mathscr{A},\ \mathscr{B} \in M_{nn}(I(\mathbb{C})), \quad a,\ b \in V_n(I(\mathbb{C})),$$
$$\mathscr{A} \subseteq \mathscr{B}, \quad a \subseteq b.$$

Отсюда следует, что

(2) $$g_{\mathrm{p}}(\mathscr{A},\ a) \subseteq g_{\mathrm{p}}(\mathscr{B},\ b).$$
$$\mathscr{A}_{\mathrm{p}} \in M_{nn}(\mathbb{C}), \quad b = u + v \in V_n(I(\mathbb{C})).$$

Отсюда следует, что

(3) $$g_{\mathrm{p}}(\mathscr{A}_{\mathrm{p}},\ b) = g_{\mathrm{p}}(\mathscr{A}_{\mathrm{p}},\ u) + g_{\mathrm{p}}(\mathscr{A}_{\mathrm{p}},\ v).$$
$$\mathscr{A}_{\mathrm{p}} \in M_{nn}(\mathbb{R}), \quad b \in V_n(I(\mathbb{R})).$$

Отсюда следует, что

$$\mathscr{A}_{\mathrm{p}}^{-1} b \subseteq g_{\mathrm{p}}(\mathscr{A}_{\mathrm{p}},\ b).$$

(4) $$\mathscr{A}_{\mathrm{p}} \in M_{nn}(\mathbb{C}), \quad a,\ b \in V_n(I(\mathbb{C})), \quad d(a) \leqslant \alpha d(b)$$
для некоторого $\alpha \geqslant 0$.

Поэтому для ширины имеет место

$$d(g_{\mathrm{p}}(\mathscr{A}_{\mathrm{p}},\ a)) \leqslant \alpha d(g_{\mathrm{p}}(\mathscr{A}_{\mathrm{p}},\ b)).$$

По поводу доказательства этих свойств заметим, что (1) сразу следует из монотонности включения, а (2)— из соотношения $a(B+C)=aB+aC$ и формул, определяющих метод Гаусса. Чтобы доказать (3), используем следующий факт (см. также обсуждение в гл. 3).

Если имеются два рациональных выражения f_1 и f_2 для одной и той же функции f: $\mathbb{R}\to\mathbb{R}$, причем f_1 содержит переменную x ровно один раз, а f_2 содержит эту переменную m раз, то для вычислений f_1 и f_2 в интервальной арифметике имеет место $f_1(X)\subseteq f_2(X)$. Аналогичное утверждение верно и для функций от нескольких переменных. Рассмотрим теперь i-е, $1\leqslant i\leqslant n$, компоненты векторов $\mathscr{A}_{\mathrm{p}}^{-1}b$ и $g_{\mathrm{p}}(\mathscr{A}_{\mathrm{p}}, b)$. Для точечных векторов b_{p} имеет место $\left(\mathscr{A}_{\mathrm{p}}^{-1}b_{\mathrm{p}}\right)_i=\left(g_{\mathrm{p}}(\mathscr{A}_{\mathrm{p}}, b_{\mathrm{p}})\right)_i$. Из формул, определяющих метод Гаусса, видно, что компоненты вектора b_{p} могут входить несколько раз в выражение $(g_{\mathrm{p}}(\mathscr{A}_{\mathrm{p}}, b_{\mathrm{p}}))_i$. Так как они входят всего один раз в $\left(\mathscr{A}_{\mathrm{p}}^{-1}b_{\mathrm{p}}\right)_i$, мы получаем (3).

Наконец, (4) получается путем использования формул, описывающих метод Гаусса, правил (2.10), (2.14), (6.12) и (6.16), а также предположения $d(a_{\mathrm{p}})\leqslant\alpha d(b_{\mathrm{p}})$.

Формулы, описывающие метод Гаусса, применимы только к таким интервальным матрицам, для которых условие $0\notin A_{ii}$ выполняется на всех шагах приведения к верхней треугольной форме. Если это не так при некотором i, $1\leqslant i\leqslant n-1$, т. е. $0\in A_{ii}$, то все еще возможно, что подходящая перестановка столбцов позволит избежать соотношения $0\in A_{ii}$. Однако это не всегда возможно, даже если предположить, что сначала $\mathscr{A}_{\mathrm{p}}^{-1}$ существует для всех $\mathscr{A}_{\mathrm{p}}\in\mathscr{A}$. Причина заключается в следующем.

Очевидно, что в предположении существования $\mathscr{A}_{\mathrm{p}}^{-1}$ для всех $\mathscr{A}_{\mathrm{p}}\in\mathscr{A}$ мы всегда можем выполнить первый шаг метода Гаусса. Если бы мы не смогли сделать этот шаг из-за того, что все элементы первого столбца содержат нуль, то исходная интервальная матрица $\mathscr{A}$ содержала бы вырожденную точечную матрицу $\mathscr{A}_{\mathrm{p}}$, а это противоречит нашему предположению. Рассмотрим теперь матрицу $\mathscr{A}^{(1)}=(A_{ij}^{(1)})$ размерности $(n-1)\times(n-1)$, для которой $A_{ij}^{(1)}=A'_{ij}$, $2\leqslant i,\ j\leqslant n$. Пусть $\mathscr{U}$ обозначает интервальную матрицу

$$\mathscr{U}=\begin{pmatrix}-A_{21}/A_{11} & 1 & & 0\\ -A_{31}/A_{11} & 0 & 1 & \\ \vdots & & & \ddots\\ -A_{n1}/A_{11} & 0 & \cdots & 0\ \ 1\end{pmatrix},$$

размерности $(n-1)\times n$, а $\mathscr{V}$ — интервальную матрицу

$$\mathscr{V}=\begin{pmatrix} A_{12} & \cdots & A_{1n} \\ \vdots & & \vdots \\ A_{n2} & \cdots & A_{nn} \end{pmatrix}.$$

размерности $n\times(n-1)$. Мы имеем $\mathscr{A}^{(1)}=\mathscr{U}\mathscr{V}$. Аналогичным образом положим

$$\mathscr{R}_{\mathrm{p}}=\begin{pmatrix} -a_{21}/a_{11} & 1 & & 0 \\ -a_{31}/a_{11} & 0 & 1 & \\ \vdots & & & \ddots \\ -a_{n1}/a_{11} & 0 & \cdots & 0 \quad 1 \end{pmatrix},$$

$$\mathscr{S}_{\mathrm{p}}=\begin{pmatrix} a_{12} & \cdots & a_{1n} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ a_{n2} & \cdots & a_{nn} \end{pmatrix},$$

где $a_{ij}\in A_{ij}$, $1\leqslant i,\ j\leqslant n$. Мы получаем, что матрица $\mathscr{Z}_{\mathrm{p}}=\mathscr{R}_{\mathrm{p}}\mathscr{S}_{\mathrm{p}}$ невырожденная, так как $\mathscr{A}_{\mathrm{p}}\in\mathscr{A}$ невырожденная. Поэтому в первом столбце матрицы $\mathscr{Z}_{\mathrm{p}}$ найдется хотя бы один ненулевой элемент. Из монотонности включения следует, что

$$\{\mathscr{Z}_{\mathrm{p}}=\mathscr{R}_{\mathrm{p}}\mathscr{S}_{\mathrm{p}}\mid\mathscr{A}_{\mathrm{p}}\in\mathscr{A}\}\subseteq\{\mathscr{A}^{(1)}_{\mathrm{p}}\mid\mathscr{A}^{(1)}_{\mathrm{p}}\in\mathscr{A}^{(1)}=\mathscr{U}\mathscr{V}\},$$

и в общем случае здесь нет равенства, как показывает простой пример из гл. 10. Поэтому нет гарантии, что в первом столбце матрицы $\mathscr{A}^{(1)}$ будет хотя бы один элемент, не содержащий нуля. Те же рассуждения проходят и для следующих шагов метода. Мы покажем это с помощью следующего простого примера.

Пример. Рассмотрим интервальную матрицу

$$\mathscr{A}=\begin{pmatrix} [1,\ 5]+i[-1,\ 1] & 1 \\ 25 & [-1,\ 1]+i[-1,\ 1] \end{pmatrix}.$$

Покажем сначала, что $\mathscr{A}$ не содержит вырожденной точечной матрицы. Так как

$$\det\mathscr{A}_{\mathrm{p}}=a_{11}a_{22}-a_{21}a_{12}\in A_{11}A_{22}-A_{21}A_{12},$$

мы получаем для любой точечной матрицы $\mathscr{A}_{\mathrm{p}}\in\mathscr{A}$, что

$$\det\mathscr{A}_{\mathrm{p}}\in[-31,\ -19]+i[-6,\ 6].$$

Из того что интервал в правой части не содержит нуля, следует, что $\mathscr{A}_{\mathrm{p}}$ невырожденная.

Теперь шаг исключения согласно методу Гаусса порождает (в силу определения 5.3 и следующего за ним замечания) ин-

тервал

$$A'_{22} = \left[-126, \frac{1}{26}\right] + i\,[-26, 26].$$

Поэтому невозможно продолжить применение метода Гаусса. Даже если мы отделим в матрице $\mathscr{A}$ вещественную и мнимую части, нам все равно не удастся решить получившуюся систему из 4 интервальных уравнений с 4 неизвестными с помощью метода Гаусса, хотя исходная интервальная матрица снова не содержит вырожденных точечных матриц.

Мы покажем теперь, что метод Гаусса всегда можно выполнить при $n \leqslant 2$, если коэффициенты — вещественные интервалы и все точечные матрицы $\mathscr{A}_p \in \mathscr{A}$ невырожденные.

Теорема 1. *Пусть* $1 \leqslant n \leqslant 2$ *и интервальная матрица* $\mathscr{A}_p = (A_{ij})$ *размерности* $n \times n$ *не содержит невырожденных матриц* $\mathscr{A}_p$. *Тогда можно выполнить метод Гаусса.*

Доказательство. При $n = 1$ наше предположение означает, что $0 \notin A_{11}$, и в этом случае наша теорема доказана для $\mathscr{A} = (A_{11})$. При $n = 2$ хотя бы один из интервалов A_{11}, A_{21} не содержит нуля. Если бы это было не так, то существовала бы вырожденная матрица $\mathscr{A}_p \in \mathscr{A}$, что противоречит условию теоремы. Не умаляя общности, считаем, что $0 \notin A_{11}$; если это не так, переставим строки матрицы $\mathscr{A}$. Теперь метод Гаусса дает

$$A'_{22} = A_{22} - (1/A_{11})\,A_{21}A_{12}.$$

Мы можем рассматривать A'_{22} как оценивание рациональной функции a'_{22} от четырех переменных a_{11}, a_{21}, a_{12} и a_{22} в интервальной арифметике (см. гл. 3), определяемое формулой

$$a'_{22}(a_{11}, a_{12}, a_{21}, a_{22}) = a_{22} - (1/a_{11})\,a_{21}a_{12},$$

По условию теоремы мы имеем для любой $\mathscr{A}_p \in \mathscr{A}$ соотношение

$$\det(\mathscr{A}_p) = a_{11}a_{22} - a_{21}a_{12} \neq 0,$$

откуда

$$a'_{22}(a_{11}, a_{12}, a_{21}, a_{22}) = (1/a_{11}) \det(\mathscr{A}_p) \neq 0.$$

Приведенное интервальное оценивание даст точную локализацию, если заменить a_{11} на A_{11}, a_{12} на A_{12}, a_{21} на A_{21} и a_{22} на A_{22}, так как каждая переменная входит в выражение для a'_{22} всего один раз (см. также рассуждения, предшествующие теореме 3.2). Поэтому мы имеем $0 \notin A'_{22}$, что и означает возможность выполнения метода Гаусса. ■

Данное выше доказательство не обобщается на случай $n \geqslant 3$. Даже для $n = 2$ теорема будет неверна, если элементы интервальных матриц берутся из $R(\mathbb{C})$ или из $K(\mathbb{C})$.

Рассмотрим теперь один конкретный класс интервальных матриц, для которого метод Гаусса всегда может быть выполнен. В этом классе можно даже не переставлять строки или столбцы. В дальнейшем мы ограничимся системами уравнений, где элементы матрицы коэффициентов и правой части принадлежат множеству вещественных интервалов или комплексных круговых интервалов. Системы уравнений, элементами которых являются комплексные прямоугольные интервалы, можно свести к первому из упомянутых случаев, отделяя вещественную и мнимую части.

Чтобы объединить приводимые ниже доказательства для вещественных интервалов и комплексных круговых интервалов, мы заметим, что вещественный интервал вида $A = [a_1, a_2]$ можно представить в виде

$$A = [a - r, a + r],$$

где

$$a = \frac{1}{2}(a_1 + a_2), \qquad r = \frac{1}{2} d(A) = \frac{1}{2}(a_2 - a_1).$$

Здесь a — центр интервала, а r — радиус, т. е. половина ширины. Мы вводим обозначение

$$A = [a - r, a + r] =: \langle a, r \rangle,$$

чтобы подчеркнуть аналогию с комплексными круговыми интервалами. Арифметические операции на вещественных интервалах также можно определить через центр и ширину. Пусть $A = \langle a, r \rangle$, $B = \langle b, s \rangle$. Тогда сложение и вычитание интервалов A и B можно записать в виде

$$A \pm B = \langle a \pm b, r + s \rangle.$$

Формально это соответствует сложению и вычитанию комплексных круговых интервалов. Для умножения нам нужно только равенство

$$[-r, r][-s, s] = \langle 0, r \rangle \langle 0, s \rangle = \langle 0, rs \rangle.$$

Если мы заметим теперь, что при $0 \notin A = [a_1, a_2]$ имеет место представление

$$\frac{1}{A} = \left[\frac{1}{a + r}, \frac{1}{a - r}\right] = \left[\frac{a}{a^2 - r^2} - \frac{r}{a^2 - r^2}, \frac{a}{a^2 - r^2} + \frac{r}{a^2 - r^2}\right],$$

то получим

$$\frac{1}{A} = \left\langle \frac{a}{a^2 - r^2}, \frac{r}{a^2 - r^2} \right\rangle.$$

Формально это соответствует обращению кругового комплексного интервала. Для представления вещественных интервалов в виде $A = \langle a, r \rangle$ верно также

$$|A| = \max\{|a_1|, |a_2|\} = |a| + r.$$

Кроме того, имеем

$$0 \notin A \Leftrightarrow |a| - r > 0.$$

Наконец, имеем

$$A = \langle a, r \rangle \subseteq \langle 0, |A| \rangle = \langle 0, |a| + r \rangle.$$

Докажем сначала следующее утверждение.

Лемма 2. *Пусть* $A = \langle a, r_1 \rangle$, $B = \langle b, r_2 \rangle$, $C = \langle c, r_3 \rangle$ и $D = \langle d, r_4 \rangle$ — *вещественные интервалы или круговые комплексные интервалы, причем* $0 \notin D$. *Тогда для*

$$Z = \langle z, r_5 \rangle = A - (1/D)\,BC$$

справедливо неравенство

$$|a| - r_1 - |B||C|/(|d| - r_4) \leqslant |z| - r_5.$$

Доказательство. Из монотонности включения следует, что

$$Z = \langle z, r_5 \rangle = A - BC\frac{1}{D} \subseteq A - \langle 0, |B| \rangle \langle 0, |C| \rangle \left\langle \frac{\bar{d}}{d\bar{d} - r_4^2}, \frac{r_4}{d\bar{d} - r_4^2} \right\rangle$$

$$= \langle a, r_1 \rangle - \left\langle 0, |B||C| \frac{|d|}{|d\bar{d} - r_4^2|} + |B||C| \frac{r_4}{d\bar{d} - r_4^2} \right\rangle$$

$$= \langle a, r_1 \rangle - \left\langle 0, |B||C| \frac{1}{|d| - r_4} \right\rangle$$

$$= \left\langle a, r_1 + |B||C| \frac{1}{|d| - r_4} \right\rangle =: \langle a, r_6 \rangle.$$

Отсюда и из $Z \subseteq \langle a, r_6 \rangle$ следует, что

$$|a| - |z| \leqslant |a - z| \leqslant r_6 - r_5$$

или

$$|z| - r_5 \geqslant |a| - r_1 - |B||C|(1/(|d| - r_4)). \blacksquare$$

Чтобы сформулировать следующее утверждение, нам нужно понятие M-матрицы, введенное в [467a]. Мы применим здесь эквивалентное определение из [325]. Вещественная матрица $\mathscr{B}_p = (b_{ij})$ называется M-матрицей, если выполнены условия

(1) $$b_{ij} \leqslant 0, \quad i \neq j,$$

(2) $$\mathscr{B}_p^{-1} \geqslant \mathscr{O}_p.$$

В (2) подразумевается покомпонентный частичный порядок. Условие (2) можно заменить следующим условием (см. [385а]):

(2′) Существует вещественный вектор $u_{\mathrm{p}} = (u_i)$, такой что $u_i > 0$, $1 \leqslant i \leqslant n$ и $\mathscr{B}_{\mathrm{p}} u_{\mathrm{p}} > o_{\mathrm{p}}$.

Этот факт, а также то обстоятельство, что диагональные элементы M-матрицы положительны, используется в следующем утверждении.

Теорема 3. *Пусть* $\mathscr{A} = (A_{ij})$ *— интервальная матрица, причем* $A_{ij} = \langle a_{ij}, r_{ij} \rangle$, $1 \leqslant i, j \leqslant n$, *и пусть* $\mathscr{B}_{\mathrm{p}} = (b_{ij})$ *— вещественная матрица, определенная соотношением*

$$b_{ij} = \begin{cases} |a_{ii}| - r_{ii}, & i = j, \\ -|A_{ij}| & \text{в противном случае.} \end{cases}$$

Если $\mathscr{B}_{\mathrm{p}}$ *является M-матрицей, то для матрицы* $\mathscr{A}_{\mathrm{p}}$ *можно выполнить алгоритм Гаусса без перестановки строк или столбцов.*

Доказательство. Предположение, что $\mathscr{B}_{\mathrm{p}}$ является M-матрицей, означает существование вектора $u_{\mathrm{p}} = (u_i)$ с положительными элементами, такого что $\mathscr{B}_{\mathrm{p}} u_{\mathrm{p}} > o_{\mathrm{p}}$, т. е. верно

$$(|a_{ii}| - r_{ii}) u_i > \sum_{j=1,\ j \neq i}^{n} |A_{ij}| u_j, \quad 1 \leqslant i \leqslant n.$$

Ввиду неравенства $|a_{11}| - r_{11} > 0$ выполнено условие $0 \notin A_{11}$ и применимы формулы из первой части этой главы. Теперь мы покажем, что условия нашей теоремы выполнены и для интервальной матрицы $\tilde{A}' = (\tilde{A}'_{ij})$ размерности $(n-1) \times (n-1)$, где $\tilde{A}'_{ij} = A'_{ij} = \langle a'_{ij}, r'_{ij} \rangle$, $2 \leqslant i, j \leqslant n$. Это позволит немедленно завершить доказательство теоремы с помощью математической индукции.

Для $i \geqslant 2$ имеет место

$$\sum_{j=2,\ j \neq i}^{n} |A'_{ij}| u_j = \sum_{j=2,\ j \neq i}^{n} \left| A_{ij} - A_{1j} \frac{A_{i1}}{A_{11}} \right| u_j$$

$$\leqslant \sum_{j=2,\ j \neq i}^{n} |A_{ij}| u_j + |A_{i1}| \left| \frac{1}{A_{11}} \right| \sum_{j=2,\ j \neq i}^{n} |A_{1j}| u_j.$$

С помощью неравенства

$$\sum_{j=2,\ j \neq i}^{n} |A_{1j}| u_j < (|a_{11}| - r_{11}) u_1 - |A_{1i}| u_i,$$

справедливого в силу условия теоремы, можно получить оценки

$$\sum_{j=2,\ j\neq i}^{n} |A'_{ij}|u_j \leqslant \sum_{j=2,\ j\neq i}^{n} |A_{ij}|u_j$$

$$+|A_{i1}|\frac{1}{|a_{11}|-r_{11}}\{(|a_{11}|-r_{11})u_1-|A_{1i}|u_i\}$$

$$=\sum_{j=1,\ j\neq i}^{n} |A_{ij}|u_j-\frac{|A_{i1}||A_{1i}|}{|a_{11}|-r_{11}}u_i$$

$$<u_i\left(|a_{ii}|-r_{ii}-\frac{|A_{i1}||A_{1i}|}{|a_{11}|-r_{11}}\right)\leqslant(|a'_{ii}|-r'_{ii})u_i,$$

где последнее неравенство получается по лемме 2. Это завершает доказательство. ■

Следующее определение вводит важный класс интервальных матриц, удовлетворяющих условиям теоремы 3.

Определение 4. Говорят, что интервальная матрица $\mathcal{A}=(A_{ij})$, компоненты которой $A_{ij}=\langle a_{ij}, r_{ij}\rangle$ являются вещественными интервалами или круговыми комплексными интервалами, имеет сильно доминирующую диагональ (или что ее диагональ сильно доминирует), если

$$|a_{ii}|-r_{ii}>\sum_{j=1,\ j\neq i}^{n}|A_{ij}|, \quad 1\leqslant i\leqslant n. \blacksquare$$

Очевидно, что элементы сильно доминирующей диагонали не содержат нулей и что для любой точечной матрицы $\hat{\mathcal{A}}_p=(\hat{a}_{ij})\in\mathcal{A}$ выполнено соотношение

$$|\hat{a}_{ii}|>\sum_{j=1,\ j\neq i}^{n}|\hat{a}_{ij}|, \quad 1\leqslant i\leqslant n.$$

Поэтому любая точечная матрица $\hat{\mathcal{A}}_p\in\mathcal{A}$ имеет сильно доминирующую диагональ в обычном смысле и потому невырождена (см., например, [325]).

Для матрицы с сильно доминирующей диагональю можно выполнить условия теоремы 3, если взять вектор $u_p=(u_i)$, такой что $1\leqslant i\leqslant n$. Мы доказали следующее утверждение.

Следствие 5. *Пусть интервальная матрица $\mathcal{A}$ имеет сильно доминирующую диагональ. Тогда метод Гаусса можно выполнить для $\mathcal{A}$ без перестановки строк или столбцов.* ■

Требование строгого доминирования диагонали можно ослабить, сохранив применимость метода Гаусса, если данная интер-

вальная матрица имеет вид

$$\mathscr{A} = \begin{pmatrix} A_1 & C_1 & & 0 \\ B_2 & A_2 & C_2 & \\ & \ddots & \ddots & \ddots \\ 0 & & & B_n \quad A_n \end{pmatrix};$$

т. е. является трехдиагональной интервальной матрицей. Мы предположим еще, что $C_i \neq 0$, $1 \leqslant i \leqslant n-1$ и $B_i \neq 0$, $2 \leqslant i \leqslant n$, так как в противном случае задача распадается на меньшие задачи, для которых эти условия выполнены.

Теорема 6. *Пусть $\mathscr{A}$ — трехдиагональная интервальная матрица, такая что*

$$A_i = \langle a_i, r_i \rangle, \qquad 1 \leqslant i \leqslant n,$$
$$B_i = \langle b_i, s_i \rangle \neq 0, \qquad 2 \leqslant i \leqslant n,$$
$$C_i = \langle c_i, t_i \rangle \neq 0, \qquad 1 \leqslant i \leqslant n-1.$$

Предположим далее, что

$$|a_1| - r_1 > |C_1|,$$
$$|a_i| - r_i \geqslant |B_i| + |C_i|, \qquad 2 \leqslant i \leqslant n-1,$$
$$|a_n| - r_n > |B_n|.$$

Тогда метод Гаусса может быть выполнен без перестановки строк или столбцов.

Замечание. В случае трехдиагональной матрицы $\mathscr{A}$ условия из определения 4 выполнены только для первой и последней строк.

Доказательство теоремы 6. Первый шаг метода Гаусса состоит в порождении трехдиагональной матрицы $\mathscr{A}'$, для которой

$$A_1' = A_1, \quad C_1' = C_1,$$
$$B_2' = 0, \quad B_i' = B_i, \qquad 3 \leqslant i \leqslant n,$$
$$A_2' = A_2 - C_1 B_2 (1/A_1), \quad A_i' = A_i, \quad 3 \leqslant i \leqslant n,$$
$$C_i' = C_i, \qquad 2 \leqslant i \leqslant n-1.$$

Мы покажем, что в матрице $\mathscr{A}'$ сильный критерий суммы по строкам выполнен не только для первой, но и для второй строки, т. е. верно

$$|a_2'| - r_2' > |C_2| = |C_2'|.$$

Мы имеем

$$A_2' = A_2 - C_1(B_2/A_1) \subseteq A_2 - |1/A_1|\langle 0, |C_1|\rangle\langle 0, |B_2|\rangle$$
$$= \langle a_2, r_2 + |C_1||B_2|/(|a_1| - r_1)\rangle,$$

т. е.

$$|a_2'| - r_2' \geqslant |a_2| - \left(r_2 + \frac{|C_1||B_2|}{|a_1| - r_1}\right).$$

Так как

$$|a_1| - r_1 > |C_1| > 0, \quad |B_2| > 0$$

и

$$-|D_2| - r_2 + |a_2| \geqslant |C_2|,$$

мы получаем, что

$$|a_2'| - r_2' \geqslant |a_2| - \left(r_2 + \frac{|C_1||B_2|}{|a_1| - r_1}\right)$$
$$> -|B_2| - r_2 + |a_2| \geqslant |C_2| = |C_2'|.$$

После $n-1$ шага такого типа мы приходим к матрице $\hat{\mathscr{A}}$, имеющей ненулевые элементы только для главной диагонали и супердиагонали (расположенной непосредственно над главной), причем никакой элемент главной диагонали не содержит нуля. Третье предположение $|a_n| - r_n > |B_n|$ применяется на $(n-1)$-м шаге. ■

Заметим, что доказательство теоремы 6 можно несколько сократить. Из условий теоремы 6 следует, что матрица $\mathscr{B}_p$, введенная в теореме 3, имеет неприводимо сильно доминирующую главную диагональ (см. [325], определение 1.7), а потому является M-матрицей. Это означает, что наше утверждение — просто частный случай теоремы 3.

Рассмотрим теперь систему, имеющую более подходящий для итерации вид

$$x = \mathscr{C}x + c,$$

где $\mathscr{C} = (C_{ij})$, $C_{ij} = \langle c_{ij}, r_{ij}\rangle$, $1 \leqslant i, j \leqslant n$ — вещественная интервальная матрица и c — интервальный вектор.

Согласно теореме 12.1, итерационный метод

$$x^{(k+1)} = \mathscr{C}x^{(k)} + c, \quad k = 0, 1, 2, \ldots$$

сходится для любого начального интервального вектора $x^{(0)}$ к единственной неподвижной точке, т. е. решению уравнения $x = \mathscr{C}x + x$, тогда и только тогда, когда спектральный радиус матрицы $|\mathscr{C}| = (|C_{ij}|)$ меньше единицы. Мы хотим показать, что при этом условия теоремы 3 всегда выполнены для матрицы $\mathscr{A} = \mathscr{I}_p - \mathscr{C} = (A_{ij})$, где $\mathscr{I}_p$ — единичная матрица.

Мы имеем

$$A_{ij} = \begin{cases} \langle 1 - c_{ii},\ r_{ii} \rangle, & i = j, \\ -C_{ij} & \text{в противном случае.} \end{cases}$$

Из $\rho(|\mathscr{C}|) < 1$ следует, что $|C_{ii}| = |c_{ii}| + r_{ii} < 1$, $1 \leqslant i \leqslant n$. Матрица $\mathscr{B}_{\mathrm{p}} = (b_{ij})$, введенная в теореме 3, имеет для $\mathscr{A} = \mathscr{I}_{\mathrm{p}} - \mathscr{C}$ элементы

$$b_{ij} = \begin{cases} |1 - c_{ii}| - r_{ii}, & i = j, \\ -|C_{ij}| & \text{в противном случае.} \end{cases}$$

Мы рассмотрим теперь вещественную матрицу $\mathscr{B}_{\mathrm{pl}} = \mathscr{I}_{\mathrm{p}} - |\mathscr{C}|$. Ввиду $\rho(|\mathscr{C}|) < 1$ обратная матрица $\mathscr{B}_{\mathrm{pl}}^{-1}$ существует в силу теоремы 3.8 из [325], причем $\mathscr{B}_{\mathrm{pl}}^{-1} \geqslant \mathscr{O}_{\mathrm{p}}$. Рассмотрим теперь представление матрицы $\mathscr{B}_{\mathrm{pl}}$ в виде $\mathscr{B}_{\mathrm{pl}} = \mathscr{M}_{\mathrm{pl}} - \mathscr{N}_{\mathrm{pl}}$, где

$$\mathscr{M}_{\mathrm{pl}} = \mathrm{diag}(1 - |C_{ii}|), \quad \mathscr{N}_{\mathrm{pl}} = -(\mathscr{B}_{\mathrm{pl}} - \mathscr{M}_{\mathrm{pl}}) = \mathscr{M}_{\mathrm{pl}} - \mathscr{B}_{\mathrm{pl}}.$$

Мы имеем $\mathscr{M}_{\mathrm{pl}}^{-1} \geqslant \mathscr{O}_{\mathrm{p}}$, $\mathscr{N}_{\mathrm{pl}} \geqslant \mathscr{O}_{\mathrm{p}}$. Отсюда в силу теоремы 3.13 из [325] следует, что $\rho(\mathscr{M}_{\mathrm{pl}}^{-1}\mathscr{N}_{\mathrm{pl}}) < 1$.

Рассмотрим, наконец, представление матрицы $\mathscr{B}_{\mathrm{p}}$ в виде $\mathscr{B}_{\mathrm{p}} = \mathscr{M}_{\mathrm{p}} - \mathscr{N}_{\mathrm{p}}$, где

$$\mathscr{M}_{\mathrm{p}} = \mathrm{diag}(|1 - c_{ii}| - r_{ii}), \quad \mathscr{N}_{\mathrm{p}} = -(\mathscr{B}_{\mathrm{p}} - \mathscr{M}_{\mathrm{p}}) = \mathscr{M}_{\mathrm{p}} - \mathscr{B}_{\mathrm{p}}.$$

Мы имеем

$$|1 - c_{ii}| - r_{ii} \geqslant 1 - |c_{ii}| - r_{ii} = 1 - |C_{ii}| > 0, \quad 1 \leqslant i \leqslant n,$$

откуда

$$\mathscr{M}_{\mathrm{p}} \geqslant \mathscr{M}_{\mathrm{pl}}, \text{ т. е. } \mathscr{M}_{\mathrm{pl}}^{-1} \geqslant \mathscr{M}_{\mathrm{p}}^{-1}.$$

Отсюда и из $\mathscr{N}_{\mathrm{p}} = \mathscr{N}_{\mathrm{pl}}$ следует, что $\mathscr{M}_{\mathrm{p}}^{-1}\mathscr{N}_{\mathrm{p}} \leqslant \mathscr{M}_{\mathrm{pl}}^{-1}\mathscr{N}_{\mathrm{pl}}$.

Теперь теорема Перрона-Фробениуса ([325, теорема 2.7]) дает $\rho(\mathscr{M}_{\mathrm{p}}^{-1}\mathscr{N}_{\mathrm{p}}) \leqslant \rho(\mathscr{M}_{\mathrm{pl}}^{-1}, \mathscr{N}_{\mathrm{pl}})$, т. е. $\rho(\mathscr{M}_{\mathrm{p}}^{-1}\mathscr{N}_{\mathrm{p}}) < 1$.

Применяя теорему 3.10 из [325], мы получаем, наконец, что $\mathscr{B}_{\mathrm{p}}$ является M-матрицей.

Замечания. Утверждения о применимости метода Гаусса (теорема 3) можно найти в [348а]. Доказательство теоремы 3 — это непосредственное обобщение доказательства таких же утверждений для точечных матриц (см. [328]). Пусть матрица $\mathscr{A}^{\mathrm{T}}$ — транспонированная для интервальной матрицы $\mathscr{A}$. Мы имеем утверждение, соответствующее следствию 5. Если $\mathscr{A}^{\mathrm{T}}$ имеет сильно доминирующую диагональ (в смысле определения 4), то метод Гаусса может быть выполнен для интервальной матрицы

$\mathscr{A}$ без перестановок строк. Доказательство получается ссылкой на теорему 3, так как введенная там матрица $\mathscr{B}_p$ снова оказывается M-матрицей. Аналогичное утверждение справедливо для трехдиагональной матрицы $\mathscr{A}$, если для $\mathscr{A}^T$ выполнены условия теоремы 6.

В гл. 19 только что доказанные утверждения будут использованы для решения систем нелинейных точечных уравнений.

Вопрос о применимости метода Гаусса не был пока что исследован удовлетворительным образом. В [485] было показано, что этот метод применим для частного класса уравнений.

Пример комплексной матрицы размерности 2×2, где метод Гаусса не срабатывает, был дан в [469]. В [484] описан класс интервальных матриц размерности 3×3, не содержащих вырожденных точечных матриц, но таких, что метод Гаусса не срабатывает ни при каком выборе главных элементов. Явное описание множества

$$\mathfrak{L} = \{x_p \mid \mathscr{A}_p x_p = b_p,\ \mathscr{A}_p \in \mathscr{A},\ b_p \in b\},$$

введенного в начале этой главы, дано недавно в [399].

* * *

Литература: [325], [328], [348a], [361], [385a], [399], [401], [467a], [469], [484], [485].

Глава 16
МЕТОД ХАНСЕНА

Если в системе линейных интервальных уравнений диагональ не является сильно доминирующей (определение 15.4), то ее можно решать с помощью преобразования, предложенного Хансеном. Цель этого преобразования — сделать данную систему интервальных уравнений системой с сильно доминирующей диагональю. Пусть дана интервальная матрица $\mathscr{A} = (A_{ij})$, где $A_{ij} = \langle a_{ij}, r_{ij} \rangle$ — элементы из $I(\mathbb{R})$ или $K(\mathbb{C})$. Будем предполагать, что существуют обращения всех точечных матриц $\mathscr{A}_p \in \mathscr{A}$. Берем обращение точечной матрицы $m(\mathscr{A}) := (a_{ij})$ и с помощью $m(\mathscr{A})^{-1}$ строим интервальную матрицу

$$\tilde{\mathscr{A}} = m(\mathscr{A})^{-1} \mathscr{A}$$

и интервальный вектор

$$\tilde{b} = m(\mathscr{A})^{-1} b.$$

Мы имеем

$$\{x_p \mid \mathscr{A}_p x_p = b_p \mathscr{A}_p \in \mathscr{A},\ b_p \in b\} \subseteq \{y_p \mid \mathscr{A}_p y_p = \tilde{b}_p \tilde{\mathscr{A}}_p \subset \tilde{\mathscr{A}},\ \tilde{b}_p \in \tilde{b}\}.$$

Чтобы показать это, предположим, что x_p принадлежит множеству из левой части, т. е.

$$\mathscr{A}_p x_p = b_p, \text{ где } \mathscr{A}_p \in \mathscr{A},\ b_p \in b.$$

Тогда

$$m(\mathscr{A})^{-1} \mathscr{A}_p x_p = m(\mathscr{A})^{-1} b_p$$

и наше утверждение следует из

$$m(\mathscr{A})^{-1} \mathscr{A}_p \in \tilde{\mathscr{A}}, \quad m(\mathscr{A})^{-1} b_p \in \tilde{b}.$$

Идея рассматриваемого преобразования состоит в том, что если элементы матрицы $\mathscr{A}$ имеют не слишком большую ширину, то диагональ матрицы $\tilde{\mathscr{A}}$ будет сильно доминирующей. Тогда можно применить метод Гаусса. В пределе мы имеем $d(\mathscr{A}) = \mathscr{O}_p$, т. е. $\tilde{\mathscr{A}} = \mathscr{I}_p$, так что в этом случае матрица $\tilde{\mathscr{A}}$ наверняка имеет сильно доминирующую диагональ. Если же ширина компонент матрицы $\mathscr{A}$ невелика, то $\mathscr{A}$ несильно отличается от $\mathscr{I}_p$.

Ясно, что сильное доминирование диагонали для матрицы $\tilde{\mathscr{A}} = m(\mathscr{A})^{-1}\mathscr{A}$ зависит не только от ширины компонент матрицы $\mathscr{A}$. Действительно, если мы представим компоненты этой интервальной матрицы (которые могут быть круговыми комплексными или вещественными интервалами) в виде

$$A_{ij} = \langle a_{ij},\ r_{ij} \rangle, \quad 1 \leqslant i,\ j \leqslant n,$$

то, вводя еще матрицу $\mathscr{D} = (D_{ij})$, $D_{ij} = \langle 0,\ r_{ij} \rangle$, $1 \leqslant i,\ j \leqslant n$, получим, что

$$\begin{aligned} \tilde{\mathscr{A}} = m(\mathscr{A})^{-1}\mathscr{A} &= m(\mathscr{A})^{-1}(m(\mathscr{A}) + \mathscr{D}) \\ &= \mathscr{I}_p + m(\mathscr{A})^{-1}\mathscr{D} = \mathscr{I}_p + |m(\mathscr{A})^{-1}|\mathscr{D} \\ &= \mathscr{I}_p + \mathscr{H}, \end{aligned}$$

где

$$\mathscr{H} = |m(\mathscr{A})^{-1}|\mathscr{D}.$$

Так как

$$\begin{aligned} \| |\mathscr{H}| \| &\leqslant \| |m(\mathscr{A})^{-1}| \| \cdot \| |\mathscr{D}| \| = \frac{1}{2} \| |m(\mathscr{A})^{-1}| \| \| d(\mathscr{A}) \| \\ &\leqslant \frac{1}{2} \| |m(A)^{-1}| \| \cdot \| d(\mathscr{A}) \| / \| m(\mathscr{A}) \|, \end{aligned}$$

то мы видим, что при данной точечной матрице $m(\mathscr{A})$ диагональ матрицы $\mathscr{A}$ будет сильно доминировать с тем большей вероятностью, чем меньше число

$$\tilde{\varkappa} = \| |m(\mathscr{A})^{-1}| \| \| m(\mathscr{A}) \|$$

Если $m(\mathscr{A})$ — вещественная точечная матрица и мы используем монотонную матричную норму, то

$$\tilde{\varkappa} = \| m(\mathscr{A})^{-1} \| \| m(\mathscr{A}) \|,$$

где $\tilde{\varkappa}$ — хорошо известная величина, обусловленность матрицы $m(\mathscr{A})$. Поэтому применимость метода Хансена к вещественной интервальной матрице зависит не только от величины $d(\mathscr{A})$, но (даже более существенным образом) и от обусловленности матрицы $m(\mathscr{A})$.

В предыдущей главе мы описали метод преобразования множества линейных интервальных уравнений к верхней треугольной форме. Другие методы, разработанные в теории вещественных систем линейных уравнений, также могут быть обобщены на линейные системы интервальных уравнений. Мы упомянем метод Гаусса — Жордана, которым данная матрица приводится к диагональному виду. После этого вычисление решения исходной системы требует еще n добавочных делений. Подробное описание этого варианта алгоритма Гаусса можно найти в [410a]. Тем же методом, что и в теореме 15.3, можно показать, что этот метод всегда применим к данной системе линейных интервальных уравнений, если ее матрица имеет сильно доминирующую диагональ.

Пусть дана вещественная интервальная матрица $\mathscr{A} = (A_{ij})$. Будем предполагать, что обратная матрица $\mathscr{A}_{\mathrm{p}}^{-1}$ существует для любой $\mathscr{A}_{\mathrm{p}} \in \mathscr{A}$. Пусть далее $\mathscr{b} = (B_i)$ — вещественный интервальный вектор.

Запишем в виде

$$\mathscr{l} = (L_1, L_2, \ldots, L_n)^{\mathrm{T}}$$

вещественный интервальный вектор с компонентами L_i, $1 \leqslant i \leqslant n$, которые получаются из множества

$$\mathfrak{L} = \{x_{\mathrm{p}} \mid \mathscr{A}_{\mathrm{p}} x_{\mathrm{p}} = \mathscr{b}_{\mathrm{p}},\ \mathscr{A}_{\mathrm{p}} \in \mathscr{A},\ \mathscr{b}_{\mathrm{p}} \in \mathscr{b}\}$$

проектированием на соответствующие оси координат. Иными словами,

$$L_i = L_i(\mathscr{A}, \mathscr{b}) = \{l_i \mid (l_1, \ldots, l_i, \ldots, l_n)^{\mathrm{T}} \in \mathfrak{L}\},\ 1 \leqslant i \leqslant n.$$

$\mathscr{l}$ — интервальный вектор наименьшей ширины, содержащий множество $\mathfrak{L}$.

Мы хотим теперь исследовать вопрос о том, насколько хорошо интервальные векторы, вычисляемые по методу Хансена, аппроксимируют вектор $\mathscr{l}$. Для «решения» преобразованной системы линейных уравнений мы используем тогда метод Гаусса — Жордана. Будет показано, что разность между шириной полученного вектора и шириной вектора $\mathscr{l}$ стремится к нулю при

стремлении к нулю ширины векторов $\mathscr{A}$ и b. Это показывает, что метод Хансена дает вектор, достаточно близкий к вектору ℓ, если ширина исходных данных не слишком велика.

Мы снова представляем вещественный интервал $A = [a_1, a_2]$ с помощью его середины $a = \frac{1}{2}(a_1 + a_2)$ и полуширины $r = \frac{1}{2} d(A) = \frac{1}{2}(a_2 - a_1)$:

$$A = \langle a, r \rangle.$$

Лемма 1. *Пусть $\mathscr{A}_{\text{р}} x_{\text{р}} = b_{\text{р}}$, где $\mathscr{A}_{\text{р}} = (a_{ij})$ — вещественная невырожденная матрица, имеющая обратную $\mathscr{A}_{\text{р}}^{-1} = \mathscr{B}_{\text{р}} = (b_{ij})$, и пусть $x_{\text{р}} = (x_j)$, $b_{\text{р}} = (b_i)$ — точечные векторы. Пусть далее $\mathscr{A} = (A_{ij})$ — интервальная матрица с элементами $A_{ij} = \langle a_{ij}, r_{ij} \rangle$, $1 \leqslant i, j \leqslant n$, а $b = (B_i)$ — интервальный вектор с элементами $B_i = \langle b_i, r_i \rangle$, $1 \leqslant i \leqslant n$. Тогда для k-й компоненты L_k введенного выше интервального вектора $\ell = (L_i)$ справедливо соотношение*

$$\frac{1}{2} d(L_k) = \sum_{i=1}^{n} \sum_{j=1}^{n} |b_{ki} x_j r_{ij}| + \sum_{i=1}^{n} |b_{ki} r_i| + O(d^2).$$

Здесь

$$d = \max \left\{ \max_{1 \leqslant i, j \leqslant n} \{r_{ij}\}, \max_{1 \leqslant i \leqslant n} \{r_i\} \right\}.$$

Запись $O(d^2)$ обозначает любую вещественную функцию f от d, для которой

$$|f/d^2| \leqslant \gamma \text{ при } d \leqslant d_0,$$

где $\gamma \geqslant 0$, $d_0 > 0$ — константы.

Доказательство. В силу правила Крамера множество L_k является образом $(n^2 + n)$-мерного гиперкуба при отображении

$$x_k = x_k(\mathscr{A}_{\text{р}}, b_{\text{р}}).$$

Из теоремы о среднем значении мы получаем, что

$$x_k(\hat{\mathscr{A}}_{\text{р}}, \hat{b}_{\text{р}}) = x_k(\mathscr{A}_{\text{р}}, b_{\text{р}}) + \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial x_k}{\partial a_{ij}} (\hat{a}_{ij} - a_{ij}) + \sum_{i=1}^{n} \frac{\partial x_k}{\partial b_i} (\hat{b}_i - b_i) + \frac{1}{2} x_k'' (u_{\text{р}} + l(v_{\text{р}} - u_{\text{р}}))(v_{\text{р}} - u_{\text{р}}) \times (v_{\text{р}} - u_{\text{р}}).$$

Здесь $l \in (0, 1)$, через x_k'' обозначен гессиан отображения x_k, а $u_{\text{р}}, v_{\text{р}}$ — это векторы из $V_{n^2+n}(\mathbb{R})$, причем значениями компонент вектора $u_{\text{р}}$ (соответственно вектора $v_{\text{р}}$) являются элемен-

ты матрицы $\mathscr{A}_p$ и вектор b_p (соответственно элементы матрицы $\hat{\mathscr{A}}_p$ и вектор $\hat{b}_p$).

Если теперь продифференцировать n уравнений

$$\sum_{j=1}^{n} a_{ij}x_j = b_i, \quad 1 \leqslant i \leqslant n,$$

по a_{ij}, то, используя соотношение

$$\mathscr{A}_p \frac{\partial}{\partial a_{ij}} x_p = -x_j e_p^i,$$

мы получим уравнение

$$\frac{\partial}{\partial a_{ij}} x_p = \left(\frac{\partial x_1}{\partial a_{ij}}, \frac{\partial x_2}{\partial a_{ij}}, \ldots, \frac{\partial x_n}{\partial a_{ij}}\right)^T,$$

где e_p^i есть i-й единичный вектор. Отсюда мы получаем

$$\frac{\partial}{\partial a_{ij}} x_p = -x_j \mathscr{A}_p^{-1} e_p^i \text{ или } \frac{\partial x_k}{\partial a_{ij}} = -b_{ki}x_j.$$

Из равенства

$$x_p = \mathscr{B}_p b_p$$

мы получаем

$$\frac{\partial x_k}{\partial b_i} = b_{ki}.$$

Если использовать эти формулы для производных в теореме о среднем значении, то мы получим

$$x_k(\hat{\mathscr{A}}_p, \hat{b}_p) \in x_k(\mathscr{A}_p, b_p) + \sum_{i=1}^{n}\sum_{j=1}^{n} |b_{ki}x_j| \langle 0, r_{ij}\rangle$$

$$+ \sum_{i=1}^{n} |b_{ki}| \langle 0, r_i\rangle + \ldots,$$

так как

$$\hat{\mathscr{A}}_p \in \mathscr{A}, \quad \hat{b}_p \in b.$$

Иными словами,

$$\frac{1}{2} d(L_k) = \sum_{i=1}^{n}\sum_{j=1}^{n} |b_{ki}x_j r_{ij}| + \sum_{i=1}^{n} |b_{ki}r_i| + O(d^2). \blacksquare$$

Лемма 2. *Пусть $\mathscr{A}_p = (a_{ij})$ — вещественная невырожденная матрица, для которой $\mathscr{A}_p^{-1} = \mathscr{B}_p = (b_{ij})$, и пусть $b_p = (b_i)$ — вещественный вектор. Пусть далее $\mathscr{A} = (A_{ij})$ — вещественная интервальная матрица с элементами $A_{ij} = \langle a_{ij}, r_{ij}\rangle$, $1 \leqslant i, j \leqslant n$, и $b = (B_i)$ — вещественный интервальный вектор с элементами $B_i = \langle b_i, r_i\rangle$, $1 \leqslant i \leqslant n$. Запишем в виде $\tilde{\mathscr{A}} = (A_{ij})$ интервальную матрицу $\mathscr{A}_p^{-1}\mathscr{A} = \mathscr{B}_p\mathscr{A}$, а в виде $\tilde{b} = (B_i)$ — интервальный*

вектор $\mathscr{A}_{\mathrm{p}}^{-1}b = \mathscr{B}_{\mathrm{p}}b$. *Предположим, что все точечные матрицы, принадлежащие* $\mathscr{A}$, *невырожденные. Тогда интервальный вектор*

$$\tilde{l} = (\tilde{L}_1,\ \tilde{L}_2,\ \ldots,\ \tilde{L}_n)^{\mathrm{T}},$$

где

$$L_i = L_i(\tilde{\mathscr{A}},\ \tilde{b}) = \{\tilde{l}_i \mid (\tilde{l}_1,\ \ldots,\ \tilde{l}_i,\ \ldots,\ \tilde{l}_n)^{\mathrm{T}} \in \Omega\},\quad 1 \leqslant i \leqslant n,$$

$$\Omega = \{\tilde{x}_{\mathrm{p}} \mid \mathscr{A}_{\mathrm{p}} = \tilde{b}_{\mathrm{p}},\ \mathscr{A}_{\mathrm{p}} \in \tilde{\mathscr{A}},\ \tilde{b}_{\mathrm{p}} \in \tilde{b}\},$$

удовлетворяет соотношению

$$d(\tilde{L}_k) = d(L_k) + O(d^2),\quad 1 \leqslant k \leqslant n.$$

Доказательство. Пусть $\mathscr{A}_{\mathrm{p}}x_p = b_{\mathrm{p}}$. Из леммы 1 следует, что

$$\frac{1}{2}d(L_k) = \sum_{i=1}^{n}\sum_{j=1}^{n}|b_{ki}x_j r_{ij}| + \sum_{i=1}^{n}|b_{ki}r_i| + O(d^2).$$

Элементы интервальной матрицы $\tilde{\mathscr{A}} = (\tilde{A}_{ij})$ имеют вид

$$\tilde{A}_{ij} = \left\langle \delta_{ij},\ \sum_{m=1}^{n}|b_{im}r_{im}| \right\rangle,\quad 1 \leqslant i,\ j \leqslant n,$$

где δ_{ij} — символ Кронекера. Элементы интервального вектора $\tilde{b} = (\tilde{B}_i)$ имеют вид

$$\tilde{B}_i = \left\langle x_i,\ \sum_{m=1}^{n}|b_{im}r_m| \right\rangle,\quad 1 \leqslant i \leqslant n.$$

Применяя лемму 1 еще раз, мы получаем

$$\begin{aligned}\frac{1}{2}d(\tilde{L}_k) &= \sum_{i=1}^{n}\sum_{j=1}^{n}\left|\delta_{ki}x_j\left(\sum_{m=1}^{n}|b_{im}r_{im}|\right)\right| \\ &\quad + \sum_{i=1}^{n}\left|\delta_{ki}\left(\sum_{m=1}^{n}|b_{im}r_m|\right)\right| + O(\tilde{d}^2) \\ &= \sum_{j=1}^{n}\left|x_j\left(\sum_{m=1}^{n}|b_{km}r_{mj}|\right)\right| + \sum_{m=1}^{n}|b_{km}r_m| + O(\tilde{d}^2) \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n}|b_{ki}x_j r_{ij}| + \sum_{i=1}^{n}|b_{ki}r_i| + O(\tilde{d}^2),\end{aligned}$$

где

$$d = \max\left\{\max_{1 \leqslant i,\ j \leqslant n}\{\tilde{r}_{ij}\},\ \max_{1 \leqslant i \leqslant n}\{\tilde{r}_i\}\right\}.$$

Из формул для элементов матрицы $\tilde{\mathscr{A}}$ и вектора $\tilde{b}$ мы сразу усматриваем, что

$$\tilde{d} = O(d).$$

Сравнение с выражением для $\frac{1}{2} d(L_k)$ дает

$$d(\tilde{L}_k) = d(L_k) + O(d^2),$$

что и доказывает лемму. ■

Теперь метод Гаусса — Жордана за конечное число шагов порождает по данной интервальной матрице $\mathscr{A}$ некоторую диагональную матрицу. В результате каждого шага в новой матрице появляется хотя бы один новый нуль вне диагонали. Предположим, что даны интервальная матрица $\mathscr{H} = (H_{ij})$ и интервальный вектор $h = (H_i)$, такие что

$$H_{ij} = \langle h_{ij}, e_{ij} \rangle, \quad 1 \leqslant i, j \leqslant n,$$
$$H_i = \langle h_i, e_i \rangle, \quad 1 \leqslant i \leqslant n,$$

где $h_{ij} = \delta_{ij}$, $h_i = x_i$ и

$$\max \left\{ \max_{1 \leqslant i, j \leqslant n} \{e_{ij}\}, \max_{1 \leqslant i \leqslant n} \{e_i\} \right\} = O(d).$$

Пусть, кроме того, $\mathscr{A} = (A_{ij})$, $b = (b_{ij})$, где

$$A_{ij} = \langle a_{ij}, r_{ij} \rangle, \quad 1 \leqslant i, j \leqslant n,$$
$$B_i = \langle b_i, r_i \rangle, \quad 1 \leqslant i \leqslant n.$$

Положим по определению

$$d = \max \left\{ \max_{1 \leqslant i, j \leqslant n} \{r_{ij}\}, \max_{1 \leqslant i \leqslant n} \{r_i\} \right\}.$$

Теперь мы можем доказать по индукции, что наши предположения о матрице $\mathscr{H}$ и векторе h верны на любом шаге алгоритма. Как показано в лемме 2, эти предположения справедливы для матрицы $\mathscr{H} := \tilde{\mathscr{A}} = \mathscr{A}_{\text{р}}^{-1} \mathscr{A}$ и интервального вектора $h := \tilde{b} = \mathscr{A}_{\text{р}}^{-1} b$. Кроме того, из леммы 2 следует, что

$$d(\tilde{L}_k) = d(L_k(\mathscr{H}, h)) = d(L_k) + O(d^2), \quad 1 \leqslant k \leqslant n.$$

Чтобы получить нуль для пары (r, s) индексов, (где $r \neq s$), по матрице $\mathscr{H}'$ и вектору h' строятся матрица $\mathscr{H}$ и вектор h согласно следующим формулам:

$$H'_{ij} = H_{ij}, \qquad i \neq r,$$
$$H'_{rj} = H_{rj} - H_{sj} H_{rs} / H_{ss}, \quad j \neq s,$$
$$H'_{rs} = 0,$$
$$H'_i = H_i, \qquad i \neq r,$$
$$H_i = H_i - H_s H_{rs} / H_{ss}.$$

Ввиду $h_{ij} = \delta_{ij}$ отсюда следует, что

$$\begin{aligned} H'_{rj} &= \langle h_{rj},\ e_{rj}\rangle - \langle h_{sj},\ e_{sj}\rangle \langle h_{rs},\ e_{rs}\rangle / \langle h_{ss},\ e_{ss}\rangle \\ &= \langle \delta_{rj},\ e_{rj}\rangle - \langle 0,\ e_{sj}\rangle \langle 0,\ e_{rs}\rangle / \langle h_{ss},\ e_{ss}\rangle \\ &= \langle \delta_{rj},\ e_{rj}\rangle - \langle 0,\ e_{sj}e_{rs}\rangle (1/1\,(-e_{ss})) \\ &= \langle \delta_{rj},\ e_{rj} + [e_{sj}e_{rs}/(1 - e_{ss})]\rangle \end{aligned}$$

и

$$\begin{aligned} H'_r &= \langle x_r,\ e\rangle - \langle x_s,\ e_s\rangle \langle 0,\ e_{rs}\rangle / \langle h_{ss},\ e_{ss}\rangle \\ &= \langle x_r,\ e_r + (((|\, x_s \,| e_{rs} + e_{rs}e_s)/(1 - e_{ss}))\rangle. \end{aligned}$$

Поэтому представление

$$H'_{ij} = \langle \delta_{ij},\ e'_{ij}\rangle, \quad H'_i = \langle x_i,\ e'_i\rangle$$

справедливо также для матрицы $\mathscr{H}' = (H'_{ij})$ и вектора $\mathscr{k}' = (H'_i)$.

Из леммы 1 следует, что

$$\frac{1}{2}\, d\,(L_k(\mathscr{H},\ \mathscr{k})) = \sum_{i=1}^{n} \sum_{j=1}^{n} |\, \delta_{ki} x_j e_{ij} \,| + \sum_{i=1}^{n} |\, \delta_{ki} e_i \,| + O\,(d^2)$$

$$= \sum_{j=1}^{n} |\, x_j e_{kj} \,| + e_k + O\,(d^2).$$

Аналогично мы получаем

$$\frac{1}{2}\, d\,(L_k(\mathscr{H}',\ \mathscr{k}')) = \sum_{j=1}^{n} |\, x_j e'_{kj} \,| + e'_k + O\,(d^2).$$

Для $k \neq r$ мы имеем $e'_{kj} = e_{kj}$ и $e'_k = e_k$. Из допущения

$$d\,(L_k(\mathscr{H},\ \mathscr{k})) = d\,(L_k) + O\,(d^2)$$

мы получаем поэтому

$$d\,(L_k(\mathscr{H}',\ \mathscr{k}')) = d\,(L_k) + O\,(d^2), \quad k \neq r.$$

Для $k = r$ мы имеем

$$\sum_{j=1}^{n} |\, x_k e'_{kj} \,| = \sum_{j=1,\ j \neq s}^{n} |\, x_j e_{kj} \,| + O\,(d^2)$$

и

$$e'_k = e_k + |\, x_s e_{ks} \,| + O\,(d^2),$$

откуда

$$\frac{1}{2}\, d\,(L_k(\mathscr{H}',\ \mathscr{k}')) = \sum_{j=1}^{n} |\, x_j e_{kj} \,| + e_k + O\,(d^2),$$

т. е.

$$d\,(L_k(\mathscr{H}',\ \mathscr{k}')) = d\,(L_k) + O\,(d^2) \text{ для } k = r.$$

Этим завершается доказательство по индукции.

Решение диагональной системы уравнений не увеличивает ширину элементов, поэтому соотношение

$$d(L_k(\hat{\mathscr{H}}, \hat{h})) = d(L_k(\mathscr{A}, b)) + O(d^2)$$

справедливо для заключительной диагональной матрицы $\mathscr{H}$ и соответствующего интервального вектора h.

Замечания. Описанный в этой главе метод «решения» систем уравнений с интервальными коэффициентами был предложен Хансеном в [108, 115]. Леммы 1, 2 и их следствия применительно к методу Хансена принадлежат Миллеру [206].

* * *

Литература: [108], [115], [206], [410a].

Глава 17
ПРОЦЕДУРА КУПЕРМАНА И ХАНСЕНА

Пусть $\mathscr{A}$ — интервальная матрица, такая, что $\mathscr{A}_p$ — невырожденная для всех $\mathscr{A}_p \in \mathscr{A}$, и пусть

$$\Omega = \{x_p \mid \mathscr{A}_p x_p = b_p,\ \mathscr{A}_p \in \mathscr{A},\ b_p = b\}$$

— множество всех решений для данного интервального вектора b. Даже простые примеры показывают, что метод Хансена, описанный в предыдущей главе, вычисляет в общем случае лишь некоторое подмножество вектора l, имеющего компоненты

$$L_k = \{l_k \mid (l_1, \ldots, l_k, \ldots, l_n)^T \in \Omega\}.$$

Куперман описал в [183] неинтервальную процедуру, которая в некоторых случаях дает лучшую локализацию множества Ω. Впоследствии Хансен обобщил эту процедуру до интервального метода.

Рассмотрим множество линейных уравнений

$$\mathscr{A}_p x_p = b_p, \quad x_p = (x_i),$$

где $\mathscr{A}_p$ — неособенная точечная матрица и b_p — вещественный вектор. Если частная производная неотрицательна

$$\frac{\partial x_k}{\partial a_{ij}} \geqslant 0,$$

то x_k — неубывающая функция от a_{ij}. Если теперь a_{ij} может изменяться в вещественном интервале

$$A_{ij} = [a_{ij}^1, a_{ij}^2],$$

то величина x_k, рассматриваемая как функция от a_{ij} на интервале $A_{ij} = [a_{ij}^1, a_{ij}^2]$, принимает свое наименьшее значение при $a_{ij} = a_{ij}^1$, а наибольшее значение при $a_{ij} = a_{ij}^2$. То же самое можно сказать и о зависимости компонент x_k от b_i.

Пусть теперь $\mathscr{A} = (A_{ij})$ — данная вещественная интервальная матрица и $\mathscr{b} = (B_i)$ — данный интервальный вектор. Чтобы локализовать интервал

$$L_k = [l_k^1, l_k^2], \quad 1 \leqslant k \leqslant n,$$

мы поступаем следующим образом.

Начинаем с вещественной интервальной матрицы $\mathscr{A} = (A_{ij})$, $A_{ij} = [a_{ij}^1, a_{ij}^2]$, $1 \leqslant i, j \leqslant n$, и вещественного интервального вектора $\mathscr{b} = (B_i)$, $B_i = [b_i^1, b_i^2]$, $1 \leqslant i \leqslant n$. Затем строим интервальные матрицы

$$\tilde{\mathscr{A}} = (\tilde{A}_{ij}) \text{ и } \hat{\mathscr{A}} = (\hat{A}_{ij})$$

и интервальные векторы

$$\tilde{\mathscr{b}} = (\tilde{B}_i) \text{ и } \hat{\mathscr{b}} = (\hat{B}_i)$$

согласно следующим правилам:

$$\tilde{A}_{ij} = \begin{cases} [a_{ij}^1, a_{ij}^1], & \text{если } \partial x_k/\partial a_{ij} \geqslant 0 \text{ для всех } \mathscr{A}_p \in \mathscr{A} \text{ и } \mathscr{b}_p \in \mathscr{b}, \\ [a_{ij}^2, a_{ij}^2], & \text{если } \partial x_k/\partial a_{ij} \leqslant 0 \text{ для всех } \mathscr{A}_p \in \mathscr{A} \text{ и } \mathscr{b}_p \in \mathscr{b}, \\ A_{ij} & \text{в противном случае.} \end{cases}$$

$$\hat{A}_{ij} = \begin{cases} [a_{ij}^2, a_{ij}^2], & \text{если } \partial x_k/\partial a_{ij} \geqslant 0 \text{ для всех } \mathscr{A}_p \in \mathscr{A} \text{ и } \mathscr{b}_p \in \mathscr{b}, \\ [a_{ij}^1, a_{ij}^1], & \text{если } \partial x_k/\partial a_{ij} \leqslant 0 \text{ для всех } \mathscr{A}_p \in \mathscr{A} \text{ и } \mathscr{b}_p \in \mathscr{b}, \\ A_{ij} & \text{в противном случае.} \end{cases}$$

$$\tilde{B} = \begin{cases} [b_i^1, b_i^1], & \text{если } \partial x_k/\partial b_i \geqslant 0 \text{ для всех } \mathscr{A}_p \in \mathscr{A} \text{ и } \mathscr{b}_p \in \mathscr{b}, \\ [b_i^2, b_i^2], & \text{если } \partial x_k/\partial b_i \leqslant 0 \text{ для всех } \mathscr{A}_p \in \mathscr{A} \text{ и } \mathscr{b}_p \in \mathscr{b}, \\ B_i & \text{в противном случае.} \end{cases}$$

$$\hat{B}_i = \begin{cases} [b_i^2, b_i^2], & \text{если } \partial x_k/\partial b_i \geqslant 0 \text{ для всех } \mathscr{A}_p \in \mathscr{A} \text{ и } \mathscr{b}_p \in \mathscr{b}, \\ [b_i^1, b_i^1], & \text{если } \partial x_k/\partial b_i \leqslant 0 \text{ для всех } \mathscr{A}_p \in \mathscr{A} \text{ и } \mathscr{b}_p \in \mathscr{b}, \\ B_i & \text{в противном случае.} \end{cases}$$

Теперь мы вычисляем локализующие интервалы

$$X_k = [x_k^1, x_k^2] \quad \text{и} \quad Y_k = [y_k^1, y_k^2]$$

для интервалов

$$L_k(\tilde{\mathscr{A}}, \tilde{b}) = [\tilde{l}_k^1, \tilde{l}_k^2] \quad \text{и} \quad L_k(\hat{\mathscr{A}}, \hat{b}) = [\hat{l}_k^1, \hat{l}_k^2],$$

используя, например, метод Хансена, описанный в предыдущей главе.

Предыдущие рассуждения показывают, что для $L_k(\mathscr{A}, b) = [l_k^1, l_k^2]$ имеет место

$$l_k^1 \geqslant \tilde{l}_k^1 \text{ и } l_k^2 \leqslant \hat{l}_k^2.$$

Отсюда следует, что верно

$$[\hat{l}_k^1, \hat{l}_k^2] \supseteq L_k(\mathscr{A}, b).$$

Поэтому мы имеем также

$$[x_k^1, y_k^2] \supseteq [\tilde{l}_k^1, \hat{l}_k^2] \supseteq L_k(\mathscr{A}, b).$$

Для построения интервальных матриц $\tilde{\mathscr{A}}$, $\hat{\mathscr{A}}$ и интервальных векторов $\tilde{b}$, $\hat{b}$ при фиксированном k нам нужны частные производные $\partial x_k/\partial a_{ij}$ и $\partial x_k/\partial b_i$. Формулы для них уже были получены в предыдущей главе: полагая $\mathscr{A}_p^{-1} = (\bar{a}_{ij})$, мы имеем

$$\frac{\partial x_k}{\partial a_{ij}} = -\bar{a}_{ki} x_j, \quad \frac{\partial x_k}{\partial b_i} = \bar{a}_{ki}.$$

Чтобы выяснить распределение знаков этих производных, мы вычисляем, используя, например, метод Хансена, интервальный вектор $\bar{l} = (\bar{L}_i)$, содержащий множество $\mathfrak{L}$, и интервальную матрицу $\bar{\mathscr{A}} = (\bar{A}_{ij})$, содержащую обращения всех $\mathscr{A}_p \in \mathscr{A}$. Если теперь нижняя граница интервала $\bar{A}_{ki} L_i$ неотрицательна для некоторого фиксированного k, то $\partial x_k/\partial a_{ij} \leqslant 0$ для всех $\mathscr{A}_p \in \mathscr{A}$ и $b_p \in b$. Аналогично, если верхняя граница интервала $\bar{A}_{ki} \bar{L}_i$ неположительна, то $\partial x_k/\partial a_{ij} \geqslant 0$. Соответствующее утверждение верно и для $\partial x_k/\partial b_i$. Если $0 \in \bar{A}_{ki} \bar{L}_i$ (соответственно $0 \in \bar{A}_{ki}$), то мы все еще можем иметь $\partial x_k/\partial a_{ij} \geqslant 0$ или $\leqslant 0$ (соответственно $\partial x_k/\partial b_i \geqslant 0$ или $\leqslant 0$), так как $\bar{l}$ только содержит множество $\mathfrak{L}$, а $\bar{\mathscr{A}}$ только локализует множество обращений матриц $\mathscr{A}_p \in \mathscr{A}$. При построении матриц $\tilde{\mathscr{A}}$, $\hat{\mathscr{A}}$ и векторов $\tilde{b}$, $\hat{b}$ элементы матрицы $\mathscr{A}$ и вектора b преобразуются в этом случае так, как будто $\partial x_k/\partial a_{ij}$ (соответственно $\partial x_k/\partial b_i$) меняет знак, потому, что вычисленные значения не позволяют нам принять иное решение.

Этот метод дает, вообще говоря, гораздо лучшую локализацию, чем метод Хансена из гл. 16. Его недостаток — большой объем вычислений. В общем случае приходится не только вы-

числять интервальную матрицу $\bar{\mathscr{A}}$, содержащую обращения всех матриц $\mathscr{A}_p \in \mathscr{A}$, но и решать две системы интервальных уравнений для каждой компоненты $L_k(\mathscr{A}, \mathscr{b})$.

Замечания. Метод Купермана и Хансена можно применить и к итерационным методам гл. 12. Подробное рассмотрение этого случая с результатами вычислений и программой на Алголе 60 имеется в [18].

Мы отметим также, что формулы для построения $2n$ систем уравнений приведены в [60] без явного использования интервальной арифметики.

Отметим еще, что можно искать метод, основанный на решении $2n$ систем уравнений методом Гаусса. Это было сделано в [336]. Однако использование метода Хансена всегда дает лучшую локализацию.

* * *

Литература: [18], [60], [111], [112], [183], [184], [336].

Глава 18
ИТЕРАЦИОННЫЕ МЕТОДЫ ДЛЯ ЛОКАЛИЗАЦИИ ОБРАТНОЙ МАТРИЦЫ И РАЗЛОЖЕНИЯ НА ТРЕУГОЛЬНЫЕ

Пусть даны невырожденная матрица $\mathscr{A}_p \in M_{nn}(\mathbb{R})$ размерности $n \times n$ и интервальная матрица $\mathscr{X}^{(0)} \in M_{nn}(I(\mathbb{R}))$, такая, что $\mathscr{X}^{(0)}$ локализует матрицу, обратную к $\mathscr{A}_p$, т. е. $\mathscr{A}_p^{-1} \in \mathscr{X}^{(0)}$. Мы рассмотрим здесь процедуры, которые итерационно улучшают локализующую матрицу $\mathscr{X}^{(0)}$.

При этом мы будем использовать преобразование m, отображающее множество интервальных матриц размерности $n \times n$ во множестве вещественных точечных матриц размерности $n \times n$. Оно переводит каждую интервальную матрицу в такую, элементами которой являются середины соответствующих элементов исходной матрицы. Иными словами, в обозначениях

$$i(X) = x_1, \quad s(X) = x_2 \quad X = [x_1, x_2] \in I(\mathbb{R})$$

мы вводим отображение.

(1) $$m: M_{nn}(I(\mathbb{R})) \to M_{nn}(\mathbb{R})$$

равенствами $m(\mathscr{X}) = \frac{1}{2}(i(X_{ij}) + s(X_{ij}))$.

Это срединное отображение интервальных матриц очевидным образом непрерывно. Оно обладает следующими свойствами:

(2) $$m(\mathscr{X} \pm \mathscr{Y}) = m(\mathscr{X}) \pm m(\mathscr{Y}), \quad \mathscr{X}, \mathscr{Y} \in M_{nn}(I(\mathbb{R})),$$

(3) $$m(\mathscr{B}_{\mathrm{p}}\mathscr{X}) = \mathscr{B}_{\mathrm{p}} m(\mathscr{X}), \quad m(\mathscr{X}\mathscr{B}_{\mathrm{p}}) = m(\mathscr{X})\mathscr{B}_{\mathrm{p}},$$
$$\mathscr{B}_{\mathrm{p}} \in M_{nn}(\mathbb{R}), \quad \mathscr{X} \in M_{nn}(I(\mathbb{R})),$$

(4) $$m(\mathscr{B}_{\mathrm{p}}) = \mathscr{B}_{\mathrm{p}}, \quad \mathscr{B}_{\mathrm{p}} \in M_{nn}(\mathbb{R}).$$

Вот краткое доказательство соотношения (3):

$$m(\mathscr{B}_{\mathrm{p}}\mathscr{X}) = m\left(\sum_{k=1}^{n} b_{lk} X_{kj}\right) = \left(\sum_{k=1}^{n} m(b_{lk}\,[i(X_{kj}),\ s(X_{kj})])\right)$$
$$= \left(\sum_{k=1}^{n} b_{lk} \frac{1}{2}\,(i(X_{kj}) + s(X_{kj}))\right) = \left(\sum_{k=1}^{n} b_{lk} m(X_{kj})\right)$$
$$= \mathscr{B}_{\mathrm{p}} m(\mathscr{X}).$$

Утверждения (2) и (4) могут быть доказаны аналогично. Теперь мы хотим сформулировать первый метод, позволяющий вычислять последовательность локализаций для обратной матрицы $\mathscr{A}_{\mathrm{p}}^{-1}$. Пусть $r > 1$ — фиксированное натуральное число. Для произвольной точечной матрицы $\mathscr{B}_{\mathrm{p}}$ положим

$$\mathscr{B}_{\mathrm{p}}^{(0)} = \mathscr{I}_{\mathrm{p}} \text{ (где } \mathscr{I}_{\mathrm{p}} \text{ — единичная матрица).}$$

Рассмотрим итерационную процедуру

(5) $$\mathscr{X}^{(k+1)} = m(\mathscr{X}^{(k)}) \sum_{\nu=0}^{r-2} (\mathscr{I}_{\mathrm{p}} - \mathscr{A}_{\mathrm{p}} m(\mathscr{X}^{(k)}))^{\nu}$$
$$+ \mathscr{X}^{(k)} (\mathscr{I}_{\mathrm{p}} - \mathscr{A}_{\mathrm{p}} m(\mathscr{X}^{(k)}))^{r-1}, \quad k \geqslant 0.$$

В случае $r = 2$ мы получаем формулу

$$\mathscr{X}^{(k+1)} = m(\mathscr{X}^{(k)}) + \mathscr{X}^{(k)} (\mathscr{I}_{\mathrm{p}} - \mathscr{A}_{\mathrm{p}} m(\mathscr{X}^{(k)})), \quad k \geqslant 0,$$

которую можно считать интервальным вариантом метода Шульца [298] для вычисления обратной матрицы.

Свойства итерационного метода (5) собраны в следующем утверждении.

Теорема 1. *Пусть $\mathscr{A}_{\mathrm{p}}$ — невырожденная матрица размерности $n \times n$ и $\mathscr{X}^{(0)}$ — интервальная матрица той же размерности, такая что $\mathscr{A}_{\mathrm{p}}^{-1} \in \mathscr{X}^{(0)}$. Пусть последовательность $\{\mathscr{X}^{(k)}\}_{k=0}^{\infty}$ интервальных матриц вычисляется по формулам* (5). *Тогда*

(6) *каждое приближение $\mathscr{X}^{(k)}$, $k \geqslant 0$, содержит $\mathscr{A}_{\mathrm{p}}^{-1}$;*

(7) *последовательность $\{\mathscr{X}^{(k)}\}_{k=0}^{\infty}$ сходится к $\mathscr{A}_{\mathrm{p}}^{-1}$ тогда и только тогда, когда спектральный радиус $\rho(\mathscr{I}_{\mathrm{p}} - \mathscr{A}_{\mathrm{p}} m(\mathscr{X}^{(0)}))$ меньше* 1;

(8) *для матричной нормы* $\|\cdot\|$ *последовательность* $\{d(\mathscr{X}^k)\}_{k=0}^{\infty}$ *удовлетворяет условию* $\|d(\mathscr{X}^{(k+1)})\| \leqslant \gamma \|d(\mathscr{X}^{(k)})\|^r$, $\gamma \geqslant 0$, *т. е. R-порядок метода* (5) *удовлетворяет неравенству* $O_R((5), \mathscr{A}^{-1}) \geqslant r$ (*см. приложение A, теорема* 2).

Доказательство. (6): Для произвольной матрицы $m(\mathscr{X}^{(k)}) \in M_{nn}(\mathbb{R})$ легко проверить соотношение

$$m(\mathscr{X}^{(k)}) \sum_{\nu=0}^{r-2} (\mathscr{I}_p - \mathscr{A}_p m(\mathscr{X}^{(k)}))^\nu = \mathscr{A}_p^{-1} - \mathscr{A}_p^{-1} (\mathscr{I}_p - \mathscr{A}_p m(\mathscr{X}^{(k)}))^{r-1}.$$

Для $k = 0$ утверждение (6) верно в силу условия теоремы. Допустим теперь, что $\mathscr{A}_p^{-1} \in \mathscr{X}^{(k)}$. Используя только что полученное равенство и соотношения (10.10′), мы получаем

$$\mathscr{A}_p^{-1} = m(\mathscr{X}^{(k)}) \sum_{\nu=0}^{r-2} (\mathscr{I}_p - \mathscr{A}_p m(\mathscr{X}^{(k)}))^\nu + \mathscr{A}^{-1} (\mathscr{I}_p - \mathscr{A}_p m(\mathscr{X}^{(k)}))^{r-1}$$

$$\in m(\mathscr{X}^{(k)}) \sum_{\nu=0}^{r-2} (\mathscr{I}_p - \mathscr{A}_p m(\mathscr{X}^{(k)}))^\nu + \mathscr{X}^{(k)} (\mathscr{I}_p - \mathscr{A}_p m(\mathscr{X}^{(k)}))^{r-1}$$

$$= \mathscr{X}^{(k+1)}.$$

Этим завершается доказательство соотношения (6) методом математической индукции.

(7) Используя равенства (2)—(4) для срединного отображения, участвующего в формулах (5), мы получаем для последовательности $\{m(\mathscr{X}^{(k)})\}_{k=0}^{\infty}$ следующую рекуррентную формулу:

$$m(\mathscr{X}^{(k+1)}) = m(\mathscr{X}^{(k)}) \sum_{\nu=0}^{r-1} (\mathscr{I}_p - \mathscr{A}_p m(\mathscr{X}^{(k)}))^\nu.$$

Это — обобщение итерационной процедуры Шульца (см. [28]). Умножая обе части этого равенства на $\mathscr{A}_p$, мы получаем

$$\mathscr{A}_p m(\mathscr{X}^{(k+1)}) = (\mathscr{I}_p - (\mathscr{I}_p - \mathscr{A}_p m(\mathscr{X}^{(k)}))) \sum_{\nu=0}^{r-1} (\mathscr{I}_p - \mathscr{A}_p m(\mathscr{X}^{(k)}))^\nu$$

$$= \mathscr{I}_p - (\mathscr{I}_p - \mathscr{A}_p m(\mathscr{X}^{(k)}))^r$$

или

$$\mathscr{I}_p - \mathscr{A}_p m(\mathscr{X}^{(k+1)}) = (\mathscr{I}_p - \mathscr{A}_p m(\mathscr{X}^{(k)}))^r$$

$$= (\mathscr{I}_p - \mathscr{A}_p m(\mathscr{X}^{(0)}))^{r(k+1)}.$$

Отсюда следует, что

$$\lim_{k\to\infty} m(\mathscr{X}^{(k)}) = \mathscr{A}_p^{-1} \Leftrightarrow \lim_{k\to\infty} (\mathscr{I}_p - \mathscr{A}_p m(\mathscr{X}^{(0)}))^k = O_p$$

$$\Leftrightarrow \rho(\mathscr{I}_p - \mathscr{A}_p m(\mathscr{X}^{(0)})) < 1.$$

Теперь мы покажем, что последовательность $\{\mathscr{X}^{(k)}\}_{k=0}^{\infty}$ сходится к $\mathscr{A}_{\mathrm{p}}^{-1}$ тогда и только тогда, когда последовательность $\{m(\mathscr{X}^{(k)})\}_{k=0}^{\infty}$ срединных матриц сходится к $\mathscr{A}_{\mathrm{p}}^{-1}$. Действительно, рассмотрим последовательность $\{d(\mathscr{X}^{(k)})\}_{k=0}^{\infty}$, которая в силу (10.12) и (10.19) удовлетворяет рекуррентному соотношению

$$d(\mathscr{X}^{(k+1)}) = d(\mathscr{X}^{(k)}) \,|\, (\mathscr{I}_{\mathrm{p}} - \mathscr{A}_{\mathrm{p}} m(\mathscr{X}^{(k)}))^{r-1} \,|.$$

Если мы имеем теперь $\lim\limits_{k\to\infty} m(\mathscr{X}^{(k)}) = \mathscr{A}_{\mathrm{p}}^{-1}$, то из последнего соотношения следует, что $\lim\limits_{k\to\infty} d(\mathscr{X}^{(k)}) = \mathscr{O}_{\mathrm{p}}$. С другой стороны, из непрерывности отображения m и равенства (4) сразу получается, что $\lim\limits_{k\to\infty} \mathscr{X}^{(k)} = \mathscr{A}_{\mathrm{p}}^{-1}$ влечет за собой $\lim\limits_{k\to\infty} m(\mathscr{X}^{(k)}) = \mathscr{A}_{\mathrm{p}}^{-1}$. Так как выше уже было показано, что условие

$$\rho(\mathscr{I}_{\mathrm{p}} - \mathscr{A}_{\mathrm{p}} m(\mathscr{X}^{(0)})) < 1$$

необходимо и достаточно для сходимости последовательности $\{m(\mathscr{X}^{(k)})\}_{k=0}^{\infty}$, мы получаем (7).

(8): Мы имеем

$$\begin{aligned} d(\mathscr{X}^{(k+1)}) &= d(\mathscr{X}^{(k)}) \,|\, (\mathscr{I}_{\mathrm{p}} - \mathscr{A}_{\mathrm{p}} m(\mathscr{X}^{(k)}))^{r-1} \,| \\ &= d(\mathscr{X}^{(k)}) \,|\, (\mathscr{A}_{\mathrm{p}} \mathscr{A}_{\mathrm{p}}^{-1} - \mathscr{A}_{\mathrm{p}} m(\mathscr{X}^{(k)}))^{r-1} \,| \\ &\leqslant d(\mathscr{X}^{(k)}) \,(|\, \mathscr{A}_{\mathrm{p}} \,|\, |\, \mathscr{A}_{\mathrm{p}}^{-1} - m(\mathscr{X}^{(k)}) \,|\,)^{r-1} \\ &\leqslant d(\mathscr{X}^{(k)}) \, 2^{-(r-1)} \,(|\, \mathscr{A}_{\mathrm{p}} \,|\, d(\mathscr{X}^{(k)}))^{r-1}. \end{aligned}$$

Мы используем монотонную и мультипликативную матричную норму $\|\cdot\|'$, поэтому из только что доказанного соотношения следует

$$\| d(\mathscr{X}^{(k+1)}) \|' \leqslant 2^{-(r-1)} \| \mathscr{A}_{\mathrm{p}} \|^{(r-1)} \| d(\mathscr{X}^{(k)}) \|'^{r}.$$

Неравенство

$$\| \mathscr{B}_{\mathrm{p}} \| \gamma_1 \leqslant \| \mathscr{B}_{\mathrm{p}} \|' \leqslant \gamma_2 \| \mathscr{B}_{\mathrm{p}} \|, \quad \gamma_1 > 0, \ \gamma_2 > 0,$$

верно для любой матричной нормы $\|\cdot\|$. Из этого неравенства следует

$$\| d(\mathscr{X}^{(k+1)}) \| \gamma_1 \leqslant 2^{-(r-1)} \gamma_2^{-(r-1)} \| \mathscr{A}_{\mathrm{p}} \|^{r-1} \gamma_2^{n} \| d(\mathscr{X}^{(k)}) \|^{r},$$

что и доказывает (8). ■

Из доказательства видно, что сходимость имеет место для произвольной матрицы $\mathscr{X}^{(0)}$, не обязательно содержащей $\mathscr{A}_{\mathrm{p}}^{-1}$. В этом случае, однако, последовательные приближения не обя-

заны содержать $\mathscr{A}_{\mathrm{p}}^{-1}$. Отметим, что критерий (7) зависел не от всей локализующей матрицы $\mathscr{X}_{\mathrm{p}}^{(0)}$, а только от ее срединной матрицы $m(\mathscr{X}^{(0)})$. При этом ширина $a(\mathscr{X}^{(0)})$ может быть произвольной. Это значит, что имея подходящую аппроксимацию $m(\mathscr{X}^{(0)})$ матрицы $\mathscr{A}_{\mathrm{p}}^{-1}$, удовлетворяющую неравенству $\rho(\mathscr{I}_{\mathrm{p}} - \mathscr{A}_{\mathrm{p}} m(\mathscr{X}^{(0)})) < 1$, мы с помощью определенных оценок по норме всегда можем построить интервальную матрицу $\mathscr{X}^{(0)}$, такую что $\mathscr{A}_{\mathrm{p}}^{-1} \in \mathscr{X}^{(0)}$. Тогда последовательные приближения, полные согласно (5), сходятся к $\mathscr{A}_{\mathrm{p}}^{-1}$ по теореме 1.

Так как последовательные приближения из (5) всегда содержат $\mathscr{A}_{\mathrm{p}}^{-1}$ в силу (6), кажется естественным брать пересечение следующего приближения с предыдущим и продолжать итерационный процесс с этим новым потенциально улучшенным приближением. Это приводит к следующей итерационной процедуре:

$$\text{(9)} \qquad \begin{cases} \mathscr{Y}^{(k+1)} = m(\mathscr{X}^{(k)}) \sum\limits_{\nu=0}^{r-2} (\mathscr{I}_{\mathrm{p}} - \mathscr{A}_{\mathrm{p}} m(\mathscr{X}^{(k)}))^{\nu} \\ \qquad\qquad + \mathscr{X}^{(k)} (\mathscr{I}_{\mathrm{p}} - \mathscr{A}_{\mathrm{p}} m(\mathscr{X}^{(k)}))^{r-1}, \\ \mathscr{X}^{(k+1)} = \mathscr{Y}^{(k+1)} \cap \mathscr{X}^{(k)}, \quad k \geqslant 0. \end{cases}$$

Применяя эту итерационную процедуру, мы получаем монотонную последовательность $\mathscr{X}^{(0)} \supseteq \mathscr{X}^{(1)} \supseteq \mathscr{X}^{(2)} \supseteq \ldots$ локализаций для матрицы $\mathscr{A}_{\mathrm{p}}^{-1}$. Следующий численный пример показывает, что в этом случае критерий (7), вообще говоря, не достаточен для сходимости.

Возьмем $r = 2$ и положим

$$\mathscr{A}_{\mathrm{p}} = \begin{pmatrix} 0.4 & 0.6 \\ -0.6 & 0.4 \end{pmatrix}, \quad \mathscr{X}^{(0)} = \begin{pmatrix} [-2, 4] & [-3, 3] \\ [-3, 3] & [-2, 4] \end{pmatrix},$$

откуда следует, что $m(\mathscr{X}^{(0)}) = \mathscr{I}_{\mathrm{p}}$. Мы получаем

$$\mathscr{I}_{\mathrm{p}} - \mathscr{A}_{\mathrm{p}} m(\mathscr{X}^{(0)}) = \begin{pmatrix} 0.6 & -0.6 \\ 0.6 & 0.6 \end{pmatrix},$$

откуда

$$\rho(\mathscr{I}_{\mathrm{p}} - \mathscr{A}_{\mathrm{p}} m(\mathscr{X}^{(0)})) < 1.$$

Поэтому процедура (5) с этим начальным приближением сходится к $\mathscr{A}_{\mathrm{p}}^{-1}$. Используя (9), мы получаем

$$\mathscr{Y}^{(1)} = m(\mathscr{X}^{(0)}) + \mathscr{X}^{(0)} (\mathscr{I}_{\mathrm{p}} - \mathscr{A}_{p} m(\mathscr{X}^{(0)})) = \begin{pmatrix} [-2, 5.2] & [-4, 2.3] \\ [-3, 4.2] & [-2, 5.2] \end{pmatrix},$$

откуда следует, что $\mathscr{X}^{(1)} = \mathscr{X}^{(0)}$. Поэтому последовательность приближений, вычисленная по формулам (9), не сходится к $\mathscr{A}_{\mathrm{p}}^{-1}$

в противоположность последовательности, вычисленной по формулам (5). Условие сходимости для итерации (9) содержится в следующей теореме.

Теорема 2. *Пусть* $\mathscr{A}_p$ *— невырожденная матрица размерности* $n \times n$, *а* $\mathscr{X}^{(0)}$ *— интервальная матрица размерности* $n \times n$, *такая что* $\mathscr{A}_p^{-1} \in \mathscr{X}^{(0)}$. *Тогда*

(6′) *каждое приближение* $\mathscr{X}^{(k)}$, $k \geqslant 0$, *содержит* $\mathscr{A}_p^{-1}$;

(10) *если неравенство* $\rho(|\mathscr{I}_p - \mathscr{A}_p\mathscr{X}|) < 1$ *выполнено для всех* $\mathscr{X}_p \in \mathscr{X}^{(0)}$, *то последовательность* $\{\mathscr{X}^k\}_{k=0}^{\infty}$ *сходится к* $\mathscr{A}_p^{-1}$;

(8′) *последовательность* $\{d(\mathscr{X}^{(k)})\}_{k=0}^{\infty}$ *может быть следующим образом ограничена в матричной норме* $\|\cdot\|$:

$$\|d(\mathscr{X}^{(k+1)})\| \leqslant \gamma' \|d(\mathscr{X}^{(k)})\|^r, \quad \gamma \geqslant 0,$$

т. е. R-порядок итерационной процедуры (9) *удовлетворяет неравенству* $O_R((9), \mathscr{A}_p^{-1}) \geqslant r$ *(см. приложение* A, *теорема* 2).

Доказательство. (6′): Как и в доказательстве утверждения (6), мы устанавливаем сначала, что $\mathscr{A}_p^{-1} \in \mathscr{Y}^{(k+1)}$, откуда ввиду $\mathscr{A}_p^{-1} \in \mathscr{X}^{(k)}$ немедленно следует $\mathscr{A}_p^{-1} \in \mathscr{X}^{(k+1)}$.
(10): В силу следствия 10.8 последовательные приближения

$$\mathscr{X}^{(0)} \supseteq \mathscr{X}^{(1)} \supseteq \mathscr{X}^{(2)} \supseteq \ldots$$

всегда сходятся к некоторой интервальной матрице $\mathscr{X}$. Теперь мы покажем, что в условиях нашей теоремы выполнено равенство $d(\mathscr{X}) = \mathscr{O}_p$. Положив

$$\mathscr{Y} = m(\mathscr{X}) \sum_{\nu=0}^{r-2} (\mathscr{I}_p - \mathscr{A}_p m(\mathscr{X}))^\nu + \mathscr{X}(\mathscr{I}_p - \mathscr{A}_p m(\mathscr{X}))^{r-1},$$

получаем $\mathscr{X} = (X_{ij} \cap Y_{ij}) \subseteq \mathscr{Y}$ в силу (9). Используя (10.11), получим $d(\mathscr{X}) \leqslant d(\mathscr{Y})$. Для $d(\mathscr{X})$ имеем из (9) соотношение

$$d(\mathscr{X})|\mathscr{I}_p - \mathscr{A}_p m(\mathscr{X})|^{r-1} \geqslant d(\mathscr{X})|(\mathscr{I}_p - \mathscr{A}_p m(\mathscr{X}))^{r-1}| = d(\mathscr{Y}) \geqslant d(\mathscr{X}),$$

откуда следует, что

$$d(\mathscr{X})(\mathscr{I}_p - |\mathscr{I}_p - \mathscr{A}_p m(\mathscr{X})|^{r-1}) \leqslant \mathscr{O}_p.$$

Из условия $\rho(|\mathscr{I}_p - \mathscr{A}_p m(\mathscr{X})|) < 1$ следует существование матрицы $(\mathscr{I}_p - |\mathscr{I}_p - \mathscr{A}_p m(\mathscr{X})|^{r-1})^{-1}$. Эта обратная матрица также неотрицательна (ср. [325, стр. 83]). Отсюда следует, что $d(\mathscr{X}) \leqslant \mathscr{O}_p$, т. е. $d(\mathscr{X}) = \mathscr{O}_p$. Ввиду (6′) мы имеем поэтому $\mathscr{X} = \mathscr{A}_p^{-1}$.

(8′): Как и в доказательстве утверждения (8), мы показываем сначала, что для монотонной и мультипликативной матричной нормы $\|\cdot\|'$ имеет место неравенство

$$\|d(\mathcal{Y}^{(k+1)})\|' \leqslant \gamma \|d(\mathcal{X}^{(k)})\|'^{r}.$$

Отсюда с помощью (10.11), монотонности нормы $\|\cdot\|'$ и включения $\mathcal{X}^{(k+1)} \subseteq \mathcal{Y}^{(k+1)}$ следует неравенство

$$\|d(\mathcal{X}^{k+1})\|' \leqslant \|d(\mathcal{Y}^{(k+1)})\|' \leqslant \gamma \|d(\mathcal{X}^{(k)})\|'^{r}.$$

Так же как и в доказательстве утверждения (8), мы используем теперь теорему об эквивалентности норм для доказательства утверждения (8). ■

В отличие от критерия (7) условие сходимости (10) зависит от ширины матрицы $\mathcal{X}^{(0)}$, локализующей $\mathcal{A}_{\mathrm{p}}^{-1}$. Эту зависимость легко охарактеризовать формулами. Если, например, матрица $\mathcal{X}^{(0)}$ удовлетворяет для монотонной мультипликативной нормы $\|\cdot\|$ неравенству $\|\mathcal{I}_{\mathrm{p}} - \mathcal{A}_{\mathrm{p}} m(\mathcal{X}^{(0)})\| < 1$, то условие

$$\|d(\mathcal{X}^{(0)})\| < 2(1 - \|\mathcal{I}_{\mathrm{p}} - \mathcal{A}_{\mathrm{p}} m(\mathcal{X}^{(0)})\|)/\|\mathcal{A}_{\mathrm{p}}\| \tag{11}$$

достаточно для того, чтобы $\|\mathcal{I}_{\mathrm{p}} - \mathcal{A}_{\mathrm{p}} \mathcal{X}\| < 1$ было верно для всех $\mathcal{X}_{\mathrm{p}} \in \mathcal{X}^{(0)}$. Рассмотрим теперь вкратце вопрос о нахождении подходящей интервальной матрицы $\mathcal{X}^{(0)}$. Допустим, что $\mathcal{A}_{\mathrm{p}}$ можно представить в виде

$$\mathcal{A}_{\mathrm{p}} = \mathcal{I}_{\mathrm{p}} - \mathcal{B}_{\mathrm{p}}, \text{ где } \|\mathcal{B}_{\mathrm{p}}\| < 1.$$

При $m(\mathcal{X}^{(0)}) := \mathcal{I}_{\mathrm{p}}$ мы имеем

$$\|\mathcal{I}_{\mathrm{p}} - \mathcal{A}_{\mathrm{p}} m(\mathcal{X}^{(0)})\| = \|\mathcal{B}_{\mathrm{p}}\| < 1,$$

так что последовательность (5) сходится в силу критерия (7) для любой интервальной матрицы $\mathcal{X}^{(0)}$, для которой $m(\mathcal{X}^{(0)}) = \mathcal{I}_{\mathrm{p}}$. Чтобы обеспечить соотношение $\mathcal{A}_{\mathrm{p}}^{-1} \in \mathcal{X}^{(0)}$, рассмотрим равенство

$$\mathcal{A}_{\mathrm{p}} \mathcal{X}_{\mathrm{p}} = (\mathcal{I}_{\mathrm{p}} - \mathcal{B}_{\mathrm{p}}) \mathcal{X}_{\mathrm{p}} = \mathcal{I}_{\mathrm{p}}$$

или

$$\mathcal{X}_{\mathrm{p}} = \mathcal{B}_{\mathrm{p}} \mathcal{X}_{\mathrm{p}} + \mathcal{I}_{\mathrm{p}}.$$

Из него следует в силу мультипликативности матричной нормы $\|\cdot\|$, что

$$\|\mathcal{X}_{\mathrm{p}}\| \leqslant a := 1/(1 - \|\mathcal{B}_{\mathrm{p}}\|).$$

Если теперь мы используем норму, задаваемую суммами по столбцам или суммами по строкам, то получим

$$-a \leqslant x_{ij} \leqslant a, \quad 1 \leqslant i, \ j \leqslant n,$$

для элементов матрицы $\mathscr{X}_\mathrm{p} = (x_{ij})$. Для матрицы $\mathscr{X}^{(0)}$ с элементами

$$X_{ij}^{(0)} = \begin{cases} [-a, a] & \text{для } i \neq j, \\ [-a, 2+a] & \text{для } i = j \end{cases}$$

имеем $\mathscr{A}_\mathrm{p}^{-1} \in \mathscr{X}^{(0)}$ и $m(\mathscr{X}^{(0)}) = \mathscr{I}_\mathrm{p}$. Поэтому итерационный метод (5) сходится к $\mathscr{A}_\mathrm{p}^{-1}$ в силу теоремы 1. Если теперь неравенство (11) выполняется после некоторого шага итерации, то процесс можно продолжать дальше по формулам (9). Подробное описание такой комбинированной процедуры можно найти в [17].

При практическом выполнении алгоритма (5) встречающиеся выражения вычисляются по аналогии со схемой Горнера.

Это дает формулу

$$(5') \quad \mathscr{X}^{(k+1)} = \ldots (\mathscr{X}^{(k)} \mathscr{T}_\mathrm{p}^{(k)} + m(\mathscr{X}^{(k)})) \mathscr{T}_\mathrm{p}^{(k)} + m(\mathscr{X}^{(k)}) \mathscr{T}_\mathrm{p}^{(k)} \ldots) \mathscr{T}_\mathrm{p}^{(k)} + m(\mathscr{X}^{(k)}),$$

где

$$\mathscr{T}_\mathrm{p}^{(k)} = \mathscr{I}_\mathrm{p} - \mathscr{A}_\mathrm{p} m(\mathscr{X}_p^{(k)}).$$

Так как умножение матриц становится неассоциативным при появлении интервальных матриц (см. также гл. 10), мы имеем в общем случае

$$\mathscr{X}^{(k)} (\mathscr{I}_\mathrm{p} - \mathscr{A}_\mathrm{p} m(\mathscr{X}^{(k)}))^{r-1} \neq (\ldots (\mathscr{X}^{(k)} (\mathscr{I}_\mathrm{p} - \mathscr{A}_\mathrm{p} m(\mathscr{X}^{(k)}))) \ldots) \times (\mathscr{I}_\mathrm{p} - \mathscr{A}_\mathrm{p} m(\mathscr{X}^{(k)})).$$

Даже при одной и той же начальной матрице формулы (5) и (5′) порождают в общем случае разные последовательности. Однако теорема 1 все же верна и для итераций (5′) (см. [17]). Рассмотрим теперь объем вычислений, нужных на каждом шаге (5′).

Если $\mathscr{A}_\mathrm{p}$ — матрица размерности $n \times n$, то (5′) требует на каждом шаге

$$rn^3 \text{ умножений и } rn^3 - n^2 + n \text{ сложений.}$$

Даже те члены в (5′), которые, как $\mathscr{T}_\mathrm{p}^{(k)}$, не содержат ничего интервального, приходится вычислять в интервальной арифметике, чтобы обеспечить локализацию матрицы $\mathscr{A}_\mathrm{p}^{-1}$. Если пренебречь более низкими степенями n, то мы увидим, что объем вычислений по алгоритму (5′) пропорционален r.

Теперь мы хотим оценить число $\bar{k}$ шагов итерации, которые требуются, чтобы, исходя из данного $\mathscr{X}^{(0)}$, достичь величины $\|d(\mathscr{X}^{(\bar{k})})\|$, меньшей, чем заранее предписанная погрешность.

Так же, как в доказательстве теоремы 1, мы получаем для (5′) следующее соотношение:

$$d(\mathscr{X}^{(k+1)} = d(\mathscr{X}^{(k)})\,|(\mathscr{I}_p - \mathscr{A}_p m(\mathscr{X}^{(k)}))\,|^{r-1}$$
$$\leqslant d(\mathscr{X}^{(k)})\,|(\mathscr{I}_p - \mathscr{A}_p m(\mathscr{X}^{(0)}))\,|^{r^k(r-1)},$$

т. е.

$$d(\mathscr{X}^{(k+1)}) \leqslant d(\mathscr{X}^{(0)} \prod_{\nu=0}^{k} |(\mathscr{I}_p - \mathscr{A}_p m(\mathscr{X}^{(0)}))\,|^{r^\nu (r-1)}.$$

Если по-прежнему мы используем монотонную и мультипликативную матричную норму $\|\cdot\|$ и допустим, что $\|\mathscr{I}_p - \mathscr{A}_p m(\mathscr{X}^{(0)})\| < 1$, то получим, что

$$\|d(\mathscr{X}^{(k+1)})\| \leqslant \|d(\mathscr{X}^{(0)})\| \left(\prod_{\nu=0}^{k} \|\mathscr{I}_p - \mathscr{A}_p m(\mathscr{X}^{(0)})\|^{r^\nu}\right)^{r-1}$$
$$= \|d(\mathscr{X}^{(0)})\| \|\mathscr{I}_p - \mathscr{A}_p m(\mathscr{X}^{(0)})\|^{r^{k+1}-1}.$$

Это выражение позволяет нам оценить $\bar{k}$ при сделанных предположениях.

Исходя из соотношения

$$\|d(\mathscr{X}^{(\bar{k})})\| \leqslant \|d(\mathscr{X}^{(0)})\| \|\mathscr{I}_p - \mathscr{A}_p m(\mathscr{X}^{(0)})\|^{r^{\bar{k}}-1},$$

мы определим, при каком значении r итерационный метод требует наименьшего объема вычислений для достижения заданной точности для $\|d(\mathscr{X}^{(k)})\|$. Согласно предшествующим рассуждениям, этот объем вычислений можно считать пропорциональным величине r. Пусть теперь даны $r^{(1)} > 1$ и $r^{(2)} > 1$, причем $r^{(1)} \neq r^{(2)}$. После $p^{(1)}$ (соответственно $p^{(2)}$) шагов итерации (5′) со значением $r = r^{(1)}$ (соответственно $r = r^{(2)}$) при одном и том же начальном значении $\mathscr{X}^{(0)}$ мы выполним один и тот же объем вычислений. Иными словами, мы имеем

$$r^{(1)}p^{(1)} = r^{(2)}p^{(2)}.$$

Точность, достигнутую при использовании этих методов, можно оценить величиной $(r^{(1)})^{p^{(1)}} - 1$ (соответственно $(r^{(2)})^{p^{(2)}} - 1$). Мы требуем от «оптимальной» итерационной процедуры $r = r^{(1)}$, чтобы для всех других значений $r^{(2)}$ и количества шагов $p^{(1)}$, $p^{(2)}$ мы имели бы

$$(r^{(1)})^{p^{(1)}} > (r^{(2)})^{p^{(2)}},$$

что ввиду $p^{(2)} = r^{(1)}p^{(1)}/r^{(2)}$ эквивалентно неравенству

$$(r^{(1)})^{1/r^{(1)}} > (r^{(2)})^{1/r^{(2)}}.$$

Так как функция $x^{1/x}$ для натуральных x имеет максимум при $x = 3$, мы получаем, что итерация (5′) оптимальна в описанном смысле при этом значении.

Заметим еще, что метод (5) можно применять и для комплексных матриц, используя арифметику в $R(\mathbb{C})$. При этом будет верна теорема 1. Мы упомянем в этой связи более общие исследования, проведенные в [19], а также статью [281].

Рассмотрим теперь описанные в [211] методы монотонной локализации обратной матрицы, обладающие свойствами, похожими на свойства метода (9). Основные вычисления в них вообще не используют интервальной арифметики. Верхняя и нижняя границы вычисляются по отдельным формулам. Этот метод, однако, применим лишь в случае, когда $\mathscr{A}_{\mathrm{p}}^{-1} \geqslant \mathscr{O}_{\mathrm{p}}$.

Итак, пусть $\mathscr{A}_{\mathrm{p}}$ — невырожденная матрица и $r \geqslant 2$ — натуральное число. Рассмотрим итерационную процедуру

$$(12)\begin{cases} \mathscr{X}_{\mathrm{p}}^{(k+1)} = \mathscr{X}_{\mathrm{p}}^{(k)} + (\mathscr{I}_{\mathrm{p}} - \mathscr{X}_{\mathrm{p}}^{(k)}\mathscr{A}_{\mathrm{p}}) \sum\limits_{\nu=0}^{r-2} (\mathscr{I}_{\mathrm{p}} - \mathscr{X}_{\mathrm{p}}^{(k)}\mathscr{A}_{\mathrm{p}})^{\nu} \mathscr{X}_{\mathrm{p}}^{(k)}, \\ \mathscr{Y}_{\mathrm{p}}^{(k+1)} = \mathscr{Y}_{\mathrm{p}}^{(k)} + (\mathscr{I}_{\mathrm{p}} - \mathscr{Y}_{\mathrm{p}}^{(k)}\mathscr{A}_{\mathrm{p}}) \sum\limits_{\nu=0}^{r-2} (\mathscr{I}_{\mathrm{p}} - \mathscr{X}_{\mathrm{p}}^{(k)}\mathscr{A}_{\mathrm{p}})^{\nu} \mathscr{X}_{\mathrm{p}}^{(k)}, \\ k \geqslant 0. \end{cases}$$

с заданными $\mathscr{X}_{\mathrm{p}}^{(0)}$, $\mathscr{Y}_{\mathrm{p}}^{(0)}$.

Исполняя эту процедуру, мы получим две последовательности точечных матриц, для которых верно следующее утверждение.

Теорема 3. *Пусть $\mathscr{A}_{\mathrm{p}}$ — невырожденная матрица размерности $n \times n$, причем $\mathscr{A}_{\mathrm{p}}^{-1} \geqslant \mathscr{O}_{\mathrm{p}}$. Пусть далее $\mathscr{X}_{\mathrm{p}}^{(0)}$, $\mathscr{Y}_{\mathrm{p}}^{(0)}$ — две матрицы размерности $n \times n$, для которых верно*

$$\mathscr{X}_{\mathrm{p}}^{(0)} \geqslant \mathscr{O}_{\mathrm{p}} \text{ и } \mathscr{X}_{\mathrm{p}}^{(0)}\mathscr{A}_{\mathrm{p}} \leqslant \mathscr{I}_{\mathrm{p}} \leqslant \mathscr{Y}_{\mathrm{p}}^{(0)}\mathscr{A}_{\mathrm{p}}.$$

Пусть последовательность $\{\mathscr{X}_{\mathrm{p}}^{(k)}\}_{k=0}^{\infty}$ и $\{\mathscr{Y}_{\mathrm{p}}^{(k)}\}_{k=0}^{\infty}$ вычислены по формулам (12). *Тогда верны следующие утверждения:*

(13) $$\mathscr{O}_{\mathrm{p}} \leqslant \mathscr{X}_{\mathrm{p}}^{(0)} \leqslant \ldots \leqslant \mathscr{X}_{\mathrm{p}}^{(k)} \leqslant \mathscr{X}_{\mathrm{p}}^{(k+1)} \leqslant \ldots \leqslant \mathscr{A}_{\mathrm{p}}^{-1} \leqslant \ldots \leqslant \mathscr{Y}_{\mathrm{p}}^{(k+1)} \leqslant \mathscr{Y}_{\mathrm{p}}^{(k)} \leqslant \ldots \leqslant \mathscr{Y}_{\mathrm{p}}^{(0)}.$$

(7′) *Обе последовательности $\{\mathscr{X}_{\mathrm{p}}^{(k)}\}_{k=0}^{\infty}$, $\{\mathscr{Y}^{(k)}\}_{k=0}^{\infty}$ сходятся к $\mathscr{A}_{\mathrm{p}}^{-1}$ тогда и только тогда, когда спектральный радиус $\rho(\mathscr{I}_{\mathrm{p}} - \mathscr{X}_{\mathrm{p}}^{(0)}\mathscr{A}_{\mathrm{p}})$ меньше* 1.

Если процедура сходится, то величины

$$d^{(k)} = \| \mathscr{Y}_{\mathrm{p}}^{(k)} - \mathscr{X}_{\mathrm{p}}^{(k)} \|$$

удовлетворяют соотношению

(14) $$d^{(k+1)} \leqslant \gamma (d^{(k)})^{r}, \quad \gamma \geqslant 0.$$

Поэтому, если понимать (12) *как метод итераций для вычисления интервальных матриц* $([x_{ij}^{(k)}, y_{ij}^{(k)}])$, *то верно*

$$O_R((12), \mathscr{A}_p^{-1}) \geqslant r$$

(*см. приложение* A, *теорема* 2).

Доказательство. (13): Мы докажем соотношения

$$\mathscr{O}_p \leqslant \mathscr{X}_p^{(0)} \leqslant \ldots \leqslant \mathscr{X}_p^{(k-1)} \leqslant \mathscr{X}_p^{(k)},$$

$$\mathscr{Y}_p^{(k)} \leqslant \mathscr{Y}_p^{(k-1)} \leqslant \ldots \leqslant \mathscr{Y}_p^{(0)},$$

$$\mathscr{X}_p^{(k)} \mathscr{A}_p \leqslant \mathscr{I}_p \leqslant \mathscr{Y}_p^{(k)} \mathscr{A}_p$$

математической индукцией по $k \geqslant 0$. Эти неравенства выполнены для $k = 0$ по условию теоремы. Из

$$\mathscr{I}_p - \mathscr{Y}_p^{(k)} \mathscr{A}_p \leqslant \mathscr{O}_p \leqslant \mathscr{I}_p - \mathscr{X}_p^{(k)} \mathscr{A}_p, \quad \mathscr{X}_p^{(k)} \geqslant \mathscr{O}_p$$

следует, что

$$(\mathscr{I}_p - \mathscr{Y}_p^{(k)} \mathscr{A}_p) \sum_{\nu=0}^{r-2} (\mathscr{I}_p - \mathscr{X}_p^{(k)} \mathscr{A}_p)^\nu \mathscr{X}_p^{(k)}$$

$$\leqslant \mathscr{O}_p \leqslant (\mathscr{I}_p - \mathscr{X}_p^{(k)} \mathscr{A}_p) \sum_{\nu=0}^{r-2} (\mathscr{I}_p - \mathscr{X}_p^{(k)} \mathscr{A}_p)^\nu \mathscr{X}_p^{(k)},$$

т. е.

$$\mathscr{X}_p^{(k)} \leqslant \mathscr{X}_p^{(k+1)}, \quad \mathscr{Y}_p^{(k+1)} \leqslant \mathscr{Y}_p^{(k)}.$$

Из

$$\mathscr{I}_p - \mathscr{X}_p^{(k+1)} \mathscr{A}_p = \mathscr{I}_p - \mathscr{X}_p^{(k)} \mathscr{A}_p - (\mathscr{I}_p - \mathscr{X}_p^{(k)} \mathscr{A}_p)$$

$$\times \sum_{\nu=0}^{r-2} (\mathscr{I}_p - \mathscr{X}_p^{(k)} \mathscr{A}_p)^\nu \mathscr{X}_p^{(k)} \mathscr{A}_p = (\mathscr{I}_p - \mathscr{X}_p^{(k)} \mathscr{A}_p)^r \geqslant \mathscr{O}_p$$

и

$$\mathscr{Y}_p^{(k+1)} \mathscr{A}_p - \mathscr{I}_p = \mathscr{Y}_p^{(k)} \mathscr{A}_p - \mathscr{I}_p - (\mathscr{Y}_p^{(k)} \mathscr{A}_p - \mathscr{I}_p)$$

$$\times \sum_{\nu=0}^{r-2} (\mathscr{I}_p - \mathscr{X}_p^{(k)} \mathscr{A}_p)^\nu \mathscr{X}_p^{(k)} \mathscr{A}_p = (\mathscr{Y}_p^{(k)} \mathscr{A}_p - \mathscr{I}_p)(\mathscr{I}_p - \mathscr{X}_p^{(k)} \mathscr{A}_p)^{r-1} \geqslant \mathscr{O}_p$$

следует, что

$$\mathscr{X}_p^{(k+1)} \mathscr{A}_p \leqslant \mathscr{I}_p \leqslant \mathscr{Y}_p^{(k+1)} \mathscr{A}_p.$$

Ввиду $\mathscr{A}_p^{-1} \geqslant \mathscr{O}_p$ мы получаем

$$\mathscr{X}_p^{(k+1)} \leqslant \mathscr{A}_p^{-1} \leqslant \mathscr{Y}_p^{(k+1)}.$$

(7′): Используя соотношения

$$\mathscr{I}_p - \mathscr{X}_p^{(k+1)} \mathscr{A}_p = (\mathscr{I}_p - \mathscr{X}_p^{(k)} \mathscr{A}_p)^r,$$

$$\mathscr{Y}_p^{(k+1)} \mathscr{A}_p - \mathscr{I}_p = (\mathscr{Y}_p^{(k)} \mathscr{A}_p - \mathscr{I}_p)(\mathscr{I}_p - \mathscr{X}_p^{(k)} \mathscr{A}_p)^{r-1},$$

установленные при доказательстве неравенства (13), мы можем показать по индукции, что

$$\mathscr{Y}_p - \mathscr{X}_p^{(k)}\mathscr{A}_p = (\mathscr{Y}_p - \mathscr{X}_p^{(0)}\mathscr{A}_p)^{r^k},$$

$$\mathscr{Y}_p^{(k)}\mathscr{A}_p - \mathscr{Y}_p = (\mathscr{Y}_p^{(0)}\mathscr{A}_p - \mathscr{Y}_p)(\mathscr{Y}_p - \mathscr{X}_p^{(0)}\mathscr{A}_p)^{r^k-1},$$

откуда и следует нужное утверждение.

(14): Снова используя соотношения, установленные при доказательстве (13), мы получаем

$$\|\mathscr{A}_p^{-1} - \mathscr{X}_p^{(k+1)}\| \leqslant \|\mathscr{A}_p\|^{r-1}\|\mathscr{A}_p^{-1} - \mathscr{X}_p^{(k)}\|^r,$$

$$\|\mathscr{Y}_p^{(k+1)} - \mathscr{A}_p^{-1}\| \leqslant \|\mathscr{A}_p\|^{r-1}\|\mathscr{Y}_p^{(k)} - \mathscr{A}_p^{-1}\|\,\|\mathscr{A}_p^{-1} - \mathscr{X}_p^{(k)}\|^{r-1}.$$

С помощью монотонной матричной нормы эту оценку можно продолжить следующим образом:

$$\begin{aligned}\|\mathscr{Y}_p^{(k+1)} - \mathscr{X}_p^{(k+1)}\| &\leqslant \|\mathscr{A}_p^{-1} - \mathscr{X}_p^{(k+1)}\| + \|\mathscr{Y}_p^{(k+1)} - \mathscr{A}_p^{-1}\| \\ &\leqslant 2\|\mathscr{A}_p\|^{r-1}\|\mathscr{Y}_p^{(k)} - \mathscr{X}_p^{(k)}\|^r,\end{aligned}$$

т. е. $d^{(k+1)} \leqslant \gamma (d^{(k)})^r$, $\gamma = 2\|\mathscr{A}_p\|^{r-1}$. Теперь нужное соотношение следует из теоремы 2 приложения А. ■

В методе (12) интервальные операции не используются. Несмотря на это, он порождает, подобно методу (9), монотонную последовательность границ для матрицы $\mathscr{A}_p^{-1}$. Был дан также необходимый и достаточный критерий сходимости, аналогичный критерию для метода (5). Однако применимость метода (12) ограничивается требованием $\mathscr{A}_p^{-1} \geqslant \mathscr{O}_p$.

Для определенных классов матриц можно указать вполне общие начальные значения $\mathscr{X}_p^{(0)}$, $\mathscr{Y}_p^{(0)}$, при которых (12) всегда будет сходиться. Если матрица $\mathscr{A}_p = (a_{ij})$ удовлетворяет условиям

$$\begin{aligned} &a_{ii} > 0, \quad 1 \leqslant i \leqslant n, \\ &a_{ij} \leqslant 0, \quad 1 \leqslant i, j \leqslant n, \quad i \neq j, \\ &\sum_{i=1}^{n} a_{ij} > 0, \quad 1 \leqslant j \leqslant n, \end{aligned}$$

то $\mathscr{A}_p$ является M-матрицей (см. определение в гл. 15, а также [325, стр. 85 и далее]). Начальные матрицы $\mathscr{X}_p^{(0)} = (x_{ij})$ и $\mathscr{Y}_p^{(0)} = (y_{ij})$, такие что

$$x_{ij} = \begin{cases} 1/a_{ii} \text{ для } i = j \\ 0 \quad \text{в противном случае,} \end{cases} \text{ и } y_{ij} = 1 \Big/ \sum_{\nu=1}^{n} a_{\nu i}, \; 1 \leqslant i, j \leqslant n,$$

удовлетворяют условиям теоремы 3.

Рассмотрим теперь невырожденную вещественную матрицу $\mathscr{A}_p = (a_{ij})$ размерности $n \times n$, строки которой были переставлены, чтобы стало возможным разложение

$$\mathscr{A}_p = (\mathscr{I}_p + \mathscr{L}_p^*)\,\mathscr{U}_p^*, \tag{15}$$

где $\mathscr{I}_p$ — единичная матрица, $\mathscr{L}_p^*$ — строго нижняя треугольная матрица, $\mathscr{U}_p^*$ — верхняя треугольная матрица. Как хорошо известно (см., например, [521a]), это разложение можно найти с помощью метода Гаусса. Мы хотим описать итерационную структуру, постепенно улучшающую границы, между которыми заключены элементы матриц $\mathscr{L}_p^*$ и $\mathscr{U}_p^*$. Такие границы, включающие все ошибки округления, можно вычислить, если использовать при исполнении метода Гаусса для точечной матрицы $\mathscr{A}_p$ машинную интервальную арифметику с округлением наружу.

Допустим поэтому, что $\mathscr{L}^{(0)}$ — строго нижняя треугольная интервальная матрица и $\mathscr{U}^{(0)}$ — верхняя треугольная матрица, для которых

$$\mathscr{L}_p^* \in \mathscr{L}^{(0)}, \quad \mathscr{U}_p^* \in \mathscr{U}^{(0)}. \tag{16}$$

Предположим сначала, что $\mathscr{L}_p$ и $\mathscr{U}_p$ — произвольные, но фиксированные треугольные матрицы, причем $\mathscr{L}_p$ — строго нижняя, а $\mathscr{U}_p$ — верхняя. Тогда мы имеем из (15)

$$\mathscr{R}_p = (r_{ik}) := \mathscr{A}_p - (\mathscr{I}_p + \mathscr{L}_p)\,\mathscr{U}_p = (\mathscr{I}_p + \mathscr{L}_p^*)\,(\mathscr{U}_p^* - \mathscr{U}_p) + (\mathscr{L}_p^* - \mathscr{L}_p)\,\mathscr{U}_p.$$

Будем считать матрицы $\mathscr{U}_p^* - \mathscr{U}_p$ и $\mathscr{L}_p^* - \mathscr{L}_p$ неизвестными, а множитель $\mathscr{I}_p + \mathscr{L}_p^*$ перед $\mathscr{U}_p^* - \mathscr{U}_p$ известным и будем затем решать это уравнение путем поочередного нахождения сначала строки матрицы $\mathscr{U}_p^* - \mathscr{U}_p$, а затем столбца матрицы $\mathscr{L}_p^* - \mathscr{L}_p$. В предложении $u_{ii} = 0$ для $1 \leqslant i \leqslant n$ мы получаем тогда

$$\begin{cases} u_{ik}^* - u_{ik} = r_{ik} - \sum_{j=1}^{i-1} l_{ij}^* (u_{jk}^* - u_{jk}) - \sum_{j=1}^{i-1} (l_{ij}^* - l_{ij})\,u_{jk}, & i \leqslant k \leqslant n, \\ l_{ki}^* - l_{ki} = \dfrac{1}{u_{ii}} \left\{ r_{ki} - \sum_{j=1}^{i} l_{kj}^* (u_{ji}^* - u_{ji}) - \sum_{j=1}^{i-1} (l_{kj}^* - l_{kj})\,u_{ji} \right\}, & i < k \leqslant n, \\ 1 \leqslant i \leqslant n. \end{cases}$$

Полагая

$$\left.\begin{aligned} r_{ik} &= a_{ik} - \sum_{j=1}^{i-1} l_{ij}u_{jk}, - u_{ik}, \quad i \leqslant k \leqslant n, \\ r_{ki} &= a_{ki} - \sum_{j=1}^{i} l_{kj}u_{ji}, \qquad\qquad i+1 \leqslant k \leqslant n, \end{aligned}\right\} \; 1 \leqslant i \leqslant n,$$

мы получаем тогда соотношения

$$\begin{cases} u^*_{ik} = a_{ik} - \sum_{j=1}^{i-1} l^*_{ij}u_{jk} - \sum_{j=1}^{i-1} \boldsymbol{l}^*_{ij}\,(u^*_{jk} - u_{jk}), \quad i \leqslant k \leqslant n, \\ l^*_{ki} = \frac{1}{u_{ii}} \left\{ a_{ki} - \sum_{j=1}^{i-1} l^*_{kj}\,u_{ji} - \sum_{j=1}^{i} \boldsymbol{l}^*_{kj}\,(u^*_{ji} - u_{ji}) \right\}, \quad i < k < n, \\ 1 \leqslant i \leqslant n. \end{cases}$$

Выделенные жирным шрифтом выражения происходят из множителя $\mathscr{I}_{\mathrm{p}} + \mathscr{L}^*_{\mathrm{p}}$, который мы предположили известным. Если воспользоваться предположением, что $\mathscr{L}^*_{\mathrm{p}} \in \mathscr{L}^{(0)}$ для этих величин и выбрать $\mathscr{U}_{\mathrm{p}} := \mathscr{U}^0_{\mathrm{p}} \in \mathscr{U}^{(0)}$, где $u^{(0)}_{ii} \neq 0$, $1 \leqslant i \leqslant n$, то получим для элементов матриц $\mathscr{L}^*_{\mathrm{p}}$ и $\mathscr{U}^*_{\mathrm{p}}$ следующие локализующие интервалы:

$$\begin{cases} u^*_{ik} \in U^{(1)}_{ik} = \left\{ a_{ik} - \sum_{j=1}^{i-1} L^{(1)}_{ij}u^{(0)}_{jk} - \sum_{j=1}^{i-1} L^{(0)}_{ij}\,(U^{(1)}_{jk} - u^{(0)}_{jk}) \right\} \cap U^{(0)}_{ik}, \\ \qquad\qquad 1 \leqslant k \leqslant n, \\ l^*_{ki} \in L^{(1)}_{ki} = \left\{ \frac{1}{u^{(0)}_{ii}} \left(a_{ki} - \sum_{j=1}^{i-1} L^{(1)}_{kj}u^{(0)}_{ji} - \sum_{j=1}^{i} L^{(0)}_{kj}\,(U^{(1)}_{ji} - u^{(0)}_{ij}) \right) \right\} \\ \qquad\qquad \cap L^{(0)}_{ki}, \quad i < k \leqslant n, \\ 1 \leqslant i \leqslant n. \end{cases}$$

Систематическое повторение этой процедуры приводит к следующему итерационному методу:

$$\begin{cases} U^{(m+1)}_{ik} = \left\{ a_{ik} - \sum_{j=1}^{i-1} L^{(m+1)}_{ij}u^{(m)}_{jk} - \sum_{j=1}^{i-1} L^{(m)}_{ij}\,(U^{(m+1)}_{jk} - u^{(m)}_{jk}) \right\} \cap U^{(m)}_{ik}, \\ \qquad\qquad i \leqslant k \leqslant n, \\ L^{(m+1)}_{ki} = \left\{ \frac{1}{u^{(m)}_{ii}} \left(a_{ki} - \sum_{j=1}^{i-1} L^{(m+1)}_{kj}u^{(m)}_{ji} - \sum_{j=1}^{i} L^{(m)}_{kj}\,(U^{(m+1)}_{ji} - u^{(m)}_{ji}) \right) \right\} \\ \qquad\qquad \cap L^{(m)}_{kj}, \quad i < k \leqslant n, \\ 1 \leqslant i \leqslant n, \\ m \geqslant 0. \end{cases}$$

С помощью рассуждений, аналогичных проведенным при описании первого шага этого метода, можно показать, что в общем случае верно следующее утверждение.

Теорема 4. *Пусть матрица* $\mathscr{A}_p$ *имеет разложение* $\mathscr{A}_p = (\mathscr{I}_p + \mathscr{L}_p^*)\mathscr{U}_p^*$. *Пусть* $\mathscr{L}_p^* \in \mathscr{L}^{(0)}$, $\mathscr{U}_p^* \in \mathscr{U}^{(0)}$. *Если* $\mathscr{L}_p^{(m)} \in \mathscr{L}^{(m)}$, $\mathscr{U}_p^{(m)} \in \mathscr{U}^{(m)}$, *то для всех* $m \geqslant 0$ *верно*

$$\mathscr{L}_p^* \in \mathscr{L}^{(m+1)}, \quad \mathscr{U}_p^* \in \mathscr{U}^{(m+1)}. \quad \blacksquare \tag{18}$$

Мы отметим, что метод (17) может быть выполнен при начальных интервалах произвольной, но конечной ширины, если выполнено предположение (16). Единственное деление встречается при вычислении $L_{ki}^{(m+1)}$. Так как $0 \neq u_{ii}^* \in U_{ii}^{(m)}$, мы всегда можем выбрать $u_{ii}^{(m)} \in U_{ii}^{(m)}$ так, чтобы $u_{ii}^{(m)} \neq 0$.

Мы покажем теперь, что метод (17) при отсутствии ошибок округления дает точное треугольное разложение за конечное число шагов.

Теорема 5. *В условиях предыдущей теоремы метод* (17) *вычисляет точное разложение матрицы* $\mathscr{A}_p$ *размерности* $n \times n$ *самое большее за* $2n - 1$ *шагов.*

Доказательство. При $m = 0$ мы получаем из (17) для $i = 1$, что

$$U_{1k}^{(1)} = a_{1k} \cap U_{1k}^{(0)} = a_{1k} = u_{1k}^*, \quad 1 \leqslant k \leqslant n.$$

Это значит, что для произвольных начальных матриц, удовлетворяющих условиям локализации (16), матрица $\mathscr{U}^{(1)}$ имеет в первой строке значения, нужные для треугольного разложения. Матрица $\mathscr{U}^{(2)}$ имеет нужные значения в первой строке точно так же, как $\mathscr{U}^{(1)}$. Поэтому для первого столбца матрицы $\mathscr{L}^{(2)}$ мы получаем из (17) при $m = 1$, $i = 1$, что

$$L_{k1}^{(2)} = \left(\frac{a_{k1}}{u_{11}}\right) \cap L_{k1}^{(1)} = \frac{a_{k1}}{a_{11}} = l_{k1}^*, \quad 2 \leqslant k \leqslant n.$$

Поэтому после второго шага итерации нужные значения имеют и первый столбец матрицы $\mathscr{L}^{(2)}$, и первая строка матрицы $\mathscr{U}^{(2)}$. Мы покажем теперь, что если первые i строк матрицы $\mathscr{U}^{(m)}$ и первые i столбцов матрицы $\mathscr{L}^{(m)}$ уже имеют нужные значения, то не менее чем $i + 1$ строк матрицы $\mathscr{U}^{(m+1)}$ (и i столбцов матрицы $\mathscr{L}^{(m+1)}$) имеют нужные значения. Предыдущее рас-

суждение показывает, что это верно при $i = 0$. Для $i > 0$ применим математическую индукцию.

Заметим прежде всего, что $\mathscr{L}^{(m+1)}$ и $\mathscr{U}^{(m+1)}$ имеют те же элементы, что $\mathscr{L}^{(m)}$ и $\mathscr{U}^{(m)}$ в первых i строках и столбцах. Таким образом эти элементы все еще имеют нужные значения. Это непосредственно следует из (17). Из (17) мы получаем также следующее соотношение для $(i+1)$-й строки матрицы $\mathscr{U}^{(m+1)}$:

$$U_{i+1,k}^{(m+1)} = \left\{ a_{i+1,k} - \sum_{j=1}^{i-1} l_{i+1,j}^{*} u_{jk}^{*} \right\} \cap U_{i+1,k}^{(m)} = u_{i+1,k}^{*}, \quad i+1 \leqslant k \leqslant n.$$

Чтобы завершить доказательство по индукции, мы должны показать, что если для некоторого $m \geqslant 0$ первые i строк матрицы $\mathscr{U}^{(m)}$ и первые $i-1$ столбцов матрицы $\mathscr{L}^{(m)}$ имеют нужные значения, то $\mathscr{L}^{(m+1)}$ имеют нужные значения в первых столбцах (а $\mathscr{U}^{(m+1)}$ — в первых i строках).

Это было доказано выше для $i = 1$. Для $i > 1$ мы заметим, что в силу (17) матрица $\mathscr{U}^{(m+1)}$ имеет нужные значения по крайней мере в тех же строках, что и $\mathscr{U}^{(m)}$, а матрица $\mathscr{L}^{(m+1)}$ — по крайней мере в тех же строках, что $\mathscr{L}^{(m)}$. Тогда для i-го столбца матрицы $\mathscr{L}^{(m+1)}$ мы получаем из (17), что

$$L_{ki}^{(m+1)} = \left\{ \frac{1}{u_{ii}^{*}} \left(a_{ki} - \sum_{j=1}^{i-1} l_{kj}^{*} u_{ji}^{*} \right) \right\} \cap L_{ki}^{(m)} = l_{ki}^{*}, \quad i+1 \leqslant k \leqslant n.$$

Таким образом, самое большее через $2n-1$ шаг мы получим точное решение. ■

Следующая теорема показывает, что наш метод обладает так называемым «квадратичным» свойством сходимости, хорошо знакомым по методу Ньютона — Рафсона.

Теорема 6. *Пусть* $d^{(m)} := \max_{1 \leqslant i,j \leqslant n} \{\max \{d(L_{ij}^{(m)}), d(U_{ij}^{(m)})\}\}$.

Тогда для метода (17) *имеет место*

$$d^{(m+1)} \leqslant \alpha (d^{(m)})^2,$$

где α *— неотрицательное вещественное число, не зависящее от* m, *т. е. на каждом шаге ширина интервала примерно возводится в квадрат.*

Доказательство (методом математической индукции). Из (17) мы получаем

$$(17')\quad \begin{cases} d\left(U_{ik}^{(m+1)}\right) \leqslant \sum_{j=1}^{i-1} d\left(L_{ij}^{(m+1)}\right) \left|u_{jk}^{(m)}\right| + \sum_{j=1}^{i-1} d\left(L_{ij}^{(m)}\right) \left|U_{jk}^{(m+1)} - u_{jk}^{(m)}\right| \\ \qquad + \sum_{j=1}^{i-1} \left|L_{ij}^{(m)}\right| d\left(U_{jk}^{(m+1)}\right), \quad i \leqslant k \leqslant n, \\ d\left(L_{ki}^{(m+1)}\right) \leqslant \frac{1}{\left|u_{ii}^{(m)}\right|} \left\{ \sum_{j=1}^{i-1} d\left(L_{kj}^{(m+1)}\right) \left|u_{ji}^{(m)}\right| \right. \\ \qquad + \sum_{j=1}^{i} d\left(L_{kj}^{(m)}\right) \left|U_{ji}^{(m+1)} - u_{ji}^{(m)}\right| \\ \qquad \left. + \sum_{j=1}^{i} \left|L_{kj}^{(m)}\right| d\left(U_{ji}^{(m+1)}\right) \right\}, \quad i < k \leqslant n, \\ 1 \leqslant i \leqslant n. \end{cases}$$

Положим теперь при $0 \notin U_{ii}^{(0)}$, $1 \leqslant i \leqslant n$,

$$(17'')\quad \begin{cases} \alpha_{ik} = \begin{cases} \sum_{j=1}^{i-1} \left(\beta_{ij} \left|U_{jk}^{(0)}\right| + 1 + \left|L_{ij}^{(0)}\right| \alpha_{jk}\right), & i \leqslant k \leqslant n, \\ 0 & \text{в противном случае,} \end{cases} \\ \beta_{ki} = \begin{cases} \left|\frac{1}{U_{ii}^{(0)}}\right| \left\{ \sum_{j=1}^{i-1} \beta_{kj} \left|U_{ji}^{(0)}\right| + \sum_{j=1}^{i} \left(1 + \left|L_{kj}^{(0)}\right| \alpha_{ji}\right) \right\}, & i < k \leqslant n, \\ 0 & \text{в противном случае,} \end{cases} \end{cases}$$

и наконец

$$\alpha = \max_{1 \leqslant i,\ k \leqslant n} \{\max\{\alpha_{ik}, \beta_{ki}\}\}.$$

Используя определение (17″), мы немедленно получаем из (17′) при $i = 1$, что

$$d\left(U_{1k}^{(m+1)}\right) \leqslant \alpha_{1k} \left(d^{(m)}\right)^2, \quad 1 \leqslant k \leqslant n,$$

и

$$d\left(L_{k1}^{(m+1)}\right) \leqslant \beta_{k1} \left(d^{(m)}\right)^2, \quad 1 < k \leqslant n.$$

Допустим теперь, что для первых $i \neq 1$ строк и столбцов имеет место при $(1 \leqslant l \leqslant i-1)$

$$(17''')\quad \begin{cases} d\left(U_{lk}^{(m+1)}\right) \leqslant \alpha_{lk} \left(d^{(m)}\right)^2, & l \leqslant k \leqslant n, \\ d\left(L_{kl}^{(m+1)}\right) \leqslant \beta_{kl} \left(d^{(m)}\right)^2, & l \leqslant k \leqslant n. \end{cases}$$

Это очевидно при $l = 1$. Теперь мы получаем из (17′) и (17″), что

$$d\left(U_{ik}^{(m+1)}\right) \leqslant \sum_{j=1}^{i-1} \beta_{ij} \left| U_{jk}^{(0)} \right| (d^{(m)})^2 + \sum_{j=1}^{i-1} (d^{(m)})^2 + \sum_{j=1}^{i-1} \left| L_{ij}^{(0)} \right| \alpha_{jk} (d^{(m)})^2$$

$$= \alpha_{ik} (d^{(m)})^2, \quad i \leqslant k \leqslant n.$$

Аналогично

$$d\left(L_{ki}^{(m+1)}\right) \leqslant \left| \frac{1}{U_{ii}^{(0)}} \right| \left\{ \sum_{j=1}^{i-1} \beta_{kj} \left| U_{ji}^{(0)} \right| (d^{(m)})^2 + \sum_{j=1}^{i} (d^{(m)})^2 \right.$$

$$\left. + \sum_{j=1}^{i} \left| L_{kj}^{(0)} \right| \alpha_{ji} (d^{(m)})^2 \right\} = \beta_{ki} (d^{(m)})^2, \quad i < k \leqslant n.$$

Из этих соотношений следует доказываемое утверждение

$$d^{(m+1)} \leqslant \alpha (d^{(m)})^2. \blacksquare$$

Если исполнять (17) на вычислительной машине, применяя машинную интервальную арифметику, то в противоположность теореме 5 мы, вообще говоря, не получим за конечное число шагов точного треугольного разложения матрицы $\mathscr{A}_p$. Рассмотрим, какой окончательной точности здесь можно достичь. В этих рассуждениях примем те же допущения, которые привели нас к формулам (4.15a) и (4.15b), а тем самым и к (4.22), (4.23). Последние две формулы были использованы при доказательстве формул (4.24) и (4.25). Теперь мы применим эти две формулы к (17). Это дает ширину вычисленных элементов

$$\begin{cases} d\left(\overline{U}_{ik}^{(m+1)}\right) \leqslant d\left(U_{ik}^{(m+1)}\right) + 2\varepsilon \sum_{j=1}^{2i-2} \left| \overline{S}_j \right| + 2\varepsilon \left(3 + 3\varepsilon + \varepsilon^2\right) \\ \qquad \times \sum_{j=1}^{i-1} \left(\left| L_{ij}^{(m)} \right| \left| \overline{U}_{jk}^{(m+1)} - u_{jk}^{(m)} \right| + \left| \overline{L}_{ij}^{(m+1)} \right| \left| u_{jk}^{(m)} \right| \right), \\ \qquad\qquad i \leqslant k \leqslant n, \\ d\left(\overline{L}_{ki}^{(m+1)}\right) \leqslant d\left(L_{ki}^{(m+1)}\right) + \left| \frac{1}{u_{ii}^{(m)}} \right| \left\{ 2\varepsilon \sum_{j=1}^{2i-1} \left| \overline{T}_j \right| + 2\varepsilon_{2i} \left| \overline{T}_{2i} \right| \right. \\ \qquad + 2\varepsilon \left(3 + 3\varepsilon + \varepsilon^2\right) \left(\sum_{j=1}^{i-1} \left| \overline{L}_{kj}^{(m+1)} \right| \left| u_{ji}^{(m)} \right| \right. \\ \qquad \left. \left. + \sum_{j=1}^{i} \left| L_{kj}^{(m)} \right| \left| \overline{U}_{ji}^{(m+1)} - u_{ji}^{(m)} \right| \right) \right\}, \quad i < k \leqslant n, \\ 1 \leqslant i \leqslant n. \end{cases}$$

В этих неравенствах $\bar{S}_j$ и $\bar{T}_j$ обозначают фактические результаты промежуточных вычислений. Эти неравенства можно интерпретировать следующим образом. Допустим, что все элементы матрицы $\mathscr{L}^*_p$ не превосходят 1 по абсолютной величине. Это предположение выполняется хотя бы приближенно, если строки матрицы $\mathscr{A}_p$ упорядочены таким образом, что не приходится переставлять строки в процессе исключения по Гауссу с выбором главных элементов по столбцам. Если ширина интервалов $L_{ij}^{(m)}$ не слишком велика, то $|L_{ij}^{(m)}|$ не намного больше 1, а в силу того, что в (17) берутся пересечения, это верно и для $|L_{ij}^{(m+1)}|$. При тех же предположениях те же рассуждения показывают, что $|\bar{U}_{jk}^{(m+1)}| - |u_{jk}^{(m)}|$ и $|\bar{U}_{ji}^{(m+1)}| - |u_{ji}^{(m)}|$ малы. Поэтому делаем вывод, что при малой ширине элементов на m-м шаге разность между $d(\bar{U}_{ik}^{(m+1)})$ и $d(U_{ik}^{(m+1)})$ существенно зависит от $|u_{jk}^{(m)}|$ и от величины промежуточных результатов. То же верно и для разности между $d(\bar{L}_{ki}^{(m+1)})$ и $d(L_{ki}^{(m+1)})$, если добавить еще, что малые величины $|u_{ii}^{(m)}|$ могут ухудшить эту разность.

Так как в силу теоремы 6 первые члены приведенных выше неравенств приближенно равны квадратам таких же членов на предыдущем шаге, мы получим небольшую ширину, если выполнены следующие условия:

(a) Элементы матрицы $\mathscr{U}^*_p$ по абсолютной величине не намного больше единицы.

(b) Диагональные элементы матрицы $\mathscr{U}^*_p$ по абсолютной величине не намного меньше единицы.

(c) Элементы матрицы $\mathscr{L}^*_p$ по абсолютной величине не больше единицы.

(d) Вычисленные промежуточные результаты в (17) не слишком велики по абсолютной величине.

* * *

Литература: [15], [16], [17], [19], [20], [28], [92], [101], [170], [211], [298], [325], [346], [351], [385a], [468a], [521a].

Глава 19

МЕТОДЫ НЬЮТОНОВСКОГО ТИПА ДЛЯ СИСТЕМ НЕЛИНЕЙНЫХ УРАВНЕНИЙ

Рассмотрим методы итерационной локализации решений для систем нелинейных уравнений. Дано множество функций $f_i(x_p)$, $1 \leqslant i \leqslant n$ векторной переменной $x_p = (x_1, \ldots, x_n)^T$, кото-

рые мы объединим в векторную функцию $\mathcal{f}_{\mathrm{p}}(x_{\mathrm{p}}) = (f_1(x_{\mathrm{p}}), \ldots$ $\ldots, f_n(x_{\mathrm{p}}))^{\mathrm{T}}$. Допустим, что производная Фреше $\mathcal{f}'_{\mathrm{p}}(x_{\mathrm{p}})$ функции $\mathcal{f}_{\mathrm{p}}(x_{\mathrm{p}})$ существует на множестве $\mathfrak{B} \subseteq V_n(\mathbb{R})$ и что $x^{(0)} = (X_1^{(0)}, \ldots, X_n^{(0)})^{\mathrm{T}} \subseteq \mathfrak{B}$. Мы предположим также, что для производной Фреше имеется интервальная оценка на $x^{(0)}$. Пусть теперь дан вектор $y_{\mathrm{p}} = (y_1, \ldots, y_n)^{\mathrm{T}} \in x^{(0)}$, удовлетворяющий уравнению

$$\mathcal{f}_{\mathrm{p}}(y_{\mathrm{p}}) = o_{\mathrm{p}}. \tag{1}$$

При сделанных предположениях мы можем линейно аппроксимировать $f_i(x_{\mathrm{p}})$ в точке $y_{\mathrm{p}} \in x^{(0)}$ с помощью формулы Тейлора

$$f_i(x_{\mathrm{p}}) = f_i(y_{\mathrm{p}}) + \sum_{r=1}^{n} \frac{\partial}{\partial x_r} f_i(y_{\mathrm{p}} + \theta_i(x_{\mathrm{p}} - y_{\mathrm{p}}))(x_r - y_r), \tag{2}$$

$$0 < \theta_i < 1, \quad 1 \leqslant i \leqslant n, \quad x_{\mathrm{p}} \in x^{(0)}.$$

Из (1) следует, что вектор y_{p} удовлетворяет уравнению

$$\mathcal{f}_{\mathrm{p}}(x_{\mathrm{p}}) = \mathcal{J}_{\mathrm{p}}(x_{\mathrm{p}})(x_{\mathrm{p}} - y_{\mathrm{p}}),$$

где матрица $\mathcal{J}_{\mathrm{p}}(x_{\mathrm{p}})$ определяется равенством

$$\mathcal{J}_{\mathrm{p}}(x_{\mathrm{p}}) := \left(\frac{\partial}{\partial x_r} f_i(z_{\mathrm{p}})\Big|_{z_{\mathrm{p}} = y_{\mathrm{p}} + \theta_i(x_{\mathrm{p}} - y_{\mathrm{p}})}\right). \tag{3}$$

Вместо $\mathcal{J}_{\mathrm{p}}(x_{\mathrm{p}})$ мы будем иногда писать

$$\mathcal{J}_{\mathrm{p}}(x_{\mathrm{p}}, y_{\mathrm{p}}, \theta_1, \ldots, \theta_n),$$

так как эта матрица зависит от всех величин, входящих в ее определение. Так как величины θ_i лежат в открытом интервале $(0, 1)$, мы имеем

$$y_{\mathrm{p}} + \theta_i(x_{\mathrm{p}} - y_{\mathrm{p}}) \in x^{(0)}.$$

Пусть интервальная матрица $\mathcal{f}'_{\mathrm{p}}(x^{(0)})$ обозначает интервальное оценивание производной Фреше на интервале $x^{(0)}$. Тогда среди решений z_{p} множества линейных уравнений

$$\mathcal{f}_{\mathrm{p}}(x_{\mathrm{p}}) = \mathcal{F}_{\mathrm{p}} \cdot (x_{\mathrm{p}} - z_{\mathrm{p}}), \quad \text{где} \quad \mathcal{F}_{\mathrm{p}} \in \mathcal{f}'_{\mathrm{p}}(x^{(0)}), \tag{4}$$

имеется и вектор y_{p}, так как $\mathcal{J}_{\mathrm{p}}(x_{\mathrm{p}}) \in \mathcal{f}'_{\mathrm{p}}(x^0)$. Теперь возникла задача — вычислить вектор $x^{(1)}$, локализующий множество решений уравнения (4), а затем использовать этот интервальный вектор как потенциально лучшую локализацию вектора-решения y_{p}. Перед тем как браться за эту задачу, мы рассмотрим еще одну возможность построения множества уравнений, аналогичного (4), описанную Хансеном [109]. В этом случае приме-

няется линейная аппроксимация значения $\mathcal{f}_{\mathrm{p}}(x_{\mathrm{p}})$ в точке y_{p}, несколько отличная от (2).

Рассмотрим $f_i(x_1, \ldots, x_n)$ как функцию только от x_1 и разложим ее в точке y_1:

$$f_i(x_1, \ldots, x_n) = f_i(y_1, x_2, \ldots, x_n) + (x_1 - y_1)\frac{\partial}{\partial x_1} f_i(y_1 + \theta_{i1}(x_1 - y_1), x_2, \ldots, x_n),$$

где $\theta_{i1} \in (0, 1)$.

Затем мы разложим $f_i(y_1, x_2, \ldots, x_n)$ как функцию от x_2 в точке y_2 и получим

$$f_i(y_1, x_2, \ldots, x_n) = f_i(y_1, y_2, x_3, \ldots, x_n) + (x_2 - y_2)\frac{\partial}{\partial x_2} f_i(y_1, y_2 + \theta_{i2}(x_2 - y_2), x_3, \ldots, x_n),$$

где $\theta_{i2} \in (0, 1)$.

Теперь мы снова разложим $f_i(y_1, y_2, x_3, \ldots, x_n)$ как функцию от x_3 в точке y_3 и т. д. и, наконец, разложим $f_i(y_1, y_2, \ldots, y_{n-1}, x_n)$ в точке y_n. Собирая эти разложения, мы получим

$$f_i(x_1, \ldots, x_n) = f_i(y_1, \ldots, y_n) + \sum_{j=1}^{n}(x_j - y_j)\frac{\partial}{\partial x_j} f_i(y_1, \ldots, y_{j-1}, y_j + \theta_{ij}(x_j - y_j), x_{j+1}, \ldots, x_n),$$

где $\theta_{ij} \in (0, 1)$, $1 \leqslant i \leqslant n$.

Таким образом, мы имеем соотношение

$$\mathcal{f}_{\mathrm{p}}(x_{\mathrm{p}}) = \tilde{\mathcal{J}}_{\mathrm{p}}(x_{\mathrm{p}})(x_{\mathrm{p}} - y_{\mathrm{p}}),$$

где матрица $\mathcal{J}_{\mathrm{p}}(x_{\mathrm{p}})$ определена равенствами

$$\tilde{\mathcal{J}}_{\mathrm{p}}(x_{\mathrm{p}}) = \left(\frac{\partial}{\partial x_j} f_i(y_1, \ldots, y_{j-1}, z_j, x_{j+1}, \ldots, x_n)\Big|_{z_j = y_j + \theta_{ij}(x_j - y_j)}\right). \tag{5}$$

Мы можем взять в качестве y_{p} решение множества линейных уравнений

$$\mathcal{f}_{\mathrm{p}}(x_{\mathrm{p}}) = \tilde{\mathcal{F}}_{\mathrm{p}}(x_{\mathrm{p}} - z_{\mathrm{p}}), \quad \text{где} \quad \tilde{\mathcal{F}}_{\mathrm{p}} \in \tilde{\mathcal{J}}_{\mathrm{p}}(x^{(0)}), \tag{6}$$

так как $y_j \in X^{(0)}$ и $y_j + \theta_{ij}(x_j - y_j) \in X_j^{(0)}$ для $1 \leqslant j \leqslant n$. Упомянутая здесь интервальная матрица $\tilde{\mathcal{J}}_{\mathrm{p}}(x^{(0)})$ определена равенствами

$$\tilde{\mathcal{J}}_{\mathrm{p}}(x^{(0)}) = \left(\frac{\partial}{\partial x_j} f_i(X_1^{(0)}, \ldots, X_{j-1}^{(0)}, X_j^{(0)}, x_{j+1}, \ldots, x_n)\right),$$

и легко видеть, что для интервальной матрицы мы имеем

$$\tilde{\mathscr{J}}_{\mathrm{p}}(x^{(0)}) \subseteq \mathscr{f}'_{\mathrm{p}}(x^{(0)}).$$

Поэтому множество (6) содержит не больше систем линейных уравнений, чем множество из (4). В случае когда $n = 1$ или когда частные производные $\partial f_i/\partial x_j$ зависят только от x_j, интервальные матрицы $\mathscr{f}'_{\mathrm{p}}(x^{(0)})$ и $\tilde{\mathscr{J}}_{\mathrm{p}}(x^{(0)})$ совпадают.

Вернемся теперь к первоначальной задаче — как улучшить локализующий вектор $x^{(0)}$ для вектора решений y_{p}. В соответствии с только что проведенными рассуждениями мы вычислим интервальный вектор, содержащий множество решений систем (4) (соответственно (6)), а потому и вектор y_{p}. Рассмотрим сначала (4). Предположим, что y_{p} — простой корень функции $\mathscr{f}_{\mathrm{p}}(x_{\mathrm{p}})$ в $x^{(0)}$ и что все матрицы $\mathscr{F}_{\mathrm{p}} \in \mathscr{f}'_{\mathrm{p}}(x^{(0)})$ невырождены. Пусть $\mathscr{V}$ — интервальная матрица, содержащая все обращения $\mathscr{F}_{\mathrm{p}}^{-}$ матриц $\mathscr{F}_{\mathrm{p}} \in \mathscr{f}'_{\mathrm{p}}(x^{(0)})$. Мы уже продемонстрировали в гл. 12—17 практические методы вычисления такой интервальной матрицы $\mathscr{V}$. Вычислим теперь с помощью $\mathscr{V}$ интервальный вектор

$$y^{(1)} = m(x^{(0)}) - \mathscr{V} \cdot \mathscr{f}_{\mathrm{p}}(m(x^{(0)})),$$

где $m(x^{(0)})$ определяется согласно (18.1). Таким образом, мы получим локализацию для множества решений уравнений (4). Здесь мы использовали соотношение $\mathscr{J}_{\mathrm{p}}(m(x^{(0)})) \in \mathscr{f}'_{\mathrm{p}}(x^{(0)})$. Аналогичным образом можно использовать уравнение (6). Ввиду $y_{\mathrm{p}} \in x^{(0)}$ мы можем построить интервальный вектор

$$x^{(1)} = \{m(x^{(0)}) - \mathscr{V} \cdot \mathscr{f}_{\mathrm{p}}(m(x^{(0)}))\} \cap x^{(0)} \subseteq x^{(0)}$$

и получить новую локализацию для y_{p}. Повторение этого вычисления приводит к итерационному методу

$$x^{(k+1)} = \{m(x^{(k)} - \mathscr{V} \cdot \mathscr{f}_{\mathrm{p}}(m(x^{(k)}))\} \cap x^{(k)}, \quad k \geqslant 0, \tag{7}$$

где $m(x^{(k)})$ — срединное отображение из (18.1).

Эта итерация порождает монотонную последовательность

$$x^{(0)} \supseteq x^{(1)} \supseteq x^{(2)} \supseteq \ldots$$

интервальных векторов, для которой мы докажем следующее утверждение.

Теорема 1. *Пусть $x^{(0)}$ — интервальный вектор и $y_{\mathrm{p}} \in x^{(0)}$ — корень функции $\mathscr{f}_{\mathrm{p}}(x_{\mathrm{p}})$. Пусть $\mathscr{V}$ — интервальная матрица, содержащая обратные матрицы $\mathscr{J}_{\mathrm{p}}(x_{\mathrm{p}})^{-1}$ (ср. (3)) для $x_{\mathrm{p}} \in x^{(0)}$. Тогда после-*

довательность $\{x^{(k)}\}_{k=0}^{\infty}$ *интервальных векторов, вычисленная согласно* (7), *удовлетворяет следующим условиям.*

(8) *Каждый интервальный вектор* $x^{(k)}$, $k \geqslant 0$, *содержит корень* y_{p}.

(9) *Если все матрицы* $\mathscr{V}_{\mathrm{p}} \in \mathscr{V}$ *невырождены, то* $\lim\limits_{k\to\infty} x^{(k)} = y_{\mathrm{p}}$.

Доказательство. Рассмотрим (8). По условию теоремы мы имеем $y_{\mathrm{p}} \in x^{(0)}$, а ввиду $m(x^{(0)}) \in x^{(0)}$ также и $\mathscr{J}_{\mathrm{p}}(m(x^{(0)}))^{-1} \in \mathscr{V}$. Поэтому из (10.10′) следует, что

$$y_{\mathrm{p}} = m(x^{(0)}) - \mathscr{J}_{\mathrm{p}}(m(x^{(0)}))^{-1} f_{\mathrm{p}}(m(x^{(0)})) \in m(x^{(0)}) - \mathscr{V} \cdot f_{\mathrm{p}}(m(x^{(0)})),$$

а потому

$$y_{\mathrm{p}} \in \{m(x^{(0)}) - \mathscr{V} \cdot f_{\mathrm{p}}(m(x^{(0)}))\} \cap x^{(0)} = x^{(1)}.$$

Это рассуждение позволяет доказать (8) методом математической индукции.

(9): В силу следствия 10.8 монотонная локализующая последовательность $x^{(0)} \supseteq x^{(1)} \supseteq x^{(2)} \ldots$ сходится к некоторому интервальному вектору x. Мы покажем, что $d(x) = o_{\mathrm{p}}$, а вместе с (8) это даст $\lim\limits_{k\to\infty} x^{(k)} = y_{\mathrm{p}}$. Допустим для приведения к противоречию, что $d(x) \neq o_{\mathrm{p}}$. Из непрерывности отображения (7) следует, что x удовлетворяет уравнению

$$x = \{m(x) - \mathscr{V} \cdot f_{\mathrm{p}}(m(x))\} \cap x.$$

Рассмотрев $m(x) \in x$, мы видим, что $m(x) \neq y_{\mathrm{p}}$, так как $d(x) \neq o$. С другой стороны, мы имеем

$$m(x) \in m(x) - \mathscr{V} \cdot f_{\mathrm{p}}(m(x)).$$

Кроме того, из (10.1) следует соотношение

$$\mathscr{A} x_{\mathrm{p}} = \{\mathscr{A}_{\mathrm{p}} x_{\mathrm{p}} \mid \mathscr{A}_{\mathrm{p}} \in \mathscr{A}\}$$

для интервальной матрицы $\mathscr{A}$ и вещественного вектора x_{p}. Отсюда получается представление

$$m(x) = m(x) - \mathscr{V}_{\mathrm{p}} \cdot f_{\mathrm{p}}(m(x))$$

для некоторой матрицы $\mathscr{V}_{\mathrm{p}} \in \mathscr{V}$. Поэтому мы имеем

$$o_{\mathrm{p}} = \mathscr{V}_{\mathrm{p}} \cdot f_{\mathrm{p}}(m(x)),$$

что дает противоречие, так как $m(x) \neq y_{\mathrm{p}}$ и матрица $\mathscr{V}_{\mathrm{p}}$ невырождена. ■

Достаточное условие сходимости (9) для итерационного метода (7) использовало более слабое условие

$$\{\mathscr{J}_{\mathrm{p}}(x_{\mathrm{p}})^{-1} \mid x_{\mathrm{p}} \in x^{(0)}\} \subseteq \mathscr{V}.$$

на $\mathscr{Y}$, чем условие, использованное при выводе формул (7). Первоначально мы потребовали, чтобы выполнялось условие

$$\{\mathscr{F}_{\mathrm{p}}^{-1} \mid \mathscr{F}_{\mathrm{p}} \in \mathscr{f}'_{\mathrm{p}}(x^{(0)})\} \subseteq \mathscr{Y},$$

которое легче поддается проверке.

Доказательство соотношения (9) можно модифицировать так, чтобы показать, что из $m(x^{(k)}) \neq y_{\mathrm{p}}$ следует

$$m(x^{(k)}) \notin x^{(k+1)}, \qquad k \geqslant 0.$$

Это означает, что в методе (7) по крайней мере одна компонента ширины $d(x^{(k)})$ уменьшается на $(k+1)$-м шаге более чем вдвое. Метод (7) можно улучшить, определяя на каждом шаге новую матрицу $\mathscr{Y}^{(k)}$. При выводе формул (7) было видно, что для соотношения (8) нужно лишь

$$\{\mathscr{J}_{\mathrm{p}}(x_{\mathrm{p}})^{-1} \mid x_{\mathrm{p}} \in x^{(k)}\} \subseteq \mathscr{Y}^{(k)}$$

и

$$\{\mathscr{F}_{\mathrm{p}}^{-1} \mid \mathscr{F}_{\mathrm{p}} \in \mathscr{f}'_{\mathrm{p}}(x^{(k)})\} \subseteq \mathscr{Y}^{(k)}.$$

Иными словами, на $(k+1)$-м шаге мы находим локализующее множество для множества уравнений

$$\mathscr{f}_{\mathrm{p}}(m(x^{(k)})) = \mathscr{F}_{\mathrm{p}} \cdot (m(x^{(k)}) - z_{\mathrm{p}}), \quad \text{где} \quad \mathscr{F}_{\mathrm{p}} \in \mathscr{f}'_{\mathrm{p}}(x^{(k)}).$$

Теперь мы покажем, как определить подходящую последовательность $\{\mathscr{Y}^{(k)}\}_{k=0}^{\infty}$. Сначала найдем локализующее множество для

$$\{\mathscr{F}_{\mathrm{p}}^{-1} \mid \mathscr{F}_{\mathrm{p}} \in \mathscr{f}'_{\mathrm{p}}(x^{(k)})\}$$

с помощью одного из итеративных методов, описанных в гл. 12 или 14. При этом мы используем сокращение $\mathscr{F}^{(k)}$ для интервальной матрицы $\mathscr{f}'_{\mathrm{p}}(x^{(k)})$ и сокращение $\tilde{\mathscr{F}}^{(k)}$ для интервальной матрицы $\tilde{\mathscr{J}}_{\mathrm{p}}(x^{(k)})$.

Мы начинаем с интервальной матрицы $\mathscr{W}^{(0)}$, которая была найдена, например, с помощью обычных оценок по норме и удовлетворяет условию

$$\{\mathscr{F}_{\mathrm{p}}^{-1} \mid \mathscr{F}_{\mathrm{p}} \in \mathscr{F}^{(k)}\} \subseteq \mathscr{W}^{(0)}.$$

Затем мы исполняем итерационную процедуру

$$\begin{aligned} \mathscr{W}^{(\nu+1)} = \{&(\mathscr{I}_{\mathrm{p}} - (m(\mathscr{F}^{(k)}))^{-1}\mathscr{F}^{(k)})\,\mathscr{W}^{(\nu)} \\ &+ (m(\mathscr{F}^{(k)}))^{-1}\} \cap \mathscr{W}^{(\nu)}, \qquad \nu \geqslant 0. \end{aligned} \tag{10}$$

Последовательность

$$\mathscr{W}^{(0)} \supseteq \mathscr{W}^{(1)} \supseteq \mathscr{W}^{(2)} \supseteq \ldots$$

всегда будет сходиться к некоторой однозначно определенной интервальной матрице $\mathscr{W} \subseteq \mathscr{W}^{(0)}$, если спектральный радиус удовлетворяет условию

$$\rho\left(\left|\mathscr{I}_{\mathrm{p}} - (m(\mathscr{F}^{(k)}))^{-1}\mathscr{F}^{(k)}\right|\right) < 1$$

(ср. гл. 12 и 14). Матрица $\mathscr{I}_{\mathrm{p}} - (m(\mathscr{F}^{(k)}))^{-1}\mathscr{F}^{(k)}$ в (10) симметрична относительно центра $\mathscr{O}_{\mathrm{p}}$. Это означает, что процедуру (10) можно разбить на две одинаковые вещественные итерационные процедуры соответственно для матриц верхних и нижних границ для $\mathscr{W}$. Поэтому для вычисления $\mathscr{W}$ нам нужно решить всего одну систему вещественных уравнений.

Положим теперь

$$\mathscr{Y}^{(k)} := \mathscr{W} \supseteq \left\{\mathscr{F}_{\mathrm{p}}^{-1} \,\middle|\, \mathscr{F}_{\mathrm{p}} \in \mathscr{F}^{(k)}\right\}. \tag{11}$$

С помощью этой матрицы $\mathscr{Y}^{(k)}$ мы делаем шаг итерации

$$x^{(k+1)} = \{m(x^{(k)}) - \mathscr{Y}^{(k)}\mathscr{f}_{\mathrm{p}}(m(x^{(k)}))\} \cap x^{(k)}.$$

Вычисляется новая матрица $\mathscr{F}^{(k+1)} = \mathscr{f}'_{\mathrm{p}}(x^{(k+1)})$ и затем запускается процедура (10). Из $x^{(k+1)} \subseteq x^{(k)}$ легко следует, что $\mathscr{F}^{(k+1)} \subseteq \mathscr{F}^{(k)}$ и $\left\{\mathscr{F}_{\mathrm{p}}^{-1} \,\middle|\, \mathscr{F}_{\mathrm{p}} \in \mathscr{F}^{(k+1)}\right\} \subseteq \mathscr{Y}^{(k)}$. Поэтому матрицу $\mathscr{Y}^{(k)}$ можно использовать в качестве нового начального элемента для (10). Мы получаем таким образом следующую итерационную процедуру, в которой $\mathscr{Y}^{(0)}$ содержит множество $\left\{\mathscr{F}_{\mathrm{p}}^{-1} \,\middle|\, \mathscr{F}_{\mathrm{p}} \in \mathscr{F}^{(0)}\right\}$:

$$x^{(k+1)} = \{m(x^{(k)}) - \mathscr{Y}^{(k)}\mathscr{f}_{\mathrm{p}}(m(x^{(k)}))\} \cap x^{(k)}, \qquad k \geqslant 0, \tag{12}$$

где $\mathscr{Y}^{(k)}$ — матрица, обладающая свойством (11) и вычисленная с помощью процедуры (10); причем $\mathscr{Y}^{(0)}$ используется в качестве начальной матрицы в (10) при $k = 1$.

Мы докажем следующее утверждение о методе (12).

Теорема 2. *Пусть $x^{(0)}$ — интервальный вектор, а $y_{\mathrm{p}} \in x^{(0)}$ — корень функции $\mathscr{f}_{\mathrm{p}}(x_{\mathrm{p}})$. Пусть $\mathscr{Y}^{(0)}$ — интервальная матрица, содержащая все обращения $\mathscr{F}_{\mathrm{p}}^{-1}$ матриц $\mathscr{F}_{\mathrm{p}} \in \mathscr{F}^{(0)} = \mathscr{f}'_{\mathrm{p}}(x^{(0)})$. Допустим еще, что производная Фреше $\mathscr{f}'_{\mathrm{p}}(x_{\mathrm{p}})$ для значения аргумента $x^{(0)}$ удовлетворяет условиям теоремы 3.5 для каждого элемента. Тогда последовательность интервальных векторов, вычисленная по формулам (12), обладает следующими свойствами.*

(8) *Каждое из приближений $x^{(k)}$, $k \geqslant 0$, содержит корень y_{p};*

(9) *если любая матрица $\mathscr{Y}_{\mathrm{p}} \in \mathscr{Y}^{(0)}$ неособенная, то мы имеем* $\lim\limits_{k\to\infty} x^{(k)} = y_{\mathrm{p}}$;

если последовательность $\{x^{(k)}\}_{k=0}^{\infty}$ *сходится к* y_p, *то для некоторой нормы* $\|\cdot\|$ *верно*

$$\|d(x^{(k+1)})\| \leqslant \gamma \|d(x^{(k)})\|^2, \quad \gamma \geqslant 0, \tag{13}$$

т. е. R-порядок метода (12) *удовлетворяет неравенству*

$$O_R((12), y_p) \geqslant 2$$

(*см. приложение* A, *теорема* 2).

Доказательство. Проверка соотношений (8) и (9) проводится совершенно так же, как в доказательстве теоремы (1). Нужно лишь рассмотреть

$$\{\mathscr{F}_p^{-1} \mid \mathscr{F}_p \in \mathscr{F}^{(k)}\} \subseteq \mathscr{Y}^{(k)} \subseteq \mathscr{Y}^{(0)}, \quad k \geqslant 0.$$

(13): Из того что $x^{(k+1)} \subseteq m(x^{(k)}) - \mathscr{Y}^{(k)} \mathscr{f}_p(m(x^{(k)}))$, следует, что

$$\begin{aligned} d(x^{(k+1)}) &\leqslant d(m(x^{(k)}) - \mathscr{Y}^{(k)} \mathscr{f}_p(m(x^{(k)}))) \\ &= d(\mathscr{Y}^{(k)} \mid \mathscr{f}_p(m(x^{(k)})) \mid = d(\mathscr{Y}^{(k)} \mid \mathscr{f}_p(m(x^{(k)})) - \mathscr{f}_p(y_p) \mid \\ &= d(\mathscr{Y}^{(k)} \mid \mathscr{J}_p(m(x^{(k)}))(m(x^{(k)}) - y_p) \mid \\ &\leqslant d(\mathscr{Y}^{(k)} \mid \mathscr{J}_p(m(x^{(k)})) \Big| \frac{1}{2} d(x^{(k)} \leqslant d(\mathscr{Y}^{(k)} \mid \mathscr{F}^{(0)} \Big| \frac{1}{2} d(x^{(k)}). \end{aligned}$$

Теперь, применяя монотонную векторную норму и совместную с ней монотонную матричную норму, мы получим

$$\|d(x^{(k+1)})\| \leqslant \|d(\mathscr{Y}^{(k)}\| \tfrac{1}{2} \| \mid \mathscr{F}^{(0)} \mid \| \| d(x^{(k)}) \| = c \| d(\mathscr{Y}^{(k)}) \| \| d(x^{(k)}) \|.$$

Наконец, нам нужно оценить $\|d(\mathscr{Y}^{(k)})\|$. Так как $\mathscr{Y}^{(k)}$ удовлетворяет равенству

$$\mathscr{Y}^{(k)} = \{(\mathscr{I}_p - (m(\mathscr{F}^{(k)}))^{-1} \mathscr{F}^{(k)}) \mathscr{Y}^{(k)} + (m(\mathscr{F}^{(k)}))^{-1}\} \cap \mathscr{Y}^{(k)},$$

мы получаем из (10.21) и (10.29), что

$$\begin{aligned} d(\mathscr{Y}^{(k)}) &\leqslant d((\mathscr{I}_p - (m(\mathscr{F}^{(k)}))^{-1} \mathscr{F}^{(k)}) \mathscr{Y}^{(k)} + (m(\mathscr{F}^{(k)}))^{-1}) \\ &= d((\mathscr{I}_p - (m(\mathscr{F}^{(k)}))^{-1} \mathscr{F}^{(k)}) \mathscr{Y}^{(k)}) \\ &= d(((m(\mathscr{F}^{(k)}))^{-1} (m(\mathscr{F}^{(k)}) - \mathscr{F}^{(k)})) \mathscr{Y}^{(k)}) \\ &= \mid m(\mathscr{F}^{(k)})^{-1} \mid d(\mathscr{F}^{(k)}) \mid \mathscr{Y}^{(k)} \mid. \end{aligned}$$

Применяя упомянутую выше монотонную матричную норму, а также соотношения

$$\mid \mathscr{Y}^{(k)} \mid \leqslant \mid \mathscr{Y}^{(0)} \mid \quad \text{и} \quad \mid (m(\mathscr{F}^{(k)}))^{-1} \mid \leqslant \mid \mathscr{Y}^{(0)} \mid,$$

мы получаем, что

$$\|d(\mathscr{Y}^{(k)})\| \leqslant \tilde{\gamma} \|d(\mathscr{F}^{(k)})\|,$$

где $\tilde{\gamma}$ — константа, не зависящая от k.

Отдельные элементы матрицы $d(\mathscr{F}^{(k)})$ имеют вид

$$d\left(\frac{\partial}{\partial x_j} f_i(X_1^{(k)}, \ldots, X_n^{(k)})\right),$$

и с помощью соотношения (3.5′) для них можно получить оценку

$$d\left(\frac{\partial}{\partial x_j} f_i(X_1^{(k)}, \ldots, X_n^{(k)})\right) \leqslant \sum_{r=1}^{n} \gamma_{ijr}^{(k)} d\left(X_r^{(k)}\right) \leqslant \sum_{r=1}^{n} \gamma_{ijr}^{(0)} d\left(X_r^{(k)}\right).$$

Следовательно, мы имеем

$$d(\mathscr{F}^{(k)}) \leqslant \mathscr{C}_{\mathrm{p}} d(x^{(k)}),$$

где $\mathscr{C}_{\mathrm{p}} = (\gamma_{jl}^{(0)})$ — билинейный оператор из $V_n(\mathbb{R}) \times V_n(\mathbb{R})$ в $V_n(\mathbb{R})$. Если определить норму для $\mathscr{C}_{\mathrm{p}}$ обычным равенством

$$\|\mathscr{C}_{\mathrm{p}}\| = \sup_{\|x_{\mathrm{p}}\|=1} \sup_{\|y_{\mathrm{p}}\|=1} \|\mathscr{C}_{\mathrm{p}} x_{\mathrm{p}} y_{\mathrm{p}}\|,$$

то получается, что

$$\|d(\mathscr{F}^{(k)})\| \leqslant \|\mathscr{C}_{\mathrm{p}}\| \|d(x^{(k)})\|,$$

т. е.

$$\|d(\mathscr{Y}^{(k)})\| \leqslant \bar{c} \|d(x^{(k)})\|, \quad \text{где} \quad \bar{c} = \|\mathscr{C}_{\mathrm{p}}\| \tilde{\gamma}.$$

Подставляя это в полученную выше оценку для $d(x^{(k+1)})$, мы получаем

$$\|d(x^{(k+1)})\| \leqslant \tilde{c} \|d(x^{(k)})\|^2, \quad \text{где} \quad \tilde{c} = c\bar{c}$$

для монотонной векторной нормы. Применяя понятие эквивалентности норм так же, как при доказательстве соотношения (18.8) в теореме 18.1, мы убеждаемся, что могли использовать произвольную норму, и имеем соотношение (13) с константой γ, не зависящей от k. ■

Теперь мы хотим рассмотреть другой способ, позволяющий вычислять решение y_{p} уравнения $f_{\mathrm{p}}(x_{\mathrm{p}}) = o_{\mathrm{p}}$ путем последовательных локализаций. Этот метод подходит для определенных классов нелинейных систем уравнений.

Мы исходим из системы (4), полагая $x_{\mathrm{p}} := m(x^{(0)})$ (см. (18.1)). Приводимые ниже соображения можно реализовать и исходя из (6). Если, например, предположить, что $\mathscr{F}^{(0)} = f'_{\mathrm{p}}(x^{(0)})$ удовлетворяет условиям теоремы 15.3, то к интервальной матрице $\mathscr{F}^{(0)}$ и вектору $f_{\mathrm{p}}(m(x^{(0)}))$ можно применить метод Гаусса. Если $w^{(0)}$ — интервальный вектор, полученный в результате, то вектор $m(x^{(0)}) - w^{(0)}$, а значит, и вектор

$$\tilde{x}^{(0)} = \{m(x^{(0)}) - w^{(0)}\} \cap x^{(0)},$$

содержат решение y_{p} системы нелинейных уравнений.

Напишем уравнение

$$f_{\mathrm{p}}(x_{\mathrm{p}}) = J_{\mathrm{p}}(x_{\mathrm{p}})(x_{\mathrm{p}} - y_{\mathrm{p}})$$

по образцу (2). Представим матрицу $J_{\mathrm{p}}(x_{\mathrm{p}})$, входящую в это уравнение, в виде

$$J_{\mathrm{p}}(x_{\mathrm{p}}) = D_{\mathrm{p}}(x_{\mathrm{p}}) - B_{\mathrm{p}}(x_{\mathrm{p}}),$$

где $D_{\mathrm{p}}(x_{\mathrm{p}})$ — содержит диагональ матрицы $J_{\mathrm{p}}(x_{\mathrm{p}})$. Полагая $x_{\mathrm{p}} = m(x^{(0)})$, получаем из монотонности включения, что

$$y_{\mathrm{p}} = m(x^{(0)}) - D_{\mathrm{p}}(m(x^{(0)}))^{-1}\{B_{\mathrm{p}}(m(x^{(0)}))(m(x^{(0)}) - y_{\mathrm{p}}) + f_{\mathrm{p}}(m(x^{(0)}))\}$$
$$\in \{m(x^{(0)}) - D_{\mathrm{p}}(x^{(0)})^{-1}\{B_{\mathrm{p}}(x^{(0)})(m(x^{(0)}) - \tilde{x}^{(0)})$$
$$+ f_{\mathrm{p}}(m(x^{(0)}))\}\} \cap \tilde{x}^{(0)} =: x^{(1)}.$$

Здесь мы использовали такое же представление матрицы $F^{(0)} = f'_{\mathrm{p}}(x^{(0)})$ в виде

$$F^{(0)} = D_{\mathrm{p}}(x^{(0)}) - B_{\mathrm{p}}(x^{(0)}),$$

как и выше, для $J_{\mathrm{p}}(x_{\mathrm{p}})$. Используя равенство $D_{\mathrm{p}}(x^{(0)}) = \mathrm{diag}(D_i(x^{(0)}))$, мы полагаем $D_{\mathrm{p}}(x^{(0)})^{-1} = \mathrm{diag}(1/D_i(x^{(0)}))$. Если матрица $F^{(1)} = f'_{\mathrm{p}}(x^{(1)})$ снова удовлетворяет условиям теоремы 15.3, то мы можем, исходя из $x^{(1)}$, вычислить тем же способом, что и раньше, новую локализацию $x^{(2)}$ для y_{p}, и т. д. Повторение этого шага приводит к следующему итерационному методу.

(14) (a) $F^{(k)} = f'_{\mathrm{p}}(x^{(k)}) = D_{\mathrm{p}}(x^{(k)}) - B_{\mathrm{p}}(x^{(k)})$.

(b) Применение метода Гаусса к $(F^{(k)}, f_{\mathrm{p}}(m(x^{(k)})))$ дает $w^{(k)}$; здесь $m(x^{(k)})$ вычисляется согласно (18.1).

(c) $\tilde{x}^{(k)} = \{m(x^{(k)}) - w^{(k)}\} \cap x^{(k)}$.

(d) $x^{(k+1)} = \{m(x^{(k)}) - D_{\mathrm{p}}(x^{(k)})^{-1}\{B_{\mathrm{p}}(x^{(k)})(m(x^{(k)}) - \tilde{x}^{(k)})$
$+ f_{\mathrm{p}}(m(x^{(k)}))\}\} \cap \tilde{x}^{(k)}, \quad k \geqslant 0.$

Шаг (d) этого алгоритма требует $\sim n^2$ операций (интервальных умножений и делений) для неразреженной матрицы. Из того что шаг (b) требует $\sim n^3/3$ операций, следует, что объем вычислений на шаге (d) несуществен при больших n.

Докажем несколько лемм, чтобы исследовать условия, при которых для метода (14) выполнено $\lim_{k\to\infty} x^{(k)} = y_{\mathrm{p}}$, и оценить R-порядок сходимости.

Лемма 3. *Пусть* $A = (A_{ij})$ *— вещественная интервальная матрица, удовлетворяющая условиям теоремы* 15.3. *Если* A *представ-*

лена в виде $\mathscr{A} = \mathscr{D} - \mathscr{B}$, *где* $\mathscr{D}$ *содержит диагональ матрицы* $\mathscr{A}$, *то матрица* $\tilde{\mathscr{D}} = \operatorname{diag}(1/A_{ii})$ *удовлетворяет неравенству*

$$\rho(|\tilde{\mathscr{D}}||\mathscr{B}|) < 1.$$

Доказательство. Мы снова представим элементы матрицы $\mathscr{A}$ через середину и полуширину в виде $A_{ij} = \langle a_{ij}, r_{ij}\rangle$, $1 \leqslant i, j \leqslant n$. Учтем теперь определение матрицы $\mathscr{B}_p$ в теореме 15.3 и равенство

$$|1/A_{ii}| = 1/(a_{ii} - r_{ii}).$$

Сразу видно, что вещественная матрица $|\tilde{\mathscr{D}}||\mathscr{B}|$ является полношаговой матрицей для полношагового метода, соответствующего матрице $\mathscr{B}_p$. По теореме 3.10 из [325] мы получаем, что $\rho(|\tilde{\mathscr{D}}||\mathscr{B}|) < 1$. ■

Пусть теперь $\mathscr{A} = (A_{ij})$ — вещественная интервальная матрица, x — вещественный интервальный вектор и выполнено

$$d(A_{ij}) \leqslant \alpha_{ij} \| d(x) \|, \quad \alpha_{ij} \geqslant 0, \quad 1 \leqslant i, \quad j \leqslant n. \tag{15}$$

Если $0 \notin A_{ij}$, то из (15) следует соотношение

$$d(1/A_{ij}) \leqslant \hat{\alpha}_{ij} \| d(x) \|, \quad \hat{\alpha}_{ij} \geqslant 0, \quad 1 \leqslant i, \quad j \leqslant n. \tag{16}$$

Пусть $h = (H_i)$ — вещественный интервальный вектор, удовлетворяющий условиям

$$|H_i| \leqslant \beta_i \| d(x) \|, \quad \beta_i \geqslant 0, \quad 1 \leqslant i \leqslant n, \tag{17}$$

$$d(H_i) \leqslant \gamma_i \| d(x) \|, \quad \gamma_i \geqslant 0, \quad 1 \leqslant i \leqslant n. \tag{18}$$

Лемма 4. *Пусть для* $\mathscr{A} = (A_{ij})$ *и* $h = (H_i)$ *выполнены неравенства* (15)—(18). *Если для интервальной матрицы* $\mathscr{A}$ *и интервального вектора* h *можно выполнить метод Гаусса, то результирующая верхняя треугольная матрица* $\tilde{\mathscr{A}} = (\tilde{A}_{ij})$ *и соответствующий вектор* $\tilde{h} = (\tilde{H}_i)$ *удовлетворяют условиям* (15)—(18) *с подходящими константами* $\tilde{\alpha}_{ij} \geqslant 0$, $\tilde{\beta}_i \geqslant 0$, $\tilde{\gamma}_i \geqslant 0$.

Доказательство. Применим математическую индукцию по шагам метода Гаусса (всего их $n-1$). Для каждого шага покажем, что если рассматриваемые условия выполнены для пары вектор — матрица на предыдущем шаге, то они выполнены и для пары, полученной на текущем шаге. Мы проведем доказательство лишь для первого шага. На этом шаге по данной матрице

$\mathscr{A}$ и вектору h вычисляются новая матрица $\mathscr{A}'$ и новый вектор h' по формулам (см. гл. 15)

$$A'_{1j} = A_{1j}, \qquad 1 \leqslant j \leqslant n, \qquad H'_1 = H_1,$$
$$A'_{i1} = 0, \qquad 2 \leqslant i \leqslant n,$$
$$A'_{ij} = A_{ij} - (A_{i1}/A_{11})\, A_{1j}, \qquad 2 \leqslant i,\ j \leqslant n,$$
$$H'_i = H_i - (A_{i1}/A_{11})\, H_1, \qquad 2 \leqslant i \leqslant n.$$

Тогда для $2 \leqslant i,\ j \leqslant n$ мы получаем с помощью (15)—(18) и (2.12), что

$$d(A'_j) = d(A_{ij}) + d((A_{i1}/A_{11})\, A_{1j})$$
$$\leqslant \alpha_{ij} \| d(x) \| + \left| \frac{A_{i1}}{A_{11}} \right| d(A_{1j}) + \left(d(A_{i1}) \left| \frac{1}{A_{11}} \right| + d\left(\frac{1}{A_{11}} \right) | A_{i1} | \right) | A_{1j} |$$
$$\leqslant \tilde{\alpha}_{ij} \| d(x) \|, \qquad \tilde{\alpha}_{ij} \geqslant 0.$$

Аналогичным образом получаем для $2 \leqslant i \leqslant n$, что

$$| H'_i | \leqslant | H_i | + | A_{i1}/A_{11} |\, | H_1 | \leqslant \tilde{\beta}_i \| d(x) \|, \qquad \tilde{\beta}_i \geqslant 0,$$

а также

$$d(H'_i) = d(H_i) + | H_1 | \left(d(A_{i1}) \left| \frac{1}{A_{11}} \right| + d\left(\frac{1}{A_{11}} \right) | A_{i1} | \right)$$
$$+ d(H_1) \left| \frac{A_{i1}}{A_{11}} \right| \leqslant \tilde{\gamma}_i \| d(x) \|^2, \qquad \tilde{\gamma}_i \geqslant 0.$$

Для остальных элементов $\mathscr{A}'$ и h' неравенства (15)—(18) выполнены тривиальным образом. ■

Лемма 5. *Пусть $\mathscr{A} = (\tilde{A}_{ij})$ — верхняя треугольная матрица, такая что $0 \notin \tilde{A}_{ii}$, $1 \leqslant i \leqslant n$, причем она и вектор $h = (\tilde{H}_i)$ удовлетворяют условиям (15)—(18). Тогда интервальный вектор $= (Y_i)$, вычисленный по формулам*

$$Y_n = \frac{\tilde{H}_n}{\tilde{A}_{nn}}, \qquad Y_i = \frac{1}{\tilde{A}_{ii}} \left(\tilde{H}_i - \sum_{j=i+1}^{n} \tilde{A}_{ij} Y_j \right), \qquad 1 \leqslant i \leqslant n-1,$$

удовлетворяет неравенству

$$\| d(y) \| \leqslant c \| d(x) \|^2, \qquad c \geqslant 0.$$

Доказательство. Из (2.12) следует, что

$$d(Y_n) \leqslant d(\tilde{H}_n) \left| \frac{1}{\tilde{A}_{nn}} \right| + d\left(\frac{1}{\tilde{A}_{nn}} \right) | \tilde{H}_n | \leqslant \delta_n \| d(x) \|^2, \qquad \delta_n \geqslant 0,$$

и что

$$| Y_n | = | 1/\tilde{A}_{nn} |\, | \tilde{H}_n | \leqslant \varkappa_n \| d(x) \|.$$

Если теперь верно, что

$$d(Y_i) \leqslant \delta_i \| d(x) \|^2, \quad |Y_i| \leqslant \varkappa_i \| d(x) \|, \quad 1 < i_0 + 1 \leqslant i \leqslant n,$$

то мы получаем

$$d(Y_{i_0}) \leqslant d\left(\frac{1}{\tilde{A}_{i_0 i_0}}\right) \left| \tilde{H}_{i_0} - \sum_{j=i_0+1}^{n} \tilde{A}_{i_0 j} Y_j \right|$$

$$+ \left| \frac{1}{\tilde{A}_{i_0 i_0}} \right| d\left(\tilde{H}_{i_0} - \sum_{j=i_0+1}^{n} \tilde{A}_{i_0 j} Y_j \right) \leqslant \delta_{i_0} \| d(x) \|^2,$$

$$|Y_{i_0}| \leqslant \left| \frac{1}{\tilde{A}_{i_0 i_0}} \right| \left(|\tilde{H}_{i_0}| + \sum_{j=i_0+1}^{n} |\tilde{A}_{i_0 j}| |Y_j| \right) \leqslant \varkappa_{i_0} \| d(x) \|.$$

Отсюда следует, что

$$d(Y_i) \leqslant \delta_i \| d(x) \|^2, \quad 1 \leqslant i \leqslant n,$$

$$\max_{1 \leqslant i \leqslant r} \{d(Y_i)\} \leqslant \max_{1 \leqslant i \leqslant n} \{\delta_i\} \| d(x) \|^2.$$

Теперь нужное утверждение получается из теоремы об эквивалентности норм. ■

Соединяя леммы 4 и 5, имеем следующее утверждение.

Лемма 6. *Пусть неравенства* (15)—(18) *выполнены для* $\mathcal{A} = (A_{ij})$ *и* $h = (H_i)$. *Если для* $\mathcal{A}$ *и* h *можно выполнить метод Гаусса, то результирующий интервальный вектор* y *удовлетворяет неравенству*

$$\| d(y) \| \leqslant c \| d(x) \|^2, \quad c \geqslant 0. \blacksquare$$

После этих подготовительных лемм мы докажем следующее утверждение о методе (14).

Теорема 7. *Пусть* $x^{(0)}$ *— интервальный вектор и* $y_p \in x^{(0)}$ *— корень функции* $f_p(x_p)$. *Пусть интервальное вычисление* $f'_p(x)$ *производной Фреше* $f'_p(x_p)$ *удовлетворяет условиям теоремы* 15.3 *для всех* $x \subseteq x^{(0)}$. *Пусть, кроме того, неравенство* (3.5′) *покомпонентно выполнено для* $f'_p(x^{(0)})$. *Тогда последовательность* $\{x^{(k)}\}_{k=0}^{\infty}$, *вычисленная согласно* (14), *удовлетворяет следующим условиям.*

(18) *Каждое приближение* $x^{(k)}$, $k \geqslant 0$, *содержит корень* y_p.

(19) *Итерационный метод* (14) *может быть выполнен для любого* $k \geqslant 0$, *и имеет место* $\lim_{k \to \infty} x^{(k)} = y_p$.

(20) *Имеет место неравенство*

$$\| d(x^{(k+1)}) \| \leqslant c \| d(x^{(k)}) \|^2, \quad c \geqslant 0,$$

т. е. R-порядок сходимости метода (14) *удовлетворяет неравенству*

$$O_R((14), y) \geqslant 2$$

(*см. приложение* A, *теорема* 2).

Доказательство. Проверка соотношения (8) сделана для $k = 1$ при выводе формул (14). В общем случае доказательство можно провести методом математической индукции.

(19): Теорема 15.3 гарантирует возможность выполнения метода Гаусса. В условиях нашей теоремы диагональ матрицы $f'(x^{(k)})$ не может содержать нулей, так что п. (d) в (14) всегда может быть выполнен. Интервальные векторы, вычисленные согласно (14), удовлетворяют соотношениям

$$x^{(0)} \supseteq x^{(1)} \supseteq \ldots \supseteq x^{(k)} \supseteq x^{(k+1)} \supseteq \ldots .$$

В силу следствия 10.8 эта последовательность сходится к некоторому предельному значению x. Отсюда следует, что $\lim_{k\to\infty} m(x^{(k)}) = m(x)$, $\lim_{k\to\infty} f_p(m(x^{(k)})) = f_p(m(x))$, $\lim_{k\to\infty} f'_p(x^{(k)}) = f'_p(x)$. Так как $w^{(k)}$ непрерывно зависит от $f_p(m(x^{(k)}))$ и $f'_p(x^{(k)})$, отсюда следует, что $\lim_{k\to\infty} w^{(k)} = w$, а также

$$\lim_{k\to\infty} \tilde{x}^{(k)} = \tilde{x} = \{m(x) - w\} \cap x,$$

$$\lim_{k\to\infty} x^{(k)} = x = \{m(x) - \mathcal{D}_p(x)^{-1}\{\mathcal{B}_p(x)(m(x) - \tilde{x}) + f_p(m(x))\}\} \cap \tilde{x}.$$

Из этих равенств следует, что

$$x \subseteq x,$$

$$x \subseteq m(x) - \mathcal{D}_p(x)^{-1}\{\mathcal{B}_p(x)(m(x) - \tilde{x}) + f_p(m(x))\},$$

т. е.

$$x \subseteq m(x) - \mathcal{D}_p(x)^{-1}\{\mathcal{B}_p(x)(m(x) - x) + f_p(m(x))\}. \tag{21}$$

В частности, тогда имеет место

$$m(x) \in m(x) - \mathcal{D}_p(x)^{-1}\{\mathcal{B}_p(x)(m(x) - x) + f_p(m(x))\},$$

откуда

$$o_p \in \mathcal{D}_p(x)^{-1}\{\mathcal{B}_p(x)(m(x) - x) + f_p(m(x))\}.$$

Отсюда следует, что

$$o_p \in \mathcal{B}_p(x)(m(x) - x) + f_p(m(x)).$$

Применяя (2.19) к (21) покомпонентно, получаем, что

$$\begin{aligned} d(x) &\leqslant |\mathscr{D}_{\mathrm{p}}(x)^{-1}|\, d(\mathscr{B}_{\mathrm{p}}(x)(m(x)-x)+\mathscr{f}_{\mathrm{p}}(m(x))) \\ &= |\mathscr{D}_{\mathrm{p}}(x)^{-1}|\,|\mathscr{B}_{\mathrm{p}}(x)|\, d(x) \\ &\leqslant |\mathscr{D}_{\mathrm{p}}(x^{(0)})^{-1}|\,|\mathscr{B}_{\mathrm{p}}(x^{(0)})|\, d(x). \end{aligned}$$

В силу леммы 3 имеем $\rho(|\mathscr{D}_{\mathrm{p}}(x^{(0)})^{-1}|\,|\mathscr{B}_{\mathrm{p}}(x^{(0)})|) < 1$; значит, $d(x) = o_{\mathrm{p}}$. Ввиду $y_{\mathrm{p}} \in x^{(k)}$, $k \geqslant 0$, отсюда следует с помощью (10.27), что $\lim\limits_{k\to\infty} x^{(k)} = y_{\mathrm{p}}$, а это и требовалось показать.

(20): Покажем, что неравенства (15)—(18) с заменой x на $x^{(0)}$ выполнены для

$$\mathscr{A} := \mathscr{f}'_{\mathrm{p}}(x^{(0)}) \text{ и } \mathscr{k} := \mathscr{f}_{\mathrm{p}}(m(x^{(0)})) = (f_i(m(x^{(0)}))).$$

По условию теоремы каждый элемент матрицы $\mathscr{f}'_{\mathrm{p}}(x^{(0)})$ удовлетворяет неравенству (3.5′), откуда следует в силу эквивалентности норм, что

$$d(A_{ij}) \leqslant \alpha_{ij}\, \|d(x^{(0)})\|, \qquad 1 \leqslant i,\ j \leqslant n,$$

т. е. (15). Из теоремы о среднем значении получаем для некоторых $0 < \theta_i < 1$ соотношение

$$\begin{aligned} |H_i| &= |f_i(m(x^{(0)}))| \leqslant |f'_i(y_{\mathrm{p}} + \theta_i(y_{\mathrm{p}} - m(x^{(0)})))|\,|m(x^{(0)}) - y_{\mathrm{p}}| \\ &\leqslant |\mathscr{f}'_i(x^{(0)})|\,|m(x^{(0)}) - y_{\mathrm{p}}|. \end{aligned}$$

Ввиду $m(x^{(0)})$, $y_{\mathrm{p}} \in x^{(0)}$ и теоремы об эквивалентности норм отсюда следует, что

$$|H_i| \leqslant \beta_i\, \|d(x^{(0)})\|, \qquad 1 \leqslant i \leqslant n,$$

т. е. верно (17). Из $d(\mathscr{f}_{\mathrm{p}}(m(x^{(0)}))) = o_{\mathrm{p}}$ тривиально следует, что верно (18). Теперь из леммы 6 следует, что

$$\|d(w^{(0)})\| \leqslant c\, \|d(x^{(0)})\|^2,$$

а п. (c), (d) *из* (14) дают тогда

$$\|d(x^{(1)})\| \leqslant c\, \|d(x^{(0)})\|^2.$$

Эти рассуждения можно повторить для любого $k \geqslant 0$, заменяя соответствующие константы на константы первого шага, так как $x^{(k)} \supseteq x^{(k+1)}$, $k \geqslant 0$. Тем самым доказано неравенство

$$\|d(x^{(k+1)})\| \leqslant c\, \|d(x^{(k)})\|^2, \qquad k \geqslant 0.$$

Теперь наша теорема получается из теоремы 2 приложения A. ■

Опишем теперь класс задач, для которых выполнены условия теоремы 7. Для этого рассмотрим краевую задачу

$$y'' = f(t, y), \quad y(0) = \bar{a}, \quad y(1) = \bar{b}, \quad f_y(t, y) \geqslant 0, \quad t \in (0, 1).$$

Применение к этой задаче обычного разностного метода приводит к системе нелинейных уравнений, удовлетворяющих всем условиям теоремы 7. (В гл. 21 эта система выписана явно.) Мы выбираем $n = 25$ точек в интервале $(0, 1)$ и приводим в табл. 1 результаты последовательных приближений к значению $y\left(\frac{1}{2}\right)$ для задачи $f(t, y) = y + \sin(y)$, $y(0) = 0$, $y(1) = 1$. Интервальный вектор $x^{(0)}$ вычислен по методу из гл. 21. В вычислениях были учтены все ошибки округления.

Таблица 1

k	*Нижняя граница*	*Верхняя граница*
0	−0.5	0.5
1	0.3966601342385	0.4046693140542
2	0.3986876258099	0.3986882685536
3	0.3986880255411	0.3986880255471
4	0.3986880255420	0.3986880255466

Мы вернемся теперь к лемме 10 гл. 11 и используем ее для доказательства некоторых теорем существования решений систем нелинейных уравнений.

Теорема 8. *Пусть отображение* f_p: $\mathfrak{B} \subseteq V_n(\mathbb{R}) \to V_n(\mathbb{R})$ *непрерывно дифференцируемо в* $\mathfrak{B}$ *и для некоторого* $x \subseteq \mathfrak{B}$ *существует вычисление производной в интервальной арифметике. Пусть также метод Гаусса применим к* $f'_p(x)$ *и его применение к* $f'_p(x)$ *и правой части* $f_p(y_p)$ *для фиксированного* $y_p \in x$ *дает в результате интервальный вектор* z. *Тогда из* $y_p - z \subseteq x$ *следует, что* f_p *имеет корень в* x, *а из* $x \cap \{y_p - z\} = \varnothing$ *следует, что* f_p *не имеет корня в* x.

Доказательство. Мы начнем с уравнения (2), записанного в матричном виде

$$f_p(x_p) - f_p(y_p) = \mathcal{J}_p(x_p)(x_p - y_p).$$

Из того что для матрицы $f'_p(x)$ можно выполнить метод Гаусса, следует, что все точечные матрицы из $f'_p(x)$, в частности матрица $\mathcal{J}_p(x_p)$, являются невырожденными. Рассмотрим отображение

$$r_p : x \subset V_n(\mathbb{R}) \to V_n(\mathbb{R}),$$

такое что

$$\mathcal{N}_{\mathrm{p}}(x_{\mathrm{p}}) = x_{\mathrm{p}} - \mathcal{J}_{\mathrm{p}}(x_{\mathrm{p}})^{-1}\, \mathcal{F}_{\mathrm{p}}(x_{\mathrm{p}}).$$

Из этой формулы следует, что

$$\begin{aligned}\mathcal{N}_{\mathrm{p}}(x_{\mathrm{p}}) &= x_{\mathrm{p}} - \mathcal{J}_{\mathrm{p}}(x_{\mathrm{p}})^{-1}\, \mathcal{F}_{\mathrm{p}}(y_{\mathrm{p}}) + \mathcal{J}_{\mathrm{p}}(x_{\mathrm{p}})^{-1}\,(\mathcal{F}_{\mathrm{p}}(y_{\mathrm{p}}) - \mathcal{F}_{\mathrm{p}}(x_{\mathrm{p}})) \\ &= y_{\mathrm{p}} - \mathcal{J}_{\mathrm{p}}(x_{\mathrm{p}})^{-1}\, \mathcal{F}_{\mathrm{p}}(y_{\mathrm{p}}) \in y_{\mathrm{p}} - z_{\mathrm{p}}.\end{aligned}$$

Поэтому из $y_{\mathrm{p}} - z \subseteq x$ следует, что $\mathcal{N}_{\mathrm{p}}(x_{\mathrm{p}}) \in x$ для $x_{\mathrm{p}} \in x$, и из леммы 11.10 следует, что $\mathcal{F}_{\mathrm{p}}$ имеет корень в x.

Чтобы доказать вторую половину теоремы, мы предположим, что $\mathcal{F}_{\mathrm{p}}$ имеет корень x^*_{p} в x. Тогда, полагая $x_{\mathrm{p}} = x^*_{\mathrm{p}}$ в формуле, определяющей отображение $\mathcal{N}_{\mathrm{p}}$, получаем

$$\mathcal{N}_{\mathrm{p}}(x^*_{\mathrm{p}}) = x^*_{\mathrm{p}} \in y_{\mathrm{p}} - z.$$

Так как $x^*_{\mathrm{p}} \in x$, мы имеем $x \cap \{y_{\mathrm{p}} - z\} \neq \varnothing$, что дает противоречие. ■

Из доказательства предыдущей теоремы видно, что можно избежать применения метода Гаусса, если известна интервальная матрица $\mathcal{Y}$, такая что

$$\mathcal{J}_{\mathrm{p}}(x_{\mathrm{p}})^{-1} \in \mathcal{Y} \quad \text{для} \quad x_{\mathrm{p}} \in x.$$

Тогда с помощью отображения $\mathcal{N}_{\mathrm{p}}$, введенного в доказательстве предшествующей теоремы, мы получим соотношение

$$\mathcal{N}_{\mathrm{p}}(x_{\mathrm{p}}) = x_{\mathrm{p}} - \mathcal{J}_{\mathrm{p}}(x_{\mathrm{p}})^{-1}\, \mathcal{F}_{\mathrm{p}}(x_{\mathrm{p}}) \in y_{\mathrm{p}} - \mathcal{Y}\mathcal{F}_{\mathrm{p}}(y_{\mathrm{p}}).$$

Итак, мы имеем следующее утверждение.

Теорема 9. *Пусть отображение* $\mathcal{F}_{\mathrm{p}}: \mathfrak{B} \subseteq V_n(\mathbb{R}) \to V_n(\mathbb{R})$ *непрерывно дифференцируемо в* $\mathfrak{B}$, *причем производная имеет вычисление в интервальной арифметике для некоторого* $x \subseteq \mathfrak{B}$. *Пусть интервальная матрица* $\mathcal{Y}$ *такова, что* $\mathcal{J}_{\mathrm{p}}(x_{\mathrm{p}})^{-1} \in \mathcal{Y}$ *для* $x_{\mathrm{p}} \in x$. *Если теперь*

$$y_{\mathrm{p}} - \mathcal{Y}\mathcal{F}_{\mathrm{p}}(y_{\mathrm{p}}) \subseteq x$$

для фиксированного $y_{\mathrm{p}} \in x$, *то в* $\mathcal{F}_{\mathrm{p}}$ *имеется нуль функции* x. *Если же*

$$\{y_{\mathrm{p}} - \mathcal{Y}\mathcal{F}_{\mathrm{p}}(y_{\mathrm{p}})\} \cap x = \varnothing,$$

то в x *нет нулей функции* $\mathcal{F}_{\mathrm{p}}$. ■

В общем случае мы будем вычислять матрицу $\mathcal{Y}$ со свойствами, нужными в этой теореме, применяя метод Гаусса n раз к интервальной матрице $\mathcal{F}'_{\mathrm{p}}(x)$ со столбцами единичной матрицы в качестве правых частей. Утверждение, соответствующее пре-

дыдущей теореме, но не требующее применения метода Гаусса к интервальной матрице, было доказано в [448]. См. также [449].

Теорема 10. *Пусть отображение* $f_p: \mathfrak{B} \subseteq V_n(\mathbb{R}) \to V_n(\mathbb{R})$ *непрерывно дифференцируемо в* $\mathfrak{B}$, *причем производная имеет вычисление в интервальной арифметике для некоторого* $x \subseteq \mathfrak{B}$.

Если включение

$$k_p(x) = y_p - \mathcal{Y}_p \cdot f_p(y_p) + (\mathcal{I}_p - \mathcal{Y}_p \cdot f_p'(x))(x - y_p) \subseteq x$$

имеет место для некоторого $y_p \in x$, *невырожденной матрицы* $\mathcal{Y}_p \in M_{nn}(\mathbb{R})$ *и единичной матрицы* $\mathcal{I}_p$, *то* f_p *имеет корень в* x. *Если же*

$$\{y_p - \mathcal{Y}_p \cdot f_p(y_p) + (\mathcal{I}_p - \mathcal{Y}_p \cdot f_p'(x))(x - y_p)\} \cap x = \varnothing,$$

то f_p *не имеет корней в* x. ■

Очевидным кандидатом на роль $\mathcal{Y}_p$ является $m(f_p'(x))^{-1}$. Нахождение этой матрицы требует обращения точечной матрицы, а также некоторого числа умножений матриц на матрицы и некоторого числа умножений матриц на векторы для проверки включений. Существенную часть этих умножений можно сэкономить, если применить метод Гаусса к некоторой системе линейных уравнений, у которой матрица коэффициентов точечная, а правая часть — интервальный вектор. Это замечание уточняется в следующем утверждении.

Теорема 11. *Пусть отображение* $f_p: \mathfrak{B} \subseteq V_n(\mathbb{R}) \to V_n(\mathbb{R})$ *непрерывно дифференцируемо в* $\mathfrak{B}$, *причем производная имеет вычисление в интервальной арифметике для некоторого* $x \subseteq \mathfrak{B}$. *Предположим еще, что* $\mathcal{Y}_p$ *— невырожденная вещественная точечная матрица и что метод Гаусса в применении к точечной матрице* $\tilde{\mathcal{Y}}_p$ *и правой части* $f_p(y_p) - (\tilde{\mathcal{Y}}_p - f_p'(x))(x - y_p)$ *для некоторого* $y_p \in x$ *дает интервальный вектор* z. *Если теперь выполнено* $y_p - z \subseteq x$, *то* f_p *имеет нуль в* x. *Если же* $x \cap \{y_p - z\} = \varnothing$, *то* f_p *не имеет нулей в* x.

Доказательство. Рассмотрим отображение

$$r_p: x \subseteq \mathfrak{B} \subseteq V_n(\mathbb{R}) \to V_n(\mathbb{R}),$$

определяемое формулой

$$r_p(x_p) = x_p - \tilde{\mathcal{Y}}_p^{-1} f_p(x_p).$$

Мы видим, что верно равенство

$$r_p(x_p) = y_p - \tilde{\mathcal{Y}}_p^{-1} f_p(y_p) + x_p - y_p - \tilde{\mathcal{Y}}_p^{-1}(f_p(x_p) - f_p(y_p))$$
$$= y_p - \tilde{\mathcal{Y}}_p^{-1} f_p(y_p) + \tilde{\mathcal{Y}}_p^{-1}(\tilde{\mathcal{Y}}_p - \mathcal{J}_p(x_p))(x_p - y_p),$$

т. е. выполнено равенство

$$\tilde{\mathcal{Y}}_p(f_p(x_p) - y_p) = -f_p(y_p) + (\tilde{\mathcal{Y}}_p - \mathcal{J}_p(x_p))(x_p - y_p).$$

Ввиду $x_p \in x$ и $\mathcal{J}_p(x_p) \in f'_p(x)$ мы получаем, что

$$-f_p(y_p) + (\tilde{\mathcal{Y}}_p - \mathcal{J}_p(x_p))(x_p - y_p) \in$$
$$-f_p(y_p) + (\tilde{\mathcal{Y}}_p - f'_p(x))(x - y_p).$$

Если теперь применить метод Гаусса к системе линейных уравнений, заданной матрицей $\tilde{\mathcal{Y}}_p$ и правой частью $f_p(y_p) - (\tilde{\mathcal{Y}}_p - f'_p(x))(x - y_p)$, то, мы получим интервальный вектор z, такой что имеет место $f_p(x_p) - y_p \in -z$ для $x_p \in x$. Если теперь $y_p - z \subseteq x$, то отсюда следует, что $f_p(x) \in x$ для $x_p \in x$. Теперь можно завершить доказательство так же, как в предыдущей теореме. ■

Основное достоинство этой теоремы заключается не в экономии вычислений по сравнению с процедурой из теоремы 10. По сравнению с аналогичным результатом из теоремы 8 главное достижение состоит в том, что метод Гаусса всегда применим к невырожденной точечной матрице. С другой стороны, условие $k_p(x) \subseteq x$ слабее, чем требование $y_p = m(x)$ в теореме 11, по крайней мере при выборе $y_p = m(x)$ (середина вектора x) и $\mathcal{Y}_p^{-1} = m(f'(x)) = \tilde{\mathcal{Y}}$. В этом можно убедиться следующим образом. Используя введенное в гл. 15 обозначение z для вектора, вычисленного методом Гаусса, мы имеем для произвольной неособенной матрицы $\mathcal{Y}_p$

$$z = g_p(\mathcal{Y}_p, f_p(y_p) - (\mathcal{Y}_p - f'_p(x))(x - y_p)).$$

Применяя доказанное в гл. 15 свойство (3) алгоритма Гаусса, мы получим

$$\mathcal{Y}_p^{-1}(f_p(y_p) - (\mathcal{Y}_p - f'(x))(x - y_p)) \subseteq$$
$$g_p(\mathcal{Y}_p, f_p(y_p) - (\mathcal{Y}_p - f'_p(x))(x - y_p)),$$

т. е. мы имеем включение

$$y_p - \mathcal{Y}_p^{-1}(f_p(y_p) - (\mathcal{Y}_p - f'_p)(x))(x - y_p)) \subseteq y_p - z.$$

Мы покажем теперь, что при $y_p = m(x)$, $\mathcal{Y}_p = m(f'_p(x))$ левая часть этого соотношения равна значению $k_p(x)$ из теоремы 10. Используя симметричность выражений $m(f'_p(x)) - f'(x)$ и $x - m(x)$, а также (10.8) и последнюю формулу из (10.9), по-

лучим

$$m(x) - m\left(f'_{\mathrm{p}}(x)\right)^{-1}\left(f_{\mathrm{p}}(m(x)) - \left(m\left(f'_{\mathrm{p}}(x)\right) - \left(f'_{\mathrm{p}}(x)\right)(x - m(x))\right)\right)$$
$$= m(x) - m\left(f'_{\mathrm{p}}(x)\right)^{-1} f_{\mathrm{p}}(m(x)) - \left(\mathcal{I}_{\mathrm{p}} - m\left(f'_{\mathrm{p}}(x)\right)^{-1} f'_{\mathrm{p}}(x)\right)$$
$$\times (x - m(x)) = k_{\mathrm{p}}(x).$$

Рассмотрим теперь некоторые теоремы существования, в которых участвует вычисление в интервальной арифметике второй производной. Для этого предположим, что отображение $f_{\mathrm{p}}: \mathfrak{B} \subseteq V_n(\mathbb{R}) \to V_n(\mathbb{R})$ дважды непрерывно дифференцируемо и что вторая производная $f''_{\mathrm{p}}(x_{\mathrm{p}}) \in M_n(\mathbb{R})$ имеет вычисление в интервальной арифметике. Для некоторого $x \subseteq \mathfrak{B}$ и произвольного вначале вектора $m(x) \in x$ мы рассмотрим следующие интервальные векторы.

(22) $k_{\mathrm{p1}}(x) := m(x) - \mathcal{Y}_{\mathrm{p}} \cdot f_{\mathrm{p}}(m(x)) - \mathcal{Y}_{\mathrm{p}}\left(f''_{\mathrm{p}}(x)(x - m(x))(x - m(x))\right)$,

где $\mathcal{Y}_{\mathrm{p}} = f'_{\mathrm{p}}(m(x))^{-1}$,

(23) $k_{\mathrm{p2}}(x) := m(x) - \mathcal{Y}_{\mathrm{p}} \cdot f_{\mathrm{p}}(m(x)) - \frac{1}{2}\left(\mathcal{Y}_{\mathrm{p}} \cdot f''_{\mathrm{p}}(x)\right)(x - m(x))$
$\times (x - m(x))$

где $\mathcal{Y}_{\mathrm{p}} = f'_{\mathrm{p}}(m(x))^{-1}$,

(24) $k_{\mathrm{p3}}(x) := m(x) - \mathcal{Y}_{\mathrm{p}} \cdot f_{\mathrm{p}}(m(x))$
$+ \left\{\mathcal{I}_{\mathrm{p}} - \mathcal{Y}_{\mathrm{p}} \cdot f'_{\mathrm{p}}(m(x)) - \frac{1}{2}\left(\mathcal{Y}_{\mathrm{p}} \cdot f''_{\mathrm{p}}(x)\right)(x - m(x))\right\}$
$\times (x - m(x))$

с произвольной невырожденной матрицей $\mathcal{Y}_{\mathrm{p}}$.

Мы докажем сначала следующее утверждение, аналогичное предыдущей теореме.

Теорема 12. *Пусть отображение $f_{\mathrm{p}}: \mathfrak{B} \subseteq V_n(\mathbb{R}) \to V_n(\mathbb{R})$ дважды непрерывно дифференцируемо в $\mathfrak{B}$. Если для некоторого i, $1 \leqslant i \leqslant 3$, и $x \subseteq \mathfrak{B}$ верно*

$$k_{\mathrm{p}i}(x) \subseteq x,$$

то f_{p} имеет нуль в x. Если же $k_{\mathrm{p}i}(x) \cap x = \varnothing$ для некоторого i, $1 \leqslant i \leqslant 3$, то f_{p} не имеет нулей в x.

Доказательство. (22): Это случай $i = 1$. Снова введем отображение

$$r_{\mathrm{p}}(x_{\mathrm{p}}) = x_{\mathrm{p}} - \mathcal{Y}_{\mathrm{p}} \cdot f_{\mathrm{p}}(x_{\mathrm{p}})$$

и воспользуемся представлением

$$r_{\mathrm{p}}(x_{\mathrm{p}}) = m(x) - \mathcal{Y}_{\mathrm{p}} \cdot f_{\mathrm{p}}(m(x)) + (x_{\mathrm{p}} - m(x)) - \mathcal{Y}_{\mathrm{p}}(f_{\mathrm{p}}(x_{\mathrm{p}})$$
$$- f_{\mathrm{p}}(m(x)))$$

для $x_p \in x$. Применяя теорему о среднем значении, получим

$$f_p(x_p) - f_p(m(x)) = \mathcal{J}_p(x_p, m(x))(x_p - m(x)).$$

Здесь i-я строка матрицы $\mathcal{J}_p(x_p, m(x))$ равна $f_i'(z_p^{(i)})$, где $z_p^{(i)} = m(x) + \theta_i(x_p - m(x))$ и $0 < \theta_i < 1$, $1 \leqslant i \leqslant n$. Далее мы имеем, что

$$f_i'(z_p^{(i)}) = f_i'(m(x)) + \int_0^1 f_i''(m(x) + t(z_p^{(i)} - m(x)))(z_p^{(i)} - m(x))\,dt,$$

т. е.

$$\mathcal{J}_p(x_p, m(x)) = f_p'(m(x)) + \mathcal{B}_p,$$

где i-я строка матрицы $\mathcal{B}_p$ равна

$$\int_0^1 f_i''(m(x) + t(z_p^{(i)} - m(x)))(z_p^{(i)} - m(x))\,dt.$$

Отсюда следует, что

$$n_p(x_p) = m(x) - \mathcal{Y}_p \cdot f_p(m(x)) + \{\mathcal{J}_p - \mathcal{Y}_p(f_p'(m(x)) - \mathcal{B}_p)\}(x_p - m(x)).$$

Полагая

$$\mathcal{Y}_p = f_p'(m(x))^{-1}$$

мы получим окончательно

$$n_p(x_p) = m(x) - \mathcal{Y}_p \cdot f_p(m(x)) - \mathcal{Y}_p\mathcal{B}_p(x_p - m(x)).$$

Так как $t \in [0, 1]$ и $z_p^{(i)} \in x$, мы имеем $\mathcal{B}_p \in f_p''(x)(x - m(x))$ и потому получаем

$$n_p(x_p) \in m(x) - \mathcal{Y}_p \cdot f_p(m(x)) - \mathcal{Y}_p(f_p''(x)(x - m(x))(x - m(x))) = k_{p1}(x).$$

(23) и (24): Для $k_{p2}(x)$ и $k_{p3}(x)$ можно рассуждать аналогично тому, как было сделано для $k_{p1}(x)$. Для $1 \leqslant i \leqslant n$ мы имеем

$$f_i(x_p) - f_i(m(x)) = f_i'(m(x))(x_p - m(x)) + \frac{1}{2} f_i''(m(x) + \theta_i(x_p - m(x)))(x_p - m(x))(x_p - m(x)),$$

$$0 < \theta_i < 1,$$

т. е.

$$f_p(x_p) - f_p(m(x)) = f_p'(m(x))(x_p - m(x)) + \frac{1}{2}\mathcal{B}_p(x - m(x))_p(x_p - m(x))$$

с билинейным оператором $\mathscr{B}_\mathrm{p}$. Далее отсюда следует, что

$$\mathscr{Y}_\mathrm{p}(\mathscr{f}_\mathrm{p}(x_\mathrm{p}) - \mathscr{f}_\mathrm{p}(m(x))) = \mathscr{Y}_\mathrm{p} \cdot \mathscr{f}'_\mathrm{p}(m(x))(x - m(x))$$
$$+ \frac{1}{2}\mathscr{Y}_\mathrm{p}\mathscr{B}_\mathrm{p}(x_\mathrm{p} - m(x))(x_\mathrm{p} - m(x)),$$

и потому ввиду $\mathscr{B}_\mathrm{p} \in \mathscr{f}''_\mathrm{p}(x)$ мы получаем

$$\mathscr{r}_\mathrm{p}(x_\mathrm{p}) = m(x) - \mathscr{Y}_\mathrm{p} \cdot \mathscr{f}_\mathrm{p}(m(x)) + \{\mathscr{I}_\mathrm{p} - \mathscr{Y}_\mathrm{p} \cdot \mathscr{f}'_\mathrm{p}(m(x))$$
$$- \frac{1}{2}(\mathscr{Y}_\mathrm{p}\mathscr{B}_\mathrm{p})(x_\mathrm{p} - m(x))\}(x_\mathrm{p} - m(x))$$
$$\in m(x) - \mathscr{Y}_\mathrm{p} \cdot \mathscr{f}_\mathrm{p}(m(x)) + \left\{\mathscr{I}_\mathrm{p} - \mathscr{Y}_\mathrm{p} \cdot \mathscr{f}'_\mathrm{p}(m(x))\right.$$
$$\left. - \frac{1}{2}(\mathscr{Y}_\mathrm{p} \cdot \mathscr{f}'_\mathrm{p}(x))(x - m(x))\right\}(x - m(x))$$
$$= k_{\mathrm{p}3}(x).$$

Полагая $\mathscr{Y}_\mathrm{p} = \mathscr{f}'_\mathrm{p}(m(x))^{-1}$, мы получаем $\mathscr{r}_\mathrm{p}(x_\mathrm{p}) \in k_{\mathrm{p}2}(x)$.

Теперь первая часть нашей теоремы доказывается с помощью теоремы Брауэра о неподвижной точке. Вторая часть устанавливается так же, как вторая часть теоремы 8. ■

Теперь мы сравним некоторые из этих теорем существования с некоторыми результатами, имеющимися в литературе. Определим замкнутый шар с центром y_p и радиусом r:

$$\zeta(y_\mathrm{p}, r) := \{x_\mathrm{p} \in V_n(\mathbb{R}) \mid \| x_\mathrm{p} - y_\mathrm{p} \| \leqslant r\}.$$

Под нормой *всегда* понимаем ∞-нормы. Поэтому можно рассматривать $\zeta(y_\mathrm{p}, r)$ как интервальный вектор, все компоненты которого имеют ширину $2r$. Нормы линейных и билинейных операторов определяются естественным образом исходя из векторной ∞-нормы. Первое из приводимых ниже уравнений содержит теорему существования Ньютона — Канторовича (см., например, [415, с. 680, теорема 6]).

Теорема 13. *Пусть* $x = \zeta(m(x), r) \in V_n(I)(\mathbb{R}))$ *отображение* $\mathscr{f}_\mathrm{p} : V_n(\mathbb{R}) \to V_n(\mathbb{R})$ *дважды непрерывно дифференцируемо и существует* $\mathscr{Y}_\mathrm{p} := \mathscr{f}'_\mathrm{p}(m(x))^{-1}$.

Если выполнены условия

$$\| \mathscr{Y}_\mathrm{p} \cdot \mathscr{f}_\mathrm{p}(m(x)) \| \leqslant \eta, \tag{25}$$

$$\| \mathscr{Y}_\mathrm{p} \cdot \mathscr{f}'_\mathrm{p}(x_\mathrm{p}) \| \leqslant k \quad \text{для} \quad x_\mathrm{p} \in \zeta(m(x), r), \tag{26}$$

$$h := k\eta \leqslant \frac{1}{2}, \tag{27}$$

$$r \geqslant ((1 - \sqrt{1 - 2h})/h)\,\eta, \tag{28}$$

то $\mathscr{f}_\mathrm{p}$ *имеет нуль в* $\zeta(m(x), r)$. ■

В следующей теореме изучается соотношение между теоремами 12 и 13. В теореме 12 в качестве $12m(x)$ всегда берется центр.

Теорема 14. (a) *Пусть* $x = \zeta(m(x), r)$, *где* $r := \frac{1}{2}\|d(x)\|$, *и для* $r_i := \frac{1}{2}\|d(k_{pi}(x))\|$ *имеет место* $\zeta(m(k_{pi}(x)), r_i) \subseteq \zeta(m(x), r) = x$ *для* $i=1$, *или* 2. *Тогда выполнены условия* (25)—(28) *теоремы* 13.

(b) *Для* $i=2$ *верно и обратное утверждение. Если выполнены условия* (25)—(28) *теоремы* 14 *и дополнительное ограничение*

$$r \leqslant ((1+\sqrt{1-2h_2})/h_2)\,\eta,$$

то $\zeta(m(k_{p2}(x)), r_2) \subseteq \zeta(m(x), r) = x$. *В частности,* $k_{p2}(x) \subseteq x$.

Доказательство. (a) Включение $\zeta(m(k_{pi}(x)), r_i) \subseteq \zeta(m(x), r)$ выполнено тогда и только тогда, когда $\|m(k_{pi}(x)) - m(x)\| + r_i \leqslant r$.

Из того, что

$$m(k_{pi}(x)) = m(x) - \mathscr{Y}_p \cdot \mathscr{f}_p(m(x))$$

следует

$$\eta := \|\mathscr{Y}_p \cdot \mathscr{f}_p(m(x))\| = \|m(k_{pi}(x)) - m(x)\|.$$

В силу теоремы 10.12(b), (c) ширина векторов $k_{p1}(x))$ и $k_{p2}(x)$ удовлетворяет неравенствам

$$d(k_{p1}(x)) = \tfrac{1}{2}|\mathscr{Y}_p||\mathscr{f}''_p(x)|\,d(x)\,d(x),$$

$$d(k_{p2}(x)) = \tfrac{1}{4}|\mathscr{Y}_p \cdot \mathscr{f}''_p(x)|\,d(x)\,d(x).$$

Полагая

$$\mathscr{C}_p = |\mathscr{Y}_p||\mathscr{f}''_p(x)| = (c_{ijk}), \quad k_1 = \max_{1\leqslant i\leqslant n} \sum_{j,k=1}^{n} c_{ijk}$$

и

$$\mathscr{D}_p = |\mathscr{Y}_p \cdot \mathscr{f}''_p(x)| = (d_{ijk}), \quad k_2 = \max_{1\leqslant i\leqslant n} \sum_{j,k=1}^{n} d_{ijk},$$

получим

$$\|d(k_{p1}(x))\| = 2k_1r^2 \quad \text{и} \quad \|d(k_{p2}(x))\| = k_2r^2,$$

так как $d(x) = 2r(1, 1, \ldots, 1)^T$. Поэтому имеем

$$r_1 = \tfrac{1}{2}\|d(k_p(x))\| = k_1r^2, \quad r_2 = \tfrac{1}{2}\|d(k_{p2}(x))\| = \tfrac{1}{2}k_2r^2.$$

Наконец, условие теоремы $\zeta(m(k_{pi}(x)), r_i) \subseteq \zeta(m(x), r)$ дает

$$\eta + k_1r^2 \leqslant r \quad \text{для} \quad i = 1$$

и

$$\eta + \frac{1}{2} k_2 r^2 \leqslant r \quad \text{для} \quad i = 2.$$

Так как

$$\eta + \frac{1}{2} k_1 r^2 \leqslant \eta + k_1 r^2,$$

мы получаем

$$\eta + \frac{1}{2} k_i r^2 \leqslant r, \quad i = 1 \text{ или } 2.$$

Отсюда следует, что

$$h_i = k_i \eta \leqslant \frac{1}{2}, \quad i = 1 \text{ или } 2,$$

и

$$\frac{1 - \sqrt{1 - 2h_i}}{h_i} \eta \leqslant r \leqslant \frac{1 + \sqrt{1 - 2h_i}}{h_i} \eta, \quad i = 1 \text{ или } 2.$$

Это показывает, что выполнены условия (25), (27) и (28) теоремы 13. Остается проверить (26). Из $x_p \in \zeta(m(x), r) = x$ следует, что

$$\mathscr{Y}_p \cdot f''_p(x_p) \in \mathscr{Y}_p \cdot f''_p(x),$$

а потому и

$$|\mathscr{Y}_p \cdot f''_p(x_p)| \leqslant |\mathscr{Y}_p \cdot f''_p(x)| \leqslant |\mathscr{Y}_p| |f''_p(x)|.$$

Отсюда следует, что

$$\|\mathscr{Y}_p \cdot f''_p(x_p)\| = \||\mathscr{Y}_p \cdot f''_p(x_p)|\| \leqslant \||\mathscr{Y}_p \cdot f''_p(x)|\| = \|\mathscr{D}_p\|$$

$$\leqslant \max_{1 \leqslant i \leqslant n} \sum_{j,k=1}^{n} d_{ijk}$$

и

$$\|\mathscr{Y}_p \cdot f''_p(x_p)\| \leqslant \||\mathscr{Y}_p| |f''_p(x_p)|\| = \|\mathscr{C}_p\|$$

$$\leqslant \max_{1 \leqslant i \leqslant n} \sum_{j,k=1}^{n} c_{ijk}.$$

Этим завершается доказательство (a).

(b) Заметим, что из условий теоремы следует неравенство

$$\eta + \frac{1}{2} k_2 r^2 \leqslant r,$$

эквивалентное включению

$$\zeta(m(k_{p2}(x)), r_2) \subseteq \zeta(m(x), r) = x.$$

Ввиду $k_{p2}(x) \subseteq \zeta(m(k_{p2}(x)), r_2)$ отсюда следует, что $k_{p2}(x) \subseteq x$. ■

Сравнение теоремы 10, где участвует $k_p(x)$, и теоремы Ньютона — Канторовича было проведено в [472]. Условие $k_p(x) \subseteq x$ приводит в случае

$$x = \zeta(m(x), r)$$

к

$$((1 - \sqrt{1-4h})/2h)\,\eta \leqslant r \leqslant ((1 + \sqrt{1-4h})/2h)\,\eta.$$

Здесь предполагается, что $h := b\lambda\eta \leqslant \frac{1}{4}$, где λ — некоторая константа Липшица для первой производной, и

$$\|\mathscr{Y}_p\| \leqslant b, \quad \|\mathscr{Y}_p \cdot f_p(m(x))\| \leqslant \eta \quad (\mathscr{Y}_p = f'_p(m(x))^{-1}).$$

Вопрос о том, когда для $k_{p1}(x) \subseteq x$ верно включение $x = \zeta(m(x), r)$, приводит по существу к тем же условиям. $k_{p1}(x) \subseteq x$ наверняка выполнено, если имеет место $\zeta(m(k_{p1}(x)), r_1) \subseteq \zeta(m(x), r)$. Как и в доказательстве п. (а) предыдущей теоремы, это эквивалентно неравенству

$$\eta + k_1 r^2 \leqslant r,$$

что приводит к оценке

$$h_1 = k_1\eta \leqslant \frac{1}{4}$$

и

$$((1 - \sqrt{1-4h_1})/2h_1)\,\eta \leqslant r \leqslant ((1 + \sqrt{1-4h_1})/2h_1)\,\eta.$$

Если теперь заметить, что λb является по существу верхней границей для k_1, то для включения $k_{p1}(x) \subseteq x$ получатся те же достаточные условия, что и для $k_p(x) \subseteq x$.

Отметим, наконец, что в случае одного уравнения с одним неизвестным условие $k_{p2}(x) \subseteq x$ выполнено тогда и только тогда, когда верно

$$\zeta(m(k_{p2}(x)), r_2) \subseteq \zeta(m(x), r).$$

Вместе с предыдущей теоремой это дает следующее утверждение.

Следствие 15. *В случае одного уравнения с одним неизвестным условия теоремы Ньютона — Канторовича выполнены тогда и только тогда, когда* $k_{p2}(x) \subseteq x = \zeta(m(x), r)$. ■

Мы воспроизведем теперь один результат, чтобы сравнить его впоследствии с теоремой 12 для случая $k_{p3}(x)$.

Теорема 16. *Пусть* $x = \zeta(m(x), r) \in V_n(I(\mathbb{R}))$ *и отображение* $f_p\colon V_n(\mathbb{R}) \to V_n(\mathbb{R})$ *дважды дифференцируемо. Допустим, что для*

$\mathscr{Y}_p \in M_{nn}(\mathbb{R})$ выполнены следующие условия

$$\|\mathscr{Y}_p \cdot \mathscr{f}_p(m(x))\| \leqslant \bar{\eta}, \tag{29}$$

$$\|\mathscr{I}_p - \mathscr{Y}_p \cdot \mathscr{f}'_p(m(x))\| \leqslant \delta < 1, \tag{30}$$

$$\|\mathscr{Y}_p \cdot \mathscr{f}''_p(x_p)\| \leqslant \bar{k} \text{ для всех } x_p \in \zeta(m(x), r), \tag{31}$$

$$h := \bar{k}\bar{\eta}/(1-\delta)^2 \leqslant \frac{1}{2}, \tag{32}$$

$$r \geqslant \bar{r}_0 = ((1-\sqrt{1-2h})/h)(\bar{\eta}/(1-\delta)). \tag{33}$$

Тогда функция $\mathscr{f}_p$ имеет нуль в $\zeta(m(x), r)$. ■

Доказательство приведено в [415] и состоит по существу в проверке условий теоремы 13.

В формулировке следующей теоремы используется возможность записать $\mathscr{k}_{p3}(x)$ в виде

$$\mathscr{k}_{p3}(x) = m(x) - \mathscr{Y}_p \cdot \mathscr{f}_p(m(x)) + (\mathscr{I}_p - \mathscr{Y}_p \cdot \mathscr{f}'_p(m(x)))(x - m(x)) + \frac{1}{2}(\mathscr{Y}_p \cdot \mathscr{f}''_p(x))(x - m(x))(x - m(x)),$$

где $m(x)$ — центр x. Чтобы показать это, заметим, что величина $x - m(x)$ симметрична, так что исходное определение $\mathscr{k}_{p3}(x)$ можно переписать в виде

$$\mathscr{k}_{p3}(x) = m(x) - \mathscr{Y}_p \cdot \mathscr{f}_p(m(x)) + |\mathscr{I}_p - \mathscr{Y}_p \cdot \mathscr{f}'_p(m(x)) - \frac{1}{2}(\mathscr{Y}_p \cdot \mathscr{f}''_p(x))(x - m(x))|(x - m(x)).$$

Далее мы имеем

$$|\mathscr{I}_p - \mathscr{Y}_p \cdot \mathscr{f}'_p(m(x)) - \frac{1}{2}(\mathscr{Y}_p \cdot \mathscr{f}''_p(x))(x - m(x))|$$

$$= |\mathscr{I}_p - \mathscr{Y}_p \cdot \mathscr{f}'_p(m(x))| + \frac{1}{4} d((\mathscr{Y}_p \cdot \mathscr{f}''_p(x))(x - m(x)))$$

$$= |\mathscr{I}_p - \mathscr{Y}_p \cdot \mathscr{f}'_p(m(x))| + \frac{1}{4} |\mathscr{Y}_p \cdot \mathscr{f}''_p(x)| d(x).$$

Так как обе матрицы, входящие в эту сумму, неотрицательны, мы можем записать $\mathscr{k}_{p3}(x)$ в виде

$$\mathscr{k}_{p3}(x) = m(x) - \mathscr{Y}_p \cdot \mathscr{f}_p(m(x)) + |\mathscr{I}_p - \mathscr{Y}_p \cdot \mathscr{f}'_p(m(x))|(x - m(x)) + \frac{1}{4}|\mathscr{Y}_p \cdot \mathscr{f}''_p(x)| d(x)(x - m(x)),$$

что и дает равенство

$$\mathscr{k}_{p3} = m(x) - \mathscr{Y}_p \cdot \mathscr{f}_p(m(x)) + (\mathscr{I}_p - \mathscr{Y}_p \cdot \mathscr{f}'_p(m(x)))(x - m(x)) + \frac{1}{2}(\mathscr{Y}_p \cdot \mathscr{f}''_p(x))(x - m(x))(x - m(x)).$$

Теперь мы можем сформулировать следующее утверждение.

Теорема 17. *Положим*

$$r' = \frac{1}{2}\| d\left(m(x) - \mathscr{Y}_p \cdot \mathscr{f}_p(m(x)) + \frac{1}{2}(\mathscr{Y}_p \cdot \mathscr{f}''_p(x)) \times (x - m(x))(x - m(x))\right)\|,$$

$$r'' = \frac{1}{2}\| d\left((\mathscr{I}_p - \mathscr{Y}_p \cdot \mathscr{f}'_p(m(x)))(x - m(x))\right)\|,$$

$$r_3 = r' + r'',$$

$$r = \frac{1}{2}\| d(x)\|,$$

$$\delta = \|\mathscr{I}_p - \mathscr{Y}_p \cdot \mathscr{f}'_p(m(x))\| < 1$$

и допустим, что имеет место

$$\zeta(m(k_{p3}(x)), r_3) \subseteq \zeta(m(x), r) = x.$$

Тогда выполнены условия (29)—(33) *теоремы* 16 (*а значит, и условия теоремы* (13)). *Заметим, что ввиду* $\frac{1}{2}\| d(k_{p3}(x))\| \leqslant r' + r''$ *наше предположение сильнее, чем требования на* $k_{p1}(x)$ *или* $k_{p2}(x)$ *в теореме* 14.

Доказательство. Мы имеем

$$d(m(x) - \mathscr{Y}_p \cdot \mathscr{f}_p(m(x)) + \frac{1}{2}(\mathscr{Y}_p \cdot \mathscr{f}''_p(x))(x - m(x))(x - m(x))$$

$$= \frac{1}{4}|\mathscr{Y}_p \cdot \mathscr{f}''_p(x)|\, d(x)\, d(x).$$

Из того что $d(x) = 2r(1, 1, \ldots, 1)^T$, следует

$$r' = \frac{1}{2}\bar{k}r^2,$$

где использованы обозначения $\mathscr{C}_p := |\mathscr{Y}_p \cdot \mathscr{f}''_p(x)|$ и $\bar{k} = \max_{1 \leqslant i \leqslant n} \sum_{j,k} c_{ijk}$.

Далее мы имеем

$$d((\mathscr{I}_p - \mathscr{Y}_p \cdot \mathscr{f}'_p(m(x)))(x - m(x))) = |\mathscr{I}_p - \mathscr{Y}_p \cdot \mathscr{f}'_p(m(x))|\, d(x),$$

откуда следует

$$r'' = \delta r.$$

Условие $\zeta(m(k_{p3}(x)), r_3) \subseteq \zeta(m(x), r)$ выполнено тогда и только тогда, когда

$$\| m(k_{p3}(x)) - m(x)\| + r_3 \leqslant r.$$

Полагая $\bar{\eta} = \| \mathscr{Y}_p \cdot \mathscr{f}_p(m(x)) \|$, получаем, что оно эквивалентно условию

$$\bar{\eta} + r' + r'' \leqslant r$$

ввиду равенства

$$m(k_{p3}(x)) = m(x) - \mathscr{Y}_p \cdot \mathscr{f}_p(m(x)).$$

Иными словами, оно эквивалентно неравенству

$$\bar{\eta} + \frac{1}{2}\bar{k}r^2 \leqslant r(1-\delta).$$

Отсюда следует, что

$$h := \bar{k}\bar{\eta}/(1-\delta)^2 \leqslant \frac{1}{2}$$

и

$$\frac{1-\sqrt{1-2h}}{h}\,\frac{\bar{\eta}}{1-\delta} \leqslant r \leqslant \frac{1+\sqrt{1-2h}}{h}\,\frac{\bar{\eta}}{1-\delta}.$$

Таким образом, мы проверили выполнение условий (29), (30) и (32), (33) теоремы 16.

Остается доказать, что

$$\| \mathscr{Y}_p \cdot \mathscr{f}''_p(x_p) \| \leqslant \bar{k} \text{ для } x_p \in x.$$

Из $x_k \in x$ следует, что

$$| \mathscr{Y}_p \cdot \mathscr{f}''_p(x_p) | \leqslant | \mathscr{Y}_p \cdot \mathscr{f}''_p(x) |,$$

а потому

$$\| \mathscr{Y}_p \cdot \mathscr{f}''_p(x_p) \| \leqslant \| | \mathscr{Y}_p \cdot \mathscr{f}''_p(x) | \| = \| \mathscr{C}_p \| \leqslant \max_{1 \leqslant i \leqslant n} \sum_{j,k} c_{ijk}. \blacksquare$$

Как уже отмечено, из $k_{pi}(x) \subseteq x$ еще не следует, что выполнены условия теоремы 13. Поэтому рассмотрим следующий простой пример.

Пусть $\mathscr{f}_p(x_p) = \begin{pmatrix} x^2 - 1 \\ y^2 - 0.01 \end{pmatrix}$, $x_p = \begin{pmatrix} x \\ y \end{pmatrix}$ и $x = \begin{pmatrix} [0.98, & 1.18] \\ [0, & 0.2] \end{pmatrix}$.

Тогда $\mathscr{f}_p$ имеет в x корень $x_p^* = \begin{pmatrix} 1 \\ 0.1 \end{pmatrix}$. Ввиду $m(x) = \begin{pmatrix} 1.08 \\ 0.1 \end{pmatrix}$ мы имеем $\mathscr{f}_p(m(x)) = \begin{pmatrix} 0.1664 \\ 0 \end{pmatrix}$. Далее получаем $\mathscr{f}'_p(x) = \begin{pmatrix} 2x & 0 \\ 0 & 2y \end{pmatrix}$, $\mathscr{f}'_p(m(x)) = \begin{pmatrix} 2.16 & 0 \\ 0 & 0.2 \end{pmatrix}$, $\mathscr{f}'_p(m(x))^{-1} = \begin{pmatrix} 1/2.16 & 0 \\ 0 & 5 \end{pmatrix}$, $\mathscr{f}'_p(m(x))^{-1}\mathscr{f}_p(m(x)) = \begin{pmatrix} 0.1664/2.16 \\ 0 \end{pmatrix}$, т. е. $\eta = 0.1664/2.16$. Вто-

рая производная равна

$$f'_p(x_p) = \left(\begin{array}{cc|cc} 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 \end{array}\right),$$

откуда следует

$$f'_p(m(x))^{-1} f''_p(x_p) = \left(\begin{array}{cc|cc} 2/2.16 & 0 & 0 & 0 \\ 0 & 0 & 0 & 10 \end{array}\right)$$

для $x_p \in x$. Простое вычисление дает $\|f'_p(m(x))^{-1} f''_p(x_p)\|_\infty = 10 =: k$. Отсюда следует $h = k\eta = 1.664/2.16 > \frac{1}{2}$, т. е. теорема 13 не применима. С другой стороны, для величины $k_{p1}(x)$ из теоремы 12 мы, полагая $\mathcal{Y}_p = m(f'_p(x))^{-1}$, получаем, что

$$k_{p1}(x) = m(x) - \mathcal{Y}_p \cdot f_p(m(x,) - \mathcal{Y}_p \cdot (f''_p(x)(x - m(x))(x - m(x)))$$

$$\subseteq \begin{pmatrix} [0.9937, & 1.0123] \\ [0, & 0.2] \end{pmatrix} \subseteq x.$$

Это означает, что применима теорема 12 при $i = 1$ (равенство (22)). Аналогичное вычисление показывает, что в нашем примере выполнено $k_{p2}(x) \subseteq x$ и $k_p(x) \subseteq x$ (см. теорему 10) при $\mathcal{Y}_p = m(f'_p(x))^{-1}$.

Если сравнивать достоинства и недостатки теорем 12 и 13, то теорема 12 кажется предпочтительнее, так как она применима к произвольным интервальным векторам, а не только к шарам, определяемым ∞-нормой. С другой стороны, результат теоремы 13 справедлив для произвольных норм. Эти замечания показывают, что теоремы существования, сформулированные в этой главе, можно считать существенным дополнением к теореме Ньютона — Канторовича.

Это верно и для утверждений о существовании, содержащихся в теоремах 8—11. Их важное преимущество по сравнению с классическим утверждением из теоремы 13 состоит в том, что не нужны ни вторая производная, ни константа Липшица.

Замечания. *Рокн и Ланкастер также занимаются задачей о локализации* решений системы уравнений $f_p(x_p) = o_p$, похожей на задачу с одним уравнением от одной переменной. Для системы делается один шаг метода Ньютона и с помощью оценок для f_p, f'_p и f''_p находится локализация. В действительности это применение теоремы Канторовича. Процедура из [274] применяется затем в [271] для локализации собственных чисел вещественных матриц. Задача о собственных числах сводится известным способом к решению системы нелинейных уравнений, а затем применяется метод из [274]. Модификация интервального метода Ньютона (7) приведена в [160]. При этой модифи-

кации нет необходимости вычислять локализации для обращений всех матриц, содержащихся в данной интервальной матрице. Для итерационной процедуры

$$(34)\quad \begin{cases} x^{(k+1)} = \{m(x^{(k)}) - \mathcal{Y}_p \cdot f_p(m(x^{(k)})) + (\mathcal{I}_p - \mathcal{Y}_p \cdot f_p'(x^{(k)})) \\ \qquad \times (x^{(k)} - m(x^{(k)}))\} \cap x^{(k)}, \\ m(x^{(k)}) \in x^{(k)} \end{cases}$$

требуется только матрица $\mathcal{Y}_p$. Если выполнено неравенство

$$\| \mathcal{I}_p - \mathcal{Y}_p \cdot f_p'(x^{(k)}) \| < \frac{1}{2},$$

то доказывается сходимость локализаций к нулю. Если $\mathcal{Y}_p$ замена на $\mathcal{Y}_p^{(k)} = f_p'(m(x^{(k)}))^{-1}$, то доказывается суперлинейная сходимость. В [226] эта процедура используется для локализации собственных чисел вещественных матриц. Выражение в левой части пересечения, определяющего $x^{(k)}$ в формулах (34),— это $k_p(x)$ из теоремы 10.

В недавней работе [449] исследуется сходимость метода (34) (см. теорему 1 в [449]). Мы докажем здесь обобщение этих результатов. В действительности мы хотим показать, что если соотношение

$$k_p(x^{(0)}) = m(x^{(0)}) \mathcal{Y}_p \cdot f_p(m(x^{(0)})) + (\mathcal{I}_p - \mathcal{Y}_p \cdot f_p'(x^{(0)}))(x^{(0)} - m(x^{(0)})) \subseteq x^{(0)}$$

имеет место для некоторой матрицы $\mathcal{Y}_p \in M_{nn}(\mathbb{R})$ и интервального вектора $x^{(0)}$ с центром $m(x^{(0)})$, то итерационный метод (34), где $x^{(k)}$ — центр вектора $m(x^{(k)})$ при $k \geqslant 0$, сходится к единственному нулю y_p функции f_p в интервале $x^{(0)}$.

Отношение $k_p(x^{(0)}) \subseteq x^{(0)}$ определяется аналогично тому, как это сделано в теореме 14.6. Сформулированное выше условие $k_p(x^{(0)}) \subseteq x^{(0)}$ влечет за собой $d(x^{(0)}) > o_p$. Мы имеем $k_p(x^{(0)}) \subset x^{(0)}$ тогда и только тогда, когда расстояние от $k_p(x^{(0)})$ до $m(x^{(0)})$ удовлетворяет неравенству

$$q(k_p(x^{(0)}), \quad m(x^{(0)})) < \frac{1}{2} d(x^{(0)}),$$

которое эквивалентно неравенству

$$|k_p(x^{(0)}) - m(x^{(0)})| < \frac{1}{2} d(x^{(0)}).$$

Ввиду равенства

$$|k_p(x^{(0)}) - m(x^{(0)})| = |\mathcal{Y}_p \cdot f_p(m(x^{(0)}))| + \frac{1}{2} |\mathcal{I}_p - \mathcal{Y}_p \cdot f(x^{(0)})| d(x^{(0)})$$

это дает

$$|\mathscr{I}_p - \mathscr{Y}_p \cdot f'_p(x^{(0)})|\, d(x^{(0)}) < d(x^{(0)}).$$

Отсюда следует, что спектральный радиус неотрицательной матрицы $|\mathscr{I}_p - \mathscr{Y}_p \cdot f'_p(x^{(0)})|$ меньше 1. (Все скалярные произведения строк на положительный вектор $d(x^{(0)})$ меньше 1.) Предельное значение x^*, полученное согласно (34), удовлетворяет равенству

$$x^* = \{m(x^*) - \mathscr{Y}_p \cdot f_p(m(x^*)) + (\mathscr{I}_p - \mathscr{Y}_p \cdot f'_p(x^*))(x^* - m(x^*))\} \cap x^*,$$

т. е.

$$d(x^*) \leqslant |\mathscr{I}_p - \mathscr{Y}_p \cdot f'_p(x^*)|\, d(x^*).$$

Из монотонности включения следует, что $f'_p(x^*) \subseteq f'_p(x^{(0)})$, и поэтому мы имеем

$$|\mathscr{I}_p - \mathscr{Y}_p \cdot f'_p(x^*)| \leqslant |\mathscr{I}_p - \mathscr{Y}_p \cdot f'_p(x^{(0)})|.$$

Из теоремы Перрона и Фробениуса о монотонности спектрального радиуса неотрицательных матриц следует, что выполняется и неравенство $\rho(|\mathscr{I}_p - \mathscr{Y}_p \cdot f'_p(x^*)|) < 1$.

Теперь мы имеем $d(x^*) = o_p$ и $\lim_{k\to\infty} x^{(k)} = y_p$. Результат Мура [449], упомянутый выше, содержит частный случай интервалов, для которых $d(X_1^{(0)}) = \ldots = d(X_n^{(0)})$ («гиперкубов»).

В диссертации [501] имеется использующий $k_p(x)$ алгоритм нахождения локализующего вектора для решения системы линейных уравнений, а также программы на Фортране. Там есть также много численных примеров и других приложений к конкретным задачам вычислительной математики. См. также [421].

Теоремы существования (теорема 12) и теоремы сравнения (теоремы 14, 17 и следствие 15) были доказаны в [469a].

По аналогии с методом (34), основанным на теореме существования 10, мы можем, исходя из теоремы 11, определить следующий итерационный метод:

$$(35)\quad \begin{cases} \tilde{x}^{(k+1)} = m(x^{(k)}) - g_p(\mathscr{Y}_p, f_p(m(x^{(k)})) \\ \quad - (\mathscr{Y}_p - f'_p(x^{(k)}))(x^{(k)} - m(x^{(k)})), \\ \qquad \text{где } m(x^{(k)}) \in x^{(k)} \text{ — центр } x^{(k)}, \\ x^{(k+1)} = \tilde{x}^{(k+1)} \cap x^{(k)}, \quad k \geqslant 0. \end{cases}$$

Объем вычислений на каждом шаге этого метода значительно меньше, чем для метода (34). (Здесь через $g_p(\mathscr{A}, b)$ снова обозначен интервальный вектор, полученный применением метода Гаусса к интервальной матрице $\mathscr{A}$ и правой части b. См. гл. 15.)

Мы используем сокращение

$$\tilde{k}_{\mathrm{p}}(x) = m(x) - g_{\mathrm{p}}(\mathcal{Y}_{\mathrm{p}}, f_{\mathrm{p}}(m(x)) - (\mathcal{Y}_{\mathrm{p}} - f'_{\mathrm{p}}(x))(x - m(x))).$$

Аналогично утверждению, доказанному для метода (34), мы покажем сейчас следующее. Если для некоторой матрицы $\mathcal{Y}_{\mathrm{p}} \in M_{nn}(\mathbb{R})$ и интервального вектора $x^{(0)}$ с центром $m(x^{(0)})$ верно соотношение $\tilde{k}_{\mathrm{p}}(x^{(0)}) \subset x^{(0)}$, то итерационный метод (35), в котором $m(x^{(k)})$ для $k \geqslant 0$ означает центр $x^{(k)}$, сходится к единственному корню $\lim_{k\to\infty} x^{(k)} = x^*_{\mathrm{p}}$.

Чтобы доказать это утверждение, мы заметим, что по теореме 11 из $\tilde{k}_{\mathrm{p}}(x^{(0)}) \subset x^{(0)}$ следует существование корня x^*_{p} функции f_{p} в $x^{(0)}$. Используя приведенные в начале гл. 15 соотношения для метода Гаусса, мы можем теперь записать

$$\begin{aligned}
x^*_{\mathrm{p}} &= x^*_{\mathrm{p}} - \mathcal{Y}_{\mathrm{p}}^{-1} f_{\mathrm{p}}(x^*_{\mathrm{p}}) \\
&= m(x^{(0)}) - \mathcal{Y}_{\mathrm{p}}^{-1} f_{\mathrm{p}}(m(x^{(0)})) + x^*_{\mathrm{p}} - m(x^{(0)}) \\
&\quad + \mathcal{Y}_{\mathrm{p}}^{-1}(f_{\mathrm{p}}(x^{(0)})) - f_{\mathrm{p}}(x^*_{\mathrm{p}})) \\
&\in m(x^{(0)}) - \mathcal{Y}_{\mathrm{p}}^{-1} f_{\mathrm{p}}(m(x^{(0)})) \\
&\quad - (\mathcal{Y}_{\mathrm{p}} - f'_{\mathrm{p}}(x^{(0)})(x^{(0)} - m(x^{(0)}))) \\
&\subseteq m(x^{(0)}) - g_{\mathrm{p}}(\mathcal{Y}_{\mathrm{p}}, f_{\mathrm{p}}(m(x^{(0)})) \\
&\quad - (\mathcal{Y}_{\mathrm{p}} - f'_{\mathrm{p}}(x^{(0)})(x^{(0)} - m(x^{(0)}))) = \tilde{k}_{\mathrm{p}}(x^{(0)}).
\end{aligned}$$

Поэтому $x^{(1)}$ существует и имеет место $x^*_{\mathrm{p}} \in x^{(1)}$. С помощью математической индукции получаем, что $x^*_{\mathrm{p}} \in x^{(k)}$, $k \geqslant 0$.

Условие $\tilde{k}_{\mathrm{p}}(x^{(0)}) \subset x^{(0)}$ эквивалентно покомпонентному собственному включению. Иными словами, найдется вещественное число $0 \leqslant \alpha < 1$, для которого

$$d(\tilde{k}_{\mathrm{p}}(x^{(0)})) \leqslant \alpha d(x^{(0)}).$$

Из

$$\tilde{x}^{(1)} = \tilde{k}_{\mathrm{p}}(x^{(0)}) \subset x^{(0)}$$

следует, что

$$x^{(1)} = \tilde{k}_{\mathrm{p}}(x^{(0)}) \cap x^{(0)} = \tilde{k}_{\mathrm{p}}(x^{(0)}) = \tilde{x}^{(1)},$$

а потому

$$d(\tilde{x}^{(1)}) = d(x^{(1)}) \leqslant \alpha d(x^{(0)}).$$

Из $x^{(1)} \subset x^{(0)}$ и монотонности включения следует, что

$$|\mathcal{Y}_{\mathrm{p}} - f'_{\mathrm{p}}(x^{(1)})| \leqslant |\mathcal{Y}_{\mathrm{p}} - f'_{\mathrm{p}}(x^{(0)})|,$$

а потому

$$d\left(f_{\mathrm{p}}\left(m\left(x^{(1)}\right)\right)-\left(\mathcal{Y}_{\mathrm{p}}-f_{\mathrm{p}}^{1}\left(x^{(1)}\right)\right)\left(x^{(1)}-m\left(x^{(1)}\right)\right)\right)$$
$$=\left|\mathcal{Y}_{\mathrm{p}}-f_{\mathrm{p}}^{\prime}\left(x^{(1)}\right)\right| d\left(x^{(1)}\right)$$
$$\leqslant \alpha\left|\mathcal{Y}_{\mathrm{p}}-f_{\mathrm{p}}^{\prime}\left(x^{(0)}\right)\right| d\left(x^{(0)}\right)$$
$$=\alpha d\left(f_{\mathrm{p}}\left(m\left(x^{(0)}\right)\right)-\left(\mathcal{Y}_{\mathrm{p}}-f_{\mathrm{p}}^{\prime}\left(x^{(0)}\right)\right)\left(x^{(0)}-m\left(x^{(0)}\right)\right)\right).$$

В гл. 15 доказано, что отсюда следует

$$d\left(\tilde{x}^{(2)}\right) \leqslant \alpha d\left(\tilde{x}^{(1)}\right),$$

т. е.

$$d\left(\tilde{x}^{(2)}\right) \leqslant \alpha^{2} d\left(x^{(0)}\right).$$

Из

$$d\left(x^{(2)}\right) \leqslant \inf\left\{d\left(\tilde{x}^{(2)}\right),\ d\left(x^{(1)}\right)\right\}$$

следует, что

$$d\left(x^{(2)}\right) \leqslant \alpha^{2} d\left(x^{(0)}\right).$$

Методом математической индукции мы доказываем, что

$$d\left(x^{(k+1)}\right) \leqslant \alpha^{k+1} d\left(x^{(0)}\right), \quad k \geqslant 0.$$

Отсюда следует, что $\lim_{k\to\infty} d\left(x^{(k)}\right)=o_{\mathrm{p}}$, а ввиду $x_{\mathrm{p}}^{*} \in x^{(k)}$ это дает $\lim_{k\to\infty} x^{(k)}=x_{\mathrm{p}}^{*}$, что и завершает доказательство.

Теперь мы хотим сделать еще несколько замечаний о методе (35). (Те же замечания справедливы и для метода (34).) Если применять метод (35), начиная с произвольной невырожденной матрицы $\mathcal{Y}_{\mathrm{p}}$ и произвольного интервального вектора $x^{(0)}$, не обязательно удовлетворяющего условию $\tilde{k}_{\mathrm{p}}\left(x^{(0)}\right) \subset x^{(0)}$, то могут представиться следующие два случая.

(i) Для некоторого $k_0 \geqslant 0$ верно $\tilde{x}^{(k_0+1)} \cap x^{(k_0)}=\phi$. Тогда f_{p} не имеет нулей в $x^{(0)}$. Действительно, если бы f_{p} имела нуль в $x^{(0)}$, то, как и в случае $\tilde{k}_{\mathrm{p}}\left(x^{(0)}\right) \subset x^{(0)}$, мы могли бы показать, что пересечение непусто для всех k.

(ii) Метод (35) определен для всех $k \geqslant 0$. Тогда верно равенство $\lim_{k\to\infty} x^{(k)}=x^{*}$. Если теперь $d\left(x^{*}\right) \neq o_{\mathrm{p}}$, то невозможно сказать что-либо о наличии или отсутствии нуля. Если же $d\left(x^{*}\right)=o_{\mathrm{p}}$, т. е. $\lim_{k\to\infty} x^{(k)}=x_{\mathrm{p}}^{*}$, то x_{p}^{*} — нуль функции f_{p}. Это получается из (35) переходом к пределу при $k \to \infty$.

Замечательно, что условие $d\left(\tilde{k}_{\mathrm{p}}\left(x^{(0)}\right)\right) \leqslant \alpha d\left(x^{(0)}\right), \quad 0 \leqslant \alpha<1$, оказывается достаточным для того, чтобы имело место (i) или (ii) с $d\left(x^{*}\right)=o_{\mathrm{p}}$. Если это условие выполнено, то те же рассуждения, что и в случае более сильного условия $\tilde{k}_{\mathrm{p}}\left(x^{(0)}\right) \subset x^{(0)}$, показывают, что для всех $k \geqslant 0$, для которых $\tilde{x}^{(k+1)} \cap x^{(k)} \neq \phi$,

верно неравенство $d(x^{(k+1)}) \leqslant \alpha^{k+1} d(x^{(0)})$. Если $\mathcal{f}_p$ не имеет корней в $x^{(0)}$, то пересечение станет пустым для некоторого $k_0 \geqslant 0$, так как иначе равенство $\lim_{k\to\infty} x^{(k)} = x_p^*$ противоречило бы равенству $\mathcal{f}_p(x_p^*) = o_p$. Если же $\mathcal{f}_p$ имеет корень x^* в $x^{(0)}$, то так же, как при условии $\tilde{k}_p(x^{(0)}) \subset x^{(0)}$, мы покажем, что пересечение непусто при любом $k \geqslant 0$, а потому верно $\lim_{k\to\infty} x^{(k)} = x_p^*$ и $\mathcal{f}_p(x_p^*) = o_p$.

Локализация решений системы линейных уравнений, например системы вида (6), используется в методах (7) и (12). Такую локализацию можно в принципе найти с помощью метода Гаусса из гл. 15. Такой метод наверняка будет сходиться, если ширина вектора $x^{(0)}$ достаточно мала. До сих пор не найдено проверяемых условий, гарантирующих сходимость. Было бы чрезвычайно полезно иметь такие критерии, так как это позволило бы избежать обращения матриц $m(\mathcal{F}^{(k)})$ и применения итерации (10). Тогда можно было бы проводить итерационную локализацию решения y_p с разумными вычислительными затратами. Хотя в (14) приходится применять метод Гаусса и делать шаг (d), эта процедура все же намного дешевле в смысле объема вычислений, чем другие методы, рассмотренные в этой главе. Важно и то, что в теореме 7 нет ограничений на величину или ширину вектора $x^{(0)}$. Поэтому утверждение о сходимости в теореме 7 дает практически глобальную сходимость. Численный пример применения метода (14) показывает важность этого утверждения. Хорошо известно, что решение дискретизированной граничной задачи методом Ньютона не представляет проблем, когда функция $f(t, y)$ выпукла по y. Однако в невыпуклом случае сходимость удается доказать лишь для начальных значений, достаточно близких к решению. Для метода (14) не нужно никаких условий выпуклости.

В гл. 21 рассматривается метод [см. (21.8)], который также решает упомянутую дискретизированную граничную задачу, но не использует метода Гаусса. Однако там для достижения квадратичной сходимости требуется на каждом шаге не только интервальное вычисление производной, но и два вычисления значений самой функции.

Применение метода Гаусса для локализации множества решений системы (4) или (6) приводит к алгоритму

$$(36)\qquad x^{(k+1)} = \{m(x^{(k)}) - g_p(\mathcal{f}_p'(x_p^{(k)}),\ \mathcal{f}_p(m(x^{(k)})))\} \cap x^{(k)}.$$

Здесь $g_p(\cdot, \cdot)$ означает отображение, определяемое методом Гаусса. Для одного уравнения от одной переменной этот метод совпадает с методом (7.11). К сожалению, теорему 7.4 не удается обобщить на системы уравнений. Это было показано

в [514a], где рассмотрен пример, для которого $x_{\mathrm{p}}^{*} \in x^{(0)}$, $f_{\mathrm{p}}(x_{\mathrm{p}}^{*}) = o_{\mathrm{p}}$ и $x^{(1)} = x^{(0)}$ при $d(x^{(1)}) \neq o_{\mathrm{p}}$. В [349d] входящая в (35) матрица $\mathcal{Y}_{\mathrm{p}}$ может меняться вместе с k. Там описана квадратично сходящаяся модификация метода (35), в которой полагают $\mathcal{Y}_{\mathrm{p}}^{(k)} = m(f_{\mathrm{p}}'(x^{(k)}))$. В [349b] можно найти достаточные условия сходимости метода (36) к некоторому решению уравнения $f_{\mathrm{p}}(x_{\mathrm{p}}) = o_{\mathrm{p}}$.

* * *

Литература: [13], [22], [74], [100], [109], [125], [160], [215], [226], [230], [240], [271], [274], [325], [349], [349b], [349d], [415], [421], [448], [449], [451], [452], [457], [469a], [472], [501], [514a].

Глава 20

МЕТОДЫ НЬЮТОНОВСКОГО ТИПА, НЕ ИСПОЛЬЗУЮЩИЕ ОБРАЩЕНИЯ МАТРИЦ

В предыдущей главе было показано, что итерация (19.12) сходится по ширине последовательных приближений не хуже, чем квадратично. На каждом шаге приходится обращать вещественную матрицу $m(\mathcal{F}^{(k)})$ и производить вычисления по формулам (19.10), пока там не будет достигнута неподвижная точка. Теперь мы изменим (19.12) так, что будет выполняться всего один шаг итерационного нахождения локализации $\mathcal{Y}^{(k)}$ для множества $\{\mathcal{F}_{\mathrm{p}}^{-1} \mid \mathcal{F}_{\mathrm{p}} \in \mathcal{F}^{(k)}\}$. Напомним, что в (19.12) приходится проводить итерацию (19.10). Мы выберем новый метод итерационной локализации обращения интервальной матрицы. Он будет обобщением на интервальные матрицы алгоритма (18.9) при $r = 2$.

Теорема 1. *Пусть даны вещественная невырожденная матрица $\mathcal{A}_{\mathrm{p}}$ и последовательность интервальных матриц $\{\mathcal{A}^{(k)}\}_{k=0}^{\infty}$, для которой $\mathcal{A}_{\mathrm{p}} \in \mathcal{A}^{(k)}$, $k \geqslant 0$. Пусть $\lim_{k\to\infty} \mathcal{A}^{(k)} = \mathcal{A}_{\mathrm{p}}$ и $\mathcal{W}^{(0)}$ — интервальная матрица, для которой $\mathcal{A}_{\mathrm{p}}^{-1} \in \mathcal{W}^{(0)}$. Для итерационного метода*

$$\text{(1)} \qquad \mathcal{W}^{(k+1)} = \{m(\mathcal{W}^{(k)}) + \mathcal{W}^{(k)}(\mathcal{I}_{\mathrm{p}} - \mathcal{A}^{(k)} m(\mathcal{W}^{(k)}))\} \cap \mathcal{W}^{(k)},$$

где $m(\mathcal{W}^{(k)})$ для $k \geqslant 0$ определено в (18.1), верно следующее:

$$\text{(2a)} \qquad \mathcal{A}_{\mathrm{p}}^{-1} \in \mathcal{W}^{(k)}, \quad k \geqslant 0.$$

Если любая матрица $\mathscr{W}_\mathrm{p} \in \mathscr{W}^{(0)}$ *невырожденная, то*

$$\lim_{k\to\infty} \mathscr{W}^{(k)} = \mathscr{A}_\mathrm{p}^{-1}. \tag{2b}$$

Доказательство. (2a): По условию имеем $\mathscr{A}_\mathrm{p}^{-1} \in \mathscr{W}^{(0)}$ и $\mathscr{A}_\mathrm{p} \in \mathscr{A}^{(0)}$. Отсюда и из (10.10′) следует, что

$$\begin{aligned}\mathscr{A}_\mathrm{p}^{-1} &= m(\mathscr{W}^{(0)}) + \mathscr{A}_\mathrm{p}^{-1}(\mathscr{I}_\mathrm{p} - \mathscr{A}_\mathrm{p} \cdot m(\mathscr{W}^{(0)})) \\ &\in \{m(\mathscr{W}^{(0)}) + \mathscr{W}^{(0)}(\mathscr{I}_\mathrm{p} - \mathscr{A}^{(0)} m(\mathscr{W}^{(0)}))\} \cap \mathscr{W}^{(0)}.\end{aligned}$$

Доказательство завершается методом математической индукции.

(2b): Монотонно убывающая последовательность матриц

$$\mathscr{W}^{(0)} \supseteq \mathscr{W}^{(1)} \supseteq \mathscr{W}^{(2)} \supseteq \ldots$$

сходится к некоторой интервальной матрице $\mathscr{W}$ (ср. следствие 10.8). В силу непрерывной зависимости центра $m(\mathscr{W}^{(k)})$ от $\mathscr{W}^{(k)}$ имеем

$$\lim_{k\to\infty} m(\mathscr{W}^{(k)}) = m(\mathscr{W}).$$

Отсюда и из непрерывности операций, входящих в (1), следует, что

$$\begin{aligned}\mathscr{W} &= \lim_{k\to\infty} (\{m(\mathscr{W}^{(k)}) + \mathscr{W}^{(k)}(\mathscr{I}_\mathrm{p} - \mathscr{A}^{(k)} m(\mathscr{W}^{(k)}))\} \cap \mathscr{W}^{(k)}) \\ &= \{m(\mathscr{W}) + \mathscr{W}(\mathscr{I}_\mathrm{p} - \mathscr{A}_\mathrm{p} \cdot m(\mathscr{W}))\} \cap \mathscr{W},\end{aligned}$$

и потому

$$m(\mathscr{W}) \in m(\mathscr{W}) + \mathscr{W}(\mathscr{I}_\mathrm{p} - \mathscr{A}_\mathrm{p} \cdot m(\mathscr{W})).$$

Обозначим j-й столбец матрицы $m(\mathscr{W})$ через $(m(\mathscr{W}))_j$. В силу (10.1) мы получаем представление

$$(m(\mathscr{W}))_j = (m(\mathscr{W}))_j + \mathscr{W}_\mathrm{p}^{(j)}((\mathscr{I}_\mathrm{p} - \mathscr{A}_\mathrm{p} \cdot m(\mathscr{W})))_j$$

с матрицей $\mathscr{W}_\mathrm{p}^{(j)} \in \mathscr{W} \subseteq \mathscr{W}^{(0)}$. Отсюда следует, что

$$\mathscr{W}_\mathrm{p}^{(j)}((\mathscr{I}_\mathrm{p} - \mathscr{A}_\mathrm{p} \cdot m(\mathscr{W})))_j = o_\mathrm{p}.$$

Матрицы $\mathscr{A}_\mathrm{p}$ и $\mathscr{W}_\mathrm{p}^{(j)}$ невырожденные по условию, поэтому, используя

$$(\mathscr{A}_\mathrm{p} m(\mathscr{W}))_j = \mathscr{A}_\mathrm{p}(m(\mathscr{W}))_j,$$

мы получаем

$$(m(\mathscr{W}))_j = (\mathscr{A}_\mathrm{p}^{-1})_j,$$

что доказывает равенство $m(\mathscr{W}) = \mathscr{A}_\mathrm{p}^{-1}$. Но отсюда следует равенство $\mathscr{W} = \mathscr{A}_\mathrm{p}^{-1}$, что и завершает доказательство теоремы. ■

Итерационный метод (18.9) при $r = 2$ получается теперь как частный случай при $\mathscr{A}^{(k)} = \mathscr{A}_\text{р}$. Если мы предположим, что последовательность $\{\mathscr{A}^{(k)}\}_{k=0}^{\infty}$ в (1) удовлетворяет условию

$$\mathscr{A}^{(0)} \supseteq \mathscr{A}^{(1)} \supseteq \mathscr{A}^{(2)} \supseteq \ldots, \text{ где } \lim_{k\to\infty} \mathscr{A}^{(k)} = \mathscr{A},$$

то аналогично (2а) получим, что из включения

$$\{\mathscr{A}_\text{р}^{-1} \mid \mathscr{A}_\text{р} \in \mathscr{A}\} \subseteq \mathscr{W}^{(0)}$$

всегда следует

$$\{\mathscr{A}_\text{р}^{-1} \mid \mathscr{A}_\text{р} \in \mathscr{A}\} \subseteq \mathscr{W}^{(k)}, \quad k \geqslant 0.$$

Если теперь выбрать $\mathscr{A}^{(k)} = \mathscr{A}$, то мы получим из (1) метод итерационной локализации обратной матрицы. Аналогичный итерационный метод был рассмотрен в (19.10).

Теперь используем (1), чтобы построить метод, аналогичный методу (19.12). Пусть $x^{(0)}$ — интервальный вектор, содержащий нуль функции $\mathscr{f}_\text{р}(x_\text{р})$. Допустим еще, что $\mathscr{Y}^{(0)}$ — интервальная матрица, такая что $\mathscr{J}_\text{р}(x_\text{р})^{-1} \in \mathscr{Y}^{(0)}$ для $x_\text{р} \in x^{(0)}$, где $\mathscr{J}_\text{р}(x_\text{р})$ определена формулами (19.3). Положим, как и раньше, $\mathscr{F}^{(k)} = \mathscr{f}'_\text{р}(x^{(k)})$.

Рассмотрим теперь итерацию

$$(3)\quad \begin{cases} x^{(k+1)} = \{m(x^{(k)}) - \mathscr{Y}^{(k)}\mathscr{f}_\text{р}(m(x^{(k)}))\} \cap x^{(k)}, \\ \mathscr{Y}^{(k+1)} = \{m(\mathscr{Y}^{(k)}) + \mathscr{Y}^{(k)}(\mathscr{J}_\text{р} - \mathscr{F}^{(k+1)}m(\mathscr{Y}^{(k)}))\} \cap \mathscr{Y}^{(k)}, \quad k \geqslant 0. \end{cases}$$

Метод (3) заключается в одновременном исполнении (19.7) и (1). Его главное отличие от итерации (19.12) состоит в вычислении новой интервальной матрицы $\mathscr{Y}^{(k+1)}$. Здесь исполняется только один шаг подходящего метода.

В методе (3) вычисляются последовательности интервальных матриц и интервальных векторов

$$x^{(0)} \supseteq x^{(1)} \supseteq x^{(2)} \supseteq \ldots \quad \text{и} \quad \mathscr{Y}^{(0)} \supseteq \mathscr{Y}^{(1)} \supseteq \mathscr{Y}^{(2)} \supseteq \ldots$$

Сейчас мы докажем некоторые свойства этих последовательностей.

Теорема 2. *Пусть $x^{(0)}$ — интервальный вектор, а $y_\text{р} \in x^{(0)}$ — нуль функции $\mathscr{f}_\text{р}(x_\text{р})$. Пусть производная функции $\mathscr{f}_\text{р}(x_\text{р})$ удовлетворяет условию Липшица при значении аргумента $x^{(0)}$, а $\mathscr{Y}^{(0)}$ — интервальная матрица, содержащая матрицы $\mathscr{J}_\text{р}(x)^{-1}$ для всех $x_\text{р} \in x^{(0)}$. Пусть, наконец, условие*

$$\| d(\mathscr{f}'_\text{р}(x)) \|' \leqslant c \| d(x) \|'$$

выполнено при $x \subseteq x^{(0)}$ для вычисления в интервальной арифметике $\mathscr{f}'_\text{р}(x)$ производной Фреше $\mathscr{f}'_\text{р}(x_\text{р})$. (Выполнение этого

условия обеспечено, если любой элемент матрицы удовлетворяет условию (3.5′).)

Тогда интервальные векторы $\{x^{(k)}\}_{k=0}^{\infty}$ *и интервальные матрицы* $\{\mathscr{Y}^{(k)}\}_{k=0}^{\infty}$, *вычисленные по формулам* (3), *удовлетворяют следующим условиям.*

(4) *Каждый интервальный вектор* $x^{(k)}$, $k \geqslant 0$, *содержит корень* y_p.

(5) *Если все матрицы* $\mathscr{Y}_p \in \mathscr{Y}^{(0)}$ *невырождены, то*

$$\lim_{k\to\infty} x^{(k)} = y_p \quad \text{и} \quad \lim_{k\to\infty} \mathscr{Y}^{(k)} = f'_p(y_p)^{-1}.$$

Пусть $\mathscr{H}^{(k)} \in M_{n,\,n+1}(I(\mathbb{R}))$, $k \geqslant 0$, — *интервальная матрица, первый столбец которой совпадает с интервальным вектором* $x^{(k)}$, *а остальные являются столбцами матрицы* $\mathscr{Y}^{(k)}$, *т. е.* $\mathscr{H}^{(k)} = (x^{(k)}, \mathscr{Y}^{(k)})$. *Тогда*

(6) $$O_R\big((3), \big(y_p, f'_p(y_p)^{-1}\big)\big) \geqslant 2$$

(*см. приложение* A, *теорема* 2).

Доказательство. Снова положим $\mathscr{F}^{(k)} = f'_p(x^{(k)})$.

(4): Точно так же, как в доказательстве соответствующего утверждения (19.8) теоремы 19.1, мы показываем, что $y_p \in x^{(1)}$. В силу $x^{(1)} \subseteq x^{(0)}$ отсюда следует, что для $x_p \in x^{(1)}$ верно

$$\mathscr{J}_p(x_p)^{-1} \in \mathscr{Y}^{(0)} \quad \text{и} \quad \mathscr{J}_p(x_p) \in \mathscr{F}^{(1)}.$$

Поэтому с помощью (10.10) получаем

$$\begin{aligned}\mathscr{J}_p(x_p)^{-1} &= m(\mathscr{Y}^{(0)}) + \mathscr{J}_p(x_p)^{-1}(\mathscr{J}_p - \mathscr{J}_p(x_p)m(\mathscr{Y}^{(0)})) \\ &\in \{m(\mathscr{Y}^{(0)} + \mathscr{Y}^{(0)}(\mathscr{J}_p - \mathscr{F}^{(1)}m(\mathscr{Y}^{(0)}))\} \cap \mathscr{Y}^{(0)} = \mathscr{Y}^{(1)}.\end{aligned}$$

Таким образом, мы имеем

$$\{\mathscr{J}_p(x_p)^{-1} \mid x_p \in x^{(1)}\} \subseteq \mathscr{Y}^{(1)}.$$

Используя этот результат, мы показываем, что $y_p \in x^{(2)}$, и (4) получается методом математической индукции.

(5): Ввиду $\mathscr{Y}^{(k)} \subseteq \mathscr{Y}^{(0)}$ мы можем доказать равенство $\lim x^{(k)} = y_p$ тем же методом, что и (19.9).

Второе утверждение о сходимости следует из теоремы 1, так как $f_p(y_p) \in \mathscr{F}^{(k)}$, $k \geqslant 0$.

(6): Используем сокращение $\hat{\mathscr{J}}_{\mathrm{p}} := \mathscr{f}'_{\mathrm{p}}(y_{\mathrm{p}})$. Из (3) с помощью (10.6), (10.7), (10.9) следует, что

$$x^{(k+1)} - y_{\mathrm{p}} \subseteq -\mathscr{V}^{(k)}\{\mathscr{f}_{\mathrm{p}}(m(x^{(k)})) - \mathscr{f}_{\mathrm{p}}(y_{\mathrm{p}}) - \mathscr{f}'_{\mathrm{p}}(y_{\mathrm{p}})(m(x^{(k)}) - y_{\mathrm{p}})\}$$
$$- \mathscr{V}^{(k)}(\hat{\mathscr{J}}_{\mathrm{p}} \cdot (m(x^{(k)}) - y_{\mathrm{p}})) + m(x^{(k)}) - y_{\mathrm{p}}$$
$$= -\mathscr{V}^{(k)}\{\mathscr{f}_{\mathrm{p}}(m(x^{(k)}) - \mathscr{f}_{\mathrm{p}}(y_{\mathrm{p}})) - \mathscr{f}'_{\mathrm{p}}(y_{\mathrm{p}})(m(x^{(k)}) - y_{\mathrm{p}})\}$$
$$+ (\hat{\mathscr{J}}_{\mathrm{p}}^{-1} - \mathscr{V}^{(k)}) \cdot (\hat{\mathscr{J}}_{\mathrm{p}} \cdot (m(x^{(k)}) - y_{\mathrm{p}})).$$

Теперь возьмем абсолютные величины (определение 10.6) и применим (10.15), (10.27), теорему 3.2.12 из [243] и соотношения

$$q(x^{(k)}, y_{\mathrm{p}}) = q(x^{(k)} - y_{\mathrm{p}}, o_{\mathrm{p}}) = |x^{(k)} - y_{\mathrm{p}}|,$$
$$q(\mathscr{V}^{(k)}, \mathscr{J}_{\mathrm{p}}^{-1}) = q(\mathscr{V}^{(k)} - \hat{\mathscr{J}}_{\mathrm{p}}^{-1}, O_{\mathrm{p}}) = |\mathscr{V}^{(k)} - \hat{\mathscr{J}}_{\mathrm{p}}^{-1}|.$$

Теперь, используя монотонную векторную норму, подчиненную ей монотонную матричную норму, теорему об эквивалентности норм и соотношение $m(x^{(k)}) \in x^{(k)}$, получим

$$\|q(x^{(k+1)}, y_{\mathrm{p}})\| \leqslant \||\mathscr{V}^{(k)}|\| c_1 \|q(x^{(k)}, y_{\mathrm{p}})\|^2$$
$$+ \|q(\mathscr{V}^{(k)}, \hat{\mathscr{J}}_{\mathrm{p}}^{-1})\| \cdot \||\hat{\mathscr{J}}_{\mathrm{p}}|\| \cdot \|q(x^{(k)}, y_{\mathrm{p}})\|.$$

Из второго равенства в (3) аналогично получим, что

$$\mathscr{V}^{(k+1)} - \hat{\mathscr{J}}_{\mathrm{p}}^{-1} \subseteq -((\hat{\mathscr{J}}_{\mathrm{p}}^{-1} - \mathscr{V}^{(k)}) \cdot \hat{\mathscr{J}}_{\mathrm{p}}) \cdot (\hat{\mathscr{J}}_{\mathrm{p}}^{-1} - m(\mathscr{V}^{(k)}))$$
$$- \mathscr{V}^{(k)}((\mathscr{F}^{(k+1)} - \hat{\mathscr{J}}_{\mathrm{p}}) m(\mathscr{V}^{(k)})).$$

Снова возьмем абсолютные величины с учетом соотношений $m(\mathscr{V}^{(k)}) \in \mathscr{V}^{(k)}$, (10.15), (10.27) и

$$|\mathscr{F}^{(k+1)} - \hat{\mathscr{J}}_{\mathrm{p}}| = q(\mathscr{F}^{(k+1)}, \hat{\mathscr{J}}_{\mathrm{p}}) \leqslant d(\mathscr{F}^{(k+1)}),$$
$$\|d(\mathscr{F}^{(k+1)})\|' \leqslant c\|d(x^{(k+1)})\|'.$$

Используя монотонную матричную норму и теорему об эквивалентности норм, получим, что

$$\|q(\mathscr{V}^{(k+1)}, \hat{\mathscr{J}}_{\mathrm{p}}^{-1})\| \leqslant c_3 \||\hat{\mathscr{J}}_{\mathrm{p}}|\| \cdot \|q(\mathscr{V}^{(k)}, \hat{\mathscr{J}}_{\mathrm{p}}^{-1})\|^2$$
$$+ \||\mathscr{V}^{(k)}|\|^2 c_4 \|q(x^{(k+1)}, y_{\mathrm{p}})\|.$$

Для $\mathscr{A}_{\mathrm{p}} = (\mathscr{b}_{\mathrm{p}}, \mathscr{B}_{\mathrm{p}}) \in M_{n, n+1}(\mathbb{R})$ введем норму $\|\mathscr{A}_{\mathrm{p}}\| = \max\{\|\mathscr{b}_{\mathrm{p}}\|, \|\mathscr{B}_{\mathrm{p}}\|\}$. Полагая $r^{(k)} = \|(q(x^{(k)}, y_{\mathrm{p}}), q(\mathscr{V}^{(k)}, \hat{\mathscr{J}}_{\mathrm{p}}^{-1}))\|$, замечая, что $\lim\limits_{k \to \infty} \mathscr{V}^{(k)} = \hat{\mathscr{J}}_{\mathrm{p}}^{-1}$, $\||\mathscr{V}^{(k)}|\| \leqslant s$ и полагая $\||\hat{\mathscr{J}}_{\mathrm{p}}|\| = p$, получаем из установленных выше неравенств, что

$$\|q(x^{(k+1)}, y_{\mathrm{p}})\| \leqslant (sc_1 + pc_2)(r^{(k)})^2,$$
$$\|q(\mathscr{V}^{(k+1)}, \hat{\mathscr{J}}_{\mathrm{p}}^{-1})\| \leqslant (c_3 p + s^2 c_4 (sc_1 + pc_2))(r^{(k)})^2.$$

Если мы теперь положим

$$\gamma = \max\{sc_1 + pc_2,\ c_3p + s^2c_4(sc_1 + pc_2)\}.$$

то получим, что

$$r^{(k+1)} \leqslant \gamma (r^{(k)})^2, \qquad k \geqslant 0.$$

Теперь из теоремы 2 приложения А следует, что мы имеем не менее чем квадратичную сходимость последовательности

$$\{(x^{(k)}, \mathscr{Y}^{(k)})\}_{k=0}^{\infty} \quad \text{к} \quad \left(y_{\text{р}}, \mathscr{f}'_{\text{р}}(y_{\text{р}})^{-1}\right). \blacksquare$$

Замечания. Теорема 2 имеется в [23]. Она улучшает результат из [21], где была установлена лишь суперлинейная сходимость. Модификация метода (3) описана в [194], где ширина приближений мажорируется квадратично сходящейся последовательностью. В этом случае на каждом шаге приходится дополнительно решать точечное уравнение. Теорема 2 показывает, что эти дополнительные усилия не оказывают никакого влияния на порядок сходимости.

Метод (1) без взятия пересечений рассмотрен в [344] для одного специального класса интервальных матриц.

* * *

Литература: [21], [23], [64], [170], [194], [243], [286], [323].

Глава 21
МЕТОДЫ НЬЮТОНОВСКОГО ТИПА ДЛЯ ЧАСТНЫХ ТИПОВ СИСТЕМ НЕЛИНЕЙНЫХ УРАВНЕНИЙ

Рассмотрим теперь ту же постановку задачи, что в двух предшествующих главах, введя дополнительные предположения о рассматриваемых системах

$$\mathscr{f}_{\text{р}}(x_{\text{р}}) = o_{\text{р}}.$$

Утверждения, доказываемые в этой главе, весьма близки к результатам из [241, 243]. Важно, однако, отметить, что не делается никаких предположений о выпуклости $\mathscr{f}_{\text{р}}$. Производная Фреше функции $\mathscr{f}_{\text{р}}$ в точке $x_{\text{р}}$ обозначается через $\mathscr{f}'_{\text{р}}(x_{\text{р}}) = (\partial f_i/\partial x_j)$. Используя естественный частичный порядок на множестве $V_n(\mathbb{R})$ и на множестве $M_{nn}(\mathbb{R})$, состоящем из всех вещественных точечных матриц размерности $n \times n$, мы сначала

докажем общую теорему, а затем покажем, как ее можно применить к конкретным системам.

Теорема 1. *Пусть отображение* f_p: $\mathfrak{D} \subseteq V_n(\mathbb{R}) \to V_n(\mathbb{R})$ *имеет производную Фреше, а множество* $\{x_p \mid x_p^{(0)} \leqslant x_p \leqslant y_p^{(0)}\}$ *содержится в* $\mathfrak{D}$. *Пусть выполнено*

(0) $$f_p(x_p^{(0)}) \leqslant o_p \leqslant f_p(y_p^{(0)}).$$

Задано отображение

$$\mathscr{A}_p\colon \{x_p \mid x_p^{(0)} \leqslant x_p \leqslant y_p^{(0)}\} \times \{x_p \mid x_p^{(0)} \leqslant x_p \leqslant y_p^{(0)}\} \to M_{nn}(R),$$
$$\mathscr{A}_p = \mathscr{A}_p(x_p, y_p),$$

для которого верно

(1) $f_p(x_p) - f_p(y_p) = \mathscr{A}_p(x_p, y_p)(x_p - y_p)$ для $x_p^{(0)} \leqslant x_p \leqslant y_p \leqslant y_p^{(0)}$.

Задано непрерывное отображение

$$\mathscr{B}_p\colon \{x_p \mid x_p^{(0)} \leqslant x_p \leqslant y_p^{(0)}\} \times \{x_p \mid x_p^{(0)} \leqslant x_p \leqslant y_p^{(0)}\} \to M_{nn}(\mathbb{R}), \quad \mathscr{B}_p = \mathscr{B}_p(x_p, y_p),$$

для которого верно

(2) $$\mathscr{B}_p(\bar{x}_p, \bar{y}_p) \leqslant \mathscr{B}_p(x_p, y_p) \quad \text{для}$$
$$x_p^{(0)} \leqslant x_p \leqslant \bar{x}_p \leqslant \bar{y}_p \leqslant y_p \leqslant y_p^{(0)},$$

(3) $\mathscr{A}_p(x_p, y_p) \leqslant \mathscr{B}_p(x_p, y_p)$ *для* $x_p^{(0)} \leqslant x_p \leqslant y_p \leqslant y^{(0)}$,

(4) *существует* $\mathscr{B}_p(x_p, y_p)^{-1}$ *и* $\mathscr{B}_p(x_p, y_p)^{-1} \geqslant O_p$ *для* $x_p^{(0)} \leqslant x_p \leqslant y_p \leqslant y_p^{(0)}$.

Матрица $\mathscr{P}_p^{(0)} \in M_{nn}(\mathbb{R})$ *невырождена, и имеет место*

(5) $\mathscr{B}_p(x_p^{(0)}, y_p^{(0)})\mathscr{P}_p^{(0)} \leqslant \mathscr{I}_p$ ($\mathscr{I}_p$ — *единичная матрица*),

(6) $$\mathscr{P}_p^{(0)}\mathscr{B}_p(x_p^{(0)}, y_p^{(0)}) \leqslant \mathscr{I}_p,$$

(7) $$\mathscr{P}_p^{(0)} \geqslant O_p.$$

Тогда итерация

(8)
$$\begin{cases} y_p^{(k+1)} = y_p^{(k)} - \mathscr{P}_p^{(k)} f_p(y_p^{(k)}), \\ x_p^{(k+1)} = x_p^{(k)} - \mathscr{P}_p^{(k)} f_p(x_p^{(k)}), \\ \mathscr{P}_p^{(k+1)} = \mathscr{P}_p^{(k)} - \mathscr{P}_p^{(k)}(\mathscr{B}_p(x_p^{(k+1)}, y_p^{(k+1)})\mathscr{P}_p^{(k)} - \mathscr{I}_p), \quad k \geqslant 0, \end{cases}$$

может быть выполнена без ограничений и имеет место

$$(9)\quad x_{\mathrm{p}}^{(0)} \leqslant x_{\mathrm{p}}^{(1)} \leqslant \ldots \leqslant x_{\mathrm{p}}^{(k)} \leqslant x_{\mathrm{p}}^{(k+1)} \leqslant y_{\mathrm{p}}^{(k+1)} \leqslant y_{\mathrm{p}}^{(k)} \leqslant \ldots \leqslant y_{\mathrm{p}}^{(1)} \leqslant y_{\mathrm{p}}^{(0)},$$

$$(10)\quad f_{\mathrm{p}}\left(x_{\mathrm{p}}^{(k)}\right) \leqslant o_{\mathrm{p}} \leqslant f_{\mathrm{p}}\left(y_{\mathrm{p}}^{(k)}\right),$$

$$(11)\quad \mathscr{P}_{\mathrm{p}}^{(0)} \leqslant \mathscr{P}_{\mathrm{p}}^{(1)} \leqslant \ldots \leqslant \mathscr{P}_{\mathrm{p}}^{(k)} \leqslant \mathscr{P}_{\mathrm{p}}^{(k+1)} \leqslant \ldots,$$

$$(12)\quad \mathscr{P}_{\mathrm{p}}^{(k)} \text{ невырождена},$$

$$(13)\quad \mathscr{P}_{\mathrm{p}}^{(k)} \mathscr{B}_{\mathrm{p}}\left(x_{\mathrm{p}}^{(k)}, y_{\mathrm{p}}^{(k)}\right) \leqslant \mathscr{I}_{\mathrm{p}},$$

$$(14)\quad \mathscr{B}_{\mathrm{p}}\left(x_{\mathrm{p}}^{(k)}, y_{\mathrm{p}}^{(k)}\right) \mathscr{P}_{\mathrm{p}}^{(k)} \leqslant \mathscr{I}_{\mathrm{p}},$$

$$(15)\quad \lim_{k\to\infty} x_{\mathrm{p}}^{(k)} = x_{\mathrm{p}}^{*} \leqslant y_{\mathrm{p}}^{*} = \lim_{k\to\infty} y_{\mathrm{p}}^{(k)},$$

$$(16)\quad \mathscr{P}_{\mathrm{p}}^{*} := \lim_{k\to\infty} \mathscr{P}_{\mathrm{p}}^{(k)} = \mathscr{B}_{\mathrm{p}}\left(x_{\mathrm{p}}^{*}, y_{\mathrm{p}}^{*}\right)^{-1}.$$

Если f_{p} *непрерывна, то* $f_{\mathrm{p}}\left(x_{\mathrm{p}}^{*}\right) = f_{\mathrm{p}}\left(y_{\mathrm{p}}^{*}\right) = o_{\mathrm{p}}$. *Все решения* z_{p}^{*} *системы* $f_{\mathrm{p}}\left(x_{\mathrm{p}}\right) = o_{\mathrm{p}}$, *принадлежащие множеству* $\left\{x_{\mathrm{p}} \mid x_{\mathrm{p}}^{(0)} \leqslant x_{\mathrm{p}} \leqslant y_{\mathrm{p}}^{(0)}\right\}$, *удовлетворяют условию* $x_{\mathrm{p}}^{*} \leqslant z_{\mathrm{p}}^{*} \leqslant y_{\mathrm{p}}^{*}$.

Если матрица $\mathscr{A}_{\mathrm{p}}\left(x_{\mathrm{p}}, y_{\mathrm{p}}\right)$ *невырожденная при* $x_{\mathrm{p}}^{(0)} \leqslant x_{\mathrm{p}} \leqslant y_{\mathrm{p}} \leqslant y_{\mathrm{p}}^{(0)}$, *то* $x_{\mathrm{p}}^{*} = y_{\mathrm{p}}^{*}$.

Если сверх того производная Фреше $f'_{\mathrm{p}}\left(x_{\mathrm{p}}\right)$ *удовлетворяет условию Липшица вида*

$$(17)\quad \left\|f'_{\mathrm{p}}\left(x_{\mathrm{p}}\right) - f'_{\mathrm{p}}\left(y_{\mathrm{p}}\right)\right\| \leqslant c_1 \left\|x_{\mathrm{p}} - y_{\mathrm{p}}\right\| \quad \text{для} \quad x_{\mathrm{p}}^{(0)} \leqslant x_{\mathrm{p}}, \quad y_{\mathrm{p}} \leqslant y_{\mathrm{p}}^{(0)}$$

и отображение $\mathscr{B}_{\mathrm{p}}\left(x_{\mathrm{p}}, y_{\mathrm{p}}\right)$ *удовлетворяет условию*

$$(18)\quad \left\|f'_{\mathrm{p}}\left(z_{\mathrm{p}}\right) - \mathscr{B}_{\mathrm{p}}\left(x_{\mathrm{p}}, y_{\mathrm{p}}\right)\right\| \leqslant c_2 \left\|x_{\mathrm{p}} - y_{\mathrm{p}}\right\| \quad \text{для}$$

$$x_{\mathrm{p}}^{(0)} \leqslant x_{\mathrm{p}} \leqslant z_{\mathrm{p}} \leqslant y_{\mathrm{p}} \leqslant y_{\mathrm{p}}^{(0)},$$

то

$$(19)\quad \lim_{k\to\infty} \mathscr{P}_{\mathrm{p}}^{(k)} = f'_{\mathrm{p}}\left(x_{\mathrm{p}}^{*}\right)^{-1}$$

и последовательность

$$\left\{\left(\left(\left[x_i^{(k)}, y_i^{(k)}\right]\right), \left(\left[r_{ij}^{(k)}, r_{ij}^{(k)}\right]\right)\right)\right\}_{k=0}^{\infty} \in M_{n,\,n+1}\left(I\left(\mathbb{R}\right)\right)$$

сходится к $\left(x_{\mathrm{p}}^{*}, \mathscr{P}_{\mathrm{p}}^{*}\right)$, *причем R-порядок сходимости не меньше 2.*

Доказательство. (9)—(14): Доказываются методом математической индукции. Идея доказательства неравенств (9) и (10) имеется уже в [241] в доказательстве теоремы 4.1, с. 177.

Пусть $k \geqslant 0$ и $x_{\mathrm{p}}^{(k)} \leqslant x_{\mathrm{p}} \leqslant y_{\mathrm{p}}^{(k)}$. Из (2) и (3) следует, что

$$\mathscr{A}_{\mathrm{p}}\left(x_{\mathrm{p}}, y_{\mathrm{p}}^{(k)}\right) \leqslant \mathscr{B}_{\mathrm{p}}\left(x_{\mathrm{p}}, y_{\mathrm{p}}^{(k)}\right) \leqslant \mathscr{B}_{\mathrm{p}}\left(x_{\mathrm{p}}^{(k)}, y_{\mathrm{p}}^{(k)}\right).$$

Ввиду $\mathscr{P}_{\mathrm{p}}^{(k)} \geqslant \mathscr{O}_{\mathrm{p}}$ из (11) следует, что

$$\mathscr{P}_{\mathrm{p}}^{(k)} \mathscr{B}_{\mathrm{p}}\left(x_{\mathrm{p}}^{(k)}, y_{\mathrm{p}}^{(k)}\right) \geqslant \mathscr{P}_{\mathrm{p}}^{(k)} \mathscr{A}_{\mathrm{p}}\left(x_{\mathrm{p}}, y_{\mathrm{p}}^{(k)}\right),$$

а потому с помощью (13) мы получаем

$$\mathscr{I}_{\mathrm{p}} - \mathscr{P}_{\mathrm{p}}^{(k)} \mathscr{A}_{\mathrm{p}}\left(x_{\mathrm{p}}, y_{\mathrm{p}}^{(k)}\right) \geqslant \mathscr{I}_{\mathrm{p}} - \mathscr{P}_{\mathrm{p}}^{(k)} \mathscr{B}_{\mathrm{p}}\left(x_{\mathrm{p}}^{(k)}, y_{\mathrm{p}}^{(k)}\right) \geqslant \mathscr{O}_{\mathrm{p}}.$$

Отсюда, используя (1) и тот факт, что $x_{\mathrm{p}} - y_{\mathrm{p}}^{(k)} \leqslant o_{\mathrm{p}}$, мы получаем неравенства

$$\begin{aligned} x_{\mathrm{p}} - \mathscr{P}_{\mathrm{p}}^{(k)} \mathscr{f}_{\mathrm{p}}\left(x_{\mathrm{p}}\right) &= y_{\mathrm{p}}^{(k+1)} + \left(x_{\mathrm{p}} - y_{\mathrm{p}}^{(k)}\right) - \mathscr{P}_{\mathrm{p}}^{(k)}\left(\mathscr{f}_{\mathrm{p}}\left(x_{\mathrm{p}}\right) - \mathscr{f}_{\mathrm{p}}\left(y_{\mathrm{p}}^{(k)}\right)\right) \\ &= y_{\mathrm{p}}^{(k+1)} + \left(x_{\mathrm{p}} - y_{\mathrm{p}}^{(k)}\right) - \mathscr{P}_{\mathrm{p}}^{(k)} \mathscr{A}_{\mathrm{p}}\left(x_{\mathrm{p}}, y_{\mathrm{p}}^{(k)}\right)\left(x_{\mathrm{p}} - y_{\mathrm{p}}^{(k)}\right) \\ &= y_{\mathrm{p}}^{(k+1)} + \left(\mathscr{I}_{\mathrm{p}} - \mathscr{P}_{\mathrm{p}}^{(k)} \mathscr{A}_{\mathrm{p}}\left(x_{\mathrm{p}}, y_{\mathrm{p}}^{(k)}\right)\right)\left(x_{\mathrm{p}} - y_{\mathrm{p}}^{(k)}\right) \leqslant y_{\mathrm{p}}^{(k+1)}. \end{aligned} \tag{20}$$

Полагая $x_{\mathrm{p}} = x_{\mathrm{p}}^{(k)}$, используя (10) и тот факт, что $\mathscr{P}_{\mathrm{p}}^{(k)} \geqslant \mathscr{O}_{\mathrm{p}}$, мы получаем неравенство

$$x_{\mathrm{p}}^{(k)} \leqslant x_{\mathrm{p}}^{(k)} - \mathscr{P}_{\mathrm{p}}^{(k)} \mathscr{f}_{\mathrm{p}}\left(x_{\mathrm{p}}^{(k)}\right) = x_{\mathrm{p}}^{(k+1)} \leqslant y_{\mathrm{p}}^{(k+1)} = y_{\mathrm{p}}^{(k)} - \mathscr{P}_{\mathrm{p}}^{(k)} \mathscr{f}_{\mathrm{p}}\left(y_{\mathrm{p}}^{(k)}\right),$$

которое доказывает (9).

Из (2), (3) и того факта, что $x_{\mathrm{p}}^{(k)} \leqslant y_{k}^{(k+1)} \leqslant y_{\mathrm{p}}^{(k)}$, следует неравенство

$$\mathscr{A}_{\mathrm{p}}\left(y_{\mathrm{p}}^{(k+1)}, y_{\mathrm{p}}^{(k)}\right) \leqslant \mathscr{B}_{\mathrm{p}}\left(y_{\mathrm{p}}^{(k+1)}, y_{\mathrm{p}}^{(k)}\right) \leqslant \mathscr{B}_{\mathrm{p}}\left(x_{\mathrm{p}}^{(k)}, y_{\mathrm{p}}^{(k)}\right).$$

Используя (14) и неравенство $\mathscr{P}_{\mathrm{p}}^{(k)} \geqslant \mathscr{O}_{\mathrm{p}}$, получаем, что

$$\mathscr{I}_{\mathrm{p}} - \mathscr{A}_{\mathrm{p}}\left(y_{\mathrm{p}}^{(k+1)}, y_{\mathrm{p}}^{(k)}\right) \mathscr{P}_{\mathrm{p}}^{(k)} \geqslant \mathscr{I}_{\mathrm{p}} - \mathscr{B}_{\mathrm{p}}\left(x_{\mathrm{p}}^{(k)}, y_{\mathrm{p}}^{(k)}\right) \mathscr{P}_{\mathrm{p}}^{(k)} \geqslant \mathscr{O}_{\mathrm{p}}.$$

Отсюда с помощью (1) и неравенства $\mathscr{f}_{\mathrm{p}}\left(y_{\mathrm{p}}^{(k)}\right) \geqslant o_{\mathrm{p}}$ (см. (10)) следует, что

$$\begin{aligned} \mathscr{f}_{\mathrm{p}}\left(y_{\mathrm{p}}^{(k+1)}\right) &= \mathscr{f}_{\mathrm{p}}\left(y_{\mathrm{p}}^{(k)}\right) + \mathscr{A}_{\mathrm{p}}\left(y_{\mathrm{p}}^{(\mathrm{p}+1)}, y_{\mathrm{p}}^{(k)}\right)\left(y_{\mathrm{p}}^{(k+1)} - y_{\mathrm{p}}^{(k)}\right) \\ &= \left(\mathscr{I}_{\mathrm{p}} - \mathscr{A}_{\mathrm{p}}\left(y_{\mathrm{p}}^{(k+1)}, y_{\mathrm{p}}^{(k)}\right) \mathscr{P}_{\mathrm{p}}^{(k)}\right) \mathscr{f}_{\mathrm{p}}\left(y_{\mathrm{p}}^{(k)}\right) \geqslant o_{\mathrm{p}}. \end{aligned}$$

Аналогичным образом мы показываем, что $\mathscr{f}_{\mathrm{p}}\left(x_{\mathrm{p}}^{(k+1)}\right) \leqslant o_{\mathrm{p}}$, откуда и следует (10). Из $x_{\mathrm{p}}^{(k)} \leqslant x_{\mathrm{p}}^{(k+1)} \leqslant y_{\mathrm{p}}^{(k+1)} \leqslant y_{\mathrm{p}+1}^{(k)}$ следует с помощью (2), что

$$\mathscr{B}_{\mathrm{p}}\left(x_{\mathrm{p}}^{(k+1)}, y_{\mathrm{p}}^{(k+1)}\right) \leqslant \mathscr{B}_{\mathrm{p}}\left(x_{\mathrm{p}}^{(k)}, y_{\mathrm{p}}^{(k)}\right).$$

С помощью (13) и (14) получаем в силу $\mathscr{P}_{\mathrm{p}}^{(k)} \geqslant \mathscr{O}_{\mathrm{p}}$, что

$$\begin{cases} \mathscr{B}_{\mathrm{p}}\left(x_{\mathrm{p}}^{(k+1)}, y_{\mathrm{p}}^{(k+1)}\right) \mathscr{P}_{\mathrm{p}}^{(k)} \leqslant \mathscr{B}_{\mathrm{p}}\left(x_{\mathrm{p}}^{(k)}, y_{\mathrm{p}}^{(k)}\right) \mathscr{P}_{\mathrm{p}}^{(k)} \leqslant \mathscr{I}_{\mathrm{p}}, \\ \mathscr{P}_{\mathrm{p}}^{(k)} \mathscr{B}_{\mathrm{p}}\left(x_{\mathrm{p}}^{(k+1)}, y_{\mathrm{p}}^{(k+1)}\right) \leqslant \mathscr{P}_{\mathrm{p}}^{(k)} \mathscr{B}_{\mathrm{p}}\left(x_{\mathrm{p}}^{(k)}, y_{\mathrm{p}}^{(k)}\right) \leqslant \mathscr{I}_{\mathrm{p}}. \end{cases} \tag{21}$$

С помощью (8) получаем, что верны равенства

$$\mathscr{I}_{\mathrm{p}} - \mathscr{B}_{\mathrm{p}}\left(x_{\mathrm{p}}^{(k+1)}, y_{\mathrm{p}}^{(k+1)}\right)\mathscr{P}_{\mathrm{p}}^{(k+1)} = \left(\mathscr{I}_{\mathrm{p}} - \mathscr{B}_{\mathrm{p}}\left(x_{\mathrm{p}}^{(k+1)}, y_{\mathrm{p}}^{(k+1)}\right)\mathscr{P}_{\mathrm{p}}^{(k)}\right)^2 \geqslant \mathscr{O}_{\mathrm{p}}$$

и

$$\mathscr{I}_{\mathrm{p}} - \mathscr{P}_{\mathrm{p}}^{(k+1)}\mathscr{B}_{\mathrm{p}}\left(x_{\mathrm{p}}^{(k+1)}, y_{k}^{(k+1)}\right) = \left(\mathscr{I}_{\mathrm{p}} - \mathscr{P}_{\mathrm{p}}^{(k)}\mathscr{B}_{\mathrm{p}}\left(x_{\mathrm{p}}^{(k+1)}, y_{\mathrm{p}}^{(k+1)}\right)\right)^2 \geqslant \mathscr{O}_{\mathrm{p}}, \tag{22}$$

а также

$$\mathscr{P}_{\mathrm{p}}^{(k+1)} = \mathscr{P}_{\mathrm{p}}^{(k)} - \mathscr{P}_{\mathrm{p}}^{(k)}\left(\mathscr{B}_{\mathrm{p}}\left(x_{\mathrm{p}}^{(k+1)}, y_{\mathrm{p}}^{(k+1)}\right)\mathscr{P}_{\mathrm{p}}^{(k)} - \mathscr{I}_{\mathrm{p}}\right) \geqslant \mathscr{P}_{\mathrm{p}}^{(k)} \geqslant \mathscr{O}_{\mathrm{p}}.$$

Отсюда получаются (11), (13) и (14). Чтобы доказать (12), мы должны показать, что матрица $\mathscr{P}_{\mathrm{p}}^{(k+1)}$ неособенная. Так как $\mathscr{P}_{\mathrm{p}}^{(k)}$ неособенная, мы имеем

$$\mathscr{B}_{\mathrm{p}}\left(x_{\mathrm{p}}^{(k+1)}, y_{\mathrm{p}}^{(k+1)}\right) = \mathscr{M}_{\mathrm{p}} - \mathscr{N}_{\mathrm{p}},$$

где $\mathscr{M}_{\mathrm{p}} = \left(\mathscr{P}^{(k)}\right)^{-1}$, $\mathscr{N}_{\mathrm{p}} = \left(\mathscr{P}^{(k)}\right)^{-1} - \mathscr{B}_{\mathrm{p}}\left(x_{\mathrm{p}}^{(k+1)}, y_{\mathrm{p}}^{(k+1)}\right)$.

Условие теоремы показывает, что верно $\mathscr{M}_{\mathrm{p}}^{-1} = \mathscr{P}_{\mathrm{p}}^{(k)} \geqslant \mathscr{O}_{\mathrm{p}}$, откуда с помощью (21) следует, что

$$\mathscr{M}_{\mathrm{p}}^{-1}\mathscr{N}_{\mathrm{p}} = \mathscr{I}_{\mathrm{p}} - \mathscr{P}_{\mathrm{p}}^{(k)}\mathscr{B}_{\mathrm{p}}\left(x_{\mathrm{p}}^{(k+1)}, y_{\mathrm{p}}^{(k+1)}\right) \geqslant \mathscr{O}_{\mathrm{p}},$$

$$\mathscr{N}_{\mathrm{p}}\mathscr{M}_{\mathrm{p}}^{-1} = \mathscr{I}_{\mathrm{p}} - \mathscr{B}_{\mathrm{p}}\left(x_{\mathrm{p}}^{(k+1)}, y_{\mathrm{p}}^{(k+1)}\right)\mathscr{P}_{\mathrm{p}}^{(k)} \geqslant \mathscr{O}_{\mathrm{p}},$$

т. е. $\mathscr{M}_{\mathrm{p}} - \mathscr{N}_{\mathrm{p}}$ является слабо регулярным разбиением матрицы $\mathscr{B}_{\mathrm{p}}\left(x_{\mathrm{p}}^{(k+1)}, y_{\mathrm{p}}^{(k+1)}\right)$ (см. [243, определение 2.4.15]). В силу (4) мы имеем $\mathscr{B}_{\mathrm{p}}\left(x_{\mathrm{p}}^{(k+1)}, y_{\mathrm{p}}^{(k+1)}\right)^{-1} \geqslant \mathscr{O}_{\mathrm{p}}$. Из теоремы 2.4.17 книги [243 с. 56] следует неравенство

$$\rho\left(\mathscr{I}_{\mathrm{p}} - \mathscr{P}_{\mathrm{p}}^{(k)}\mathscr{B}_{\mathrm{p}}\left(x_{\mathrm{p}}^{(k+1)}, y_{\mathrm{p}}^{(k+1)}\right)\right) < 1,$$

где ρ обозначает спектральный радиус. Это означает, что матрица

$$\mathscr{I}_{\mathrm{p}} - \left(\mathscr{I}_{\mathrm{p}} - \mathscr{P}_{\mathrm{p}}^{(k)}\mathscr{B}_{\mathrm{p}}\left(x_{\mathrm{p}}^{(k+1)}, y_{\mathrm{p}}^{(k+1)}\right)\right)^2$$

обратима. В силу (22) отсюда следует, что $\mathscr{P}_{\mathrm{p}}^{(k+1)}$ обратима. Соотношение (15) выводится из (9) с помощью стандартных рассуждений.

Из (13) или (14) следует с помощью (4), что

$$\mathscr{P}_{\mathrm{p}}^{(k)} \leqslant \mathscr{B}_{\mathrm{p}}\left(x_{\mathrm{p}}^{(k)}, y_{\mathrm{p}}^{(k)}\right)^{-1}.$$

Из непрерывности отображения $\mathscr{B}_{\mathrm{p}}(x_{\mathrm{p}}, y_{\mathrm{p}})$ следует, что $\lim\limits_{k\to\infty} \mathscr{B}_{\mathrm{p}}\left(x_{\mathrm{p}}^{(k)}, y_{\mathrm{p}}^{(k)}\right) = \mathscr{B}_{\mathrm{p}}\left(x_{\mathrm{p}}^{*}, y_{\mathrm{p}}^{*}\right)$, и ввиду $x_{\mathrm{p}}^{(k)} \leqslant x_{\mathrm{p}}^{*} \leqslant y_{\mathrm{p}}^{*} \leqslant y_{\mathrm{p}}^{(k)}$ из (2) следует, что

$$\mathscr{B}_{\mathrm{p}}\left(x_{\mathrm{p}}^{*}, y_{\mathrm{p}}^{*}\right) \leqslant \mathscr{B}_{\mathrm{p}}\left(x_{\mathrm{p}}^{(k)}, y_{\mathrm{p}}^{(k)}\right).$$

Отсюда с помощью (4) следует, что

$$\mathscr{B}_{\mathrm{p}}\left(x_{\mathrm{p}}^{(k)},\ y_{\mathrm{p}}^{(k)}\right)^{-1} \leqslant \mathscr{B}_{\mathrm{p}}\left(x_{\mathrm{p}}^{*},\ y_{\mathrm{p}}^{*}\right)^{-1},$$

т. е. верно

$$\mathscr{P}_{\mathrm{p}}^{(k)} \leqslant \mathscr{B}_{\mathrm{p}}\left(x_{\mathrm{p}}^{*},\ y_{\mathrm{p}}^{*}\right)^{-1}.$$

Поэтому последовательность (11) ограничена сверху, а значит она сходится.

Из формул (8), описывающих итерацию, следует с помощью (21) и (11), что

$$\begin{aligned}\mathscr{P}_{\mathrm{p}}^{(k)} - \mathscr{P}_{\mathrm{p}}^{(k+1)} &= \mathscr{P}_{\mathrm{p}}^{(k)}\left(\mathscr{B}_{\mathrm{p}}\left(x_{\mathrm{p}}^{(k+1)},\ y_{\mathrm{p}}^{(k+1)}\right)\mathscr{P}_{\mathrm{p}}^{(k)} - \mathscr{I}_{\mathrm{p}}\right)\\ &\leqslant \mathscr{P}_{\mathrm{p}}^{(0)}\left(\mathscr{B}_{\mathrm{p}}\left(x_{\mathrm{p}}^{(k+1)},\ y_{\mathrm{p}}^{(k+1)}\right)\mathscr{P}_{\mathrm{p}}^{(k)} - \mathscr{I}_{\mathrm{p}}\right) \leqslant \mathscr{O}_{\mathrm{p}},\end{aligned}$$

откуда получается, что

$$\begin{aligned}\mathscr{O}_{\mathrm{p}} &= \lim_{k\to\infty}\left(\mathscr{P}_{\mathrm{p}}^{(k)} - \mathscr{P}_{\mathrm{p}}^{(k+1)}\right) \leqslant \lim_{k\to\infty}\mathscr{P}_{\mathrm{p}}^{(0)}\left[\mathscr{B}_{\mathrm{p}}\left(x_{\mathrm{p}}^{(k+1)},\ y_{\mathrm{p}}^{(k+1)}\right)\mathscr{P}_{\mathrm{p}}^{(k)} - \mathscr{I}_{\mathrm{p}}\right]\\ &= \mathscr{P}_{\mathrm{p}}^{(0)}\left(\mathscr{B}_{\mathrm{p}}\left(x_{\mathrm{p}}^{*},\ y_{\mathrm{p}}^{*}\right)\mathscr{P}_{\mathrm{p}}^{*} - \mathscr{I}_{\mathrm{p}}\right) \leqslant \mathscr{O}_{\mathrm{p}}.\end{aligned}$$

Так как матрица $\mathscr{P}_{\mathrm{p}}^{(0)}$ невырожденная, мы обосновали (16).

Пусть для некоторого $k \geqslant 0$ верно $x_{\mathrm{p}}^{(k)} \leqslant z_{\mathrm{p}} \leqslant y_{\mathrm{p}}^{(k)}$, где $\mathscr{f}_{\mathrm{p}}(z_{\mathrm{p}}) = o_{\mathrm{p}}$. Тогда мы получаем из (20), что

$$z_{\mathrm{p}} = z_{\mathrm{p}} - \mathscr{P}_{\mathrm{p}}^{(k)}\mathscr{f}_{\mathrm{p}}\left(z_{\mathrm{p}}\right) \leqslant y_{\mathrm{p}}^{(k+1)}.$$

Из соотношения

$$x_{\mathrm{p}} - \mathscr{P}_{\mathrm{p}}^{(k)}\mathscr{f}_{\mathrm{p}}\left(x_{\mathrm{p}}\right) \geqslant x_{\mathrm{p}}^{(k+1)},$$

которое справедливо для $x_{\mathrm{p}}^{(k)} \leqslant x_{\mathrm{p}} \leqslant y_{\mathrm{p}}^{(k)}$ и может быть доказано аналогично (20), мы получаем, полагая $x_{\mathrm{p}} := z_{\mathrm{p}}$, что

$$z_{\mathrm{p}} = z_{\mathrm{p}} - \mathscr{P}_{\mathrm{p}}^{(k)}\mathscr{f}_{\mathrm{p}}\left(z_{\mathrm{p}}\right) \geqslant x_{\mathrm{p}}^{(k+1)},$$

т. е.

$$x_{\mathrm{p}}^{(k)} \leqslant x_{\mathrm{p}}^{(k+1)} \leqslant z_{\mathrm{p}} \leqslant y_{\mathrm{p}}^{(k+1)} \leqslant y_{\mathrm{p}}^{(k)}$$

для всех $k \geqslant 0$. Переходя в этих неравенствах к пределу, мы видим, что решения уравнения $\mathscr{f}_{\mathrm{p}}\left(x_{\mathrm{p}}\right) = o_{\mathrm{p}}$, локализованные между $x_{\mathrm{p}}^{(0)}$ и $y_{\mathrm{p}}^{(0)}$, локализованы также и между x_{p}^{*} и y_{p}^{*}.

Из формул (8), описывающих итерацию, следует, что для непрерывной $\mathscr{f}_{\mathrm{p}}$ имеет место

$$o_{\mathrm{p}} = \lim_{k\to\infty}\left(y_{\mathrm{p}}^{(k)} - y_{\mathrm{p}}^{(k+1)}\right) = \lim_{k\to\infty}\mathscr{P}_{\mathrm{p}}^{(k)}\mathscr{f}_{\mathrm{p}}\left(y_{\mathrm{p}}^{(k)}\right) = \mathscr{P}_{\mathrm{p}}^{*}\mathscr{f}_{\mathrm{p}}\left(y_{\mathrm{p}}^{*}\right).$$

Из того что $\mathscr{P}_{\mathrm{p}}^{*} = \mathscr{B}_{\mathrm{p}}\left(x_{\mathrm{p}}^{*},\ y_{\mathrm{p}}\right)^{-1}$ невырожденная, следует, что $\mathscr{f}_{\mathrm{p}}\left(y_{\mathrm{p}}^{*}\right) = o_{\mathrm{p}}$. Аналогично устанавливается, что $\mathscr{f}_{\mathrm{p}}\left(x_{\mathrm{p}}^{*}\right) = o_{\mathrm{p}}$.

Если матрица $\mathscr{A}_p(x_p, y_p)$ для $x_p^{(0)} \leqslant x_p \leqslant y_p \leqslant y_p^{(0)}$ невырожденная, то из (1) следует, что

$$o_p = \mathscr{f}_p(x_p^*) - \mathscr{f}_p(y_p^*) = \mathscr{A}_p(x_p^*, y_p^*)(x_p^* - y_p^*),$$

т. е. $x_p^* = y_p^*$.

Если мы предположим теперь, что выполнено (18), то $\lim_{k\to\infty} x_p^{(k)} = x_p^{(k)} = y_p^* = \lim_{k\to\infty} y_p^{(k)}$ влечет за собой $\lim_{k\to\infty} \mathscr{B}_p(x_p^{(k)}, y_p^{(k)}) = \mathscr{f}'_p(x_p^*)$.

С помощью (16) мы устанавливаем (19).

Доказательство квадратичной сходимости аналогично доказательству теоремы 2 из [286]. Введем сокращения

$$\mathscr{B}_p^{(k)} := \mathscr{B}_p(x_p^{(k)}, y_p^{(k)}), \quad p := \|\mathscr{f}'_p(x_p^*)\|, \quad \mathscr{P}_p^* = \mathscr{f}'_p(x_p^*)^{-1}.$$

Ввиду $\lim_{k\to\infty} \mathscr{P}_p^{(k)} = \mathscr{P}_p^* = \mathscr{f}'_p(x_p^*)^{-1}$ существует константа s, такая что верно $\|\mathscr{P}_p^{(k)}\| \leqslant s$, $k \geqslant 0$.

Из формул (8) следует, с помощью (17), что

$$\begin{aligned}(23)\quad \|y_p^{(k+1)} - x_p^*\| &= \|y_p^{(k)} - x_p^* - \mathscr{P}_p^{(k)}\mathscr{f}_p(y_p^{(k)})\| \\ &= \|-\mathscr{P}_p^{(k)}(\mathscr{f}_p(y_p^{(k)} - \mathscr{f}_p(x_p^*) - \mathscr{f}'_p(x_p^*)(y_p^{(k)} - x_p^*)) \\ &\quad + (\mathscr{f}'_p(x_p^*)^{-1} - \mathscr{P}_p^{(k)})\mathscr{f}'_p(x_p^*)(y_p^{(k)} - x_p^*)\| \\ &\leqslant (c_1/2)\|\mathscr{P}_p^{(k)}\|\|y_p^{(k)} - x_p^*\|^2 + r\|\mathscr{P}_p^* - \mathscr{P}_p^{(k)}\|\|y_p^{(k)} - x_p^*\|.\end{aligned}$$

Аналогично устанавливается, что

$$\begin{aligned}(24)\quad \|x_p^{(k+1)} - x_p^*\| &\leqslant (c_1/2)\|\mathscr{P}_p^{(k)}\|\|x_p^{(k)} - x_p^*\|^2 \\ &\quad + p\|\mathscr{P}_p^* - \mathscr{P}_p^{(k)}\|\|x_p^{(k)} - x_p^*\|.\end{aligned}$$

Используя монотонную норму и теорему об эквивалентности норм, мы можем записать (23) и (24) в виде

$$(23')\quad \|y_p^{(k+1)} - x_p^*\| \leqslant \gamma_1\|y_p^{(k)} - x_p^{(k)}\|^2 + \gamma_2\|\mathscr{P}_p^* - \mathscr{P}_p^{(k)}\|\|y_p^{(k)} - x_p^{(k)}\|,$$

$$(24'')\quad \|x_p^{(k+1)} - x_p^*\| \leqslant \gamma_3\|y_p^{(k)} - x_p^{(k)}\|^2 + \gamma_4\|\mathscr{P}_p^* - \mathscr{P}_p^*\|\|y_p^{(k)} - x_p^{(k)}\|.$$

Из (8) с помощью равенства $\mathscr{I}_p = \mathscr{f}'_p(x_p^*)\mathscr{P}_p^*$ мы получаем

$$\begin{aligned}\mathscr{P}_p^{(k+1)} - \mathscr{P}_p^* &= \mathscr{P}_p^{(k)} - \mathscr{P}_p^* - \mathscr{P}_p^{(k)}(\mathscr{B}_p^{(k+1)}\mathscr{P}_p^{(k)} - \mathscr{I}_p) \\ &= (\mathscr{P}_p^* - \mathscr{P}_p^{(k)})\mathscr{f}'_p(x_p^*)(\mathscr{P}_p^{(k)} - \mathscr{P}_p^*) + \mathscr{P}_p^{(k)}(\mathscr{f}'_p(x_p^*) - \mathscr{B}_p^{(k+1)})\mathscr{P}_p^{(k)}.\end{aligned}$$

С помощью (18) мы получаем теперь

$$\begin{aligned}(25)\quad \|\mathscr{P}_p^{(k+1)} - \mathscr{P}_p^*\| &\leqslant p\|\mathscr{P}_p^{(k)} - \mathscr{P}_p^*\|^2 + c_2\|\mathscr{P}_p^{(k)}\|^2\|x_p^{(k+1)} - y_p^{(k+1)}\| \\ &\leqslant \gamma_5\|\mathscr{P}_p^{(k)} - \mathscr{P}_p^*\|^2 + \gamma_6\|y_p^{(k+1)} - x_p^{(k+1)}\|.\end{aligned}$$

Введем норму для матриц $\mathscr{A}_{\mathrm{p}} = (\mathscr{b}_{\mathrm{p}}, \mathscr{B}_{\mathrm{p}}) \in M_{n,\, n+1}(\mathbb{R})$ с помощью равенства

$$\| \mathscr{A}_{\mathrm{p}} \| = \max \{ \| \mathscr{b}_{\mathrm{p}} \|, \| \mathscr{B}_{\mathrm{p}} \| \}$$

и положим

$$d^{(k)} = \| \left(y_{\mathrm{p}}^{(k)} - x_{\mathrm{p}}^{(k)},\ \mathscr{P}_{\mathrm{p}}^{*} - \mathscr{P}_{\mathrm{p}}^{(k)} \right) \|.$$

Из (23′), (24′) и (25) мы получаем теперь

$$\| y_{\mathrm{p}}^{(k+1)} - x_{\mathrm{p}}^{(n+1)} \| \leqslant \left(\sum_{i=1}^{4} \gamma_i \right) (d^{(k)})^2,$$

$$\| \mathscr{P}_{\mathrm{p}}^{*} - \mathscr{P}_{\mathrm{p}}^{(k+1)} \| \leqslant \left(\gamma_5 + \gamma_6 \left(\sum_{i=1}^{4} \gamma_i \right) \right) (d^{(k)})^2,$$

т. е.

$$(d^{(k+1)} \leqslant \gamma (d^{(k)})^2,$$

где

$$\gamma = \max \left\{ \sum_{i=1}^{4} \gamma_i,\ \gamma_5 + \gamma_6 \sum_{i=1}^{4} \gamma_i \right\}.$$

Теперь мы получаем нашу теорему, сославшись на теорему 2 из приложения А. ■

Содержание теоремы 1 легко перенести на методы с более высокой скоростью сходимости [286]. Для этого мы сохраняем матрицу $\mathscr{P}_{\mathrm{p}}^{(k)}$ неизменной в течение нескольких шагов. Приближенное обращение матрицы $\mathscr{P}_{\mathrm{p}} \left(x_{\mathrm{p}}^{(k+1)},\ y_{\mathrm{p}}^{(k+1)} \right)$ делается с помощью метода высокого порядка. В результате это дает следующую итерацию:

$$\begin{cases} y_{\mathrm{p}}^{(k,\,0)} = y_{\mathrm{p}}^{(k)}, \\ y_{\mathrm{p}}^{(k,\,r+1)} = y_{\mathrm{p}}^{(k,\,r)} - \mathscr{P}_{\mathrm{p}}^{(k)} \mathscr{f}_{\mathrm{p}} \left(y_{\mathrm{p}}^{(k,\,r)} \right), \qquad 0 \leqslant r \leqslant m, \\ y_{\mathrm{p}}^{(k+1)} = y_{\mathrm{p}}^{(k,\,m+1)}, \\ x_{\mathrm{p}}^{(k,\,0)} = x_{\mathrm{p}}^{(k)}, \\ x_{\mathrm{p}}^{(k,\,r+1)} = x_{\mathrm{p}}^{(k,\,r)} - \mathscr{P}_{\mathrm{p}}^{(k)} \mathscr{f}_{\mathrm{p}} \left(x_{k}^{(k,\,r)} \right), \qquad 0 \leqslant r \leqslant m, \\ x_{\mathrm{p}}^{(k+1)} = x_{\mathrm{p}}^{(k,\,m+1)}, \\ \mathscr{P}_{\mathrm{p}}^{(k+1)} = \mathscr{P}_{\mathrm{p}}^{(k)} + \mathscr{P}_{\mathrm{p}}^{(k)} \sum_{\mu=1}^{m+1} (-1)^{\mu} \left[\mathscr{B}_{\mathrm{p}} \left(x_{\mathrm{p}}^{(k+1)},\ y_{\mathrm{p}}^{(k+1)} \right) \mathscr{P}_{\mathrm{p}}^{(k)} - \mathscr{I}_{\mathrm{p}} \right]^{\mu}, \\ k \geqslant 0. \end{cases}$$

Этот метод сходится в условиях теоремы 1, и порядок сходимости равен $m + 2$.

Следующее утверждение содержит частный случай теоремы 1, который получается, когда обращение матрицы $\mathscr{B}_{\mathrm{p}}\left(x_{\mathrm{p}}^{(k)}, y_{\mathrm{p}}^{(k)}\right)$ производится точно, а не приближенно.

Теорема 2. *Пусть в теореме* 1 $\mathscr{P}_{\mathrm{p}}^{(k)} = \mathscr{B}_{\mathrm{p}}\left(x_{\mathrm{p}}^{(k)}, y_{\mathrm{p}}^{(k)}\right)^{-1}$, *а итерация* (8) *заменена на*

$$(8') \qquad \begin{cases} y_{\mathrm{p}}^{(k+1)} = y_{\mathrm{p}}^{(k)} - \mathscr{P}_{\mathrm{p}}^{(k)} \mathscr{f}_{\mathrm{p}}\left(y_{\mathrm{p}}^{(k)}\right), \\ x_{\mathrm{p}}^{(k+1)} = x_{\mathrm{p}}^{(k)} - \mathscr{P}_{\mathrm{p}}^{(k)} \mathscr{f}_{\mathrm{p}}\left(x_{\mathrm{p}}^{(k)}\right), \qquad k \geqslant 0. \end{cases}$$

Тогда верны все утверждения теоремы 1.

Доказательство. (5), (6), (7), а также (12), (13) и (14) выполнены тривиальным образом ввиду (4). Неравенство (11) также следует из (4). Остальную часть доказательства можно повторить без изменений. ■

Следует заметить, что при численном исполнении метода (8′) треугольное разложение матрицы $\mathscr{B}_{\mathrm{p}}\left(x_{\mathrm{p}}^{(k)}, y_{\mathrm{p}}^{(k)}\right)$ по методу Гаусса следует производить только один раз.

Рассмотрим теперь класс систем нелинейных уравнений, к которому можно применить предыдущие теоремы.

Пусть

$$(26) \quad \mathscr{f}_{\mathrm{p}}\colon \mathfrak{D}_{\mathrm{p}} \subseteq V_n(\mathbb{R}) \to V_n(\mathbb{R}), \qquad \mathscr{f}_{\mathrm{p}}(x_{\mathrm{p}}) = \mathscr{H}_{\mathrm{p}} x_{\mathrm{p}} + \varepsilon h_{\mathrm{p}}(x_{\mathrm{p}}) + b_{\mathrm{p}},$$

где $\varepsilon \geqslant 0$, $b_{\mathrm{p}} \in V_n(\mathbb{R})$ и $\mathscr{H}_{\mathrm{p}}$ является M-матрицей, т. е. $h_{ij} \leqslant 0$ для $i \neq j$ и $\mathscr{H}_{\mathrm{p}}^{-1} \geqslant O_{\mathrm{p}}$. Отображение

$$h_{\mathrm{p}}\colon \mathfrak{D} \subseteq V_n(\mathbb{R}) \to V_n(\mathbb{R}), \quad h_{\mathrm{p}}(x_{\mathrm{p}}) = (h_i(x_{\mathrm{p}}))$$

непрерывно дифференцируемо и

$$h_i(x_{\mathrm{p}}) = \sum_{j=1}^{n} \alpha_{ij} g_j(x_j), \qquad 1 \leqslant i \leqslant n,$$

где $g_i\colon \mathfrak{D}_i \subseteq \mathbb{R} \to \mathbb{R}$, $g_i = g_i(s)$, $1 \leqslant i \leqslant n$,

$$\frac{d}{ds} g_i(s) \geqslant 0, \qquad 1 \leqslant i \leqslant n,$$

и выполнено

$$\alpha_{ij} \geqslant 0, \qquad 1 \leqslant i, j \leqslant n$$

Полагая

$$g_{\mathrm{p}}(x_{\mathrm{p}}) = (g_i(x_i)) \quad \text{и} \quad \mathscr{U}_{\mathrm{p}} = (\alpha_{ij}),$$

мы можем записать $\mathscr{f}_{\mathrm{p}}$ в виде

$$\mathscr{f}_{\mathrm{p}}(x_{\mathrm{p}}) = \mathscr{H}_{\mathrm{p}} x_{\mathrm{p}} + \varepsilon \mathscr{U}_{\mathrm{p}} g_{\mathrm{p}}(x_{\mathrm{p}}) + b_{\mathrm{p}}.$$

Пусть $\mathscr{A}_{\mathrm{p}}(x_{\mathrm{p}}, y_{\mathrm{p}})$ — матрица, i-я строка которой равна

$$f_i'(x_{\mathrm{p}}+\theta_i(y_{\mathrm{p}}-x_{\mathrm{p}})) \in \left(h_{i1}+\varepsilon\alpha_{i1}\frac{d}{dx_1}g_1(x_1+\theta_i(y_1-x_1)),\ \ldots,\right.$$
$$\left. h_{in}+\varepsilon\alpha_{in}\frac{d}{dx_n}g_n(x_n+\theta_i(y_n-x_n))\right)$$

для $\theta_i \in (0, 1)$, $1 \leqslant i \leqslant n$. Если $\mathfrak{D}$ выпукла, то из теоремы о среднем значении для отображений, определенных на $\mathfrak{D} \subseteq V_n(\mathbb{R}) \to \mathbb{R}$, следует, что

$$f_{\mathrm{p}}(x_{\mathrm{p}}) - f_{\mathrm{p}}(y_{\mathrm{p}}) \in \mathscr{A}_{\mathrm{p}}(x_{\mathrm{p}}, y_{\mathrm{p}})(x_{\mathrm{p}} - y_{\mathrm{p}})$$

для $x_{\mathrm{p}}, y_{\mathrm{p}} \in \mathfrak{D}$.

Пусть $v_j(x_j, y_j)$ — верхняя граница в интервальном вычислении производной dg_j/dx_j для интервала $[x_j, y_j]$ с $x_j \leqslant y_j$. Из монотонности включения следует, что

$$\frac{d}{dx_j}g_j(\bar{x}_j+\bar{\theta}_i(\bar{y}_j-\bar{x}_j)) \in \frac{d}{dx_j}g_j([\bar{x}_j, \bar{y}_j]) \subseteq \frac{d}{dx_j}g_j([x_j, y_j])$$

для $[\bar{x}_j, \bar{y}_j] \subseteq [x_j, y_j]$, т. е.

$$v_j(x_j, y_j) \geqslant \frac{d}{dx_j}g_j(\bar{x}_j+\bar{\theta}_i(\bar{y}_j-\bar{x}_j)) \geqslant 0.$$

Если мы теперь положим

$$\overline{\mathscr{U}}_{\mathrm{p}} = (\bar{u}_{ij}),\ \text{где}\ \bar{u}_{ij} = \alpha_{ij}v_j(x_j, y_j),\quad 1 \leqslant i,\ j \leqslant n,$$

то в обозначениях

$$\mathscr{B}_{\mathrm{p}}(x_{\mathrm{p}}, y_{\mathrm{p}}) = (h_{ij}+\varepsilon\alpha_{ij}v_j(x_j, y_j))$$

получим, что

$$\mathscr{B}_{\mathrm{p}}(\bar{x}_{\mathrm{p}}, \bar{y}_{\mathrm{p}}) \leqslant \mathscr{B}_{\mathrm{p}}(x_{\mathrm{p}}, y_{\mathrm{p}})\quad \text{для}\quad x_{\mathrm{p}} \leqslant \bar{x}_{\mathrm{p}} \leqslant \bar{y}_{\mathrm{p}} \leqslant y_{\mathrm{p}},$$

а также

$$\mathscr{A}_{\mathrm{p}}(x_{\mathrm{p}}, y_{\mathrm{p}}) \leqslant \mathscr{B}_{\mathrm{p}}(x_{\mathrm{p}}, y_{\mathrm{p}}).$$

Теперь выполнены условия (1), (2), (3) теоремы 1. Так как $\mathscr{H}_{\mathrm{p}}$ — M-матрица, то из теоремы 2.4.10 книги [243] и того факта, что $\varepsilon \geqslant 0$, следует при достаточно малых $\mathscr{U}_{\mathrm{p}} \geqslant \mathscr{O}_{\mathrm{p}}$, что $\mathscr{B}_{\mathrm{p}}(x_{\mathrm{p}}, y_{\mathrm{p}})$ является M-матрицей. Это означает, что выполнено (4) (при условии, что ε удовлетворяет неравенству

$$h_{ij}+\varepsilon\alpha_{ij}v_j(x_j^{(0)}, y_j^{(0)}) \leqslant 0,\quad i \neq j,\ 1 \leqslant i,\ j \leqslant n).$$

Пусть $\mathscr{D}_{\mathrm{p}}$ обозначает диагональную часть матрицы $\mathscr{B}_{\mathrm{p}}(x_{\mathrm{p}}^{(0)}, y_{\mathrm{p}}^{(0)})$. Ввиду $\mathscr{D}_{\mathrm{p}}^{-1} \geqslant \mathscr{O}_{\mathrm{p}}$ и $\mathscr{B}_{\mathrm{p}}(x_{\mathrm{p}}^{(0)}, y_{\mathrm{p}}^{(0)}) = \mathscr{D}_{\mathrm{p}} - (\mathscr{D}_{\mathrm{p}} -$

$-\mathscr{B}_{\mathrm{p}}\left(x_{\mathrm{p}}^{(0)}, y_{\mathrm{p}}^{(0)}\right)$) мы, полагая $\mathscr{P}_{\mathrm{p}}^{(0)}=\mathscr{D}_{\mathrm{p}}^{-1}$, получаем, что

$$\mathscr{B}_{\mathrm{p}}\left(x_{\mathrm{p}}^{(0)}, y_{\mathrm{p}}^{(0)}\right)\mathscr{P}_{\mathrm{p}}^{(0)}=\mathscr{I}_{\mathrm{p}}-\left(\mathscr{D}_{\mathrm{p}}-\mathscr{B}_{\mathrm{p}}\left(x_{\mathrm{p}}^{(0)}, y_{\mathrm{p}}^{(0)}\right)\right)\mathscr{D}_{\mathrm{p}}^{-1}\leqslant\mathscr{I}_{\mathrm{p}},$$

$$\mathscr{P}_{\mathrm{p}}^{(0)}\mathscr{B}_{\mathrm{p}}\left(x_{\mathrm{p}}^{(0)}, y_{\mathrm{p}}^{(0)}\right)=\mathscr{I}_{\mathrm{p}}-\mathscr{D}_{\mathrm{p}}^{-1}\left(\mathscr{D}_{\mathrm{p}}-\mathscr{B}_{\mathrm{p}}\left(x_{\mathrm{p}}^{(0)}, y_{\mathrm{p}}^{(0)}\right)\right)\leqslant\mathscr{I}_{\mathrm{p}},$$

так как $\mathscr{D}_{\mathrm{p}}-\mathscr{B}_{\mathrm{p}}\left(x_{\mathrm{p}}^{(0)}, y_{\mathrm{p}}^{(0)}\right)\geqslant\mathscr{O}_{\mathrm{p}}$.

Наш выбор матрицы $\mathscr{P}_{\mathrm{p}}^{(0)}$ обеспечивает выполнение условий (5), (6) и (7) теоремы 1.

Если матрица $\mathscr{B}_{\mathrm{p}}\left(x_{\mathrm{p}}^{(0)}, y_{\mathrm{p}}^{(0)}\right)$ представлена в виде $\mathscr{D}_{\mathrm{p}}-\mathscr{L}_{\mathrm{p}}-\mathscr{R}_{\mathrm{p}}$, где $\mathscr{L}_{\mathrm{p}}$ — строго нижняя треугольная матрица, а $\mathscr{R}_{\mathrm{p}}$ — строго верхняя треугольная матрица, то $\mathscr{P}_{\mathrm{p}}^{(0)}=(\mathscr{D}_{\mathrm{p}}-\mathscr{L}_{\mathrm{p}})^{-1}\geqslant\mathscr{O}_{\mathrm{p}}$, а ввиду $\mathscr{R}_{\mathrm{p}}\geqslant\mathscr{O}_{\mathrm{p}}$ следует, что

$$\mathscr{B}_{\mathrm{p}}\left(x_{\mathrm{p}}^{(0)}, y_{\mathrm{p}}^{(0)}\right)\mathscr{P}_{\mathrm{p}}^{(0)}=\mathscr{I}_{\mathrm{p}}-\mathscr{R}_{\mathrm{p}}\left(\mathscr{D}_{\mathrm{p}}-\mathscr{L}_{\mathrm{p}}\right)^{-1}\leqslant\mathscr{I}_{\mathrm{p}},$$

$$\mathscr{P}_{\mathrm{p}}^{(0)}\mathscr{B}_{\mathrm{p}}\left(x_{\mathrm{p}}^{(0)}, y_{\mathrm{p}}^{(0)}\right)=\mathscr{I}_{\mathrm{p}}-\left(\mathscr{D}_{\mathrm{p}}-\mathscr{L}_{\mathrm{p}}\right)^{-1}\mathscr{R}_{\mathrm{p}}\leqslant\mathscr{I}_{\mathrm{p}}.$$

Это указывает еще одного возможного кандидата на роль $\mathscr{P}_{\mathrm{p}}^{(0)}$.

Условие (17) будет выполнено, если производные отображений g_i удовлетворяют условию Липшица. Тогда выполнено и условие (18). Действительно, для

$$\mathscr{f}'_{\mathrm{p}}\left(z_{\mathrm{p}}\right)=\left(h_{ij}+\varepsilon\alpha_{ij}\frac{d}{dx_j}g_j\left(z_j\right)\right)$$

и

$$\mathscr{B}_{\mathrm{p}}\left(x_{\mathrm{p}}, y_{\mathrm{p}}\right)=\left(h_{ij}+\varepsilon\alpha_{ij}v_j\left(x_j, y_j\right)\right)$$

очевидным образом выполнено соотношение

$$\mathscr{f}'_{\mathrm{p}}\left(z_{\mathrm{p}}\right)-\mathscr{B}_{\mathrm{p}}\left(x_{\mathrm{p}}, y_{\mathrm{p}}\right)=\left(\varepsilon\alpha_{ij}\left(\frac{d}{dx_j}g_j\left(z_j\right)-v_j\left(x_j, y_j\right)\right)\right).$$

Если мы положим теперь

$$\left[u_j\left(x_j, y_j\right), v_j\left(x_j, y_j\right)\right]:=\frac{d}{dx_j}g_j\left(\left[x_j, y_j\right]\right),$$

то из монотонности включения следует, что

$$\frac{d}{dx_j}g_j\left(z_j\right)\in\left[u_j\left(x_j, y_j\right), v_j\left(x_j, y_j\right)\right].$$

Теперь из теоремы 3.5 следует, что

$$v_j\left(x_j, y_j\right)-\frac{d}{dx_j}g_j\left(z_j\right)\leqslant v_j\left(x_j, y_j\right)-u_j\left(x_j, y_j\right)\leqslant c_j\left(y_j-x_j\right).$$

Поэтому мы получаем

$$\mathscr{B}_{\mathrm{p}}\left(x_{\mathrm{p}}, y_{\mathrm{p}}\right)-\mathscr{f}'_{\mathrm{p}}\left(z_{\mathrm{p}}\right)\leqslant\max_{1\leqslant i\leqslant n}\left\{y_i-x_i\right\}\mathscr{C}_{\mathrm{p}}$$

для матрицы $\mathscr{C}_{\mathrm{p}} = (c_{ij})$, $c_{ij} = c_j \varepsilon \alpha_{ij} \geqslant 0$, $1 \leqslant i, j \leqslant n$, откуда следует (18).

Системы нелинейных уравнений вида (26) получаются при численном решении граничных задач вида

$$y'' = f(t, y), \quad y(0) = \bar{a}, \quad y(1) = \bar{b}.$$

Применение метода конечных разностей к такой системе дает n уравнений

$$x_{i-2} - 2x_i + x_{i+1} - h^2 \{\alpha f(t_{i-1}, x_{i-1}) + \beta f(t_i, x_i) + \gamma f(t_{i+1}, x_{i+1})\} = 0, \quad 1 \leqslant i \leqslant n,$$

где $t_i = ih$, $0 \leqslant i \leqslant n+1$, $(n+1)h = 1$ и $x_0 = \bar{a}$, $x_{n+1} = \bar{b}$. Числа x_i — приближения к $y(t_i)$. При $\alpha = \gamma = 0$, $\beta = 1$ получается стандартный метод конечных разностей. При $\alpha = \gamma = \frac{1}{2}$, $\beta = \frac{10}{12}$ получается эрмитовский метод (см. [341, с. 461]). Вводя обозначения

$$x_{\mathrm{p}} = \begin{Bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{Bmatrix}, \quad g_{\mathrm{p}}(x_{\mathrm{p}}) = \begin{Bmatrix} f(t_1, x_1) \\ f(t_2, x_2) \\ \vdots \\ f(t_n, x_n) \end{Bmatrix}, \quad f_{\mathrm{p}} = \begin{Bmatrix} -\bar{a} + \alpha h^2 f(0, \bar{a}) \\ 0 \\ \vdots \\ -\bar{b} + \gamma h^2 f(1, \bar{b}) \end{Bmatrix},$$

$$\mathscr{H}_{\mathrm{p}} = \begin{Bmatrix} 2 & -1 & & & 0 \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ 0 & & & -1 & 2 \end{Bmatrix}, \quad \mathscr{U}_{\mathrm{p}} = \begin{Bmatrix} \beta & \gamma & & & 0 \\ \alpha & \beta & \gamma & & \\ & \ddots & \ddots & \ddots & \\ & & \alpha & \beta & \gamma \\ 0 & & & \alpha & \beta \end{Bmatrix}$$

и полагая $\varepsilon = h^2$, можно записать эти n уравнений в виде

$$f_{\mathrm{p}}(x_{\mathrm{p}}) = \mathscr{H}_{\mathrm{p}} x_{\mathrm{p}} + \varepsilon \mathscr{U}_{\mathrm{p}} g_{\mathrm{p}}(x_{\mathrm{p}}) - b_{\mathrm{p}}.$$

Из предположения $\partial f / \partial y \geqslant 0$ для $t \in (0, 1)$ следует, что f_{p} удовлетворяет всем условиям, наложенным на системы вида (26). В случае эрмитовского метода нужно еще выбрать h достаточно малым.

Мы рассмотрим в качестве конкретного примера граничную задачу

$$y'' = \sin y + y, \quad y(0) = 1, \quad y(1) = 1.$$

Так как $\partial^2 f / \partial y^2 = -\sin y$ меняет знак, то очевидно, что дискретизированная задача не является выпуклой на всем пространстве $V_n(\mathbb{R})$. Последовательные приближения к величине $y\left(\frac{1}{2}\right)$ для $n = 5, 25, 51$ и 101 приведены в табл. 1 для обычного метода конечных разностей и в табл. 2 для эрмитовского метода.

Таблица 1

Обычный метод конечных разностей

k	$n=5$	
0	−0.5	0.5
1	0.3940299983760	0.4000335866235
2	0.3989339526997	0.3989347005095
3	0.3989344659822	0.3989344659822

	$n=25$	
0	−0.5	0.5
1	0.3935413781128	0.3997788906381
2	0.3986874733555	0.3986882610402
3	0.3986880255447	0.3986880255452
4	0.3986880255452	0.3986880255452

k	$n=51$	
0	−0.5	0.5
1	0.3935206369679	0.3997680696930
2	0.3986771185469	0.3986779080124
3	0.3986776725106	0.3986776725169
4	0.3986776725137	0.3986776725153

k	$n=101$	
0	−0.5	0.5
1	0.3935155168238	0.3997653993461
2	0.3986745645657	0.3986753545055
3	0.3986751189564	0.3986751189564

Таблица 2

Эрмитовский метод

k	$n=5$	
0	−0.5	0.5
1	0.3938048950831	0.3997635541509
2	0.3986758325059	0.3986765352929
3	0.3986763144021	0.3986763144021

k	$n=25$	
0	−0.5	0.5
1	0.3935292233327	0.3997644587939
2	0.3986736779243	0.3986744630652
3	0.3986742283178	0.3986742283191

k	$n=51$	
0	−0.5	0.5
1	0.3935175960669	0.3997644612118
2	0.3986736691791	0.3986744580066
3	0.3986742226748	0.3986742226762
4	0.3986742226762	0.3986742226762

k	$n=101$	
0	−0.5	0.5
1	0.3935147300836	0.3997644611468
2	0.3986736680718	0.3986744577855
3	0.3986742223087	0.3986742223447
4	0.3986742223155	0.3986742223174

Начальные векторы $x_p^{(0)}$ и $y_p^{(0)}$ были найдены согласно п. 13.4.6c из [243]. При этих значениях для эрмитовского метода имеет место

$$f_p(x_p^{(0)}) \leqslant o_p \leqslant f_p(y_p^{(0)}).$$

Замечания. Как уже упоминалось, результаты этой главы тесно связаны с соответствующими результатами из [241, 243]. Все соответствующие утверждения доказаны при определенных предположениях о выпуклости. Простые примеры показывают, что для методов, рассмотренных в [241, 243], эти предположения нельзя ослабить. Вычисления в интервальной арифметике используются в этой главе только для того, чтобы вычислить по системам вида (26) матрицы $\mathscr{B}_p(x_p^{(k)} y_p^{(k)})$, нужные для применения методов (8) и (8′).

n-мерные аналогии правила regula falsi и метода секущих рассмотрены в [283, 284, 287] и применены для локализации корней выпуклых отображений. Не вполне ясно, до какой степени эти методы можно модифицировать, используя интервальную арифметику, чтобы они могли локализовать решения невыпуклых систем.

Заметим, что в условиях теоремы 1 можно вывести (17) из (18) с помощью неравенства треугольника.

* * *

Литература: [13], [21], [43], [74], [215], [241], [243], [283], [284], [286], [287], [324], [341].

Глава 22

ПОЛНОШАГОВЫЕ И КОРОТКОШАГОВЫЕ МЕТОДЫ НЬЮТОНОВСКОГО ТИПА

В этой главе мы рассмотрим итерационные методы локализации решений вещественных систем нелинейных уравнений, где не требуется обращать матрицы ни точно (как в гл. 19), ни приближенно (как в гл. 20). Мы снова исходим из системы нелинейных уравнений

$$\mathcal{f}_{\mathrm{p}}(x_{\mathrm{p}}) = o_{\mathrm{p}}$$

и предполагаем, что $\mathcal{f}_{\mathrm{p}}$ непрерывно дифференцируема над данным интервальным вектором $x^{(0)}$. Применяя теорему о среднем значении и полагая $\mathcal{f}_{\mathrm{p}}(x_{\mathrm{p}}) = (f_i(x_{\mathrm{p}}))$, мы получаем тогда

$$f_i(x_{\mathrm{p}}) = f_i(y_{\mathrm{p}}) + \sum_{j=1}^{n} \frac{\partial}{\partial x_j} f_i(y_{\mathrm{p}} + \theta_i(x_{\mathrm{p}} - y_{\mathrm{p}}))(x_j - y_j),$$

$$0 < \theta_i < 1, \quad 1 \leqslant i \leqslant n, \quad y_{\mathrm{p}}, \ x_{\mathrm{p}} \in x^{(0)},$$

и, полагая

$$\mathcal{J}_{\mathrm{p}}(x_{\mathrm{p}}) = \left(\frac{\partial}{\partial x_j} f_i(y_{\mathrm{p}} + \theta_i(x_{\mathrm{p}} - y_{\mathrm{p}}))\right),$$

получаем уравнение

$$\mathcal{f}_{\mathrm{p}}(x_{\mathrm{p}}) - \mathcal{f}_{\mathrm{p}}(y_{\mathrm{p}}) = \mathcal{J}_{\mathrm{p}}(x_{\mathrm{p}})(x_{\mathrm{p}} - y_{\mathrm{p}}).$$

Если y_{p} — решение исходной системы уравнений, принадлежащее $x_{\mathrm{p}}^{(0)}$, и мы представили матрицу $\mathcal{J}_{\mathrm{p}}(x_{\mathrm{p}})$ в виде

$$\mathcal{J}_{\mathrm{p}}(x_{\mathrm{p}}) = \mathcal{D}_{\mathrm{p}}(x_{\mathrm{p}}) - \mathcal{L}_{\mathrm{p}}(x_{\mathrm{p}}) - \mathcal{U}_{\mathrm{p}}(x_{\mathrm{p}}),$$

где $\mathcal{D}_{\mathrm{p}}(x_{\mathrm{p}})$ — диагональная матрица, $\mathcal{L}_{\mathrm{p}}(x_{\mathrm{p}})$ — строго нижняя треугольная матрица $\mathcal{U}_{\mathrm{p}}(x_{\mathrm{p}})$ — строго верхняя треугольная ма-

трица, то из последнего уравнения следует для невырожденной матрицы $\mathscr{D}_{\mathrm{p}}(x_{\mathrm{p}})$ соотношение

$$y_{\mathrm{p}} = x_{\mathrm{p}} - \mathscr{D}_{\mathrm{p}}(x_{\mathrm{p}})^{-1}\{\mathscr{B}_{\mathrm{p}}(x_{\mathrm{p}})(x_{\mathrm{p}} - y_{\mathrm{p}}) + \mathscr{f}_{\mathrm{p}}(x_{\mathrm{p}})\},$$

где $\mathscr{B}_{\mathrm{p}}(x_{\mathrm{p}}) = \mathscr{L}_{\mathrm{p}}(x_{\mathrm{p}}) + \mathscr{U}_{\mathrm{p}}(x_{\mathrm{p}})$.

Ввиду $y_{\mathrm{p}} + \theta_i(x_{\mathrm{p}} - y_{\mathrm{p}}) \in x^{(0)}$, $1 \leqslant i \leqslant n$, из монотонности включения следует, что

$$\frac{\partial}{\partial x_j} f_i(y_{\mathrm{p}} + \theta_i(x_{\mathrm{p}} - y_{\mathrm{p}})) \in \frac{\partial}{\partial x_j} f_i(x^{(0)}), \quad 1 \leqslant i,\ j \leqslant n.$$

Представим теперь интервальную матрицу

$$\mathscr{f}'_{\mathrm{p}}(x^{(0)}) = \left(\frac{\partial}{\partial x_j} f_i(x^{(0)})\right)$$

в том же виде, что и матрицу $\mathscr{J}_{\mathrm{p}}(x_{\mathrm{p}})$:

$$\mathscr{f}'_{\mathrm{p}}(x^{(0)}) = \mathscr{D}_{\mathrm{p}}(x^{(0)}) - \mathscr{L}_{\mathrm{p}}(x^{(0)}) - \mathscr{U}_{\mathrm{p}}(x^{(0)}).$$

Положим еще $\mathscr{B}_{\mathrm{p}}(x^{(0)}) := \mathscr{L}_{\mathrm{p}}(x^{(0)}) + \mathscr{U}_{\mathrm{p}}(x^{(0)})$ и $x_{\mathrm{p}} := m(x^{(0)})$ (см. (18.1)). Если никакой диагональный элемент матрицы $\mathscr{f}'_{\mathrm{p}}(x^{(0)})$ не содержит нуля, то с помощью монотонности включения мы получим

$$y_{\mathrm{p}} \in m(x^{(0)}) - \mathscr{D}_{\mathrm{p}}(x^{(0)})^{-1}\{\mathscr{B}_{\mathrm{p}}(x^{(0)})(m(x^{(0)}) - y_{\mathrm{p}}) + \mathscr{f}_{\mathrm{p}}(m(x^{(0)}))\}.$$

Здесь подразумевается, что $\mathscr{D}_{\mathrm{p}}(x^{(0)})^{-1}$ — это диагональная матрица, полученная из диагональной матрицы

$$\mathscr{D}_{\mathrm{p}}(x^{(0)}) = \operatorname{diag}\left(\frac{d}{dx_i} f_i(x^{(0)})\right)$$

обращением диагональных элементов.

Если теперь выбран интервальный вектор $z^{(0)}$, содержащий y_{p}, то с помощью математической индукции по номеру шага итерации в итерационном методе

$$(1)\qquad z^{(\nu+1)} = \{m(x^{(0)}) - \mathscr{D}_{\mathrm{p}}(x^{(0)})^{-1}\{\mathscr{B}_{\mathrm{p}}(x^{(0)}(m(x^{(0)}) - z^{(\nu)}) + \mathscr{f}_{\mathrm{p}}(m(x^{(0)}))\}\} \cap z^{(\nu)}$$

можно показать, что имеет место $y_{\mathrm{p}} \in z^{(\nu)}$, $\nu \geqslant 0$. Так как на каждом шаге берется пересечение, последовательность $\{z^{(\nu)}\}_{\nu=0}^{\infty}$ сходится к некоторому пределу z, и имеем $y_{\mathrm{p}} \in z$. Исходя из $x^{(1)} := z$, вычисляем интервальную матрицу $\mathscr{f}'_{\mathrm{p}}(x^{(1)})$, а с ее помощью — новую локализацию $x^{(2)}$ решения y_{p} и т. д. Однако мы выберем сейчас другой путь. В итерационном методе (1) положим $z^{(0)} := x^{(0)}$. Из предшествующих рассуждений следует, что

$$y_{\mathrm{p}} \in z^{(1)} = \{m(x^{(0)}) - \mathscr{D}_{\mathrm{p}}(x^{(0)})^{-1}\{\mathscr{B}_{\mathrm{p}}(x^{(0)})(m(x^{(0)}) - x^{(0)}) + \mathscr{f}_{\mathrm{p}}(m(x^{(0)}))\}\} \cap x^{(0)}$$

и $z^{(1)} \subseteq x^{(0)}$. Кажется естественным взять $x^{(1)} := z^{(1)}$ и вычислить новую интервальную матрицу $f'_{\mathrm{p}}(x^{(1)})$ с помощью этого интервального вектора, который локализует y_{p} не хуже, чем $x^{(0)}$. В итерационном методе (1) для вычисления новой локализации нужен ровно один шаг. Соединяя все это, мы окончательно получаем следующую итерационную процедуру:

$$(2)\quad \begin{cases} u^{(k+1)} = m(x^{(k)}) - \mathscr{D}_{\mathrm{p}}(x^{(k)})^{-1}\{\mathscr{B}_{\mathrm{p}}(x^{(k)})(m(x^{(k)}) \\ \qquad\qquad - x^{(k)}) + f_{\mathrm{p}}(m(x^{(k)}))\} \\ x^{(k+1)} = x^{(k)} \cap u^{(k+1)}, \quad k \geqslant 0. \end{cases}$$

Таким образом, данная система нелинейных уравнений преобразуется, как и в методе Ньютона, в систему линейных уравнений с интервальными коэффициентами. По этой линейной системе новая локализация для y_{p} вычисляется с помощью метода, похожего на полношаговый. Поэтому мы называем его полношаговым методом ньютоновского типа.

По аналогии с предыдущими рассуждениями мы устанавливаем, что (в обозначениях $u^{(k)} = (U_i^{(k)})$, $x^{(k)} = (X_i^{(k)})$) итерационный метод

$$(3)\quad \begin{cases} U_i^{(k+1)} = m(X_i^{(k)}) - \dfrac{1}{(\partial/\partial x_i) f_i(x^{(k)})}\left\{\sum_{j=1}^{i-1} \dfrac{\partial}{\partial x_j} f_i(x^{(k)})(m(X_j^{(k)}) - X_j^{(k+1)})\right. \\ \qquad\qquad \left. + \sum_{j=i+1}^{n} \dfrac{\partial}{\partial x_j} f_i(x^{(k)})(m(X_j^{(k)}) - X_j^{(k)}) + f_i(m(x^{(k)}))\right\}, \\ X_i^{(k+1)} = U_i^{(k+1)} \cap X_i^{(k)}, \quad 1 \leqslant i \leqslant n, \quad k \geqslant 0, \end{cases}$$

порождает векторы $x^{(k)}$, содержащие решение y_{p}. Мы назовем его короткошаговым методом ньютоновского типа со взятием покомпонентных пересечений.

Теперь мы исследуем для этих двух методов условия сходимости к решению $y_{\mathrm{p}} \in x^{(0)}$. Так как мы берем пересечение, каждый из этих методов порождает последовательность

$$x^{(0)} \supseteq x^{(1)} \supseteq \ldots \supseteq x^{(k)} \supseteq x^{(k+1)} \supseteq \ldots,$$

которая в силу следствия 10.8 сходится к некоторому интервальному вектору x. Далее мы имеем

$$y_{\mathrm{p}} \in x^{(k)}, \quad k \geqslant 0 \quad \text{и} \quad y_{\mathrm{p}} \in x.$$

Без дополнительных предположений мы имеем в общем случае равенства $\lim\limits_{k\to\infty} x^{(k)} = y_{\mathrm{p}}$. Следующая теорема дает достаточные условия равенства $d(x) = o_{\mathrm{p}}$, из которого ввиду $y_{\mathrm{p}} \in x$ следует $\lim\limits_{k\to\infty} x^{(k)} \leqslant y_{\mathrm{p}}$.

Теорема 1. *Пусть система нелинейных уравнений* $f_p(x_p) = o_p$ *имеет решение* y_p, *принадлежащее интервальному вектору* $x^{(0)}$. *Представим интервальную матрицу*

$$f'_p(x^{(0)}) = \left(\frac{\partial}{\partial x_j} f_i(x^{(0)})\right)$$

в виде

$$f'_p(x^{(0)}) = \mathcal{D}_p(x^{(0)}) - \mathcal{L}_p(x^{(0)}) - \mathcal{U}_p(x^{(0)})$$

и предположим, что

$$0 \notin \frac{\partial}{\partial x_i} f_i(x^{(0)}), \quad 1 \leqslant i \leqslant n.$$

Пусть $\mathcal{B}_p(x^{(0)}) = \mathcal{L}_p(x^{(0)}) + \mathcal{U}_p(x^{(0)})$. *Если выполнено условие*

$$\rho(|\mathcal{D}_p(x^{(0)})^{-1}||\mathcal{B}_p(x^{(0)})|) < 1,$$

где ρ *обозначает спектральный радиус, то последовательность* $\{x^{(k)}\}_{k=0}^{\infty}$ *для методов* (2) *и* (3) *сходится к решению* y_p.

Доказательство. Мы докажем это утверждение для полношагового метода (2). При $k \to \infty$ из формул (2) следует, что

$$u = m(x) - \mathcal{D}_p(x)^{-1}\{\mathcal{L}_p(x)(m(x) - x) + \mathcal{U}_p(x)(m(x) - x) + f_p(m(x))\},$$
$$x = x \cap u.$$

Из $x = x \cap u$ следует, что $x \subseteq u$, в частности

$$m(x) \subseteq m(x) - \mathcal{D}_p(x)^{-1}\{\mathcal{L}_p(x)(m(x) - x) + \mathcal{U}_p(x)(m(x) - x) + f_p(m(x))\},$$

т. е.

$$o_p \in \mathcal{D}_p(x)^{-1}\{\mathcal{L}_p(x)(m(x) - x) + \mathcal{U}_p(x)(m(x) - x) + f_p(m(x))\}.$$

Из того, что диагональные элементы матрицы $\mathcal{D}_p(x)^{-1}$ не содержат нуля, следует

$$o_p \in \mathcal{L}_p(x)(m(x) - x) + \mathcal{U}_p(x)(m(x) - x) + f_p(m(x)).$$

С помощью (2.19) получаем теперь, что

$$\begin{aligned} d(x) \leqslant d(u) &= |\mathcal{D}_p(x)^{-1}|(|\mathcal{L}_p(x)|d(x) + |\mathcal{U}_p(x)|d(x)) \\ &\leqslant |\mathcal{D}_p(x)^{-1}|(|\mathcal{L}_p(x)|d(u) + |\mathcal{U}_p(x)|d(u)) \\ &= |\mathcal{D}_p(x)^{-1}||\mathcal{B}_p(x)|d(u). \end{aligned}$$

Из того что $\rho(|\mathcal{D}_p(x)^{-1}||\mathcal{B}_p(x)|) \leqslant \rho(|\mathcal{D}_p(x^{(0)})^{-1}||\mathcal{B}_p(x^{(0)})|) < 1$, следует, что $d(u) = o_p$, т. е. $d(x) = o_p$. Тем самым доказано,

что последовательность $\{x^{(k)}\}_{k=0}^{\infty}$ сходится к y_{p}. Доказательство для короткошагового метода аналогично. ■

Из условий теоремы 1 следует, что система уравнений $\mathscr{F}_{\mathrm{p}}(x_{\mathrm{p}}) = o_{\mathrm{p}}$ не может иметь отличных от y_{p} решений, содержащихся в $x^{(0)}$.

Описанные выше методы можно модифицировать разными способами. Мы опишем эти модификации для полношагового метода ньютоновского типа. В случае когда вычисление $\mathscr{F}'_{\mathrm{p}}(x^{(k)})$ очень трудоемко, имеет смысл делать в методе (1) более одного шага вычисления локализации для y_{p}. Количество r шагов, которые выполняются после вычисления $\mathscr{F}'_{\mathrm{p}}(x^{(k)})$, тоже может зависеть от k: $r = r_k$. Это приводит к следующей итерации:

$$
(4)\quad \begin{cases} x^{(0,\,0)} = x^{(0)}, \\ u^{(k+1,\,i)} = m\left(x^{(k)}\right) - \mathscr{D}_{\mathrm{p}}\left(x^{(k)}\right)^{-1}\left\{\mathscr{B}_{\mathrm{p}}\left(x^{(k)}\right)\left(m\left(x^{(k)}\right)\right.\right. \\ \qquad\qquad \left.\left. - x^{(k,\,i-1)}\right) + \mathscr{F}_{\mathrm{p}}\left(m\left(x^{(k)}\right)\right)\right\}, \\ x^{(k,\,i)} = u^{(k+1,\,i)} \bigcap x^{(k,\,i-1)}, \quad 1 \leqslant i \leqslant r_k, \\ x^{(k+1)} = x^{(k,\,r_k)}, \\ x^{(k+1,\,0)} = x^{(k+1)}, \qquad\qquad k \geqslant 0. \end{cases}
$$

Аналогичным образом можно модифицировать и второй из описанных выше методов — короткошаговый метод (3) ньютоновского типа со взятием покомпонентных пересечений. Полученный метод сходится к y_{p} в условиях теоремы 1.

Модифицированные методы интересны еще и потому, что сходятся суперлинейно при подходящих условиях на последовательность $\{r_k\}_{k=0}^{\infty}$. Мы сформулируем и докажем это только для полношагового метода ньютоноскоого типа. Формулировка и доказательство следующего утверждения сохраняются и для короткошагового метода.

Теорема 2. *Пусть выполнены условия теоремы* 1. *Если модифицированный полношаговый метод* (4) *удовлетворяет условию* $r_k \to \infty$ *при* $k \to \infty$, *то*

$$
\| d\left(x^{(k+1)}\right)\| \leqslant c_k \| d\left(x^{(k)}\right)\|, \quad k \geqslant 0,
$$

причем $\lim\limits_{k\to\infty} c_k = 0$, *если для элементов матрицы* $\mathscr{F}'_{\mathrm{p}}(x^{(k)})$ *выполнено* (3.5′).

Доказательство. Перед тем как переходить к подробному доказательству, заметим, что в условиях теоремы 1 метод (4) сходится к y_{p}. Это можно показать так же, как в доказательстве теоремы 1.

Мы покажем теперь, что имеет место

$$d\left(x^{(k+1)}\right) = d\left(x^{(k,\,r_k)}\right) \leqslant \left(\mathscr{M}_{\mathrm{p}}^{r_k} + \mathscr{R}_{\mathrm{p}r_k}\right) d\left(x^{(k)}\right),$$

где

$$\mathscr{M}_{\mathrm{p}} = \left|\mathscr{D}_{\mathrm{p}}\left(x^{(0)}\right)^{-1}\right| \left|\mathscr{B}_{\mathrm{p}}\left(x^{(0)}\right)\right|$$

и где $\mathscr{R}_{\mathrm{p}r_k}$ — вещественная матрица, зависящая от $x^{(k)}$ и такая, что $\mathscr{R}_{\mathrm{p}r_k} \to \mathscr{O}_{\mathrm{p}}$ при $k \to \infty$. После этого утверждения теоремы сразу получатся из $\rho(\mathscr{M}_{\mathrm{p}}) < 1$. В доказательстве нашего неравенства мы воспользуемся тем, что из (3.5′) следует

$$d\left(\mathscr{D}_{\mathrm{p}}\left(x^{(k)}\right)^{-1} \mathscr{B}_{\mathrm{p}}\left(x^{(k)}\right)\right) \leqslant \mathscr{C}_{\mathrm{p}} d\left(x^{(k)}\right),$$

а также

$$d\left(\mathscr{D}_{\mathrm{p}}\left(x^{(k)}\right)^{-1}\right) \left|\mathscr{f}'_{\mathrm{p}}\left(x^{(k)}\right)\right| \leqslant d\left(\mathscr{D}_{\mathrm{p}}\left(x^{(k)}\right)^{-1} \left|\mathscr{f}'\left(x^{(0)}\right)\right|\right) \leqslant \mathscr{E}_{\mathrm{p}} d\left(x^{(k)}\right).$$

Здесь $\mathscr{C}_{\mathrm{p}}$ и $\mathscr{E}_{\mathrm{p}}$ — симметрические билинейные операторы, не зависящие от k.

Мы используем также включение $\mathscr{A}(\mathscr{B} c_{\mathrm{p}}) \subseteq (\mathscr{A}\mathscr{B})\, c_{\mathrm{p}}$ для произвольных интервальных матриц $\mathscr{A}$, $\mathscr{B}$ и точечного вектора c_{p}. (Его доказательство основано на определении 10.3 и законе дистрибутивности $(A + B)c = Ac + Bc$ для $A,\ B \in I(R),\ c \in \mathbb{R}$.)

Ввиду $x^{(k,\,0)} = x^{(k)}$ мы получаем из (5), что

$$\begin{aligned} d\left(x^{(k,\,1)}\right) &\leqslant d\left(u^{(k+1,\,1)}\right) \\ &\leqslant d\left(\mathscr{D}_{\mathrm{p}}\left(x^{(k)}\right)^{-1} \left\{\mathscr{B}_{\mathrm{p}}\left(x^{(k)}\right)\left(m\left(x^{(k)}\right) - x^{(k,\,0)}\right)\right\}\right) \\ &\quad + d\left(\mathscr{D}_{\mathrm{p}}\left(x^{(k)}\right)^{-1} \mathscr{f}_{\mathrm{p}}\left(m\left(x^{(k)}\right)\right)\right) \\ &< \left(\mathscr{M}_{\mathrm{p}} + \mathscr{R}_{\mathrm{p}1}\right) d\left(x^{(k)}\right), \end{aligned}$$

где

$$\mathscr{R}_{\mathrm{p}1} = d\left(\mathscr{D}_{\mathrm{p}}\left(x^{(k)}\right)^{-1}\right) \left|\mathscr{f}'_{\mathrm{p}}\left(x^{(k)}\right)\right|$$

и $\mathscr{R}_{\mathrm{p}1} \to \mathscr{O}_{\mathrm{p}}$ при $k \to \infty$.

Допустим теперь, что для некоторого $i \geqslant 1$ мы имеем

$$d\left(x^{(k,\,i)}\right) \leqslant \left(\mathscr{M}_{\mathrm{p}}^{i} + \mathscr{R}_{\mathrm{p}i}\right) d\left(x^{(k)}\right)$$

и $\mathscr{R}_{\mathrm{p}i} \to \mathscr{O}_{\mathrm{p}}$ при $k \to \infty$. Тогда получаем из (5), что

$$\begin{aligned} d\left(x^{(k,\,i+1)}\right) &\leqslant d\left(u^{(k+1,\,i+1)}\right) \\ &\leqslant d\left(\mathscr{D}_{\mathrm{p}}\left(x^{(k)}\right)^{-1} \left\{\mathscr{B}_{\mathrm{p}}\left(x^{(k)}\right)\left(m\left(x^{(k)}\right) - m\left(x^{(k,\,i)}\right)\right.\right.\right. \\ &\quad \left.\left.\left. + m\left(x^{(k,\,i)}\right) - x^{(k,\,i)}\right)\right\}\right) + d\left(\mathscr{D}_{\mathrm{p}}\left(x^{(k)}\right)^{-1} \left|\mathscr{f}'_{\mathrm{p}}\left(x^{(k)}\right)\right| \left| m\left(x^{(k)}\right)\right.\right. \\ &\quad \left.\left. - m\left(x^{(k,\,i)}\right) + m\left(x^{(k,\,i)}\right) - y_{\mathrm{p}}\right|\right) \\ &\leqslant \mathscr{C}_{\mathrm{p}} d\left(x^{(k)}\right) \left| m\left(x^{(k)}\right) - m\left(x^{(k,\,i)}\right)\right| \\ &\quad + \left|\mathscr{D}_{\mathrm{p}}\left(x^{(k)}\right)^{-1}\right| \left|\mathscr{B}_{\mathrm{p}}\left(x^{(k)}\right)\right| d\left(x^{(k,\,i)}\right) \end{aligned}$$

$$+ \mathscr{E}_{\mathrm{p}} d(x^k) | m(x^{(k)}) - m(x^{(k,\,i)}) | + \mathscr{R}_{\mathrm{p}1} d(x^{(k,\,i)})$$
$$\leqslant ((\mathscr{E}_{\mathrm{p}} + \mathscr{C}_{\mathrm{p}}) | m(x^{(k)}) - m(x^{(k,\,i)}) |$$
$$+ \mathscr{M}_{\mathrm{p}} (\mathscr{M}_{\mathrm{p}}^{i} + \mathscr{R}_{\mathrm{p}i})) d(x^{(k)}) + \mathscr{R}_{\mathrm{p}1} (\mathscr{M}_{\mathrm{p}}^{i} + \mathscr{R}_{\mathrm{p}i}) d(x^{(k)})$$
$$= (\mathscr{M}_{\mathrm{p}}^{i+1} + \mathscr{R}_{\mathrm{p}i+1}) d(x^{(k)}),$$

где

$$\mathscr{R}_{\mathrm{p}i+1} = (\mathscr{E}_{\mathrm{p}} + \mathscr{C}_{\mathrm{p}}) | m(x^{(k)}) - m(x^{(k,\,i)}) | + \mathscr{M}_{\mathrm{p}} \mathscr{R}_{\mathrm{p}i} + \mathscr{R}_{\mathrm{p}1} (\mathscr{M}_{\mathrm{p}}^{i} + \mathscr{R}_{\mathrm{p}i}).$$

Ввиду того что

$$| m(x^{(k)}) - m(x^{(k,\,i)}) | \to o_{\mathrm{p}} \quad \text{при} \quad k \to \infty,$$
$$\mathscr{R}_{\mathrm{p}1} \to O_{\mathrm{p}} \quad \text{при} \quad k \to \infty$$

и

$$\mathscr{R}_{\mathrm{p}i} \to O_{\mathrm{p}} \quad \text{при} \quad k \to \infty,$$

мы получаем, что

$$\mathscr{R}_{\mathrm{p}i+1} \to O_{\mathrm{p}} \quad \text{при} \quad k \to \infty.$$

Отсюда можно вывести утверждение теоремы. ■

Теперь мы исследуем один класс систем нелинейных уравнений, к которому можно применить описанные методы.

Ищется решение нелинейной краевой задачи

$$\Delta u \equiv u_{ss} + u_{tt} = f(s,\, t,\, u), \quad (s,\, t) \in \Omega$$

при $u(s, t) = \varphi(s, t)$ для $(s, t) \in \dot{\Omega}$. Здесь Ω — односвязная ограниченная открытая область плокости (s, t), а $\dot{\Omega}$ обозначает границу области Ω. Мы предположим для простоты, что $\Omega = (0, 1) \times (0, 1)$. Если $f\colon \Omega \times \mathbb{R} \to \mathbb{R}$ непрерывно дифференцируема и $f_u(s, t, u) \geqslant 0$, $(s, t) \in \Omega$, $u \in \mathbb{R}$, то при относительно слабых условиях на φ рассматриваемая краевая задача имеет единственное решение. Для нахождения численных приближений к этому решению мы можем, например, преобразовать краевую задачу в систему нелинейных уравнений с помощью разностного метода. Если выбран один и тот же шаг по s и по t, то замена частных производных подходящими разностными отношениями приводит к системе из n^2 нелинейных уравнений

$$4x_{ij} - x_{i-1,\,j} - x_{i+1,\,j} - x_{i,\,j+1} - x_{i,\,j-1} + h^2 f(ih,\, jh,\, x_{ij}) = 0,$$
$$1 \leqslant i,\ j \leqslant n.$$

Здесь $h = 1/(n+1)$ и x_{ij} — приближенное значение величины $u(ih, jh)$. Для $i = 0, n+1$ и $j = 0, n+1$ значения x_{ij} определяются функцией $\varphi(s, t)$.

Наши n^2 уравнений можно записать в виде

$$\mathscr{f}_{\mathrm{p}}(x_{\mathrm{p}}) = \mathscr{A}_{\mathrm{p}} x_{\mathrm{p}} + g_{\mathrm{p}}(x_{\mathrm{p}}) - b_{\mathrm{p}} = o_{\mathrm{p}}, \tag{5}$$

где $x_p = (x_{ij}) \in V_{n^2}(\mathbb{R})$. Хорошо известно, что $\mathscr{A}_p$ — блочно треугольная матрица и $g_{ij}(x_p) = g_{ij}(x_{ij})$. В предположении $f_u(s, t, u) \geqslant 0$ нетрудно показать, что условия теоремы 1 выполнены независимо от ширины $d(x^{(0)})$ данной локализации $x^{(0)}$ единственного решения системы (5).

В качестве конкретного примера мы рассмотрим уравнение

$$u_{ss} + u_{tt} = u^3/(s^2 + t^2 + 1), \quad (s, t) \in (0, 1) \times (0, 1).$$

$$u(s, 0) = 1, \qquad s \in [0, 1],$$
$$u(0, t) = 1, \qquad t \in [0, 1],$$
$$u(1, t) = 2 - e^t, \quad t \in [0, 1],$$
$$u(s, 1) = 2 - e^s, \quad s \in [0, 1]$$

при $n = 5$. Легко убедиться, что векторы

$$y_p = (y_{ij}), \quad z_p = (z_{ij}) \in V_{n^2}(\mathbb{R}),$$

такие что $y_{ij} = -1$, $z_{ij} = 2$, $1 \leqslant i, j \leqslant n$, удовлетворяют неравенству

$$\mathscr{f}_p(y_p) \leqslant o_p \leqslant \mathscr{f}_p(z_p)$$

для функции $\mathscr{f}_p$ из (5). Так как $\mathscr{f}_p$ является M-функцией, отсюда следует, что интервальный вектор $x^{(0)} = (X_{ij}^{(0)}) \in V_{n^2}(I(\mathbb{R}))$ для $X_{ij}^{(0)} = [-1, 2]$ содержит решение системы (5).

Мы приведем теперь результаты вычисления приближений к значению $u\left(\frac{1}{2}, \frac{1}{2}\right)$ с помощью такой же модификации короткошагового метода ньютоновского типа со взятием покомпонентных пересечений, как модификация полношагового метода, рассмотренная в теореме 2.

Мы выбираем $r_1 = 1$, $r_{k+1} = r_k + 1$, $k \geqslant 1$ и приводим результаты в табл. 1.

Итерация прекращалась, как только половинная ширина всех локализующих интервалов становилась меньше, чем 10^{-10}. С другой стороны, методу (3) потребовался 81 шаг для достижения той же точности.

Простые соображения показывают, что в этом примере количество арифметических операций на один шаг итерации примерно одно и то же. Однако в методе (3) потребовалось 81 раз производить вычисление функции и интервальное вычисление производной по сравнению с 13 вычислениями для модифицированного метода. Это отразилось и на времени вычисления для нашего примера; для метода (3) оно оказалось в 6—7 раз больше.

Замечания. Приведенные здесь методы тесно связаны с соответствующими методами из [242]. Там всегда предполагалась выпуклость, чтобы можно было вычислить монотонные последо-

Таблица 1

k	*Нижняя граница*	*Верхняя граница*
0	[− 1	2]
1	[− 0.8683495081969,	1.932431741809]
2	[− 0.4334368762048,	1.662891859166]
3	[0.1747052174969,	1.110238786145]
4	[0.4964693372903,	0.7889353835863]
5	[0.6094825964828,	0.6756608548531]
6	[*0.6370682656242*,	*0.6480110082876*]
7	[0.6418668139331,	0.6432014513712]
8	[0.6424732262048,	0.6425936520955]
9	[0.6425293531665,	0.6425373968087]
10	[0.6425331717500,	0.6425335694949]
11	[0.6425333631204,	0.6425333776870]
12	[0.6425333701936,	0.6425333705975]
13	[0.6425333703873,	0.6425333704038]

вательности, локализующие решение. Простые примеры показывают, что выпуклость необходима. Результаты, приведенные в этой главе, были доказаны в [10]. В этой же статье имеются *аналогичные утверждения для релаксационных* методов ньютоновского типа.

Если принять для рассматриваемой системы такие же допущения, как в гл. 22, то и здесь можно отказаться от явного использования вычислений в интервальной арифметике при исполнении итерационной процедуры. В этом случае вычисления в интервальной арифметике используются только для вычисления матрицы $\mathscr{B}_{\mathrm{p}}\left(x_{\mathrm{p}}^{(k)}, y_{\mathrm{p}}^{(k)}\right)$ (см. гл. 21). Мы не будем рассматривать этот вопрос подробнее.

Если на каждом шаге производить итерацию (1) полностью, то получится последовательность локализаций для y_{p}, которая, как можно показать, сходится к y_{p} в условиях теоремы 1. При этом ширина сходится к нулю квадратично. Таким образом, для *систем, удовлетворяющих условиям теоремы 1, этот метод дает* большую экономию вычислений по сравнению с системами, для которых приходится использовать метод, исследованный в теореме 2 гл. 19. Ведь там приходится на каждом шаге обращать точечную матрицу и производить итеративную локализацию матрицы, обратной к некоторой интервальной матрице.

В [383а] рассмотрена модификация метода (3), в которой вектор $m(x^{(k)})$ заменен на произвольный точечный вектор, принадлежащий $x^{(k)}$. Выбирая этот точечный вектор подходящим образом, можно получить методы, сходящиеся гораздо быстрее,

чем в случае, когда в качестве $m(x^{(k)})$ берется центр интервала $x^{(k)}$, вычисленного по методу (3). В [514a] введен и исследован ряд методов, аналогичных (2) и (3). Введен также интервально-арифметический вариант хорошо известного алгоритма Бинемана, который затем применен к системам нелинейных уравнений, возникающих при решении некоторых нелинейных краевых задач. Численные результаты производят большое впечатление.

* * *

Литература: [10], [242], [383a], [514a].

Приложение А
ПОРЯДОК СХОДИМОСТИ ИТЕРАЦИОННЫХ МЕТОДОВ в $V_n(I(\mathbb{C}))$ и $M_{mn}(I(\mathbb{C}))$

Пусть I — итерационный метод, порождающий последовательности $\{x^{(k)}\}$, членами которых являются комплексные интервальные векторы $x^{(k)}$ из $V_n(I(\mathbb{C}))$, $k \geqslant 0$. Кроме того, предположим, что каждая такая последовательность сходится к пределу x^*, принадлежащему $V_n(I(\mathbb{C}))$. Поскольку большинство реальных итерационных методов используется для локализации конкретных решений, мы будем предполагать, что

$$x^* \subseteq x^{(k)}, \quad k \geqslant 0. \tag{1}$$

Имеются два способа, позволяющих измерить отклонение элемента $x^{(k)}$ от предела x^*:

$$e_{\mathrm{p}}^{(k)} = q(x^{(k)}, x^*) \tag{2}$$

и

$$h_{\mathrm{p}}^{(k)} = d(x^{(k)}) - d(x^*). \tag{3}$$

Обе величины представляют собой неотрицательные вещественные векторы $(e_{\mathrm{p}}^{(k)}, h_{\mathrm{p}}^{(k)} \geqslant o_{\mathrm{p}})$, удовлетворяющие соотношению

$$e_{\mathrm{p}}^{(k)} = o_{\mathrm{p}} \Leftrightarrow h_{\mathrm{p}}^{(k)} = o_{\mathrm{p}} \Leftrightarrow x^{(k)} = x^*. \tag{4}$$

Кроме того, $e_{\mathrm{p}}^{(k)}$ и $h_{\mathrm{p}}^{(k)}$ являются монотонными по включению отображениями:

$$x^{(m)} \subseteq x^{(k)} \Rightarrow e_{\mathrm{p}}^{(m)} \leqslant e_{\mathrm{p}}^{(k)}, \quad h_{\mathrm{p}}^{(m)} \leqslant h_{\mathrm{p}}^{(k)}. \tag{5}$$

Справедливость (5) непосредственно вытекает из следующего свойства метрики q:

$$A \subseteq B \subseteq C \Rightarrow q(B, A) \leqslant q(C, A), \quad A, B, C \in I(\mathbb{C}),$$

а также из свойства (2.9), истинность которого на $I(\mathbb{C})$ очевидна. Таким образом, меньшая локализация x^* означает и меньшую величину отклонения от x^*.

Часто проще бывает выразить отклонение $x^{(k)}$ от x^* через неотрицательное вещественное число. Для этого можно использовать монотонную векторную норму $\|\cdot\|$, подставив в нее $e_{\mathrm{p}}^{(k)}$ (соответственно $h_{\mathrm{p}}^{(k)}$):

$$e^{(k)} = \| q(x^{(k)}, x^*) \|$$

(соответственно $h^{(k)} = \| d(x^{(k)}) - d(x^*) \|$).

Снова имеем

$$e^{(k)} = 0 \Leftrightarrow h^{(k)} = 0 \Leftrightarrow x^{(k)} = x^*,$$

а также свойство монотонности

$$x^{(m)} \subseteq x^{(k)} \Rightarrow e^{(m)} \leqslant e^{(k)}, \quad h^{(m)} \leqslant h^{(k)}.$$

Если $X = [x^*, x^*]$, то в интервальном пространстве $I(\mathbb{R})$ введенные нами величины получают простую интерпретацию. В этом случае

$$h^{(k)} = d(X^{(k)})$$

представляет собой расстояние между включающими границами x^*, а

$$e^{(k)} = q(X^{(k)}, x^*) = \max \{|x_1^{(k)} - x^*|, |x_2^{(k)} - x^*|\}$$

— максимальное отклонение элемента x из $X^{(k)}$ от x^*.

Теперь предположим, что последовательности $\{x^{(k)}\}$, порожденные итерационной процедурой I, сходятся к x^* и удовлетворяют (1). Тогда соответствующие последовательности $\{e_{\mathrm{p}}^{(k)}\}$ и $\{h_{\mathrm{p}}^{(k)}\}$ сходятся к нулю. К ним можно применить хорошо известное по работе Ортеги и Рейнбольдта [242] понятие порядка сходимости.

Предварительно определим так называемый R-фактор:

$$R_t\{u_{\mathrm{p}}^{(k)}\} = \begin{cases} \lim\limits_{k\to\infty} \sup \| u_{\mathrm{p}}^{(k)} \|^{1/k}, & t = 1, \\ \lim\limits_{k\to\infty} \sup \| u_{\mathrm{p}}^{(k)} \|^{1/t^k}, & t > 1, \end{cases}$$

где $\{u_{\mathrm{p}}^{(k)}\}$ обозначает одну из последовательностей $\{e_{\mathrm{p}}^{(k)}\}$ или $\{h_{\mathrm{p}}^{(k)}\}$. Ортега и Рейнбольдт, опираясь на теорему об эквивалентности норм, доказали [242], что значение R_t не зависит от того, какая норма используется в качестве $\|\cdot\|$.

Справедлива также следующая лемма.

Лемма 1. *Пусть $\{x^{(k)}\}$ — последовательность, порожденная процедурой I. Предположим, что эта последовательность сходится*

к x^* и обладает свойством (1). Тогда для соответствующих последовательностей $\{e_p^{(k)}\}$ и $\{h_p^{(k)}\}$ выполнено равенство

$$R_t\{e_p^{(k)}\} = R_t\{h_p^{(k)}\}, \quad t \geqslant 1.$$

Доказательство. Исходя из (10.27) для $k \geqslant 0$, получаем

$$\frac{1}{2} h_p^{(k)} \leqslant e_p^{(k)} \leqslant h_p^{(k)}.$$

При использовании монотонной векторной нормы имеем

$$\frac{1}{2}\|h_p^{(k)}\| \leqslant \|e_p^{(k)}\| \leqslant \|h_p^{(k)}\|.$$

Так как очевидно, что

$$\lim_{k\to\infty}\left(\frac{1}{2}\right)^{1/t^k} = 1, \quad t > 1$$

то

$$\limsup_{k\to\infty}\|h_p^{(k)}\|^{1/t^k} \geqslant \limsup_{k\to\infty}\|e_p^{(k)}\|^{1/t^k}$$

$$\geqslant \limsup_{k\to\infty}\left(\frac{1}{2}\right)^{1/t^k}\|h_p^{(k)}\|^{1/t^k}$$

$$= \limsup_{k\to\infty}\|h_p^{(k)}\|^{1/t^k}.$$

Поскольку доказательство для $t = 1$ можно провести аналогичным образом, оно здесь опущено. ■

Из леммы 1 следует, что в дальнейших рассуждениях можно не указывать, какая именно из последовательностей — $\{e_p^{(k)}\}$ или $\{h_p^{(k)}\}$ — имеется в виду. Это не так, если (1) не выполнено. Проиллюстрируем сказанное на примере.

Пример. Последовательность

$$\left\{[0, 1] + \left(\frac{1}{2}\right)^k\right\}_{k=0}^{\infty}$$

имеет предел $[0, 1]$. Очевидно, что

$$e^{(k)} = \left(\frac{1}{2}\right)^k, \quad k \geqslant 0,$$

и, следовательно, при $t = 1$

$$R_1\{e^{(k)}\} = \lim_{k\to\infty}\sup\left(\left(\frac{1}{2}\right)^k\right)^{1/k} = \frac{1}{2}.$$

С другой стороны, $h^{(k)} = 0$ при всех $k \geqslant 0$, откуда следует, что

$$R_1\{h^{(k)}\} = 0. \blacksquare$$

Обозначим через $\mathfrak{C}(I, x^*)$ множество порожденных итерационной процедурой I последовательностей $\{x^{(k)}\}$, для которых справедливо

$$\lim_{k\to\infty} x^{(k)} = x^* \text{ и } x^* \subseteq x^{(k)}, \quad k \geqslant 0.$$

Под R-фактором процедуры I будем понимать

$$R_t(I, x^*) = \sup\{R_t\{e_{\mathrm{p}}^{(k)}\} \mid \{x^{(k)}\} \in \mathfrak{C}(I, x^*)\}, \quad t \geqslant 1.$$

Из доказанных выше утверждений следует, что этот R-фактор не зависит от того, какая именно норма была использована в его определении; кроме того, с равным успехом он мог быть определен через последовательность $\{h_{\mathrm{p}}^{(k)}\}$.

Теперь мы можем определить R-порядок итерационной процедуры I как

$$O_R(I, x^*) = \begin{cases} +\infty, \text{ если } R_t(I, x^*) = 0 \text{ и } t \geqslant 1, \\ \inf\{t \mid t \in [1, \infty),\ R_t(I, x^*) = 1\} \\ \qquad\qquad \text{в остальных случаях.} \end{cases} \tag{6}$$

Это определение применительно к последовательностям векторов из $V_n(\mathbb{R})$ предложили Ортега и Рейнбольдт [242].

А сейчас мы хотим привести несколько полезных утверждений, касающихся вычисления (и оценивания) R-порядка (ср. у Ортеги и Рейнбольдта [242, с. 297]).

Теорема 2. *Пусть I — итерационная процедура с пределом x^*, и пусть $\mathfrak{C}(I, x^*)$ — множество всех порожденных процедурой I последовательностей, для которых* $\lim_{k\to\infty} x^{(k)} = x^*$ *и* $x^* \subseteq x^{(k)}$, *$k \geqslant 0$. Если существует $t \geqslant 1$ и константа γ_2, такая что для всех $\{x^{(k)}\}$ из $\mathfrak{C}(I, x^*)$ и нормы $\|\cdot\|$ выполнено*

$$\|e_{\mathrm{p}}^{(k+1)}\| \leqslant \gamma_2 \|e_{\mathrm{p}}^{(k)}\|^t, \quad k \geqslant k(\{x^{(k)}\}),$$

либо

$$\|h_{\mathrm{p}}^{(k+1)}\| \leqslant \gamma_2 \|h_{\mathrm{p}}^{(k)}\|^t, \quad k \geqslant k(\{x^{(k)}\}),$$

то R-порядок процедуры I удовлетворяет неравенству

$$O_R(I, x^*) \geqslant t.$$

С другой стороны, если существуют положительная константа γ_1 и последовательность $\{x^{(k)}\}$ из $\mathfrak{C}(I, x^*)$, такие что

$$\|e_{\mathrm{p}}^{(k+1)}\| \geqslant \gamma_1 \|e_{\mathrm{p}}^{(k)}\|^t > 0, \quad k \geqslant k_0,$$

либо

$$\|h_{\mathrm{p}}^{(k+1)}\| \geqslant \gamma_1 \|h_{\mathrm{p}}^{(k)}\|^t > 0, \quad k \geqslant k_0,$$

то R-порядок процедуры I удовлетворяет неравенству

$$O_R(I, x^*) \leqslant t.$$

Если с некоторыми константами γ_1 и γ_2 выполнены оба указанных условия, то

$$O_R(I, x^*) = t. \blacksquare$$

Доказательство этой теоремы строится аналогично соответствующему доказательству у Ортеги и Рейнбольдта [242].

Теорема 3. *Пусть I — итерационная процедура с пределом x^*, и пусть $\mathfrak{C}(I, x^*)$ — множество всех порожденных I последовательностей, для которых $\lim\limits_{k\to\infty} x^{(k)} = x^*$ и $x^* \subseteq x^{(k)}$, $k \geqslant 0$. Пусть $r = \sum\limits_{i=0}^{n} m_i > 0$, где все m_i — целые неотрицательные числа. Тогда, если существует положительная константа γ, такая что для всех $\{x^{(k)}\}$ из $\mathfrak{C}(I, x^*)$ выполнено одно из неравенств*

$$\| e_{\mathrm{p}}^{(k+1)} \| \leqslant \gamma \prod_{i=0}^{n} \| e_{\mathrm{p}}^{(k-i)} \|^{m_i}, \quad k \geqslant k(\{x^{(k)}\}),$$

либо

$$\| h_{\mathrm{p}}^{(k+1)} \| \leqslant \gamma \prod_{i=0}^{n} \| h_{\mathrm{p}}^{(k-i)} \|^{m_i}, \quad k \geqslant k(\{x^{(k)}\}),$$

то R-порядок процедуры I удовлетворяет неравенству

$$O_R(I, x^*) \geqslant s,$$

где s — единственное положительное решение уравнения

$$s^{n+1} - \sum_{i=0}^{n} m_i s^{n-i} = 0. \blacksquare$$

Доказательство теоремы 3 может быть основано на идеях, высказанных Шмидтом в [507, разд. 4].

Часто оказывается возможным проверить какое-либо из приведенных в теоремах 2 и 3 неравенств и тем самым получить оценку R-порядка на данной итерации.

Сделанные в этом приложении утверждения могут быть перенесены на случай $M_{mn}(I(\mathbb{C}))$. В частности, останутся справедливыми лемма 1 и теорема 2 и 3.

* * *

Литература: [349с], [242], [507].

Приложение В

РЕАЛИЗАЦИЯ МАШИННОЙ ИНТЕРВАЛЬНОЙ АРИФМЕТИКИ НА АЛГОЛЕ 60

Теперь мы хотим дополнить материал гл. 4, более внимательно рассмотрев вопросы реализации машинной интервальной арифметики. Наше изложение будет в основном следовать работе Криста [72]; из прочей литературы использованы результаты Дьюара [77], Гуда и Лондона [98].

Формулы (1.2) и определение 4.1 задают рамки, в которых реализуются машинные интервальные операции:

логическую часть, определяющую порядок вычисления границ результата по формулам (1.2),

и

арифметическую часть, в которой пара границ результирующего интервала вычисляется при помощи арифметических операций с направленными округлениями (см., например, формулы (4.8) и (4.9)).

Первая часть может быть без труда описана на Алголе 60 и не нуждается поэтому в более подробном объяснении. Вторая часть, однако, требует бóльших усилий, поскольку в реализациях Алгола 60 направленные округления обычно не доступны. Поэтому сейчас мы займемся моделированием машинной арифметики, целиком основанной на использовании направленного округления вниз ($\downarrow$). Соответствующая процедура LOW будет определена через обычные машинные операции. При употреблении знака * для обозначения машинных операций предполагается выполнение установленных в гл. 4 соглашений. В частности, считается, что

(а) множество машинных чисел $\mathbb{R}_м$ симметрично относительно нуля и состоит из нормализованных чисел с плавающей точкой, т. е. имеющих вид

$$x = mb^e,$$

где m — мантисса, b — основание степени, e — порядок;

(б) все арифметические операции с плавающей точкой оптимальны, т. е. результат вещественной операции округляется с помощью отображения fl, обладающего свойством (4.3) (см. (4.9)).

Как было указано в гл. 4, направленное округление вверх $\uparrow$ получается из $\downarrow$ по формуле (4.5). Эта формула сохраняет силу и для машинных операций:

$$\uparrow(x * y) = \begin{cases} -\downarrow((-x)*(-y)), & \text{где } * \in \{+, -\}, \\ -\downarrow((-x)*y), & \text{где } * \in \{\cdot, /\}. \end{cases}$$

Здесь ↓ обозначает результат выполнения процедуры LOW, а * — машинную операцию.

Запишем теперь процедуру LOW на Алголе 60. Эта процедура позволяет вычислить нижнюю границу вещественного результата операции над числами с плавающей точкой. Ее заголовок выглядит так:

'REAL''PROCEDURE'LOW(X);

'VALUE'X; 'REAL'X;

Каким будет тело процедуры, зависит от некоторых особенностей машинной арифметики. К их числу относятся

(I) тип округления *fl*, используемого для перехода от вещественных операций к операциям с плавающей точкой. Мы различаем

(1) округление к ближайшему со стороны нуля машинному числу, получаемое отсечением младших разрядов мантиссы, и

(2) округление к ближайшему машинному числу;

(II) поведение вблизи нуля. Если результат вещественной операции оказывается меньше по абсолютной величине, чем минимальное положительное машинное число, то мы различаем

(3) установление результата в нуль и

(4) установление результата равным минимальному по абсолютной величине машинному числу со знаком, совпадающим со знаком вещественного результата;

(III) поведение в случае выхода за диапазон машинных чисел. Осмысленное расширение машинных интервальных операций возможно лишь в исключительных случаях[1]). Поэтому в подобной ситуации мы будем всегда останавливать вычисления. В теле процедуры LOW можно применить следующую конструкцию, вызывающую останов:

'IF'X'NOTGREATER' ⟨lowerbound⟩ 'THEN'

'BEGIN' X := 0; X := 1/X 'END'

Здесь ⟨lowerbound⟩ обозначает наименьшее число с плавающей точкой, т. е. $\min\{x \mid x \in \mathbb{R}_M\}$.

Пусть k — количество цифр в мантиссе, основание степени b равно 2 или 10. В тексте процедуры LOW будут использоваться

---

[1]) С этим тезисом авторов трудно согласиться. В работе [143] содержатся носящие достаточно универсальный характер идеи, касающиеся способов продолжения вычислений после выхода за допустимый диапазон. Впоследствии эти идеи получили значительное развитие Подробнее об этом см. п. 1.2.2 Добавления. — *Прим. перев.*

следующие обозначения:

$$\langle \text{minpos} \rangle := \lim \{x \mid x \in \mathbb{R}_M,\ x > 0\},$$

$$\langle \text{factor } A \rangle := 1 - b^{1-k},$$

$$\langle \text{factor } B \rangle := 1 - \frac{1}{2} b^{1-k}.$$

Две последние константы на двоичной машине должны быть порождены как частное от деления двух целых чисел:

$$\langle \text{factor } A \rangle := (2^{k-1} - 1)/2^{k-1},$$

$$\langle \text{factor } B \rangle := (2^k - 1)/2^k.$$

Пункты (I) и (II) дают четыре возможные комбинации их подпунктов: (1)—(3), (1)—(4), (2)—(3) и (2)—(4). Поэтому мы приводим четыре варианта для тела процедуры LOW.

У Криста [72] можно найти доказательство того, что приведенные тела процедур действительно реализуют округление $\downarrow$ в случае, когда X получен в результате выполнения обычной машинной операции.

Теперь, используя процедуру-функцию LOW и определяемую через нее машинную арифметику с округлениями, мы можем без труда реализовать машинные интервальные операции. Интервал $A = [a_1, a_2]$ представляется в виде 'ARRAY'A, где $A[1]$ равно a_1, $A[2]$ равно a_2. Заголовки процедур содержат как описание интервалов-операндов A и B, так и описание результирующего интервала C.

(Текст вышеупомянутых процедур приведен на стр. 294—295.)

А теперь вкратце обсудим, как осуществить ввод интервалов. Трудность здесь состоит в том, что при чтении десятичных дробных чисел может появиться погрешность при их преобразовании к внутреннему двоичному представлению. Величина этой погрешности, вообще говоря, неизвестна. Поэтому мы рекомендуем до ввода умножить границы на величины 10^n с тем, чтобы сделать их целыми числами. Эти целые числа считаются без погрешности. После ввода считанные границы делятся на 10^n, а применение процедуры LOW гарантирует корректное округление. Ниже (стр. 296) следует текст процедуры ввода, в котором $\langle 10^n \rangle$ обозначает выбранную соответствующим образом степень десяти.

Арифметика (1), (3)

```
'BEGIN'
      'IF' X 'NOTGREATER' ∅ 'THEN'
         'BEGIN'
            X:= X/⟨        A⟩;
            'IF' X 'GREATER' –⟨minpos ⟩ 'THEN' X:= – ⟨minpos
         'END';
      LOW:= X
'END'
```

Арифметика(1), (4)

```
'BEGIN'
      'IF' X 'NOTLESS' ∅ 'THEN'
         'BEGIN'
            'IF' X 'NOTGREATER' ⟨minpos ⟩ 'THEN' X:= ∅
         'END'
      'ELSE'
         'BEGIN'
            X:= X/⟨factor A⟩;
            'IF' X 'GREATER' – ⟨minpos ⟩ 'THEN' X:= – ⟨minpos ⟩
         'END';
      LOW:= X
'END'
```

Арифметика(2), (3)

```
'BEGIN'
      'IF' X 'GREATER' ∅ 'THEN'
         'BEGIN'
            X:= X × ⟨factor B⟩;
            'IF' X 'NOTGREATER' ⟨minpos ⟩ 'THEN' X:= ∅
         'END'
      'ELSE'
         'BEGIN'
            X:= X/⟨factor B⟩;
            'IF' X 'GREATER' – ⟨minpos⟩ 'THEN' X:= – ⟨minpos⟩
         'END';
      LOW:= X
'END'
```

Арифметика(2), (4)

```
'BEGIN'
      'IF' X 'NOTLESS' ∅ 'THEN'
         'BEGIN'
            X:= X × ⟨factor B⟩;
            'IF' X 'NOTGREATER' ⟨minpos⟩ 'THEN' X:= ∅
         'END'
      'ELSE'
         'BEGIN'
            X:= X/⟨factor B⟩;
            'IF' X 'GREATER' – ⟨minpos ⟩ 'THEN' X:= – ⟨minpos⟩
         'END';
      LOW:= X
'END'
```

Сложение $C = A + B$:

```
'PROCEDURE'ADD(A, B, C); 'ARRAY'A, B, C;
    'BEGIN'C[1] := LOW(A[1] + B[1]);
           C[2] := – LOW(– A[2] – B[2]) 'END'
```

Вычитание $C = A - B$:

```
'PROCEDURE'SUB(A, B, C); 'ARRAY'A. B, C;
    'BEGIN'C[1] := LOW(A[1] – B[2]);
           C[2] := – LOW(B[1] – A[2]) 'END'
```

Умножение $C = A \times B$:

```
'PROCEDURE'MUL(A, B, C); 'ARRAY'A, B, C;
    'BEGIN' 'REAL'A1, A2, B1, B2, C1, C2, P; 'BOOLEAN'BA, BB;
    A2 := – A[2]; B2 := – B[2];
    BA := A2 'LESS' ∅; BB := B2 'LESS' ∅;
    'IF'BA  'THEN'A1 := A[1]
            'ELSE' 'BEGIN'A1 := A2; A2 := A[1] 'END';
    'IF'BB  'THEN'B1 := B[1]
            'ELSE' 'BEGIN'B1 := B2; B2 := B[1] 'END';
    C2 := LOW(– A2 × B2);
    'IF'B1  'LESS' ∅
    'THEN' 'BEGIN'C1 := LOW(– A2 × B1);
                'IF'A1  'LESS' ∅
                'THEN' 'BEGIN'P := LOW(– A1 × B2);
                              'IF'P 'LESS'C1 'THEN'C1 := P;
                              P := LOW(–A1 × B1);
                              'IF'P 'LESS'C2 'THEN'C2 := P
                    'END'
        'END'
'ELSE'C1 := 'IF'A1 'LESS' ∅ 'THEN' LOW(– A1 × B2)
                             'ELSE' LOW(A1 × B1);
'IF'BA 'EQUIV'BB
'THEN' 'BEGIN'C[1] := C1; C[2] := – C2 'END'
'ELSE' 'BEGIN'C[1] := C2; C[2] := – C1 'END'
    'END'
```

Деление $C = A/B$:

```
'PROCEDURE'DIV(A, B, C); 'ARRAY'A, B, C;
    'BEGIN' 'REAL'A1, A2, B1, B2, C1, C2; 'BOOLEAN'BA, BB;
    A2 := – A[2]; B2 := – B[2];
    BA := A2 'LESS' ∅; BB := B2 'LESS' ∅;
    'IF'BA  'THEN'A1 := A[1]
            'ELSE' 'BEGIN'A1 := A2; A2 := A[1] 'END';
    'IF'BB 'THEN'B1 := B[1]
            'ELSE' 'BEGIN'B1 := B2; B2 := B[1] 'END';

    'IF'B1'NOTGREATER' ∅
    'THEN' 'BEGIN' 'COMMENT'ОСТАНОВ ПРИ ДЕЛЕНИИ;
                           C1 := ∅; C1 := 1/C1 'END'
    'ELSE' 'BEGIN'C2 := LOW(A2/B1);
                   C1 := 'IF'A1 'LESS' ∅ 'THEN' LOW(A1/B1)
                                         'ELSE'  LOW(– A1/B2)
        'END';
    'IF'BA 'EQUIV'BB
    'THEN' 'BEGIN'C[1] := C1; C[2] := – C2 'END'
    'ELSE' 'BEGIN'C[1] := C2; C[2] := – C1 'END'
    'END'
```

Ввод *A*

```
'PROCEDURE' READIN(A); 'ARRAY'A;
    'BEGIN' 'REAL'A1, A2;
            READ(A1, A2); A[1]:= LOW(A1/⟨10"⟩);
                          A[2]:= - LOW(- A2/⟨10"⟩)
    'END'
```

Рассмотрим небольшой пример программирования интервальных операций. Приведенные выше процедуры используются в виде, ориентированном на компьютер ELX8. Их разработал Х. Крист в Вычислительном центре университета г. Карлсруэ. Для простоты пример оформлен таким образом, что не возникает необходимости во вводе и выводе интервалов.

Пример. Имеется многочлен $p(x)$, записанный в виде

$$p(x) = (((\ldots(a_{10}x + a_9)x + a_8)\ldots + a_1)x + a_0.$$

Это выражение должно быть вычислено в интервальной арифметике. Все коэффициенты и переменная представлены интервалами $A_{10} = 10$, $A_9 = 9$, ..., $A_1 = 1$, $A_0 = [-0.02, -0.01]$, а $X = [-0.1, -0.02]$. Мы хотим определить: $0 \in p(X)$ или нет? (см. стр. 297).

Замечания. Вопросы реализации машинной интервальной арифметики на более современном уровне и с учетом последних достижений в этой области были исследованы Кулишем и Миранкером в [430a].

* * *

Литература: [27a], [38], [72], [77], [98], [124], [181], [277], [430a]

Приложение C
ПРОЦЕДУРЫ НА АЛГОЛЕ

A. Локализация собственных значений

Приведем теперь текст программы, представляющей собой алгол-процедуру для итерационного уточнения границ, в которых лежат собственные значения симметричных трехдиагональных вещественных матриц. Процедура реализует метод одновременной локализации в форме (8.5). Постановка задачи приведена в конце главы 8. Как и в гл. 8, собственные значения симметричной трехдиагональной матрицы $\mathscr{A}_p$ мы будем обозначать через $\lambda_1, \ldots, \lambda_n$, а характеристический многочлен для $\mathscr{A}_p$ — че-

```
'BEGIN'
'REAL' 'PROCEDURE'LOW(A);'VALUE'A; 'REAL'A;
'BEGIN'
'IF' A 'NOTLESS' Ø
'THEN' 'BEGIN'A := A × Ø.999999999999; 'IF' A 'NOTGREATER' 10 - 6Ø4
   'THEN'A := Ø 'END'
'ELSE' 'BEGIN'A := A/Ø.999999999999; 'IF' A 'NOTLESS' - 10 - 604
   'THEN'A := - 10 - 6Ø4 'END';
LOW := A;
'IF'A'NOTGREATER' - 10625'THEN''BEGIN'WRITE("OUTOFRANGE"); A := 0;
   A := 1/A 'END'
'END';
'PROCEDURE'ADD(A, B, C); 'ARRAY' A, B, C;
'BEGIN'C[1] := LOW(A[1] + B[1]); C[2] := - LOW(- A[2] - B[2]) 'END';
'PROCEDURE'MUL(A, B, C); 'ARRAY'A, B, C;
   'BEGIN' 'REAL'A1, A2, B1, B2, C1, C2, P; 'BOOLEAN'BA, BB;
A2 := - A[2]; B2 := - B[2]; BA := A2 'LESS' Ø; BB := B2 'LESS' Ø;
'IF'BA'THEN'A1 := A[1] 'ELSE' 'BEGIN' A1 := A2; A2 := A[1] 'END';
'IF'BB 'THEN'B1 := B[1] 'ELSE' 'BEGIN' B1 := B2; B2 := B[1] 'END';
C2 := LOW(- A2 × B2);
'IF' B1 'LESS' Ø
'THEN' 'BEGIN' C1 := LOW(- A2 × B1);
                  'IF' A1 'LESS' Ø
                  'THEN' 'BEGIN' P := LOW(- A1 × B2); 'IF' P 'LESS' C1
                                    'THEN'C1 := P;
                                 P := LOW(- A1 × B1); 'IF' P 'LESS' C2
                                    'THEN'C2 := P
                        'END'
        'END'
'ELSE'C1 := 'IF' A1 'LESS' Ø 'THEN'LOW(- A1 × B2) 'ELSE'LOW(A1 × B1);
'IF' BA 'EQUIV'BB
'THEN' 'BEGIN'C[1] := C1; C[2] := - C2 'END'
'ELSE' 'BEGIN'C[1] := C2; C[2] := - C1 'END'
'END';
'INTEGER'I; 'BOOLEAN'B; 'ARRAY'A[Ø : 1Ø, 1 : 2], X, F, H[1 : 2];
'FOR'I := 1 'STEP' 1 'UNTIL' 1Ø 'DO'A[I, 1] := A[I, 2] := I;
A[Ø, 1] := LOW(- 2/1ØØ); A[Ø, 2] := - LOW(1/1ØØ);
X[1] := LOW(- 1/1Ø); X[2] := - LOW(1/5Ø);
F[1] := A[1Ø, 1]; F[2] := A[1Ø, 2];
'FOR'I := 9 'STEP' - 1 'UNTIL' Ø 'DO'
                  'BEGIN'
                  MUL(F, X, F);
                  H[1] := A[I, 1]; H[2] := A[I, 2];
                  ADD(F, H, F)
               'END';
B := F[1] 'NOTGREATER' Ø 'AND' F[2] 'NOTLESS' Ø;
PRINT(B)
'END'
```

рез $p(x)$. Его значения могут быть вычислены с помощью (8.10).

При реализации метода (8.5) необходимо позаботиться о том, чтобы все встречающиеся в его описании операции были заменены на соответствующие машинные интервальные операции. Замену можно осуществить с помощью алгол-процедур из приложения В.

Необходимо также заметить, что все вещественные операции в (8.5) должны быть выполнены как операции над точечными интервалами. Лишь при соблюдении этого требования можно быть уверенными в том, что локализация собственных значений будет гарантирована. Так, например, если $p([x^{(k,i)}, x^{(k,i)}])$ вычисляется как интервальное выражение, то результат вычисления объемлет вещественное значение $p(x^{(k,i)})$. Поэтому при реализации (8.5) следует различать случаи, используя функцию sign в виде

$$\operatorname{sign}(p([x^{(k,i)}, x^{(k,i)}]));$$

функция эта определена в гл. 8. Встречающиеся в процедуре элементы матрицы представлены как интервалы по той же самой причине.

Если метод (8.5) выполняется так, как это указано выше, то мы получаем последовательность интервалов

$$X^{(0,i)} \supset X^{(1,i)} \supset X^{(2,i)} \supset \ldots \supset X^{(k(i),i)} = X^{(k(i)+1,i)} = \ldots,$$
$$1 \leqslant i \leqslant n,$$

которые перестают изменяться после конечного числа шагов и которые на каждом шаге содержат собственное значение λ_i. В процедуре EIGIMP, реализующей этот метод, в качестве критерия остановки выбрано выполнение равенства

$$X^{(k,i)} = X^{(k+1,i)}.$$

Перечислим формальные параметры процедуры EIGIMP

N Целое число, задающее размерность матрицы.

A Диагональ матрицы. Компоненты $A[I, 1]$ и $A[I, 2]$ служат для представления границ, лежащих на диагонали интервалов $[a_{i1}, a_{i2}]$, $1 \leqslant i \leqslant n$.

B Вектор элементов матрицы, лежащих над ее диагональю. Компоненты $B[I, 1]$ и $B[I, 2]$ служат для представления границ составляющих этот вектор интервалов $[b_{i1}, b_{i2}]$, $1 \leqslant i \leqslant n-1$.

X Вектор, состоящий из пар $X[I, 1]$, $X[I, 2]$. Служит для представления интервалов $[x_{i,1}, x_{i,2}]$, локализующих собственные значения λ_i, $1 \leqslant i \leqslant n$. В момент вызова процедуры должен содержать первоначальные прибли-

жения $X^{(0, i)}$, $1 \leqslant i \leqslant n$. После выполнения процедуры содержит локализации, уточненные согласно (8.5).

BO Булевский вектор размерности n, у которого BO[*I*] имеет значение 'TRUE', если на следующей итерации требуется построить новую локализацию для λ_i. В противном случае *BO*[*I*] = 'FALSE'. Если очередная локализация оказалась равной предыдущей, то соответствующая компонента *BO* устанавливается в 'FALSE'. Перед вызовом EIGIMP комопненты *BO*, соответствующие уточняемым компонентам *X*, должны быть установлены в 'TRUE', а остальные — в 'FALSE'. После выхода из EIGIMP *BO*[*I*] = 'FALSE' для всех *I*.

EPC Вещественная неотрицательная компонента ε, такая что если ширина локализации $d(X^{(k, i)})$ удовлетворяет соотношению

$$d(X^{(k, i)}) < \varepsilon \max\{|s(X^{(k, i)})|, |i(X^{(k, i)})|\},$$

где $d(X^{(k, i)}) = s(X^{(k, i)}) - i(X^{(k, i)})$, то *BO*[*I*] устанавливается в 'FALSE' (т. е. дальнейшее уточнение не производится). Если достигнуть выполнения данного неравенства не удается (например, в случае, когда ε оказалось равным 0), тогда применяется критерий остановки, упомянутый при описании *BO*.

EMPTY Это метка, на которую передается управление, когда при некотором *i* попытка взять пересечение, согласно (8.5), приводит к получению пустого множества. Необходимость в этом возникает, когда $\lambda_i \notin X^{(0, i)}$.

EXIT На эту метку передается управление, если какая-либо из интервальных компонент входных массивов *A*, *B* или *X* имеет нижнюю границу больше верхней. То же самое происходит, если начальные локализации не являются попарно непересекающимися, как того требует метод (8.5). Предполагается, что собственные значения пронумерованы в порядке их роста. Если по каким-либо причинам более удобен другой порядок, то необходимо изменить следующий фрагмент:

```
'FOR'I:=1 'STEP' 1 'UNTIL' N-1 'DO'
    'IF'(BO[I] 'OR'BO[I + 1]) 'AND'X[I, 2] 'NOTLESS'X[I + 1, 1]
    'THEN' 'GOTO'EXIT;
```

Глобальные имена. Процедура EIGIMP использует процедуры LOW и MUL как глобальные.

B. Короткошаговый метод

Теперь дадим описание алгол-процедуры ITERATION. Эта процедура реализует короткошаговый метод с взятием пересечения после вычисления каждой компоненты (SIC). Метод предназначен для нахождения неподвижной точки x^* уравнения $x = \mathscr{A}x + b$ и удовлетворяет построчному или постолбцовому критерию (обсуждение метода см. в гл. 14). Элементы $\mathscr{A}$ и b — вещественные интервалы. Как показал Алефельд [5], в этом случае легко находится вектор $x^{(0)}$, содержащий x^*. Построение локализующего множества для систем уравнений, к которым не удается применить такого рода итерации, обсуждалось Алефельдом и Херцбергером в [14].

Перейдем к перечислению формальных параметров.

N Целое число, задающее размерность системы уравнений.

MAT Соответствует интервальной матрице $\mathscr{A}$, имеющей размерность $n \times n$. Компонента *MAT*[*I*, *J*, 1] (соответственно MAT[*I*, *J*, 2]) представляет нижнюю (соответственно верхнюю) границу A_{ij} — элемента матрицы $\mathscr{A}$.

```
'PROCEDURE'EIGIMP (N,A,B,X,BO,EPS,EMPTY,EXIT);
                'VALUE'N,EPS;
                'INTEGER'N;
                'REAL'EPS;
                'ARRAY'A,B,X;
                'BOOLEAN' 'ARRAY'BO;
                'LABEL'EMPTY,EXIT;
'BEGIN'
    'INTEGER'I,J;
    'REAL'N1,N2,U,V,Y,Z;
    'BOOLEAN'BIT;
    'ARRAY'BB,F[1:N,1:2],XM,D[1:N],H,L,M,F1,F2[1:2];


'COMMENT'ВХОДНЫЕ ДАННЫЕ;
'FOR'I:= 1 'STEP' 1 'UNTIL' N 'DO'
'BEGIN'
    'IF'A[I,2] 'LESS'A[I,1] 'OR'B[I,2] 'LESS'B[I,1]
    'THEN' 'GOTO'EXIT;
    H[1]:= B[I,1];
    H[2]:= B[I,2];
    MUL(H,H,H);
    BB[I,1]:= H[1];
    BB[I,2]:= H[2]
'END'I;


'FOR'I:= 1 'STEP' 1 'UNTIL' N - 1 'DO'
'IF'(BO[I] 'OR'BO[I+1]) 'AND'X[I,2] 'NOTLESS' X[I+1,1]
'THEN' 'GOTO'EXIT;
```

```
'COMMENT'НАЧАЛЬНЫЕ ЗНАЧЕНИЯ;
'FOR'I:= 1 'STEP' 1 'UNTIL' N 'DO'
'IF'BO[I] 'THEN'
'BEGIN'
        XM[I]:= (X[I,1] + X[I,2])/2;
        D[I]:= -LOW(X[I,1] - X[I,2]);
        F1[1]:= F1[2]:= 1;
        F2[1]:= LOW(XM[I] - A[1,2]);
        F2[2]:= -LOW(A[1,1] - XM[I]);
        'FOR'J:= 2 'STEP' 1 'UNTIL' N 'DO'
        'BEGIN'
            L[1]:= BB[J-1,1];
            L[2]:= BB[J-1,2];
            MUL(L,F1,M);
            L[1]:= LOW(XM[I]-A[J,2]);
            L[2]:= -LOW(A[J,1]-XM[I]);
            MUL(L,F2,L);
            F1[I]:= F2[1];
            F1[2]:= F2[2];
            F2[I]:= LOW(L[1]-M[2]);
            F2[2]:= -LOW(M[1]-L[2])
    'END'J;
    F[I,1]:= F2[1];
        F[I,2]:= F2[2]
'END'I;


'COMMENT'ИТЕРАЦИЯ;
REPEAT:
BIT:= 'FALSE';
'FOR' I:= 1 'STEP' 1 'UNTIL' N 'DO'
'IF'BO[I] 'THEN'
'BEGIN'
        'COMMENT'ЗНАМЕНАТЕЛЬ;
    H[1]:= H[2]:= 1;
    'FOR'J:= 1 'STEP' 1 'UNTIL'I-1,I+1 'STEP' 1 'UNTIL' N 'DO'
    'BEGIN'
        L[1]:= LOW(XM[I]-X[J,2]);
        L[2]:= -LOW(X[J,1]-XM[I]);
        MUL(H,L,H)
'END'J;


'COMMENT'НЬЮТОН;
N1:= 'IF'H[1] 'GREATER' ∅
'THEN'
('IF'F[I,1] 'LESS' ∅ 'THEN'LOW(F[I,1]/H[1])
                    'ELSE'LOW(F[I,1]/H[2]))
'ELSE'
'IF'F[I,2] 'GREATER' ∅ 'THEN'LOW(F[I,2]/H[2])
                    'ELSE'LOW(F[I,2]/H[1]);
```

```
N2:= 'IF'H[1] 'GREATER' ∅
     'THEN'
     ('IF'F[1,2] 'GREATER' ∅ 'THEN'-LOW(-F[1,2]/H[1])
                             'ELSE'-LOW(-F[1,2]/H[2]))
     'ELSE'
     'IF'F[1,1] 'LESS' ∅ 'THEN'-LOW(-F[1,1]/H[2])
                         'ELSE'-LOW(-F[1,1]/H[1]);
Y:=N1;
N1:=LOW(XM[1]-N2);
N2:=-LOW(Y-XM[1]);
'IF'N1 'LESS'X[1,1] 'THEN'N1:=X[1,1];
'IF'N2 'GREATER'X[1,2] 'THEN'N2:=X[1,2];
'IF'N2 'LESS' N1 'THEN' 'GOTO'EMPTY;

'COMMENT'ДЕЛЕНИЕ ПОПОЛАМ;
XM[1]:=(N1+N2)/2;
F1[1]:=F1[2]:=1;
F2[1]:=LOW(XM[1]-A[1,2]);
F2[2]:=-LOW(A[1,1]-XM[1]);
'FOR'J:=2 'STEP' 1 'UNTIL' N 'DO'
'BEGIN'
     L[1]:=BB[J-1,1];
     L[2]:=BB[J-1,2];
     MUL(L,F1,M);
     L[1]:=LOW(XM[1]-A[J,2]);
     L[2]:=-LOW(A[J,1]-XM[1]);
     MUL(L,F2,L);
     F1[1]:=F2[1];
     F1[2]:=F2[2];
     F2[1]:=LOW(L[1]-M[2]);
     F2[2]:=-LOW(M[1]-L[2])
'END'J;
F[I,1]:=F2[1];
F[I,2];=F2[2];
'IF'F[I,1] 'GREATER' ∅
'THEN'
     'BEGIN'
          'IF'H[1] 'GREATER' ∅
          'THEN'
               'BEGIN'
                    X[I,2]:=XM[I];
                    X[I,1]:=N1
               'END'
          'ELSE'
               'BEGIN'
                    X[I,1]:=XM[I];
                    X[I,2]:=N2
               'END'
     'END'
'ELSE'
'IF'F[I,2] 'LESS' ∅
'THEN'
```

```
'BEGIN'
    'IF'H[I] 'GREATER' ∅
    'THEN'
        'BEGIN'
            X[I,1]:= XM[I];
            X[I,2]:= N2
        'END'
    'ELSE'
        'BEGIN'
            X[I,2]:= XM[I];
            X[I,1]:= N1
        'END'
    'END'
'ELSE'
    'BEGIN'
        X[I,1]:= N1
        X[I,2]:= N2
    'END',

'COMMENT'ПРОВЕРКА;
    Z:= -LOW(X[I,1]-X[I,2]);
    U:= ABS(X[I,1]);
    V:= ABS(X[I,2]);
    'IF' V 'LESS' U 'THEN' V:= U;
    'IF' Z 'NOTLESS'D[I] 'OR'Z 'NOTGREATER'EPS × V
    'THEN'BO[I]:='FALSE' 'ELSE'BIT:='TRUE';
    D[I]:= Z
'END'КОНЕЦ ИТЕРАЦИИ;
'IF'BIT 'THEN' 'GOTO'REPEAT
'END'КОНЕЦ ПРОЦЕДУРЫ
```

VEC Соответствует n-мерному интервальному вектору $\mathbf{b}$. Его компоненты $VEC[I, 1]$ и $VEC[I, 2]$ представляют границы b_{i1} и b_{i2} компоненты B_i вектора b.

IXIT Метка, на которую передается управление, если $\mathcal{A}$ не удовлетворяет построчному или постолбцовому критерию.

X Возвращаемое значение неподвижной точки x^*.

Процедура ITERATION использует процедуры LOW, ADD и MUL как глобальные (см. стр. 304, 305).

C. Обращение матрицы

Перейдем теперь к описанию алгол-процедуры, которая реализует изложенный в гл. 18 итерационный метод для локализации матрицы, обратной к $\mathcal{A}_p$. Чтобы сохранить гарантированность включений при вычислениях в машинной интервальной арифметике, нужно заменить вещественные операции на операции с точечными интервалами. Только в этом случае мы сможем учесть все погрешности, возникающие из-за округлений. По этой

причине, а также из-за того, что при вводе тоже нельзя избежать появления погрешностей, все вещественные константы и матрицы заменяются на интервальные. Для получения направленных округлений при выполнении интервальных операций мы будем пользоваться процедурами, описанными в приложении В.

Сначала ознакомимся с процедурой MATINV, которая имеет следующие формальные параметры (см. стр. 306):

N Целое число, задающее размерность матрицы.

A Служит для представления обращаемой интервальной матрицы $\mathscr{A}$ размерности $n \times n$. Компоненты $A[I, J, 1]$ и $A[I, J, 2]$ соответствуют нижней и верхней границам элемента A_{ij} матрицы $\mathscr{A}$. Если A_{ij} представим в виде точечного интервала, то, конечно, эти границы равны.

```
'PROCEDURE'ITERATION(MAT,VEC,N,EXIT)RESULT:(X);
'VALUE'MAT,VEC,N; 'ARRAY'MAT,VEC,X; 'INTEGER'N; 'LABEL'EXIT;
'BEGIN' 'INTEGER'I,J; 'REAL'MAX,HMAX,NORM,HNORM; 'BOOLEAN'BV;
    'ARRAY'Y[1:N],IK,HMAT,HX[1:2];
    'REAL' 'PROCEDURE'UP(A); 'VALUE'A; 'REAL'A; UP:= -LOW(-A);
    'REAL' 'PROCEDURE'ABSVAL(LBD,VBD); 'REAL'LBD,UBD;
    'IF'ABS(LBD) 'LESS'ABS(UBD) 'THEN'ABSVAL:= ABS(UBD) 'ELSE'
    ABSVAL:= ABS(LBD);
MAX:= 0;
'COMMENT'ОЦЕНИВАНИЕ С ПОМОЩЬЮ НОРМЫ, ВЫЧИСЛЯЕМОЙ
    СУММИРОВАНИЕМ ПО СТРОКАМ;
'FOR'I:= 1 'STEP' 1 'UNTIL' N 'DO'
    'BEGIN'Y[I]:= 0;
        'FOR'J:= 1 'STEP' 1 'UNTIL' N 'DO'
        Y[I]:= UP(Y[I] + ABSVAL (MAT[I,J,1],MAT[I,J,2]));
        'IF'Y[I] 'NOTLESS' 1 'THEN' 'GOTO'M1
    'END';
'COMMENT'НАХОЖДЕНИЕ НАЧАЛЬНОГО ВЕКТОРА, ЕСЛИ
    УДОВЛЕТВОРЕН КРИТЕРИЙ СУММЫ ПО СТРОКАМ;
'FOR'I:= 1 'STEP' 1 'UNTIL' N 'DO'
'BEGIN'HMAX:= 0;
    'FOR'J:= 1 'STEP' 1 'UNTIL' N 'DO'
    HMAX:= UP(HMAX+UP(ABSVAL(MAT[I,J,1],MAT[I,J,2]) ×
    ABSVAL(VEC[J,1],VEC[J,2])));
    HMAX:= UP(HMAX/LOW(1-Y[I]));
    'IF'MAX 'LESS'HMAX 'THEN'MAX:= HMAX
'END'; 'GOTO'M2;
'COMMENT'ОЦЕНИВАНИЕ С ПОМОЩЬЮ НОРМЫ, ВЫЧИСЛЯЕМОЙ
    СУММИРОВАНИЕМ ПО СТОЛБЦАМ;
M1 : NORM:= 0;
'FOR'J:= 1 'STEP' 1 'UNTIL' N 'DO'
'BEGIN'HNORM:= 0;
    'FOR'I:= 1 'STEP' 1 'UNTIL' N 'DO'
    HNORM:= UP(HNORM + ABSVAL(MAT[I,J,1],MAT[I,J,2]));
    'IF'HNORM 'NOTLESS' 1 'THEN' 'GOTO'EXIT;
    'IF'HNORM 'GREATER'NORM 'THEN'NORM:= HNORM
'END';
```

```
'COMMENT'НАХОЖДЕНИЕ НАЧАЧАЛЬНОГО ВЕКТОРА,ЕСЛИ
    УДОВЛЕТВОРЕН КРИТЕРИЙ СУММЫ ПО СТОЛБЦАМ;
'FOR'I:= 1 'STEP' 1 'UNTIL' N 'DO'
'FOR'J:= 1 'STEP' 1 'UNTIL' N 'DO'
MAX:= UP(MAX+UP(ABSVAL(MAT[I,J,1],MAT[I,J,2]) ×
 ABSVAL(VEC[J,1],VEC[J,2])));
MAX:= UP(MAX/LOW(1-NORM));
M2 : 'FOR'I:= 1 'STEP' 1 'UNTIL' N 'DO'
    'BEGIN'X[I,1]:= LOW(VEC[I,1]-MAX);
    X[I,2]:= UP(VEC[I,2]+MAX) 'END';
'COMMENT'НАЧАЛО ОЧЕРЕДНОЙ ИТЕРАЦИИ;
M3 : BV:='TRUE';
'FOR'I:= 1 'STEP' 1 'UNTIL' N 'DO'
'BEGIN'IK[1]:= VEC[I,1]; IK[2]:= VEC[I,2];
    'FOR'J:= 1 'STEP' 1 'UNTIL' N 'DO'
    'BEGIN'HMAT[1]:= MAT[I,J,1]; HMAT[2]:= MAT[I,J,2];
        HX[1]:= X[J,1]; HX[2]:= X[J,2];
        MUL(HMAT,HX,HX); ADD(HX,IK,IK)
    'END';
    HX[1]:= X[I,1]; HX[2]:= X[I,2];
    'IF'IK[1] 'LESS'HX[1] 'THEN'IK[1]:= HX[1];
        'IF'IK[2] 'GREATER'HX[2] 'THEN'IK[2]:= HX[2];
        'IF'BV 'THEN' 'BEGIN' 'IF'IK[1] 'NOTEQUAL'X[I,1] 'OR'IK[2]
          'NOTEQUAL'
        X[I,2] 'THEN'BV:= 'FALSE' 'END';
        X[I,1]:= IK[1]; X[I,2]:= IK[2]
    'END';
    'IF' 'NOT'BV 'THEN' 'GOTO'M3
'END'
```

X Служит для представления интервальной матрицы размерности $n \times n$. Компоненты $X[I, J, 1]$ и $X[I, J, 2]$ соответствуют нижней и верхней границам I-го элемента матрицы. Параметр используется и как входной, и как выходной. При вызове процедуры этот формальный параметр заменяется фактическим параметром $\mathscr{X}^{(0)}$, таким что

$$\{\mathscr{A}_p^{-1} \mid \mathscr{A}_p \in \mathscr{A}\} \in \mathscr{X}^{(0)} \quad \text{и} \quad \| \mathscr{I}_p - \mathscr{A}m(\mathscr{X}^{(0)}) \| < 1,$$

где $\|\cdot\|$ — норма матрицы. После подстановки $\mathscr{X}$ итерация (18.5) повторяется (с $r = 2$) до тех пор, пока не будет сформирована матрица, удовлетворяющая неравенству (18.11). В качестве нормы используется $\|\mathscr{X}\| = n \max_{1 \leqslant i, j \leqslant n} |x_{ij}|$.

Когда выполнено (18.11), тогда при $r = 2$ итерации продолжаются в соответствии с (18.9), пока две следующие подряд локализации не окажутся равными друг другу. Тогда итерации останавливаются, а последняя интервальная матрица сохраняется в X, служащем выходным параметром. Для нахождения матрицы $\mathscr{X}^{(0)}$ применяется алгол-процедура INCLUSIONSET.

Процедуры LOW, ADD и MUL из приложения В используются нами как глобальные.

Последней мы рассмотрим алгол-процедуру INCLUSIONSET (см. стр. 309), которая вычисляет начальную матрицу $\mathscr{X}^{(0)}$ для последующей передачи ее в процедуру MATINV. Процедура имеет следующие формальные параметры:

N Задает размерность обращаемой матрицы.

B Соответствует интервальной матрице $\mathscr{B}$ размерности $n \times n$. Компонента $B[I, J, 1]$ (соответственно $B[I, J, 2]$) представляет нижнюю (соответственно верхнюю) границу B_{ij} — элемента матрицы $\mathscr{B}$. Фактическим параметром для *B* служит машинное представление обращаемой матрицы $\mathscr{A}_{\text{р}}$.

X Соответствует интервальной матрице $\mathscr{X}$ размерности $n \times n$. Компонента $X[I, J, 1]$ (соответственно $X[I, J, 2]$) представляет нижнюю (соответственно верхнюю) границу X_{ij} — элемента матрицы $\mathscr{X}$. После выполнения процедуры *X* соответствует интервальной матрице $\mathscr{X}$, обладающей свойством

$$\{\mathscr{B}^{-1} \mid \mathscr{B}_{\text{р}} \in \mathscr{B}\} \subseteq \mathscr{X}.$$

Вычисление $\mathscr{X}$ производится с помощью формул, следующих из теоремы 18.2. В качестве $|\mathscr{B}|$ используются как построчная, так и постолбцовая нормы.

L Это метка, на которую передается управление, если ни построчная, ни постолбцовая нормы $|\mathscr{B}|$ не меньше 1 (в этом случае матрицу $\mathscr{X}$ невозможно вычислить с помощью упомянутых выше формул).

Процедура LOW из прилож. В используется как глобальная.

```
'PROCEDURE'MATINV(A,N,X);
'VALUE'N;
'INTEGER'N;
'ARRAY'A,X;
'BEGIN' 'INTEGER'I,J,K;
        'REAL'U,V,W,NA,NDX,NDR,H;
        'BOOLEAN'B1,B2;
        'ARRAY'S,H1,H2[1:2],Y[1:N,1:2],M[1:N,1:N],
        Z,R[1:N,1:N,1:2];
U:= LOW(1/N);
B1:= B2:= 'FALSE';

'COMMENT'НОРМА A;
NA:= 0;
'FOR'I:= 1 'STEP' 1 'UNTIL' N 'DO'
'FOR'J:= 1 'STEP' 1 'UNTIL' N 'DO'
'BEGIN'V:= ABS(A[I,J,1]); W:= ABS(A[I,J,2]):
    'IF'W 'LESS'V 'THEN'W:= V;
    'IF'W 'GREATER'NA 'THEN'NA:= W;
    Z[I,J,1]:= X[I,J,1]; Z[I,J,2]:= X[I,J,2]
'END'КОНЕЦ НОРМЫ A;
L1:B1:= 'FALSE';
```

```
'COMMENT'ВЫЧИСЛЕНИЕ СРЕДНЕЙ ТОЧКИ;
'FOR'I:= 1 'STEP' 1 'UNTIL' N 'DO'
'FOR'J:= 1 'STEP' 1 'UNTIL' N 'DO'
M[I,J]:= (X[I,J,1] + X[I,J,2])/2;

'COMMENT'РАЗНОСТЬ;
'FOR'I:= 1 'STEP' 1 'UNTIL' N 'DO'
'BEGIN' 'FOR'J:= 1 'STEP' 1 'UNTIL' N 'DO'
        'BEGIN'S[1]:= S[2]:= 0;
                'FOR'K:= 1 'STEP' 1 'UNTIL' N 'DO'
                'BEGIN'H1[1]:= A[I,K,1]; H1[2]:= A[I,K,2];
                        H2[1]:= H2[2]:= M[K,J];
                        MUL(H1,H2,H1); ADD(S,H1,S)
                'END'K;
                R[I,J,1]:= -S[2]; R[I,J,2]:= -S[1]
        'END'J;
        R[I,I,1]:= LOW(1 + R[I,I,1]);
        R[I,I,2]:= -LOW(-1 - R[I,I,2])
'END' КОНЕЦ РАЗНОСТИ;

'COMMENT'ОЧЕРЕДНАЯ ИТЕРАЦИЯ;
'IF'B2 'THEN' 'FOR'I:= 1 'STEP' 1 'UNTIL' N 'DO'
                'FOR'J:= 1 'STEP' 1 'UNTIL' N 'DO'
                'BEGIN'Z[I,J,1]:= X[I,J,1];
                        Z[I,J,2]:= X[I,J,2]
                'END';
'FOR'I:= 1 'STEP' 1 'UNTIL' N 'DO'
'BEGIN' 'FOR'J:= 1 'STEP' 1 'UNTIL' N 'DO'
        'BEGIN'S[1]:= S[2]:= 0;
                'FOR'K:= 1 'STEP' 1 'UNTIL' N 'DO'
                'BEGIN'H1[1]:= Z[I,K,1]; H1[2]:= Z[I,K,2];
                        H2[1]:= R[K,J,1]; H2[2]:= R[K,J,2];
                        MUL(H1,H2,H1); ADD(S,H1,S)
                'END'K;
                H1[1]:= H1[2]:= M[I,J]; ADD(S,H1,S);
                Y[J,1]:= S[1]; Y[J,2]:= S[2]
        'END'J;
        'FOR'J:= 1 'STEP' 1 'UNTIL' N 'DO'
        'BEGIN'Z[I,J,1]:= Y[J,1];
                Z[I,J,2]:= Y[J,2]
        'END'J
'END'I;

'COMMENT'ПЕРЕСЕЧЕНИЕ;
'IF'B2 'THEN'
L2 : 'BEGIN' 'FOR'I:= 1 'STEP' 1 'UNTIL' N 'DO'
                'FOR'J:= 1 'STEP' 1 'UNTIL' N 'DO'
                'BEGIN' 'IF'X[I,J,1] 'LESS'Z[I,J,1] 'THEN'
                        'BEGIN'X[I,J,1]:= Z[I,J,1]; B1:= 'TRUE' 'END';
                        'IF'X[I,J,2] 'GREATER'Z[I,J,2] 'THEN'
                        'BEGIN'X[I,J,2]:= Z[I,J,2]; B1:= 'TRUE' 'END'
                'END'I,J;
                'GOTO' 'IF'B1 'THEN'L1 'ELSE'L3
'END';
```

```
'COMMENT'ПРОВЕРКА ШИРИНЫ;
NDR:= NDX:= 0;
'FOR'I:= 1 'STEP' 1 'UNTIL' N 'DO'
'FOR'J:= 1 'STEP' 1 'UNTIL' N 'DO'
'BEGIN'V:= ABS(R[I,J,1]); W:= ABS(R[I,J,2]);
        'IF'W 'LESS'V 'THEN'W:= V;
        'IF'W 'GREATER'NDR 'THEN' NDR:= W;
        H:= -LOW(X[I,J,1]-X[I,J,2]);
        'IF' H 'GREATER'NDX 'THEN'NDX:= H
'END'КОНЕЦ РАЗНОСТИ НОРМ И ШИРИНЫ НОРМЫ;
W:= LOW(U-NDR);
U:= LOW(2/N);
W:= LOW(W/NA);
W:= LOW(W×U);
'IF'NDX 'LESS'W 'THEN' 'BEGIN'B2:= 'TRUE';
                                'GOTO'L2
                        'END'
                        'ELSE'
                        'BEGIN' 'FOR'I:= 1 'STEP' 1 'UNTIL' N 'DO'
                                'FOR'J:= 1 'STEP' 1 'UNTIL' N 'DO'
                                'BEGIN'X[I,J,1]:= Z[I,J,1];
                                        X[I,J,2]:= Z[I,J,2]
                                'END';
                                'GOTO'L1
                        'END'КОНЕЦ ПРОВЕРКИ ШИРИНЫ
ITERATION LOOP;
L3 : 'END'КОНЕЦ ТЕЛА ПРОЦЕДУРЫ
```

Детальное описание обеих процедур, а также численные примеры можно найти у Алефельда и Херцбергера в [17].

* * *

Литература: [5], [14], [17], [18], [72], [340]

```
'PROCEDURE' INCLUSIONSET(B,N,X,L);
        'VALUE'N;
        'INTEGER'N;
        'ARRAY'B,X;
        'LABEL'L;
'BEGIN' 'INTEGER'I,J;
         'REAL'NOR,A1,A2,HI;
         NOR:=Ø;

'COMMENT' ПОСТРОЧНАЯ НОРМА;
'FOR' I:=1 'STEP' 1 'UNTIL' N 'DO'
'BEGIN'HI:=Ø;
    'FOR'J:=1 'STEP' 1 'UNTIL' N 'DO'
    'BEGIN'A1:=ABS(B[I,J,1]); A2:=ABS(B[I,J,2]);
            HI:=-LOW(-HI-('IF'A1 'GREATER'A2 'THEN'A1 'ELSE'A2))
    'END'J;
    'IF'HI 'GREATER'NOR 'THEN'NOR:=HI;
    'IF'NOR 'NOTLESS' 1 'THEN' 'GOTO'COLNORM
'END'I,КОНЕЦ ПОСТРОЧНОЙ НОРМЫ;
'GOTO'INCL;
COLNORM : NOR:=Ø;
'FOR'J:=1 'STEP' 1 'UNTIL' N 'DO'
'BEGIN'HI:=Ø;
    'FOR'I:=1 'STEP' 1 'UNTIL' N 'DO'
    'BEGIN'A1:=ABS(B[I,J,1]); A2:=ABS(B[I,J,2]);
            HI:=-LOW(-HI-('IF'A1 'GREATER'A2 'THEN'A1 'ELSE'A2))
    'END'I;
    'IF'HI 'GREATER'NOR 'THEN'NOR:=HI;
    'IF'NOR 'NOTLESS' 1 'THEN' 'GOTO'L
'END'J,КОНЕЦ ПОСТОЛБЦОВОЙ НОРМЫ;

INCL :
A1:=LOW(1-NOR);
A1:=-LOW(-1/A1);
'FOR'I:=1 'STEP' 1 'UNTIL' N 'DO'
'BEGIN' 'FOR'J:=1 'STEP' 1 'UNTIL'I-1,I+1 'STEP' 1 'UNTIL' N 'DO'
      BEGIN'X[I,J,1]:=-A1;
            X[I,J,2]:=A1
    'END'J;
    X[I,I,1]:=-A1;
    X[I,I,2]:=-LOW(-2-A1)
'END'I
'END'
```

ЛИТЕРАТУРА

[1] O. Aberth, Iteration methods for finding all zeros of a polinomial simultaneously. Math. Comp. 27, 339—344 (1973).

[2] I. Ahrens, Über Folgen und Reihen von Punktmengen. Arch. Math. 23, 92—103 (1972).

[3] J. Albrecht, Monotone Iterationsfolgen und ihre Verwendung zur Lösung linearer Gleichungssysteme. Numer. Math. 3, 345—358 (1961).

[4] G. Alefeld, Intervallrechnung über den komplexen Zahlen und einige Anwendungen. Ph. D. Thesis, Univ. Karlsruhe (1968).

[5] G. Alefeld, Über die aus monoton zerlegbaren Operatoren gebildeten Iterationsverfahren. Computing 6, 161—172 (1970).

[6] G. Alefeld, Eine Modifikation des Newtonverfahrens zur Bestimmung der reellen Nullstellen einer reellen Funktion. Z. Angew. Math. Mech. 50, T32—T33 (1970).

[7] G. Alefeld, Über Eigenschaften und Anwendungsmoglichkeiten einer Intervallarithmetik über den komplexen Zahlen. Z. Angew. Math. Mech. 50, 455—465 (1970).

[8] G. Alefeld, Anwendungen des Fixpunktsatzes für pseudometrische Räume in der Intervallrechnung. Numer. Math. 17, 33—39 (1971).

[9] G. Alefeld, Quadratisch konvergente Einschließung von Lösungen bei nichtkonvexen Gleichungssystemen. Z. Angew. Math. Mech. 54, 335—342 (1974).

[10] G. Alefeld, Über die Existenz einer eindeutigen Lösung bei einer Klasse nichtlinearer Gleichungssysteme und deren Berechnung mit Iterationsverfahren. Apl. Mat. 17. 267—294 (1972).

[11] G. Alefeld and N. Apostolatos, Auflösung nichtlinearer Gleichungssysteme mit zwei Unbekannten. Bericht des Rechenzentrums Inst. Angew. Math., Univ. Karlsruhe (1967).

[12] G. Alefeld and N. Apostolatos, Praktische Anwendung von Abschätzungsformeln bei Iterationsverfahren. Z. Angew. Math. Mech. 48. T46—T49 (1968).

[13] G. Alefeld and J. Herzberger, Über das Newton-Verfahren bei nichtlinearen Gleichungssystemen. Z. Angew. Math. 50. 773—774 (1970).

[14] G. Alefeld and J. Herzberger, ALGOL-60 Algorithmen zur Auflösung linearer Gleichungssysteme mit Fehlererfassung. Computing 6, 28—34 (1970).

[15] G. Alefeld and J. Herzberger, Über die Berechnung der inversen Matrix mit Hilfe der Intervallrechnung. Elektron. Rech. 12, 259—261 (1970).

[16] G. Alefeld and J. Herzberger, Verfahren höherer Ordnung zur iterativen Einschteißung der inversen Mitrix. Z. Angew. Math. Phys. 21, 819—824 (1970).

[17] G. Alefeld and J. Herzberger, Matrizeninvertierung mit Fehlererfassung. Elektron. Datenverarbeit. 12, 410—416 (1970).

[18] G. Alefeld and J. Herzberger, Über die Verbesserung von Schranken für die Lösung bei linearen Gleichungssystemen. Angew. Informat. 13, 107—112 (1971).

[19] G. Alefeld and Herzberger, Zur Invertierung linearen und beschränkter Operatoren. Math. Z. 120, 309—317 (1971).
[20] G. Alefeld and J. Herzberger, Einschließungsverfahren zur Berechnung des inversen Operators. Z. Angew. Math. Mech. 52. T197—T198 (1972).
[21] G. Alefeld and J. Herzberger, Nullstelleneinschließung mit dem Newton — Verfahren ohne Invertierung von Intervallmatrizen. Numer. Math. 19, 56—64 (1972).
[22] G. Alefeld and J. Herzberger, Ein Verfahren zur monotonen Einschließung von Lösungen nichtlinearer Gleichngssysteme. Z. Angew. Math. Mech. 53, T176—T177 (1973).
[23] G. Alefeld and J. Herzberger, Über die Konvergenzordnung intervallarithmetischer iterationsverfahren. Kolloq. Konstrukt. Theorie Funktionen. Conf. Proc., Cluj (1973).
[24] G. Alefeld and J. Herzberger. Über Simultanverfahren zur Bestimmung reeller Polynomwurzeln. Z. Angew. Math. Mech. 54, 413—420 (1974).
[25] G. Alefeld and J. Herzberger, Über die Verbesserung von Schranken für die Eigenwerte symmetrischer Tridiagonalmatrizen. Angew. Informat. 16, 27—35 (1974).
[26] G. Alefeld and J. Herzberger. On the convergence speed of some algorithms for the simultaneous approximation of polynomial roots. SIAM J. Numer. Anal. 11, 237—243 (1974).
[27] G. Alefeld, J. Herzberger, and O. Mayer, Über neuere Gesichtspunkte beim numerischen Rechnen. Math. Naturwiss. Unterricht 24, 458—467 (1971).
[27a] G. Alefeld, J. Herzberger, and O. Mayer. Einfuhrung in das Programmieren mit ALGOL 60. Bibliograph. Inst. Mannheim, Hochschultaschenbuch 777 (1972).
[28] M. Altmann, An optimum cubically convergent iterative method of inverting a linear bounded operator in Hilbert space. Pacific J. Math. 10, 1107—1113 (1960).
[29] N. Apostolatos. Ein Einschließungsverfahren für Nullsstellen. Z. Angew. Math. Mech. 47, T80 (1967).
[30] N. Apostolatos, Nullstellenbestimmung mit Fehlererfassung. Bericht des Rechenzentrums. Inst. Angew. Math., Univ. Kalsruhe (1967).
[31] N. Apostolatos Allgemeine Intervallarithmetiken und Anwendungen. Bull. Soc. Math. Athens. N. Ser. 10, 136—180 (1969).
[32] N. Apostolatos, see G. Alefeld.
[33] N. Apostolatos and U. Kulisch, Grundlagen einer Maschinenintervallarithmetik. Computing 2, 89—104 (1967).
[34] N. Apostolatos and U. Kulisch, Approximation der erweiterten Intervallarithmetik durch die einfache Maschinenintervallarithmetik. Computing 2, 181—194 (1967).
[35] N. Apostolatos and U. Kulisch, Über die Konvergenz des Relaxationsverfahrens bei nichtnegativen und diagonaldominanten Matrizen. Computing 2, 17—24 (1967).
[36] N. Apostolatos and U. Kulisch, Gruntzuge einer Intervallrechnung für Matrizen und einige Anwendungen. Elektron. Rech. 10, 73—83 (1968).
[37] N. Apostolatos, U. Kulisch, and K. Nickel, Ein Einschließungsverfahren für Nullstellen. Computing 2, 195—201 (1967).
[38] N. Apostolatos, U. Kulisch, R. Krawczyk B. Loriz, K. Nickel, and H.-W. Wippermann. The algorithmic language Triplex-Algol 60. Numer. Math. 11, 175—180 (1968).
[39] W. Appelt, Fehlereinschliebung bei der numerischen Lösung elliptischer Differentialgleichungen unter Verwendung eines intervallarithmetischen Defektverfahrens. Ph. D. Thesis, Univ. Bonn (1972).
[40] D. W. Arthur, The use of interval arithmetic to bound the zeros of real polynomials. J. Inst. Math. Appl. 10, 231—237 (1972).

[41] J. Avenhaus, Zur numerischen Behandlung des Anfangswertproblems mit exakter Fehlererfassung. Ph. D. Thesis, Univ. Karlsruhe (1970).
[42] J. Avenhaus, Ein Verfahren zur Einschließung der Lösung des Anfangswertproblems. Computing 8, 182—190 (1971).
[43] A. N. Baluev, Zur Methode von Chaplygin (Russian). Dokl. Akad. Nauk SSSR 83, 781—784 (1952). — Оригинал: Балуев А. Н. К абстрактной теории метода С. А. Чаплыгина//Доклады Акад. наук СССР. — 1952, т. 83, № 6, с. 781—784.
[44] W. Barth, Nullstellenbestimmung mit der Intervallrechnung. Computing 8, 320—328 (1971).
[45] W. Barth, Ein Algorithmus zur Berechnung aller reellen Nullstellen in einem Intervall. Computing 9, 327—333 (1972).
[46] M. Bauer, Zur Bestimmung der reellen Wurzeln einer algebraischen Gleichung durch Iteration. Jber. Deutsch. Math.-Verein. 25, 294—301 (1917).
[47] H. Beeck, Über intervallanalytische Methoden bei linearen Gleichungssystemen mit Intervallkoeffizient und Zusammenhänge mit der Fehleranalysis. Ph. D. Thesis. TH München (1971).
[48] H. Beeck, Über die Struktur und Abschätzungen der Lösungsmenge von linearen Gleichungssystemen mit Intervallkoeffizienten. Computing 10, 231—244 (1972).
[49] S. Berti, The solution of an interval equation. Mathematica 11, 189—194 (1969).
[50] S. Berti, The geometry of the relations in the set of real numbers. Mathematica 11, 29—43 (1969).
[51] S. Berti, Intervalele si aritmetica lor in analiza numerica. Gaz. Mat. A75, 309—313 (1970).
[52] S. Berti, Some relations between interval functions (I). Mathematica 14, 9—26 (1972).
[53] S. Berti, Aritmetica si Analiza Intervalelor. Rev. Anal. Numer. Teoria Aproximatiei 1, 21—39 (1972).
[54] M. Blumhofer, Über einige Algorithmen zur numerischen Einschließung der Lösung bei linearen Gleichungssystmen. Diplomarbeit, Inst. Angew. Math., Univ. Karlsruhe (1972).
[55] R. Boche, Some observations on the economics of interval arithmetic. Comm. ACM 8, 649 (1965).
[56] R. Boche, Complex interval arithmetic with some applications. Lockheed Missiles and Spase Company 4-22-66-1, Sunnyvale. California (1966).
[57] W. Börsch-Supan, A post eriori error bounds for the zeros of polynomials. Numer. Math. 5, 380—398 (1963).
[58] D. Braess, Die Konstruction monotoner Iterationsfolgen zur Lösungseinschließung bei linearen Gleichungssystemen. Arch. Rational Mech. Anal. 9, 97—106 (1962).
[59] D. Braess, Monotone Iterationsfolgen bei Gleihungssystemen mit fehlerhaften Koeffizienten und Iterationsbeschleunigung. Numer. Math. 7, 32—41 (1965).
[60] D. Braess, Die Berehnung der Fehlergrenzen bei linearen Gleichungssystemen mit fehlerhaften Koeffizienten. Arch. Rational Mech. Anal. 19, 74—80 (1965).
[61] D. Braess and K. P. Hideler, Simultaneous inclusion of the zeros of a polynomial. Numer. Math. 21, 161—165 (1973).
[62] R. P. Brent, An algorithm with guaranteed convergence for finding a zero of a function. Comput. J. 14, 422—425 (1972).
[63] M. Brockhaus, B. Rothmaier, and P. Schroth, Benutzeranleitung für TRIPLEX-ALGOL im System HYDRA 2. Interner Bericht 11, Inst. f. Informatik. Univ. Karlsruhe (1969).
[64] W. Burmeister, Inversionzfreie Verfahren zur Lösung nichtlinearer Operatorgleichungen. Z. Amgew. Math. Mech. 52. 101—110 (1972).

[65] P. Byrnev, see K. Dochev.
[66] W. P. Champagne, On finding roots of polynomials by hook or by crook. TTN-37, Univ. of Texas, Computation Center (1964).
[67] B. A. Chartres, Automatic controlled precision calculations. J. ACM 13, 386—403 (1966).
[68] B. A. Chartres, Computable error bounds for direct solution of linear equations. J. ACM 14, 63—71 (1967).
[69] F. L. Chernous'ko, An optimal algorithm for finding the roots of an approximately computed function. Ž. Vyčisl. Mat. i Mat. Fiz. 8, 706—724 (1968). — Оригинал: Черноусько Ф. Л. Оптимальный алгоритм поиска корня функции, вычисляемой приближенно.//Ж. выч. мат. и мат. физ. — 1968, т. 8, № 4, с. 706—724.
[70] H. H. Y. Chien, A multiphase algorithm for single variable equation solving. J. Inst. Math. Appl. 9, 290—298 (1972).
[71] H. Crist, Realisierung einer Maschinenintervallarithmetik auf beliebigen ALGOL-Compilern. Z. Angew. Math. Mech. 48, T110 (1968).
[72] H. Crist, Realisierung einer Maschinenintervallarithmetik mit beliebigen ALGOL-60 Compilern. Elektron. Rech. 10, 217—222 (1968).
[73] W. Chuba and W. Miller, Quadratic convergence in interval arithmetic, Part. I. BIT 12, 284—290 (1972).
[74] L. Collatz, «Fynktionalanalysis und numerische Mathematik». Springer-Verlag, Berlin and New York, 1964. — Рус. перев.: Коллатц Л. Функциональный анализ и вычислительная математика. — М.: Мир, 1969, 447 с.
[75] R. H. Dargel, F. R. Loscalzo, and T. H. Witt, Automatic error bounds on real zeros of rational functions. Comm. ACM 9, 806—809 (1966).
[76] D. P. Davey and N. F. Stewart, Guaranteed error bounds for the initial value problem using polytope arithmetic. Publication #109, Departement d'Informatique, Univ. de Montreal (1972).
[77] J. K. S. Dewar, Procedures for interval arithmetic. Comp. J. 14, 447—450 (1970).
[78] K. Dochev and P. Byrnev, Certain modifications of Newton's method for the approximate solution of algebraic equations. Ž. Vyčisl. Mat. i Mat. Fiz. 4, 915—920 (1964). — Оригинал: Дочев К., Бырнев П. О некоторых модификациях метода Ньютона для приближенного решения алгебраических уравнений. Ж. выч. мат. и мат. физ. — 1964, т. 4, N 5, с. 915—920.
[79] M. Dowell and P. Jarratt, A modified regula falsi method for computing the root of an equation. BIT 11, 168—174 (1971).
[80] H. Dressel, see J. W. Schmidt.
[81] E. Durand, «Solutions Numeriques des Equations Algebriques», Vol. I. Masson, Paris, 1960.
[82] R. Dussel, Einschließung des Minimalpunktes einer streng konvexen Funktion auf einem n-dimensionalen Quaber. Ph. D. Thesis, Univ. Karlsruhe (1972).
[83] R. Dussel and B. Schmitt, Die Berechnung von Schranken fur den Wertebereich emes Polynoms in einem Intervall. Computing 6, 35—60 (1970).
[84] P. S. Dwyer, Matrix inversion with the square root method. Technometrics 6, 197—213 (1964).
[85] H.-P. Dyck, Algorithmen zur numerischen Behandlung von Funktionen mit Fehlererfassung. Diplomarbeit, Inst. Angew. Math., Univ. Karlsruhe (1969).
[86] K. Ecker and H. Ratschek, Intervallarithmetik mit Wahrscheinlichkeitsstruktur. Angew. Informat. 14, 313—320 (1972).
[87] L. W. Ehrlich, A modified Newton method for polynomials. Comm. ACM 10, 107—108 (1967).
[88] H. Ehrmann, Konstruktion und Durchführung von Iterationsverfahren höherer Ordnung. Arch. Rational Mech. Anal. 4, 65—88 (1959).

[89] H. Fischer, Eine Methode zur Einschließung der Nullstellen von Polynomen mit komplexen Koeffizienten. Computing 4, 202—206 (1969).
[90] H. Fischer, Intervall-Arithmetiken fur komplexe Zahlen. Z. Angew. Math. Mech. 53, T190—T191 (1973).
[91] H. Fischer, Hypernormballe als abstrakte Schrankenzahlen. Computing 12, 67—73 (1974).
[92] P. Forster, Bemerkungen zum Iterationsverfahren von Schulz zur Bestimmung der Inversen einer Matrix. Numer. Math. 12, 211—214 (1968).
[93] R. Franzen, Die intervallanalytische Behandlung parameterabhängiger Gleichungssysteme. GMD-Bericht No. 47, Bonn (1971).
[94] G. S. Ganshin, Extension of the convergence region of Newton's method. Ž. Vyčisl. Mat. i Mat. Fiz. 11, 1294—1296 (1971). — Оригинал: *Ганшин Г. С.* Расширение области сходимости метода Ньютона. Ж. выч. мат. и мат. физ. — 1971. — Т. 11. — С. 1294—1296.
[95] I. Gargantini and P. Henrici, Circular arithmetic and the determination of polynomial zeros. Numer. Math. 18, 305—320 (1972).
[96] A. Gibb, Algorithm 61. Procedures for range arithmetic. Comm. ACM 4, 319—320 (1961).
[97] A. J. Goldstein and P. L. Richman, A. midpoint phenomenon. J. ACM 20, 301—304 (1973).
[98] D. I. Good and R. L. London, Computer interval arithmetic: Definition and proof of correct implementation. J. ACM 17, 603—612 (1970).
[99] S. Gorn, The automatic analysis and control of computing errors. SIAM J. Appl. Math. 2, 69—81 (1954).
[100] J. A. Grant and G. D. Hitchins, The solution of polynomial equations in interval arithmetic. Comput. J. 16, 69—72 (1973).
[101] R. T. Gregory and D. L. Karney, «A Collection of Matrices for Testing Computational Algorithms». Wiley (Interscience), New York, 1969.
[102] W. Gröbner, Matrizenrechnung. Bibliographisches Institut, Mannheim (1966).
[103] N. B. Haaser and J. A. Sullivan, «Real Analysis», Chapter 5, Section 2. Van Nostrand Reinhold, New York, 1971.
[104] K. P. Hadeler, see D. Braess.
[105] W. Hahn, Intervallarithmetik in normierten Räumen und Algebren. Ph. D. Thesis, Graz (1971).
[106] K. Hainer, see F. Stummel.
[107] H. Haller, Externe Zahlendarstellung bei Intervallrechnung. Tagung uber Intervallrechnung in Overwolfach, Tagungsbericht 12/68 (1968).
[108] E. R. Hansen, Interval arithmetic in matrix computations Part I. SIAM J. Numer. Anal. 2, 308—320 (1965).
[109] E. R. Hansen, On solving systems of equations using interval arithmetic. Math. Comp. 22, 374—384 (1968).
[110] E. R. Hansen, On the solution of linear algebraic equations with interval coefficients. Linear Algebra Appl. 2, 153—165 (1969).
[111] E. R. Hansen, «Topics in Interval Analysis». Oxford Univ. Press (Clarendon), London and New York, 1969.
[112] E. R. Hansen, On linear algebraic equations with interval coefficients. In «Topics in Interval Analysis». Oxford Univ. Press (Clarendon), London and New York, 1969.
[113] E. R. Hansen, On solving two-point boundary-value problems using interval arithmetic. In «Topics in Interval Analysis». Oxford Univ. Press (Clarendon), London and New York, 1969.
[114] E. R. Hansen, The centered form. In «Topics in Interval Analysis». Oxford Univ. Press (Clarendon)), London and New York, 1969.
[115] E. R. Hansen and R. Smith, Interval arithmetic in matrix computations, Part II. SIAM J. Numer. Anal. 4, 1—9 (1967).
[116] R. J. Hanson, Automatic error bounds for real roots of polynomials having interval coefficients. Comput. J. 13, 284—288 (1970).

[117] F. Hausdorff, «Mengenlehre». De Gruyter, Leipzig, 1927. — Рус. перев.: Хаусдорф Ф. Теория множеств. — М. — Л.: Объединенное науч.-техн. изд-во НКТП СССР, 1937, 304 л.
[118] M. Heidt, Zur numerischen lösung von gewöhnlichen Differentialgleichungen zweiter Ordnung. Ph. D. Thesis, Univ. Karlsruhe (1971).
[119] P. Henrici, Methods of search for solving polynomial equations. J. ACM 17, 273—283 (1970).
[120] P. Henrici, Circular arithmetic and the determination of polynomial zeros. Conf. Appl. Numer. Anal. Lecture Notes in Mathematics 228, pp. 86—92. Springer-Verlag, Berlin and New York, 1971.
[121] P. Henrici, see I. Gargantini.
[122] J. Herzberger, Metrische Eingenschaften von Mengensystemen und einige Anwendungen. Ph. D. Thesis, Univ. Karlsruhe (1969).
[123] J. Herzberger, Definition und Eigenschaften allgemeiner Intervallräume. Z. Angew. Math. Mech. 50, T50—T51 (1970).
[124] J. Herzberger, Intervallmäßige Auswertung von Standardfunktionen in ALGOL 60 *Computing* 5, 377—384 (1970).
[125] J. Herzberger, Zur Konvergenz untervallmäßiger Iterationsverfahren. Z. Angew. Math. Mech. 51, T56 (1971).
[126] J. Herzberger, Über optimale Verfahren zur numerischen Bestimmung reeller Nullstellen mit Fehlerschranken. Habilitationsschrifft, Univ. Karlsruhe (1972).
[127] J. Herzberger, Über ein Verfahren zur Bestimmung reeller Nullstellen mit Anwendung auf Parallelrechnung. Elektron. Rech. 14, 250—254 (1972).
[128] J. Herzberger, Über die Nullstellenbestimmung bei näherungsweise berechneten Funktionen. Computing 10, 23—31 (1972).
[129] J. Herzberger, Bemerkungen zu einem Verfahren von R. E. Moore Z. Angew. Math. Mech. 53, 356—358 (1973).
[130] J. Herzberger, see G. Alefeld.
[131] G. D. Hitchins, The numerical solution of polynomial equations with special reference to the determination of initial approximations to the roots. Ph. D. Thesis, Univ. of Leeds (1971).
[132] G. D. Hitchins. An interval arithmetic package and some applications. Technical Report 10, Centre for Computer Studies, Univ. of Leeds (1972).
[133] G. D. Hitchins, see J. A. Grant.
[134] A. S. Householder, «The Numerical Treatment of a Single Nonlinear Equation». McGraw-Hill, New York, 1970.
[135] S. Hunger, Intervallanalytische Defektabschätzung bei Anfangswertaufgaben für Systeme gewöhlnlicher Differentialgleichungen. GMD-Bericht No. 41. Bonn (1971).
[136] L. W. Jackson, Automatic error analysis for the solution of ordinary differential equations. Technical Report #28, Department of Computer Science, Univ. of Toronto (1971).
[137] K.-U. Jahn, Aufbau einer 3-wertigen linearen Algebra und affinen Geometrie auf der Grundlage der Intervallarithmetik. Ph. D. Thesis, Univ. Leipzig (1971).
[138] K.-U. Jahn, Eine Theorie der Gleichungssysteme mit Intervallkoeffizienten. Z. Angew. Math. Mech. 54, 405—412 (1974).
[139] K.-U. Jahn, Eine auf den Intervallzahlen fußenge 3-wertige lineare Algebra. Math. Nachr. 65, 105—116 (1975).
[140] P. Jarratt, see M. Dowell.
[141] W. M. Kahan, A computable error-bound for linear systems of ordinary differential equations. Abstract, SIAM Rev. 8, 568—569 (1966).
[142] W. M. Kahan, An ellipsoidal error-bound for linear systems of ordinary differential equations. Unpublished manuscript (1967).
[143] W. M. Kahan, A more complete interval arithmetic. Report, Univ. of Toronto (1968).

[144] K. Kansy, Ableitungsvertragliche Verallgemeinerung der Intervallpolynome. GMD-Bericht No. 70, Bonn (1973).
[145] D. L. Karney, see R. T. Gregory.
[146] V. Kartheus, Zur intervallanalytischen Behandlung linearer Gleichungssysteme. GMD-Mitteilungen No. 16, Bonn (1972).
[147] P. Katzan, Faktorzerleagung von Polynomen mit Fehlererfassung. GMD-Bericht No. 18, Bonn (1969).
[148] E. Kaucher, Uber metrische und algebraische Eigenschaften einiger beim numerischen Rechenen auftretender Raume. Ph. D. Thesis, Univ. Karlsruhe (1973).
[149] I. O. Kerner, Ein Gesamtschrittverfahren zur Berechung der Nullstellen von Polynomen. Numer. Math. 8, 290—294 (1966).
[150] R. Klatte, Anwendung einer Verallgemeinerung der Modifikation des Newtonverfahrens. Diplomarbeit, Inst. f. Angew. Math., Univ. Karlsruhe (1969).
[151] D. Klaua, Partiell definierte Mengen. Mber. Dt. Akad. Wiss. 10 (1968).
[152] D. Klaua, Partielle Mengen mit mehrwertigen Glundbeziehungen. Mber. Dt. Akad. Wiss. 11 (1969).
[153] D. Klaua, Partelle Mengen und Zahlen. Mber. Dt. Akad. Wiss. 11, 585—599 (1969).
[154] H.-O. Klein, Eine intervallanalytische Methode zur Lösung von Integralgleichungen mit mehrparametrigen Funktionen. GMD-Bericht No. 69, Bonn (1973).
[155] D. E. Knuth, «The Art of Computer Programming». Vol. 2. Addison — Wesley, Reading, Massachusetts, 1969. — Рус. перев.: Кнут Д. Искусство программирования для ЭВМ: В 3-х т. Т. 2. Получисленные алгоритмы. — М.: Мир, 1977. — 726 с.
[156] M. Kracht and G. Schröder, Eine Einführung in die Theorie der quasilinearen Räume mit Anwendung auf die in der Intervallrechung auftretenden Räume. Math.-Phys. Semesterber. 20, 226—242 (1973).
[157] M. Kracht and G. Schröder, Zur Intervallrechung in linearen Räumen. Computing 11, 73—79 (1973).
[158] R. Krawczyk, Iterative Verbesserung von Schranken für Eigenwerte und Eigenvektoren reeller Matrizen. Z. Angew. Math. Mech. 48, 780—783 (1968).
[159] R. Krawczyk, Fehlerabschätzung reeller Eigenwerte und Eigenvektoren von Matrizen. Computing 4, 281—293 (1969).
[160] R. Krawczyk, Newton-Algorithmen zur Bestimmung von Nullstellen mit Fehlerschranken. Computing 4, 187—201 (1969).
[161] R. Krawczyk, Verbesserung von Schranken für Eigenwerte und Eigenvektoren von Matrizen. Computing 5, 200—206 (1970).
[162] R. Krawczyk, Einschließung von Nullstellen mit Hilfe einer Interwallarithmetik. Computing 5, 356—370 (1970).
[163] R. Krawczyk, Gleichugen in halbgeordneten Rãumen. Abh. Math. Sem., Univ. Hamburg 36, 150—165 (1971).
[164] R. Krawczyk, Zur Einschließung inversen linearer Operatoren. Methoden Verfahren Math. Phys. 7, 1—19 (1972).
[165] R. Krawczyk, see N. Apostolatos.
[166] T. Kreifelts. Über vollautomatische Erfassung und Abschätzung von Rundungsfehlern in arithmetischen Prozessen. GMD-Bericht No. 62, Bonn (1972).
[167] O. Kress. Über Fehlerabschätzungen für Eigenwerte und Eigenvektoren reellsymmetrischer Matrizen. Ph. D. Thesis, Univ. Karlsruhe (1973).
[168] N. Krier. Komplexe Kreisarithmetik. Ph. D. Thesis, Univ. Karlsruhe (1973).
[169] F. Krückeberg, Numerische Intervallrechnung und deren Anwendung, Bonn (1966).
[170] F. Krückeberg, Inversion von Matrizen mit Fehlererfassyng. Z. Angew. Math. 46, T69—T71 (1966).

[171] F. Krückeberg, Ordinary differential equations. In «Topics in Interval Analysis», Oxford Univ. Press (Clarendon), London and New York, 1969.
[172] F. Krückeberg, Bemerkungen zur Intervallanalysis. Apl. Mat. 13, 152—153 (1968).
[173] F. Krückeberg, Partial differential equations. In «Topics in Interval Analysis», Oxford Univ. Press (Clarendon), London and New York, 1969.
[174] U. Kulisch, Grundlagen einer Maschinenintervallarithmetik. Z. Angew. Math. Mech. 47, T81 (1967).
[175] U. Kulisch, Grundzüge der Intervallrechnung. In «Überbliche Mathematik», Vol. 2, Bibliograph. Inst. Mannheim, 1969.
[176] U. Kulisch, Rounding invariant structures. MRC Technical summary Rep. 1103, Univ. of Wisconsin, Madison, Wisconsin (1970).
[177] U. Kulisch, Interval arithmetic over completely ordered ringoids. MRC Technical Summary Rep. 1105, Univ. of Wisconsin, Madison, Wisconsin (1970).
[178] U. Kulisch, An axiomatic approach to rounded computations. Numer. Math. 18, 1—17 (1971).
[179] U. Kulisch, Axiomatik des numerischen Rechnens. Z. Angew. Math. 52, 7211 (1972).
[180] U. Kulisch. On the concept of a creen. Z. Angew. Math. Mech. 53, 115—119 (1973).
[181] U. Kulisch, Implementation and formalization of floatingpoint arithmetic. Caratheodogy Symp., Athen (1973).
[182] U. Kulisch, see N. Apostolatos.
[183] I. B. Kupermann, Approximate linear algebraic equations and rounding error estimation. Ph. D. Thesis, Dept. of Applied Math., Univ. of Witwatersrand, Johannesburg, South Africa (1967).
[184] I. B. Kupermann, «Approximate Linear Algebraic Equations», Van Nostrand Reinhold, New York, 1971.
[185] I. Kupka, Simulation reeller Arithmetik und reeller Funktionen in endlichen Mengen. Numer. Math. 17, 143—152 (1971).
[186] P. Lancaster, see J. Rokne.
[187] D. Leder, see J. W. Schmidt.
[188] H. Leonhardt, see J. W. Schmidt.
[189] R. L. London, see D. J. Good.
[190] B. Lortz, Eine Langzahlarithmetik mit optimaler einseitiger Rundung. Ph. D. Thesis, Univ. Karlsruhe (1971).
[191] B. Lortz, see N. Apostolatos.
[192] F. R. Loscalzo, see R. H. Dargel.
[193] B. Machost, Numerische Behandlung des Simplexverfahrens mit intervallanalytischen Methoden. GMD-Bericht No. 30, Bonn (1970).
[194] K. Madsen, On the solution of nonlinear equations in interval arithmetic. BIT 13, 428—433 (1973).
[195] O. Mayer, Über die in der Intervallrechnung auftretenden Räume und einige Anwendungen. Ph. D. Thesis, Univ. Karlsruhe (1968).
[196] O. Mayer, Über Eigenschaften von Intervallprozessen. Z. Angew. Math. Mech. 48, T86—T87 (1968).
[197] O. Mayer, Algebraische und metrische Strukturen in der Intervallrechnung und einige Anwendungen. Computing 5, 144—162 (1970).
[198] O. Mayer, Über die Bestimmung von Einschießungsmengen für die Lösungen linearer Gleichungssysteme mit fehlerbehafteten Koenffizienten. Elektron. Datenverarbeit. 12, 164—167 (1970).
[199] O. Mayer, Über die itervallmaßige Durchfuhrung einiger Iterationsverfahren. Z. Angew. Math. Mech. 50, 765—T66 (1970).
[200] O. Mayer, Über eine Klasse komplexer Intervallgleichungssysteme mit iterationsfähiger Gestalt. Computing 6, 104—106 (1970).

[618] J. G. Rokne. Low Complexity k-dimensional Centered Forms. Computing 37, 247—253 (1986).
[619] J. G. Rokne, T. Wu, The Circular Complex Centered Form. Computing 28, 17—30 (1982).
[620] J. Rokne, T. Wu, A Note on the Circular Complex Centered Form. Computing 30, 201—211 (1983).
[621] S. M. Rump, See E. Kaucher.
[622] W. Sauer, see H. Ratschek.
[623] H. Schwandt, Krawczyk-like Algorithms for the Solution of Systems of Nonlinear Equations. SIAM J. Numer. Anal. 22. 792—810 (1985).
[624] H. Schwandt, Accelerating Krawczyk-like Interval Algorithms for the Solution of Nonlinear Systems of Equations by Using Second Derivatives. Computing 35, 355—367 (1985).
[625] H. Schwandt, Almost Globally Convergent Interval Methods for Discretizations of Nonlinear Elliptic Partial Differential Equations. SIAM J. Numer. Anal. 23, 304—324 (1986).
[626] J. M. Shearer, M. A. Wolfe, Some Algorithms for the Solution of a Class of Nonlinear Algebraic Equations. Computing 35, 63—72 (1985).
[627] F. S. Sisser, Inverting an Interval Hessian of a Factorable Function. Computing 29, 63—72 (1982).
[628] H. Spreuer, E. Adams, Zur Wahrscheinlichkeitsverteilung in intervallarithmetischen Ausdrücken. Z. Angew. Math. Mech. 65, T325—T326 (1985).
[629] H. Spreuer, see G. Alefeld.
[630] L. V. Stefanović, see M. S. Petković.
[631] P. Wildenauer, A New Method for Automatical Computation of Error Bounds for the Set of All Solutions of Nonlinear Boundary Value Problems. Computing 35, 131—154 (1985).
[632] J. Wolff v. Gudenberg, see R. Lohner.
[633] T. Wu, see J. Rokne.
[634] Shen Zuhe, see N. S. Asaithambi.

КНИГИ ПО ИНТЕРВАЛЬНОЙ МАТЕМАТИКЕ [1])

1. Калмыков С. А., Шокин Ю. И., Юлдашев З. Х. Методы интервального анализа. — Новосибирск: Наука, 1986. — 224 с.
2. Назаренко Т. И., Марченко Л. В. Введение в интервальные методы вычислительной математики: Учеб. пособие. — Иркутск: Издательство Иркутского университета, 1982. — 108 с.
3. Шокин Ю. И. Интервальный анализ. — Новосибирск: Наука, 1981.— 112 с.
4. A new approach to scientific computation/Ed.: U. W. Kulisch, W. L. Miranker. — New York etc.: Academic Press, 1983. — xv, 384 p.
5. Alefeld G., Herzberger J. Introduction to interval computations. — New York etc.: Academic Press, 1983. — xviii, 333 p. Рус. перев.: настоящая книга.
6. Berti S. N. Aritmetic Intervalelor = Arithmetic of intervals = Арифметика интервалов. — Bucureşti: Editura Academiei Republicii Socialiste România, 1977. — 201 p. — (Teoria calculului, analiza numerică şi informatică; N 2).
7. Fundamentals of numerical computation (computer-oriented numerical analysis)/Ed.: G. Alefeld, R. D. Grigorieff. — Wien etc.: Springer-Verlag, 1980. — 230 p. — (Computing; Suppl. 2)
8. Interval and fuzzy mathematics: Proc. Polish Symp., Technical Univ. of Poznan, Inst. of Mathematics, Poland, Aug. 1983/Ed.: J. Albricht, H. Wiisniewski. — Poznan: Technical Univ. of Poznan, 1985.

---

[1]) Этот список был подготовлен переводчиком А. Г. Яковлевым. — *Прим. ред.*

9. Interval mathematics/Ed.: K. Nickel. — New York etc.: Springer Verlag, 1975. — vi, 331 p. — (Lecture notes in computer science; Vol. 29)
10. Interval mathematics 1980/Ed.: K. Nickel. — New York etc.: Academic Press, 1980. — XV, 554 p.
11. Interval mathematics 1985 Ed.: K. Nickel. — New York etc.: Springer Verlag, 1986. — 227 p. — (Lecture notes in computer science; Vol. 212)
12. Kaucher E. W., Miranker W. L. Self validating numerics for function space problems. Computation with guarantees for differential and integral equations. — Orlando (Fla.): Academic Press, 1984. — xiii, 255 p. — Рец.: H. J. S.//Mathematics of Computation. — 1986. — Vol. 47, N 175. — P. 369—370
13. Kulisch N. W., Miranker W. L. Computer arithmetic in theory and practice. — New York etc.: Academic Press, 1981. — xiii, 251 p. — Рец.: Ступин Ю. В.//Новые книги за рубежом. Серия А. — 1982. — N 7. — С. 46—49; Rall L. B. Book review of «Computer arithmetic in theory and practice» by U. W. Kulisch and W. L. Miranker//SIAM Rev. — 1983. — Vol. 25.— P. 585—588
14. Moore R. E. Interval analysis. — Englewood Cliffs: Prentice Hall, 1966. — 145 p.
15. Moore R. E. Methods and applications of interval analisys. — Philadelfia: SIAM, 1979, — xi, 190 p. — Рец.: Матиясевич Ю. В.//Новые книги за рубежом. Серия А. — 1984. — N 7. — С. 38—42
16. Moore R. E. Computational functional analysis. — Chichester; New York: Halsted Press: Wiley, 1985. — 156 p. — (Ellis Horwood Series, Mathematics and its Applications). — Рец.: Fink J. P.//Mathematics of Computation. — 1986. — Vol. 47, N 175. — P. 372—374
17. Ratschek H., Rokne J. Computer methods for the range of functions. — Chichester: Horwood, 1984. — 168 p. — Рец.: Курякин В. Ф.//Новые книги за рубежом. Серия А. — 1985. — N 12. — С. 30—32
18. Topics in interval analysis/Ed.: Hansen E. — Oxford: Clarendon Press, 1969. — viii, 130 p.

ОГЛАВЛЕНИЕ

Предисловие редактора перевода 5
Предисловие к английскому изданию 8
Предисловие к немецкому изданию 9

Глава 1. Вещественная интервальная арифметика 15
Глава 2. Дальнейшие понятия и свойства 23
Глава 3. Интервальное оценивание и множество значений в случае вещественных функций 31
Глава 4. Машинная интервальная арифметика 50
Глава 5. Комплексная интервальная арифметика 61
A. Прямоугольники в качестве комплексных интервалов . . . 61
B. Круги в качестве комплексных интервалов 64
Глава 6. Метрика, абсолютная величина и ширина в $I(\mathbb{C})$ 69
Глава 7. Локализация нулей функций одной вещественной переменной 75
A. Методы ньютоновского типа 76
B. Определение оптимального метода 80
C. Квадратично сходящиеся методы 84
D. Методы более высоких порядков 90
E. Интерполяционные методы 95
Глава 8. Методы одновременной локализации вещественных корней многочленов 109
Глава 9. Методы одновременной локализации комплексных корней многочленов 121
Глава 10. Операции над интервальными матрицами 126
Глава 11. Итерационная локализация неподвижной точки для систем нелинейных уравнений 137
Глава 12. Системы линейных уравнений, поддающиеся методу итерации 148
Глава 13. Методы релаксации 164
Глава 14. Оптимальность симметрического короткошагового метода со взятием пересечения на каждом шаге 171
Глава 15. О применимости метода Гаусса к системам уравнений с интервальными коэффициентами 183
Глава 16. Метод Хансена 196
Глава 17. Процедура Купермана и Хансена 204
Глава 18. Итерационные методы для локализации обратной матрицы и разложения на треугольные 207
Глава 19. Методы ньютоновского типа для систем нелинейных уравнений 225
Глава 20. Методы ньютоновского типа, не использующие обращения матриц 259

Глава 21. Методы ньютоновского типа для частных типов систем нелинейных уравнений 264
Глава 22. Полношаговые и короткошаговые методы ньютоновского типа 277
Приложение А. Порядок сходимости итерационных методов в $V_n((IC))$ и $M_{mn}(I(C))$ 286
Приложение В. Реализация машинной интервальной арифметики на Алголе 60 291
Приложение С. Процедуры на Алголе 296
A. Локализация собственных значений 296
B. Короткошаговый метод 300
C. Обращение матрицы 303
Литература 310
Дополнительная литература к английскому изданию 323
Дополнительная литература к русскому изданию 331
Книги по интервальной математике 334
Добавление. Интервальные вычисления на ЭВМ. *А. Г. Яковлев* 336
Предметный указатель 353