Высшее профессиональное образование
БАКАЛАВРИАТ

Е. Н. ЖИДКОВ

ВЫЧИСЛИТЕЛЬНАЯ МАТЕМАТИКА

Учебник

Допущено
Учебно-методическим объединением вузов
по университетскому политехническому образованию
в качестве учебника для студентов высших учебных
заведений, обучающихся по направлениям подготовки
«Информатика и вычислительная техника»
и «Информационные системы и технологии»

2-е издание, переработанное

ACADEMIA

Москва
Издательский центр «Академия»
2013

УДК 519.6(075.8)
ББК 22.19я73
Ж696

Рецензенты:
проф. кафедры «Нелинейный анализ и оптимизация»
Российского университета дружбы народов, д-р физ.-мат. наук *Е.Б.Ланеев*;
зав. кафедрой «Математическое моделирование» Московского
государственного технического университета им. Н.Э.Баумана,
д-р физ.-мат. наук, проф. *Л.П.Крищенко*;
зав. кафедрой «Высшая математика» Московского государственного
университета леса, проф., канд. физ.-мат. наук *К.К.Рыбников*

Жидков Е.Н.

Ж696 Вычислительная математика: учебник для студ. учреждений высш. проф. образования / Е.Н.Жидков. — 2-е изд., перераб. — М.: Издательский центр «Академия», 2013. — 208 с. — (Сер. «Бакалавриат»).

ISBN 978-5-4468-0222-7

Учебник создан в соответствии с Федеральным государственным образовательным стандартом по направлениям подготовки «Информатика и вычислительная техника» и «Информационные системы и технологии» (квалификация «бакалавр»).

В учебнике рассмотрены вопросы применения численных методов к решению стандартных задач математического анализа и дифференциальных уравнений, в частности, основы теории погрешностей, численные методы линейной алгебры, решение систем нелинейных уравнений, теория интерполяции, численное дифференцирование и интегрирование, аппроксимация функций, решение дифференциальных уравнений.

Для студентов учреждений высшего профессионального образования.

УДК 519.6(075.8)
ББК 22.19я73

ISBN 978-5-4468-0222-7

ПРЕДИСЛОВИЕ

В учебнике изложены основные численные методы решения прикладных задач. Книга написана на основе лекций, читаемых автором в течение последних лет на факультете «Радиоэлектроника и лазерная техника» Московского государственного технического университета им. Н. Э. Баумана.

Книга состоит из введения и семи глав.

Во введении излагаются вспомогательные материалы из курсов анализа и алгебры.

Первая глава посвящена численным методам алгебры: методам решения линейных систем и нахождения собственных чисел матрицы.

Во второй главе описаны методы приближения функций: построение многочленов Лагранжа, Ньютона и Эрмита, сплайн-интерполяция, а также численное дифференцирование.

Третья глава посвящена методам численного интегрирования — вычислению определенных интегралов, несобственных интегралов и интегралов от быстроосциллирующих функций.

Четвертая глава описывает численные методы решения нелинейных уравнений и систем.

В пятой главе рассмотрены методы приближения в нормированных пространствах — равномерные и среднеквадратичные приближения.

Шестая глава посвящена методам решения задачи Коши для обыкновенных дифференциальных уравнений: Рунге — Кутты, Адамса, прогноза и коррекции.

В седьмой главе представлены методы решения краевой задачи для обыкновенных дифференциальных уравнений: методам прогонки, вариационные, а также методы решения нелинейных краевых задач — пристрелки, квазилинеаризации, разностные.

В книге приняты следующие обозначения:
! — замечания;
■ — конец доказательства теоремы.

Учебник предназначен для студентов технических вузов, аспирантов и преподавателей. Может быть также полезен инженерам и научным работникам, использующим в практической деятельности численные методы.

ВВЕДЕНИЕ

В.1. Символы O и o

Пусть x — действительная переменная. В дальнейшем изложении поведение интересующей нас функции $f(x)$ при $x \to x_0$ будет получать с помощью более простой функции $\varphi(x)$, наиболее часто степенной.

Определение В.1. Символ O (читается «о большое»). Будем говорить, что функция $f(x)$ является ***о большое от*** $\varphi(x)$ при $x \to x_0$, если существует такая постоянная $M > 0$ и такое число $\delta(\varepsilon) > 0$, что при $|x - x_0| < \delta(\varepsilon)$ справедливо:

$$|f(x)| \le M|\varphi(x)|. \qquad \text{(В.1)}$$

Этот факт коротко записывают

$$f(x) = O(\varphi(x)). \qquad \text{(В.2)}$$

Если при прочтении исключены недоразумения, то пишут: $f = O(\varphi)$. В частности, запись $f(x) = O(1)$ означает ограниченную функцию.

Аналогично можно определить понятие «о большое» при $x \to \infty$, $x \to +\infty$, $x \to -\infty$, $x \to x_0 + 0$, $x \to x_0 - 0$.

Определим это понятие, например, при $x \to \infty$.

Определение В.2. Будем говорить, что функция $f(x)$ является ***о большое от*** $\varphi(x)$ при $x \to \infty$, если существует такая постоянная $M > 0$ и такое число $N > 0$, что при $|x| > N$ справедливо:

$$|f(x)| \le M|\varphi(x)|, \qquad \text{(В.3)}$$

что также записывается

$$f(x) = O(\varphi(x)). \qquad \text{(В.4)}$$

Примеры:

а) $\sin x = O(1)$ при $x \to x_0$. Очевидно, что $|\sin x| \le 1$, $M = 1$, $\varphi(x) = 1$;

б) $\sin x = O(x)$ при $x \to 0$. Это следует из неравенства $|\sin x| \le x$, $M = 1, \varphi(x) = x$;

в) $\operatorname{sh} x = O(e^x)$ при $x \to \infty$.

Так как $\operatorname{sh} x$ — нечетная функция, то достаточно доказать это утверждение для $x \to +\infty$:

$$|\operatorname{sh} x| = \left|\frac{e^x - e^{-x}}{2}\right| = \frac{1}{2}e^x\left|1 - e^{-2x}\right| \le \frac{1}{2}e^x,\ M = \frac{1}{2},\ \varphi(x) = e^x.$$

***Свойства символа O*:**

1) $f = O(f)$. Для обоснования этого свойства положим в (В.1) $M = 1$;

2) если $f = O(\varphi)$ и $\varphi = O(g)$, то $f = O(g)$. В самом деле $|f| \le M|\varphi|$, $|\varphi| \le L|g|$, поэтому

$$|f| \le M|\varphi| \le LM|g|.$$

! Из $f = O(\varphi)$ не следует $\varphi = O(f)$.

Примеры: $\sin x = O(1)$, но $1 \ne O(\sin x)$;

! Равенство, содержащее символ «о большое», должно читаться только слева направо;

3) из неравенства (В.1) следует, что функция $f(x)$ должна иметь в окрестности точки x_0 те же нули, что и $\varphi(x)$. Если это не так, то неравенства (В.1) не может быть;

4) если в окрестности точки x_0 функция $\varphi(x) \ne 0$, то можно дать модификацию определения В.1: функция $f(x) = O(\varphi(x))$, если

$$\lim_{x \to x_0}\left|\frac{f(x)}{\varphi(x)}\right| = c > 0. \qquad \text{(В.5)}$$

В качестве константы M можно взять любое число, большее, чем c.

Отметим, что обратное утверждение неверно, т. е. если справедливо (В.1), то необязательно, что существует предел (В.5). Например, можно привести $\sin x = O(1)$;

5) для данной функции $f(x)$ можно установить бесконечно много соотношений вида (В.1), в частности если $f(x) = O(\varphi(x))$, то $f(x) = O(c\varphi(x))$, $c \ne 0$;

6) функции $f(x)$ и $\varphi(x)$ не обязаны принадлежать одному и тому же классу гладкости; они могут быть разрывными или же в некоторой части области определения обращаться в нуль.

Будем теперь сравнивать две бесконечно малые функции $f(x)$ и $\varphi(x)$.

Определение В.3. Символ *o* (читается «о малое»). Если для любого $\varepsilon > 0$ существует число $\delta(\varepsilon) > 0$, такое, что при $|x - x_0| < \delta$ справедливо

$$|f(x)| \le \varepsilon |\varphi(x)|, \qquad \text{(B.6)}$$

то пишут

$$f(x) = o(\varphi(x)) \qquad \text{(B.7)}$$

и говорят: $f(x)$ ***есть о малое от*** $\varphi(x)$ ***при*** $x \to x_0$.

Если при прочтении исключены недоразумения, то пишут $f = o(\varphi)$.

В частности, запись $f(x) = o(1)$ означает, что $f(x) \to 0$ при $x \to x_0$.

Аналогично вводится понятие *о* малое при $x \to \infty$.

Определение В.4. Если для любого $\varepsilon > 0$ существует число N, такое, что при $|x| > N$ справедливо

$$|f(x)| \le \varepsilon |\varphi(x)|, \qquad \text{(B.8)}$$

то пишут

$$f(x) = o(\varphi(x)) \qquad \text{(B.9)}$$

и говорят: $f(x)$ ***есть о малое от*** $\varphi(x)$ ***при*** $x \to \infty$.

Примеры:

а) $\dfrac{1}{x^2} = o\left(\dfrac{1}{x}\right)$ при $x \to \infty$;

б) $e^{-x} = o(x^{\alpha})$, $-\infty < \alpha < 0$ при $x \to +\infty$;

в) $1 - \cos x = o(x)$ при $x \to 0$.

Свойства символа о:

1) из $f = o(\varphi)$ и $\varphi = o(g)$ следует $f = o(g)$.

$f = o(f)$ имеет место тогда и только тогда, когда $f \equiv 0$;

! Равенство, содержащее символ «о малое», должно читаться только слева направо;

2) из неравенства (В.8) следует, что функция $f(x)$ должна иметь те же нули, что и $\varphi(x)$. Если, как и в случае определения В.1, функция $\varphi(x)$ не имеет нулей при $x \to x_0$, то неравенство можно переписать в виде

$$\lim_{x \to x_0} \frac{f(x)}{\varphi(x)} = 0; \qquad \text{(B.10)}$$

3) из определений В.1 и В.3 следует, что $o(\varphi(x)) = O(\varphi(x))$;

4) соотношение (В.9) никогда не указывает точный порядок функции $f(x)$. Это соотношение всегда можно уточнить, построив такую функцию $\varphi_1(x)$, что $f(x) = o(\varphi_1(x))$ и $\varphi_1(x) = o(\varphi(x))$. Для этого достаточно взять $\varphi_1(x) = \left(|f(x)||\varphi|\right)^{1/2}$.

Часто приходится выполнять арифметические действия с функциями, связанными соотношениями o и O. Если рассматривать сумму нескольких функций, то, строго говоря, при каждом символе надо ставить свой индекс, так как каждому символу соответствует своя функция. Но если не возникает никаких недоразумений, то индексами обычно пренебрегают и ради удобства для разных функций применяют один и тот же символ порядка.

Например, будем писать $O(\varphi(x))+O(\varphi(x))=O(\varphi(x))$, хотя, строго говоря, надо было бы писать так: $O_1(\varphi)+O_2(\varphi)=O_3(\varphi)$.

Опираясь на определения В.1 и В.3, легко доказать следующие формулы, читаемые слева направо:

1) $o(\varphi(x))=O(\varphi(x))$, $O(O(\varphi(x)))=O(\varphi(x))$, $o(o(\varphi(x)))=o(\varphi(x))$, $O(o(\varphi(x)))=o(\varphi(x))$, $o(O(\varphi(x)))=o(\varphi(x))$;

2) $O(\varphi)+O(\varphi)=O(\varphi)$, $o(\varphi)+o(\varphi)=o(\varphi)$, $O(\varphi)+o(\varphi)=O(\varphi)$;

3) $O(\varphi_1)O(\varphi_2)=O(\varphi_1\varphi_2)$, $o(\varphi_1)o(\varphi_2)=o(\varphi_1\varphi_2)$, $O(\varphi_1)o(\varphi_2)=o(\varphi_1\varphi_2)$;

4) $O(\varphi_1)+O(\varphi_2)=O(\varphi_1+\varphi_2)$, $o(\varphi_1)+o(\varphi_2)=o(|\varphi_1|+|\varphi_2|)$, $O(\varphi_1)+o(\varphi_2)=O(|\varphi_1|+|\varphi_2|)$;

5) в частном случае в качестве одного множителя можно применять обычное обозначение функции, тогда

$h(x)O(\varphi)=O(h(x)\varphi(x))$, $h(x)o(\varphi)=o(h(x)\varphi(x))$.

Если $h(x)=c\neq 0$, то

$cO(\varphi)=O(\varphi(x))$, $co(\varphi)=o(\varphi(x))$, $O(\varphi(x))=O(c\varphi)$, $o(\varphi)=o(c\varphi(x))$.

Отметим, что формулы 1 — 5 соответствующим образом легко переносятся на конечное число слагаемых и множителей; при дополнительных условиях допускается также бесконечное число слагаемых.

Нас особо будет интересовать случай $x\to 0$ и степенные функции x^α, $\alpha>0$. В этом случае формулы 2 — 4 можно переписать следующим образом:

2′) $O(x^\alpha)+O(x^\alpha)=O(x^\alpha)$;

3′) $O(x^\alpha)O(x^\beta)=O(x^\alpha x^\beta)=O(x^{\alpha+\beta})$;

4′) $O(x^\alpha)+O(x^\beta)=O(x^\alpha)$ при $0<\alpha\leq\beta$.

В.2. Разложение по формуле Тейлора функций одной и нескольких переменных

Как известно, если функция имеет $n+1$ непрерывную производную на отрезке $[a, b]$, то ее можно разложить по формуле Тейлора [16]:

$$f(x+h)=\sum_{k=0}^{n}\frac{d^k f(x)}{dx^k}\frac{h^k}{k!}+\frac{d^{n+1}f(x+h\xi)}{dx^{n+1}}\frac{h^{n+1}}{(n+1)!},\ \xi\in[0,1]. \qquad \text{(В.11)}$$

В случае если функция имеет бесконечное число производных, то ее можно разложить в ряд Тейлора:

$$f(x+h)=\sum_{k=0}^{\infty}\frac{d^k f(x)}{dx^k}\frac{h^k}{k!}. \qquad \text{(B.12)}$$

При этом существует число $R\geq 0$, такое, что ряд (B.12) сходится для всех $h, |h|<R$. Очевидно, что при $h=0$ ряд (B.12) сходится.

Используя формулу (B.11), можно получить аналогичную формулу для функции двух переменных.

Пусть функция $f(x,y)$ имеет непрерывные производные $\frac{\partial^n f(x,y)}{\partial x^2 \partial y^{n-k}}$ при $k\leq n\leq m$ в области $D=\{[a,b]\times[c,d]\}$. Требуется найти разложение этой функции по формуле Тейлора в произвольной точке $M_0=(x,y)$ области D по степеням приращений h по x и H по y.

Пусть $M_1=(x+h,y+H)$. Соединим точки M_0 и M_1 отрезком прямой так, что точка M на этой прямой имеет параметризацию $M=\lambda M_1+(1-\lambda)M_0$, $\lambda\in[0,1]$.

Обозначим

$$F(\lambda)=f[\lambda(x+h)+(1-\lambda)x,\lambda(y+H)+(1-\lambda)y]=f(x+\lambda h,y+\lambda H).$$

Очевидно, что функция $F(\lambda)\in C^{n+1}[0,1]$. Разложим ее в точке 0 по формуле Тейлора по степеням λ:

$$F(\lambda)=\sum_{k=0}^{n}\frac{d^k F(0)}{d\lambda^k}\frac{\lambda^k}{k!}+\frac{d^{n+1}F(\xi)}{d\lambda^{n+1}}\frac{\lambda^{n+1}}{(n+1)!},\ \xi\in[0,1].$$

Так как

$$\frac{d^k F(0)}{d\lambda^k}=\sum_{j=0}^{k}C_k^j\frac{\partial^k f(x,y)}{\partial x^j\partial y^{k-j}}h^j H^{k-j},$$

то

$$f(x+h,y+H)=\sum_{k=0}^{n}\frac{1}{k!}\sum_{j=0}^{k}C_k^j\frac{\partial^k f(x,y)}{\partial x^j\partial y^{k-j}}h^j H^{k-j}+$$
$$+\frac{1}{(n+1)!}\sum_{j=0}^{n+1}C_{n+1}^j\frac{\partial^{n+1}f(x+\xi h,y+\xi H)}{\partial x^j\partial y^{n+1-j}}h^j H^{n+1-j},\ \xi\in[0,1] \qquad \text{(B.13)}$$

Формулу (B.13) можно обобщить на случай нескольких переменных. Впоследствии нам потребуется разложение функции до второго порядка

$$f(x_1+h_1,\ \ldots,x_n+h_n)=f(x_1,\ \ldots,h_n)+\sum_{j=0}^{n}\frac{\partial f(x_1,\ \ldots,x_n)}{\partial x_j}h_j+$$
$$+\frac{1}{2}\sum_{k=0}^{n}\sum_{j=0}^{n}\frac{\partial^2 f(x_1+\xi h_1,\ \ldots,x_n+\xi h_n)}{\partial x_j\partial x_k}h_j h_k. \qquad \text{(B.14)}$$

В.3. Погрешность результата численного решения задачи

В.3.1. Источники погрешности при математических вычислениях

При решении практических задач возникают следующие виды погрешности:

1) *неустранимая погрешность* — исходные данные для решения задачи являются приближенными;

2) *погрешность метода* — математическое описание задачи является приближенным. Такое описание называется *математической моделью задачи*. Для решения задачи с помощью данной математической модели ее приходится заменять другой, более простой, но более удобной для вычислений;

3) *вычислительная погрешность* — при вводе данных в ЭВМ производится отбрасывание лишних разрядов, при выполнении арифметических действий — округление.

Более подробно последний вид погрешности будет рассмотрен в следующем пункте.

Иногда под неустранимой погрешностью понимают ту часть погрешности, которая возникает при введении исходных данных в ЭВМ. Погрешность же, обусловленная несоответствием математической модели реальности, называется *погрешностью модели*.

В.3.2. Представление чисел в ЭВМ. Погрешность арифметических действий

В отличие от классической математики, которая оперирует с бесконечными дробями и производит точные вычисления, при вычислении на ЭВМ приходится пользоваться числами с конечным числом разрядов, что накладывает свою специфику на результат вычисления.

В ЭВМ любое число представляется в виде

$$x = \theta m 2^p, \qquad \text{(В.15)}$$

где m — мантисса — двоичное число из диапазона $0 \le m < 1$, имеющее n разрядов; θ — знак числа x; p — порядок числа x — двоичное число, имеющее k разрядов.

Мантисса m должна быть нормализована, т.е. если $m \neq 0$, то старший разряд мантиссы не равен нулю.

Покажем, как перевести число в двоичную систему. Представим в машинном виде число 7,0625:

• запишем его в двоичном виде. Целая часть числа равна 7; разделим ее на 2; частное равно 3, остаток — 1. Следовательно,

последний разряд целой части в двоичном представлении будет равен 1;

• разделим 3 на 2; частное равно 1, остаток — 1. Так как оба результата меньше 1, то дальнейшее деление прекращаем. В результате получим $7_{10} = 111_2$.

> **!** Индекс, стоящий справа от числа, означает основание системы счисления.

Для нахождения дробной части поступим следующим образом:

• умножим 0,0625 на 2: $0{,}0625 \cdot 2 = 0{,}125$. Таким образом, первый двоичный разряд мантиссы после запятой равен 0;

• умножим теперь 0,125 на 2: $0{,}125 \cdot 2 = 0{,}25$. Второй двоичный разряд — 0;

• умножим 0,25 на 2: $0{,}25 \cdot 2 = 0{,}5$. Очередной разряд — 0;

• поступая также далее, получим $0{,}5 \cdot 2 = 1$. Следующий разряд — 1. Таким образом, $7{,}0625_{10} = 111{,}0001_2$.

После нормализации получим следующий вид числа: $111{,}0001 = 0{,}1110001 \cdot 2^3$. Порядок полученного числа равен 3.

Так как на каждый из параметров в ЭВМ отводится определенное количество разрядов, то не всякое действительное число из допустимого диапазона можно представить точно в виде (B.15), и это приводит к особенностям при выполнении арифметических вычислений.

Для пояснения воспользуемся грубой моделью. Пусть на мантиссу m отведено три двоичных разряда, а на порядок — один, т. е. число представляется в следующем виде:

Знак числа	Мантисса			Знак порядка	Порядок

Поскольку мы предположили, что мантисса содержит три разряда, то при вычислениях остаток либо округляется, либо отбрасывается. Для простоты будем считать, что он отбрасывается.

Перечислим все двоичные числа, представимые с трехзначной мантиссой (порядок p равен 0):

1	0	0	0,5
1	0	1	5/8 = 0,625
1	1	0	3/4 = 0,75
1	1	1	7/8 = 0,875

Для получения чисел при $p = 1$ умножим результат на 2. Следовательно, при $p = 1$ имеем значения 1; 1,25; 1,5; 1,75.

Аналогично, при $p = -1$ получаем 0,25; 0,3125; 0,375; 0,4375.

Поэтому все допустимые положительные числа, которые можно точно отобразить таким образом, заключены в следующей таблице:

1/4	5/16	3/8	7/16	1/2	5/8	3/4	7/8	1	5/4	3/2	7/4

Числа, по модулю меньшие 1/4, называются машинным 0. Числа, по модулю бóльшие 7/4, считаются бесконечно большими. Следовательно, число 7,0625 считается бесконечно большим.

Остальные числа отображаются лишь приближенно. Пусть, например, требуется записать в память ЭВМ $\sqrt{2}$. В результате будем иметь $\sqrt{2} = 5/4$.

Все числа из диапазона (1/2, 5/8) будут изображаться числом 1/2.

Отсюда вытекают особенности машинной арифметики:

$$-0{,}8 + 0{,}9 = 1/8;$$
$$5/4 \cdot 3/2 = \infty;$$
$$1/4 \cdot 5/16 = 0.$$

Предположим, что мантисса состоит из четырех десятичных разрядов.

Задача В.1. Требуется найти сумму десяти чисел [2]:

0,2897; 0,4976; 2,488; 7,259; 16,38; 62,49; 216,2; 523,3; 1403; 5291.

Точное значение суммы равно 7522,9043.

Вычислим ее теперь в четырехразрядной сетке сначала слева направо, учитывая, что пятый разряд отбрасывается. Сумма равна 7522.

Вычислим сумму справа налево и получим 7520.

В обоих случаях получен неточный результат. Для того чтобы оценить результат вычислений, введем понятия абсолютной и относительной погрешности.

Определение В.5. Абсолютная величина разности точного x и приближенного x^* значений называется ***абсолютной погрешностью***, $A = |x - x^*|$.

> **!** Так как нам неизвестно точное значение, то в качестве абсолютной погрешности берется оценка этой величины.

Определение В.6. Отношение абсолютной погрешности к приближенному значению называется ***относительной погрешностью***, $\delta = A/|x^*|$.

Вычислим абсолютную и относительную погрешности результата. В первом случае абсолютная погрешность равна $A_1 = 0,9043$, относительная $\delta_1 = 0,0001202$; во втором $A_2 = 2,9043$, $\delta_2 = 0,000386$.

Относительная погрешность меньше при суммировании от меньшего числа к большему.

Отсюда вытекают ***правила действий с числами***:

1) при сложении или вычитании последовательности чисел нужно суммировать их в порядке возрастания по модулю;

2) следует, по возможности, избегать вычитания почти одинаковых чисел;

3) выражение $a(b-c)$ можно записать в виде $ab-ac$, а $(b-c)/a$ — $b/a-c/a$. Если числа b и c почти равны друг другу, то нужно выполнить вычитание до умножения и деления;

4) рекомендуется свести к минимуму количество арифметических операций.

Продемонстрируем применение перечисленных правил, решив квадратное уравнение $x^2 + 0,4002x + 0,00008 = 0$.

Точные значения корней $x_1 = -0,4$ и $x_2 = -0,0002$.

Формула для решения этого уравнения имеет вид

$$x_{12} = \frac{-0,4002 \pm \sqrt{(0,4002)^2 - 0,00032}}{2}.$$

Вычислим оба корня при длине мантиссы, равной 4:

$$x_{12} = \frac{-0,4002 \pm \sqrt{0,1601 - 0,00032}}{2} =$$

$$= \frac{-0,4002 \pm \sqrt{0,1597}}{2} = \frac{-0,4002 \pm 0,3996}{2}.$$

Таким образом, приближенные значения корней равны:

$$x_1 = -0,3999 \text{ и } x_2 = -0,0003.$$

Абсолютные погрешности корней соответственно $A_1 = 0,0001$ и $A_2 = 0,0001$. Относительные погрешности $\delta_1 = 0,00025$, $\delta_2 = 0,5$.

Ясно, что точность второго корня слишком мала. Это произошло вследствие нарушения правила 2. Попробуем исправить положение. Для этого преобразуем формулу для вычисления второго корня так, чтобы не вычитать, а складывать почти одинаковые числа:

$$x_2 = \frac{-b + \sqrt{b^2 - 4c}}{2} = \frac{(-b + \sqrt{b^2 - 4c})(-b - \sqrt{b^2 - 4c})}{2(-b - \sqrt{b^2 - 4c})} =$$

$$= \frac{b^2 - 4c - b^2}{-2(\sqrt{b^2 - 4c} + b)} = \frac{2c}{\sqrt{b^2 - 4c} + b}.$$

Используя эту формулу, получим

$$x_2 = -\frac{0{,}00016}{0{,}3996 + 0{,}4002} = -\frac{0{,}00016}{0{,}7998} = -0{,}0002,$$

т. е. получили практически точное значение корня.

К сожалению, не существует общих правил получения такого улучшения результата.

Задача В.2. Проверить правило 3 на примере $a = 0{,}9364$; $b = 0{,}6392$; $c = 0{,}6375$ (умножение) и $a = 0{,}41$; $b = 0{,}36$; $c = 0{,}7$ (деление).

Правила действий с погрешностями

Постановка задачи. Пусть задана дифференцируемая функция

$$z = f(x_1, \ldots, x_n).$$

Требуется, зная A_i — абсолютные погрешности аргументов, найти абсолютную A и относительную δ погрешности вычисления функции.

Абсолютная погрешность результата равна

$$A = \left| f(x_1, \ldots, x_n) - f(x_1^*, \ldots, x_n^*) \right|. \qquad \text{(В.16)}$$

Так как погрешности исходных данных обычно малы, применив к (В.16) формулу Лагранжа, получим

$$A = \left| \sum_{i=1}^{n} \frac{\partial f(\bar{x}_1, \ldots, \bar{x}_n)}{\partial x_i} \Delta x_i \right| \le \sum_{i=1}^{n} \left| \frac{\partial f(\bar{x}_1, \ldots, \bar{x}_n)}{\partial x_i} \right| A_i. \qquad \text{(В.17)}$$

Для относительной погрешности имеем следующую формулу:

$$\delta = \frac{A}{f} \le \sum_{i=1}^{n} \left| \frac{1}{f} \frac{\partial f(\bar{x}_1, \ldots, \bar{x}_n)}{\partial x_i} \right| A_i = \sum_{i=1}^{n} \left| \frac{\partial \ln f(\bar{x}_1, \ldots, \bar{x}_n)}{\partial x_i} \right| A_i. \qquad \text{(В.18)}$$

Вычислим погрешности для основных арифметических формул:

- $f = x \pm y$. Из формулы (В.17) следует

$$A(x^* \pm y^*) \le A(x^*) + A(y^*);$$

- $f = xy$. Аналогично,

$$A(x^* y^*) \le y^* A(x^*) + x^* A(y^*),$$
$$\delta(x^* y^*) \le \delta(x^*) + \delta(y^*) + \delta(x^*)\delta(y^*);$$

- $f = \dfrac{x}{y}$:

$$\delta\left(\frac{x^*}{y^*}\right) \le \frac{\delta(x^*) + \delta(y^*)}{1 - \delta(y^*)}.$$

Если $\overline{\delta}(x^*),\ \overline{\delta}(y^*) \ll 1$, то можно воспользоваться приближенными равенствами

$$\overline{\delta}\left(\frac{x^*}{y^*}\right) \approx \overline{\delta}(x^*) + \overline{\delta}(y^*), \overline{\delta}(x^* y^*) \approx \overline{\delta}(x^*) + \overline{\delta}(y^*);$$

- $f = x^n$:

$$A((x^*)^n) \le n x^* A(y^*),$$
$$\delta((x^*)^n) \le n\delta(x^*).$$

B.3.3. Обратная задача теории погрешностей

Постановка задачи. Определить, какими должны быть погрешности аргументов, чтобы обеспечить заданную точность результата.

Простейшим решением будет принцип равных влияний, т. е. каждое из слагаемых в формуле (B.17) оказывает одинаковое влияние на результат. Следовательно,

$$\left|\frac{\partial f(\overline{x}_1, \ldots, \overline{x}_n)}{\partial x_i}\right| A_i = \frac{A}{n},$$

откуда

$$A_i = \frac{A}{n\left|\dfrac{\partial f(\overline{x}_1, \ldots, \overline{x}_n)}{\partial x_i}\right|}.$$

Пример B.1. Пусть радиус шара равен примерно 1. С какой точностью нужно вычислить радиус и число π, чтобы объем шара был вычислен с точностью 0,1?

Объем шара определяем по формуле $V = \frac{4}{3}\pi R^3$.

Абсолютные погрешности:

$$A_\pi = \frac{0{,}1}{2\left|\frac{4}{3}R^3\right|} = \frac{0{,}3}{8} = 0{,}0375; \quad A_R = \frac{0{,}1}{2\left|4\pi R^2\right|} = \frac{0{,}1}{8\pi} = 0{,}00398.$$

B.4. Нормы вектора и матрицы

Обобщим понятие длины вектора на случай многомерного пространства.

Определение B.7. ***Нормой вектора*** $\overline{x}$ называют число, обозначаемое $\|\overline{x}\|$ и удовлетворяющее следующим условиям:

1) $\|\overline{x}\| \ge 0,\ \|\overline{x}\| = 0 \Leftrightarrow \overline{x} = \overline{\theta}$, здесь $\overline{\theta}$ — нулевой вектор;

2) $\|\alpha\bar{x}\| = |\alpha| \times \|\bar{x}\|$, где α — любое действительное число;
3) $\|\bar{x} + \bar{y}\| \le \|\bar{x}\| + \|\bar{y}\|$ — неравенство, называемое ***неравенством треугольника***.

Очевидно, что свойства 1 — 3 являются обобщением свойств длины трехмерного вектора.

Примеры:

1) $\|\bar{x}\|_1 = \sum_{i=1}^{n} |x_i|$;

2) $\|\bar{x}\|_2 = \sqrt{x_i^2} = \sqrt{(\bar{x}, \bar{x})}$;

3) $\|\bar{x}\|_\infty = \max_i |x_i|$.

Определение В.8. ***Евклидовой нормой*** вектора $\bar{x}$ называется $\|\bar{x}\|_2$.

Примеры: пусть задан вектор $\bar{x} = (1, -2, 3, -4)^T$, тогда:
1) $\|\bar{x}\|_1 = 1 + 2 + 3 + 4 = 10$;
2) $\|\bar{x}\|_2 = \sqrt{1 + 4 + 9 + 16} = \sqrt{30}$;
3) $\|\bar{x}\|_\infty = 4$.

В конечномерном пространстве справедливо следующее важное утверждение.

Теорема В.1. Все конечномерные нормы эквивалентны между собой, т. е. для любых двух норм $\|\bar{x}\|_\alpha$ и $\|\bar{x}\|_\beta$, введенных ранее, справедливо неравенство

$$\gamma \|\bar{x}\|_\alpha \le \|\bar{x}\|_\beta \le \delta \|\bar{x}\|_\alpha,$$

где γ и δ — некоторые положительные числа.

Это означает, что если доказана сходимость последовательности векторов $\{\bar{x}_n\}$ в некоторой норме, то сходимость будет иметь место и в любой другой норме.

Перенесем понятие нормы вектора на матрицы.

Определение В.9. ***Нормой матрицы*** A называют число, обозначаемое $\|A\|$, удовлетворяющее следующим условиям:
1) $\|A\| \ge 0$, $\|A\| = 0 \Leftrightarrow A$ — нулевая матрица;
2) $\|\alpha A\| = |\alpha| \times \|A\|$, где α — любое действительное число;
3) $\|A + B\| \le \|A\| + \|B\|$. Здесь A и B — матрицы одинаковой размерности;
4) $\|AB\| \le \|A\| \times \|B\|$.

Поскольку в дальнейшем нас будет интересовать произведение матрицы и вектора, то должно быть согласование в определениях соответствующих норм.

Определение В.10. Будем говорить, что норма матрицы $\|A\|$ ***согласована с нормой вектора*** $\|\bar{x}\|$, если $\|A\bar{x}\| \le \|A\| \times \|\bar{x}\|$.

Обычно норма матрицы вводится с помощью введенной ранее нормы вектора.

Определение В.11. Норма матрицы A называется ***подчиненной*** норме вектора $\bar{x}$, если $\|A\|$ определена таким образом:

$$\|A\| = \sup_{\|\bar{x}\|\neq 0} \frac{\|A\bar{x}\|}{\|\bar{x}\|} = \sup_{\|\bar{x}\|=1} \|A\bar{x}\|.$$

Очевидно, что подчиненная норма согласована с соответствующей нормой метрикой вектора:

$$\frac{\|A\bar{x}\|}{\|\bar{x}\|} \le \sup_{\|\bar{x}\|\neq 0} \frac{\|A\bar{x}\|}{\|\bar{x}\|} = \|A\|.$$

Сравнив правую и левую части неравенства, получим $\|A\bar{x}\| \le \|A\|\|\bar{x}\|$.

Без доказательства приведем выражения для норм матриц, подчиненным соответствующим векторным нормам (через a_{ij} обозначим элементы матрицы A):

1) $\|\bar{x}\|_1 = \sum_{i=1}^{n} |x_i|$,

$$\|A\|_1 = \max_i \left| \sum_{j=1}^{n} a_{ij} \right|; \qquad \text{(B.19)}$$

2) $\|\bar{x}\|_2 = \sqrt{\sum_{l=1}^{n} x_i^2}$,

$$\|A\|_2 = \sqrt{\lambda}, \qquad \text{(B.20)}$$

где λ — наибольшее собственное значение матрицы $A^T A$;

3) $\|\bar{x}\|_\infty = \max|x_i|$,

$$\|A\|_\infty = \max_j \left| \sum_{i=1}^{n} a_{ij} \right|. \qquad \text{(B.21)}$$

Например, пусть задана матрица

$$A = \begin{pmatrix} 1 & -2 & 3 \\ -4 & 5 & -6 \\ 7 & -8 & 9 \end{pmatrix}.$$

Рассчитаем норму заданной матрицы.

1) $\|A\|_1 = \max_i \left| \sum_{j=1}^{n} a_{ij} \right| = \max(|1-2+3|, |-4+5-6|, |7-8+9|) =$

$$= \max(2,5,8) = 8;$$

2) $\|A\|_2 = \sqrt{\lambda}$,

$$A^T A = \begin{pmatrix} 1 & -4 & 7 \\ -2 & 5 & -8 \\ 3 & -6 & 9 \end{pmatrix} \begin{pmatrix} 1 & -2 & 3 \\ -4 & 5 & -6 \\ 7 & -8 & 9 \end{pmatrix} = \begin{pmatrix} 66 & -87 & 90 \\ -78 & 93 & -108 \\ 90 & -108 & 126 \end{pmatrix}.$$

Ее собственные значения $\{0, \frac{1}{2}(285 - \sqrt{8881}), \frac{1}{2}(285 + \sqrt{8881})\}$.

Поэтому $\|A\|_2 = \sqrt{\frac{1}{2}(285 + \sqrt{8881})}$;

3) $\|A\|_\infty = \max\limits_j \left| \sum\limits_{i=1}^{n} a_{ij} \right|$,

$$\|A\|_\infty = \max_j \left| \sum_{i=1}^{n} a_{ij} \right| = \max\{|1-4+7|, |-2+5-8|, |3-6+9|\} =$$

$$= \max\{4,5,6\} = 6.$$

B.5. Сжимающее отображение

B.5.1. Метрические пространства

Определение B.12. ***Метрикой***, или ***расстоянием***, на множестве D называют числовую функцию $\rho(x,y)$, $x,y \in D$, определяемую следующим образом:

1) $\rho(x,y) \geq 0$, $\forall x,y \in D$;
2) $\rho(x,y) = \rho(y,x)$;
3) $\rho(x,y) \leq \rho(x,z) + \rho(z,x)$, $\forall x,y,z \in D$ — это неравенство называется неравенством треугольника.

Определение B.13. ***Метрическим пространством*** называют совокупность $\{D, \rho\}$.

Например:

1) $D = R^1$, $\rho(x,y) = |x-y|$;
2) $D = R^2$, $x^T = (x_1, x_2)$, $\rho(x,y) = \sqrt{(x_1-y_1)^2 + (x_1-y_2)^2}$;
3) $D = R^2$, $x^T = (x_1, x_2)$, $\rho(x,y) = \max\{|x_1-y_1|, |x_2-y_2|\}$,
4) $D = R^2$, $x^T = (x_1, x_2)$, $\rho(x,y) = |x_1-y_1| + |x_2-y_2|$.

Определение B.14. Множество U линейного пространства L называют ***выпуклым множеством***, если вместе с любыми двумя точками A и B оно содержит отрезок AB.