Санкт-Петербургский государственный университет
Факультет прикладной математики — процессов управления

А. П. Иванов, Л. Т. Позняк, А. С. Еремин

Практикум на ЭВМ по численным методам

Тема 1. Вычисление функций

Методические указания

Санкт-Петербург
2021

Основы теории погрешностей

На погрешность результата приближенного решения задачи влияют следующие причины:

а) Неточность информации о решаемой задаче. Ошибки в начальных данных дают ту часть погрешности в решении, которая не зависит от математической стороны решения задачи и называется неустранимой погрешностью.

б) Погрешность аппроксимации (методическая погрешность). При решении задачи численными методами необходимо считаться с тем, что неизбежно придётся иметь дело только с конечным количеством чисел, и с ними можно выполнить только конечное число операций. Поэтому вместо точного решения задачи приходится прибегать к приближенному методу.

в) Погрешность округления. Всякое положительное число a может быть представлено в виде конечной или бесконечной десятичной дроби

$$a = \alpha_m 10^m + \alpha_{m-1} 10^{m-1} + \cdots + \alpha_{m-n+1} 10^{m-n+1} + \cdots ,$$

где α_i — цифры числа a, причём старшая цифра $\alpha_m \neq 0$, а m — некоторое число (старший десятичный разряд числа a).

Например,

$$3141.59\ldots = 3 \cdot 10^3 + 1 \cdot 10^2 + 4 \cdot 10^1 + 1 \cdot 10^0 + 5 \cdot 10^{-1} + 9 \cdot 10^{-2} + \cdots$$

На практике имеют дело с приближёнными числами, представляющими собой конечные десятичные дроби

$$\bar{b} = \beta_m 10^m + \beta_{m-1} 10^{m-1} + \cdots + \beta_{m-n+1} 10^{m-n+1}, \quad \beta_m \neq 0.$$

Определение 1. Цифра β_k в изображении числа $\bar{b}$ называется *верной*, если имеет место неравенство $|b - \bar{b}| \leqslant \omega 10^k$, $\omega \leqslant 1$, чаще всего, $\omega = 0.5$. Здесь b — точное значение величины, представленной приближённой записью через $\bar{b}$.

Очевидно, что если цифра β_k верная, то и все цифры в записи числа b, расположенные левее неё, тоже верны.

Определение 2. *Значащей цифрой* числа называется всякая его цифра в десятичном изображении, кроме нулей, стоящих слева в записи числа до первой ненулевой цифры.

Число, являющееся решением конкретной задачи, принято записывать только с *верными значащими цифрами.* Например, в числе 0.002080 первые три нуля не являются значащими цифрами, так как они служат только для установления десятичных разрядов других цифр. Остальные два нуля являются значащими. В случае, если в данном числе 0.002080 последняя цифра не является верной, то её не следует использовать в записи числа.

Определение 3. Число $\Delta a = a - \bar{a}$ называется *абсолютной погрешностью* приближённого значения $\bar{a}$. Часто абсолютной погрешностью называют $|\Delta a|$, однако для работы с *оценками* погрешностей разницы нет.

Определение 4. Число Δ_a такое, что $|\Delta a| \leqslant \Delta_a$, называется *верхней границей (оценкой) абсолютной погрешности* приближённого значения $\bar{a}$.

Определение 5. Число $\delta a = \frac{\Delta a}{|a|}$ называется *относительной погрешностью* приближенного значения $\bar{a}$ (при $a \neq 0$). Обычно значение a неизвестно, и практически используется близкое значение $\delta a = \frac{\Delta a}{|\bar{a}|}$ при $\bar{a} \neq 0$. Конечно, близость этих значений зависит от соотношения абсолютных величин числа и его погрешности.

Определение 6. Число δ_a такое, что $|\delta a| \leqslant \delta_a$, называется *верхней границей (оценкой) относительной погрешности* приближённого значения $\bar{a}$.

Часто в определениях 4 и 6 слово «верхняя» опускают для краткости. Если известна граница абсолютной погрешности Δ_a, то в качестве δ_a, очевидно, можно взять $\frac{\Delta_a}{|\bar{a}|}$. Если же известна верхняя граница относительной погрешности δ_a, то за Δ_a можно взять $|\bar{a}|\delta_a$. Эту связь между Δ_a и δ_a выражают формулой $\Delta_a = |\bar{a}|\delta_a$.

Замечание 1. Для $\bar{a} = 0$ относительная погрешность не определена.

Использование различных характеристик точности

Чаще в практике вычислений при работе с числами с плавающей точкой используют не абсолютную, а относительную погрешность, так как более значимым является получение нужного числа верных цифр всех компонентов решения, насколько бы они ни различались по абсолютной величине. Однако в тех случаях, когда абсолютные значения становятся очень малы или, тем более, обращаются в нуль, все же проверяется малость абсолютной погрешности.

Обычно задают различные допуски на относительную и абсолютную погрешности. Обозначим их $rtol$ и $atol$ (от англ. *relative tolerance* и *absolute tolerance*) соответственно. Можно удобно объединить в одной формуле проверку относительной и абсолютной погрешностей и одновременно избежать деления на малую величины (если $|\bar{a}|$ мало). Считаем, что точность удовлетворена, если

$$|\Delta a| \leqslant rtol \cdot |\bar{a}| + atol. \tag{1}$$

Поскольку на практике $atol$ на порядки меньше чем $rtol$, то в случае, когда значение $|a|$ (и $|\bar{a}|$) достаточно велико, величина правой части определяется первым слагаемым и мы проверяем малость относительной погрешности. Если же $|a|$ достаточно мало, то начинает доминировать $atol$ и проверяется абсолютная погрешность.

Например, пусть мы работаем с точностью *double*, которая позволяет хранить примерно шестнадцать десятичных цифр в записи числа. Пусть $rtol = 10^{-6}$, то есть мы требуем, чтобы ответ содержал шесть верных десятичных знаков, но было бы наивно надеяться, что можно получить шесть верных цифр, если сама величина близка к нулю. Потому, зададим $atol = 10^{-12}$ и тем самым будем проверять только те цифры, которые соответствуют разрядам бóльшим чем 10^{-12}.

Так если

$$\bar{a} = 0.033\,166\,247\,903\,554,$$

то

$$rtol \cdot |\bar{a}| + atol = 0.000\,000\,033\,167\,248.$$

Таким образом при проверке абсолютная погрешность не должна превышать 10^{-8}, что гарантирует шесть верных цифр в $\bar{a}$.

Если же

$$\bar{a} = 0.000\,000\,003\,316\,625,$$

то

$$rtol \cdot |\bar{a}| + atol = 0.000\,000\,000\,001\,0033$$

и абсолютная погрешность не должна превышать 10^{-12}, что гарантирует лишь три верных десятичных знака для $\bar{a}$. Если же $|\bar{a}| < atol$, то мы вообще не проверяем его погрешность, считая его равным вычислительному нулю.

Если мы работаем с вектором, то вместо модуля используется норма вектора, например проверку (1) проводится покоординатно, что соответствует норме-максимум. Часто используются следующие нормы векторов:

$$\begin{aligned} \|\sigma\|_\infty &= \max_{1\leqslant i\leqslant m} |\sigma^i| \quad \text{(норма-максимум)}, \\ \|\sigma\|_1 &= \sum_{i=1}^{m} |\sigma^i|, \\ \|\sigma\|_2 &= \left(\sum_{i=1}^{n} (\sigma^i)^2\right)^{1/2} \quad \text{(евклидова норма)}. \end{aligned} \tag{2}$$

Однако, порой при решении больших систем приходится накладывать разные ограничения на погрешности для разных компонент решения, например, если одна часть неизвестных соответствует координатам объекта, а другая — скоростям их изменения, или одна — зарядам на узлах электрической цепи, а другая — токам через эти узлы. Характерные величины разных по физической природе искомых функций могут сильно различаться, и добиваться одинаковой точности (абсолютной или относительной) становится нецелесообразно. Весь вектор решения разделяется тогда на подвекторы w_j, для каждого из которых погрешность Δ_i сравнивается со своими собственными $rtol_j$ и $atol_j$ по норме (в предельном случае каждая компонента образует свой подвектор и сравнивается со своими допусками). Может быть и так, что для некоторых из подвекторов решения мы вообще не ведем контроль погрешности.

Прямая задача теории погрешностей

В дальнейшем изложении будем считать, что погрешность округлений пренебрежимо мала по сравнению с методической погрешностью и работать с абсолютными прогрешностями.

Пусть в выпуклой области $G \in \mathbb{R}^n$ рассматривается непрерывно дифференцируемая функция $y = f(\cdot)$. Предположим, что в точке $x = (x_1, x_2, \ldots, x_n)$ области G нужно вычислить значение $y = f(x)$. Пусть нам известны лишь приближённые значения $\bar{x}_1$, $\bar{x}_2$, ..., $\bar{x}_n$ такие, что точка $\bar{x} = (\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_n) \in G$.

Необходимо найти оценку погрешности приближённого значения функции $\bar{y} = f(\bar{x})$, обусловленную погрешностями аргументов. Через погрешности $\Delta x_i = x_i - \bar{x}_i$ аргументов она выражается следующим образом:

$$\Delta y = f(\bar{x}_1 + \Delta x_1, \bar{x}_2 + \Delta x_2, \ldots, \bar{x}_n + \Delta x_n) - f(\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_n).$$

Или, если воспользоваться формулой Лагранжа[1],

$$\Delta y = \sum_{i=1}^{n} \frac{\partial}{\partial x_i} f(\bar{x}_1 + \theta_1 \Delta x_1, \bar{x}_2 + \theta_2 \Delta x_2, \ldots, \bar{x}_n + \theta_n \Delta x_n), \quad \theta_i \in [0, 1].$$

Отсюда получается оценка для границы погрешности вычисления функции, порождённой погрешностью её аргументов:

$$|\Delta y| \leqslant \sum_{i=1}^{n} B_i \Delta_{x_i}, \tag{3}$$

где $|\Delta x_i| \leqslant \Delta_{x_i}$,

$$B_i = \max_{\substack{\theta_j \in [0,1] \\ j=1,\ldots,n}} \left| \frac{\partial}{\partial x_i} f(\bar{x}_1 + \theta_1 \Delta x_1, \bar{x}_2 + \theta_2 \Delta x_2, \ldots, \bar{x}_n + \theta_n \Delta x_n) \right|.$$

Таким образом решается *прямая задача теории погрешностей*: известны погрешности некоторой системы величин. Требуется определить погрешность вычисления заданной функции f этих величин, порождённой их погрешностями.

Здесь учтена лишь неустранимая погрешность вычисления функции, порождённая погрешностями её аргументов. Если же считать,

[1] Жозеф Луи Лагранж

что значение функции $f(x)$ не может быть вычислено точно (например, $\sqrt{2}$) и его вычисление заменяется вычислением другой функции $\bar{f}(x)$ (например, отрезка ряда Тейлора), то возникает и *методическая погрешность вычисления функции.* Для совокупной (полной) погрешности (без учёта ошибок округления) имеем:

$$|f(x) - \bar{f}(\bar{x})| \leqslant |f(x) - f(\bar{x})| + |f(\bar{x}) - \bar{f}(\bar{x})| \leqslant |\Delta y| + \Delta_{f(\bar{x})}.$$

В итоге, для полной погрешности Δy вычисления функции $f(x)$ следует написать оценку:

$$|\Delta f| \leqslant \Delta_{f(\bar{x})} + \sum_{i=1}^{n} B_i \Delta_{x_i}, \tag{4}$$

Неравенство (4) назовём *решением полной задачи теории погрешностей.*

Полная обратная задача теории погрешностей

Рассмотрим вопрос: каковы должны быть абсолютные погрешности аргументов функции и методическая погрешность функции, чтобы абсолютная величина полной погрешности вычисления функции не превышала заданной величины?

Эта задача математически неопределена, так как заданную предельную погрешность (допуск ε на верхнюю границу абсолютной погрешности Δf) можно обеспечить, устанавливая по-разному предельные абсолютные погрешности Δ_{x_i} её аргументов и вычисления функции $\Delta_{f(\bar{x})}$ лишь бы они удовлетворяли условию:

$$\Delta_{f(\bar{x})} + \sum_{i=1}^{n} B_i \Delta_{x_i} \leqslant \varepsilon. \tag{5}$$

Требуется наложить дополнительное условие, чтобы задача имела единственное решение.

Одно из двух простейших решений обратной задачи даётся так называемым *принципом равных влияний.* Предполагается, что все слагаемые в правой части (5) имеют одинаковую величину. Тогда

$$\Delta_{f(\bar{x})} \leqslant \frac{\varepsilon}{n+1}, \quad \Delta_{x_i} \leqslant \frac{\varepsilon}{(n+1)B_i}, \quad i = \overline{1, n}.$$

Другой столь же простой способ носит название *принципа равных погрешностей*: считается, что $\Delta_{f(\bar{x})} = \Delta_{x_i}$, $i = \overline{1, n}$, и тогда из (5) немедленно получаем:

$$\Delta_{f(\bar{x})} \leqslant \frac{\varepsilon}{1 + \sum_{j=1}^{n} B_j}, \quad \Delta_{x_i} \leqslant \frac{\varepsilon}{1 + \sum_{j=1}^{n} B_j}, \quad i = \overline{1, n}.$$

Исходя из особенностей задачи и вычисляемой функции можно выставлять и другие требования к уровню погрешностей аргументов, например, сгруппировать аргументы по их физическому смыслу и использовать принцип равных влияний для разных групп, а внутри групп применить принцип равных погрешностей, или исходя из неких соображений задать погрешности для части аргументов, а погрешности остальных найти из условия выполнения равенства (5) с учётом положительности искомых величин $\Delta_{f(\bar{x})}$ и Δ_{x_i}, $i = \overline{1, n}$.

Пример. Требуется с погрешностью $\Delta_M = 1$ г определить массу металла, потребного для изготовления диска толщиной $h \approx 1$ см и радиусом $r \approx 10$ см (плотность металла ϱ). Для определенности будем считать, что $\varrho = 10$ г/см^3, $h = 1 \pm 0.1$ см, $\pi = 3.14 \pm 0.002$, $r = 10 \pm 0.1$ см.

Решение. Поскольку масса диска (цилиндра) вычисляется по формуле $M = \varrho h \pi r^2$, то поставленный вопрос эквивалентен вопросу: с какой погрешностью должны быть измерены радиус и толщина диска, а также сколько следует взять знаков в числе π, чтобы выполнить условия по точности вычисления массы диска (считая, что плотность — величина точная)?

Согласно формуле (5), имеем:

$$\Delta_M \leqslant B_h \Delta_h + B_\pi \Delta_\pi + B_r \Delta_r,$$

где

$$B_h = \max_{\pi, r}(\varrho \pi r^2) = 3205.52,$$

$$B_\pi = \max_{h, r}(\varrho h r^2) = 1122.11,$$

$$B_r = \max_{h, \pi, r}(2 \varrho h \pi r) = 698.15.$$

Применяя принцип равных погрешностей, получаем:

$$\Delta_h = \Delta_\pi = \Delta_r \approx 1/5000 = 2 \cdot 10^{-4}.$$

т. е. линейные размеры диска должны быть определены с точностью 2 микрона, а значение числа π следует взять с точностью 5 знаков после запятой.

Обратная задача для вычисления элементарных функций

Общие положения

Пусть $z(x) = f(u(x), v(x))$, $x \in [a, b]$, и требуется построить таблицу значений этой функции для узлов $x = x_i$, $i = 1, ..., n$. Здесь x — скалярный аргумент ($x_i < x_{i+1}$). Предполагается, что каждая из трёх функций f, u и v не может быть вычислена точно и вычисляется приближённо:

$$u(x) \approx \bar{u}(x), \quad v(x) \approx \bar{v}(x), \quad f(x) \approx \bar{f}(x).$$

Таким образом, реально вместо искомых точных значений функции $z_i = f(u(x_i), v(x_i))$ будут вычислены приближённые значения $\bar{z}_i = \bar{f}(\bar{u}(x_i), \bar{v}(x_i))$. Требуется определить с какой точностью должны быть вычислены $\bar{u}(x)$, $\bar{v}(x)$ и $\bar{f}(x)$, чтобы обеспечивалась заданная точность приближённых значений $\bar{z}_i$, т. е. чтобы $|z_i - \bar{z}_i| \leqslant \varepsilon$. Как нетрудно убедиться, здесь мы имеем дело с рассмотренной ранее обратной задачей теории погрешностей. В самом деле, нам требуется вычислить значения функции двух переменных $f(u, v)$ при некоторых значениях аргументов u и v, которые известны нам не точно, но могут быть найдены их приближённые значения $\bar{u}$ и $\bar{v}$ с требуемой точностью.

Пусть $u_* \leqslant u(x) \leqslant u^*$, $v_* \leqslant v(x) \leqslant v^*$ при $x \in [x_1; x_n]$. В таком случае область G есть прямоугольник $[u_*, u^*] \times [v_*, v^*]$. Считаем далее, что производные $\frac{\partial f}{\partial u}$ и $\frac{\partial f}{\partial v}$ мало изменяются в G, и что

$$\left|\frac{\partial f}{\partial u}(u, v)\right| \leqslant c_u, \quad \left|\frac{\partial f}{\partial v}(u, v)\right| \leqslant c_v \quad (u, v) \in G.$$

Из предыдущих рассуждений о решении обратной задачи вытекает, что требуемая точность ε приближённых табличных значений $\bar{z}_i$ обеспечивается тогда, когда приближённые значения аргументов $\bar{u}$ и $\bar{v}$ удовлетворяют неравенствам

$$|u(x_i) - \bar{u}(x_i)| \leqslant \frac{\varepsilon}{3c_u}, \quad |v(x_i) - \bar{v}(x_i)| \leqslant \frac{\varepsilon}{3c_v},$$

а приближённо вычисленное значение $\bar{f}(\bar{u}, \bar{v})$ удовлетворяет неравенству

$$|f(\bar{u}, \bar{v}) - \bar{f}(\bar{u}, \bar{v})| \leqslant \frac{\varepsilon}{3}.$$

Предполагаем, что мы умеем оценивать методическую погрешность вычисления функций u, v и f, т. е. можем дать оценки:

$$\begin{aligned}
&|u(x) - \bar{u}(x)| \leqslant \Delta_u, \\
&|v(x) - \bar{v}(x)| \leqslant \Delta_v, \\
&|f(\bar{u}, \bar{v}) - \bar{f}(\bar{u}, \bar{v})| \leqslant \Delta_f.
\end{aligned}$$

Поставленная задача будет решена, когда будет обеспечено выполнение следующих неравенств:

$$\Delta_u \leqslant \frac{\varepsilon}{3c_u}, \quad \Delta_v \leqslant \frac{\varepsilon}{3c_v}, \quad \Delta_f \leqslant \frac{\varepsilon}{3}.$$

Примеры построения таблицы значений функции

Пример 1. Требуется построить таблицу значений функции

$$z(x) = \sqrt{\frac{\sin(0.9x + 0.51)}{x\mathrm{e}^{x+0.3}}}$$

для $x = 0.5(0.01)0.6$ (т. е. для $x \in [0.5, 0.6]$ с шагом 0.01) с заданной точностью $\varepsilon = 10^{-6}$. Источниками погрешности (при точных значениях x) являются три функции: синус, экспонента и квадратный корень. Положим

$$u(x) = \sin(0.9x + 0.51), \quad v(x) = \mathrm{e}^{x+0.3}, \quad f(u, v, x) = \sqrt{\frac{u}{xv}}.$$

Упростим обозначения, не указывая явную зависимость f от x, так как этот аргумент мы считаем точным. Тогда $z(x) = f(u(x), v(x))$. Найдём пределы изменения величин u и v при $x \in [0.5, 0.6]$. Поскольку функции u и v монотонны на $[0.5, 0.6]$, то $\sin 0.96 \leqslant u \leqslant \sin 1.05$, а $\mathrm{e}^{0.8} \leqslant v \leqslant \mathrm{e}^{0.9}$.

Интервалы изменения u и v можно расширить, чтобы не вычислять верхние и нижние границы изменения этих функций с большой точностью. Положим с недостатком $\sin 0.96 \approx 0.8$, $\exp 0.8 \approx 2.2$ и с избытком $\sin 1.05 \approx 0.9$, $\exp 0.9 \approx 2.5$. Таким образом, можно положить

$$G = \{(u, v) \mid 0.8 \leqslant u \leqslant 0.9,\ 2.2 \leqslant v \leqslant 2.5\}$$

Оценим в G частные производные

$$\frac{\partial f}{\partial u}(u, v) = \frac{1}{2\sqrt{uvx}}, \quad \frac{\partial f}{\partial v}(u, v) = -\frac{\sqrt{u}}{2v\sqrt{vx}}\ :$$

$$\left|\frac{\partial f}{\partial u}\right| \leqslant \frac{1}{2\sqrt{0.8 \cdot 2.2 \cdot 0.5}} < 0.54, \quad \left|\frac{\partial f}{\partial v}\right| \leqslant \frac{\sqrt{0.9}}{2 \cdot 2.2 \cdot \sqrt{2.2 \cdot 0.5}} < 0.21,$$

Итак, в данном примере $c_u = 0.54$, $c_v = 0.21$ и, следовательно, функцию $u(x)$ нужно вычислять с точностью $\varepsilon_u = 10^{-6}/1.62$, функцию $v(x)$ — с точностью $\varepsilon_v = 10^{-6}/0.63$, а $f(u, v)$ — с точностью $\varepsilon_f = 10^{-6}/3$.

Функции $u(x)$ и $v(x)$ предлагается вычислять, ракладывая функции $\cos y$ и e^t в ряд Маклорена по аргументам $y = \frac{\pi}{2} - (0.9x + 0.51)$ и $t = x + 0.3$ (при этом будет $y \in (0; \frac{\pi}{4}) \subset [0; 1]$ и ряд для $\cos y$ станет лейбницевым). Функцию $f(u, v)$ предлагается вычислять, определяя приближённое значение функции $\sqrt{s}$ по формуле Герона (частного случая формулы Ньютона):

$$w_{k+1} = \frac{1}{2}\left(w_k + \frac{s}{w_k}\right),$$

где начальное w_0 — приближённое значение $\sqrt{s}$ с избытком. Можно, к примеру, взять в данном случае $w_0 = 1$.

Для всех трёх функций мы умеем оценивать абсолютную величину методической погрешности (с учётом того, что элементарные функции вычисляются с помощью разложения в ряд Маклорена):

$$\begin{aligned}
&|u(x) - \bar{u}(x)| \leqslant \left|\frac{y^{2n}}{(2n)!}\right| = \Delta_u \leqslant \varepsilon_u,\\
&|v(x) - \bar{v}(x)| \leqslant \left|\frac{t^m}{m!}\right| = \Delta_v \leqslant \varepsilon_v,\\
&|f(\bar{u}, \bar{v}) - \bar{f}(\bar{u}, \bar{v})| \leqslant |w_{k+1} - w_k| = \Delta_f \leqslant \varepsilon_f.
\end{aligned}$$

Следовательно, требуемая точность табличных значений функции $z(x)$ будет обеспечена тогда, когда номера n, m и k будут удовлетворять неравенствам

$$\begin{aligned}
&\left|\frac{y^{2n}}{(2n)!}\right| \leqslant \frac{10^{-6}}{1.62},\\
&\left|\frac{t^m}{m!}\right| \leqslant \frac{10^{-6}}{0.63},\\
&|w_{k+1} - w_k| \leqslant \frac{10^{-6}}{3}.
\end{aligned}$$

Пример 2. Рассмотрим пример, в котором последнее действие в функции f выполняется точно. В этом случае удобнее все источники погрешности считать её аргументами и в формуле (4) $\Delta_{f(\bar{x})}$ будет равно нулю.

Требуется построить таблицу значений функции

$$z(x) = \frac{\sqrt{\sin(0.9x + 0.51)}}{x\mathrm{e}^{x+0.3}}$$

для $x = 0.5(0.01)0.6$ с заданной точностью $\varepsilon = 10^{-6}$. Источниками погрешности являются те же три функции, что и в примере 1. Положим

$$u(x) = \sin(0.9x + 0.51), \quad v(x) = \mathrm{e}^{x+0.3}, \quad w(u) = \sqrt{u}.$$

Ещё одним отличием от примера 1 здесь является сложная зависимость f от u, а именно $z(x) = f(w(u(x)), v(x))$. Найдём пределы изменения величин u, v и w при $x \in [0.5, 0.6]$. Все три функции монотонны на рассматриваемом отрезке, и к двум полученным в примере 1 оценкам $\sin 0.96 \leqslant u \leqslant \sin 1.05$, $\mathrm{e}^{0.8} \leqslant v \leqslant \mathrm{e}^{0.9}$ добавляется $\sqrt{\sin 0.96} \leqslant w \leqslant \sqrt{\sin 1.05}$.

Получим область

$$G = \{(u, v, w) \mid 0.8 \leqslant u \leqslant 0.9,\ 2.2 \leqslant v \leqslant 2.5,\ 0.9 \leqslant w \leqslant 0.95\}$$

Оценим в G частные производные функции $f = w(u)/xv$:

$$\frac{\partial f}{\partial u} = \frac{\partial f}{\partial w}\frac{\partial w}{\partial u} = \frac{1}{2vx\sqrt{u}}, \quad \frac{\partial f}{\partial v} = -\frac{w}{2xv^2}, \quad \frac{\partial f}{\partial w} = \frac{1}{xv} :$$

$$\left|\frac{\partial f}{\partial u}\right| \leqslant \frac{1}{2 \cdot 0.5 \cdot 2.2 \cdot \sqrt{0.8}} < 0.51, \quad \left|\frac{\partial f}{\partial v}\right| \leqslant \frac{0.95}{2 \cdot 0.5 \cdot 2.2^2} < 0.4,$$

$$\left|\frac{\partial f}{\partial w}\right| \leqslant \frac{1}{0.5 \cdot 2.2} < 0.91.$$

Отсюда $c_u = 0.51$, $c_v = 0.4$, $c_w = 0.91$. Следовательно, функцию $u(x)$ нужно вычислять с точностью $\varepsilon_u = 10^{-6}/1.53$, функцию $v(x)$ — с точностью $\varepsilon_v = 10^{-6}/1.2$, а $w(u)$ — с точностью $\varepsilon_f = 10^{-6}/2.73$.

Последние действия, а именно вычисления $d = xv$ и $f = w/d$, считаются выполняемыми точно.

Содержание задания

1. Решить обратную задачу теории погрешностей для заданной функции $z(x)$ с допуском $\varepsilon = 10^{-6}$. Выполняется аналитически.

2. Найти с требуемой точностью приближенные значения $\bar{z}(x)$ этой функции (квадратный корень вычислять по формуле Герона, а остальные элементарные функции вычислять с использованием степенных рядов, указанных ниже).

3. Найти «точные» значения $z(x)$, используя встроенные функции языка программирования. Убедиться, что абсолютная погрешность вычислений удовлетворяет заданному допуску ε.

   Итоговый вывод программы должен представлять собой таблицу

x	$z(x)$	$\bar{z}(x)$	$\lvert\Delta z(x)\rvert$
$\ldots$	$\ldots$	$\ldots$	$\ldots$
$\ldots$	$\ldots$	$\ldots$	$\ldots$
$\ldots$	$\ldots$	$\ldots$	$\ldots$

   причём погрешности должны быть меньше 10^{-6}, однако не должны быть слишком маленькими (не меньше 10^{-10}).

Задания для самостоятельного выполнения

1. $z(x) = \sqrt{1 + \operatorname{arctg}(16.7x + 0.1)} / \cos(7x + 0.3)$, $x = 0.01(0.005)0.05$;

2. $z(x) = \sqrt{1 + \operatorname{arctg}(6.4x + 1.1)} / \sin(2x + 1.05)$, $x = 0.01(0.005)0.06$;

3. $z(x) = \sqrt{(1 + x)\exp(x + 0.5) + \sin(x + 0.4)}$, $x = 0.5(0.01)0.6$;

4. $z(x) = \exp(1 + x)\cos\sqrt{2 + x^2}$, $x = 0.01(0.005)0.06$;

5. $z(x) = \sqrt{2x + 0.4}\,\operatorname{arctg}\cos(3x + 1)$, $x = 0.01(0.005)0.06$;

6. $z(x) = \sqrt{\operatorname{sh}(2x + 0.45)}\,\operatorname{arctg}(6x + 1)$, $x = 0.01(0.005)0.06$;

7. $z(x) = \operatorname{tg}(2x + 0.6) / \exp\sqrt{(1 + x - 12x^2)}$, $x = 0.1(0.01)0.2$;

8. $z(x) = \sqrt{\cos(2.6x + 0.1)} / \exp(1 + x)$, $x = 0.1(0.01)0.2$;

9. $z(x) = \sqrt{1 + \operatorname{arctg}(0.8x + 0.2)} / \exp(2x + 1)$, $x = 0.1(0.01)0.2$;

10. $z(x) = \sqrt{\sin(x + 0.74)}\,\mathrm{sh}(0.8x^2 + 0.1)$, $x = 0.1(0.01)0.2$;

11. $z(x) = \cos(2.8x + \sqrt{1 + x})\,\mathrm{arctg}(1.5x + 0.2)$, $x = 0.1(0.01)0.2$;

12. $z(x) = \mathrm{ch}(1 + \sqrt{1 + x})/\cos\sqrt{1 + x - x^2}$, $x = 0.1(0.01)0.2$;

13. $z(x) = \sqrt{1 + x^2}(\sin(3x + 0.1) + \cos(2x + 0.3))$, $x = 0.2(0.01)0.3$;

14. $z(x) = \mathrm{arctg}(\sqrt{0.9x + 1}/(1 - x^2)) + \sin(3x + 0.6)$, $x = 0.2(0.01)0.3$;

15. $z(x) = (\mathrm{arctg}\sqrt{1 + 0.6x})/\sin(1 + 0.4x)$, $x = 0.2(0.01)0.3$;

16. $z(x) = \mathrm{sh}(\sqrt{1 + x^2}/(1 - x))/\sin(x^2 + 0.4)$, $x = 0.2(0.01)0.3$;

17. $z(x) = \mathrm{ch}(\sqrt{x^2 + 0.3}/(1 + x))\sin((1 + x)/0.6x)$, $x = 0.2(0.01)0.3$;

18. $z(x) = \sqrt{(1 + x)}\exp(x + 0.5)\sin(0.3x + 0.7)$, $x = 0.5(0.01)0.6$;

19. $z(x) = \mathrm{ch}(2x^2 + \sqrt{x})/\sin(0.3 + \sqrt{x})$, $x = 0.5(0.01)0.6$;

20. $z(x) = \cos(0.5 + \sqrt{x})/\mathrm{arctg}(1 + 2x\sqrt{x})$, $x = 0.5(0.01)0.6$.

Степенные ряды для элементарных функций и оценки их остатков

Через u_k обозначен k-й член ряда, а через R_n — остаток частичной суммы $S_n = \sum_{k=0}^{n} u_k$.

1. $\exp x = \sum\limits_{k=0}^{\infty} \dfrac{x^k}{k!}, \quad |R_n(x)| \leqslant |u_n(x)|, \quad |x| < n + 2;$

2. $\mathrm{sh}\, x = \sum\limits_{k=0}^{\infty} \dfrac{x^{2k+1}}{(2k+1)!}, \quad |R_n(x)| \leqslant |u_n(x)|/3, \quad |x| \leqslant n;$

3. $\mathrm{ch}\, x = \sum\limits_{k=0}^{\infty} \dfrac{x^{2k}}{(2k)!}, \quad |R_n(x)| \leqslant 2|u_n(x)|/3, \quad |x| \leqslant n;$

4. $\sin x = \sum\limits_{k=0}^{\infty} (-1)^k \dfrac{x^{2k+1}}{(2k+1)!}, \quad |R_n(x)| \leqslant |u_n(x)|, \quad |x| < \dfrac{\pi}{4};$

5. $\cos x = \sum\limits_{k=0}^{\infty} (-1)^k \dfrac{x^{2k}}{(2k)!}, \quad |R_n(x)| \leqslant |u_n(x)|, \quad |x| < \dfrac{\pi}{4};$

6. $$\operatorname{arctg} x = \begin{cases} \displaystyle\sum_{k=0}^{\infty}(-1)^k \frac{x^{2k+1}}{2k+1}, & |x| < 1, \\ \displaystyle\frac{\pi}{2}\operatorname{sgn}(x) - \sum_{k=0}^{\infty}(-1)^k \frac{x^{-(2k+1)}}{2k+1}, & |x| \geqslant 1, \end{cases}$$

в обоих случаях $|R_n(x)| \leqslant |u_n(x)|$.

Литература

1. Бахвалов Н. С., Жидков Н. П., Кобельков Г. М. Численные методы, 9-е изд. — М.: Лаборатория знаний, 2020. 636 с.

2. Вержбицкий В. М. Основы численных методов: учебник для вузов — М.: Директ-Медиа, 2013. 847 с.

3. Демидович Б. П., Марон И. А. Основы вычислительной математики. М.: Изд. Наука, 1970. 664 с.