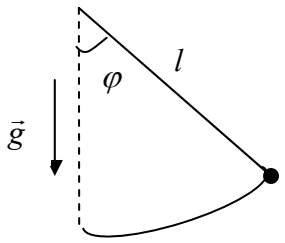
Погрешность численного решения. Специфические особенности вычислительной математики.

Вычислительная математика сталкивается со следующими понятиями.

1) Вычислительная математика имеет дело не только с непрерывными, но и **дискретными** объектами. Обычно вместо отрезка прямой $[t_0;t_N]$ рассматривается система точек $\{t_k\}_{k=0}^{N}$, вместо непрерывной функции $f(x)$ — табличная (или сеточная) функция $\{f_k\}_{k=0}^{N}$. Такие замены порождают **погрешность метода**.
2) Математическая модель, описывающая задачу, может быть неточной. Эта неточность порождает погрешность **математической модели**. Неточное задание данных, входящих в математическую модель, порождает **неустранимую погрешность**.
3) Большое значение имеет **обусловленность задачи**. Под обусловленностью задачи понимается чувствительность ее решения к малым изменениям и входным данным.
4) Численное решение математической модели не является точным. Точным оно будет в пределе, при неограниченном количестве математических итераций. В отличие от «классической» математики, выбор (численного) алгоритма влияет на результат вычислений. Существенная черта численного метода — **экономичность**, то есть, минимизация числа элементарных операций. Численный метод **порождает погрешность метода**.
5) Вычислительная машина способна хранить ограниченное количество знаков. Поэтому возникает округление чисел. Округления производятся также при всех арифметических операциях. Округление порождает **вычислительную погрешность**.



Для лучшего представления о сказанном рассмотрим пример с маятником. Предположим, требуется найти угол отклонения маятника в момент времени t_N при заданном значении начального угла — $\varphi(t_0)$. Уравнение, описывающее колебание маятника

$$l\frac{d^2\varphi}{dt^2}+g\sin\varphi+\mu\frac{d\varphi}{dt}=0. \qquad (1.1)$$

Поскольку принята такая математическая модель, то автоматически появляется неустранимая погрешность, например, хотя бы потому, что сила трения, вообще говоря, не является линейной относительно скорости. Для численного приближения производных можно применять различные численные методы. Какой бы метод не был выбран, он внесет погрешность метода. Поскольку численный метод будет решаться на ЭВМ с конечноразрядным представлением чисел, то в результате округлений возникает вычислительная погрешность. Ниже представлен пример применения численного метода.

$$l\frac{\varphi_{n+2}-2\varphi_{n+1}+\varphi_n}{\tau^2}+g\sin\varphi_{n+1}+\mu\frac{\varphi_{n+2}-\varphi_n}{2\tau}=0. \qquad (1.2)$$

Пусть, V — точное значении решения исходной (физической) задачи, $\bar{V}$ — решение математической задачи (выбранной математической модели) (1.1), $\bar{V}_\tau$ — решение численной задачи, определяемой выбранным вычислительным алгоритмом, (1.2), $\bar{V}_\tau^*$ — приближение к решению численной задачи, полученное на реальной вычислительной системе. Тогда согласно определениям

$$\rho_1=V-\bar{V} \text{ — неустранимая погрешность,} \qquad (1.3)$$

$$\rho_2=\bar{V}-\bar{V}_\tau \text{ — погрешность метода,} \qquad (1.4)$$

$$\rho_3 = \bar{V}_\tau - \bar{V}_\tau^* \text{ — вычислительная погрешность.} \quad (1.5)$$

Полная погрешность складывается из всех погрешностей — $\rho_0 = \rho_1 + \rho_2 + \rho_3$. Произведя суммирование, получаем

$$\rho_0 = V - \bar{V}_\tau^*, \quad (1.6)$$

то есть разность точного значения параметра и приближенного решения, полученного на ЭВМ при помощи численного метода.

Однако на практике в качестве погрешности берут норму величины. В скалярном случае — это модуль.

$$\rho_0 = \left|V - \bar{V}_\tau^*\right|, \rho_1 = \left|V - \bar{V}\right|, \ \rho_2 = \left|\bar{V} - \bar{V}\right|_\tau, \rho_3 = \left|\bar{V}_\tau - \bar{V}_\tau^*\right|.$$

В таком случае $\rho_0 \le \rho_1 + \rho_2 + \rho_3$.

Рассмотрим несколько примеров плохо обусловленных задач.

Пример 1.1. Округление

Рассмотрим алгебраическое уравнение

$$x^4 - 4x^3 + 8x^2 - 16x + 15,\underbrace{99999999}_{8\ штук} = 0.$$

Точное решение уравнения

$$(x-2)^4 - 10^{-8} = 0, \ (x-2)^2 = \pm 10^{-4}.$$

Отсюда

1) $x - 2 = \pm 10^{-2}$,
2) $x - 2 = \pm i 10^{-2}$.

В итоге получаем 4 различных решения

$$x_1 = 2{,}01, \ x_2 = 1{,}99, \ x_{3,4} = 2 \pm i 10^{-2}.$$

Теперь, предположим, что в силу возмущенности или машинной ошибки округления $\delta_\mu > 10^{-8}$, свободный член будет иметь значение 16. Тогда решение $x_{1,2,3,4} = 2$, что отражает принципиально неверный характер.

Пример 1.2. Расходимость

Рассмотрим простейший итерационный процесс.

$$y_{n+1} + 1000 y_n = 1001, \ n = 0 \div 4.$$

Пусть, начальное приближение $x_0 = 1$. Тогда приближения, полученные при помощи итерационного процесса, имеют следующий вид.

$$x_1 = 1, \ x_2 = 1, \ x_3 = 1, \ x_4 = 1.$$

Пусть начальное приближение возмущено $x_0 = 1 + 10^{-6}$, тогда

$$x_1 = 1 - 10^{-3}, \ x_2 = 2, \ x_3 = -999, \ x_4 = 10000001.$$

Пример 1.3. Плохая обусловленность матрицы системы

Требуется решить линейную систему уравнений

$$\begin{cases} 0{,}99x + y = 1, \\ x + 0{,}99y = 1. \end{cases}$$

Вычитаем из второго уравнения первое, помноженное на $0{,}99$.

$$\left(1 - 0{,}99^2\right)x = 1 - 0{,}99, \ 1{,}99\left(1 - 0{,}99\right)x = 1 - 0{,}99, \ 1{,}99x = 1.$$

Отсюда решение системы

$$x = y = \frac{1}{1,99}.$$

Возмутим правую часть

$$\begin{cases} 0,99x + y = 1,01, \\ x + 0,99y = 0,99. \end{cases}$$

Повторяем действия над матрицей

$$\left(1 - 0,99^2\right)x = 0,99\left(1 - 1,01\right),\ 1,99 \cdot 0,01 \cdot x = -0,99 \cdot 0,01,\ 1,99x = -0,99.$$

Решение системы

$$x = -\frac{0,99}{1,99},\ x = \frac{2,99}{1,99}.$$

Теперь рассмотрим несколько примеров, относящихся к экономичности вычислительного метода.

Пример 1.4. Сумма ряда

Требуется вычислить сумму $S = 1 + x + x^2 + ... + x^{2^n - 1}$. Для вычисления потребуется сделать $2^n - 2$ операций умножения (следующее слагаемое получается через предыдущее) и столько же операций суммирования. Всего $2^{n+1} - 4$ арифметических операций.

Если применить формулу сложения для геометрической прогрессии, то получим

$$S = \frac{1 - x^{2^n}}{1 - x}.$$

Для вычисления последнего требуется $2^n - 1$ операций умножений и 3 операции вычитания и деления. Всего $2^n + 2$. А если оптимизировать умножение,

$$x^2 = x \cdot x,\ x^4 = x^2 \cdot x^2,\ \ldots,\ x^{2n} = x^{2^{n-1}} \cdot x^{2^{n-1}},$$

то $n - 1$ операций умножения. Всего $n + 2$ арифметических операций.

Пример 1.5. Выбор метода решения линейной системы

Пусть линейная алгебраическая система $A\vec{u} = \vec{b}$ имеет ленточную матрицу $A = A_{nxn}$. Напомним, что ленточной называется матрица, состоящая из нескольких диагоналей, соседних с главной диагональю. Наиболее просто изобразить ленточную матрицу на рисунке.

$$A = \begin{pmatrix} & & & & 0 \\ & & & & \\ & & & & \\ & & & & \\ 0 & & & & \end{pmatrix}$$

Решение такой системы методом Гаусса безотносительно к типу матрицы системы потребует порядка n^3 арифметических операций ($O\left(n^3\right)$). Однако, в случае если матрица 3-х или 5-ти диагональная, то можно воспользоваться методом $3 - x$ или 5-ти точечной прогонки, который учитывает структуру матрицы. Методы прогонки требуют до n арифметических операций ($O(n)$). Подробно о методах решения систем линейных уравнений будет рассказано в соответствующем параграфе.