

Тема «Приближенные числа. Абсолютная и относительная погрешности приближенного числа, погрешность округления. Погрешность функции».

Возникающие при математическом моделировании погрешности условно делят на три типа: **неустранимую, погрешность аппроксимации и вычислительную**. Источниками первой из них являются неточности в составлении математической модели, а также в задании исходных данных. Вычислитель, как правило, лишён возможности влиять на величину этой погрешности, по-видимому, поэтому она так и называется, **неустраняемая**. Источником **погрешности аппроксимации** является замена исходной математической модели иной, более удобной для исследования. Такой переход обычно производится с целью упрощения или видоизменения модели, сокращающего, или открывающего путь к получению решения. Погрешность, возникающая вследствие ошибок округления, производимого в процессе вычислений, называется **вычислительной**.

Поясним это на **примере**.

Предположим, что вычисление значения $\sin(x)$ решено проводить по формуле

$$f(x) = x - \frac{x^3}{3!}.$$

Примечание. Из теории рядов известна формула разложения в ряд

$$f(x) = \sum_{n=1}^{\infty} \frac{(-1)^{n+1} x^{2n-1}}{(2n-1)!} = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \dots$$

Тогда неточность в задании значения x , если она есть, образует неустранимую погрешность. Величина $\sin(x) - f(x)$ представит собой погрешность аппроксимации. А проводимые в процессе вычислений округления, обусловят вычислительную погрешность.

Целью данного раздела является разработка методов численного анализа готовых математических моделей. Поэтому в первую очередь в поле зрения раздела попадают погрешность аппроксимации и вычислительная. Величина и особенности их поведения в процессе вычислений относятся к числу важных показателей, характеризующих качество вычислительных схем.

Обозначим через A точное значение некоторой величины, через a - его приближенное значение. Тогда разность $\varepsilon(x) = A - a$ называется **погрешностью приближённого числа a** . При действиях с приближёнными числами обычно известно, что абсолютная величина погрешности $\varepsilon(a)$ не превосходит некоторой величины $\Delta(a)$, т.е.

$$|\varepsilon(a)| \leq \Delta(a).$$

Величина $\Delta(a)$ представляет собой оценку абсолютной величины погрешности и называется **абсолютной погрешностью приближённого числа a** .

Естественно, в качестве $\Delta(a)$, по имеющейся информации, выбирают наименьшую величину, удовлетворяющую указанному условию. Отметим, что при наличии $\Delta(a)$ может быть установлен и диапазон расположения точного значения a . Действительно, т.к.

$$|A - a| \leq \Delta(a), \text{ то } a - \Delta(a) \leq A \leq a + \Delta(a).$$

Отношение

$$\frac{\varepsilon(a)}{a} = \frac{A - a}{a}$$

представляет собой относительную погрешность приближенного числа a . Но т.к. $\varepsilon(a)$, вообще говоря, не известна, то в качестве относительной принимают верхнюю оценку $\delta(a)$ модуля этого отношения. Т.е. величина $\delta(a)$, такая, что

$$\left| \frac{A - a}{a} \right| \leq \delta(a)$$

называется **относительной погрешностью** числа x .

Очевидно, что

$$\delta(a) = \frac{\Delta(a)}{|a|} \quad \text{и} \quad \Delta(a) = |a| \delta(a).$$

Иногда величину $\delta(a)$ выражают и в процентах. Заметим, что при наличии $\delta(a)$ также может быть установлен диапазон расположения точного значения. Действительно, так как

$$|A - a| \leq \delta(a) |a|,$$

то

$$a - |a| \delta(a) \leq A \leq a + |a| \delta(a).$$

Приближённые числа принято записывать, указывая лишь **верные значащие цифры**. Рассмотрим $x = \pm a_n a_{n-1} \dots a_1 a_0 . b_1 \dots b_{m-1} b_m$ - число, представленное в виде десятичной, а вообще говоря, n -значной дроби. Его цифры, начиная с первой отличной от нуля слева, называются **значащими**.

Определение. Значащей цифрой приближенного числа называется всякая цифра в его десятичном изображении, отличная от нуля, и нуль, если он содержится между значащими цифрами или является представителем сохраненного десятичного разряда.

Например, в числе 0.010230 пять значащих цифр, это - 1, 0, 2, 3, 0. Значащая цифра называется **верной**, если модуль погрешности числа не превышает единицу разряда, соответствующей этой цифре. Так как погрешность числа, как правило, неизвестна, то при определении верных знаков используют абсолютную погрешность.

Количество **верных знаков** числа a в широком смысле отсчитывается от первой значащей числа до первой значащей цифры его абсолютной погрешности $\Delta(a)$.

Значащую цифру называют верной в узком смысле, если абсолютная погрешность не превосходит половины единицы разряда соответствующего этой цифре.

Для записи числа с верными значащими цифрами используют как n - значную, так и показательную формы. Пусть, например, исходное приближённое число $a = 1203.045$, а $\Delta(a) = 0.03$.

Тогда верными являются цифры 1, 2, 0, 3, 0 и возможны записи: $a = 1203.0$ или $120.30 \cdot 10$ или $0.12030 \cdot 10^4$ и т.д.

Заметим, что запись без последнего справа нуля, например, $120.3 \cdot 10$ неверна. Если же $\Delta(a) = 5$, то верными являются цифры 1, 2, 0 и число следует записать так: $a = 120 \cdot 10$ или $0.120 \cdot 10^4$, т. е. сохраняя его в записи ровно три значащие цифры. Запись $x = 1200$ или $x = 0.12 \cdot 10^4$ будет неверна.

Запись чисел, содержащая лишь верные значащие цифры, обладает определённым достоинством, т.к. дает возможность составлять суждения о величине абсолютной и относительной погрешности. Действительно, пусть запись $a=0.012$ содержит лишь верные знаки. Тогда, при отсутствии иной информации, следует считать $\Delta(a)=0.001$, отсюда

$$\delta(a) = \frac{\Delta(a)}{a} = \frac{1}{12}.$$

Пример 1

Число $a = 3,7834$ задано всеми своими верными знаками в узком смысле, это означает, что абсолютная погрешность $\Delta a = 0,00005$. Найдем относительную погрешность $\delta a = \frac{\Delta a}{|a|} = 0,0000132156$.

Пример 2

Число $a = 0,07834$ округлим до трех значащих цифр $a_1 = 0,0783$. Так как цифры числа a верны в широком смысле, то $\Delta a = 0,00001$.

Погрешность округления находим по формуле $\Delta = |a - a_1| = 0,00004$. Вычислим абсолютную погрешность округленного числа по формуле

$$\Delta a_1 = \Delta a + \Delta = 0,00005.$$

Относительную погрешность округленного числа определим по формуле

$$\delta a_1 = \frac{\Delta a_1}{|a_1|} = 0,00063857.$$

Правила действий над приближёнными числами

Полученные выше соотношения, описывающие характер распространения погрешностей при единичных вычислениях, непригодны для оценки погрешностей при значительных объёмах вычислений. Они не учитывают возможность взаимного погашения погрешностей, имеющее место при реальных вычислениях и дают сильно завышенные оценки. Поэтому на практике рекомендуют поступать в соответствии с определёнными правилами, подтверждёнными специально проведёнными вычислениями. Приведём некоторые из них.

Правило 1. При небольшом объёме вычислений (несколько десятков операций) и различном числе верных значащих цифр в исходных данных необходимо провести округление до наименьшего числа значащих цифр, оставив, по возможности, одну дополнительную. Также следует поступать и с промежуточными результатами, отбросив дополнительную цифру в окончательном, проведя округление.

Правило 2. При значительном объёме вычислений (сотни и более операций) и при необходимости получить в результате n верных значащих цифр рекомендуется провести округление исходных данных до $(n+2)$ -х верных цифр. Также следует поступать и с промежуточными результатами, проводя округление заключительного до n знаков.

В связи с этим сделаем следующее замечание.

При проведении машинных вычислений, как правило, не предусматривается возможность учёта значащих цифр в промежуточных вычислениях. Однако наличие большого числа знаков в окончательном результате не должно создавать иллюзию высокой точности произведённых вычислений. Доверие, по – прежнему, должны вызывать знаки, указанные в приведённых правилах. Иногда для оценки достоверности полученного результата задачу решают различными методами и совпадающие знаки полагают верными.

В заключение напомним одно из основных правил округления, называемое **округлением с поправкой**.

Если отбрасываемая часть числа превышает половину младшего оставляемого разряда, то соответствующая ему цифра увеличивается на единицу. Если отбрасываемая часть равна половине указанного разряда, то соответствующая цифра увеличивается на единицу, если является нечётной.

Так, например,

$$1.3\mathbf{46} \rightarrow 1.3$$

$$1.3\mathbf{46} \rightarrow 1.3\mathbf{5} \rightarrow 1.4 \quad ,$$

отбрасываемые части выделены. Обратим внимание на различный итоговый результат, зависящий от порядка проводимых округлений. Однако в обоих случаях оставшиеся цифры являются верными.

Погрешность вычисления функции

С целью упрощения математических преобразований, рассмотрим функцию двух аргументов $z = f(x, y)$, предполагаю её достаточно гладкой, т.е. имеющей в достаточной окрестности точки (x, y) непрерывные производные первого порядка. Предположим также, что известны абсолютные $\Delta(x)$, $\Delta(y)$ и относительные $\delta(x)$, $\delta(y)$ погрешности её аргументов. Поставим задачу определить абсолютную $\Delta(z)$ и относительную $\delta(z)$ погрешности вычисления функции в точке (x, y) .

Тогда

$$\Delta(z) = |f(x + \Delta(x), y + \Delta(y)) - f(x, y)|$$

Учтем далее связь между дифференциалом и приращением

$$|\Delta(z)| = |df(x, y)| = \left| \frac{\partial f(x, y)}{\partial x} \Delta(x) + \frac{\partial f(x, y)}{\partial y} \Delta(y) \right| \leq \left| \frac{\partial f(x, y)}{\partial x} \right| \Delta(x) + \left| \frac{\partial f(x, y)}{\partial y} \right| \Delta(y)$$

Аналогично, для функции n аргументов

$$z = f(x_1, x_2, \dots, x_n)$$

получаем

$$|\Delta(z)| \leq \sum_{k=1}^n \left| \frac{\partial f}{\partial x_k} \right| \Delta(x_k)$$

Тогда относительную погрешность можно найти как

$$\delta = \left| \frac{\Delta(z)}{f} \right| \leq \sum_{k=1}^n \left| \frac{\partial f}{\partial x_k} \right| \frac{\Delta(x_k)}{|f|} = \sum_{k=1}^n \left| \frac{\partial \ln f}{\partial x_k} \right| \Delta(x_k)$$

Пример. Вычислить абсолютную и относительную погрешности объема шара $V = \frac{\pi d^3}{6}$, если $\pi \approx 3,14$, $\Delta\pi = 0,0016$, а диаметр равен $d = 3,7 \pm 0,05$

Пусть $d = 3,7$, тогда $V = \frac{\pi d^3}{6} = 27,4$. Найдем абсолютную погрешность по формуле

$$\Delta V = \left| \frac{\partial V}{\partial \pi} \right| \Delta\pi + \left| \frac{\partial V}{\partial d} \right| \Delta d = 8,44 \cdot 0,0016 + 21,5 \cdot 0,05 \approx 1,1,$$

вычислим относительную погрешность $\delta V = \frac{\Delta V}{|V|} = 0,0397$.

Пользуясь этими формулами можно получить погрешности основных арифметических операций:

- 1) Для арифметической операции $a + b$ абсолютная погрешность равна $\Delta a + \Delta b$
- 2) Для арифметической операции $a - b$ абсолютная погрешность равна $\Delta a + \Delta b$
- 3) Для арифметической операции $a \cdot b$ абсолютная погрешность равна $|a|\Delta b + |b|\Delta a$.
- 4) Для арифметической операции $\frac{a}{b}$ абсолютная погрешность равна
$$\frac{|a|\Delta b + |b|\Delta a}{b^2}.$$
- 5) Для арифметической операции a^n абсолютная погрешность равна $na^{n-1}\Delta a$.