Элементы теории погрешности

Классификация погрешностей

Поскольку численные методы предназначены для отыскания приближенного решения задач, не решаемых точными методами, такому решению всегда свойственна некоторая погрешность. Рассмотрим здесь источники погрешности.

1) **Погрешность модели.** Природа слишком сложна и многообразна, чтобы пытаться изучать ее во всей полноте присущих ей в том числе и малозначимых взаимосвязей. Любая (естественная) наука изучает не природу непосредственно, а те модели, которые создаются самой этой наукой для описания природных явлений. **Модель** — это идеализированное описание явления, в котором выявлены основные и игнорируются второстепенные свойства явления. Хорошая модель — это верный шарж, меткая карикатура на изучаемое явление. Естественно, что моделирование, сопровождаемое огрублением и упрощением, вносит погрешность в результат описания явления. Математическая модель создается на языке математики, но оценка погрешности математической модели есть прерогатива не математики, а той науки, в рамках которой изучается явление.

2) **Погрешность исходных данных**. Как правило, математическая модель содержит некоторые параметры, зависящие от исходных данных. Поскольку последние определяются обычно из экспериментов, неизбежно сопровождаемых ошибками измерений, возникает погрешность исходных данных.

Погрешности в решении, обусловленные моделированием и исходными данными, называются **неустранимыми**. Они не зависят от математики и присутствуют, даже если решение поставленной математической задачи найдено точно.

3) **Погрешность метода**. После того как математическая модель создана, вычисления в рамках модели обычно можно выполнять по-разному. Сложная математическая задача заменяется более простой. Например, вычисление определенного интеграла заменяется вычислением интегральной суммы. При этом неизбежно возникает погрешность метода вычислений, которой в дальнейшем мы будем уделять большое внимание при рассмотрении конкретных численных методов.

4) **Погрешность округления**. Любые расчеты, выполняемые как вручную, так и с помощью вычислительной техники, производятся с конечным числом цифр, поэтому приходится прибегать к округлению промежуточных и окончательного ответа. Так возникает погрешность округления, которая может накапливаться в ходе вычислений (опасный процесс, способный обесценить результат вычислений!). Даже те результаты, которые получены точными аналитическими методами, испытывают влияние погрешности округлений и в действительности могут оказаться приближенными.

Полная погрешность является результатом взаимодействия разных видов погрешностей и не может быть меньше, чем наибольшая из составляющих ее погрешностей.

Абсолютная и относительная погрешности

Для оценки погрешности вводятся понятия абсолютной и относительной погрешности.

Пусть x — точное значение некоторой величины (нам оно неизвестно и никогда не будет известно, поскольку определяется с помощью измерений, страдающих неточностями); a — приближенное значение той же величины ($a \approx x$). Абсолютная погрешность приближенного числа a определяется как $\Delta_a = |x - a|$. Но поскольку x неизвестно, то и абсолютную погрешность мы узнать не можем! Чтобы разрешить парадокс, вводят предельную абсолютную погрешность Δ_a^* — такое значение, которое абсолютная погрешность заведомо не превзойдет при данном способе измерений

$$|x - a| \le \Delta_a^*. \qquad (1)$$

Из выражения (1) следует, что $a-\Delta_a^* \le x \le a+\Delta_a^*$, поэтому желательно возможно меньшее значение Δ_a^* — это уменьшит длину интервала, содержащего искомое значение x и, следовательно, понизит неопределенность в наших знаниях об этой величине.

В технике формулу (1) часто записывают в виде $x = a \pm \Delta_a^*$, причем Δ_a^* называется *допусками*. Никакое изделие не может быть изготовлено с абсолютно точным соблюдением номинальных размеров, допуски показывают возможные (*допустимые*) отклонения от номинала.

Итак, абсолютная погрешность оценивает точность измерений, но эта оценка *неполна*, поскольку не учитывает характерный размер изучаемого явления (объекта). Так, например, абсолютная погрешность в 1 см при измерении длины комнаты — вероятно, вполне приемлемая точность, но при измерении роста человека эта же погрешность будет сочтена непозволительно грубой.

Более информативным показателем качества измерений является *относительная погрешность* δ_a (соответственно *предельная относительная погрешность* δ_a^*) приближенного числа a как отношение абсолютной погрешности (предельной абсолютной погрешности) к модулю числа a

$$\delta_a = \frac{\Delta_a}{|a|}, \quad \delta_a^* = \frac{\Delta_a^*}{|a|}.$$

Относительная погрешность является величиной *безразмерной*, т. е. не зависит от выбора системы единиц измерения, что позволяет сравнивать качество измерений разнородных величин (бессмысленным является вопрос о том, что больше: 1 кг или 1 м, — но сравнение *качества* измерений массы и длины в терминах относительной погрешности вполне допустимо). Измеряется δ_a (δ_a^*) в долях единицы или в процентах.

Пример. Согласно ныне действующим (2015 г.) определениям международного Комитета по константам для науки и технологии входящая в закон всемирного тяготения гравитационная постоянная

$$\gamma = (6.67259 \pm 0.00085) \cdot 10^{-11}\ \text{м}^3 \cdot \text{кг}^{-1} \cdot \text{с}^{-2},$$

а заряд электрона

$$e = (1.602\,17733 \pm 0.000\,00049) \cdot 10^{-19}\ \text{Кл}.$$

Сравнить точность определения этих фундаментальных физических постоянных.

Решение. Для гравитационной постоянной предельная относительная погрешность

$$\delta_\gamma^* = \frac{0.00085}{6.67259} = 1.27 \cdot 10^{-4},$$

а для заряда электрона

$$\delta_e^* = \frac{0.000\,000\,49}{1.602\,177\,33} = 3.1 \cdot 10^{-7}.$$

Таким образом, в последнем случае относительная погрешность оказывается на три порядка меньшей, т. е. заряд электрона определен существенно точнее, чем гравитационная постоянная.

С понятиями абсолютной и относительной погрешности связаны понятия верных и значащих цифр.

Если абсолютная погрешность приближенного числа не превышает единицы последнего (самого правого) разряда его десятичной записи, то цифры числа называют **верными** (или **точными**).

По умолчанию десятичная запись приближенного числа должна содержать только верные цифры, и тогда по записи числа сразу можно узнать предельную абсолютную погрешность, с которой оно известно.

Цифры, не являющиеся верными, называются **сомнительными**.

Пример. Даны приближенные числа $a = 8.6$, $b = 8.60$, $c = 3200$, $d = 3.2 \cdot 10^3$. Указать предельную абсолютную погрешность для каждого числа.

Решение. Для числа a погрешность $\Delta_a^* \le 0.1$, для числа b $\Delta_b^* \le 0.01$, для числа c $\Delta_c^* \le 1$, для числа d $\Delta_d^* \le 0.1 \cdot 10^3 = 100$.

Итак, числа a и b, c и d, равные с точки зрения «обычной» математики, существенно различны в вычислительной математике: из абсолютной погрешности мы заключаем, что число b известно точнее, чем число a, а число c — точнее, чем d. Кроме того, нуль, стоящий справа в дробной части десятичного числа, важен, и им нельзя пренебрегать, если мы хотим составить верное суждение о точности числа.

Значащими цифрами приближенного числа называются все цифры его десятичной записи, кроме нулей, находящихся левее первой отличной от нуля цифры.

Пример. Числа 0.001 307 и 6.0400 имеют соответственно четыре и пять значащих цифр. Итак, нули, находящиеся слева, значащими не являются, а нуль, записанный в конце десятичной дроби, всегда является значащей цифрой.

Действия с приближенными числами

Теорема 1. Абсолютная погрешность алгебраической суммы нескольких приближенных чисел не превышает суммы абсолютных погрешностей этих чисел.

В частности, для суммы двух чисел a и b любого знака получаем $\Delta_{a\pm b} \le \Delta_a + \Delta_b$.

Из этой теоремы следует, что абсолютная погрешность алгебраической суммы не меньше абсолютной погрешности наименее точного из слагаемых, т. е. увеличение точности за счет других слагаемых невозможно. Поэтому бессмысленно сохранять излишние десятичные знаки в более точных слагаемых. Отсюда вытекает следующее.

Правило сложения и вычитания приближенных чисел:

1) выделить наименее точное число (или числа), т. е. такое, в десятичной записи которого наименьшее число верных десятичных знаков;

2) округлить остальные числа так, чтобы каждое из них содержало на один (запасной) знак больше, чем выделенное число;

3) выполнить сложение и вычитание с учетом сохраненных знаков;

4) полученный результат округлить до предпоследнего знака.

Напомним правила округления числа, т. е. его замены числом с меньшим количеством значащих цифр:

1) если первая из отбрасываемых цифр меньше 5, то сохраняемые десятичные знаки оставляют без изменения;

2) если первая из отбрасываемых цифр больше 5, то последний из сохраняемых знаков увеличивают на 1;

3) если первая из отбрасываемых цифр равна 5, а среди следующих за ней цифр есть отличные от нуля, то последний из сохраняемых знаков увеличивают на 1;

4) если первая из отбрасываемых цифр равна 5, а все последующие — нули, то последний из сохраняемых десятичных знаков увеличивают

на 1, когда он нечетен, и сохраняют неизменным, когда он четен (правило четной цифры).

Пример. Округляя число 53.471 до одного знака после запятой, получим 53.5 (правило 2), а при округлении до двух знаков после запятой получим 53.47 (правило 1). Округляя число 7.825 001 до трех знаков после запятой, получим 7.825 (правило 3). Округляя число 8.465 до сотых долей, получим 8.4**6**; сохраняемая цифра не увеличивается на единицу, поскольку она четна. При округлении числа 8.475 до сотых долей получим 8.4**8** — нечетная цифра увеличилась на единицу (правило 4).

Смысл правила 4 в том, что при многочисленных округлениях избыточные числа будут встречаться примерно с той же частотой, что и недостаточные, и произойдет частичная взаимная компенсация погрешностей округления; результат окажется более точным.

Теперь проиллюстрируем правило сложения и вычитания приближенных чисел.

Пример. Найти сумму приближенных чисел $a = 414.8$, $b = 0.025$, $c = 24.17$, $d = 0.000326$. По умолчанию все цифры в этих числах считать верными.

Решение. Наименее точное слагаемое — a, поскольку в нем только один верный десятичный знак. Округлим остальные слагаемые до двух знаков после запятой: $b \to 0.02$, $c \to 24.17$, $d \to 0.00$. Теперь сложим округленные числа: $414.8 + 0.02 + 24.17 + 0.00 = 438.99$. Округляя результат до одного знака после запятой, получим окончательный ответ: 439.0.

Теорема 2. Относительная погрешность произведения (частного) приближенных чисел не превышает суммы относительных погрешностей этих чисел.

В частности, для трех чисел $\delta_{ab/c} \le \delta_a + \delta_b + \delta_c$.

Из теоремы следует, что относительная погрешность произведения и частного не может быть меньше относительной погрешности наименее точного из исходных чисел (т.е. имеющего меньше всего верных значащих цифр). Поскольку относительная погрешность числа определяется количеством его верных значащих цифр, то при умножении и делении бессмысленно оставлять значащих цифр больше, чем их было в исходном числе с наименьшим количеством верных значащих цифр.

Отсюда вытекает следующее правило.

Правило умножения и деления приближенных чисел:

1) из всех чисел, которые предстоит умножать и делить, выделить наименее точное — то, в котором меньше всего верных значащих цифр;

2) округлить остальные числа так, чтобы каждое из них содержало на одну (запасную) значащую цифру больше, чем выделенное число;

3) выполнить умножение и деление округленных чисел с учетом сохраненных значащих цифр;

4) оставить в ответе столько значащих цифр, сколько их было в наименее точном числе.

Пример. Найти произведение приближенных чисел $a = 3.5$ и $b = 83.368$, все цифры которых верные.

Решение. В первом числе две верные значащие цифры, а во втором — пять. Второе число округлим до трех значащих цифр: $b \to 83.4$. После округления перемножим числа: $ab = 3.5 \cdot 83.4 = 291.9 \approx 2.9 \cdot 10^2$. В ответе оставлены две значащие цифры — столько, сколько их было во множителе с наименьшим количеством верных значащих цифр.