ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ

ЭВМ и методы решения задач в численном виде (*численные методы*) глубоко проникли в самые разные сферы человеческой деятельности. Особо отметим задачи инженерной практики, когда интересующие исследователя показатели недоступны непосредственному наблюдению и могут быть получены лишь косвенно, в результате математической обработки измерений некоторых других параметров. При этом неизбежно возникают погрешности, как на этапе выполнения замеров, так и на этапе собственно вычислений.

Роль ЭВМ в научных исследованиях раскрывает научно-популярное издание [1]. В [2] излагаются методы теории вероятностей и математической статистики в применении к задачам обработки результатов измерений на ЭВМ. Современный учебник высшей школы по численным методам представляет издание [3]. В частности, в первой главе дано изложение этапов вычислительного эксперимента и основ теории погрешностей. В методичке [4] доступно изложено правило вычисления фактической погрешности на последовательности вложенных сеток. Основные сведения о численных методах содержатся также в [5]. Отметим, что в библиотеках ТГАСУ имеется большое количество экземпляров этого издания. Задачи и упражнения из раздела "Элементы теории погрешностей" включены, в частности, в пособие [6]. Большое количество задач (с вариантами и решенными примерами) и основные сведения по применению численных методов предлагает сборник [7].

В первой части данных методических указаний излагаются основы теории погрешностей. Излагаемый материал иллюстрируется разбором примеров и сопровождается образцами выполнения контрольных работ. Вопросы к самопроверке даются во второй части, и в третьей части содержатся варианты заданий к контрольным работам. В качестве вычислительного средства допустимо использовать микрокалькулятор.

1. ЭЛЕМЕНТЫ ТЕОРИИ ПОГРЕШНОСТЕЙ

1.1. Основные источники и классификация погрешностей

Под *погрешностью* понимается некоторая величина, характеризующая точность результата. Анализ погрешностей должен являться непременной составной частью любого серьезного вычисления, независимо от того, производится это вычисление вручную или с помощью ЭВМ. Известны, по крайней мере, четыре источника погрешностей.

1. В курсе математического моделирования [1] обсуждаются погрешности, возникающие в силу неточности самой модели физического явления/процесса. Величина погрешности, вносимой в результат математической моделью, может возрасти, если в модели не учтены какие-либо важные характеристики изучаемого явления.

2. Исходная информация очень редко бывает точной, т.к. исходные величины являются экспериментальными результатами или основаны на экспериментальных оценках. Такую погрешность относят к *неустранимой погрешности*, однако при аккуратном учете вероятностной природы ошибок измерения/наблюдения ее можно значительно уменьшить и в пределе (при бесконечном увеличении числа измерений) свести к нулю [2]. Не следует стремиться к уменьшению погрешности одних данных, оставляя без внимания другие. Это не приводит к повышению точности результата, поэтому на практике исходные данные задаются примерно с одинаковой точностью.

3. Любой численный метод в силу его конечности связан с возникновением погрешности. Тогда задача состоит в том, чтобы оценить скорость убывания погрешности данного метода в зависимости от количества проведенных вычислений. В результате удается либо вычислить погрешность на каждом шаге численного метода и тем самым сравнить разные методы по их производительности с тем, чтобы выбрать наиболее приемлемый метод для решения поставленной задачи, либо оценить требуемое количество вычислений для обеспечения заданной погрешности. Эта задача решается с применением аналитических построений [3–5].

4. Кроме того, процессы вычислений могут вносить в результат погрешности, возникающие из-за конечности разрядной сетки ЭВМ – *ошибки округления*.

1.2. Правила округления чисел

Всякое нецелое число можно представить в виде конечной или бесконечной десятичной дроби. В основном на практике имеют дело с приближенными числами – конечными дробями. *Значащими цифрами* числа называют все сохраняемые цифры в его десятичном представлении, начиная с первой ненулевой слева.

Например, в числе $x = 0.00267$ значащие только цифры 2, 6, 7; первые три нуля – незначащие, ибо они служат вспомогательной цели – определению положения цифр 2, 6, 7. Поэтому может быть принята запись· $x = 2.67 \cdot 10^{-3}$. Если при записи числа 237000 нужно показать, что три последних нуля не являются значащими цифрами, то пишут это число следующим образом:

$$237 \cdot 10^3 \text{ или } 23.7 \cdot 10^4 \text{ или } 2.37 \cdot 10^5 \text{ или } 0.237 \cdot 10^6.$$

Предположим, что вычисления производятся на машине, в которой каждое число представляется пятью значащими цифрами, и что необходимо сложить два числа:

$$\begin{array}{r} 9.2654 \\ +\,7.1625 \\ \hline 16.4279 \end{array}$$

– оба числа точные, но сумма их содержит 6 значащих цифр и не помещается в разрядной сетке вычислительной машины. Поэтому 6-значный результат будет усечен до 16.427. В процессе *усечения* просто отбрасываются все цифры числа за последним сохраняемым разрядом.

При применении правила округления по усечению погрешность округления не превосходит единицы того разряда, в котором записана последняя значащая цифра.

При ручном счете под округлением числа, записанного в десятичной форме, понимается замена числа x (точного или приближенного) на число x_1 с меньшим количеством значащих цифр. Число x_1 выбирают так, чтобы погрешность округления $|x - x_1|$ была минимальной.

Правило округления (по дополнению). Чтобы округлить число до n значащих цифр, отбрасывают все его цифры, стоящие справа от n-й значащей цифры, или, если это нужно для сохранения разрядов, заменяют их нулями. При этом:

1) если первая из отброшенных цифр меньше 5, то оставшиеся десятичные знаки сохраняются без изменения;

2) если первая из отброшенных цифр больше 5, то к последней оставшейся цифре прибавляется единица;

3) если первая из отброшенных цифр равна 5 и среди отброшенных цифр имеются ненулевые, то последняя оставшаяся цифра увеличивается на единицу;

3а) если же первая из отброшенных цифр равна 5 и все остальные отброшенные цифры являются нулями, то последняя оставшаяся цифра сохраняется неизменной, если она четная, и увеличивается на единицу, если она нечетная (правило четной цифры).

Иными словами, если при округлении числа отбрасывается меньше половины единицы последнего сохраняемого разряда, то цифры всех сохраненных разрядов остаются неизменными; если больше, то число увеличивается на единицу. В исключительном случае, когда отброшенная часть в точности равна половине последнего сохраненного разряда, для компенсации ошибок округления используется правило четной цифры.

При применении правила округления по дополнению погрешность округления не превосходит половины единицы десятичного разряда, определяемого последней оставленной значащей цифрой.

Вопросы округления относятся только к действительным числам. При выполнении арифметических операций с целыми числами потребность в округлении не возникает. Сумма, разность и произведение целых чисел сами являются целыми числами: если результат слишком велик и не вмещается в разрядную сетку ЭВМ, то возникает ошибка в программе. При делении целых чисел возникающая дробная часть просто игнорируется.

1.3. Неустранимая погрешность приближенного числа

Приближенным числом $\tilde{x}$ называется число, незначительно отличающееся от точного x и заменяющее его в вычислениях. Если известно, что $\tilde{x} < x$, то $\tilde{x}$ называется приближенным значением числа x *по недостатку*, если же $\tilde{x} > x$, то – *по избытку*.

Например, для $\sqrt{2}$ число 1.41 – приближенное значение по недостатку, а число 1.42 – по избытку, так как $1.41 < \sqrt{2} < 1.42$.

В этом случае пишут $1.415 = \tilde{x} \approx x = \sqrt{2}$.

Абсолютная ошибка есть разность между истинным значением величины и ее приближенным значением $e_x = x - \tilde{x}$. Таким образом, $x = \tilde{x} + e_x$.

Абсолютной погрешностью числа $\tilde{x}$ называется величина Δ_x, удовлетворяющая условию $|x - \tilde{x}| \le \Delta_x$. Тогда $|e_x| \le \Delta_x$.

Границы, в которых заключено точное значение числа x:

$$\tilde{x} - \Delta_x \le x \le \tilde{x} + \Delta_x.$$

Часто полученное неравенство, задающее область неопределенности точного значения, записывают в виде: $x = \tilde{x} \pm \Delta_x$.

Пример 1.

Даны приближенные значения числа x = 2/3: $\tilde{x}_1 = 0.6$; $\tilde{x}_2 = 0.66$; $\tilde{x}_3 = 0.67$. Какое из этих приближений лучше?

Решение. Находим

$$\Delta x_1 = \left|\frac{2}{3} - 0.6\right| = \left|\frac{2}{3} - \frac{3}{5}\right| = \frac{1}{15},$$

$$\Delta x_2 = \left|\frac{2}{3} - 0.66\right| = \left|\frac{2}{3} - \frac{33}{50}\right| = \frac{1}{150},$$

$$\Delta x_3 = \left|\frac{2}{3} - 0.67\right| = \left|\frac{2}{3} - \frac{67}{100}\right| = \frac{1}{300}.$$

Лучшим приближением числа x является $\tilde{x}_3 = 0.67$.

Пример 2.

Длина детали x (см) заключена в границах $33 \le x \le 34$. Найти предельную абсолютную погрешность измерения длины детали.

Решение. Среднее арифметическое $\tilde{x} = \frac{33 + 34}{2} = 33.5$ (см). Тогда $\Delta_x \le 0.5$ (см). Длина детали x, найденная с точностью $\Delta_x = 0.5$ (см), заключена в границах:

$$33.5 - 0.5 \le x \le 33.5 + 0.5; \quad x = 33.5 \pm 0.5 \text{ (см)}.$$

Пример 3.

Определить предельную абсолютную погрешность числа $\tilde{x} = 3.14$, заменяющего число π.

Решение. Так как имеет место неравенство $3.14 < p < 3.15$, то $|\tilde{x} \;\; p| < 0.01$ и, следовательно, можно принять $\Delta_x = 0.01$.

Если учесть, что $3.14 < p < 3.142$, то будем иметь лучшую оценку: $\Delta_x = 0.002$.

Отметим, что абсолютная погрешность – величина размерная, имеющая размерность $\tilde{x}$. *При вычислении абсолютной погрешности результат принято округлять только в большую сторону* (в отличие от обычного правила округления), так как при округлениях границы неопределенности числа, как правило, увеличиваются. Скажем, Вы собрались в Новосибирск (расстояние 225 ± 7 км) на серийном автомобиле ВАЗ 2109 (расход горючего 8 ± 0.3 л на 100 км). Тогда, залив в бензобак $\frac{232 \cdot 8.3}{100} = 19.256 \approx 19$ литров бензина, Вы рискуете не доехать $\frac{0.256 \cdot 100}{8} = 3.2 \approx 3$ км до ближайшей новосибирской заправки.

С помощью абсолютной погрешности можно отразить количественную сторону погрешности некоторого результата, но не качественную. Например, измерение расстояния, преодоленного прыгуном в длину, проведено с точностью 0.5 см. С такой же погрешностью измерена длина беговой дорожки стадиона. Очевидно, что второе измерение выполнено более качественно, чем первое. Или, чтобы было еще заметнее, усилие, которое требуется, чтобы перейти с быстрого шага 5 км/час на бег 10 км/час, заставит Вас вспотеть. На такое же приращение скорости движения автомобиля со 100 км/час Вы, пожалуй, и не обратите внимание.

Относительная ошибка определяется как отношение абсолютной ошибки к приближению. Казалось бы, что более естественно определить ее как отношение абсолютной ошибки к точному значению, но обычно точное значение нам неизвестно. Все, что обычно бывает известно – это приближенное значение величины и оценка ошибки (абсолютная погрешность) или границы максимальной возможной величины ошибки (сверху и снизу, если они несимметричны).

Относительной погрешностью называется величина δ*x*, удовлетворяющая условию $\left|\dfrac{x-\widetilde{x}}{\widetilde{x}}\right| \le \text{д}x$.

Относительная погрешность суть величина безразмерная и обычно выражается в процентах. Ее принято записывать с двумя, тремя знаками в сторону увеличения. Для величин, близких по значению к единице, абсолютные и относительные погрешности почти одинаковы. Для очень больших или для очень малых величин относительная и абсолютная погрешности представляются совершенно разными числами. Так, если точное значение некоторой величины равно 0.00006, а приближенное значение ≈ 0.00005, то абсолютная ошибка равна 10^{-5}, в то время как относительная ошибка составляет 0.2 = 20%. С другой стороны, если точное значение равно 100500, а приближенное ≈ 100000, то, хотя абсолютная ошибка составляет 500 >> 10^{-5}, относительная ошибка составляет всего 0.005 = 0.5% < 20%. *Относительная погрешность лучше характеризует точность результата.*

Пример 4.

Вес 1 дм^3 воды при t = 0°C равен p = 1000.14 (± 0.01) г. Определить предельную относительную погрешность результата взвешивания.

Решение. Очевидно, что $\Delta_p = 0.01$ г и $p \ge 1000.004$ г. Следовательно

$$\text{д}_p = \frac{0.01}{1000.014} \le \frac{0.01}{1000.004} < \frac{0.01}{1000.0} = 10^{-6},\ \text{д}_p \approx 10^{-4}\%.$$

Пример 5.

а) Пусть $x = 3.141592$ и $\widetilde{x} = 3.14$. Тогда

$$\Delta_x = |x-\widetilde{x}| = |3.141592 - 3.14| = 0.001592,$$

$$\text{д}_x = \frac{|x-\widetilde{x}|}{|x|} = \frac{0.001592}{3.141592} \approx 0.000507 \text{ или } 0.051\%.$$

б) Пусть $x = 1000\,000$ и $\widetilde{x} = 999996$. Тогда

$$\Delta_x = 4;\ \text{д}_x = \frac{4}{1000\,000} = 0.000004 \text{ или } 0.0004\%.$$

в) Пусть $x = 0.000012$ и $\widetilde{x} = 0.000009$. Тогда

$$\Delta_x = 0.000003 \text{ и } \text{д}_x = \frac{0.000003}{0.000012} = 0.25 \text{ или } 25\%.$$

Пример 6.

Для *записи действительных чисел в ЭВМ* применяется двоичная система с плавающей запятой: $x \approx \pm 2^p \sum_{k=1}^{t} \text{б}_k 2^{-k} = \widetilde{x}$, где параметр p определяет порядок числа, а параметр t – количество значащих цифр в числе. При этом $|p| \le p_0 = 63$ (в зависимости от числа позиций в машинной ячейке, отводимых под порядок), $0 \ne \alpha_1 = 1$ (т.е. число записывается в нормализованном виде) и параметр $t = 24$ определяет возможную длину мантиссы. Абсолютная погрешность такой записи не больше единицы последнего разряда в $\widetilde{x}$, т.е. $|x-\widetilde{x}| \le 2^{p-t} = \Delta_x$, а относительная погрешность может быть найдена следующим образом:

$$\text{д}_x = \left|\frac{x-\widetilde{x}}{\widetilde{x}}\right| \le \frac{\Delta_x}{|\,\widetilde{x}\,|} = \frac{2^{p-t}}{2^p \sum_{k=1}^{t} \text{б}_k 2^{-k}} \le 2^{-t}.$$

1.4. Понятие о верных значащих цифрах числа

Точность приближенного числа зависит не от количества значащих цифр, а от количества *верных значащих цифр*. Приближенное число $\tilde{x}$ называется *верным в широком смысле*, если предельная абсолютная погрешность Δ_x не превосходит единицы того разряда, в котором записана последняя значащая цифра.

Приближенное число $\tilde{x}$ называется *верным в узком (строгом) смысле*, если предельная абсолютная погрешность Δ_x не превосходит половины единицы того разряда, в котором записана последняя значащая цифра.

При ручных вычислениях желательно сохранять такое количество значащих цифр, которое больше верных на 1, 2 единицы. Эти 1, 2 последние цифры относят к разряду сомнительных цифр и используют для округления.

В силу неравенства

$$|x - x_1| \le |x - \tilde{x}| + |\tilde{x} - x_1|$$

предельная абсолютная погрешность числа x_1 складывается из абсолютной погрешности числа $\tilde{x}$ и погрешности округления.

Заметим, что если в исходных данных задачи дано число без указания его абсолютной или относительной погрешности, то предполагается, что все цифры данного числа являются верными значащими цифрами. Если число имеет лишь верные цифры, его округленное значение имеет также лишь верные цифры.

1.5. Образцы выполнения контрольной работы 1

Задание 1.

Определить, какое равенство точнее:

$$9/11 = 0.818; \sqrt{18} = 4.24.$$

Решение. Находим значения данных выражений с большим числом десятичных знаков:

$$a_1 = 9/11 = 0.81818\ldots,\ a_2 = \sqrt{18} = 4.2426\ldots.$$

Затем вычисляем предельные абсолютные погрешности, округляя их с избытком:

$$\Delta_{a_1} = |0.81818 - 0.818| \le 0.00019,\ \Delta_{a_2} = |4.2426 - 4.24| \le 0.0027\ .$$

Предельные относительные погрешности составляют

$$\text{д}_{a_1} = \frac{\Delta_{a_1}}{a_1} = \frac{0.00019}{0.818} = 0.00024 = 0.024\%\ ;$$

$$\text{д}_{a_2} = \frac{\Delta_{a_2}}{a_2} = \frac{0.0027}{4.424} = 0.00064 = 0.064\%\ .$$

Так как $\text{д}_{a_1} < \text{д}_{a_2}$, то равенство 9/11 = 0.818 является более точным.

Задание 2.

Даны приближенные значения и погрешности чисел:

а) 72.353 (± 0.062); б) 2.3544; δ = 0.2%.

Округлить сомнительные цифры числа, оставив верные знаки: а) в узком смысле; б) в широком смысле. Определить предельную абсолютную погрешность результата.

Решение. а) Пусть 72.353 (± 0.026) = a. Согласно условию, абсолютная погрешность $\Delta_a = 0.026 < 0.05$; это означает, что в числе 72.353 верными в узком смысле являются цифры 7, 2, 3. По правилам округления найдем приближенное значение числа, сохранив десятые доли: a_1= 72.4; при этом совершаем ошибку

$$\Delta_{a_1} = \Delta_a + \Delta_{\text{окр}} = 0.026 + 0.047 = 0.073.$$

Полученная погрешность больше 0.05; значит, нужно уменьшить число цифр в приближенном числе до двух:

$$a_2 = 72;\ \Delta_{a_2} = \Delta_a + \Delta_{\text{окр}} = 0.026 + 0.353 = 0.379.$$

Так как $\Delta_{a_2} < 0.5$, то обе оставшиеся цифры верны в узком смысле.

б) Пусть b = 2.3544, причем относительная погрешность δ_b = 0.2%; тогда $\Delta_b = b \cdot д_b = 0.00471$. В данном числе верными в широком смысле являются три цифры, поэтому округляем его, сохраняя эти три цифры:

$$b_1 = 2.35;\ \Delta_{b_1} = 0.0044 + 0.00471 = 0.00911 < 0.01.$$

Значит, и в округленном числе 2.35 все три цифры верны в широком смысле.

Задание 3.

Даны числа: а) 0.4357; б) 12.384.

Найти предельные абсолютные и относительные погрешности чисел, если они имеют только верные цифры: а) в узком смысле; б) в широком смысле.

Решение. а) Так как все четыре цифры числа a = 0.4357 верны с точностью до половины пятого разряда, то абсолютная погрешность Δ_a = 0.00005, а относительная погрешность

$$\delta_a = \Delta_a/a < 1/(2\cdot4\cdot10^3) = 0.000125 = 0.0125\%.$$

б) Так как все пять цифр числа b = 12.384 верны с точностью до единицы третьего разряда, то Δ_b = 0.001, а

$$\delta_b = \Delta_b/b < 1/(1\cdot10^4) = 0.0001 = 0.01\%.$$

1.6. Вычислительная погрешность

Одним из наиболее важных вопросов в численном анализе является вопрос о том, как ошибка, возникшая в определенном месте в ходе вычисления, распространяется дальше, т.е. становится ли ее влияние больше или меньше по мере того, как производятся последующие операции.

Пример 7.

Пусть требуется найти один из корней уравнения $x^2 + 0.4002x + 0.00008 = 0$, используя вычисление с точностью до четвертой значащей цифры. Используя обычную формулу

$x = \dfrac{-b + \sqrt{b^2 - 4ac}}{2a}$, мы получаем ответ –0.00015. Поскольку правильный корень, найденный при вычислении с восемью значащими цифрами, равен –0.0002, то ошибка ответа составляет

$$\frac{0.00005}{0.0002}100\% = 25\%.$$

В данном примере источником погрешности явилось применение четырехзначной арифметики. Но не нужно думать, что увеличение количества значащих цифр решит все проблемы.

Пример 8.

Рассмотрим ряд Тейлора для синуса:

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots.$$

Считается, что этот ряд годится для любого конечного угла, а ошибка, происходящая от ограничения ряда конечным числом членов, не превосходит по модулю первого отброшенного члена ряда. Это было бы справедливо, если бы существовал способ производить арифметические вычисления с бесконечным числом значащих цифр. Практически, для больших углов этот ряд совершенно бесполезен. Предположим, что нужно вычислить sin 1470° (≈ 25.7 рад), и мы намерены производить вычисления с 8 значащими цифрами и остановиться тогда, когда очередной член ряда будет по абсолютной величине $<10^{-8}$. Выданный машиной результат будет равен 24.25401855, с большим числом десятичных знаков и совершенно без всякого смысла. Даже если производить вычисления с 16 значащими цифрами, то вместо sin 2500° машина выдаст число 295.

Существуют источники ошибок, и не связанные с конечным числом значащих цифр.

Пример 9.

Рассмотрим следующую систему уравнений:

$$\begin{cases} 5x - 331y = 3.5; \\ 6x - 397y = 5.2. \end{cases}$$

Точное решение этот системы x = 331.7, y = 5.000. Попытаемся оценить количество достоверных значащих цифр в этом решении. Изменим константу в правой части второго уравнения с 5.2 до 5.1, т.е. примерно на 2%. Решение будет x = 298.6, y = 4.5 – результат изменился на 10%. Более того, если подставить в уравнения x = 358.173, y = 5.4, то при округлении вычисленные значения левых частей дадут те же самые правые части, что и в исходной системе уравнений.

По-видимому, в этом случае можно считать, что величины x и y имеют не более одной достоверной значащей цифры, независимо от общего количества цифр в результате. Причина неточности в данном случае – малая величина определителя системы уравнений. Геометрически это означает, что две прямые, представленные указанными уравнениями, почти параллельны.

Из приведенных примеров становится ясно, что без анализа ошибок вычисления нельзя сказать что-либо определенное относительно точности результата.

1.7. Неустранимая погрешность функции

Рассмотрим оценку погрешностей простейших действий с приближенными числами.

Сложение.

Пусть $x = \tilde{x} + e_x$, $y = \tilde{y} + e_y$. Тогда в результате сложения имеем

$$a = x + y = \tilde{x} + e_x + \tilde{y} + e_y = (\tilde{x} + \tilde{y}) + (e_x + e_y) = \tilde{a} + \varepsilon .$$

Поэтому $|\varepsilon| \le |e_x| + |e_y| \le \Delta_x + \Delta_y$. *Граница погрешности суммы не больше суммы границ погрешностей слагаемых.*

С другой стороны, абсолютная погрешность суммы не может быть меньше погрешности наиболее точного из слагае-

мых. Известно *практическое правило*: при выполнении приближенных вычислений число значащих цифр промежуточных результатов не должно превышать числа верных цифр более чем на одну или две единицы. Окончательный результат может содержать не более чем одну излишнюю значащую цифру.

Пример 10.

Поведение вычислительной погрешности зависит от правила округлений и алгоритма численного решения задачи. Поэтому сложение нескольких приближенных чисел, например: S = 2.17+12.3871+1.198683+0.006732, следует производить в следующем порядке:

1) выбираем число с наименьшим числом знаков после запятой: 2.17;

2) другие числа округляем до этого числа, сохранив два запасных разряда: 12.3171; 1.1987; 0.0067;

3) суммируем: S = 15.7725;

4) сумму S округляем на один разряд, $\tilde{S} = 15.772$.

Относительная погрешность суммы находится по правилу

$$\text{д}_a = \frac{1}{\tilde{a}}\left(\tilde{x}\,\text{д}_x + \tilde{y}\,\text{д}_y\right).$$

Пусть, в частности, $\max\{\delta_x, \delta_y\} = M$, $\min\{\delta_x, \delta_y\} = m$. Тогда

$$\text{д}_a \le \frac{1}{\tilde{a}}\left(\tilde{x}M + \tilde{y}M\right) = M;$$

$$\text{д}_a \ge \frac{1}{\tilde{a}}\left(\tilde{x}m + \tilde{y}m\right) = m.$$

Таким образом, $m \le \text{д}_a \le M$.

При сложении нескольких слагаемых, если $\Delta_x = \ldots = \Delta_y = \Delta$, то $\varepsilon \le n \cdot \Delta$, где n – число слагаемых.

Заметим, что обычно погрешности имеют разные знаки и при их сложении происходит частичная компенсация. В таких случаях можно пользоваться правилом *статистической оценки* абсолютной погрешности [3, 6]: $\varepsilon \approx \sqrt{3n} \cdot \Delta$.

Аналогично можно получить следующую рекомендацию. При сложении нескольких положительных чисел *максимальная возможная ошибка (абсолютная или относительная), обусловленная округлением, будет меньше, если сначала складывать меньшие числа*. Это кажется удивительным, поскольку классическая математическая подготовка инженера основана на утверждении, что от перемены мест слагаемых сумма не меняется. Однако если вспомнить, что вычисления на ЭВМ производятся с конечной мантиссой, то можно понять, что если вначале малые числа можно складывать, не опасаясь потери значащих цифр, то при обратном порядке на фоне уже полученного большого результата добавление малых чисел окажется незаметным.

Вычитание.

Нетрудно видеть, что *при вычитании приближенных чисел их абсолютные погрешности складываются*.

Вычитание двух почти равных чисел является еще одним случаем потери верных значащих цифр: даже при очень маленьких ошибках обоих чисел относительная ошибка разности (после деления на результат) может оказаться очень большой. Покажем, как повышение точности может быть достигнуто за счет несложного алгебраического преобразования.

Пример 11.

Отыскивается наименьший корень уравнения $y^2 - 140y + 1 = 0$. Пусть снова используется четырехзначная арифметика. Имеем $y = 70 - \sqrt{4899} \approx 70 - 69.99 = 0.01$.

То же самое значение y можно, избавивших от приближенного вычитания близких величин, представить в виде

$$y = \frac{1}{70 + \sqrt{4899}} = \frac{1}{70 + 69.99} = \frac{1}{140} = 0.00714285$$

и после округления получить $y \approx 0.007143$. Нетрудно проверить, что во втором случае точность результата существенно выше.

Умножение.

$$x \cdot y = \left(\tilde{x} + e_x\right)\left(\tilde{y} + e_y\right) = \tilde{x} \cdot \tilde{y} + \tilde{x} e_y + \tilde{y} e_x + e_x e_y = \tilde{x} \cdot \tilde{y} + \text{е}.$$

Таким образом, $\text{е} = \tilde{x} \cdot e_y + \tilde{y} \cdot e_x + e_x \cdot e_y$.

Если теперь разделить обе части равенства на $\tilde{x} \cdot \tilde{y}$, то получим:

$$\frac{\varepsilon}{\tilde{x} \cdot \tilde{y}} = \frac{e_y}{\tilde{y}} + \frac{e_x}{\tilde{x}} + \frac{e_x}{\tilde{x}} \cdot \frac{e_y}{\tilde{y}}.$$

Отсюда вытекает оценка границы относительной погрешности

$$\left|\frac{\text{е}}{\tilde{x} \cdot \tilde{y}}\right| \le \text{д}_x + \text{д}_y + \text{д}_x \text{д}_y.$$

В большом числе реальных задач погрешность является малой величиной сравнительно с точным и приближенным значениями, поэтому произведением ошибок можно пренебречь и пользоваться для погрешности ε неточным равенством $\text{е} \approx \tilde{x} e_y + \tilde{y} e_x$. Аналогично,

$$\left|\frac{\varepsilon}{\tilde{x} \cdot \tilde{y}}\right| \approx \delta_x + \delta_y$$

– *относительные погрешности сомножителей складываются*.

В частности, при умножении приближенного числа на множитель k относительная погрешность δ не меняется, а абсолютная увеличивается в $|k|$ раз.

Пример 12.

Найти p = 2.25·1.0113. Вычисления проводим по следующему правилу:

1) выбираем число с меньшим числом значащих цифр после запятой: 2.25;

2) второе число округляем до этого разряда, сохранив один запасной знак: 1.011;

3) вычисляем 2.25·1.011 = 2.27475;

4) округляем: p = 2.27.

Деление.

Пусть $a = x/y$ и сделаем дополнительное предположение

$\left|\Delta_y\right| < \left|\tilde{y}\right|$, так как иначе приближенное значение делителя y может оказаться равным нулю. Тогда справедливы соотношения

$$\left|\frac{\Delta_a}{a}\right| \approx \text{д}_x + \text{д}_y, \quad \left|\Delta_a\right| \approx \frac{1}{\left|\tilde{y}\right|^2}\left(\Delta_x\left|\tilde{y}\right| + \Delta_y\left|\tilde{x}\right|\right).$$

Для того, чтобы получить оценку погрешности приближенного значения *произвольной функции*, можно поступить следующим образом. Предположим, что $f(x_1, x_2, \ldots, x_n)$ – непрерывно дифференцируемая функция. Погрешность

$$\varepsilon = f\left(\tilde{x}_1 + \varepsilon_1, \ldots, \tilde{x}_n + \varepsilon_n\right) - f\left(\tilde{x}_1, \ldots, \tilde{x}_n\right)$$

можно вычислить по формуле Лагранжа

$$\text{e} = \sum_{i=1}^{n} \text{e}_i \frac{\partial}{\partial x_i} f\left(\tilde{x}_1 + \text{и}_1 \text{e}_1, \ldots, \tilde{x}_n + \text{и}_n \text{e}_n\right), 0 \le \text{и}_i \le 1.$$

Отсюда получается оценка для границы погрешности

$$\left|\text{e}\right| \le \Delta \le \sum_{i=1}^{n} B_i \Delta_i, \text{ где } \left|\text{e}_i\right| \le \Delta_i,\ B_i = \max_{\text{и}_i}\left|\frac{\partial}{\partial x_i} f\left(\tilde{x}_1 + \text{и}_1 \text{e}_1, \ldots, \tilde{x}_n + \text{и}_n \text{e}_n\right)\right|.$$

Когда погрешности ε_i достаточно малы, а частные производные – плавно изменяющиеся функции, числа B_i допустимо заменить абсолютными значениями производных в точке $\left(\tilde{x}_1, \ldots, \tilde{x}_n\right)$. После этого получится приближенное, более простое неравенство, оценивающее Δ:

$$\Delta \approx \sum_{i=1}^{n} \Delta_i \left|\frac{\partial}{\partial x_i} f\left(\tilde{x}_1, \ldots, \tilde{x}_n\right)\right|.$$

С этим правилом согласуются и все полученные ранее непосредственным путем формулы вычисления погрешностей.

Полученные оценки могут быть, вообще говоря, включены в стандартные программы для выполнения всех арифметических действий. В этом состоит суть так называемого *интервального анализа*, когда в качестве решения выдается не отдельное значение, а верхняя и нижняя границы решения. Но обычно этого не делают на основании следующих соображений: во-первых,

подобные вычисления влекут за собой усложнение программ, замедляют работу компьютера и отнимают дополнительное место в памяти. Во-вторых, при достаточно большом числе операций эти оценки, рассчитанные на учет самых неблагоприятных случаев, дадут сильно завышенную оценку погрешности. В практике вычислений часто используют следующие приемы: 1) решение задачи другим методом или повторное применение того же метода, но с иной последовательностью операций; 2) малое изменение входных данных и повторное решение задачи с измененными данными. Можно надеяться на то, что совпадающие в результатах двух разных решений задачи значения будут верными.

1.8. Образцы выполнения контрольной работы 2

Задание 1.

Вычислить выражение $X = \dfrac{m^2 n^3}{\sqrt{k}}$, где m = 28.3 (± 0.02), $n = 7.45(\pm 0.01)$, $k = 0.678(\pm 0.003)$ и определить абсолютную погрешность результата.

Решение. Находим $m^2 = 800.9$; $n^3 = 413.5$; $\sqrt{k} = 0.8234$;

$$X = \frac{800.9 \cdot 413.5}{0.8234} = 402200 = 4.02 \cdot 10^5.$$

Далее, что в данном случае выгоднее, переходим к относительным погрешностям. Имеем $\delta_m = 0.02/28.3 = 0.00071$;

$\delta_n = 0.01/7.45 = 0.00135$; $\delta_k = 0.003/0.678 = 0.00443$,

откуда

$$\delta_X = 2\ \delta_m + 3\ \delta_n + 0.5\ \delta_k =$$

$$= 0.00142 + 0.00405 + 0.00222 = 0.00769 = 0.77\%.$$

Теперь $\Delta_X = 4.02 \cdot 10^5 \cdot 0.0077 = 3.1 \cdot 10^3 < 0.5 \cdot 10^4$.

Ответ: $X = 4.02 \cdot 10^5$ $(\pm 3.1 \cdot 10^3) = (40.2 \pm 0.31)10^4$.

Если требуется оставить только верные знаки, то допустимо записать $(40 \pm 0.51)\cdot 10^4$ (в широком смысле) или продолжить вычисления, чтобы получить ответ, верный в узком смысле.

Задание 2.

Вычислить выражение

$$N = \frac{(n-1)(m+n)}{(m-n)^2}, \text{ где } n = 3.0567\,(\pm 0.0004),\ m = 5.72\,(\pm 0.02),$$

и определить абсолютную погрешность результата.

Решение. Имеем $n - 1 = 2.0567\ (\pm 0.0004)$;

$$m + n = 3.057\ (\pm 0.0004) + 5.72\ (\pm 0.02) = 8.777\ (\pm 0.0204);$$

$$m - n = 5.72\ (\pm 0.02) - 3.057\ (\pm 0.0004) = 2.663\ (\pm 0.0204);$$

$$N = \frac{2.0567\cdot 8.777}{2.663^2} = \frac{2.0567\cdot 8.777}{7.092} = 2.5454 \approx 2.55\,;$$

$$\delta_N = \frac{0.0004}{2.0567} + \frac{0.0204}{8.777} + 2\frac{0.0204}{2} =$$

$$= 0.000196 + 0.00233 + 2\cdot 0.00766 =$$

$$= 0.00253 + 0.01532 = 0.01785 = 1.79\%;$$

$$\Delta_N = 2.55\cdot 0.0179 = 0.046 < 0.5\cdot 10^{-1}.$$

Ответ: $N \approx 2.55\ (\pm 0.046)$ или 2.6 ± 0.096 (в широком смысле).

1.9. Образец выполнения контрольной работы 3

Задание.

Вычислить выражение

$$V = \pi h^2\left(R - \frac{h}{3}\right), \text{ где } h = 11.8,\ R = 23.67,$$

пользуясь правилами подсчета цифр. Определить абсолютную погрешность результата, предполагая, что все цифры данных чисел являются верными значащими цифрами.

Решение. Находим

$$V = 3.142\cdot 11.8^2\cdot(23.67 - 3.933) = 3.142\cdot 11.8^2\cdot 19.737 =$$

$$= 3.142\cdot 139.2\cdot 19.737 = 437.37\cdot 19.737 = 8632.37 \approx 8.63\cdot 10^3.$$

$$\delta_V = \frac{0.0005}{3.142} + 2\frac{0.05}{11.8} + \frac{0.005 + 0.05/3}{19.737} =$$

$$= 0.00016 + 0.008475 + 0.001098 = 0.009733 = 0.98\%;$$

$$\Delta_V = 8.63\cdot 10^3\cdot 0.0098 = 84.6 < 0.5\cdot 10^3.$$

Ответ: $V = 8.63\cdot 10^3$ или, оставляя только верные знаки в узком смысле, $(9 \pm 0.46)\cdot 10^3$.

2. ВОПРОСЫ К САМОПРОВЕРКЕ

1. Что понимается под погрешностью?
2. Какие источники погрешностей Вы знаете?
3. Какую погрешность относят к неустранимой погрешности? Можно ли ее уменьшить?
4. С чем связана погрешность численного метода? Как ее уменьшить?
5. Из-за чего возникают ошибки округления?
6. Назовите значащие цифры числа.
7. В чем состоит процесс усечения числа?
8. Определить погрешность округления, возникающую в результате процесса усечения.
9. Пояснить правило округления по дополнению.
10. Определить погрешность округления, возникающую в результате процесса округления по дополнению.
11. Возникает ли погрешность при выполнении арифметических операций с целыми числами?
12. Дайте определение приближенного значения числа по недостатку и по избытку.