ЛЕКЦИЯ 1.1 ЭЛЕМЕНТАРНАЯ ТЕОРИЯ ПОГРЕШНОСТЕЙ

1. Источники и классификация погрешностей

Практически все числовые величины, с которыми оперируют при практических расчётах, являются приближёнными, следовательно, содержащими погрешности — отличия от истинных значений. Естественно, что тогда и численное решение любой задачи заключает в себе некоторые погрешности. В основном они обусловлены следующими причинами:

1. Приближённость математических моделей объектов, систем, процессов (любая, даже очень сложная модель является лишь приближением реальности, поэтому в ней самой её природой заложена погрешность);
2. Погрешности исходных данных;
3. Приближённость методов численного решения (очень немногие классы задач имеют точные методы решения, в основном же численные методы являются приближёнными, т.е. дают погрешности даже при гипотетическом отсутствии других их источников);
4. Неизбежные потери точности при машинном представлении чисел и арифметических операциях над ними в компьютере.

Погрешности результатов вычислений можно разделить на *неустранимые* (обусловленные причинами 1 и 2) и *устранимые* (причины 3, 4). Следует отметить, что устранение не означает возможность полного обнуления погрешностей, которые всегда будут присутствовать в какой-то степени. Их можно *уменьшить*, в чём, собственно, и заключается одна из важнейших задач вычислительной математики. Погрешности, вызванные причиной 4, устраняются (т.е. уменьшаются) программными и аппаратными средствами, т.е. применением более совершенной вычислительной техники и рациональной организацией счёта. С другой стороны, неустранимость погрешностей не означает, что их нельзя уменьшить. Это возможно, но средствами, находящимися вне методов вычислений, алгоритмизации и программирования, т.е. *внешними* по отношению к поставленной вычислительной задаче и методам её решения. Погрешности, вызванные несовершенством математиче-

ской модели, снижаются выбором более адекватной модели, вызванные неточностями исходных данных — применением более совершенных методов измерений и инструментов. Но это уже компетенции других наук.

2. Абсолютная и относительная погрешности

Точные значения числовых скалярных величин будем обозначать латинскими буквами (a, b и т.д.). Как правило, они неизвестны. Вместо них при расчётах используются вычисленные или измеренные приближённые значения, которые будем помечать «звёздочкой» (a^*, b^* и т.д.). Теперь определим числовые величины, характеризующие степень близости приближённых и точных значений. Будем рассматривать только скалярные величины.

Определение. *Погрешностью* приближённого значения a^* называется величина

$$\varepsilon a^* = a - a^*,$$

т.е. разность между точным и приближённым числами.

Погрешность неудобна для операций и для оценки, поскольку может быть как положительной, так и отрицательной. Поэтому определим другую, всегда положительную характеристику.

Определение. *Абсолютной погрешностью* a^* называется

$$\Delta a^* = |\varepsilon a^*| = |a - a^*|.$$

Вычислить абсолютную погрешность невозможно, поскольку неизвестно точное значение. По этой причине в приближённых вычислениях всегда оперируют с *верхними оценками* абсолютных погрешностей, которые будем обозначать $\bar{\Delta}$, то есть $\bar{\Delta}a^*$ — такое число, про которое заведомо известно, что $\Delta a^* \leq \bar{\Delta}a^*$. Верхние оценки погрешностей или предельные погрешности вполне поддаются вычислениям.

Верхняя оценка абсолютной погрешности даёт величину отклонения приближённого значения от неизвестного точного числа, т.е. если $\bar{\Delta}a^*$ — верхняя оценка абсолютной погрешности a, в таком случае применяется запись

$$a = a^* \pm \bar{\Delta}a^*,$$

что означает принадлежность величины a отрезку $[a^* - \bar{\Delta}\, a^*, a^* + \bar{\Delta}\, a^*]$.

По одной лишь абсолютной погрешности a^* трудно судить о точности приближён-

ного значения, так как она зависит от величин a, a^*, и абсолютная погрешность имеет ту же размерность, что и оцениваемая. Например, при измерении длины спортзала и расстояния от дома до метро получены значения абсолютных погрешностей, равные 1 метру. Насколько хорошо сделаны приближения? Понятно, что приближение расстояния до метро лучше, чем приближенная длина спортзала, но как можно об этом судить только по величине 1 метр? Поэтому требуется безразмерная, не зависящая от масштаба, характеристика. Таковой является *относительная погрешность* приближённого значения a^*.

Определение. *Относительной погрешностью* приближённого значения a^* называется величина

$$\delta a^* = \frac{|a - a^*|}{|a|} = \frac{\Delta a^*}{|a|}.$$

Относительную погрешность также невозможно вычислить, поэтому при расчётах используют её *верхнюю оценку* $\bar{\delta} a^*$. Обычно относительную погрешность обычно дают в процентах. В приведенном выше примере относительная погрешность длины спортзала может составлять не менее 1-2 процентов, а погрешность расстояния до метро, скорее всего, не более 0,2-0,4 процента.

3. Значащие цифры в десятичной записи числа

Понятно, что приближённое число не может быть бесконечной дробью, поэтому все приближённые числа записываются в виде *конечных* десятичных дробей, возможно, с порядком, т.е. в виде

$$\underbrace{\alpha_n \alpha_{n-1} \dots \alpha_1 \alpha_0, \beta_{-1} \dots \beta_{-m}}_{M} \cdot 10^p \qquad (1)$$

где α_i, β_{-j} — десятичные цифры, $M = \alpha_n \alpha_{n-1} \dots \alpha_1 \alpha_0, \beta_{-1} \dots \beta_{-m}$— *мантисса*. Её значение вычисляется по формуле представления числа в десятичной позиционной системе счисления:

$$M = \alpha_n \cdot 10^n + \alpha_{n-1} \cdot 10^{n-1} + \dots + \alpha_1 \cdot 10 + \alpha_0 + \beta_{-1} \cdot 10^{-1} + \dots + \beta_{-m} \cdot 10^{-m}.$$

Целое число p — *порядок* приближённой величины. В такой записи различают *значащие* и *незначащие цифры*.

Определение. *Значащими* называются все цифры мантиссы в представлении (1), начиная с первой ненулевой слева. Остальные цифры (нули) называются *незначащими***.**

Примеры. В следующих приближённых числах значащие цифры подчёркнуты:

$$a^* = 0{,}00\underline{231},\ b^* = 000\underline{140},\ c^* = -0{,}00\underline{840} \cdot 10^6,\ d^* = 0{,}0\underline{610}.$$

Нули слева можно убрать без потери значений, при необходимости изменив порядок, поэтому они и называются незначащими. В этих примерах числа можно записать без незначащих нулей так:

$$a^* = 2{,}31 \cdot 10^{-3},\ b^* = 140,\ c^* = -8{,}40 \cdot 10^3,\ d^* = 6{,}10 \cdot 10^{-2}.$$

Замечание 1. Значащие нули справа убирать нельзя, поскольку они означают разряды числа, например, $0{,}56$ и $0{,}560$ — разные приближённые числа, так как первое дано с двумя знаками после запятой, а второе — с тремя.

Замечание 2. Особый случай представляет собой нулевая величина. Как бы ни записывали нуль, в её записи единственной значащей цифрой считают последний нуль справа: $0{,}000\underline{0}$, $0,\underline{0} \cdot 10^{-2}$, $00\underline{0}$.

4. Округление

При вычислениях для получения приближённых значений некоторые значащие цифры числа отбрасываются. Такая операция называется *округлением*. Есть два правила округления: *усечением* и *по дополнению*. Опишем каждое из них.

При округлении *усечением* отбрасываются все значащие цифры правее той, до которой производится округление. При необходимости изменяется порядок числа.

Примеры. Число $12{,}01297$, округлённое до второго знака после запятой (разряд 10^{-2}), равно $12{,}01$, до третьего (разряд 10^{-3}) — $12{,}012$, число $0{,}5627 \cdot 10^4$, округлённое до второго знака до запятой (разряд 10^1 — десятки), есть $5{,}62 \cdot 10^3$, но не 5620 (т.к. последнее число имеет точность до единиц, а мы округляли до десятков).

Таким образом, округление усечением есть округление в меньшую сторону для положительных чисел и в большую для отрицательных чисел. Очевидно, что абсолютная погрешность при нём – это величина отброшенной части - не превышает единицы разряда, до которого производилось округление.

Более точным, а следовательно, более применимым на практике является округление *по дополнению*, когда также отбрасываются все значащие цифры правее той, до которой производится округление, при необходимости изменяется порядок, при этом руко-

водствуются правилами:

1. Если первая слева отбрасываемая цифра меньше пяти, то оставшиеся не меняются, как при усечении.
2. Если эта цифра больше пяти или она равна пяти и среди остальных отбрасываемых есть ненулевые, то последняя сохраняемая цифра увеличивается на единицу.
3. Если эта цифра равна пяти и все остальные отбрасываемые есть нули, то последняя сохраняемая цифра не меняется, если она чётная, и увеличивается на единицу, если она нечётная.

Примеры. Число $0{,}04252$, округленное до сотых долей, равно $0{,}04$, до тысячных — $0{,}043$. Число 99500, округленное до тысяч, равно $100 \cdot 10^3$.

Очевидно, что округление по дополнению обеспечивает величину абсолютной погрешности, не превосходящей половины единицы разряда, в котором находится последняя оставляемая цифра; округление идёт в ближайшую сторону.

В дальнейшем будем считать округление по дополнению применяемым по умолчанию.

5. Верные значащие цифры

Значащие цифры с помощью абсолютных погрешностей делятся на *верные* и *сомнительные*. Пусть дано представление приближённого числа (1), в котором все цифры значащие.

Определение. *Верными в широком смысле* называются те значащие цифры, для которых абсолютная погрешность не превосходит единицы разряда, в котором они находятся. Значащие цифры, для которых это условие не выполняется, называются *сомнительными в широком смысле*.

Очевидно, что это определение связано с округлением усечением, а именно, при таком округлении все оставшиеся цифры будут верными в широком смысле при условии учёта только погрешности округления.

Пример. Пусть

$$a = 25{,}01257 \pm 0{,}0008,$$

т.е. $a^* = 25{,}01257$, $\Delta a^* = 0{,}0008$. Определим верные в широком смысле цифры. Понятно, что цифры 2 и 5 до запятой будут верными. Проверим цифру 0. Её разряд $10^{-1} > \Delta a^* = 0{,}0008$, поэтому она является верной. Также проверяется верность следующих цифр 1, 2. А вот 5 будет сомнительной, поскольку её разряд $10^{-4} < \bar{\Delta} a^* = 0{,}0008$. Очевидно, что 7 — тоже сомнительная цифра. Итак, в записи a^* все разряды до 10^{-3} включительно верны, остальные сомнительны.

Из определений верной цифры и абсолютной погрешности следует, что верные цифры с точностью до абсолютной погрешности совпадают с соответствующими цифрами точного значения, поэтому только они несут информацию о нём. Сомнительные цифры являются бесполезными, «лишними», в записи числа при данной абсолютной погрешности.

Теперь определим верность и сомнительность в узком смысле.

Определение. Значащая цифра называется *верной в узком смысле*, если абсолютная погрешность не превосходит половины единицы разряда, в котором она находится. В противном случае она называется *сомнительной в узком смысле*.

Здесь также очевидно, что это определение связано с округлением по дополнению, поскольку, при таком округлении все оставшиеся цифры верны в узком смысле при условии учёта только погрешности округления.

Пример. Возьмём значение из предыдущего примера:

$$a = 25{,}01257 \pm 0{,}0008.$$

Определим верные в узком смысле цифры. Понятно, что цифры до запятой будут верными. Проверим цифру 0. Половина её разряда $0{,}5 \cdot 10^{-1} = 0{,}05 > \bar{\Delta} a^* = 0{,}0008$. Далее, цифра 1 также верна: $0{,}5 \cdot 10^{-2} = 0{,}005 > \bar{\Delta} a^*$. А вот следующая цифра 2 сомнительна, так как $0{,}5 \cdot 10^{-3} = 0{,}0005 < \bar{\Delta} a^*$. Получаем, что в записи a^* все разряды до 10^{-2} включительно верны в узком смысле, остальные сомнительны.

Замечание. Очевидно, что верная в узком смысле цифра будет верной и в широком, обратное не всегда имеет место, как показано в примерах.

Поскольку выше было оговорено, что все округления будут производиться по дополнению, будем отныне считать верность и сомнительность понимаемыми по умолчанию в узком смысле.

Определение верных цифр по разрядам делается не так быстро, как хотелось бы, поэтому было выведено простое правило выявления верных цифр без проверки каждой значащей цифры. Оно формулируется следующим образом:

1. Абсолютная погрешность округляется с избытком до одной значащей цифры.
2. Если эта цифра не превосходит пяти, то все разряды левее её разряда считаем верными, если же она больше пяти, то соседний слева разряд будет сомнительным, все остальные левее него — верными.

Примеры. Пусть

$$a = 0{,}5482 \pm 0{,}0045.$$

Округляем абсолютную погрешность с избытком, получаем $\bar{\Delta}a^* = 0{,}005$, единственная значащая цифра $\bar{\Delta}a^*$ равна пяти, поэтому по правилу верными цифрами будут 5 и 4.

Если

$$a = 0{,}5482 \pm 0{,}0077,$$

то округлённая предельная абсолютная погрешность $\bar{\Delta}a^*$ равна $0{,}008$, по правилу цифра 5 верна, 4 сомнительна (левый соседний разряд от разряда 8).

Пусть $b^* = 98321,\ \bar{\Delta}b^* = 6 \cdot 10^2$. Тогда верной цифрой будет только 9, остальные сомнительны.

Выше было отмечено, что только верные цифры дают представление об истинном значении вычисляемой величины, ибо они совпадают с его соответствующими цифрами в пределах данной абсолютной погрешности.

Замечание. Совпадение в пределах погрешности верных цифр с точными значениями не обязательно означает буквальное совпадение. Например, пусть $a = 0{,}9999,\ a^* = 1{,}000$, тогда $\bar{\Delta}a^* = 0{,}0001$. По правилу все цифры a^* верны, но, как видно, ни одна из них не совпадает с цифрами a.

Поэтому вычисленные значения обязательно округляют до верных цифр, избавляясь от ненужных сомнительных. Если при этом применяется округление усечением, то верные цифры понимаются в широком смысле, если округление по дополнению, то в узком смысле. Но надо учитывать, что к исходной погрешности, по которой определялись верные цифры, добавляется ещё погрешность округления, что может привести к тому, что последняя верная цифра может стать сомнительной. В этом случае надо продолжить

округление, двигаясь справа налево по разрядам, пока все цифры не будут верными с учётом полной погрешности.

Примеры. Пусть

$$a = 4{,}01350 \pm 0{,}00068.$$

Применяем округление усечением. Непосредственной проверкой убеждаемся, что все разряды до 10^{-3} включительно верны, остальные сомнительны. Округляя число до верных цифр, получаем $a_1^* = 4{,}013$. Полная погрешность равна

$$\Delta a_1^* = \overline{\Delta} a^* + \Delta_{\text{окр.}} = 0{,}00068 + |4{,}01350 - 4{,}013| = 0{,}00118 > 10^{-3} = 0{,}001.$$

Как видно, добавленная погрешность округления привела к тому, что цифра 3 стала сомнительной. Округляя дальше до 10^{-2}, получаем $a_2^* = 4{,}01$, общая погрешность:

$$\Delta a_2^* = \overline{\Delta} a^* + \Delta_{\text{окр.}} = 0{,}00068 + |4{,}01350 - 4{,}01| = 0{,}00418 \leq$$

$$\leq 10^{-2} = 0{,}01.$$

Все цифры верные. Итак, значение $a^* = 4{,}01350$, округлённое до верных цифр, равно $4{,}01$.

Пусть теперь

$$b = 0{,}9950 \cdot 10^5 \pm 4{,}8 \cdot 10^2.$$

Применяем округление по дополнению. Согласно правилу верными будут разряды до 10^3 включительно, округляя b^* до верных цифр, получаем $b_1^* = 1{,}00 \cdot 10^5$. Полная погрешность:

$$\Delta b_1^* = 4{,}8 \cdot 10^2 + |0{,}9950 \cdot 10^5 - 1{,}00 \cdot 10^5| = 9{,}8 \cdot 10^2 > 0{,}5 \cdot 10^3,$$

разряд 10^3 стал сомнительным в узком смысле. Округляя дальше, получаем $b_2^* = 1{,}0 \cdot 10^5$, итоговая погрешность

$$\Delta b_2^* = 9{,}8 \cdot 10^2 \leq 0{,}5 \cdot 10^4,$$

все цифры в $b_2^* = 1{,}0 \cdot 10^5$ верные.

Для решения обратной задачи оценки абсолютной погрешности числа по известным и верным в узком смысле цифрам применяется следующее очевидно вытекающее из определений правило: за предельную абсолютную погрешность приближённой величины принимается половина единицы разряда, в котором находится последняя слева верная значащая цифра.

Примеры. Если $a^* = 5{,}985$, причём все цифры верные, то $\overline{\Delta} a^* = 0{,}5 \cdot 10^{-3} = 0{,}0005$ -

последний верный разряд — 10^{-3}. Если в $b^* = 5{,}210 \cdot 10^4$ все цифры верные, то $\overline{\Delta} b^* = 0{,}5 \cdot 10^1 = 5$.

6. Правила записи приближённых чисел

Из изложенного выше понятно, что только верные цифры имеют ценность как несущие информацию о вычисляемых величинах. Точность ответа определяется не количеством значащих цифр, а *количеством верных цифр*, именно их надо оставлять в промежуточных и окончательных результатах. Однако округлять только до верных цифр всё-таки нецелесообразно по следующим причинам:

1. Оценки погрешностей, по которым определяются верные цифры, завышены, некоторые сомнительные цифры могут оказаться верными и потому можно потерять верные цифры из-за завышенных погрешностей.
2. Погрешности округлений могут привести к тому, что последняя верная цифра станет сомнительной, это значит, что желательно иметь хотя бы одну сомнительную цифру в запасе.

Поэтому при приближённых вычислениях руководствуются следующими правилами:

1. В промежуточных результатах оставляют, кроме верных, одну-две сомнительные цифры.
2. Окончательный результат округляют с сохранением не более одной сомнительной цифры.

Для таких округлений необходимо уметь оценивать погрешности при арифметических операциях и вычислениях функций. Этим вопросам посвящены следующие пункты.

7. Погрешности представления чисел

Представление числовой информации в компьютере, как правило, влечет за собой появление погрешностей, величина которых зависит от формы представления числа и от длины разрядной сетки компьютера. Для представления числа в компьютере также определяются абсолютная и относительная погрешности.

Определение. *Абсолютная погрешность ΔA^* представления числа A* – это модуль разности между истинным значением величины A и ее машинным представлением A^*, т.е.

$$\Delta A^* = |A - A^*|.$$

Определение. *Относительная погрешность δA^* представления числа A* – это отношение абсолютной погрешности к модулю машинного представления числа A^*, т.е.

$$\delta A^* = \frac{\Delta A^*}{|A^*|}.$$

Основным источником погрешности является ограниченная разрядная сетка. При вычислениях часто возникает ситуация, когда полученный результат, а именно, его дробная часть, имеет больше разрядов, чем имеется в разрядной сетке. В этом случае приходится округлять результат до нужного числа разрядов. Так, если разрядная сетка имеет длину n разрядов, то максимальное значение абсолютной погрешности будет равно 2^{-n}, а минимальное значение равняется 0. Часто при оценке итоговой погрешности используют усредненную абсолютную погрешность

$$\Delta_{\text{ср}} = \frac{0 + 2^{-n}}{2} = 0{,}5 \cdot 2^{-n}.$$

Абсолютное значение представления дробного числа в форме с фиксированной запятой находится в диапазоне от 2^{-n} до $1 - 2^{-n}$. Относительная погрешность представления для максимального значения числа равняется

$$\delta A^*_{\text{ф макс}} = \frac{\Delta_{\text{ср}}}{|A^*_{\text{ф макс}}|} = \frac{0{,}5 \cdot 2^{-n}}{1 - 2^{-n}},$$

где $A^*_{\text{ф макс}}$ – максимальное значение представления дробного числа с фиксированной запятой. Обычно длина разрядной сетки $n = 16 \div 64$, тогда $2^{-n} \ll 1$ и $\delta A^*_{\text{ф макс}} \approx 0{,}5 \cdot 2^{-n}$.

Относительная погрешность представления для минимального значения числа равняется

$$\delta A^*_{\text{ф мин}} = \frac{\Delta_{\text{ср}}}{|A^*_{\text{ф мин}}|} = 0{,}5 \cdot \frac{2^{-n}}{2^{-n}} = 0{,}5,$$

где $A^*_{\text{ф мин}}$ – минимальное значение представления дробного числа с фиксированной запятой.

Видно, что погрешности представления малых чисел в форме с фиксированной запятой могут быть очень значительными, то есть соизмеримыми с самими числами.

Для числа в форме с плавающей запятой принята так называемая *нормализованная форма записи*: мантисса M десятичного числа в (1) находится в диапазоне от 0 включительно до 1 исключительно, причём первый разряд после запятой всегда ненулевой. Абсолютное значение нормализованной двоичной мантиссы при использовании n разрядов находится в диапазоне от 2^{-1} (первый бит после запятой всегда единичный) до $1 - 2^{-n}$. Для нахождения относительной погрешности представления числа в форме с плавающей запятой необходимо погрешность мантиссы умножить на величину порядка числа p_A:

$$\delta A^*_{\text{пл мин}} = 0{,}5 \cdot \frac{2^{-n} p_A}{2^{-1} p_A} = 2^{-n},$$

$$\delta A^*_{\text{пл макс}} = 0{,}5 \cdot \frac{2^{-n} p_A}{(1 - 2^{-n}) p_A} \approx 0{,}5 \cdot 2^{-n}.$$

Относительная погрешность представления чисел в форме с плавающей запятой почти не зависит от величины числа. Поэтому все математические вычисления над дробными числами проводят, когда эти числа представлены в форме с плавающей запятой. Вычисления над целыми числами можно проводить, когда эти числа представлены в форме с фиксированной запятой. Но следует контролировать, чтобы полученный результат не превысил пороговые значения, определяемые длиной разрядной сетки.

8. Погрешности арифметических операций

Сформулируем для четырёх арифметических операций — сложения, вычитания, умножения, деления — утверждения, которые позволят оценивать их погрешности.

Начнём с оценок погрешностей суммы и разности.

Пример. Рассмотрим две заданные величины $a = 3{,}26$ и $b = 2{,}67$. Предположим, что они приближенно представляются числами $a^* = 3{,}24$ и $b^* = 2{,}63$. Тогда

$$\Delta a^* = |3{,}26 - 3{,}24| = 0{,}02,\ \Delta b^* = |2{,}67 - 2{,}63| = 0{,}04.$$

Чему равны абсолютные погрешности сложения и вычитания этих чисел? Находим точную и приближённую сумму:

$$a + b = 3{,}26 + 2{,}67 = 5{,}93;\ a^* + b^* = 3{,}24 + 2{,}63 = 5{,}87.$$

Абсолютная погрешность сложения равна $\Delta(a + b)^* = 5{,}93 - 5{,}87 = 0{,}06$. Погрешность не превосходит суммы абсолютных погрешностей:

$$\Delta(a + b)^* = 0{,}06 \leq 0{,}02 + 0{,}04 = 0{,}06.$$

Находим погрешность разности:

$$a - b = 3{,}26 - 2{,}67 = 0{,}59;$$

$$a^* - b^* = 3{,}24 - 2{,}63 = 0{,}61;$$

$$\Delta(a - b)^* = |0{,}59 - 0{,}61| = 0{,}02.$$

Погрешность разности также не превосходит суммы абсолютных погрешностей:

$$\Delta(a - b)^* = 0{,}02 \le 0{,}02 + 0{,}04 = 0{,}06.$$

Теорема 1. Для абсолютных погрешностей суммы и разности имеет место следующая оценка:

$$\Delta(a \pm b)^* \le \Delta a^* + \Delta b^*.$$

Доказательство. По определению абсолютной погрешности имеем:

$$\Delta(a \pm b)^* = |(a \pm b) - (a \pm b)^*| = |(a \pm b) - (a^* \pm b^*)| =$$

$$= |(a - a^*) \pm (b - b^*)|.$$

Далее, применяя известное свойство модуля $|x \pm y| \le |x| + |y|$, получаем

$$\Delta(a \pm b)^* = |(a - a^*) \pm (b - b^*)| \le |a - a^*| + |b - b^*| =$$

$$= \Delta a^* + \Delta b^*,$$

что и доказывает теорему. ■

Следствие. Методом математической индукции можно доказать обобщение теоремы на произвольное число величин:

$$\Delta(a_1 \pm a_2 \pm \cdots \pm a_n)^* \le \Delta a_1^* + \Delta a_2^* + \cdots + \Delta a_n^*.$$

Пример. Рассмотрим величины из предыдущего примера $a = 3{,}26$ и $b = 2{,}67$, которые заданы приближенно: $a^* = 3{,}24$ и $b^* = 2{,}63$. Тогда абсолютные и относительные погрешности представления величин следующие:

$$\Delta a^* = 0{,}02, \Delta b^* = 0{,}04, \delta a^* = \frac{\Delta a^*}{a} = 0{,}006, \delta b^* = \frac{\Delta b^*}{b} = 0{,}015.$$

Вычислим относительные погрешности сложения и вычитания:

$$\delta(a + b)^* = \frac{\Delta(a + b)^*}{|\,a + b|} = \frac{0{,}06}{5{,}93} = 0{,}01, \delta(a - b)^* = \frac{\Delta(a - b)^*}{|\,a - b|} = \frac{0{,}02}{0{,}59} = 0{,}03.$$

Относительная погрешность сложения меньше максимальной из относительных погрешностей слагаемых, относительная погрешность разности больше относительных погрешностей исходных величин.

Теорема 2. Если a, b— ненулевые числа одного знака, то:

1. Для относительной погрешности суммы имеет место оценка

$$\delta(a+b)^* \le M,$$

где $M = \max\{\delta a^*;\ \delta b^*\}$.

2. Для относительной погрешности разности справедлива оценка

$$\delta(a-b)^* \le M \cdot \vartheta,$$

где

$$\vartheta = \left|\frac{a+b}{a-b}\right|.$$

Доказательство. По определению относительной погрешности и теореме 1 имеем:

$$\delta(a \pm b)^* = \frac{\Delta(a \pm b)^*}{|\,a \pm b|} \le \frac{\Delta a^* + \Delta b^*}{|a \pm b|} = \frac{\delta a^* \cdot |a| + \delta b^* \cdot |b|}{|a \pm b|}.$$

Теперь заменим δa^*, δb^* на M, т.е. максимальную из них величину. Получим

$$\delta(a \pm b)^* = \frac{\delta a^* \cdot |a| + \delta b^* \cdot |b|}{|a \pm b|} \le \frac{M \cdot |a| + M \cdot |b|}{|a \pm b|} = M\frac{|a| + |b|}{|a \pm b|}.$$

Поскольку по условию числа a и b одного знака, то для них выполняется равенство $|a| + |b| = |a + b|$ (докажите его самостоятельно). Итак, имеем

$$\delta(a \pm b)^* \le M\frac{|a+b|}{|a \pm b|}.$$

Если теперь выписать отдельно неравенства для суммы и разности, то получатся доказываемые оценки. ■

Следствие. При вычитании близких чисел погрешность результата может быть очень большой ($\vartheta \to \infty$ при $a - b \to 0$). Происходит катастрофическая потеря точности.

Замечание. Значение ϑ точно вычислить невозможно (a и b неизвестны), поэтому вместо него при оценках используют приближённое значение

$$\vartheta^* = \left|\frac{a^* + b^*}{a^* - b^*}\right|.$$

Теорема 3. Для относительной погрешности произведения имеет место оценка

$$\delta(ab)^* \le \delta a^* + \delta b^* + \delta a^* \cdot \delta b^*.$$

Доказательство. Начиная с определения относительной погрешности, проделываем простые преобразования:

$$\delta(ab)^* = \frac{\Delta(ab)^*}{|ab|} = \frac{|ab - a^*b^*|}{|ab|} = \frac{|a(b-b^*) + b(a-a^*) - (a-a^*)(b-b^*)|}{|ab|}.$$

Применяя свойство модуля, заключающееся в том, что модуль суммы и разности не превосходит суммы модулей, получаем с помощью несложных преобразований

$$\delta(ab)^* = \frac{|a(b-b^*)+b(a-a^*)-(a-a^*)(b-b^*)|}{|ab|} \le$$
$$\le \frac{|a(b-b^*)|+|b(a-a^*)|+|(a-a^*)(b-b^*)|}{|ab|} =$$
$$= \frac{|a|\cdot|b-b^*|+|b|\cdot|a-a^*|+|a-a^*|\cdot|b-b^*|}{|a|\cdot|b|} =$$
$$= \frac{|b-\ b^*|}{|b|} + \frac{|a-a^*|}{|a|} + \frac{|a-a^*|}{|a|}\cdot\frac{|b-b^*|}{|b|} =$$
$$= \delta a^* + \delta b^* + \delta a^* \cdot \delta b^*$$

(здесь было применено известное свойство модуля $|xy| = |x|\cdot|y|$). ■

Следствие. Если хотя бы одна из погрешностей δa^* или δb^* близка к нулю, то можно использовать приближённую оценку

$$\delta(ab)^* \le \delta a^* + \delta b^*,$$

которая обобщается на любое конечное число сомножителей:

$$\delta(a_1 a_2 \cdots a_n)^* \le \delta a_1^* + \delta a_2^* + \cdots + \delta a_n^*.$$

Как правило, для оценки погрешности произведения применяют это следствие.

Теорема 4. Если b^* — ненулевое число и $\delta b^* < 1$, то для относительной погрешности частного имеет место оценка

$$\delta\left(\frac{a}{b}\right)^* \le \frac{\delta a^* + \delta b^*}{1-\delta b^*}.$$

Доказательство. Снова начинаем преобразования с определения относительной погрешности и применяем известные свойства модуля:

$$\delta\left(\frac{a}{b}\right)^* = \frac{\left|\frac{a}{b}-\frac{a^*}{b^*}\right|}{\left|\frac{a}{b}\right|} = \frac{|ab^*-a^*b|}{|ab^*|} = \frac{|a(b^*-b)+b(a-a^*)|}{|ab^*|} \le$$
$$\le \frac{|a(b^*-b)|+|b(a-a^*)|}{|ab^*|} = \frac{|a|\cdot|b-b^*|+|b|\cdot|a-a^*|}{|a|\cdot|b^*|} =$$
$$= \frac{|a|\Delta(b^*)+|b|\Delta(a^*)}{|a|\cdot|b^*|}. \qquad (2)$$

Теперь для оценки $|b^*|$ воспользуемся свойством модуля $|a-b| \ge |a|-|b|$ (докажите его самостоятельно):

$$|b^*| = |b-(b-b^*)| \ge |b| - |b-b^*| = |b| - \Delta b^* =$$
$$= |b|\left(1-\frac{\Delta b^*}{|b|}\right) = |b|(1-\delta b^*).$$

Если теперь заменить в знаменателе (2) $|b^*|$ полученной оценкой, то дробь тем самым увеличится, поэтому далее получаем

$$\delta\left(\frac{a}{b}\right)^* \le \frac{|a|\Delta b^* + |b|\Delta a^*}{|a|\cdot|b^*|} \le \frac{|a|\Delta b^* + |b|\Delta a^*}{|a|\cdot|b|(1-\delta b^*)} =$$

$$= \frac{\frac{\Delta b^*}{|b|} + \frac{\Delta a^*}{|a|}}{1-\delta b^*} = \frac{\delta a^* + \delta b^*}{1-\delta b^*}.$$

Теорема доказана. ■

Следствие. Если погрешность δb^* близка к нулю, то можно использовать приближённую оценку

$$\delta\left(\frac{a}{b}\right)^* \le \delta a^* + \delta b^*.$$

Для оценки погрешности частного также чаще применяется следствие.

Пример. Найти значение выражения

$$s = \frac{a^2 + ab + b^2}{2(a+b)}$$

с соблюдением правил приближённых вычислений, если $a^* = 2{,}31$, $b^* = 3{,}102$. Оценить абсолютную и относительную погрешности результата. Все цифры в a^*, b^* верные.

По правилу определяем предельные абсолютные погрешности a^* и b^*:

$$\overline{\Delta}\, a^* = 0{,}5\cdot 10^{-2} = 0{,}005,\ \overline{\Delta}\, b^* = 0{,}5\cdot 10^{-3} = 0{,}0005.$$

Вычисляем их относительные погрешности:

$$\overline{\delta} a^* = \frac{\overline{\Delta}\, a^*}{|a|} \approx \frac{0{,}005}{2{,}31} = 0{,}0022,\ \overline{\delta} b^* = \frac{\overline{\Delta}\, b^*}{|b|} \approx \frac{0{,}0005}{3{,}102} = 0{,}00016.$$

Пусть $c = a^2$, тогда $c^* = 2{,}31^2 = 5{,}3361$. Относительную погрешность c оцениваем по следствию из теоремы 3: $\overline{\delta} c^* = \overline{\delta}(a^*)^2 = \overline{\delta} a^* + \overline{\delta} a^* = 2\overline{\delta}(a^*) = 2\cdot 0{,}0022 = 0{,}0044$.

Абсолютная погрешность равна $\overline{\Delta} c^* = \overline{\delta} c^* |c| \approx 0{,}0044\cdot 5{,}3361 = 0{,}023 \le 0{,}03$. Следовательно, верными в c^* будут разряды до 10^{-1} включительно; округляя по дополнению, получаем $c^* = 5{,}34$ с одной сомнительной запасной цифрой. Далее во всех промежуточных результатах будем оставлять одну сомнительную цифру. Проверять, не стала ли последняя верная цифра сомнительной из-за добавленной погрешности округления, не будем, так как оставлена одна дополнительная сомнительная цифра.

Пусть теперь $d = b^2$, $f = ab$. Аналогично оцениваем их погрешности и проводим округление:

$$d^* = 3{,}102^2 = 9{,}622404,\ \overline{\delta}d^* = 2\overline{\delta}b^* = 2 \cdot 0{,}00016 = 0{,}00032,$$

$$\overline{\Delta}d^* \approx 0{,}00032 \cdot 9{,}622404 = 0{,}0031 \le 0{,}004 \ \Rightarrow\ d^* = 9{,}622;$$

$$f^* = 2{,}31 \cdot 3{,}102 = 7{,}16562,\ \overline{\delta}f^* = \overline{\delta}a^* + \overline{\delta}b^* = 0{,}0022 + 0{,}00016 = 0{,}00236,$$

$$\overline{\Delta}f^* \approx 0{,}00236 \cdot 7{,}16562 = 0{,}017 \le 0{,}02 \ \Rightarrow\ f^* = 7{,}17.$$

Обозначим числитель дроби $k = a^2 + ab + b^2 = c + f + d$. Тогда

$$k^* = c^* + f^* + d^* = 5{,}34 + 7{,}17 + 9{,}622 = 22{,}132.$$

Абсолютную погрешность k^* оценим по следствию из теоремы 1:

$$\overline{\Delta}k = \overline{\Delta}c^* + \overline{\Delta}f^* + \overline{\Delta}d^* = 0{,}023 + 0{,}017 + 0{,}0031 = 0{,}0431 \le\ 0{,}05 \ \Rightarrow\ k^* = 22{,}13;$$

относительная погрешность k^* равна

$$\overline{\delta}k^* = \frac{\overline{\Delta}k^*}{|k|} \approx \frac{0{,}0431}{22{,}132} = 0{,}00195.$$

Вычисляем знаменатель $t = 2(a + b)$: $t^* = 2(2{,}31 + 3{,}102) = 10{,}824$. Оцениваем абсолютную погрешность:

$$\overline{\Delta}t^* = \overline{\Delta}(2a^* + 2b^*) = 2\left(\overline{\Delta}a^* + \overline{\Delta}b^*\right) = 2(0{,}005 + 0{,}0005) =$$

$$= 2 \cdot 0{,}0055 = 0{,}011 \le 0{,}02 \ \Rightarrow\ t^* = 10{,}82;$$

относительная погрешность знаменателя:

$$\overline{\delta}t^* \approx \frac{0{,}011}{10{,}824} = 0{,}00102.$$

Наконец, вычисляем результат:

$$s = \frac{a^2 + ab + b^2}{2(a + b)} = \frac{k}{t}\ ,\ s^* = \frac{k^*}{t^*} = \frac{22{,}13}{10{,}82} = 2{,}045289.$$

Относительная погрешность s^* оценивается по следствию из теоремы 4:

$$\overline{\delta}s^* = \overline{\delta}k^* + \overline{\delta}t^* = 0{,}00195 + 0{,}00102 = 0{,}00297 \approx 0{,}3\%.$$

Вычисляем абсолютную погрешность и округляем:

$$\overline{\Delta}s^* \approx \overline{\delta}s^* \cdot s^* = 0{,}00297 \cdot 2{,}045289 = 0{,}0061 \le 0{,}007 \ \Rightarrow\ \ s^* = 2{,}05.$$

Ответ: $s^* = 2{,}05$ (с одной запасной сомнительной цифрой), $\overline{\delta}s^* = 0{,}3\%$, $\overline{\Delta}s^* = 0{,}0061$.

Пример. Пусть дана величина $a^* = 3{,}167$ с абсолютной погрешностью $\Delta a^* = 0{,}004$. Вычислить приближённо величину

$$f = \frac{(a - 3)^2}{a + 3},$$

относительную и абсолютную погрешности f^*.

Подставляем a^* и находим $f^* = 0{,}004522$, относительная погрешность $\delta a^* = 0{,}0013$. Относительная погрешность величины $a^* - 3$ вычисляется по формуле относительной погрешности разности двух чисел одного знака (предполагаем, что константа 3 дана без погрешности):

$$\delta 1 = \delta a^* \frac{|a^*|}{|a^* - 3|} = 0{,}048.$$

Погрешность числителя $\delta 2 = 2\delta 1 = 0{,}096$, предельная относительная погрешность знаменателя равна погрешности δa^*. Предельная погрешность величины f^* равна $\delta f^* = \delta 2 + {+}\delta a^* = 0{,}097$. Абсолютная погрешность величины f^* $\Delta f^* = 0{,}097 \cdot 0{,}004522 = 0{,}00044 \approx$ $\approx 0{,}0005$. Тогда величина f^* с верными знаками $f^* = 0{,}005$. Ответ с одним сомнительным запасным знаком: $f^* = 0{,}0045$, погрешности $\Delta f^* = 0{,}00044$, $\delta f^* = 9{,}7\%$.

Пример. Пусть дана величина $a^* = 3{,}167$ с абсолютной погрешностью $\Delta a^* = 0{,}004$ и $b^* = -0{,}873$ с абсолютной погрешностью $\Delta b^* = 0{,}001$. Надо вычислить

$$f = \frac{2(a+b)}{a-b}$$

и её относительную и абсолютную погрешности.

Подставляем a^* и b^* и находим $f^* = 1{,}13564$; относительная погрешность $\delta a^* = 0{,}0013$, относительная погрешность $\delta b^* = 0{,}0011$, относительная погрешность величины $(a+b)^*$ оценивается по формуле относительной погрешности разности двух чисел одного знака (b^* отрицательно, поэтому разность):

$$\delta 1 = \delta a^* \frac{|a^* - b^*|}{|a^* + b^*|} = 0{,}0023.$$

Погрешность числителя равна $\delta 1$, поскольку погрешность константы 2 нулевая; погрешность знаменателя равна погрешности δa^* (максимальная из δa^*, δb^*). Погрешность f^* равна $\delta f^* = \delta 1 + \delta a^* = 0{,}0023 + 0{,}0013 = 0{,}0036$, абсолютная погрешность f^* $\Delta f^* =$ $= 0{,}0036 \cdot 1{,}13564 \approx 0{,}0041 \leq 0{,}005$. Тогда f^* с верными знаками $f^* = 1{,}14$ или с одним сомнительным $f^* = 1{,}136$, погрешности $\Delta f^* = 0{,}0041$, $\delta f^* = 0{,}36\%$.

9. Погрешности функций

Имеется функция n аргументов $y = f(x_1; \ldots; x_n)$. Для краткости обозначим вектор аргументов $\overline{x} = (x_1; \ldots; x_n)^T$ ($(\cdot)^T$ означает транспонирование) и функцию $y = f(\overline{x})$. Пусть

$\overline{x}^* = (x_1^*; \ldots; x_n^*)^T$ — вектор приближённых значений аргументов, для которых известны предельные абсолютные и относительные погрешности $\overline{\Delta}x_1^*$, … , $\overline{\Delta}x_n^*$, $\overline{\delta}x_1^*$, … , $\overline{\delta}x_n^*$. Решим задачу оценивания по этим данным погрешностей приближённого значения функции $y^* = f(\overline{x}^*)$.

Пусть f непрерывно дифференцируема по всем аргументам в точке $\overline{x}$ — точном векторе. Тогда по формуле Лагранжа для функций нескольких переменных приращение функции выражается следующим образом:

$$y - y^* = f(\overline{x}) - f(\overline{x}^*) = \sum_{j=1}^{n} b_j(\theta)(x_j - x_j^*),$$

где $b_j(\theta) = f'_{x_j}\left(\overline{x}^* + \theta(\overline{x} - \overline{x}^*)\right)$ — частная производная f по переменной x_j, вычисленная на векторе $\overline{x}^* + \theta(\overline{x} - \overline{x}^*)$, θ — параметр, принимающий значение в отрезке $[0; 1]$. Из этой формулы сразу следует оценка абсолютной погрешности y^*:

$$\Delta y^* = |y - y^*| = \left|\sum_{j=1}^{n} b_j(\theta)(x_j - x_j^*)\right| \le \sum_{j=1}^{n} |b_j(\theta)| \cdot |x_j - x_j^*| =$$

$$= \sum_{j=1}^{n} |b_j(\theta)| \cdot \Delta x_j^* \le \sum_{j=1}^{n} B_j \cdot \overline{\Delta} x_j^* = \overline{\Delta}_0 y^*,$$

где

$$B_j = \sup_{\theta \in [0;1]} |b_j(\theta)| = \sup_{\theta \in [0;1]} \left|f'_{x_j}\left(\overline{x}^* + \theta \cdot (\overline{x} - \overline{x}^*)\right)\right|.$$

На практике применение оценки $\overline{\Delta}_0 y^*$ невозможно, так как величины B_j не поддаются хотя бы приближённому эффективному вычислению. Однако при некоторых дополнительных условиях на f можно вывести более применимую оценку погрешности.

Обозначим

$$\rho = \sqrt{\sum_{j=1}^{n} \left(\overline{\Delta} x_j^*\right)^2}$$

Если все частные производные f'_{x_j} непрерывны и функция f достаточно гладкая в некоторой области G n-мерного пространства, содержащей отрезок, соединяющий точки $\overline{x}^*$ и $\overline{x}$, то $b_j(\theta) = b_j(0) + o(1)$ и $B_j = \left|f'_{x_j}(\overline{x}^*)\right| + o(1)$, где $o(1)$ — бесконечно малая первого по-

рядка по ρ (эти равенства подразумеваются в том смысле, что $x = y + o(1)$ тогда и только тогда, когда $x - y \to 0$ при ρ → 0). Следовательно,

$$\overline{\Delta}_0 y^* = \sum_{j=1}^{n} B_j \cdot \overline{\Delta} x_j^* = \sum_{j=1}^{n} \left(\left| f'_{x_j}(\overline{x}^*) \right| + o(1) \right) \overline{\Delta} x_j^* =$$

$$= \sum_{j=1}^{n} \left| f'_{x_j}(\overline{x}^*) \right| \overline{\Delta} x_j^* + o(1) \sum_{j=1}^{n} \overline{\Delta} x_j^* = \sum_{j=1}^{n} \left| f'_{x_j}(\overline{x}^*) \right| \overline{\Delta} x_j^* + \widetilde{o},$$

где

$$\tilde{o} = o(1) \sum_{j=1}^{n} \overline{\Delta} x_j^*.$$

Очевидно, что $\tilde{o}$ также является бесконечно малой по ρ. Величина $\overline{\Delta}_0 y^*$ называется *дифференциальной погрешностью*. Её можно представить в виде суммы $\overline{\Delta}_0 y^* = \overline{\Delta}^0 y^* + \tilde{o}$, где

$$\overline{\Delta}^0 y^* = \sum_{j=1}^{n} \left| f'_{x_j}(\overline{x}^*) \right| \overline{\Delta} x_j^* -$$

главная (линейная) часть, которая называется *линейной оценкой дифференциальной погрешности*.

Итак, при поставленных выше условиях в практических расчётах для абсолютной погрешности y^* можно применять приближённую оценку

$$\Delta y^* = |y - y^*| \le \overline{\Delta}^0 y^* = \sum_{j=1}^{n} \left| f'_{x_j}(\overline{x}^*) \right| \overline{\Delta} x_j^*. \qquad (3)$$

Именно она используется для вычисления предельных погрешностей функций.

Замечание. Линейная оценка является довольно грубой оценкой погрешности $\overline{\Delta}_0 y^*$, поэтому при её применении надо следить за выполнением изложенных выше условий.

Из (3) можно вывести приближённую оценку относительной погрешности функции:

$$\overline{\delta} y^* = \frac{\overline{\Delta}^0 y^*}{|y|} = \frac{1}{|y|} \sum_{j=1}^{n} \left| f'_{x_j}(\overline{x}^*) \right| \overline{\Delta} x_j^* = \sum_{j=1}^{n} \frac{\left| f'_{x_j}(\overline{x}^*) \right| \overline{\Delta} x_j^*}{|y|} =$$

$$= \sum_{j=1}^{n} \left| \frac{x_j f'_{x_j}(\overline{x}^*)}{y} \right| \frac{\overline{\Delta} x_j^*}{|x_j|} = \sum_{j=1}^{n} \vartheta_j \, \overline{\delta} x_j^*,$$

где

$$\vartheta_j = \left|\frac{x_j f'_{x_j}(\overline{x}^*)}{y}\right|.$$

Вычислить ϑ_j невозможно, так как неизвестны точные значения аргументов и функции, поэтому применяется приближённая формула для оценки $\overline{\delta} y^*$:

$$\overline{\delta} y^* \approx \sum_{j=1}^{n} \vartheta_j^* \, \overline{\delta} x_j^*, \qquad (4)$$

где

$$\vartheta_j^* = \left|\frac{x_j^* f'_{x_j}(\overline{x}^*)}{y^*}\right|.$$

Пример. Найти значение функции

$$z = \sqrt{\frac{x^3}{y-x}},$$

если $x^* = 0{,}318 \pm 0{,}0005$, $y^* = 1{,}175 \pm 0{,}0003$. Оценить абсолютную и относительную погрешности результата.

Нетрудно убедиться в том, что значения x^*, y^* даны с верными цифрами. Решим задачу двумя способами: с помощью линейной оценки (3) и оценки относительной погрешности (4).

Способ 1. Из условия следует, что $\overline{\Delta} x^* = 0{,}0005$, $\overline{\Delta} y^* = 0{,}0003$. Вычисляем z^*:

$$z^* = \sqrt{\frac{0{,}318^3}{1{,}175 - 0{,}318}} = 0{,}19372.$$

Оценка (3) для данной функции записывается так:

$$\overline{\Delta} z^* = |z'_x(x^*;\, y^*)|\overline{\Delta} x^* + \left|z'_y(x^*;\, y^*)\right|\overline{\Delta} y^*.$$

Вычисляем частные производные:

$$z'_x(x;\, y) = \frac{1}{2\sqrt{\frac{x^3}{y-x}}} \cdot \frac{3x^2(y-x) + x^3}{(y-x)^2} = \frac{\sqrt{x}(3y-2x)}{2(y-x)^{3/2}} \Rightarrow$$

$$\Rightarrow |z'_x(x^*;\, y^*)| = \left|\frac{\sqrt{0{,}318}(3 \cdot 1{,}175 - 2 \cdot 0{,}318)}{2(1{,}175 - 0{,}318)^{3/2}}\right| = 1{,}0268;$$

$$z'_y(x;\, y) = -\frac{1}{2\sqrt{\frac{x^3}{y-x}}} \cdot \frac{x^3}{(y-x)^2} = -\frac{x^{3/2}}{2(y-x)^{3/2}} \Rightarrow$$

$$\Rightarrow \left|z'_y(x^*;\, y^*)\right| = \left|\frac{(0{,}318)^{3/2}}{2(1{,}175-0{,}318)^{3/2}}\right| = 0{,}1130.$$

Теперь можно найти абсолютную погрешность и округлить z^*:

$$\overline{\Delta}z^* = 1{,}0268 \cdot 0{,}0005 + 0{,}1130 \cdot 0{,}0003 = 0{,}00055 \le 0{,}0006 \;\Rightarrow\; z^* = 0{,}194.$$

Относительная погрешность

$$\overline{\delta}z^* \approx \frac{\overline{\Delta}z^*}{|z^*|} = \frac{0{,}00055}{0{,}19372} = 0{,}00284 \approx 0{,}29\%.$$

Способ 2. Вычислим относительные погрешности:

$$\overline{\delta}x^* \approx \frac{0{,}0005}{0{,}318} = 0{,}00157, \overline{\delta}y^* \approx \frac{0{,}0003}{1{,}175} = 0{,}000256.$$

Записываем формулу (4) для функции z: $\overline{\delta}z^* = \vartheta_x^*\overline{\delta}x^* + \vartheta_y^*\overline{\delta}y^*$. Вычисляем множители ϑ_x^* и ϑ_y^*:

$$\vartheta_x^* = \left|\frac{x^* \cdot z'_x(x^*;\, y^*)}{z(x^*;\, y^*)}\right| = \left|\frac{3y^* - 2x^*}{2(y^* - x^*)}\right| = 1{,}6855;$$

$$\vartheta_y^* = \left|\frac{y^* \cdot z'_y(x^*;\, y^*)}{z(x^*;\, y^*)}\right| = \left|\frac{y^*}{2(y^* - x^*)}\right| = 0{,}6855$$

(несложные алгебраические преобразования здесь пропущены).

Теперь находим относительную и абсолютную погрешности и округляем z^*:

$$\overline{\delta}z^* = 1{,}6855 \cdot 0{,}00157 + 0{,}6855 \cdot 0{,}000256 = 0{,}002825 \approx 0{,}29\%;$$

$$\overline{\Delta}z^* \approx \overline{\delta}z^* \cdot |z^*| = 0{,}002825 \cdot 0{,}19372 = 0{,}00055 \le 0{,}0006 \;\Rightarrow\; z^* = 0{,}194.$$

Итак, обоими способами получили ответ: $z^* = 0{,}194$ (с одной сомнительной цифрой), $\overline{\delta}z^* = 0{,}29\%$, $\overline{\Delta}z^* = 0{,}00055$.

Пример. Найдём линейную оценку дифференциальной погрешности для функции $y = \alpha_1 x_1 + \cdots + \alpha_n x_n$, где $\alpha_i \in \{-1; 1\}$. Частные производные равны $y'_{x_j} = \alpha_j$, поэтому $\left|y'_{x_j}\right| = 1$, откуда получаем оценку абсолютной погрешности:

$$\overline{\Delta}y^* = \overline{\Delta}((\pm x_1 \pm x_2 \pm \cdots \pm x_n)^*) = \overline{\Delta}^0 y^* = \overline{\Delta}x_1^* + \overline{\Delta}x_2^* + \cdots + \overline{\Delta}x_n^*.$$

Таким образом, другим способом получена оценка абсолютной погрешности суммы и разности.

Пусть теперь $y = x_1^{p_1} \cdots x_n^{p_n}$. Тогда

$$y'_{x_j} = \frac{p_j}{x_j} y.$$

Найдём по формуле (4) оценку относительной погрешности y^*:

$$\vartheta_j^* = \left|\frac{x_j^*}{y^*}\cdot\frac{p_j}{x_j^*}y^*\right| = |p_j| \;\Rightarrow\; \overline{\delta}y^* = \sum_{j=1}^{n}|p_j|\overline{\delta}\,x_j^*. \qquad (5)$$

В частности, если $n = 2$, $p_1 = p_2 = 1$, то из (5) получается следствие из теоремы 3 для двух сомножителей, при $p_1 = \cdots = p_n = 1$ — для n сомножителей; если $n = 2$, $p_1 = 1$, $p_2 = -1$, то (5) даёт следствие из теоремы 4.

Пример. Обратная задача погрешности. Дана функция

$$z = \sqrt{\frac{x^3}{y-x}}.$$

Вычислено ее значение при $x^* = 0{,}318$, $y^* = 1{,}175$. С какой точностью должны быть заданы аргументы, если значение функции имеет абсолютную погрешность $\Delta z^* = 0{,}003$?

Используем принцип *равного влияния*: вклады в погрешность функции от каждой переменной должны быть равными. Тогда вклад каждого аргумента в абсолютную погрешность z^* равен

$$\frac{\overline{\Delta}z^*}{2} = |z'_x(x^*;\, y^*)|\overline{\Delta}x^* = \left|z'_y(x^*;\, y^*)\right|\overline{\Delta}y^* = 0{,}0015.$$

Так как $|z'_x(x^*;\, y^*)| = 1{,}0268$; $\left|z'_y(x^*;\, y^*)\right| = 0{,}1130$, то

$$\overline{\Delta}x^* = \frac{0{,}0015}{1{,}0268} = 0{,}00146 \approx 0{,}0015; \overline{\Delta}y^* = \frac{0{,}0015}{0{,}1130} = 0{,}0133 \approx 0{,}013,$$

$$\overline{\delta}x^* = \frac{\overline{\Delta}x^*}{|x^*|} = 0{,}005, \overline{\delta}y^* = \frac{\overline{\Delta}y^*}{|y^*|} = 0{,}01.$$