Курс лекций для групп 111 и 113
«ВВЕДЕНИЕ В ЧИСЛЕННЫЕ МЕТОДЫ»

 А. В. Абраменков

Лекция 7. Численное дифференцирование

1. Погрешность численного дифференцирования

Простейшая формула численного дифференцирования получается непосредственно из определения производной как предела отношения приращения функции к приращению аргумента, когда последнее стремится к нулю. Поэтому можно ожидать, что при достаточно малом h

$$f'(x) \approx \frac{f(x+h) - f(x)}{h}\,. \tag{1}$$

Погрешность такого приближения можно оценить, воспользовавшись разложением в ряд Тейлора:

$$f(x+h) = f(x) + hf'(x) + \frac{h^2}{2}f''(x) + \frac{h^3}{6}f'''(x) + \cdots \tag{2}$$

Перенося $f(x)$ в левую часть и деля обе части на h, получаем:

$$\frac{f(x+h) - f(x)}{h} = f'(x) + \frac{h}{2}f''(x) + \frac{h^2}{6}f'''(x) + \cdots$$

Таким образом, разница δ_t между приближенным выражением (1) и точным значением производной определяется высшими членами ряда (2):

$$\delta_t = \frac{h}{2}f''(x) + \frac{h^2}{6}f'''(x) + \cdots \tag{3}$$

Очевидно, что $\delta_t \to 0$ при $h \to 0$; кроме того, при малых h величина δ_t приблизительно пропорциональна h, поскольку в этом случае можно пренебречь членами высших порядков. По мере роста h начинает сказываться квадратичный член, а затем и более высокие степени h.

Формула (1) получается из разложения (2) в результате отбрасывания членов второго и более высоких порядков. По этой причине величина δ_t называется *ошибкой усечения*, так как она вызвана усечением бесконечного ряда. (Буква t в обозначении δ_t является первой буквой слова *truncation* — усечение или обрыв ряда). Существует и другое название для δ_t — *ошибка дискретизации*. Это название отражает тот факт, что ошибка связана с заменой предела отношением конечных приращений, т. е. с переходом от непрерывно меняющегося аргумента к сетке дискретных значений (с шагом h).

На первый взгляд может показаться, что с помощью формулы численного дифференцирования (1) можно получить значение производной с любой желаемой степенью точности — надо лишь взять шаг h достаточно малым. Действительно, выбором шага можно сделать ошибку усечения (3) меньше любой заранее заданной величины. Однако не следует забывать, что кроме ошибки самой формулы (1) существует еще один источник погрешностей, а именно, *ошибки округления*. Влияние ошибок округления может быть весьма значительным, поскольку по мере уменьшения h значения функции $f(x)$ и $f(x+h)$ становятся все более близкими между собой, и в результате их вычитания в числителе формулы (1) происходит сокращение старших значащих цифр и увеличение относительной погрешности получаемой разности (явление потери точности, которое обсуждалось в лекции 1).

Пусть ε — относительная погрешность вычисления значений функции $f(x)$, так что $|\Delta f| \leqslant \varepsilon \cdot |f(x)|$, где Δf — абсолютная погрешность $f(x)$. Величина ε зависит от точности машинной арифметики и в лучшем случае находится на уровне «машинного эпсилон», т. е. относительной погрешности машинного представления чисел с плавающей точкой. Если вычисление функции требует большого числа арифметических операций, то из-за накопления ошибок округления ε будет больше уровня предельной машинной точности.

Как известно, максимальная абсолютная погрешность суммы или разности равна сумме абсолютных погрешностей слагаемых. Поэтому оценка сверху для абсолютной погрешности числителя формулы (1) равна $2\varepsilon \cdot |f(x)|$ (мы пренебрегаем различием между $f(x)$ и $f(x+h)$, поскольку они близки и уровень ошибок округления у них практически одинаков). Таким образом, абсолютная ошибка округления δ_r при использовании формулы (1) в первом приближении определяется выражением

$$\delta_r \leqslant \frac{2\varepsilon \cdot |f(x)|}{h} \tag{4}$$

Мы не учитываем здесь погрешность значения h и ошибку при вычислении частного, поскольку эти вклады несущественны для нашего рассмотрения (операция деления не может заметно повысить относительную погрешность результата; кроме того, шаг h всегда можно выбрать так, чтобы он допускал точное машинное представление).

Как следует из выражения (4), ошибка округления при численном дифференцировании обратно пропорциональна величине шага h. По мере уменьшения шага ошибка усечения снижается, а ошибка округления неограниченно растет. Если шаг увеличивать, то ошибка округления будет уменьшаться, зато возрастет ошибка усечения. Общая погрешность приближенного вычисления производной равна сумме ошибок усечения (дискретизации) и округления. Поведение общей погрешности формулы (1) и ее составляющих показано на рис. 1.

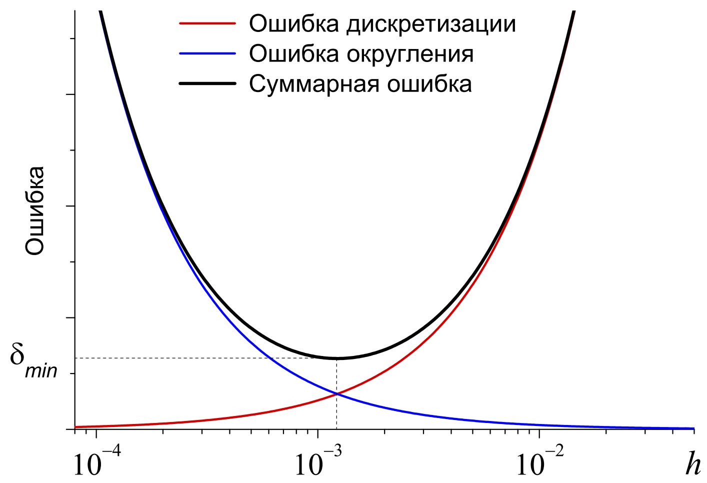

Рис. 1. Зависимость общей погрешности и ее составляющих (ошибок усечения и округления) от величины шага численного дифференцирования h.

Как видно из рисунка, существует некоторый оптимальный шаг, при котором суммарная ошибка имеет наименьшую величину (δ_{min}). Это шаг, при котором ошибки усечения и округления приблизительно одинаковы. При меньших величинах шага доминирующий вклад в погрешность дифференцирования дает ошибка округления, при бóльших — ошибка усечения. Оптимальный шаг и величина δ_{min} зависят от свойств дифференцируемой функции и точности машинной арифметики, но в каждом случае существует некоторая предельная точность численного дифференцирования, которую уже нельзя улучшить.

Из сказанного выше следует, что оптимальный шаг определяется условием $|\delta_r| \approx |\delta_t|$, то есть

$$\frac{2\varepsilon \cdot |f(x)|}{h} = \left| \frac{h}{2} f''(x) + \frac{h^2}{6} f'''(x) + \cdots \right|. \tag{5}$$

Однако, чтобы оценить отсюда величину h, необходимо иметь значения второй и более высоких производных, а их можно получить не раньше, чем будет найдена первая производная. На практике обычно выбирают шаг, ориентируясь по оценке ошибки округления, а именно, берут минимальное значение h, при котором δ_r не превышает максимальной допустимой погрешности результата (хотя вполне возможно, что фактическая ошибка при этом окажется все же слишком большой из-за того, что велика ошибка усечения).

2. Формулы более высоких порядков точности

Есть две принципиальных возможности повысить точность численного дифференцирования:

1. Увеличить точность вычисления значений функции (уменьшить величину ε). Тогда понизится уровень ошибок округления δ_r, можно будет работать в области малых h и тем самым сократить ошибку усечения, а вместе с ней и общую погрешность дифференцирования.

2. Использовать формулу, обладающую меньшей ошибкой усечения. Тогда без ухудшения δ_t можно взять большее значение шага h и тем самым сократить вклад ошибки округления, получая в результате меньшую величину общей погрешности.

Первый путь требует более точного представления машинных чисел. Например, в языке C существуют по крайней мере два типа вещественных чисел (`float` и `double`), которые различаются величиной машинного эпсилон, а следовательно, имеют разные уровни ошибок округления. Предусмотрен также тип `long double`, который может иметь еще более высокую точность, чем `double` (при условии, что процессор в самом деле умеет работать с такими числами). В других языках программирования также предусмотрена возможность работы с числами разной степени точности[1]. Недостатки этого подхода связаны с тем, что вычисления повышенной точности, как правило, происходят с меньшей скоростью, а числа занимают больше места в памяти.

На практике чаще применяется второй путь, основанный на уменьшении ошибок усечения. Рассмотрим простейшую из таких формул:

$$f'(x) \approx \frac{f(x+h) - f(x-h)}{2h}. \tag{6}$$

По смыслу эта формула не отличается от (1) — производная по-прежнему оценивается как отношение приращения функции к приращению аргумента. Разница состоит лишь в том, что значения функции вычисляются в двух точках, взятых на равных расстояниях h слева и справа от точки x, где оценивается производная.

Оценим ошибку усечения формулы (6) тем же способом, что и в случае формулы (1). Для этого представим значения функции в точках $x+h$ и $x-h$ в виде разложений в ряд:

$$f(x+h) = f(x) + hf'(x) + \frac{h^2}{2}f''(x) + \frac{h^3}{6}f'''(x) + \frac{h^4}{24}f^{(4)}(x) + \frac{h^5}{120}f^{(5)}(x) + \cdots$$

$$f(x-h) = f(x) - hf'(x) + \frac{h^2}{2}f''(x) - \frac{h^3}{6}f'''(x) + \frac{h^4}{24}f^{(4)}(x) - \frac{h^5}{120}f^{(5)}(x) + \cdots$$

Вычитая второе уравнение из первого и деля обе части полученного равенства на $2h$, имеем:

$$\frac{f(x+h) - f(x-h)}{2h} = f'(x) + \frac{h^2}{6}f'''(x) + \frac{h^4}{120}f^{(5)}(x) + \cdots$$

[1]Можно также программным путем реализовать арифметику многократной точности. Одна из наиболее известных библиотек для этой цели — GMP (**G**NU **M**ultiple **P**recision Arithmetic Library), которая распространяется свободно вместе с исходными текстами. GMP позволяет выполнять вычисления с любой точностью — 50, 100, 1000 значащих цифр и более. Существуют и другие библиотеки многократной точности. Однако скорость счета при таких вычислениях падает во много раз, поскольку каждая арифметическая операция выполняется с помощью соответствующей библиотечной функции, как правило, весьма сложной.

Таким образом, выражение для ошибки усечения формулы (6) имеет вид:

$$\delta_t = \frac{h^2}{6} f'''(x) + \frac{h^4}{120} f^{(5)}(x) + \cdots \tag{7}$$

Сравнивая выражения (7) и (3), мы видим, что у формулы (6) главный член погрешности зависит от h квадратично, тогда как в случае формулы (1) эта зависимость линейная. Поэтому по мере уменьшения h ошибка усечения формулы (6) убывает гораздо быстрее, чем ошибка формулы (1).

Введем понятие *порядка точности* формулы численного дифференцирования: порядок точности равен максимальной степени многочлена, для которого формула дает точный результат (то есть нулевую ошибку усечения). Формула (1) имеет первый порядок точности, так как ее погрешность (3) содержит вторую и более высокие производные дифференцируемой функции, которые тождественно равны нулю у многочленов нулевой и первой степеней. У многочлена второй степени вторая производная отлична от нуля, и формула (1) в этом случае уже не даст точного результата. В то же время ошибка формулы (6) определяется третьей и более высокими производными. Поэтому формула (6) даст точный результат и для многочлена второй степени, то есть она имеет второй порядок точности.

Более высокая точность формулы (6) по сравнению с (1) иллюстрируется рисунком 2, где показано дифференцирование квадратичной функции. Геометрический смысл обеих формул одинаков. Каждая из них дает тангенс угла наклона секущей, проведенной через две точки на кривой $f(x)$. Из рисунка видно, что секущая 2, построенная по узлам $x - h$ и $x + h$, параллельна касательной в точке x, тогда как секущая 1, построенная по узлам x и $x + h$, значительно отличается по наклону от касательной.

Формулы (1) и (6) имеют одинаковую структуру, поэтому ошибки округления для них практически совпадают. Строго говоря, знаменатель у формулы (6) вдвое больше по сравнению с (1), так что максимальная ошибка округления у (6) в два раза меньше. Однако, это несущественное различие, поскольку оно относится к оценкам сверху, тогда как реальная ошибка с достаточно высокой вероятностью может оказаться меньше в 2 и более раз. На рис. 3 показано поведение ошибок усечения и округления и суммарной погрешности формул (1) и (6) в зависимости от величины шага h. Видно, что для формулы второго порядка оптимальный шаг

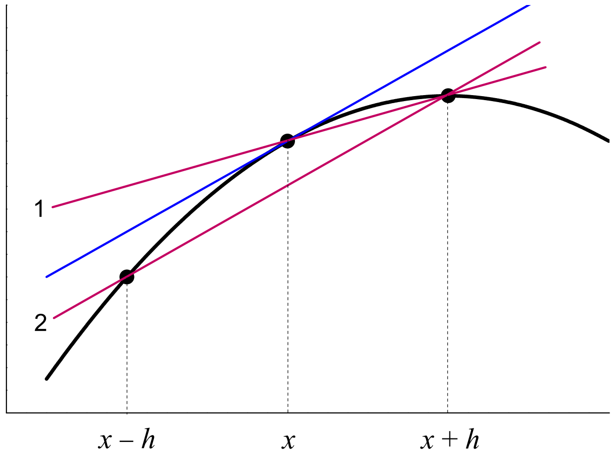

Рис. 2. Оценка производной квадратичной функции с помощью формул первого (1) и второго (2) порядков точности.

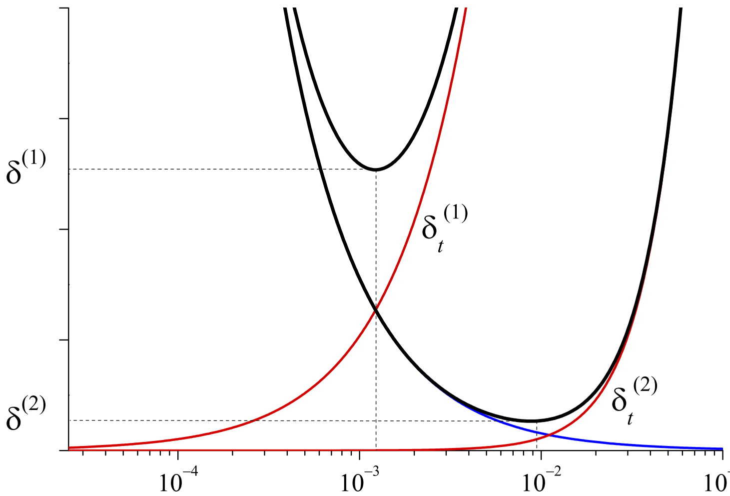

Рис. 3. Погрешность численного дифференцирования и ее составляющие в случае формул первого и второго порядков точности.

больше, а предельный уровень погрешности, соответственно, ниже.

Рассмотрим общий путь получения формул численного дифференцирования произвольного порядка точности. Возьмем несколько узлов x_k вблизи точки x, в которой оценивается производная, вычислим в этих узлах значения дифференцируемой функции $f(x)$, и построим интерполяционный многочлен $P(x)$ соответствующей степени. Многочлен служит приближенной заменой функции $f(x)$, поэтому его производную $P'(x)$ естественно принять в качестве оценки $f'(x)$. Если интерполяционный многочлен записать в форме Лагранжа, то получится следующая структура формул численного дифференцирования:

$$\begin{aligned} f'(x) \approx \frac{dP_n(x)}{dx} &= \frac{d}{dx}\sum_{k=0}^{n} f(x_k)\, L_k^{(n)}(x) = \\ &= \sum_{k=0}^{n} f(x_k)\, \frac{dL_k^{(n)}(x)}{dx} = \sum_{k=0}^{n} w_k\, f(x_k)\,, \end{aligned} \tag{8}$$

причем w_k — значение $dL_k^{(n)}(x)/dx$ в той точке, где оценивается производная функции. Таким образом, приближенное значение производной вычисляется как сумма значений дифференцируемой функции в заданных узлах, умноженных на некоторые коэффициенты w_k, называемые весами. Так как $L_k^{(n)}(x)$ не зависят от $f(x)$, а определяются исключительно выбранными узлами, точнее, расстояниями между ними, то w_k зависят лишь от числа и расположения узлов. В частности, при равноотстоящих узлах w_k зависят от h и от числа узлов, т. е. от степени многочлена $P(x)$. Используя интерполяционный многочлен n-й степени, получим формулу численного дифференцирования n-го порядка точности.

Например, чтобы получить формулу 2-го порядка точности, возьмем три равноотстоящих узла $x_0 = x - h$, $x_1 = x$, $x_2 = x + h$ и построим по ним многочлен Лагранжа второй степени. Найдем веса для вычисления производной в точке x, совпадающей со средним узлом:

$$w_0 = \frac{d}{dx}\left[\frac{(x-x_1)(x-x_2)}{(x_0-x_1)(x_0-x_2)}\right] = \frac{(x-x_1)+(x-x_2)}{2h^2} = \frac{-h}{2h^2} = -\frac{1}{2h}\,,$$

$$w_1 = \frac{d}{dx}\left[\frac{(x-x_0)(x-x_2)}{(x_1-x_0)(x_1-x_2)}\right] = \frac{(x-x_0)+(x-x_2)}{-h^2} = \frac{h-h}{-h^2} = 0,$$

$$w_2 = \frac{d}{dx}\left[\frac{(x-x_0)(x-x_1)}{(x_2-x_0)(x_2-x_1)}\right] = \frac{(x-x_0)+(x-x_1)}{2h^2} = \frac{h}{2h^2} = \frac{1}{2h}\,.$$

Таким образом, имеем формулу:

$$f'(x) \approx -\frac{1}{2h} f(x-h) \; + \; 0 \cdot f(x) \; + \; \frac{1}{2h} f(x+h) = \frac{f(x+h) - f(x-h)}{2h}\,.$$

Как и следовало ожидать, мы получили формулу (6).

Теперь становится понятно, почему формула (6), использующая значения функции лишь в двух точках, имеет второй порядок точности: она фактически соответствует интерполяционному многочлену второй степени, построенному по трем узлам, но один из весов равен нулю, так что явным образом в формулу входят только два узла. Если тот же многочлен использовать для оценки производной в точке, не совпадающей с узлом x_1, то все веса окажутся отличными от нуля. Полученная формула по-прежнему будет иметь второй порядок, но теперь в нее войдут значения функции в трех разных узлах.

При симметричном размещении узлов относительно точки x (совпадающей со средним узлом) всегда получается нулевой вес для среднего узла. Поэтому для формулы четного порядка n достаточно значений функции в n точках, отличных от той, где оценивается производная. Так, взяв 5 узлов $x_0 = x-2h$, $x_1 = x-h$, $x_2 = x$, $x_3 = x+h$, $x_4 = x+2h$ и построив по ним многочлен 4-й степени, получим формулу 4-го порядка точности

$$f'(x) \approx \frac{1}{12h}\Big[f(x-2h) - 8f(x-h) + 8f(x+h) - f(x+2h)\Big], \tag{9}$$

куда входят лишь 4 значения дифференцируемой функции, поскольку $w_2 = 0$.

3. Формулы для высших производных

Высшие производные можно получить либо многократным последовательным численным дифференцированием, либо непосредственным использованием n-й производной интерполяционного многочлена. Результаты при этом будут одинаковыми.

Возьмем, например, три узла $x_0 = x - h$, $x_1 = x$, $x_2 = x + h$. Применяя формулу (6), найдем первую производную в двух средних точках, расположенных между соседними узлами:

$$f'(x-\tfrac{h}{2}) \approx \frac{f(x)-f(x-h)}{h}\,,$$
$$f'(x+\tfrac{h}{2}) \approx \frac{f(x+h)-f(x)}{h}\,.$$

Применяя формулу (6) к полученным значениям первой производной, найдем вторую производную в средней точке x:

$$f''(x) \approx \frac{f'(x+\frac{h}{2})-f'(x-\frac{h}{2})}{h} \approx \frac{f(x+h)-2f(x)+f(x-h)}{h^2}\,. \tag{10}$$

Оценка погрешности показывает, что формула (10) имеет второй порядок точности.

Тот же результат можно получить, записывая вторую производную для интерполяционного многочлена второй степени и вычисляя веса аналогично тому, как это было сделано выше:

$$w_0 = \frac{d^2}{dx^2}\left[\frac{(x-x_1)(x-x_2)}{(x_0-x_1)(x_0-x_2)}\right] = \frac{2}{2h^2} = \frac{1}{h^2}\,,$$
$$w_1 = \frac{d^2}{dx^2}\left[\frac{(x-x_0)(x-x_2)}{(x_1-x_0)(x_1-x_2)}\right] = \frac{2}{-h^2} = -\frac{2}{h^2}\,,$$
$$w_2 = \frac{d^2}{dx^2}\left[\frac{(x-x_0)(x-x_1)}{(x_2-x_0)(x_2-x_1)}\right] = \frac{2}{2h^2} = \frac{1}{h^2}\,.$$

Точно так же может быть получена формула 4-го порядка точности для второй производной:

$$f''(x) \approx \frac{1}{12h^2}\Big[-f(x-2h)+16f(x-h)-30f(x)+16f(x+h)-f(x+2h)\Big]. \tag{11}$$

4. Численное дифференцирование экспериментальных данных

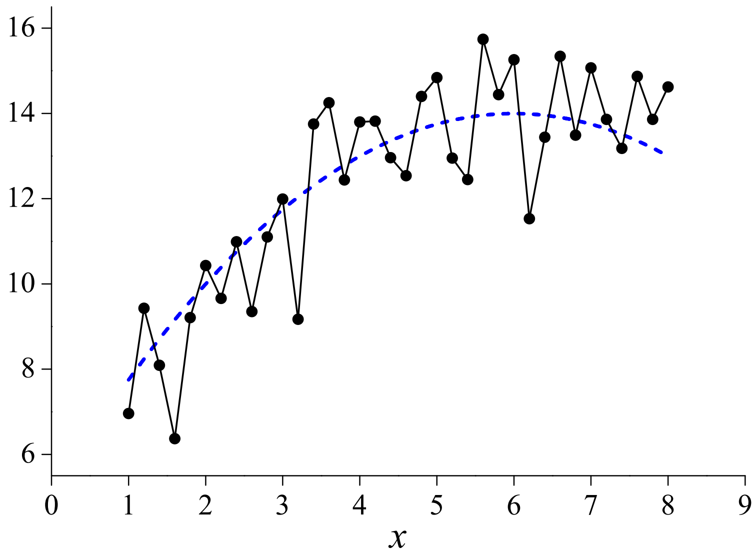

Рис. 4. Пример численного дифференцирования данных, содержащих случайные ошибки со средним уровнем 10%.