

Точность вычислений при программировании алгоритмов численного интегрирования

77-48211/671065

12, декабрь 2013

Папшев И. С., Булатова И. Г.

УДК 519.6

Россия, МГТУ им. Н.Э. Баумана

ispapshev@rambler.ru

Точность вычислений при программировании алгоритмов численного интегрирования

Вычислительная погрешность складывается из следующих составляющих:

- Методическая погрешность, связанная с заменой дифференциальных уравнений их разностными аналогами;
- Погрешность округления, связанная с конечной длиной разрядной сетки компьютера.

Наиболее популярными методами, используемыми в алгоритмах численного интегрирования, являются методы, в которых область интегрирования делится на равные отрезки. К ним относятся метод прямоугольников, метод трапеций и метод парабол (метод Симпсона). Эти алгоритмы являются итерационными, и время выполнения программ, их реализующих, сильно зависит от требуемой точности вычислений и от вида интегрируемой функции.

Оценка точности производилась путем сравнения вычисленной интегральной суммы и значения интеграла, вычисленного аналитически.

Если погрешность вычислений не должна превышать $10^{-6} \div 10^{-10}$, то программа, реализующая алгоритм численного интегрирования, может заикнуться или завершиться, не достигнув требуемой точности (неустойчивость решения). Причиной такого поведения программы является ограниченная точность выполнения операций в ПК. При описании математических методов численного интегрирования возможные погрешности вычислений обычно не учитывают.

Чтобы избежать зависания программы и повысить достижимую точность интегрирования, при программировании алгоритмов необходимо принимать специальные меры для уменьшения влияния погрешностей вычислений на конечный результат.

С точки зрения математической теории при $\Delta x \rightarrow 0$ интегральная сумма по определению будет равна интегралу Римана (здесь Δx – основание элементарного прямоугольника или высота элементарной трапеции). Но при выполнении программы, вычисляющей интегральную сумму, Δx не может бесконечно стремиться к нулю.

Рассмотрим две проблемы: защита от зависания и повышение точности вычисления интегральной суммы.

1. Защита от зависания программы

Для вычисления интегральной суммы используются два вложенных цикла:

- внешний цикл `while((k<20)&&(abs(i1-i2)/3.0)>eps)`, использующий для завершения итераций правило Рунге (в этом цикле dx делится на 2 при каждом проходе) ;
- внутренний цикл `for(x=a+dx;x<b;x=x+dx)`, вычисляющий интегральную сумму.

Оценим $k_{\max}$ - максимальное число повторений внешнего цикла вычисления интегральной суммы, после превышения которого выполнение итераций не имеет смысла и приводит к зависанию алгоритма. Минимальное x , такое, что $1,0 + x \neq 1,0$ для типа `double` в C++ определено как 10^{-9} . Если предположить, что на каждом шаге цикла вычисления интегральной суммы Δx уменьшается в два раза, интервал интегрирования равен 1 и считать, что $2^{10} \approx 10^3$, мантисса чисел типа `double` = 53 бита и результаты операций нормализуются, то разумное число повторений внешнего цикла не должно превышать 30, так как иначе не выполнится условие окончания цикла `for ( x=a; x<b; x+=dx )`. Чем больше x , тем при больших dx выражение $x+=dx$ будет равно x .

Для защиты от заикливания нужно ограничить число повторений внешнего цикла. Если число повторений цикла будет равно $k_{\max}$, то требуемая точность вычислений может быть не достигнута, но программа не зависнет и выполнит вычисления с максимально возможной точностью. Очевидно, что при больших значениях интегральной суммы для оценки точности вычислений должна использоваться относительная погрешность, которая оценивается числом точных старших цифр в значении интегральной суммы.

2. Повышение точности вычисления интегральной суммы

Рассмотрим две причины, приводящие к увеличению числа повторений цикла и, как следствие, к снижению максимально допустимой точности вычислений:

- погрешность, связанная с вычислением правой границы интервала интегрирования b ;
- погрешность вычисления площади элементарного прямоугольника.

2.1. Вычисление площади последнего элемента на правой границе интервала интегрирования

Из-за погрешностей вычислений x по формуле $x=x+dx$ правое основание последней трапеции не будет равно b , и вычисления закончатся при $x < b$, то есть вычисляется не вся площадь под графиком функции. Использование в цикле `for` оператора $x=a+i*dx$; вместо $x=x+dx$ не дает видимых изменений результатов. Можно скорректировать вычисление площади последнего элемента таким образом, чтобы x всегда равнялось b , а высота последней трапеции (или основание последнего прямоугольника) соответственно изменялась. Формула вычисления площади последней трапеции:

$$S = (f(x-dx)+f(b))*(b-x)/2;$$

Здесь:

S – интегральная сумма,

$f(x)$ – подынтегральная функция,

a, b – границы интегрирования.

Формула вычисления площади последнего прямоугольника:

$$S = f(x-dx)*(b-x+dx/2);$$

2.2. Вычисление площади элементарного прямоугольника

Погрешность вычисления площади элементарного прямоугольника относительно площади под кривой может быть существенно уменьшена, если ее вычислять как произведение основания на значение функции в его середине, а не на одной из границ:

$$S = dx * f(x+dx/2).$$

Это очевидно, так как на большинстве элементарных участках функция или монотонно убывает, или монотонно возрастает, и вычисленная таким образом площадь прямоугольника будет меньше отличаться от площади, ограниченной графиком функции.

Таким образом, оператор цикла вычисления интегральной суммы методом прямоугольников (если это цикл for) должен иметь вид:

for ($x=a+dx/2$; $x<b$; $x+=dx$), где a, b – границы интегрирования.

В большинстве случаев трапеция более точно представляет площадь под кривой, чем прямоугольник с высотой, вычисленной на одной из границ основания. Но если вычислять высоту в середине основания, то результат может измениться. Если для вычисления интегральной суммы использовать метод парабол, то для сложных функций можно существенно повысить достижимую точность и сократить число повторений внешнего цикла.

Для оценки сказанного рассмотрим результаты вычисления интегральной суммы для четырех функций с различной точностью (eps) двумя методами с учетом и без учета поправок. В графе Integral приводятся значения интеграла Римана, вычисленные по формулам, в графе IntSum – значения интегральной суммы, а в графе N – число элементарных отрезков Δx , на которые разбивается область интегрирования при достижении требуемой точности вычислений (оно не равно числу повторений цикла).

Метод прямоугольников: eps = 0.0001000; a = -1.000000, b = 3.000000

Function	Integral	IntSum	N
$y=x$	4.00000000000	3.9999389648	131072
$y=\sin(22x)$	-0.0000142441	0.0000549101	1024
$y=x^4$	48.8000000000	48.7999237061	2097152
$y=\arctg(x)$	2.1570201976	2.1569581112	65536

Метод прямоугольников: eps = 0.0000001; a = -1.000000, b = 3.000000

Function	Integral	IntSum	N
$y=x$	4.00000000000	3.9999999404	134217728
$y=\sin(22x)$	-0.0000142441	-0.0000141766	1048576
$y=x^4$	48.8000000000	48.7999999255	-2147483648
$y=\arctg(x)$	2.1570201976	2.1570201369	67108864

Метод трапеций: $\text{eps} = 0.0001000$; $a = -1.000000$, $b = 3.000000$

Function	Integral	IntSum	N
$y=x$	4.00000000000	4.00000000000	8
$y=\sin(22x)$	-0.0000142441	0.0000786901	8
$y=x^4$	48.8000000000	48.8000356038	2048
$y=\arctg(x)$	2.1570201976	2.1569876461	128

Метод трапеций: $\text{eps} = 0.0000001$; $a = -1.000000$, $b = 3.000000$

Function	Integral	IntSum	N
$y=x$	4.00000000000	4.00000000000	8
$y=\sin(22x)$	-0.0000142441	-0.0000142090	512
$y=x^4$	48.8000000000	48.8000000348	65536
$y=\arctg(x)$	2.1570201976	2.1570201658	4096

Рис. 1. Без поправки на правой границе, высота в начале интервала

В методе прямоугольников для функции $y = x^4$ наблюдается переполнение разрядной сетки ($N > 2^{31}-1$). Время вычислений – несколько минут.

Метод прямоугольников: $\text{eps} = 0.0001000$; $a = -1.000000$, $b = 3.000000$

Function	Integral	IntSum	N
$y=x$	4.00000000000	3.9999389574	32768
$y=\sin(22x)$	-0.0000142441	0.00000005673	4096
$y=x^4$	48.8000000000	48.8012731737	1024
$y=\arctg(x)$	2.1570201976	2.1569243188	16384

Метод прямоугольников: $\text{eps} = 0.0000001$; $a = -1.000000$, $b = 3.000000$

Function	Integral	IntSum	N
$y=x$	4.00000000000	3.9999999404	33554432
$y=\sin(22x)$	-0.0000142441	-0.0000141740	262144
$y=x^4$	48.8000000000	48.80000000596	33554432
$y=\arctg(x)$	2.1570201976	2.1570201040	16777216

Метод трапеций: $\text{eps} = 0.0001000$; $a = -1.000000$, $b = 3.000000$

Function	Integral	IntSum	N
$y=x$	4.00000000000	4.00000000000	8
$y=\sin(22x)$	-0.0000142441	0.0000786901	8
$y=x^4$	48.80000000000	48.8000356038	2048
$y=\arctg(x)$	2.1570201976	2.1569876461	128

Метод трапеций: $\text{eps} = 0.0000001$; $a = -1.000000$, $b = 3.000000$

Function	Integral	IntSum	N
$y=x$	4.00000000000	4.00000000000	8
$y=\sin(22x)$	-0.0000142441	-0.0000142090	512
$y=x^4$	48.80000000000	48.8000000348	65536
$y=\arctg(x)$	2.1570201976	2.1570201658	4096

Рис. 2. С поправкой на правой границе, высота в начале интервала

Число элементарных отрезков в методе прямоугольников существенно уменьшилось, а в методе трапеций осталось без изменений.

Метод прямоугольников: $\text{eps} = 0.0001000$; $a = -1.000000$, $b = 3.000000$

Function	Integral	IntSum	N
$y=x$	4.00000000000	4.00000000000	4
$y=\sin(22x)$	-0.0000142441	0.0001110392	8
$y=x^4$	48.80000000000	48.7999821981	2048
$y=\arctg(x)$	2.1570201976	2.1570364731	128

Метод прямоугольников: $\text{eps} = 0.0000001$; $a = -1.000000$, $b = 3.000000$

Function	Integral	IntSum	N
$y=x$	4.00000000000	4.00000000000	4
$y=\sin(22x)$	-0.0000142441	-0.0000142617	512
$y=x^4$	48.80000000000	48.7999999826	65536
$y=\arctg(x)$	2.1570201976	2.1570202135	4096

Рис. 3. С поправкой на правой границе, высота в середине dx .

Function	Integral	IntSum	N
$y=x$	4.00000000000	4.00000000000	4
$y=\sin(22x)$	-0.0000142441	0.0001110392	8
$y=x^4$	48.80000000000	48.7999821981	2048
$y=\arctg(x)$	2.1570201976	2.1570364731	128

Метод прямоугольников: $\text{eps} = 0.0000001$; $a = -1.000000$, $b = 3.000000$

Function	Integral	IntSum	N
$y=x$	4.00000000000	4.00000000000	4
$y=\sin(22x)$	-0.0000142441	-0.0000142617	512
$y=x^4$	48.80000000000	48.7999999826	65536
$y=\arctg(x)$	2.1570201976	2.1570202135	4096

Рис. 4. Без поправки на правой границе, высота в середине dx .

Результаты совпадают с предыдущим экспериментом. Влияние коррекции незаметно.

Метод трапеций $a=0.000000$; $b=100.000000$; $\text{eps}=0.0000100000$; $k_{\max}=20$

Function	Integral	IntSum	N
$y=x$	5000.00000000000	5000.00000000000	4
$y=\sin(22x)$	0.0166736717	0.0166674081	32768
$y=x^4$	2000000000.0000000000	2000000000.0007560000	2097152
$y=\arctg(x)$	151.4744458273	151.4744427232	16384

Метод трапеций $a=0.000000$; $b=100.000000$; $\text{eps}=0.0000000001$; $k_{\max}=20$

Function	Integral	IntSum	N
$y=x$	5000.00000000000	5000.00000000000	4
$y=\sin(22x)$	0.0166736717	0.0166736702	2097152
$y=x^4$	2000000000.0000000000	2000000000.0007560000	2097152
$y=\arctg(x)$	151.4744458273	151.4744458271	2097152

Рис. 5. С ограничением числа повторений внешнего цикла $k_{\max}=20$

Если число повторений внешнего цикла не ограничивать, то для приведенных на рис. 5 значениях границ области интегрирования и точности вычислений программа зависала.

Значение $N=2097152$ получается при завершении цикла по условию $k_{\max}=20$.

Достижимая точность вычислений ограничивается 13-ю десятичными знаками.

Выводы.

1. В общем трапеция более точно представляет площадь под кривой, чем прямоугольник с высотой, вычисленной на одной из границ основания. Но если вычислять высоту в середине основания, то результат может измениться. Если для вычисления интегральной суммы использовать метод парабол, то для немонотонных функций можно существенно повысить достижимую точность и сократить число повторений внешнего цикла.
2. Для защиты от заикливания нужно ограничить число повторений внешнего цикла. Если число повторений цикла будет равно $k_{\max}$, то требуемая точность вычислений может быть не достигнута, но программа не зависнет и выполнит вычисления с максимально возможной точностью. $k_{\max}$ следует передавать в функцию численного интегрирования в качестве параметра, чтобы в случае зависания программы можно было легко изменять его значение.
3. При больших значениях интегральной суммы для завершения итераций следует использовать относительную погрешность. Максимальное число точных десятичных цифр в вычисленной интегральной сумме зависит от типа функции и равно 12 – 15.
4. Коррекция площади последнего элемента оказывает заметное влияние только при большом числе элементарных отрезков. Если выполнить условие п.1, то в рассмотренных случаях оно незаметно. С другой стороны, использование коррекции лишь незначительно увеличивает время вычислений и поэтому ее также следует использовать при программировании алгоритмов численного интегрирования.

Список литературы

1. Бахвалов Н., Жидков Н., Кобельков Г. Численные методы. М.: Лаборатория базовых знаний, 2001. 632 с.
2. Г.Корн, Т.Корн. Справочник по математике (для научных работников и инженеров). М, 1974 г., 832 стр. G. Korn, T. Korn. Mathematical handbook for scientists and engineers. NY, 1968