Вяцков Михаил, КН-401

Лабораторная работа №1

По условиям лабораторной

$$k = 3$$
$$a = 1 + 0.1k = 1.3$$
$$b = 2 - 0.1k = 1.7$$

Рассматриваемый ряд имеет вид:

$$\sum_{n=1}^{\infty} \frac{1}{0.2n^2 + 1.3n + 1.7}$$

Рассмотрим последовательно исходный ряд и все улучшения

1 Вычисления с погрешностью

Теперь, допустим мы использем арифметику с плавающей запятой с экспонентой 10 и мантиссой длны t. Допустим мы работаем только с положительными числами. Пусть все числа изначально хранятся с относительной точностью $\varepsilon = 5 \cdot 10^{-t}$. Пусть есть два числа x и y, вычисленных с относительной точностью ε_x и ε_y. Тогда в машинном представлении на самом деле хранятся числа $x(1 + \delta_x)$ и $y(1 + \delta_y)$, где $|\delta_x| \le \varepsilon_x$ и $|\delta_y| \le \varepsilon_y$.

Рассмотрим сумму:

$$x(1 + \delta_x) + y(1 + \delta_y) = (x + y)(1 + \frac{x}{x+y}\delta_x + \frac{y}{x+y}\delta_y)$$

$$\left|\frac{x}{x+y}\delta_x + \frac{y}{x+y}\delta_y\right| \le \left|\frac{x}{x+y}\delta_x\right| + \left|\frac{y}{x+y}\delta_y\right| \le \frac{x}{x+y}\varepsilon_x + \frac{y}{x+y}\varepsilon_y$$

$$\frac{x}{x+y}\varepsilon_x + \frac{y}{x+y}\varepsilon_y \le \frac{x}{x+y}\max\{\varepsilon_x, \varepsilon_y\} + \frac{y}{x+y}\max\{\varepsilon_x, \varepsilon_y\} = \max\{\varepsilon_x, \varepsilon_y\}\left(\frac{x}{x+y} + \frac{y}{x+y}\right)$$

$$\frac{x}{x+y} + \frac{y}{x+y} = \frac{x+y}{x+y} = 1 \implies \left|\frac{x}{x+y}\delta_x + \frac{y}{x+y}\delta_y\right| \le \max\{\varepsilon_x, \varepsilon_y\}$$

Таким образом, точно вычисленная сумма приближенных значений будет иметь относительную погрешность $\max\{\varepsilon_x, \varepsilon_y\}$, но кроме того, нам необходимо теперь сохранить это значение, а значит в худшем случае к относительной погрешности добавится еще ε.

Рассмотрим произведение:

$$x(1 + \delta_x)y(1 + \delta_y) = xy(1 + \delta_x + \delta_y + \delta_x\delta_y)$$
$$|\delta_x + \delta_y + \delta_x\delta_y| \le \varepsilon_x + \varepsilon_y + \varepsilon_x\varepsilon_y$$

Опять же, после вычисления точного значения может произойти округление, которое добавит к относительное погрешности ε.

Рассмотрим отношение:

$$\frac{x(1 + \delta_x)}{y(1 + \delta_y)} = \frac{x}{y}\frac{1 + \delta_x}{1 + \delta_y}$$

Разложим знаменатель второй дроби в ряд Тейлора:

$$(1 + \delta_y)^{-1} = 1 - \delta_y + \delta_y^2 - \delta_y^3 + .. = 1 - \delta_y(1 - \delta_y + \delta_y^2 - ..) = 1 - \frac{\delta_y}{1 + \delta_y}$$

$$\frac{x}{y}\frac{1+\delta_x}{1+\delta_y} = \frac{x}{y}(1+\delta_x)(1-\frac{\delta_y}{1+\delta_y}) = \frac{x}{y}(1+\delta_x-\frac{\delta_y}{1+\delta_y}-\frac{\delta_x\delta_y}{1+\delta_y})$$

$$\left|\delta_x-\frac{\delta_y}{1+\delta_y}-\frac{\delta_x\delta_y}{1+\delta_y}\right| \leq \varepsilon_x+\frac{\varepsilon_y}{|1+\delta_y|}+\frac{\varepsilon_x\varepsilon_y}{|1+\delta_y|}$$

Пусть $\max\{\varepsilon_x,\varepsilon_y\}<1$, тогда

$$|1+\delta_y| \geq 1+\delta_y \geq 1-\varepsilon_y$$

$$\frac{1}{|1+\delta_y|} \leq \frac{1}{1-\varepsilon_y}$$

$$\varepsilon_x+\frac{\varepsilon_y}{|1+\delta_y|}+\frac{\varepsilon_x\varepsilon_y}{|1+\delta_y|} \leq \varepsilon_x+\frac{\varepsilon_y}{1-\varepsilon_y}+\frac{\varepsilon_x\varepsilon_y}{1-\varepsilon_y}$$

2 Анализ исходного ряда

Погрешность метода заключается в том, что для достижения требуемой точности нам достаточно рассмотреть только m первых членов ряда, так как сумма оставшихся членов будет не превзойдет точности. Оценим сверху остаточную сумму ряда:

$$err_m(m) = \sum_{n=m+1}^{\infty}\frac{1}{0.2n^2+1.3n+1.7} \leq \sum_{n=m+1}^{\infty}\frac{1}{0.2n^2} \leq 5\sum_{n=m+1}^{\infty}\frac{1}{n^2} \leq 5\int_{n=m}^{\infty}\frac{dx}{x^2} = 5(-\frac{1}{x})\Big|_m^{\infty} = \frac{5}{m}$$

Пусть числа хранятся с относительной погрешностью ε, посчитаем относительную погрешность вычисления одного слагаемого, обозначим относительную погрешность величины x как δx. Заметим, что целое число, умещающееся в размер мантиссы, имеет нулевую погрешность, как, например, 1 или n.

$$\delta(0.2\cdot n) = \varepsilon+\varepsilon = 2\varepsilon$$

$$\delta((0.2\cdot n)+1.3) = 2\varepsilon+\varepsilon = 3\varepsilon$$

$$\delta(n\cdot(0.2n+1.3)) = 4\varepsilon+\varepsilon = 5\varepsilon$$

$$\delta((n(0.2n+1.3))+1.7) = 5\varepsilon+\varepsilon = 6\varepsilon$$

$$\delta\left(\frac{1}{n(0.2n+1.3)+1.7}\right) = \frac{6\varepsilon}{1-6\varepsilon}+\varepsilon$$

Если мы складываем m слагаемых, то будет еще $m-1$ сложений, каждое из которых добавит к относительной погрешности еще столько же ε. Итоговая вычислительная погрешность, учитывая относительный характер погрешностей, которые мы оценивали и тот факт, что итоговая сумма не превосходит 2, будет выглядеть таким образом:

$$err_c(m) = 2\left(m\varepsilon+\frac{6\varepsilon}{1-6\varepsilon}\right)$$

Теперь общую погрешность можно оценить как

$$err(m) \leq err_c(m)+err_m(m) = \frac{5}{m}+2m\varepsilon+\frac{12\varepsilon}{1-6\varepsilon} \leq 0.5\cdot10^{-4}$$

$$\frac{5}{m}+10m10^{-t}+\frac{60}{10^t-30} \leq 5\cdot10^{-5}$$

$$\frac{1}{m}+2m10^{-t}+\frac{12}{10^t-30} \leq \cdot10^{-5}$$

Минимальный размер мантиссы, при котором решение все-таки существует $t = 11$, в таком случае нам необходимо $m \approx 1.1 \cdot 10^5$ слагаемых для достижения необходимой точности. В условиях двойной точности, однако, согласной этой формуле, достаточно 10^5+1 слагаемых. Максимальная точность, которую можно достигнуть — $0.5 \cdot 10^{-6}$, для этого необходимо $m \approx 1.1 \cdot 10^8$ слагаемых.

3 Улучшение исходного ряда

3.1 Улучшение рядом $\sum_{n=1}^{\infty} \frac{1}{n^2}$

Оценим погрешность метода. Для этого заметим для начала.

$$\sum_{n=1}^{\infty} \frac{1}{0.2n^2 + 1.3n + 1.7} = \sum_{n=1}^{\infty} \frac{5}{n^2 + 6.5n + 8.5} = \sum_{n=1}^{\infty} \frac{5}{n^2 + 6.5n + 8.5} + \frac{5}{n^2} - \frac{5}{n^2} =$$

$$5\sum_{n=1}^{\infty} \frac{1}{n^2} + 5\sum_{n=1}^{\infty} \frac{1}{n^2 + 6.5n + 8.5} - \frac{1}{n^2} = \frac{5\pi^2}{6} + 5\sum_{n=1}^{\infty} \frac{n^2 - (n^2 + 6.5n + 8.5)}{n^2(n^2 + 6.5n + 8.5)} =$$

$$\frac{5\pi^2}{6} - 5\sum_{n=1}^{\infty} \frac{6.5n + 8.5}{n^2(n^2 + 6.5n + 8.5)} = \frac{5\pi^2}{6} - \frac{5 \cdot 13}{2} \sum_{n=1}^{\infty} \frac{n + \frac{17}{13}}{n^2(n^2 + 6.5n + 8.5)}$$

Оценим остаток суммы ряда, если мы взяли только m слагаемых.

$$\sum_{n=m+1}^{\infty} \frac{n + \frac{17}{13}}{n^2(n^2 + 6.5n + 8.5)} \le \sum_{n=m+1}^{\infty} \frac{n + \frac{17}{13}}{n^4} = \sum_{n=m+1}^{\infty} \frac{1}{n^3} + \frac{17}{13n^4} \le \int_m^{\infty} \left(\frac{1}{x^3} + \frac{17}{13x^4} \right) dx$$

$$\int_m^{\infty} \left(\frac{1}{x^3} + \frac{17}{13x^4} \right) dx = \left(-\frac{1}{2x^2} - \frac{17}{39x^3} \right) \Big|_m^{\infty} = \frac{1}{2m^2} + \frac{17}{39m^3}$$

$$err_m(m) = \left| -\frac{65}{2} \sum_{n=m+1}^{\infty} \frac{n + \frac{17}{13}}{n^2(n^2 + 6.5n + 8.5)} \right| \le \frac{65}{4m^2} + \frac{17 \cdot 65}{78m^3}$$

Можно организовать вычисления таким образом, чтобы позже показать, что

$$\delta\left(\frac{n + \frac{17}{13}}{n^2(n^2 + 6.5n + 8.5)} \right) \le \frac{7\varepsilon}{1 - 7\varepsilon} + \varepsilon$$

По старой схеме вычислим ошибку

$$err(m) \le err_m(m) + err_c(m) \le \frac{65}{4m^2} + \frac{17 \cdot 65}{78m^3} + 2m\varepsilon + \frac{14\varepsilon}{1 - 7\varepsilon} \le 0.5 \cdot 10^{-4}$$

$$\frac{13}{4m^2} + \frac{17 \cdot 13}{78m^3} + 2m10^{-t} + \frac{14}{10^t - 35} \le 10^{-5}$$

Минимальная точность, которой достаточно, чтобы это уравнение имело решение — 9 знаков в мантиссе. В этом случае необходимо 610 слагаемых. В случае двойной точности достаточно 570 слагаемых.

3.2 Улучшение рядом $\sum_{n=1}^{\infty} \frac{1}{n^3}$

Улучшим далее наши оценки следующим образом

$$\frac{5}{2}\sum_{n=1}^{\infty}\left[\frac{13n+17}{n^2(n^2+6.5n+8.5)} - \frac{13}{n^3} + \frac{13}{n^3}\right] =$$

$$\frac{5\cdot 13}{2}\sum_{n=1}^{\infty}\frac{1}{n^3} + \frac{5}{2}\sum_{n=1}^{\infty}\frac{\cancel{13n^4} + 17n^3 - \cancel{13n^4} - 13\cdot 6.5n^3 - 13\cdot 8.5n^2}{n^5(n^2+6.5n+8.5)} =$$

$$\frac{5\cdot 13}{2}\zeta(3) + \frac{5}{4}\sum_{n=1}^{\infty}\frac{34n^3 - 13\cdot 13n^3 - 13\cdot 17n^2}{n^5(n^2+6.5n+8.5)} =$$

$$\frac{5\cdot 13}{2}\zeta(3) - \frac{5}{4}\sum_{n=1}^{\infty}\frac{135n^3 + 221n^2}{n^5(n^2+6.5n+8.5)}$$

Где $\zeta(3)$ — дзета-функция Рейманна.

Аналогичным образом посчитаем погрешность метода через остаток ряда

$$err_m(m) = \left| -\frac{5}{4}\sum_{n=m+1}^{\infty}\frac{135n^3+221n^2}{n^5(n^2+6.5n+8.5)}\right| \le \frac{5\cdot 135\cdot 2}{4}\sum_{n=m+1}^{\infty}\frac{1}{n^4} \le \frac{675}{6m^3} \le 0.5\cdot 10^{-4}$$

$$m \ge \sqrt[3]{\frac{135\cdot 10^5}{6}} \approx 131$$

3.3 Улучшение рядом $\sum_{n=1}^{\infty} \frac{1}{n^4}$

Улучшим далее наши оценки следующим образом

$$\frac{5}{4}\sum_{n=m+1}^{\infty}\left[\frac{135n^3+221n^2}{n^5(n^2+6.5n+8.5)} - \frac{135}{n^4} + \frac{135}{n^4}\right] =$$

$$\frac{5\cdot 135}{4}\sum_{n=1}^{\infty}\frac{1}{n^4} + \frac{5}{8}\sum_{n=1}^{\infty}\frac{\cancel{270n^7} + 442n^6 - \cancel{270n^7} - 135\cdot 13n^6 - 135\cdot 17n^5}{n^9(n^2+6.5n+8.5)} =$$

$$\frac{5\cdot 135}{4}\frac{\pi^4}{90} - \frac{5}{8}\sum_{n=1}^{\infty}\frac{1313n^6 + 2295n^5}{n^9(n^2+6.5n+8.5)} =$$

$$\frac{15\pi^4}{8} - \frac{5}{8}\sum_{n=1}^{\infty}\frac{1313n^6 + 2295n^5}{n^9(n^2+6.5n+8.5)}$$

Аналогичным образом посчитаем погрешность метода через остаток ряда

$$err_m(m) = \left| -\frac{5}{8}\sum_{n=m+1}^{\infty}\frac{1313n^6+2295n^5}{n^9(n^2+6.5n+8.5)}\right| \le \frac{5\cdot 1313\cdot 2}{8}\sum_{n=m+1}^{\infty}\frac{1}{n^5} \le \frac{5\cdot 1313\cdot 2}{8\cdot 4\cdot m^4} = \frac{5\cdot 1313}{16m^4} \le 0.5\cdot 10^{-4}$$

$$m \ge \sqrt[4]{\frac{1313\cdot 10^5}{16}} \approx 54$$

4 Вычислительный эксперимент

Кол-во слагаемых	Исходный ряд	Улучшение (1)	Улучшение (2)	Улучшение (3)
20	1.1837195539	1.4275736686	1.3889291796	1.3954506421
100	1.3462762433	1.3960270766	1.3944182454	1.3944736572
1000	1.3894903965	1.3944878974	1.3944716636	1.3944717198
рассчитанное	1.3944217214	1.3944965344	1.3944476502	1.3944934544

5 Выводы

Практические результаты сошлись с аналитическими расчетам, однако оказалось, что при увеличении количества слагаемых в исходном ряду точность не сильно увеличивается. Скорее всего, ошибка вошла в расчеты, связанные с точностью вычисления отдельного слагаемого в сумме. Дело в том, что для того, чтобы вычислить очередной член ряда, необходимо сложить достаточно большие числа, значительно потеряв в точности, прежде чем брать обратное к этому числу. В данном случае вольно сделанные допущения не сделали результат в корне неверными, однако такое могло случится.