Math-Net.Ru

Общероссийский математический портал

Ю. В. Матиясевич, Альтернативы формуле Эйлера–Маклорена для вычисления бесконечных сумм, *Матем. заметки*, 2010, том 88, выпуск 4, 543–548

DOI: <https://doi.org/10.4213/mzm8689>

Использование Общероссийского математического портала Math-Net.Ru подразумевает, что вы прочитали и согласны с пользовательским соглашением <http://www.mathnet.ru/rus/agreement>

Параметры загрузки:

IP: 52.90.164.192

26 апреля 2023 г., 14:46:45

Альтернативы формуле Эйлера–Маклорена для вычисления бесконечных сумм

Ю. В. Матиясевич

Предлагается новый класс формул, которые подобно формуле Эйлера–Маклорена могут быть использованы для вычисления приближенного значения бесконечной суммы путем ее аппроксимации соответствующим интегралом.

Библиография: 4 названия.

1. Введение. Одним из эффективных средств численного анализа является формула Эйлера–Маклорена, связывающая значение суммы сходящегося ряда

$$\sum_{k=1}^{\infty} f(k) \quad (1)$$

со значением определенного интеграла

$$\int_1^{\infty} f(s) ds. \quad (2)$$

Формула была введена независимо Эйлером [1] и Маклореном [2], причем для противоположных целей – у Эйлера это было средство вычислить сумму (1) посредством интеграла (2), в то время как Маклорен вычислял интеграл (2), приближая его суммой (1). В настоящей работе предлагаются другие формулы, которые могут быть использованы для целей Эйлера, т.е. как средство приближенного вычисления суммы (1). Соответственно, мы будем предполагать, что в нашем распоряжении есть эффективный способ вычислять значение интеграла

$$F(t) = \int_t^{\infty} f(s) ds. \quad (3)$$

при любом t с заданной точностью. Это имеет место, например, в случае, когда для F удалось найти явное выражение в виде суперпозиции эффективно вычисляемых функций.

Формула Эйлера–Маклорена может быть формально записана как

$$\sum_{k=1}^{\infty} f(k) = \frac{1}{2}f(1) + F(1) - \sum_{n=1}^{\infty} \frac{B_{2n}}{(2n)!} f^{(2n-1)}(1), \quad (4)$$

где $B_2 = 1/6$, $B_4 = -1/30$, ... – числа Бернулли. Непосредственно эта формула не может быть использована для вычисления значения суммы с произвольной точностью, поскольку

$$|B_{2n}| \approx \frac{2(2n)!}{(2\pi)^{2n}}$$

и, соответственно, ряд в правой части (4) расходится, если только f не стремится к нулю чрезвычайно быстро. По этой причине этот ряд приходится обрывать:

$$\sum_{k=1}^{\infty} f(k) \approx \frac{1}{2} f(1) + F(1) - \sum_{n=1}^N \frac{B_{2n}}{(2n)!} f^{(2n-1)}(1), \quad (5)$$

при этом погрешность и ошибка при этом типично имеет порядок первого отброшенного члена, т.е. примерно равна

$$\frac{B_{2N+2}}{(2N+2)!} f^{(2N+1)}(1).$$

Чтобы сделать эту ошибку произвольно малой, обычно достаточно просуммировать начальный отрезок ряда (1) явно и применить (5) к остатку:

$$\sum_{k=1}^{\infty} f(k) \approx \sum_{k=1}^L f(k) + \frac{1}{2} f(L+1) + F(L+1) - \sum_{n=1}^N \frac{B_{2n}}{(2n)!} f^{(2n-1)}(L+1). \quad (6)$$

Одна из трудностей в практическом применении формулы (6) для вычисления суммы (1) с большой точностью состоит в необходимости вычислять производные функции f больших порядков. Если эта функция задана, скажем, в виде суперпозиции трех-четырех пусть даже простых функций, количество слагаемых в выражении ее n -й производной очень резко увеличивается с ростом n . Предлагаемые в качестве альтернативы формулы для приближенного вычисления суммы (1) не содержат производных и выглядят следующим образом:

$$\sum_{k=1}^{\infty} f(k) \approx \sum_{k=1}^K f(k) + \sum_{m=1}^M A_m f(K+m) + \sum_{m=1}^M B_m F(K+m). \quad (7)$$

Ниже указаны два способа выбора численных коэффициентов A_m и B_m (для заданного значения M).

2. Первый способ. Возьмем произвольную квадратурную формулу, имеющую $M+1$ равноотстоящих узлов

$$\sum_{m=0}^M Q_m g(m) = \int_0^M g(s) ds + c_M g^{(L)}(\tilde{s}), \quad 0 \leq \tilde{s} \leq M, \quad (8)$$

где $L > 0$ (мы не требуем, чтобы L имело максимально возможное значение, т.е. чтобы формула была точна в случае, когда f – многочлен степени M). Полагая $g(s) \equiv 1$ в (8), получаем, что коэффициенты Q_m должны удовлетворять условию

$$\sum_{m=0}^M Q_m = M. \quad (9)$$

Подставим $g(s) = f(s + j)$ в (8) и просуммируем это равенство по $j = K + 1, K + 2, \dots$. В левой части мы, учитывая (9), получим

$$\sum_{j=K+1}^{\infty} \sum_{m=0}^M Q_m f(m+j) = \sum_{m=1}^M \left(\sum_{i=0}^{m-1} Q_i \right) f(K+m) + M \sum_{j=M+1}^{\infty} f(K+j), \quad (10)$$

а в правой –

$$\sum_{j=K+1}^{\infty} \left(\int_0^M f(s+j) ds + c_M f^{(L)}(\tilde{s}_j) \right) = \sum_{m=1}^M F(K+m) + c_M \sum_{j=K+1}^{\infty} f^{(L)}(\tilde{s}_j), \quad (11)$$

где $j \leq \tilde{s}_j \leq j + M$. Поделив (10) на M , мы получим величину, которая отличается от требуемой бесконечной суммы (1) на конечную сумму

$$\sum_{k=1}^K f(k) + \sum_{m=1}^M A_m f(K+m), \quad (12)$$

где

$$A_m = 1 - \frac{1}{M} \sum_{i=0}^{m-1} Q_i = \frac{1}{M} \sum_{i=m}^M Q_i, \quad m = 1, \dots, M. \quad (13)$$

Соответственно, при

$$B_1 = \dots = B_m = \frac{1}{M} \quad (14)$$

левая и правая части (7) отличаются на величину

$$\frac{c_M}{M} \sum_{j=K+1}^{\infty} f^{(L)}(\tilde{s}_j), \quad (15)$$

которая по модулю примерно равна

$$\frac{c_M}{M} \left| \int_{K+1+M/2}^{\infty} f^{(L)}(s) ds \right| \approx \frac{c_M}{M} \left| f^{(L-1)} \left(K + 1 + \frac{M}{2} \right) \right|. \quad (16)$$

ПРИМЕР 1. При $M = 1$ формула (8) – это формула трапеций с $Q_0 = Q_1 = 1/2$, а получающаяся из (7) формула совпадает с формулой Эйлера–Маклорена (4) при $N = 0$.

ПРИМЕР 2. При $M = 2$ формула (8) – это формула Симсона

$$\frac{1}{3} g(0) + \frac{4}{3} g(1) + \frac{1}{3} g(2) = \int_0^2 g(s) ds + \frac{1}{90} g'''(\tilde{s}), \quad 0 \leq \tilde{s} \leq 2. \quad (8_{M=2})$$

При, скажем, $K = 5$ полагаем, как указано выше, $j = 6, 7, \dots$, соответственно,

$$g(s) = f(s+6), \quad g(s) = f(s+7), \quad \dots,$$

и суммируем левые и правые части (8_{M=2}). Для левых частей получаем величину

$$\begin{aligned} & \frac{1}{3} f(6) + \frac{4}{3} f(7) + \frac{1}{3} f(8) \\ & + \frac{1}{3} f(7) + \frac{4}{3} f(8) + \frac{1}{3} f(9) \end{aligned}$$

$$\begin{aligned}
& + \frac{1}{3} f(8) + \frac{4}{3} f(9) + \dots \\
& \quad + \frac{1}{3} f(9) + \dots = \\
& \frac{1}{3} f(6) + \frac{5}{3} f(7) + 2f(8) + 2f(9) + \dots, \quad (10_{M=2, K=5})
\end{aligned}$$

половине (т.е. $1/M$ -й части) которой недостает до искомой бесконечной суммы (1) конечного числа слагаемых, а именно

$$f(1) + f(2) + f(3) + f(4) + f(5) + \frac{5}{6} f(6) + \frac{1}{6} f(7). \quad (12_{M=2, K=5})$$

Соответственно, согласно (13) полагаем $A_1 = 5/6$, $A_2 = 1/6$.

Суммирование правой части (8_{M=2}) дает, после изменения пределов интегрирования, величину

$$\begin{aligned}
& \int_6^8 f(s) ds + \int_8^{10} f(s) ds + \int_{10}^{12} f(s) ds + \dots \\
& \quad + \int_7^9 f(s) ds + \int_9^{11} f(s) ds + \int_{11}^{13} f(s) ds + \dots \\
& \quad + \frac{1}{90} (f''''(\tilde{s}_6) + f''''(\tilde{s}_7) + f''''(\tilde{s}_8) + f''''(\tilde{s}_9) + \dots) \\
& = F(6) + F(7) + \frac{1}{90} (f''''(\tilde{s}_6) + f''''(\tilde{s}_7) + f''''(\tilde{s}_8) + f''''(\tilde{s}_9) + \dots) \quad (11_{M=2, K=5})
\end{aligned}$$

Соответственно, согласно (14) полагаем $B_1 = 1/2$, $B_2 = 1/2$, и при выбранных значениях A_1 , A_2 , B_1 , B_2 левая и правая части (7) отличаются на величину

$$\frac{1}{180} \sum_{j=6}^{\infty} f''''(\tilde{s}_j). \quad (15_{M=2, K=5})$$

3. Второй способ. Описанное выше применение квадратурных формул дает формулы вида (7) только с равными значениями B_m . Это ограничение снимается при приводимом ниже способе выбора коэффициентов в (7), что позволяет повысить порядок производной, от которой зависит оценка погрешности.

На этот раз мы будем исходить из формулы Эйлера–Маклорена (4) при $N < M$ и $L = K + M$. Согласно (4) и (7) должно выполняться примерное равенство

$$\begin{aligned}
& \sum_{m=1}^M A_m f(K+m) + \sum_{m=1}^M B_m F(K+m) \\
& \approx \sum_{k=K+1}^{K+M} f(k) + \frac{1}{2} f(K+M+1) + F(K+M+1) - \sum_{n=1}^N \frac{B_{2n}}{(2n)!} f^{(2n-1)}(K+M+1). \quad (17)
\end{aligned}$$

Предположим, что мы нашли числа $A_1, \dots, A_M$, $B_1, \dots, B_M$, а также многочлены $P_1(t), \dots, P_m(t)$ некоторой степени D , где $D \geq 2N$, такие, что для произвольной

функции $g(s)$ выполняется тождество

$$\begin{aligned} \sum_{m=1}^M A_m g(m) + \sum_{m=1}^M B_m G(m) + \sum_{m=1}^M \int_0^1 P_m(t) g^{(D)}(t+m) dt \\ = \sum_{k=1}^M g(k) + \frac{1}{2} g(M+1) + G(M+1) - \sum_{n=1}^N \frac{B_{2n}}{(2n)!} g^{(2n-1)}(M+1) \end{aligned} \quad (18)$$

где, аналогично (3),

$$G(t) = \int_t^\infty g(s) ds. \quad (19)$$

Подставляя $g(s) = f(K+s)$ в (18) и добавляя к обеим частям этого равенства сумму $\sum_{k=1}^K f(k)$, мы видим, что при этих значениях $A_1, \dots, A_M, B_1, \dots, B_M$ приближение, даваемое формулой (7), отличается от классического приближения (4) с $L = K+M$ на величину

$$\sum_{m=1}^M \int_0^1 P_m(t) f^{(D)}(K+t+m) dt,$$

которая оценивается сверху величиной

$$\sum_{m=1}^M \max_{0 \leq t \leq 1} |P_m(t)| \max_{0 \leq t \leq 1} |f^{(D)}(K+m+t)|. \quad (20)$$

Для нахождения чисел $A_1, \dots, A_M, B_1, \dots, B_M$ и многочленов $P_1(t), \dots, P_m(t)$ можно записать последние с неопределенными коэффициентами

$$P_m(t) = p_{m,D} t^D + \dots + p_{m,0},$$

подставить эти выражения в (18) и устранить интегралы путем многократного интегрирования по частям. Приравнявая затем коэффициенты при

$$G(1), \dots, G(M+1), \quad g(1), \dots, g(M+1), \quad \dots, \quad g^{(D-1)}(1), \dots, g^{(D-1)}(M+1)$$

в левой и правой частях, мы получим систему из $(D+1)(M+1)$ уравнений с $(D+3)M$ неизвестными

$$A_1, \dots, A_M, \quad B_1, \dots, B_M, \quad p_{1,0}, \dots, p_{1,D}, \quad \dots, \quad p_{M,0}, \dots, p_{M,D}.$$

При $(D+1)(M+1) \leq (D+3)M$, т.е. когда $D \leq 2M-1$, мы вправе ожидать наличие решения у этой системы, однако выбор $D = 2M-1$ может оказаться неудачным из-за большой величины норм $\|P_1\|_{L^\infty[0,1]}, \dots, \|P_M\|_{L^\infty[0,1]}$, входящих в (20). Взяв меньшее значение D , мы можем найти общее решение построенной линейной системы и затем выбрать значения входящих в него параметров, минимизируя нормы многочленов $P_1(t), \dots, P_m(t)$. С вычислительной точки зрения при этой минимизации проще использовать вместо $L^\infty[0,1]$ норму $L^2[0,1]$ и минимизировать $\|P_1\|_{L^2[0,1]}^2 + \dots + \|P_M\|_{L^2[0,1]}^2$, так как это сводится к решению еще одной линейной системы.

ПРИМЕР 3. При $M = 3$, $D = 3$, $N = 1$ тождество (18) имеет место при

$$A_1 = \frac{23}{24}, \quad A_2 = \frac{7}{12}, \quad A_3 = \frac{1}{12}, \quad B_1 = \frac{7}{48}, \quad B_2 = \frac{7}{12}, \quad B_3 = \frac{13}{48},$$

$$P_1(t) = \frac{7}{288}t^3 - \frac{1}{48}t^2, \quad P_2(t) = \frac{35}{288}t^3 - \frac{5}{32}t^2 + \frac{1}{32}t + \frac{1}{288},$$

$$P_3(t) = \frac{1}{6}t^3 - \frac{1}{4}t^2 + \frac{1}{12}t,$$

при этом

$$\max_{0 \leq t \leq 1} |P_1(t)| = \frac{1}{288} \approx 0.00347, \quad \max_{0 \leq t \leq 1} |P_2(t)| = \frac{5 + 3\sqrt{30}}{2205} \approx 0.00972,$$

$$\max_{0 \leq t \leq 1} |P_3(t)| = \frac{\sqrt{3}}{72} \approx 0.00802.$$

Заключение. Вместо (8) можно было бы использовать и квадратурную формулу “открытого типа”

$$\sum_{m=1}^M Q_m g(m) = \frac{1}{M} \int_{1/2}^{M+1/2} g(s) ds + c'_M g^{(L)}(\tilde{s}), \quad \frac{1}{2} \leq \tilde{s} \leq M + \frac{1}{2}.$$

Формулы (6) и (7) представляют крайние случаи – в первой используются много производных, но вычисляемых в одной точке, во второй формуле производные не участвуют вообще, но требуется вычисление $F(t)$ для многих значений t . Можно ввести и “гибридные” формулы, в которых используются значения многих производных, вычисляемые во многих точках.

Общим для формулы (6) и (7) является использование интеграла (3) в качестве главного члена приближения. Отметим, что имеются способы вычисления суммы (1), не использующие ни производных, ни интеграла в явном виде – см., например, [3].

Автор [4] применял формулы (7), получающиеся из квадратурных формул, для вычисления так называемых констант Стильбеса, участвующих в коэффициентах разложения дзета-функции Римана в полюсе.

СПИСОК ЦИТИРОВАННОЙ ЛИТЕРАТУРЫ

- [1] L. Eulero, “Methodus generalis summandi progressionem”, *Commentarii academiae scientiarum Petropolitanae*, **6** (1738), 68–97, <http://www.math.dartmouth.edu/~euler/docs/originals/E025.pdf>; англ. пер.: <http://www.17centurymaths.com/contents/euler/e025tr.pdf>.
- [2] C. MacLaurin, *A Treatise of Fluxions*. т. 1, 2, T. W. and T. Ruddimans, Edinburgh, 1742, <http://www.archive.org/details/atreatisefluxio00conggooq>.
- [3] H. Monien, *Gaussian Summation: An Exponentially Converging Summation Scheme*, arXiv: [math.NA/0611057](https://arxiv.org/abs/math.NA/0611057).
- [4] Yu. V. Matijasevich, “The Riemann hypothesis from a logician’s point of view”, *Number Theory* (Banff, AB, 1988), de Gruyter, Berlin, 1990, 387–400.

Ю. В. Матиясевич

Санкт-Петербургское отделение

Математического института им. В. А. Стеклова РАН

E-mail: yumat@pdmi.ras.ru

Поступило

07.12.2009

Исправленный вариант

07.07.2010