

УДК 519.6

УЛУЧШЕНИЕ ОЦЕНОК ПОГРЕШНОСТИ ДЛЯ ЭКСПОНЕНЦИАЛЬНО СХОДЯЩИХСЯ КВАДРАТУР

© 2021 г. В. С. Хохлачев^{1, *}, А. А. Белов^{1, 2}, Н. Н. Калиткин³

¹Федеральное государственное бюджетное образовательное учреждение высшего образования “Московский государственный университет имени М.В. Ломоносова”, Физический факультет, Москва, Россия

²Федеральное государственное автономное образовательное учреждение высшего образования “Российский университет дружбы народов”, Москва, Россия

³Федеральное государственное бюджетное учреждение науки “Федеральный исследовательский центр Институт прикладной математики имени М.В. Келдыша Российской академии наук”, Москва, Россия

*E-mail: valentin.mycroft@yandex.ru

Поступила в редакцию 20.07.2020 г.

После доработки 28.08.2020 г.

Принята к публикации 28.09.2020 г.

Для некоторых практически важных классов функций возможно кардинальное увеличение точности и уменьшение трудоемкости квадратур при вычислении интегралов. Разработан соответствующий математический аппарат с новейшими улучшениями, что позволяет в сотни раз и более сократить трудоемкость вычислений. Приводятся примеры физических задач, к которым он хорошо применим.

DOI: 10.31857/S0367676521010166

ВВЕДЕНИЕ

Прикладные задачи

В физических задачах часто требуется численно находить интегралы. Выделим два класса подынтегральных функций. Первый – периодические функции. Такие задачи возникают при разложении в ряды или интегралы Фурье; они являются некорректными [5]. Поэтому для них необходимо вычислять квадратуры с высокой точностью. Второй класс – функции, быстро убывающие на бесконечности. Приведем примеры таких задач.

1) Передача сигнала. Передать сигнал большого объема дорого. Используют следующий подход. Сигнал разлагают в интеграл Фурье, и в памяти хранят дискретный Фурье образ. Его, в отличие от аналогового сигнала, можно программно сжать без потери информации. Сжатый сигнал передают. По нему восстанавливают Фурье образ и воспроизводят исходный сигнал.

2) Функции Ферми–Дирака. Они зависят от индекса k и параметра x :

$$I_k(x) = \int_0^{\infty} \frac{t^k dt}{1 + \exp(t - x)}, \quad x \in (-\infty, +\infty). \quad (1)$$

Подынтегральная функции быстро убывает на бесконечности. Функции (1) были введены в [6, 7] для описания проводимости твердых метал-

лов. Они являются моментами фермиевского распределения электронов. Затем эти функции были использованы для описания термодинамики плотной горячей плазмы [8]. Сейчас они используются во многих областях физики. Наиболее важны функции с полужелыми k , но их трудно вычислять.

3) Плазменное микрополе. Хаотическое движение заряженных частиц в плазме создает в ней поля, флуктуирующие на расстояниях порядка межатомных. Именно этими микрополями определяются многие оптические свойства плазмы [9, 10]. Распределение микрополя часто находят, представляя суммарное действие заряженных частиц интегралом Фурье, при этом Фурье образ корректируют с помощью различных модельных предположений. Первой была работа [11], в которой функция распределения равна

$$H(\beta) = \frac{2}{\pi} \beta \int_0^{\infty} \exp(-x^{3/2}) \sin(\beta x) x dx. \quad (2)$$

4) Скорости реакций. Поведение скоростей термоядерных реакций при низких температурах важно для расчета зажигания дейтериевых мишеней. Для многих ядерных реакций измерены экспериментальные зависимости сечений от скоростей частиц $\sigma(v)$. Тогда зависимость скорости ре-

акции от температуры можно найти, усредняя сечение по распределению Максвелла

$$K(T) = \langle \sigma(v)v \rangle = \frac{\pi}{\sqrt{m}} \left(\frac{2}{\pi T} \right)^{3/2} \int_0^{\infty} \sigma(E) E \times \exp\left\{-\frac{E}{T}\right\} dE, \quad E = \frac{mv^2}{2}. \quad (3)$$

Подынтегральная функция быстро убывает при $v \rightarrow \infty$.

Вычисление квадратур

Для вычисления квадратур подобных типов в последние годы появился математический метод, ускоряющий вычисления в сотни раз. Оказалось, что для указанных классов функций квадратуры на равномерных сетках сходятся не по степенному закону, а по экспоненциальному. Наиболее полно данная проблема изложена в [1]; там же приведен исчерпывающий список литературы. В [2, 3] экспериментально было показано, что теоретические оценки работы [1] справедливы лишь для функций с полюсами первого порядка.

Опишем полученные результаты. 1) Мажорантную оценку работы [1] можно заметно улучшить. 2) Приведенное в [1] доказательство справедливо лишь для функций с полюсами первого порядка. 3) Найдена эмпирическая оценка погрешности для функций с полюсами порядка выше первого. 4) Эмпирически показано, что зависимость от числа узлов в этих оценках является не мажорантной, а асимптотически точной. Для иллюстрации данных результатов найдены новые точно берущиеся интегралы. Приведены примеры интегралов, для которых погрешность квадратурных формул зависит от шага сетки немонотонно и не описывается существующими теориями.

ЭКСПОНЕНЦИАЛЬНО СХОДЯЩИЕСЯ КВАДРАТУРЫ

В многочисленных учебниках по вычислительной математике строятся квадратурные формулы трапеций, средних, Симпсона и т.д. на сетках с шагом $h \sim 1/N$, где N – число узлов сетки. Доказывается, что для достаточно гладких функций погрешность таких квадратур есть $O(h^p) = O(N^{-p})$,

где p есть порядок сходимости. Такую сходимость называют степенной.

В работе [1] была доказана теорема, что для неограниченно дифференцируемой функции интеграл по периоду аппроксимируется формулой трапеций на равномерной сетке не со степенной погрешностью, а с экспоненциальной. Пусть

$$I = \int_0^{2\pi} u(x) dx, \quad I_N = \sum_{n=0}^{N-1} \frac{h}{2} (u_n + u_{n+1}). \quad (4)$$

Здесь $u_n = u(x_n)$, $x_n = nh$, $h = 2\pi/N$ – равномерная сетка. Пусть L есть наименьшее расстояние от всех особых точек функции $u(x)$ на комплексной плоскости до вещественной оси. Тогда справедлива мажорантная оценка

$$\delta = |I - I_N| \leq \frac{2\pi M_1}{\exp(LN) - 1}, \quad M_1 = \max_{0 \leq x \leq 2\pi} |u|. \quad (5)$$

Зависимость погрешности от N имеет необычный вид. Если $LN \ll 1$, то $\delta \sim (LN)^{-1}$, то погрешность степенная с первым порядком точности. Если же $LN \gg 1$, то погрешность $\delta \sim \exp(-LN)$, т.е. убывание погрешности не степенное, а экспоненциальное.

В [1] доказаны аналогичные теоремы для интеграла по окружности на комплексной плоскости и для интеграла по числовой оси от непериодической функции, достаточно быстро убывающей на бесконечности.

Видно, что зависимость от N в (5) является не степенной, а экспоненциальной. Такая сходимость гораздо быстрее степенной и имеет качественно другой характер. Для функций, удовлетворяющих требованиям указанной теоремы, трудоемкость квадратур очень сильно уменьшается. Это указывает на большую практическую ценность данной теоремы.

ПОЛЮСЫ ПЕРВОГО ПОРЯДКА

В работах [2–4] экспоненциально сходящиеся квадратуры применялись к важной практической задаче – построению способов быстрого вычисления функций Ферми–Дирака. Для тщательно исследования квадратур был найден интеграл [12, стр. 383]

$$\int_0^{\pi} \frac{\cos rx dx}{(1 - 2a \cos x + a^2)^q} = \begin{cases} \frac{a^{2q+r-2} \pi}{(1-a^2)^{2q-1}} \sum_{k=0}^{q-1} \binom{q+r-1}{k} \binom{2q-k-2}{q-1} \left(\frac{1-a^2}{a^2}\right)^k, & a^2 < 1; \\ \frac{\pi}{a^r (a^2-1)^{2q-1}} \sum_{k=0}^{q-1} \binom{q+r-1}{k} \binom{2q-k-2}{q-1} (a^2-1)^k, & a^2 > 1. \end{cases} \quad (6)$$

При целочисленных r и q он берется точно в элементарных функциях.

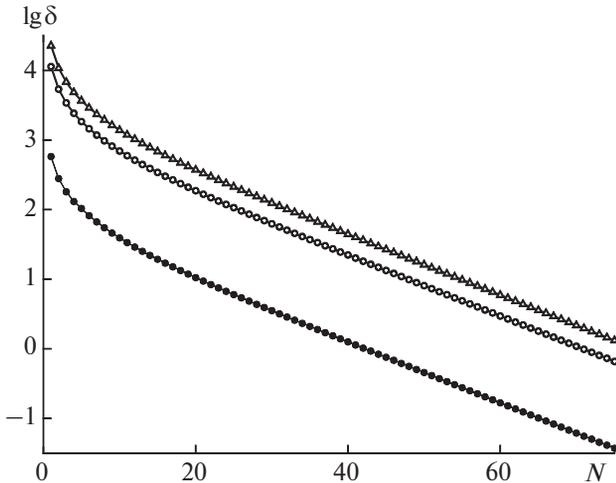


Рис. 1. Погрешность интеграла (6) для $q = 1, r = 10$. Обозначения – см. текст.

Численные расчеты показали, что для $q = 1$ при увеличении N действительно наблюдалась экспоненциальная сходимость. Однако при $q = 2$ наблюдалось отличие сходимости от экспоненциальной. Хотя оно было небольшим, но сходимость оказалась более медленной. Это нарушало теорему [1]. Это противоречие заставило нас более тщательно проанализировать детали работы [1]. Опишем результаты проведенного исследования.

$$M_2 = \frac{1}{2} \sqrt{(\max \operatorname{Re} u(x) - \min \operatorname{Re} u(x))^2 + (\max \operatorname{Im} u(x) + \min \operatorname{Im} u(x))^2}. \tag{9}$$

Для вещественных функций оценка (9) переходит в (7). Заметим, что авторы [1] считали свою оценку M_1 неулучшаемой.

Апробация оценок

Хотя авторы [1] называют свою оценку мажорантной, фактически они получили гораздо более сильный результат. Наши численные расчеты показали, что если особые точки $u(x)$ являются полюсами первого порядка, то указанная в [1] зависимость от N является асимптотически точной. Опишем эти расчеты.

Для теста был выбран интеграл (6) с $q = 1, r = 10$. Все особые точки являются полюсами первого порядка. Они расположены в точках

$$x = \arccos \frac{1+a^2}{2a} + 2\pi k, \quad k = 0, \pm 1, \pm 2, \dots \tag{10}$$

Приведем типичный расчет, в котором выбрано очень малое $L = 0.05$ (ему соответствует $a = e^L \approx 1.051$) и все значения $N = 1, 2, 3, \dots$. Ре-

Мажорирующая константа

Пусть функция постоянна: $u(x) = C$. Очевидно, для такой функции формула трапеций точна. Поэтому погрешность (5) для функции $u(x) - C$ такова же, как для функции $u(x)$. Однако константа M для функции $u(x) - C$ будет другой. Выберем такую константу C , чтобы минимизировать величину M для новой функции. Для вещественной функции $u(x)$ этот выбор очевиден: $C = 0.5(\max u(x) + \min u(x))$. Это дает

$$M_2 = \frac{1}{2}(\max u(x) - \min u(x)). \tag{7}$$

Легко видеть, что всегда $M_2 \leq M_1/2$. Если же $u(x) > 0$, причем $\max u(x)/\min u(x)$ близко к 1, то $M_2 \ll M_1$.

Аналогичную оценку можно построить для комплексной функции $u(x)$. Рассмотрим значения функции только на вещественном отрезке $x \in [0, 2\pi]$. Положим

$$C = \frac{1}{2}(\max \operatorname{Re} u(x) + \min \operatorname{Re} u(x)) + i \frac{1}{2}(\max \operatorname{Im} u(x) + \min \operatorname{Im} u(x)). \tag{8}$$

Тогда получим оценку

результаты показаны на рис. 1 и 2. На рис. 1 показана зависимость логарифма погрешности $\lg \delta$ от N . Фактическая погрешность показана точками, оценка [1] – треугольниками, оценка [1] с константой M_2 – кружками. На начальных участках линий $LN \ll 1$, и эти участки искривлены. При $LN \gg 1$ линии асимптотически переходят в прямые, что соответствует практически экспоненциальной сходимости. Видно, что наше уточнение константы M улучшает теоретическую оценку.

В данном примере выбрано очень малое L , т.е. особая точка расположена очень близко к вещественной оси. Такой пример считается очень трудным для обычных расчетов. Тем не менее, погрешность стремительно убывает при увеличении N . Уже при $N \approx 400$ она уменьшается до $\delta \sim 10^{-16}$. Если же взять $L \sim 1$, то 16 верных знаков (предельная компьютерная точность) будут получены при $N \approx 20$. Это наглядно показывает выгоды экспоненциальной сходимости.

На рис. 2 показано отношение оценки [1] с константой M_2 к фактической погрешности.

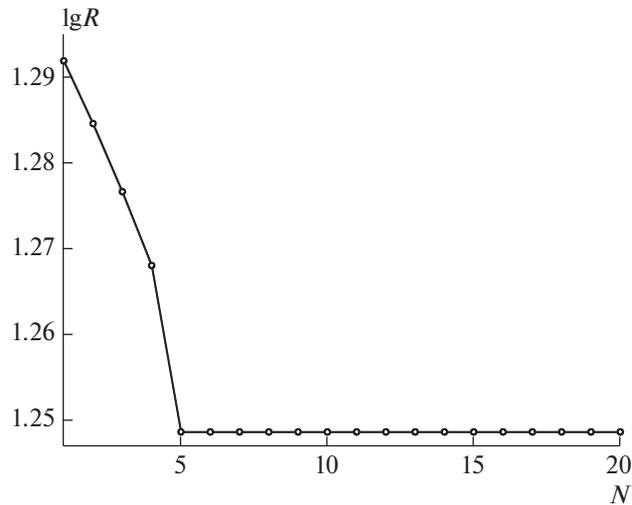


Рис. 2. Расчет интеграла (6) для $q = 1$, $r = 10$. Отношение R оценки (5) с константой M_2 (7) к фактической погрешности.

Видно, что указанное отношение слегка отличается от константы при $LN < 0.5$ и очень быстро стремится к константе при увеличении N . Это убедительно свидетельствует, что зависимость от N в оценке [1] является не просто мажорантной, а асимптотически точной, и вдобавок применимой при очень малых значениях LN . Это показывает высокую ценность оценки [1].

Практические рекомендации

Из практики расчетов известно, что при степенном характере сходимости $\delta \sim O(N^{-p})$ знак реальной погрешности, начиная с некоторого N_0 не меняется при дальнейшем увеличении N . Это является важным свойством, которое позволяет строить оценки точности методом Ричардсона по сгущению сеток.

В данных расчетах мы наблюдали аналогичную картину: погрешность сохраняла свой знак при сгущении сеток. Поэтому здесь также возможно апостериорное определение погрешности по расчетам на сгущающихся сетках. При этом удобно проводить расчеты не при всех N подряд, а только при последовательно удваивающихся N .

Однако сами эти оценки будут другими. Мы используем их при $LN \gg 1$, когда можно считать $\delta_N \approx \text{const} \cdot \exp(-LN)$. В этом случае $\delta_{2N} \sim O(\delta_N^2)$. Это напоминает ньютоновскую сходимость, которую обычно называют квадратичной (зачастую не совсем корректно говорят, что число верных знаков удваивается с удвоением N). Это позволяет сформулировать следующую практическую рекомендацию.

Если требуется получить точность ϵ , то остановим вычисления, как только выполнится условие $|I_N - I_{2N}| < \epsilon^v$, где показатель степени выбирают в пределах $v \approx 0.65-0.75$. Меньший показатель может не дать требуемой точности, больший — к избыточным вычислениям или отсутствию сходимости из-за ошибок округления.

ПОЛЮСЫ ЦЕЛЫХ ПОРЯДКОВ

Обобщение оценки погрешности

Мы провели тщательный анализ доказательства теоремы [1]. Удалось обнаружить, что при доказательстве теоремы сделано неявное предположение о том, что ближайшая особенность является полюсом первого порядка. Поэтому мы экспериментально исследовали, какова будет погрешность для полюсов целого порядка $q > 1$. Для исследования был взят тестовый интеграл (6) при $r = 0$.

Численные расчеты показали, что фактическая погрешность при больших N превышает мажорантную оценку [1]. Поскольку эта оценка мажорантна, то превышение означает, что функциональная зависимость погрешности от N должна быть несколько иной. Более тщательный анализ позволил предложить следующую зависимость погрешности от N :

$$\delta = 2\pi M \frac{(LN/q + \exp(-LN/q))^{q-1}}{\exp(LN) - 1}. \quad (11)$$

Для полюса первого порядка ($q = 1$) она переходит в оценку [1]. Видно, что при $LN \gg 1$ главным членом является экспонента в знаменателе, то

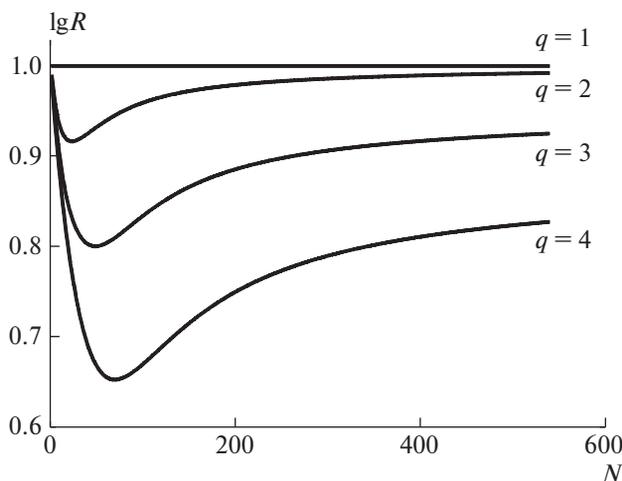


Рис. 3. Расчет интеграла (6) для $r = 0$. Отношение R оценки (11) с константой M_2 (7) к фактической погрешности. Значения q указаны у кривых.

есть погрешность будет близка к экспоненциальной. Наличие множителя $(LN)^{q-1}$ в числителе несколько замедляет скорость сходимости тем сильнее, чем больше q , т.е. чем сильнее особенность. Тем не менее, сходимость очень быстра.

Апробация оценки

Результаты расчетов тестового интеграла показаны на рис. 3. По оси абсцисс, как и на рис. 2, отложен логарифм отношения теоретической оценки (11) к фактической погрешности расчета. Около каждой кривой указана кратность полюса q . Для сравнения приведена кривая для $q = 1$.

Видно, что линия $q = 1$ горизонтальна. Остальные линии – кривые с минимумом, переходящие в горизонтальные линии при увеличении N . Это показывает, что предложенная оценка (11) является асимптотически точной по N . Выход на асимптоту происходит тем медленнее, чем больше q . Из высот горизонтальных асимптот видно, что при $q = 1$ и $q = 2$ теоретическая оценка в ~ 10 раз больше фактической погрешности. При $q = 3$ превышение составляет ~ 8.5 раз; при $q = 4$ оно равно ~ 6.6 раз. Такие превышения показывают, что функциональная зависимость погрешности от N, L, q близка к истинной, но использование константы M_2 приводит к слишком большому превышению (константа M_1 даст еще худшие результаты).

Отметим, что в этих расчетах реальная погрешность сохраняет свой знак при увеличении N , как это было для полюса кратности 1. Поэтому здесь остаются в силе те же практические рекомендации по выбору сеток и критерию окончания расчета для получения заданной точности.

НЕПЕРИОДИЧЕСКИЕ ФУНКЦИИ

Новые тестовые интегралы

В [1] были рассмотрены непериодические функции на бесконечной прямой. Сформулированы условия, при которых соответствующие квадратуры сходятся экспоненциально. Найдены мажорантные оценки их погрешности.

Для верификации этих оценок необходимо иметь тестовые примеры с такими подынтегральными функциями, при которых интегралы выражаются через элементарные функции. Мы построили четыре таких интеграла, значения которых ранее были неизвестны. Приведем их без вывода.

$$I = \int_0^{\infty} \exp\left(-\left(x + \frac{1}{x}\right)^2\right) dx = \frac{\sqrt{\pi}}{2e^4}, \quad (12)$$

$$I = \int_0^{\infty} \left(x + \frac{1}{x}\right)^2 \exp\left(-\left(x + \frac{1}{x}\right)^2\right) dx = \frac{9\sqrt{\pi}}{4e^4}, \quad (13)$$

$$I = \int_0^{\infty} \frac{1}{x^2} \exp\left(-\left(x + \frac{1}{x}\right)^2\right) dx = \frac{\sqrt{\pi}}{2e^4}, \quad (14)$$

$$I = \int_0^{\infty} x^2 \exp\left(-\left(x + \frac{1}{x}\right)^2\right) dx = \frac{3\sqrt{\pi}}{4e^4}. \quad (15)$$

В этих интегралах функция имеет существенно особую точку при $x = 0$ (то есть эта точка совпадает с концом отрезка интегрирования), а также бесконечно удаленную особую точку. Интегралы с такими особенностями возникают, например, при вычислении скоростей химических или ядерных реакций по их сечению в термодинамически равновесной среде.

Этот случай не подпадает под теорему [1], поскольку в ней предполагалось, что особая точка

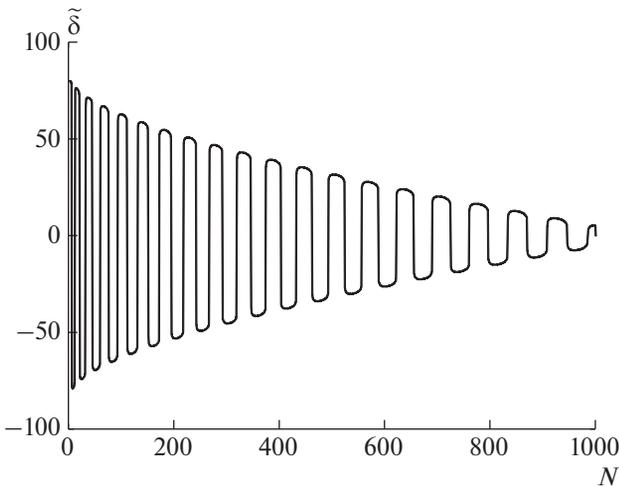


Рис. 4. Погрешность интеграла (15).

не лежит на отрезке интегрирования (хотя расстояние до отрезка интегрирования может быть сколь угодно малым). В связи с отсутствием теоретических оценок численный расчет таких примеров становится особенно интересным.

Пример расчета

Для интеграла (15) были проведены численные расчеты для сеток с $N = 1, 2, \dots, 1000$. Было найдено точное значение погрешности $\delta = |I - I_N|$. Оказалось, что зависимость этой величины от N кардинально отличается от ранее рассмотренных случаев. Эта зависимость знакопеременна и затухает по амплитуде.

Для передачи такой зависимости был найден специфический масштаб. По оси ординат откладывалась величина

$$\tilde{\delta} = \operatorname{arcsch} \frac{\delta}{\mu} = \operatorname{sgn}(\delta) \ln \left(\frac{|\delta|}{\mu} + \sqrt{\frac{\delta^2}{\mu^2} + 1} \right); \quad (16)$$

здесь μ — величина наименьшего из экстремумов. На рис. 4 показана погрешность в данном масштабе. Такой масштаб линейно изображает погрешность в малой окрестности нулевых значений, но логарифмически уменьшает ее вдали от нулей.

Такой вид погрешности, во-первых, не описывается ни одной из известных теорий. Во-вторых, к нему не применимы все известные методы апостериорного нахождения погрешности расчета. Видно, что подобные задачи требуют развития новых теоретических подходов.

Бесконечно удаленная особенность

Если подынтегральная функция имеет только бесконечно удаленную особую точку, то это должно быть антиподом только что рассмотрен-

ного примера. Следует ожидать очень быстрой сходимости квадратур. Рассмотрим следующий тестовый пример:

$$I = \int_0^{\infty} x^2 \exp(-x^2) dx = \frac{\sqrt{\pi}}{4}. \quad (17)$$

Этот пример также выходит за рамки теоремы [1]. Расчет показывает, что сходимость квадратурных формул оказывается качественно быстрее экспоненциальной. Она подчиняется закону $\delta_{2N} \sim O(\delta_N^4)$.

Особая точка подынтегральной функции (17) является существенно особой. Тем не менее, сходимость оказывается исключительно быстрой. Это показывает, что характер сходимости определяется не столько типом особенности, сколько ее расположением. Чем дальше особая точка находится от отрезка интегрирования, тем сходимость быстрее. Расположение особой точки на отрезке кардинально ухудшает сходимость.

ЗАКЛЮЧЕНИЕ

Изложенный метод является мощным инструментом для решения физических задач. Если удастся найти преобразования переменных, сводящие интегралы к указанным выше видам, то вычисления ускорятся в сотни раз.

Проведенные исследования позволили: 1) улучшить теоретические оценки погрешности экспоненциально сходящихся квадратур; 2) найти оценку погрешности для случая кратных полюсов подынтегральной функции; 3) установить, что попадание особой точки на отрезок интегрирования качественно меняет поведение погрешности, что не объясняется известными теориями.

Полученные результаты применены к вычислению функций Ферми–Дирака полуцелого индекса [2–4]. Замена переменных $t = \tau^2$ в (1) приводит интеграл к требуемой форме. Квадратура трапеций обеспечивает 16 верных знаков уже при $N \sim 10$ –100, а традиционные квадратуры требуют $N \sim 10^4$.

Работа поддержана РНФ (проект № 16-11-10001-П).

СПИСОК ЛИТЕРАТУРЫ

1. Trefethen L.N., Weideman J.A.C. // SIAM Rev. 2014. V. 56. No 3. P. 385.
2. Kalitkin N.N., Kolganov S.A. // Doklady Math. 2017. V. 95. No 2. P. 157.
3. Kalitkin N.N., Kolganov S.A. // Math. Models Comp. Simul. 2017. V. 9. No 5. P. 554.
4. Kalitkin N.N., Kolganov S.A. // Math. Models Comp. Simul. 201. V. 10. No 4. P. 472.
5. Тихонов А.Н., Арсенин В.Я. Методы решения некорректных задач. М.: Наука, 1979.
6. Pauli W. // Z. Phys. 1927. V. 41. P. 81.

7. *Sommerfeld A.* // *Z. Phys.* 1928. V. 47. P. 1.
8. *Feynman R.P., Metropolis N., Teller E.* // *Phys. Rev.* 1949. V. 75. P. 1561.
9. *Собельман И.И.* Введение в теорию атомных спектров. М.: Физматлит, 1963.
10. *Грим Г.* Уширение спектральных линий в плазме. М.: Мир, 1978.
11. *Holtmark J.* // *Ann. Phys.* 1919. V. 58. P. 577.
12. *Градиштейн И.С., Рыжик И.М.* Таблицы интегралов, сумм, рядов и произведений. М.: Физматиз, 1963.

Improvement of error estimates for exponentially convergent quadratures

V. S. Khokhlachev^{a,*}, A. A. Belov^{a,b}, N. N. Kalitkin^c

^a*Lomonosov Moscow State University, Moscow, 119991 Russia*

^b*Peoples' Friendship University of Russia, Moscow, 117198 Russia*

^c*Keldysh Institute of Applied Mathematics RAS, Moscow, 125047 Russia*

^{*}*e-mail: valentin.mycroft@yandex.ru*

For some practically important classes of functions, a dramatic increase in accuracy and a decrease in the complexity of quadratures are arise when calculating the integrals. A corresponding mathematical apparatus with the latest improvements has been developed, which allows to reduce by hundreds of times and more the complexity of calculations. Physical examples of such issues are given.