ДИФФЕРЕНЦИАЛЬНЫЕ УРАВНЕНИЯ, 2019, том 55, № 7, с. 907–918

ЧИСЛЕННЫЕ МЕТОДЫ

УДК 519.622

ЭКОНОМИЧНЫЕ МЕТОДЫ ЧИСЛЕННОГО ИНТЕГРИРОВАНИЯ ЗАДАЧИ КОШИ ДЛЯ ЖЁСТКИХ СИСТЕМ ОБЫКНОВЕННЫХ ДИФФЕРЕНЦИАЛЬНЫХ УРАВНЕНИЙ

© 2019 г. А. А. Белов, Н. Н. Калиткин

Уточняется понятие жёсткости системы обыкновенных дифференциальных уравнений. Указаны основные трудности, возникающие при решении задачи Коши для жёстких систем. Показаны преимущества перехода к новому аргументу – длине дуги интегральной кривой. Обсуждаются различные критерии выбора шага и рекомендуется использовать критерий кривизны интегральной кривой. Указаны наиболее надёжные неявные и явные схемы, пригодные для решения жёстких задач. Изложена стратегия расчёта, позволяющая одновременно с решением вычислить асимптотически точное значение погрешности численного решения. В качестве иллюстрации приведён расчёт химической кинетики горения водорода в кислороде с учётом 9 компонент и 50 реакций между ними.

DOI: 10.1134/S037406411907001X

Введение. До конца 40-х гг. прошлого столетия задачи Коши для систем обыкновенных дифференциальных уравнений (ОДУ) успешно решались явными численными методами Адамса, Рунге–Кутты и др. Однако в конце 40-х гг. появился ряд новых прикладных задач (например, химическая кинетика горения ракетного топлива), на которых явные схемы требовали неприемлемо малого шага. Такие задачи получили название жёстких, и для их решения потребовалась разработка новых численных методов. Этим методам посвящена обширная литература, обзор которой дан в классической монографии [1]. Однако в этой монографии, во-первых, не дано рекомендаций, какие из методов лучше использовать в том или ином случае. Во-вторых, имеется немало задач, для которых ни один из методов, описанных в [1], не позволяет найти их решение, “срываясь” задолго до конечной точки. В-третьих, даже если расчёт завершён, остаётся открытым вопрос о фактически достигнутой точности. Решению этих проблем посвящена данная работа.

Точного отражающего существо вопроса математического определения жёсткости систем ОДУ пока не предложено. Предлагались формальные определения жёсткости системы по спектру матрицы Якоби её правой части: если все спектральные числа отрицательны и среди них есть большие по модулю, то задачу относили к жёстким. При этом неявно предполагалось, что для линейных систем, удовлетворяющих этому условию, все компоненты решения затухают в соответствии с величинами спектральных чисел. Последнее, однако, опровергается примером Винограда [2; 3, с. 123–126] линейной неавтономной системы ОДУ, все собственные значения которой отрицательны и постоянны (не зависят от аргумента t), а решение имеет экспоненциально нарастающую компоненту. Очевидно, что для нелинейных задач строго определить понятие жёсткости ещё труднее.

Разумнее определять жёсткость как качественное свойство, описывающее структуру решения. Ю.В. Ракитский [4] определял жёсткость как наличие у решения хотя бы одной такой компоненты, скорости изменения которой сильно отличаются между собой (от очень быстрых до медленных). Такое понимание жёсткости концептуально близко к теории пограничного слоя, разработанной в работах А.Н. Тихонова и его учеников (см., например, [5–7]).

Таким образом, решение жёсткой задачи имеет участки быстрого изменения, которые называются *пограничными слоями*, и участки плавного изменения, называемые *регулярными участками*. Это хорошо иллюстрируется на простейшем примере уравнения

$$\frac{du}{dt} = -\lambda u, \tag{1}$$

где $\lambda \gg 1$ (см. рис. 1). В работе [8] нами предложено выделять ещё один участок решения – *переходную зону*: это участок перехода пограничного слоя в регулярное решение. Он характеризуется большой кривизной кривой $u(t)$. Строго говоря, термин "пограничный слой" относится к быстрому изменению решения только в начальный момент времени $t = 0$. Если быстрое изменение происходит в момент $t > 0$, то его называют *внутренним пограничным слоем* или *контрастной структурой*.

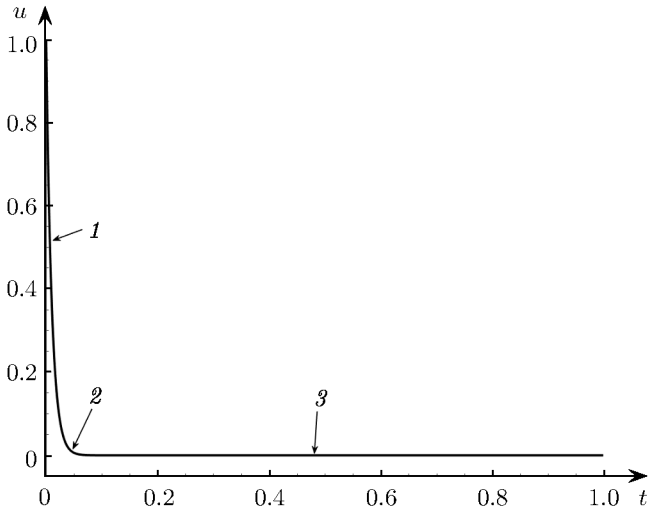

Рис. 1. Решение задачи (1) в аргументе t: *1* – пограничный слой, *2* – переходная зона, *3* – регулярный участок.

Очевидно, что численный расчёт жёстких задач сопряжён со значительными трудностями.

1. Пограничный слой требует очень мелкого шага.

2. Не всякая схема даже при хорошем выборе шага позволяет считать без срывов (т.е. позволяет выполнить расчёт до конца, давая при этом хотя бы внешне правдоподобные результаты). Нужно выделить классы достаточно надёжных и одновременно точных схем.

3. Недостаточно только провести расчёт до заданного момента. Нужно ещё уметь надёжно оценить достигнутую точность. Теоретические мажорантные оценки точности при этом неэффективны: они используют значения высоких производных решения, которые априори неизвестны и для жёстких задач очень велики. Поэтому особое значение приобретают асимптотически точные вычисления погрешности, проводимые одновременно с нахождением самого решения.

1. Переход к длине дуги. Существует общий приём, позволяющий существенно облегчить решение жёстких задач. Он заключается в переходе к новому аргументу – длине дуги интегральной кривой.

Для системы ОДУ порядка M рассмотрим исходную задачу Коши

$$\frac{du_m}{dt} = f_m(t, u_1, u_2, \dots, u_M), \quad 1 \leqslant m \leqslant M, \quad 0 \leqslant t < T, \quad u_m(0) = u_m^0. \tag{1.1}$$

Для решения задачи (1.1) используют численные методы некоторого порядка точности p. Их применение оправдано, если у решения существует $p+1$ непрерывных производных. Напомним, что для этого правые части должны иметь p-е непрерывные производные по всем аргументам. Будем предполагать это условие выполненным, если не сказано противное.

Жёсткость означает, в частности, что у решения задачи (1.1) имеются быстро изменяющиеся компоненты, и для них $|f_m| \gg 1$. Система (1.1) в общем случае неавтономна. Формально мы можем считать аргумент t также некоторой функцией $u_0(t) \equiv t$. Тогда ей соответствует правая часть $f_0(t) \equiv 1$ и начальное условие $u_0^0 = 0$. Это простейшая автономизация системы (1.1).

Введём для автономизированной системы длину дуги интегральной кривой

$$dl = \left[\sum_{m=0}^{M}(du_m)^2\right]^{1/2} = \left[\sum_{m=0}^{M} f_m^2(u_0, u_1, \ldots, u_M)\right]^{1/2} dt. \tag{1.2}$$

Заменяя в системе (1.1) dt на dl с помощью формулы (1.2), получаем

$$\frac{d\mathbf{u}}{dl} = \mathbf{F}(\mathbf{u}), \quad \mathbf{u} = (u_0, \ldots u_M)^{\mathrm{T}}, \quad \mathbf{F} = \mathbf{f}/\rho, \quad \rho = (\mathbf{f}, \mathbf{f})^{1/2}, \tag{1.3}$$

где $\mathbf{f} = (f_0, \ldots, f_M)^{\mathrm{T}}$. Если правые части системы (1.1) имеют p-е непрерывные производные по всем аргументам, то это же справедливо и для системы (1.3).

Новая система (1.3) является автономной. Очевидно, что $(\mathbf{F}, \mathbf{F}) = 1$. Тем самым, все компоненты правых частей невелики, так что пограничные слои системы (1.1) превращаются в регулярные участки системы (1.3). Это хорошо иллюстрируется рис. 2 на примере (1).

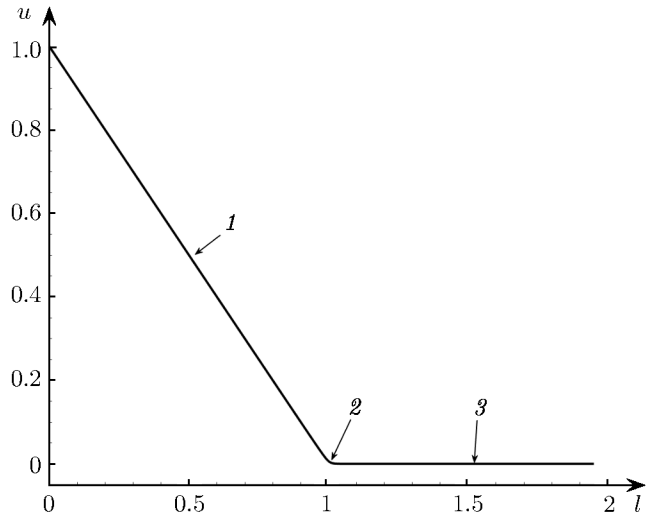

Рис. 2. Решение задачи (1) в аргументе l, обозначения соответствуют рис. 1.

Таким образом, переход к длине дуги позволяет преодолеть трудности, связанные с численным расчётом пограничных слоёв. Однако между участками бывших пограничных слоёв и регулярными решениями по-прежнему остаётся переходная зона, в которой интегральная кривая для жёстких задач имеет большую кривизну. Расчёт переходной зоны по-прежнему представляет определённую трудность и требует существенного измельчения шага.

Переход к длине дуги был, по-видимому, впервые предложен в работе [9]. Его изложение имеется в монографии [10]. В ней доказано, что такая параметризация является наилучшей с точки зрения обусловленности системы. Поэтому для жёстких систем переход к длине дуги должен стать обязательной частью решения задачи.

Однако этот метод пока остаётся малоизвестным широкому кругу вычислителей. В монографиях [1] и [11] он отсутствует. В учебной литературе его изложение имеется, по-видимому, только в [12].

2. Рекомендуемые схемы. Существует очень много явных и неявных схем, и почти все они описаны в монографиях [1] и [11].

2.1. Многошаговые схемы. Практика расчётов показала, что многошаговые схемы менее удобны. Во-первых, как сказано выше, использование равномерного шага нецелесообразно, а запись таких схем на произвольных неравномерных сетках достаточно сложна. Во-вторых, расчёты по таким схемам оказались менее надёжными: в них легче возникали срывы. Поэтому сейчас продолжают использоваться, пожалуй, только программы Гира [13], основанные на многошаговых неявных схемах Гиршфельдера–Кертисса с дифференцированием назад. Детали этих программ очень тщательно отшлифованы временем.

Подавляющее большинство современных программ основано на одношаговых многостадийных методах. В них смена шага тривиальна: каждый новый шаг не связан с предыдущим, а необходимый порядок точности достигается за счёт числа промежуточных стадий. Схемы можно разделить на три группы: явные, явно-неявные и чисто неявные.

2.2. Явные схемы. Явные схемы не требуют вычисления матрицы Якоби правой части системы. Поэтому их трудоёмкость невелика. Принято считать, что эти схемы пригодны лишь для мягких (т.е. не жёстких) задач. Однако это относится только к вычислениям с постоянным шагом. При использовании удачных формул выбора шага эти схемы позволяют находить решения задач средней и даже высокой жёсткости.

В этом случае можно рекомендовать общеизвестные явные схемы Рунге–Кутты. Схема с p стадиями имеет следующий вид:

$$\hat{\mathbf{u}} = \mathbf{u} + h\sum_{s=1}^{p} b_s \mathbf{w}_s, \quad \mathbf{w}_s = \mathbf{f}\biggl(\mathbf{u} + h\sum_{q=1}^{s-1} a_{sq}\mathbf{w}_q\biggr). \tag{2.1}$$

С увеличением числа стадий точность этих схем увеличивается, но надёжность ухудшается. Поэтому мы рекомендуем, чтобы число стадий не превышало четырёх. Для таких схем порядок точности равен числу стадий. Однако заметим, что среди стандартных пакетов есть неплохая шестистадийная схема Дормана–Принса [14] пятого порядка точности.

2.3. Явно-неявные схемы. Более высокую надёжность имеют явно-неявные схемы. Первое семейство таких схем с произвольным числом стадий предложено Розенброком [15]. Среди них уникальной надёжностью отличается одностадийная схема с комплекснозначными коэффициентами. Эта схема малоизвестна, она отсутствует в монографии [1], а в учебной литературе приведена лишь в [12]. Поэтому запишем её. Чтобы опустить номер шага, обозначим численное решение на исходном шаге через $\mathbf{u}$, а на новом шаге – через $\hat{\mathbf{u}}$. Тогда переход на новый шаг происходит по формулам

$$\hat{\mathbf{u}} = \mathbf{u} + h\operatorname{Re}\mathbf{w}, \quad \biggl(E - \frac{1+i}{2}h\,\mathbf{f}_{\mathbf{u}}\biggr)\mathbf{w} = \mathbf{f}(\mathbf{u}). \tag{2.2}$$

Здесь h – величина шага, $\mathbf{f}_{\mathbf{u}}$ – матрица Якоби правой части системы. Хотя эта схема одностадийная, она имеет второй порядок точности.

Приведённая схема требует нахождения матрицы Якоби и решения системы линейных уравнений, поэтому она гораздо более трудоёмка, чем явные схемы. Зато её надёжность настолько высока, что эта схема пригодна даже для сверхжёстких задач.

Построены и двухстадийные схемы с комплекснозначными коэффициентами, имеющие четвёртый порядок точности [16, 17]. По надёжности они несколько уступают схеме (2.2), но если расчёт идёт без срывов, то точность оказывается существенно выше.

Любые явно-неявные и неявные схемы для обеспечения высокой надёжности требуют аналитического вычисления матрицы Якоби. Если производные в матрице Якоби вычислять с помощью конечных разностей, то расчёт сохраняет надёжность лишь при повышенной разрядности чисел (64 бит может оказаться недостаточно).

2.4. Чисто неявные схемы. При использовании чисто неявных схем сталкиваются с определёнными трудностями. В них для нахождения решения на новом шаге нужно решать систему нелинейных алгебраических уравнений. Метод простых итераций для жёстких задач, как правило, не сходится. Ньютоновский итерационный процесс требует на каждой итерации решения систем линейных уравнений аналогичных (2.2), причём на каждой итерации матрицу Якоби нужно находить заново. Из-за этого неявные схемы оказываются достаточно трудоёмкими. При этом на сверхжёстких задачах сходимость итерационного процесса не гарантирована. Поэтому такие методы применяются редко.

3. Выбор шага. В настоящее время практически во всех пакетах программ используется один из двух методов выбора шага, описанных в [1, 11]. Чаще всего выбирают шаг с помощью вложенной схемы. Расчёт проводят по схеме порядка точности p, из промежуточных стадий

которой можно составить схему порядка точности $p-1$. Эта схема называется вложенной. Вычисление каждого шага проводят по обеим схемам и сравнивают результаты. Разность этих результатов считают локальной погрешностью вложенной схемы. Если она примерно равна заданной пользователем погрешности tol (её называют tolerance), то с этой же величиной h выполняется следующий шаг. В противном случае h увеличивают либо уменьшают по некоторому правилу. Именно так работает программа Дормана–Принса.

Второй метод – локальное сгущение сетки. Используют только одну схему и проводят вычисления с шагом h и с шагом $h/2$. Локальную погрешность определяют по разности результатов этих двух расчётов и дальше поступают аналогично предыдущему методу. Поскольку этот метод требует вычисления двух дополнительных полушагов, его трудоёмкость в среднем втрое выше предыдущего. Поэтому он менее употребителен.

Оригинальный алгоритм выбора шага, напоминающий локальное сгущение, предложен в работе [18]. Он основан на анализе сходимости метода Ньютона при решении системы линейных алгебраических уравнений относительно $\hat{\mathbf{u}}$: если итерации сходятся медленно, то значение $\mathbf{u}$ (выбираемое в качестве начального приближения) далеко от $\hat{\mathbf{u}}$ и шаг следует уменьшить.

В последние годы предложен третий метод, основанный на геометрических характеристиках интегральных кривых [19–21]. В нём адаптивная сетка строится по кривизне интегральной кривой. Данный метод можно применять только для аргумента l, для аргумента t он непригоден. Этот метод является наиболее перспективным, но он пока малоизвестен. Поэтому опишем его на примере явной схемы Эйлера. Предварительно введём необходимые понятия.

3.1. Квазиравномерные сетки. Они были предложены А.А. Самарским в 1952 г. и впервые опубликованы А.Ф. Сидоровым в 1966 г. [22]. Математическое определение таких сеток дано в [23]. Напомним его.

Семейство сеток $\omega_N = \{l_n : 0 \leqslant n \leqslant N\}$, $N \in \mathbb{N}$, на отрезке $[a,b]$ называется *квазиравномерным*, если существует такая строго возрастающая достаточно гладкая функция $\psi(\xi)$, $0 \leqslant \xi \leqslant 1$, $\psi(0) = a$, $\psi(1) = b$, что для любого N выполняются равенства $l_n = \psi(\xi_n)$, $1 \leqslant n \leqslant N$, где $\xi_n = n/N$. Функция ψ называется *производящей*.

Отметим некоторые следствия из этого определения. Шаг сетки равен $h_n = l_n - l_{n-1}$, $1 \leqslant n \leqslant N$. Отношение соседних шагов $h_n/h_{n-1} \to 1$ при $N \to \infty$ (отсюда название квазиравномерные для таких сеток).

Середина интервала сетки определяется как $l_{n-1/2} = \psi((n-1/2)/N)$. Если из семейства сеток выбрать семейство с удваивающимися значениями N, то узлы любой из этих сеток являются чётными узлами следующей сетки с удвоенным числом шагов, а середины её интервалов являются нечётными узлами следующей сетки.

Заметим, что квазиравномерные сетки можно строить в неограниченной области, что позволяет решать задачи в таких областях сравнительно простыми средствами.

3.2. Близость кривых. Традиционно близость кривых $\mathbf{u}(l)$ и $\mathbf{v}(l)$ определяют с помощью нормы разности $\mathbf{u}(l) - \mathbf{v}(l)$. Такой подход неудобен для разрывных решений, а также для жёстких задач, поскольку их пограничные слои очень напоминают сильные разрывы. Поэтому используем иное определение близости кривых.

Рассмотрим интегральную кривую $\mathbf{u}(l)$ как множество точек в евклидовом $(M+1)$-мерном пространстве. Расстояние между двумя кривыми определим как расстояние между соответствующими множествами в метрике Хаусдорфа. Напомним это определение. Пусть множества U и V состоят из точек $\mathbf{u}$ и $\mathbf{v}$ соответственно. Тогда расстояние $D(U,V)$ между U и V – это

$$D(U,V) = \max\{\sup_{\mathbf{v}\in V}\inf_{\mathbf{u}\in U}|\mathbf{u}-\mathbf{v}|, \sup_{\mathbf{u}\in U}\inf_{\mathbf{v}\in V}|\mathbf{u}-\mathbf{v}|\}. \tag{3.1}$$

Использование sup в этом определении делает введённое расстояние аналогом C-нормы. Если в приведённом определении заменить sup на интеграл по dl, взять $|\mathbf{u}-\mathbf{v}|$ в квадрате и извлечь из результата квадратный корень, то получим аналог L_2-нормы.

Оба определения достаточно естественно используются для решений с разрывами или пограничными слоями.

3.3. Адаптивная сетка. Пусть требуется решить задачу (1.3) на отрезке $0 \leqslant l \leqslant L$. Выбирая шаг, построим на этом отрезке сетку l_n, $0 = l_0 < l_1 < \ldots < l_N = L$ из N интервалов. Обозначим через $\varkappa_n$ кривизну и через $R_n = 1/\varkappa_n$ радиус кривизны интегральной кривой в узле l_n, $0 \leqslant n \leqslant N$, сетки.

Наша задача – построить сетку, обеспечивающую как можно более высокую точность. Такую сетку будем называть оптимальной. Интуитивно представляется, что шаги такой сетки должны сгущаться в областях с большой кривизной, но соседние шаги таких сеток не должны сильно различаться между собой. Поэтому естественно искать оптимум в классе квазиравномерных сеток.

Потребуем выполнения двух условий. Во-первых, поскольку кривизна выражается через вторые производные решения, то будем считать, что правые части системы (1.3) имеют вторые непрерывные производные. Тогда кривизна $\varkappa(l)$ будет иметь первую непрерывную производную. Во-вторых, ограничимся классом квазиравномерных сеток l_n с дважды непрерывно дифференцируемой производящей функцией.

Построим оптимальную сетку для схемы Эйлера. Шаг в этой схеме – это движение по касательной. Сравнивая расхождение кривой и касательной на шаге h_n, получаем величину локальной ошибки на одном шаге

$$\delta_n = \frac{h_n^2}{2R_n}.$$

Сама ошибка представляет собой вектор, перпендикулярный кривой. Таким образом, эту ошибку нужно рассматривать в смысле метрики Хаусдорфа. Тогда аналог сеточной L_2-нормы погрешности Δ определяется выражением

$$\Delta^2 = \sum_{n=1}^{N} \delta_n^2 h_n = \frac{1}{4} \sum_{n=1}^{N} \frac{h_n^5}{R_n^2}.$$

Будем искать набор шагов h_n, минимизирующий величину Δ. При этом нужно учитывать, что значения R_n сами зависят от положения узлов l_n и, тем самым, от набора шагов. Удобнее приближённо перейти к непрерывному индексу n, тогда $h_n \approx dl/dn$ и $R_n = R(l)$. При сделанных выше предположениях о гладкости функций такой переход является асимптотически точным. При этом должно выполняться условие

$$\sum_{n=1}^{N} h_n \approx \int_0^N \frac{dl}{dn}\, dn = L.$$

Задача на условный экстремум $\Delta \to \min$ методом Лагранжа сводится к задаче на безусловный экстремум

$$\frac{1}{4} \int_0^N \frac{1}{R^2(l)} \left(\frac{dl}{dn}\right)^5 dn - \frac{\mu}{4} \left(\int_0^N \frac{dl}{dn}\, dn - L \right) \to \min, \tag{3.2}$$

где $\mu/4$ – множитель Лагранжа.

Вариационное уравнение Эйлера с краевыми условиями для задачи (3.2) принимает вид

$$\frac{d}{dn} \left[\frac{5}{R^2(l)} \left(\frac{dl}{dn}\right)^4 - \mu \right] + \frac{1}{2R^3(l)} \left(\frac{dl}{dn}\right)^5 \frac{dR}{dl} = 0, \quad l(0) = 0, \quad l(N) = L.$$

Последнее уравнение приводится к следующему:

$$\frac{d^2 l}{dn^2} - \frac{2}{5} \left(\frac{dl}{dn}\right)^2 \frac{d \ln R}{dl} = 0, \quad l(0) = 0, \quad l(N) = L.$$

Нетрудно найти первый интеграл

$$h \equiv \frac{dl}{dn} = CR^{2/5}, \quad C = \text{const}. \tag{3.3}$$

Отсюда для положения узлов получаем равенство

$$n(l) = C^{-1} \int_0^l R^{-2/5}(\tilde{l})\, d\tilde{l},$$

константа C в котором определяется из условия $n(L) = N$. Находя эту константу, сформулируем полученный результат следующим образом.

Теорема. *При сделанных выше предположениях о гладкости оптимальная сетка для схемы Эйлера при $N \to \infty$ асимптотически удовлетворяет условию*

$$h_n = \frac{1}{N} \varkappa_n^{-2/5} \int_0^L \varkappa^{2/5}(l)\, dl, \quad 1 \leqslant n \leqslant N.$$

3.4. Расчётная формула. Построим расчётную формулу для шага. Обозначим через $N_{\max}$ число шагов на всей сетке с учётом переходных зон, а через $N_{\min}$ – число шагов на регулярных участках (без учёта переходных зон); очевидно, $N_{\min} \ll N_{\max}$. Тогда шаг ограничивается сверху выражением

$$h \leqslant L/N_{\min}. \tag{3.4}$$

В формуле (3.3) перейдём от радиуса кривизны к кривизне и подставим явное значение константы $\text{const} = 1/N_{\max}$. В качестве расчётной формулы выберем простую интерполяцию выражений (3.3) и (3.4):

$$h = \left[\frac{N_{\min}}{L} + N_{\max} \varkappa^{2/5} \left(\int_0^L \varkappa^{2/5}(l)\, dl \right)^{-1} \right]^{-1}. \tag{3.5}$$

Способ вычисления интеграла в (3.5) будет описан ниже.

Очевидно, сетка, построенная указанным образом, адаптирована к решению. Будем называть её *геометрически-адаптивной* (GEAM – Geometrically Adaptive Mesh).

Замечание. Переход от аргумента t к аргументу l представляется, на первый взгляд, усложнением задачи. Однако такой переход следует делать даже для мягких задач. Поясним причину этого.

Во-первых, правые части уравнения (1.3) очень просто выражаются через правые части уравнения (1.1). Мягкие задачи решают явными схемами, в которых требуется вычислять только правые части (но не матрицу Якоби), поэтому для них никакого усложнения фактически не происходит.

Во-вторых, приведённый выше удачный выбор шага удалось построить только для аргумента l. Формулы выбора шага по аргументу t, используемые в методах вложенных схем или локального сгущения шага, не столь надёжны, особенно в случае жёстких задач.

В-третьих, в аргументе l пограничные слои перестают быть трудными участками и легко рассчитываются крупными шагами. Измельчение шага требуется лишь в переходных зонах.

4. Вычисление кривизны. Для построения геометрически-адаптивной сетки нужно уметь вычислять кривизну многомерной кривой. Мы не встречали в литературе конструктивных формул для вычисления кривизны в многомерном пространстве. Напомним, что, по определению, кривизной называется производная от единичного вектора направления касательной к кривой по длине дуги (тем самым, это вторая производная радиус-вектора кривой по длине дуги). Реализации этого определения для явных и неявных схем различны. Опишем эти реализации.

4.1. Простейшее выражение. Вводя длину дуги в качестве аргумента, мы переходим от системы (1.1) к системе (1.3). В системе (1.3) правые части $F_m = f_m/\rho$ – это компоненты вектора касательной к интегральной кривой. Напомним, что вектор $\mathbf{F}$ имеет единичную длину. Таким образом, кривизна $\varkappa$ получается дифференцированием вектора $\mathbf{F}$ по скаляру l, т.е. это вектор

$$\varkappa = \frac{d\mathbf{F}}{dl}. \tag{4.1}$$

Правые части в (4.1) вычисляются на каждом шаге. Запишем простейшую разностную аппроксимацию

$$\hat{\varkappa} = \varkappa(l_n) = [\mathbf{F}(u_n) - \mathbf{F}(u_{n-1})]/h_n; \quad h_n = l_n - l_{n-1}. \tag{4.2}$$

Эта аппроксимация имеет первый порядок точности, что хорошо согласуется с точностью схемы Эйлера. Поэтому такая формула пригодна для построения геометрически-адаптивных сеток.

Кривизна (4.2) вычисляется после завершения текущего шага, поэтому она может использоваться для определения величины только следующего шага. На первом шаге у нас ещё нет значения кривизны. Поэтому расчёт первого шага нужно повторять дважды: сначала найти величину шага по кривизне, взятой “с потолка”, а по завершении шага найти кривизну и скорректировать шаг.

4.2. Явные схемы Рунге–Кутты. Выражение (4.2) по существу получено для одностадийной схемы Рунге–Кутты. В многостадийных схемах можно использовать для построения кривизны величины $\mathbf{w}_s$ из промежуточных стадий (2.1), а также величину $\hat{\mathbf{w}} = \mathbf{F}(\hat{\mathbf{u}})$. Использование $\hat{\mathbf{w}}$ не увеличивает объём расчётов: выражение для кривизны относится к следующему шагу, на котором $\hat{\mathbf{w}}$ всё равно нужно вычислять. Поэтому выражение для кривизны ищем в следующем виде:

$$\hat{\varkappa} = h^{-1} \sum_{q=1}^{S+1} c_q \mathbf{w}_q, \quad \mathbf{w}_{S+1} = \mathbf{F}(\hat{\mathbf{u}}). \tag{4.3}$$

Коэффициенты c_q в формуле (4.3) и b_q, a_{sq} в формуле (2.1) нужно подбирать так, чтобы решение имело аппроксимацию $O(h^p)$, а аппроксимация кривизны (4.3) имела максимально возможный порядок точности. Этот анализ делается стандартным методом разложения схемы (2.1) и выражения для кривизны (4.3) по степеням h и сравнением с соответствующими разложениями для точного решения дифференциального уравнения (1.3).

При этом оказывается, что для двухстадийной схемы возможно построить выражение кривизны лишь с порядком точности не выше первого. Однако при этом остаётся один свободный параметр, выбором которого можно уменьшить коэффициент остаточного члена в выражении для кривизны. Для трёх- и четырёхстадийных схем возможно построить выражение кривизны лишь со вторым порядком точности. Такие ограничения напоминают известные пороги Бутчера для схем Рунге–Кутты. Рекомендуемые наборы коэффициентов приведены в таблице.

Отметим, что нахождение кривизны по приведённым выше формулам не увеличивает трудоёмкости расчётов по схемам Рунге–Кутты.

4.3. Явно-неявные схемы. Проведём преобразование формального определения кривизны

$$\varkappa = \frac{d\mathbf{F}}{dl} = \mathbf{F}_\mathbf{u} \frac{d\mathbf{u}}{dl} = \mathbf{F}_\mathbf{u}\mathbf{F}. \tag{4.4}$$

Таким образом, кривизна равна произведению матрицы Якоби правой части и вектора правой части. Поскольку в явно-неявных и неявных схемах матрицу Якоби всё равно приходится вычислять (это самая трудоемкая часть расчёта), то попутное нахождение кривизны по формуле (4.4) не увеличивает общую трудоемкость вычислений. Поскольку матрица Якоби вычисляется *до* выполнения шага, то выражение (4.4) для кривизны можно использовать для определения величины текущего шага, а не следующего. В этом заключается качественное преимущество неявных схем перед явными.

Таблица. Коэффициенты схемы (2.1) и кривизны (4.3) при различных значениях p

c_q	b_s	a_{sq}			
		$p=1$			
-1	1	–			
1	–	–			
		$p=2$			
0	0	0	0		
2	1	1/2	0		
2	–	–	–		
		$p=3$			
2/3	2/9	0	0	0	
-2	1/3	1/2	0	0	
$-8/3$	4/9	0	3/4	–	
4	–	–	–	–	
		$p=4$			
1	1/6	0	0	0	0
-2	1/3	1/2	0	0	0
-2	1/3	0	1/2	0	0
0	1/6	0	0	1	0
3	–	–	–	–	–

5. Расчёт с гарантированной точностью. В известных пакетах программ также строятся некоторые адаптивные сетки. В них пользователь задаёт требуемую точность tol, и шаги сетки выбираются на основе вложенной схемы либо локального сгущения шага. На этой сетке проводится единственный расчёт. Предполагается, что погрешность этого расчёта равна заданной tol. Это предположение никак не обосновывается.

Тестирование таких пакетов на задачах с известными точными решениями показывает, что для мягких задач фактическая погрешность может превышать tol, но не во много раз. Однако на жёстких задачах фактическая погрешность может быть на несколько порядков больше, чем tol.

Расчёт на единственной сетке в принципе не может дать гарантированную оценку погрешности. Единственный способ получения надёжной оценки – это расчёт на последовательности сгущающихся сеток и сравнение решений на этих сетках по методу Ричардсона. Этот способ даёт асимптотически точное значение погрешности. Первоначально этот способ был предложен для равномерных сеток. Впоследствии было показано, что он применим и на квазиравномерных сетках [11, 23, 24], а также на кусочно-равномерных и кусочно-квазиравномерных. Поэтому для гарантированной оценки погрешности нужно найти такую сетку, которая была бы геометрически-адаптивной и одновременно квазиравномерной. Для этого приходится строить процедуру расчёта, состоящую из двух стадий. Опишем её.

5.1. Построение адаптивной сетки. Это первая стадия расчёта. Возьмём некоторые начальные не особенно большие значения $N_{\min}$ и $N_{\max}$ и проведём расчёт по явной схеме Эйлера, используя для шага формулу (3.5). Перед началом расчёта нам известно полное время T, но длина дуги L и значение интеграла в (3.5) пока неизвестны. Поэтому зададим их “с потолка”. Расчёт на первой сетке будем вести до тех пор, пока текущее расчётное время не станет большим либо равным T. В ходе этого расчёта найдём L и вычислим значение интеграла от кривизны по любой квадратурной формуле. Затем удвоим $N_{\min}$ и $N_{\max}$, воспользуемся уже найденным значением интеграла и повторим расчёт уже найденной адаптивной сетки. Она не будет сгущением первой сетки, так как её чётные узлы не будут совпадать с узлами первой сетки. Поэтому снова удвоим $N_{\min}$ и $N_{\max}$ и повторим расчёт. Такое удвоение будем повторять до тех пор, пока чётные узлы новой сетки не окажутся достаточно близкими к узлам предыдущей сетки.

В тестовых расчётах было опробовано несколько критериев близости сеток. Наиболее удачными оказались два следующих критерия. В первом критерии требовалась малость величин $(l_n - \hat{l}_{2n})(\hat{h}_{2n}^{-1} + \hat{h}_{2n+1}^{-1})$. Выполнение этого критерия означает, что разность положений узлов должна быть мала по сравнению с соседними шагами. Здесь l относится к более грубой сетке, $\hat{l}$ – к более подробной сетке. Во втором критерии сравнивались только отношения соответствующих шагов на двух сетках

$$\sqrt{\xi_n} - \frac{1}{\sqrt{\xi_n}}, \quad \xi_n = \frac{\hat{h}_{2n} + \hat{h}_{2n+1}}{h_n}.$$

В обоих случаях требовалась малость среднеквадратичной нормы этих величин.

Схема Эйлера используется по следующим причинам. Во-первых, она наименее трудоёмка среди всех известных схем. Во-вторых, среди явных схем она наиболее надёжна. Её низкая точность несущественна, поскольку результат расчёта нужен только для построения геометрически-адаптивной сетки.

5.2. Квазиравномерное сгущение сетки. Напомним, что априорные мажорантные оценки погрешности через производные решения неконструктивны и обычно даже так называемые неулучшаемые оценки сильно превышают фактические погрешности. Асимптотически точную величину погрешности даёт только метод Ричардсона. Однако для применения этого метода к дифференциальным уравнениям необходимо сгущать сетки в точности вдвое, причём так, чтобы все узлы предыдущей сетки совпадали с чётными узлами новой сетки (тогда возможно поточечное сравнение решений на соседних сетках и вычисление сеточных норм погрешности). Сами сетки должны быть равномерными или квазиравномерными.

На стадии построения адаптивной сетки узлы соседних сеток не совпадают. Однако если выполнен критерий совпадения сеток, то построенная сетка хорошо соответствует кривизне точного решения. Тем самым, эта сетка получается из равномерной сетки некоторым гладким преобразованием, т.е. является квазиравномерной. Поэтому её можно взять за основу для квазиравномерного сгущения и применения метода Ричардсона.

Для квазиравномерного сгущения между каждой парой узлов старой сетки нужно поставить узел новой сетки, соответствующий той же производящей функции. Приведём формулы для нахождения такого узла. Пусть шаг h_n является внутренним интервалом исходной сетки. Тогда он делится на два интервала с шагами

$$\hat{h}_{2n-1} = h_n \frac{\sqrt[4]{h_{n-1}}}{\sqrt[4]{h_{n-1}} + \sqrt[4]{h_{n+1}}}, \quad \hat{h}_{2n} = h_n \frac{\sqrt[4]{h_{n+1}}}{\sqrt[4]{h_{n-1}} + \sqrt[4]{h_{n+1}}}. \tag{5.1}$$

Если интервал примыкает к левой границе, то его шаг h_1 делится на два новых шага по правилу

$$\hat{h}_1 = h_1 \frac{\sqrt{h_1}}{\sqrt{h_1} + \sqrt{h_2}}, \quad \hat{h}_2 = h_1 \frac{\sqrt{h_2}}{\sqrt{h_1} + \sqrt{h_2}}. \tag{5.2}$$

Для интервала h_N, примыкающего к правой границе, проводим аналогичное деление

$$\hat{h}_{2N-1} = h_N \frac{\sqrt{h_{N-1}}}{\sqrt{h_{N-1}} + \sqrt{h_N}}, \quad \hat{h}_{2N} = h_N \frac{\sqrt{h_N}}{\sqrt{h_{N-1}} + \sqrt{h_N}}. \tag{5.3}$$

Формулы сгущения (5.1)–(5.3) имеют третий порядок точности. Однако их можно применять к одношаговым схемам решения задачи Коши любого порядка точности (в том числе выше третьего). Это объясняется тем, что в одношаговых схемах нет дифференцирования сквозь узел. Но для многошаговых схем (например, схем Адамса) такое сгущение сеток может ограничить предельный порядок точности.

Результаты расчета тестовых задач с такими сетками показали, что нерекуррентный метод Ричардсона (оценивающий погрешность по двум соседним сеткам) даёт асимптотически

точное значение погрешности. Такую оценку погрешности можно прибавить к полученному результату и получить повышение порядка точности на единицу.

Однако вопрос о применимости рекуррентного метода Ричардсона, использующего несколько последовательных сеток и повышающего порядок точности на несколько единиц, остаётся открытым.

6. Пример расчёта. В качестве примера расчёта рассмотрим важную и достаточно сложную прикладную задачу – кинетику химических реакций горения смеси водорода с кислородом. Система содержит 25 прямых и стольких же обратных реакций и учитывает 9 компонент: O, H, O_2, H_2, OH, H_2O, HO_2, O_3, H_2O_2. Сама система довольно громоздка и приведена в работе [25]. Приведём пример расчёта "гремучей смеси" (2 части молекулярного водорода и 1 часть молекулярного кислорода) при температуре $T = 2000$ К. Время горения – 10 мкс. Расчёт проводился со сгущением сеток до выхода на заданную высокую точность.

На рис. 3 приведены распределения концентраций важнейших компонент в зависимости от времени. Хорошо виден внутренний пограничный слой, соответствующий вспышке смеси, и выход концентраций на стационарные значения. Все графики являются гладкими кривыми без пилообразных осцилляций. Это означает хорошее качественное поведение численного решения и свидетельствует о надёжности схемы.

Точное решение данной задачи неизвестно. Поэтому погрешность решения определялась методом Ричардсона по сгущению сеток. Вычислялась погрешность в L_2-норме, в которой усреднение проводилось не только по времени, но и по всем компонентам. На рис. 4 приведены погрешности расчёта с геометрически-адаптивной сеткой для трёх явных схем: четырёхстадийная схема Рунге–Кутты, двухстадийная схема Рунге–Кутты и схема точности $O(h^2)$, названная химической [25]. Последняя схема является специализированной. Она построена исключительно для задач химической кинетики и учитывает специфическую структуру правых частей их уравнений, а также неотрицательность всех функций (поскольку их значения – это химические концентрации). График дан в двойном логарифмическом масштабе. Все три линии выходят на прямые, а их наклоны соответствуют теоретическим порядкам точности схем.

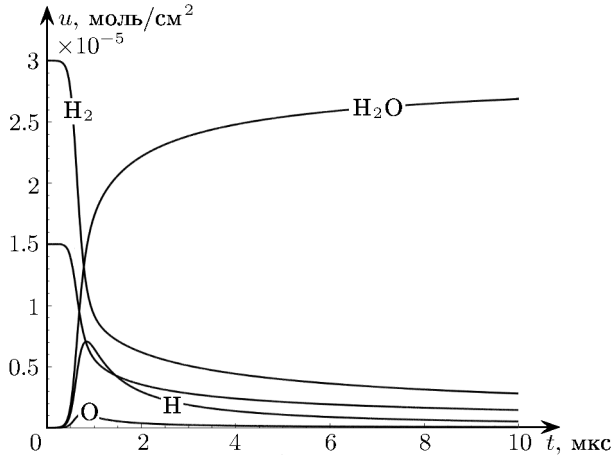

Рис. 3. Поведение концентраций при горении водорода в кислороде при $T = 2000$ К.

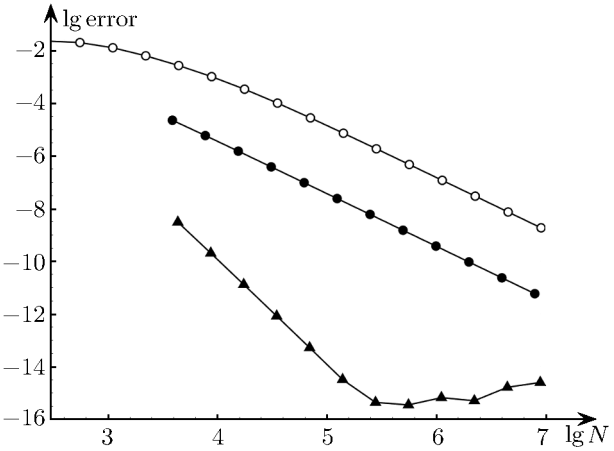

Рис. 4. Горение водорода в кислороде при $T = 2000$ К. Погрешности схем: ▲ – РК4, ● – РК2, ○ – специальная химическая схема.

Видно, что наилучшую точность при хорошей надёжности обеспечивает четырёхстадийная схема Рунге–Кутты. Точность двухстадийной схемы Рунге–Кутты существенно хуже. Химическая схема несколько уступает по точности двухстадийной схеме Рунге–Кутты. Тем самым, все описанные явные схемы вполне пригодны для расчёта сложных прикладных задач, если используются геометрически-адаптивные сетки.

Однако если вести расчёты на равномерной сетке с принудительно заданным шагом, то схемы Рунге–Кутты требуют неприемлемо малого шага. При увеличении шага счёт быстро срывается. Это убедительно показывает преимущества, которые доставляют выбор длины дуги в качестве независимой переменной и построение геометрически-адаптивной сетки.

СПИСОК ЛИТЕРАТУРЫ

1. *Хайрер Э., Ваннер Г.* Решение обыкновенных дифференциальных уравнений. Жесткие и дифференциально-алгебраические задачи. М., 1999.
2. *Виноград Р.Э.* Об одном критерии неуйстойчивости в смысле Ляпунова решений линейной системы обыкновенных дифференциальных уравнений // Докл. АН СССР. 1952. Т. 84. № 2. С. 201–204.
3. *Былов Б.Ф., Виноград Р.Э., Гробман Д.М., Немыцкий В.В.* Теория показателей Ляпунова и ее приложения к вопросам устойчивости. М., 1966.
4. *Ракитский Ю.В., Устинов С.М., Черноруцкий И.Г.* Численные методы решения жёстких систем. М., 1979.
5. *Тихонов А.Н.* О зависимости решений дифференциальных уравнений от малого параметра // Мат. сб. 1948. Т. 22 (64). № 2. С. 193–204.
6. *Васильева А.Б., Бутузов В.Ф.* Асимптотические методы в теории сингулярных возмущений. М., 1990.
7. *Нефедов Н.Н.* Метод дифференциальных неравенств для некоторых сингулярно возмущенных задач в частных производных // Дифференц. уравнения. 1995. Т. 31. № 4. С. 719–722.
8. *Белов А.А., Калиткин Н.Н.* Проблема нелинейности при численном решении сверхжёстких задач Коши // Мат. моделирование. 2016. Т. 28. № 4. С. 16–32.
9. *Riks E.* The application of Newton's method to the problem of elastic stability // J. of Appl. Mechanics. 1972. V. 39. № 4. P. 1060–1065.
10. *Шалашилин В.И., Кузнецов Е.Б.* Метод продолжения решения по параметру и наилучшая параметризация. М., 1999.
11. *Хайрер Э., Нерсет С., Ваннер Г.* Решение обыкновенных дифференциальных уравнений. Нежёсткие задачи. М., 1990.
12. *Калиткин Н.Н., Корякин П.В.* Численные методы. В 2 кн. Кн. 2: Методы математической физики М., 2013.
13. *Shampine L.F., Reichelt M.W.* The Matlab ODE suite // SIAM J. on Sci. Comp. 1997. V. 18. № 1. P. 1–22.
14. *Dormand J.R., Prince P.J.* A family of embedded Runge–Kutta formulae // J. of Comp. and Appl. Math. 1980. V. 6. P. 19–26.
15. *Rosenbrock H.H.* Some general implicit processes for the numerical solution of differential equations // Comput. J. 1963. V. 5. № 4. P. 329–330.
16. *Ширков П.Д.* Оптимально затухающие схемы с комплексными коэффициентами для жёстких систем ОДУ // Мат. моделирование. 1992. Т. 4. № 8. С. 47–57.
17. *Альшин А.Б., Альшина Е.А., Лимонов А.Г.* Двухстадийные комплексные схемы Розенброка для жёстких систем // Журн. вычислит. математики и мат. физики. 2009. Т. 49. № 2. С. 270–287.
18. *Галанин М.П., Конев С.А.* Об одном численном методе решения обыкновенных дифференциальных уравнений // Препринты ИПМ им. М.В. Келдыша. 2017. № 18.
19. *Белов А.А., Калиткин Н.Н., Пошивайло И.П.* Геометрически-адаптивные сетки для жёстких задач Коши // Докл. РАН. 2016. Т. 466. № 3. С. 276–281.
20. *Белов А.А., Калиткин Н.Н.* Выбор шага по кривизне для жёстких задач Коши // Мат. моделирование. 2016. Т. 28. № 11. С. 97–112.
21. *Белов А.А., Калиткин Н.Н.* Численные методы решения задач Коши с контрастными структурами // Моделирование и анализ информац. систем. 2016. Т. 23. № 5. С. 528–537.
22. *Сидоров А.Ф.* Об одном алгоритме расчета оптимальных разностных сеток // Тр. Мат. ин-та им. В.А. Стеклова АН СССР. 1966. Т. 74. С. 147–151.
23. *Калиткин Н.Н.* Численные методы. М., 1978.
24. *Калиткин Н.Н., Альшин А.Б., Альшина Е.А., Рогов Б.В.* Вычисления на квазиравномерных сетках. М., 2005.
25. *Белов А.А., Калиткин Н.Н., Кузьмина Л.В.* Моделирование химической кинетики в газах // Мат. моделирование. 2016. Т. 28. № 8. С. 46–64.

Московский государственный университет
им. М.В. Ломоносова,
Российский университет дружбы народов
им. П. Лумумбы, г. Москва,
Институт прикладной математики
им. М.В. Келдыша РАН, г. Москва

Поступила в редакцию 08.02.2019 г.
После доработки 08.02.2019 г.
Принята к публикации 12.02.2019 г.