

УДК 519.62

В. В. БОБКОВ, П. А. МАНДРИК, В. И. РЕПНИКОВ

**АДАПТИВНЫЕ МЕТОДЫ ЧИСЛЕННОГО РЕШЕНИЯ
ЖЕСТКИХ СИСТЕМ**

1. Введение. Хотя основное внимание в данной работе будет уделено задаче Коши для системы обыкновенных дифференциальных уравнений вида

$$u' = f(t, u), \quad (1)$$

где $u = (u_1, u_2, \dots, u_n)^T$, $f = (f_1, f_2, \dots, f_n)^T$, $f_i = f_i(t, u)$, $u_i = u_i(t)$, $i = 1, 2, \dots, n$, однако многое из сказанного ниже может быть использовано при конструировании разностных схем и в случае других эволюционных задач с дифференциальными операторами более сложной структуры, а также при построении итерационных процессов для стационарных задач, привязанных к соответствующей системе вида (1) через идею установления решения во времени.

При разработке вычислительных алгоритмов для решения исходной дифференциальной задачи обычно обязательным этапом является переход к аппроксимирующей ее разностной задаче, многие свойства которой (например, устойчивость) часто рассматривают [1, с. 19] как внутренние свойства разностной схемы. Очевидно, что замена дифференциальной модели разностной, не претерпевающей конструктивных изменений во времени, значительно снижает, вообще говоря, возможности так организованного численного моделирования. И если адекватность дифференциальной модели исследуемому процессу может сохраняться при этом на всем рассматриваемом временном промежутке, то для разностной модели (особенно в случае явных схем) такой адекватности часто удается достигать лишь при значениях шагов сетки, значительно меньших тех, которые являются естественными для наблюдения данной стадии процесса. Это особенно актуально в том случае, когда мы сталкиваемся с явлением жесткости [2, с. 16] и вынуждены численно анализировать медленно изменяющуюся стадию процесса при наличии быстро затухающих малых возмущений. Существенная разномасштабность составляющих процесса обычно заставляет нас при выборе шага сетки ориентироваться на более «быстрые» составляющие и в той стадии, когда их вкладом в пределах требуемой точности наблюдения уже можно было бы пренебречь. Более экономичным мог бы быть такой вычислительный алгоритм, в основе которого лежит требование хорошей аппроксимации только тех составляющих решения, которые наблюдаемы в пределах заданной точности, а для других составляющих обеспечен лишь некоторый уровень их качественного поведения. При построении такого типа разностных схем, привязанных к интересующему нас решению, следует, естественно, обеспечить выполнение ряда новых условий, на часть из которых будет обращено внимание ниже. При этом будут упомянуты лишь те требования, выполнение которых в дальнейшем будет конструктивно обеспечено в рамках данной работы.

В качестве системы дифференциальных уравнений, на которую будем здесь ориентироваться при постановке дополнительных требований к кон-

струируемым численным методам (в том числе и универсального назначения), возьмем систему

$$u' = Au + b \quad (2)$$

с постоянными матрицей A и вектором b . Такой выбор обусловлен не только тем, что проблема численного решения систем вида (2) представляет значительный самостоятельный интерес, а основные задачи линейной алгебры, построение итерационных процессов для которых может быть непосредственно связано с численным интегрированием (2), наиболее часто встречаются в вычислительной практике, но и тем, что в общем случае системы (1) часто бывает оправданной предварительная аппроксимация на каждом шаге сетки исходной задачи последовательностью линейных задач (в сочетании с известным приемом «замораживания» коэффициентов, а также с оценкой области адекватности таких задач).

Если в рассматриваемом n -мерном векторном пространстве решений системы (2) из собственных векторов матрицы A можно выбрать, например, ортонормированный базис $\xi^1, \xi^2, \dots, \xi^n$, отвечающий собственным значениям $\lambda_1, \lambda_2, \dots, \lambda_n$ соответственно, то любая траектория системы (2) с неособенной матрицей A может быть записана в виде

$$u(t) = -A^{-1}b + \sum_{i=1}^n c_i \xi^i \exp(\lambda_i t) \quad (3)$$

со своими значениями коэффициентов c_1, c_2, \dots, c_n , определяемыми ее начальным состоянием. В случае, скажем, $\text{Re } \lambda_i < 0, i = 1, 2, \dots, n$, структура избранной траектории системы (2) с течением времени в пределах требуемой точности наблюдений упрощается и при $t \gg 1$ (если, например, $c_1 \neq 0, \text{Re } \lambda_i < \lambda_1 < 0, i = 2, 3, \dots, n$) можно говорить о так называемой стадии регулярного режима [1, с. 301], когда

$$u(t) \approx -A^{-1}b + c_1 \xi^1 \exp(\lambda_1 t), \quad u'(t) \approx c_1 \xi^1 \lambda_1 \exp(\lambda_1 t) \approx \lambda_1 u(t) + \lambda_1 A^{-1}b$$

и система (2) в пределах точности наблюдений распадается на n независимых линейных обыкновенных дифференциальных уравнений с постоянными коэффициентами. При большом разбросе собственных значений матрицы A (случай жестких [3, с. 17] систем) подобная ситуация возможна и в других стадиях. Поэтому конструируемому численному методу естественно предъявить дополнительное требование — быть точным на простейших решениях

$$u(t) = u'(t) = -A^{-1}b + c_i \xi^i \exp(\lambda_i t) \quad (4)$$

системы (2) для случая вещественных λ_i . Решения вида (4) являются более характерными для систем обыкновенных дифференциальных уравнений (см. (3)), чем полиномиальные решения, требование точности на которых (до определенной степени полинома) обеспечено для многих классических методов. Это обстоятельство требует иных подходов к построению численных методов (см., например, [4, 5]), при которых полиномиальная аппроксимация не является определяющей. Некоторые из этих подходов будут использованы и ниже.

Обычно, наблюдая за поведением избранной траектории, мы имеем возможность следить лишь за изменением некоторого функционала вдоль нее. В качестве такого функционала часто используют норму (скажем, евклидову) решения. Заметим, что посредством такого функционала неразличимы, например, два следующих принципиально разных решения однородной (для простоты записи) системы (2):

$$u^* = a_1 \xi^1 + a_n \xi^n, \quad u^{**} = a_n \xi^1 + a_1 \xi^n, \quad |a_n| \ll |a_1|, \quad \lambda_n \ll \lambda_1 < 0.$$

В то же время, к примеру, значения $\mu_0(u^*)$ и $\mu_0(u^{**})$ функционала $\mu_0(z) = (Az, z) / (z, z)$, называемого обычно отношением Рэлея, уже существен-

но различны. Естественно использовать такой функционал для получения информации об избранной траектории в данный момент времени. Кроме того, известно (см., например, [5]), что в случае $A=A^T$ вдоль любой нестационарной траектории системы (2), отличной от (4), отношение Рэлея

$$\mu_0(v) = (Av, v)/(v, v), \quad v=v(t)=u'(t), \quad (5)$$

монотонно возрастает (в пределах спектра матрицы), при этом $\mu_0(v) \equiv \lambda_i$, если $u(t) = u^i(t)$. В качестве $v(t)$ на приближенном решении $y(t)$ может быть взята величина $Ay(t) + b$. Тогда при $y \neq u^i(t)$ естественным является требование

$$\mu_0(Ay + b) < \mu_0(A\hat{y} + b), \quad (6)$$

где $y=y(t)$, $\hat{y}=y(t+\tau)$, $\tau > 0$. Иногда вместо (5) бывает целесообразным использовать обобщенное отношение Рэлея

$$\mu_k(v) = (A^{k+1}v, v)/(A^k v, v), \quad (7)$$

обладающее аналогичными (5) свойствами (см. [5]). Для функционала (7) на приближенном решении также является естественным требование типа (6).

Ставя перед собой цель — построить такие вычислительные алгоритмы, для которых выбор шага сетки предопределяется главным образом требованиями заданной точности аппроксимации избранного решения, мы должны априори обеспечить достаточно высокий уровень согласованности дифференциальной и соответствующей разностной задач. Укажем здесь еще на ряд условий, выполнение которых при этом должно быть предусмотрено. Если применительно к (2) конструируемый метод принимает вид

$$\hat{y} = Sy + g = y + Qy + g, \quad (8)$$

где $y \approx u(t)$, $\hat{y} \approx u(t+\tau)$, то посредством матрицы $S = E + Q$ и вектора g можно судить о достоинствах метода, руководствуясь, например, следующими соображениями. Заметим, что в силу (3)

$$u(t+\tau) = -A^{-1}b + \sum_{i=1}^n c_i \xi^i \exp(\lambda_i t) \exp(\lambda_i \tau). \quad (9)$$

Чтобы при $y=u(t)$ в n -мерном векторном пространстве, натянутом на $\xi^1, \xi^2, \dots, \xi^n$, для \hat{y} метод (8) приводил к аналогичной конструкции

$$\hat{y} = -A^{-1}b + \sum_{i=1}^n c_i \xi^i \exp(\lambda_i t) s_i, \quad (10)$$

достаточно, очевидно, обеспечить выполнение условий

$$Q^{-1}g = A^{-1}b, \quad (11)$$

$$S\xi^i = s_i \xi^i, \quad i = 1, 2, \dots, n. \quad (12)$$

Для случая перестановочных матриц A и Q , что обычно не является обременительным ограничением на конструкцию метода, вместо (11) можно записать условие

$$Ag = Qb. \quad (13)$$

Наряду с (12), (13) естественны также следующие требования к множителям $s_i = s_i(\tau)$ в (10), призванным играть здесь роль множителей $\exp(\lambda_i \tau)$ из (9):

$$\text{если } \operatorname{Re} \lambda_i < 0, \text{ то } s_i(\tau) \rightarrow 0 \text{ при } \tau \rightarrow \infty; \quad (14)$$

$$\text{если } \lambda_i > 0, \text{ то } s_i(\tau) \rightarrow \infty \text{ при } \tau \rightarrow \infty, \quad (15)$$

$$\text{если } \operatorname{Re} \lambda_i > 0, \text{ то } |s_i(\tau)| \rightarrow \infty \text{ при } \tau \rightarrow \infty, \quad (16)$$

$$\text{если } \lambda_i < 0, \text{ то } 0 < s_i(\tau) < 1, \quad (17)$$

$$\text{если } \operatorname{Re} \lambda_i < 0, \text{ то } |s_i(\tau)| < 1, \quad (18)$$

$$\text{если } \lambda_i > 0, \text{ то } s_i(\tau) > 1, \quad (19)$$

$$\text{если } \operatorname{Re} \lambda_i > 0, \text{ то } |s_i(\tau)| > 1, \quad (20)$$

$$\text{если } \lambda_j < \lambda_k, \text{ то } s_j(\tau) < s_k(\tau), \quad (21)$$

$$\text{если } \operatorname{Re} \lambda_i = 0, \text{ то } |s_i(\tau)| = 1. \quad (22)$$

Заметим, что ни один из известных классических методов (как явных, так и неявных) требованиям (14) — (22), а также другим сформулированным выше условиям одновременно не удовлетворяет.

2. Вычислительные алгоритмы, основанные на построении результирующего разностного оператора. При решении задачи Коши для системы дифференциальных уравнений вида (2) наряду с шагом дискретизации $\tau > 0$ введем новый параметр $h > 0$, постоянный для данного τ и связанный с последним соотношением $\tau = 2^k h$, $k \geq 0$. На отрезке длиной h независимой переменной рассмотрим некий численный метод, который в применении к (2) может быть записан в виде

$$y^{j+1} = y^j + Q_0 y^j + g^0 = S_0 y^j + g^0, \quad (23)$$

где $y^{j+\alpha} \approx u(t_j + \alpha h)$, $t_j = t + jh$, $j = 0, 1, \dots, 2^k$, а матрица $S_0 = E + Q_0$ и вектор g^0 постоянны на шаге τ . Опираясь на (23), зададим следующий способ построения вектора $\hat{y} \approx u(t + \tau)$ по известному значению $y \approx u(t)$ искомого решения:

$$y^0 = y, \quad y^{j+1} = S_0 y^j + g^0, \quad j = 0, 1, \dots, 2^k - 1, \quad \hat{y} = y^{2^k}. \quad (24)$$

Используя идею [6] построения результирующего разностного оператора, определяемого алгоритмом (24), можно записать равенства

$$\begin{aligned} \hat{y} &= S_0^{2^k} y + (E + S_0 + S_0^2 + \dots + S_0^{2^k - 1}) g^0 = \\ &= S_0^{2^k} y + (E - S_0^{2^k}) (E - S_0)^{-1} g^0 = S_0^{2^k} y + \prod_{i=1}^k (E + S_0^{2^i - 1}) g^0, \quad k \geq 1, \end{aligned} \quad (25)$$

которые в свою очередь определяют метод вида

$$\hat{y} = S_k y + g^k = y + Q_k y + g^k, \quad (26)$$

где $Q_i = 2Q_{i-1} + Q_{i-1}^2$, $g^i = 2g^{i-1} + Q_{i-1} g^{i-1}$, $i = 1, 2, \dots, k$.

Сопоставляя вычислительные затраты на шаге τ алгоритмов (24) и (26), нетрудно видеть, что эффективность последнего, измеряемая, например, лишь количеством операций умножения матрицы на вектор, связана с неравенством $(n+1)k + 1 < 2^k$. Подобная эффективность, очевидно, с увеличением k быстро нарастает. Это особенно важно для случая жестких систем в стадии, близкой к (4).

Рассмотрим сейчас вопрос выбора Q_0 и g^0 в (26), который обеспечивает выполнение ряда требований согласованности дифференциальной и разностной задач, обсуждавшихся выше.

Отметим прежде всего вытекающие из (25) равенства

$$s_i(\tau) = (s_{0i}(h))^{2^k}, \quad i = 1, 2, \dots, n, \quad (27)$$

связывающие множители $s_i = s_i(\tau)$ (см. (10), (12)) конструируемого метода (26) и соответствующие множители $s_{0i}(h)$ базового метода (23).

Равенство (27) делает очевидным утверждение, что если при некоторых h для спектральных множителей $s_{0i}(h)$ базового метода справедливы неравенства типа (17) — (21), то для спектральных множителей $s_i(\tau)$ результирующего метода будут выполнены требования (14) — (21).

Из равенств (25) также вытекает, что при выполнении условий (см. (12), (13))

$$S_0 \xi^i = s_{0i} \xi^i, \quad i=1, 2, \dots, n, \quad (28)$$

$$Ag^0 = Q_{0b} \quad (29)$$

и условий типа (18) для множителей $s_{0i}(h)$, $i=1, 2, \dots, n$, в асимптотически устойчивом случае системы (2) будет иметь место важное свойство

$$\hat{y} \rightarrow -A^{-1}b \quad \text{при } k \rightarrow \infty \quad (\tau = 2^k h), \quad (30)$$

которое особенно существенно в том случае, если на основе подобных методов будут строиться итерационные процессы для решения стационарных задач вида $Au + b = 0$.

Заметим, что выполнения для множителей $s_{0i}(h)$ условий типа (17) — (21) можно добиться, например, ограничив значения параметра h некоторой предельной величиной h_0 . Выполнение же для множителя $s_{0i}(h)$ равенства типа (22), а также требования (29) и свойства точности на решениях вида (4) зависит от выбора базового численного метода вида (23).

Последнее свойство, например, может быть достигнуто на пути выделения из матричной экспоненты главной спектральной составляющей [5] или реализацией способа пошагового выделения и точного обращения главной части исходного дифференциального оператора [4], или, как в настоящем случае, применением принципа мультипликативной корректировки спектральных свойств известных численных методов [7].

Ориентируя, скажем, конструируемый вычислительный алгоритм на случай вещественного спектра матрицы A , в (26) матрицу S_0 и вектор g^0 можно выбирать, например, в виде

$$S_0 = E + h\rho A \left(E + \frac{hA}{2!} + \dots + \frac{(hA)^{p-1}}{p!} \right), \quad (31)$$

$$g^0 = h\rho A \left(E + \frac{hA}{2!} + \dots + \frac{(hA)^{p-1}}{p!} \right) b, \quad (32)$$

где

$$\rho = \frac{\exp(\mu h) - 1}{\mu h + \frac{(\mu h)^2}{2!} + \dots + \frac{(\mu h)^p}{p!}}, \quad \mu = \mu_0(Ay + b). \quad (33)$$

Подобный вариант выбора S_0 и g^0 в базовом модуле (23) может быть получен мультипликативной корректировкой, например, некоторых явных методов, основанных на принципе последовательного повышения порядка точности [8, с. 31]. Непосредственно проверяется, что в этом случае выполняются равенство (29) и свойство точности на решениях вида (4), а главная часть локальной погрешности соответствующего метода (26), (31) — (33) может быть записана в виде

$$r = \frac{2^{-kp}}{(\rho + 1)!} \tau^{p+1} (A^p - \mu^p E) (Ay + b) = \frac{2^k}{(\rho + 1)!} h^{p+1} (A^p - \mu^p E) (Ay + b).$$

Для соответствующих множителей $s_{0i}(h)$ с учетом (28) справедливо представление

$$s_{0i}(h) = 1 + \rho \left(h\lambda_i + \frac{(h\lambda_i)^2}{2!} + \dots + \frac{(h\lambda_i)^p}{p!} \right),$$

из которого, в частности, непосредственно следует, что в случае метода первого порядка точности ($p=1$) выбор

$$h < h_0 \leq \frac{1}{\mu} \ln \left(1 + \frac{\mu}{\|A\|} \right), \quad (34)$$

где $\|A\|$ есть, например, евклидова норма матрицы A , обеспечивает выполнение для $s_0(h)$ неравенств типа (17), (19), (21). Следовательно, в этом случае при любых шагах $\tau > 0$ обеспечивается выполнение всех рассмотренных выше требований согласованности в поведении точного и приближенного решений, в том числе и (30).

Аналогично (34) могут быть оценены значения параметра h_0 в случае методов высших порядков точности ($p > 1$). Так, например, для метода второго порядка точности можно взять $h_0 \leq 1/\|A\|$, а для случая четвертого порядка точности h_0 может быть ограничена величиной $1.59/\|A\|$.

Заметим, что варианты выбора базового модуля для (26) могут быть различными (см., например, [9]). Здесь же мы отметим лишь один вариант, основанный на мультипликативной корректировке неявных методов Рунге — Кутты (методов Гаусса — Лежандра [2, с. 77]) и предлагаемый для повышения уровня согласованности дифференциальной и соответствующей разностной задач в случае наличия в спектре матрицы A комплексных собственных значений (сохраняя при этом достигнутый уровень согласованности в случае вещественного спектра матрицы A). Корректировка неявного метода трапеций, например, приводит к выбору базового метода (23), где

$$S_0 = (E - 0,5h\hat{\rho}A)^{-1}(E + 0,5h\hat{\rho}A), \quad (35)$$

$$g^0 = h\hat{\rho}(E - 0,5h\hat{\rho}A)^{-1}, \quad (36)$$

$$\hat{\rho} = \frac{2}{\mu h} \cdot \frac{\exp(\mu h) - 1}{\exp(\mu h) + 1}, \quad \mu = \mu_0(Ay + b). \quad (37)$$

Непосредственно проверяется, что все рассмотренные выше требования согласованности, в том числе и условие (22), выполняются для метода вида (26), (35) — (37), если параметр h удовлетворяет условию

$$h < h_0 \leq \frac{1}{\mu} \ln \left(\frac{\|A\| + \mu}{\|A\| - \mu} \right). \quad (38)$$

Заметим далее, что если о локальной на шаге τ ошибке приближенного решения судить, например, по величине главной части локальной погрешности метода, то, ограничив допустимое значение такой ошибки величиной ε , нетрудно построить примеры вычислительных алгоритмов с автоматическим выбором шага и порядка точности метода.

Рассмотрим, например, алгоритм, основанный на методе первого порядка точности вида (26), (31) — (33), где $p = 1$. Евклидова норма главной части локальной погрешности данного метода имеет вид $\|\tau\| = \tau^2 \omega / 2^{k+1} = 2^{k-1} h^2 \omega$, где $\omega = \|(A - \mu E)(Ay + b)\|$, учитывая который можно записать (см. (34))

$$h_e \leq \sqrt{\frac{2\varepsilon}{\omega}}, \quad h \leq \min(h_e, h_0), \quad (39)$$

$$k \leq \log_2 \left(\frac{2\varepsilon}{h^2 \omega} \right), \quad \tau = 2^k h. \quad (40)$$

Последние неравенства позволяют автоматически выбирать значения параметров h и k , адаптируя во времени результирующий метод к избранной траектории не только по ее качественным характеристикам, но и по требованиям к точности ее наблюдения. Заметим, кроме того, что значение ω , используемое в приведенных выше неравенствах, обладает, например, следующими свойствами: $\omega = \min \| (A - \eta E)(Ay + b) \|$, $\omega = 0$ при $y = u'(t)$.

Один из вариантов алгоритма автоматического выбора шага численного интегрирования, основанный на методах вида (26), (31) — (33) порядка точности $p > 1$, может опираться на неравенства $h_e \leq \sqrt{(p+1)! \varepsilon / p \omega}$, $h \leq \min(h_e, 1/\|A\|)$, $k \leq \log_2((p+1)! \varepsilon \|A\|^2 / p \omega)$, $\tau = 2^k h$.

Предложенный подход к построению адаптивного типа вычислительных алгоритмов может быть дополнен способом выбора порядка точности используемого численного метода, в определенном смысле оптимального на данном отрезке интегрирования. Например, если значения шага $\tau = 2^k h$ ограничены лишь заданными точками наблюдения избранной траектории системы вида (2) или размером окрестности точки t независимой переменной, в которой свойства интересующего нас решения нелинейной системы (1) адекватно отражаются посредством соответствующей линеаризованной системы вида (2), то обеспечение допустимой величины ε евклидовой нормы главной части локальной ошибки метода может быть достигнуто за счет выбора базового модуля соответствующего порядка точности p при заданных значениях параметров

$$k > \log_2(\tau \|A\|) \text{ и } h = 2^{-k} \tau < 1/\|A\| \quad (41)$$

с использованием неравенства

$$\frac{p}{(p+1)!} \leq \frac{\varepsilon}{2^k h^2 \omega} \quad (42)$$

3. Вычислительные алгоритмы, основанные на пошаговой корректировке спектральных свойств разностного оператора. Рассмотрим сначала случай однородной системы вида (2) с симметричной отрицательно-определенной матрицей A . Тогда, как известно [5], евклидова норма вектора $u(t)$ монотонно убывает во времени. Выполнение аналогичного свойства для разностных решений традиционные явные методы (например, типа Рунге — Кутты) обеспечивают, как уже отмечалось выше, лишь при весьма сильных ограничениях на шаг сетки. Так, для явной схемы Эйлера известно условие [1, с. 103]

$$\tau < -2/\lambda_n \quad (\lambda_n \leq \lambda_{n-1} \leq \dots \leq \lambda_1 < 0), \quad (43)$$

которое при большой жесткости исходной системы обычно становится весьма обременительным. Заметим, однако, что условие (43) обеспечивает выполнение обсуждаемого свойства для всего семейства разностных решений, доставляемых явной схемой Эйлера, тогда как пользователя, как правило, интересует лишь вполне конкретная разностная траектория, определяемая начальными данными. Легко видеть, что этот факт позволяет ослабить ограничение (43).

Действительно, неравенство $\|\hat{y}\|^2 < \|y\|^2$ приводит к условию вида $\tau < -2/\mu_1(y)$, которое, вообще говоря, является более слабым по сравнению с (43), так как отношение Рэлея $\mu_1(y)$ лежит в пределах спектра матрицы A . Более того, в силу монотонного возрастания $\mu_1(u(t))$ вдоль дифференциальной траектории исходной системы можно надеяться, что разностные схемы, обеспечивающие выполнение аналогичного свойства для отношения Рэлея вдоль приближенного решения, будут работать при более естественных ограничениях на шаг сетки.

Покажем, как эти требования можно реализовать, прежде всего отметив, что основанный на пошаговом выделении точно обрабатываемой части дифференциального оператора подход [4] позволяет получить явные методы, для которых свойство монотонного убывания евклидовой нормы приближенного решения выполняется при любых шагах сетки дискретизации.

Выбрав в качестве обрабатываемой части исходного дифференциального оператора вектор $u'(t) - \alpha_1 u(t)$, где α_1 — постоянный на отрезке $[t, t + \tau]$ числовой параметр, запишем точное равенство

$$u(t + \tau) = \exp(\alpha_1 \tau) u(t) + \tau \int_0^1 (A - \alpha_1 E) u(t + \beta_1 \tau) \exp(\alpha_1 \tau (1 - \beta_1)) d\beta_1. \quad (44)$$

Заменяв интеграл по квадратурной формуле левых прямоугольников, построим метод первого порядка точности

$$\hat{y} = \exp(\alpha_1 \tau) y + \tau \exp(\alpha_1 \tau) z, \quad z = (A - \alpha_1 E) y. \quad (45)$$

Полагая $\alpha_1 = \mu_1(y)$ (что обеспечивает, в частности, выполнение свойства $\|z\|_1^2 = (-Az, z) = \min_{\eta} \|Ay - \eta y\|_1^2$), получим разностную схему, обладающую свойством точности на решениях вида $u(t) = u^i(t) = c_i \xi^i \exp(\lambda t)$ (см. (4)). Кроме того, в этом случае

$$\begin{aligned} \exp(-2\alpha_1 \tau) (\hat{y}, \hat{y}) &= (y, y) (1 - 2\alpha_1 \tau + \alpha_1^2 \tau^2) + 2\tau (Ay, y) - \tau^2 (Ay, Ay) \leq \\ &\leq \exp(-2\alpha_1 \tau) (y, y), \end{aligned}$$

что подтверждает безусловное убывание евклидовой нормы данного приближенного решения. В то же время можно показать, что требование монотонного возрастания отношения Рэлея $\mu_1(y)$ уже накладывает ограничения на выбор шага численного интегрирования.

Заметим в качестве обобщения, что выбор $\alpha_1 = \mu_k(y)$ в (45) приводит к методу с аналогичными свойствами, причем $\|y\|_0$ заменяется $\|y\|_{k-1} = ((-A)^{k-1} y, y)^{1/2}$.

Для дальнейшего улучшения свойств конструируемых методов применим описанный выше прием выделения точно обрабатываемой части дифференциального оператора еще раз на отрезке длиной в один шаг дискретизации. Тогда аналогично (44) можно записать равенство

$$\begin{aligned} u(t + \beta_1 \tau) &= \exp(\alpha_2 \beta_1 \tau) u(t) + \beta_1 \tau \int_0^1 (A - \alpha_2 E) u(t + \beta_1 \beta_2 \tau) \times \\ &\times \exp(\alpha_2 \beta_1 \tau (1 - \beta_2)) d\beta_2, \end{aligned}$$

где α_2 — постоянный на отрезке $[t, t + \beta_1 \tau]$ числовой параметр.

Подставив последнее соотношение в (44), получим обобщенное интегральное соотношение

$$\begin{aligned} u(t + \tau) &= \exp(\alpha_1 \tau) u(t) + \tau \exp(\alpha_1 \tau, \alpha_2 \tau) (A - \alpha_1 E) u(t) + \\ &+ \tau^2 \int_0^1 \beta_1 e^{\alpha_1 \tau (1 - \beta_1)} \int_0^1 (A - \alpha_1 E) (A - \alpha_2 E) u(t + \beta_1 \beta_2 \tau) e^{\alpha_2 \beta_1 \tau (1 - \beta_2)} d\beta_2 d\beta_1. \quad (46) \end{aligned}$$

Заметим, что в (46) через $\exp(x_1, x_2)$ обозначена разделенная разность функции $\exp(x)$.

Простейшей разностной схемой, которую можно получить на основе равенства (46) путем отбрасывания интегрального слагаемого, является схема

$$\hat{y} = \exp(\alpha_1 \tau) y + \exp(\alpha_1 \tau, \alpha_2 \tau) z, \quad (47)$$

имеющая первый порядок точности при любых значениях параметров α_1 и α_2 . Если же в (47) положить $\alpha_1 = \mu_k(y)$, а $\alpha_2 = \mu_k(z)$, то соответствующий метод, как и метод (45), будет обладать свойством точности на решениях вида (4), а также будет обеспечивать монотонное убывание нормы $\|y\|_{k-1}$ приближенного решения и монотонное возрастание обобщенного отношения Рэлея $\mu_k(y)$ при любых шагах численного интегрирования.

Дальнейшего улучшения свойств численных методов, построенных на пути выделения точно обрабатываемой части дифференциального оператора, можно достичь применением принципа корректировки их спектральных свойств [7]. Например, в случае метода (45) выберем $\alpha_1 = \mu_0(y)$ и введем корректирующий множитель ρ следующим образом:

$$\hat{y} = \exp(\alpha_1 \tau) y + \tau \rho \exp(\alpha_1 \tau) z. \quad (48)$$

Непосредственно проверяется, что при $0 \leq \rho \leq 1$ численный метод (48) сохраняет все обсуждавшиеся выше свойства метода (45).

Значение корректирующего множителя ρ предлагается выбирать таким образом, чтобы обеспечить монотонное возрастание отношения Рэлея $\mu_0(y)$ вдоль приближенного решения. Последнее равносильно выполнению неравенства

$$\text{tr} \sigma(y) \geq -1, \quad (49)$$

где

$$\sigma(y) = 0,5 \mu_1(y) \frac{\mu_2(y) - \mu_0(y)}{\mu_1(y) - \mu_0(y)} - \mu_0(y). \quad (50)$$

Если, например, корректирующий множитель ρ выбрать в виде

$$\rho = 1/(1 + \nu \tau), \quad (51)$$

где $\nu \geq 0$ — числовой параметр, то метод (48), (51), как и метод (45), будет иметь первый порядок точности. Если, кроме того, выбор параметра ν в (51) подчинить требованию $\nu \geq \max(-\sigma(y), 0)$ или

$$\nu \geq \max(-\sigma(y) - 1/\tau, 0), \quad (52)$$

то неравенство (49) будет выполнено при любых значениях τ . Использование ограничения (52) позволяет, в частности, в случае естественной необходимости дробления шага сетки автоматически обращать корректирующий множитель ρ в единицу, так как монотонное возрастание отношения Рэлея $\mu_0(y)$ при этом обеспечивается за счет малости шага τ .

Рассмотренный здесь подход к построению алгоритмов несложным образом распространяется и на методы высших порядков точности (см., например, [10]). При этом, конечно, можно использовать и отличные от (51) способы задания корректирующего множителя ρ . В частности, подобным образом на основе (48) может быть получена разностная схема (47).

Рассмотрим далее неоднородную систему вида (2) с симметричной отрицательно определенной матрицей A . Очевидно, что изложенные выше результаты могут быть непосредственно перенесены на случай подобных систем с помощью замены $\tilde{u}(t) = u(t) - u^*$, где $u^* = -A^{-1}b$ — положение равновесия системы (2). Здесь же мы рассмотрим подход к конструированию численных методов решения таких систем, не использующий, вообще говоря, вектора u^* .

Выбрав в качестве обрабатываемой части дифференциального оператора вектор $u'(t) - \mu(u(t) - y^*)$, где y^* — некоторый постоянный на отрезке $[t, t + \tau]$ вектор, можем записать точное равенство

$$u(t + \tau) = y^* + e^{\mu\tau}(u(t) - y^*) + \tau \int_0^1 (Au(t + \alpha\tau) + b - \mu u(t + \alpha\tau) + \mu y^*) e^{\mu\tau(1-\alpha)} d\alpha,$$

на основе которого способом, аналогичным рассмотренному выше, можно построить численный метод первого порядка точности

$$\hat{y} = y^* + e^{\mu\tau}(1 - \tau\rho\mu)(y - y^*) + \tau\rho e^{\mu\tau}(Ay + b), \quad (53)$$

где корректирующий множитель ρ имеет, скажем, вид (51).

Потребуем, чтобы вектор y^* в каждый момент времени t аппроксимировал положение равновесия u^* системы (2). Простейшей формулой, позволяющей достичь подобного, может служить, например, выражение

$$y^* = y - w/\mu_0(w), \quad (54)$$

где для нахождения вектора $w \approx u'(t)$ предлагается применять соответствующий численный метод вида (48), (51).

Таким образом, один из возможных вариантов метода первого порядка точности для систем вида (2) с матрицей $A = A^T < 0$ может быть записан (см. (53) с учетом (54), а также (48), (50) — (52)) в виде

$$\hat{y} = y + \frac{\exp(\mu\tau) - 1}{\mu} \omega + \tau \exp(\mu\tau) (Ay + b - \omega),$$

$$\hat{\omega} = \exp(\mu\tau) \omega + \tau \exp(\mu\tau) (A - \mu E) \omega, \quad (55)$$

$$\mu = \mu_0(\omega), \quad \rho = \frac{1}{1 + \nu\tau}, \quad \nu \geq \max(-\sigma(\omega) - 1/\tau, 0),$$

где $\hat{\omega} \approx u'(t + \tau)$, а в начальный момент времени $\omega = Ay + b$.

4. **Результаты численных экспериментов.** В заключение приведем здесь некоторые результаты численного эксперимента. Первые две таблицы содержат данные по решению системы вида (2) с матрицей A , ортогонально подобной матрице $\Lambda = -\text{diag}(\lambda_5, \lambda_4, \lambda_3, \lambda_2, \lambda_1)$, и вектором b , удовлетворяющим условию $-A^{-1}b = (1, 1, 1, 1, 1)^T$. Задача Коши решалась ($\varepsilon = 1.e-5$) на отрезке $[0, 0; 10, 0]$, начальный вектор выбран равным $u(0) = (1, 68; 1, 68; 1, 68; 0, 72; -0, 88)^T$, а ортогональная матрица подобия позаимствована из [11, с. 66].

В табл. 1 представлены результаты решения начальной задачи при $\lambda_5 = 10^6$, $\lambda_4 = 10^4$, $\lambda_3 = 10$, $\lambda_2 = 5$, $\lambda_1 = 1$ адаптивного типа вычислительным алгоритмом, основанным на явном методе вида (26), (31) — (33) при $p = 1$ и формулах (34), (38) — (40) (метод 1), и неявным методом Эйлера с автоматическим выбором шага τ на основе оценки главной части локальной погрешности (метод 2). Заметим, что неявный метод Эйлера удовлетворяет при любых значениях шага τ всем выдвинутым в рассматриваемом случае требованиям согласованности дифференциальной и разностной задач, за исключением свойства точности на решениях вида (4). При этом также вид оператора перехода на шаге τ в методе 2, конечно, фиксирован и не адаптируется во времени к избранной траектории.

Таблица 1

Метод	t_n	k_c	δ	Δ	Время, с
1	10.0	150	2.47e-10	9.99e-04	126.8
2	10.0	1723	1.91e-05	1.09e-03	1078.5

В табл. 2 представлены результаты решения начальной задачи при $\lambda_5 = 10^5$, $\lambda_4 = 10^3$, $\lambda_3 = 10^2$, $\lambda_2 = 10$, $\lambda_1 = 1$ адаптивного типа вычислительным алгоритмом переменного порядка точности, основанным на явных методах вида (26), (31) — (33) и формулах (38), (41), (42) (метод 1), и алгоритмом переменного порядка точности [12], основанным на формулах дифференцирования назад (метод 2).

Таблица 2

Метод	t_n	k_c	p	δ	Время, с
1	0.0001	1	6	1.31e-06	4.2
	0.0100	2	7	6.96e-08	11.0
	0.1000	3	4	2.04e-08	17.7
	1.0000	4	3	1.52e-08	25.1
	10.0000	5	1	1.76e-11	32.4
2	0.0001	53	4	8.65e-06	27.0
	0.0100	105	5	4.39e-06	48.8
	0.1000	139	5	5.01e-06	62.0
	1.0000	183	5	2.06e-06	80.2
	10.0000	218	5	1.97e-06	93.6

В табл. 3 приведены результаты решения однородной системы вида (2) возникающей в результате применения метода прямых ($n = 25$) к однородной задаче уравнения теплопроводности с крайними условиями первого

Тест	Метод	k_c	δ	Δ	Время, с
1	1	229	4.54e-13	4.64e-04	246.5
	3	2404	3.00e-06	6.20e-03	10292.1
2	2	714	8.51e-04	1.22e-03	744.8
	3	2404	3.00e-06	6.20e-03	10408.4

рода и начальным условием $u(x, 0) = 100x(1-x)$ (тест 1), а также аналогичной неоднородной системы, в которой вектор b выбирался из условия $-A^{-1}b = (1, 1, \dots, 1)^T$ (тест 2). Задачи решались на отрезке $[0, 0; 2, 0]$ с помощью алгоритмов (48), (50) — (52) (метод 1), (55), (50) (метод 2) и неявного метода Эйлера (метод 3) с выбором шага численного интегрирования на основании оценки локальной ошибки путем сравнения по координатной близости решений, полученных с шагами τ и $\tau/2$ (начальный шаг $\tau_0 = 1.e-5$, $\varepsilon = 1.e-5$).

В таблицах использованы следующие обозначения: t_m — точка наблюдения на отрезке интегрирования; k_c — число шагов интегрирования τ ; p — текущий порядок метода в точке t_m ; $\delta = \|u(t_m) - y(t_m)\|_c$; $\Delta = \max \|u(t_j) - y(t_j)\|_c$, $1 \leq j \leq N$.

Литература

1. Самарский А. А. Теория разностных схем. М., 1977.
2. Деккер К., Вервер Я. Устойчивость методов Рунге — Кутты для жестких нелинейных дифференциальных уравнений. М., 1988.
3. Ракитский Ю. В., Устинов С. М., Черноруцкий И. Г. Численные методы решения жестких систем. М., 1979.
4. Бобков В. В. // Дифференц. уравнения. 1983. Т. 19, № 7. С. 1115—1122.
5. Бобков В. В. // Дифференц. уравнения. 1985. Т. 21, № 7. С. 1117—1126.
6. Бобков В. В. // Вестн. Бел. ун-та. Сер. 1. Физика, математика, механика. 1987. № 2. С. 72—74.
7. Бобков В. В. // Вестн. Бел. ун-та. Сер. 1. Физика, математика, механика. 1981. № 3. С. 61—65.
8. Крылов В. И., Бобков В. В., Монастырский П. И. Начала теории вычислительных методов. Дифференциальные уравнения. Минск, 1982.
9. Бобков В. В., Мандрик П. А. // Дифференц. уравнения. 1989. Т. 25, № 7. С. 1171—1176.
10. Бобков В. В., Мандрик П. А., Репников В. И. // Вестн. АН БССР. Сер. физ.-мат. наук. 1989. № 5. С. 10—14.
11. Марчук Г. И., Шайдулов В. В. Повышение точности разностных схем. М., 1979.
12. Hindmarsh A. C. GEAR: ordinary differential equations system solver // Rept. UCID-30001. Rev. 3. Lawrence Livermore Lab., Univ. of California. 1972.