Аннотация

Решение жестких систем реакций свертывания крови и моделирование образования тромба в придатке левого желудочка

Бутаков Иван Дмитриевич

При патологиях в сердце характер течения в придатке предсердия левого желудочка меняется, повышается риск образования в нем тромба. Для моделирования процесса образования тромба требуется решать систему переноса-диффузии-реакции, где реакционная часть представлена жёсткой системой каскада свёртывания крови. Применение традиционных численных схем при интегрировании данной системы может вести к неустойчивости, нефизичным осцилляциям, к отсутствию сходимости итерационных методов решения возникающих нелинейных уравнений. В данной работе предложены два метода для неявного численного интегрирования жёстких нелинейных систем, способные решить указанные проблемы: модифицированный метод Ньютона и взвешенный метод Эйлера. Методы основаны на «фильтрации» спектра матрицы Якоби правой части системы. Фильтрация производится путём комбинирования явного и неявного метода Эйлера с матричным весовым коэффициентом, и позволяет разделить части спектра, имеющие разный знак действительной части. Весовая матрица вычисляется путём применения специально подобранной функции к спектру матрицы Якоби правой части. Подбор функции осуществлён с целью получения экспоненциального интегратора. В ходе работы полученные методы были проверены на следующих жёстких системах: модель Лотки-Вольтерры, модель осциллятора Ван-дер-Поля, модель каскада свёртывания крови. Во всех случаях предложенные методы показали улучшение устойчивости в сравнении со стандартными подходами к аппроксимации: неявным методом Эйлера и методом трапеций. В дополнение к этому модифицированный метод Ньютона позволил кратно увеличить шаг интегрирования системы по времени, что критически важно для

практических задач. В работе также описаны детали программной реализации предложенных методов. Наконец, описано дальнейшее применение предложенных методов к практической задаче.

Abstract

Solving stiff blood coagulation system and modeling clot formation in left atrial appendage

Оглавление

Глава 1

Введение

Написание данной работы мотивировано необходимостью решать жёсткие нелинейные уравнения реакции при моделировании образования тромба в придатке предсердия левого желудочка. В качестве реакционной части выступает существенно упрощённая модель каскада свёртывания крови, описываемая девятью компонентами [1]. Каскад свёртывания крови — это жёсткая система дифференциальных уравнений, имеющая пороговый отклик на изменение параметров модели [31]. Это вынуждает использовать малый шаг по времени при неявном численном интегрировании [17]. В свою очередь, это приводит к непрактично большому времени расчёта упомянутой модели. С целью увеличения допустимого шага по времени в работе [25] были предложены некоторые модификации реакционной части модели. Данные модификации заключались в замене нескольких компонент, дающих большой вклад во внедиагональную часть матрицы Якоби, на их экстраполированные значения. В данной работе мы фокусируемся исключительно на реакционной части модели, исследуя некоторые обобщения описанного выше метода. Предложенный в настоящей работе численный метод основан на замечании, что матрица Якоби правой части уравнения реакции обладает седловой структурой, которая проявляет себя при численном интегрировании с большим шагом по времени. Мы рассматриваем одношаговые схемы интегрирования по времени, чтобы в дальнейшем встроить полученный метод в полностью неявный интегратор уравнений переноса-диффузии.

Согласно теореме Канторовича [15; 24], для липшицевых в окрестности корня функ-

ций метод Ньютона локально сходится с квадратичной скоростью. В случае уравнений, возникающих при решении жёстких систем, односторонняя константа Липшица может оказаться сколь угодно большой, и сходимость ухудшается [6; 7]. Среди методов по улучшению сходимости метода Ньютона можно перечислить линейный поиск [4; 32], метод доверительных областей [29] и методы ускорения [3; 9; 23]. Линейный поиск минимизирует невязку вдоль выбранного направления путём подбора оптимального шага. Метод доверительных областей изменяют направление шага, используя информацию о производных высшего порядка. Методы ускорения используют историю шагов при решении задачи оптимизации. Возможна также комбинация упомянутых методов [10]. Квазиньютоновские методы активно используются для решения уравнений, возникающих при интегрировании жёстких систем [2; 8; 22; 27]. Данная группа методов решает задачу оптимизации или поиска корней уравнения, используя аппроксимации производных, а не их точные значения. Все эти методы отличаются необходимым количеством вычислений невязки, якобиана или гессиана в ходе поиска решения. Первый подход заключается в улучшении сходимости метода Ньютона в случае неявного метода Эйлера численного интегрирования. Предложенный способ можно рассматривать как вариант квазиньютоновского метода, где на каждой итерации используется модифицированная матрица Якоби. В настоящей работе предложен способ получения модифицированной матрицы Якоби, основанный на решении вспомогательной линеаризованной задачи, возникающей на каждом шаге Ньютона и связанной с построением численной схемы специального вида, соответствующей экспоненциальному интегратору.

Много работ посвящено устойчивости и выбору численных схем [6; 11; 12; 18]. Среди популярных схем можно перечислить метод трапеций, семейство многостадийных методов Рунге-Кутты, формулу дифференцирования назад, методы Розенброка и многие другие. Простейшим методом является явный метод Эйлера, но он имеет малую область абсолютной устойчивости. Неявный метод Эйлера обладает гораздо большей областью абсолютной устойчивости, однако требует решения нелинейного уравнения на каждом шаге. Метод трапеций — арифметическое среднее между явным и неявным методом Эйлера — всё еще достаточно простая схема, дающая, однако, хорошие результаты для некоторых жёстких систем [5]. В настоящей работе строится аналогичная

схема, где явная и неявная части комбинируются при помощи весовой матрицы. Данная матрица зависит от производной правой части системы дифференциальных уравнений и подбирается так, чтобы итоговая схема давала экспоненциальный интегратор.

В работе также поднимается вопрос определения понятия жёсткости системы дифференциальных уравнений. С этой целью приведены основные положения линейной и нелинейной теории устойчивости, взятые из работ [11; 12; 16; 35]. В частности, рассмотрена обобщённая линейная задача Коши с ограниченным линейным оператором, действующим в банаховом пространстве. Приведены спектральные признаки устойчивости, а также формально доказан набор утверждений, связывающих область устойчивости численного метода и асимптотические свойства операторной экспоненты. Это позволяет ввести понятие линейной жёсткости и связать с ним линейную теорию устойчивости. Результаты нелинейной теории устойчивости позволяют обобщить это понятие на произвольные системы. В работе, однако, показано, что проблемы устойчивости интегрирования систем не всегда связаны только с линейной жёсткостью. Поэтому также предлагается ввести понятие нелинейной жёсткости, которое можно связать со сложностью оптимизационных задач, возникающих при использовании неявных численных методов.

Глава 2

Теоретические сведения

Прежде чем перейти к описанию численных методов, остановимся подробнее на теории жёстких систем дифференциальных уравнений. В данном разделе приведены элементы линейной и нелинейной теории устойчивости численных методов. Отметим, что данная теория гораздо более полно описана в работах [16; 35].

В дальнейшем часто будет рассматриваться задача Коши вида

$$\begin{cases} \dfrac{d\mathbf{x}}{dt} = f(t, \mathbf{x}), \\ \mathbf{x}(0) = \mathbf{x}_0, \end{cases} \qquad \mathbf{x} \in \mathbb{R}^d, \qquad f : [0; T] \times D \to \mathbb{R}^d,$$

где $D \subseteq \mathbb{R}^d$. Поскольку любую систему обыкновенных дифференциальных уравнений можно свести к автономной, запись задачи можно упростить:

$$\begin{cases} \dfrac{d\mathbf{x}}{dt} = f(\mathbf{x}), \\ \mathbf{x}(0) = \mathbf{x}_0 \end{cases} \tag{2.1}$$

Раскладывая $f(\mathbf{x})$ в ряд Тейлора в окрестности $\mathbf{x}_0$ и пренебрегая членами, содержащими производные порядка выше второго, можно получить линеаризованную задачу Коши:

$$\begin{cases} \dfrac{d\mathbf{x}}{dt} = f_0 + F_0 \cdot (\mathbf{x} - \mathbf{x}_0), \\ \mathbf{x}(0) = \mathbf{x}_0, \end{cases} \tag{2.2}$$

где $\frac{\partial f}{\partial \mathbf{x}}\big|_{\mathbf{x}_0} = F(\mathbf{x}_0) \equiv F_0$ — матрица Якоби правой части уравнения в точке $\mathbf{x}_0$. Для линеаризованной задачи известно точное решение:

$$\mathbf{x}(t) = \mathbf{x}_0 + (\exp(t \cdot F_0) - I)F_0^{-1} \cdot f_0 \tag{2.3}$$

Стоит отметить, что выражение $(\exp(A) - I)A^{-1}$ можно определить для всех матриц, так как

$$\varphi_1(z) = \begin{cases} \dfrac{e^z - 1}{z}, & z \neq 0 \\ 1, & z = 0 \end{cases}$$

— регулярная в $\mathbb{C}$ функция. Для этого введём следующее определение:

Определение 2.1. *Пусть A — линейный ограниченный оператор, действующий в банаховом пространстве $\mathcal{B}$ над $\mathbb{C}$, $\sigma(A)$ — спектр оператора A, а f — регулярная в области $U \supset \sigma(A)$ функция. Тогда можно определить*

$$f(A) = \frac{1}{2\pi i} \int\limits_{\Gamma} f(\xi) \left(\xi I - A\right)^{-1} d\xi,$$

где Γ — произвольный гладкий контур в U такой, что $\sigma(A)$ целиком находится по левую сторону при положительном обходе Γ.

Корректность данного определения доказана в [30]. Также справедливо следующее замечание:

Замечание 2.2. *Пусть в условиях определения 2.1 оператор A диагонализуем: $A = V\Lambda V^{-1}$, $\Lambda = \operatorname{diag}(\lambda_\alpha)$. Тогда $f(A) = V f(\Lambda) V^{-1} = V \operatorname{diag}\left(f(\lambda_\alpha)\right) V^{-1}$.*

2.1 Линейная теория устойчивости

При численном решении систем дифференциальных уравнений неизбежны неточности и численные погрешности, возмущения. Поведение систем и численных методов под действием данных возмущений во многом определяет возможность получения точного численного решения, а также вычислительные ресурсы, которые для этого потребуются. Линейная теория устойчивости численных методов подходит к данному вопросу со стороны линеаризации динамической системы в окрестности некоторой точки или решения. В рамках данной теории рассматривается поведение численного метода при решении уравнения (2.2). В случае невырожденной F_0 при помощи линейной замены $\mathbf{x} := \mathbf{x} - \mathbf{x}_0 + F_0^{-1} f_0$ задача сводится к проверочному *уравнению Далквиста* [12]:

$$\begin{cases} \dfrac{d\mathbf{x}}{dt} = F_0 \cdot \mathbf{x}, \\ \mathbf{x}(0) = \mathbf{x}_0 \end{cases} \qquad \Longleftrightarrow \qquad \mathbf{x}(t) = \exp(t \cdot F_0) \cdot \mathbf{x}_0 \tag{2.4}$$

Для линейных (возможно, многошаговых) численных схем в случае $f(\mathbf{x}) = F_0 \cdot \mathbf{x}$ один шаг можно записать в виде

$$\mathbf{x}^{n+1} = R(\Delta t \cdot F_0) \cdot \mathbf{x}^n, \tag{2.5}$$

где $R(z)$ — *функция устойчивости*, а $R(\Delta t \cdot F_0)$ — *матрица перехода*. Обычно $R(z)$ рассматривается как функция комплексного переменного, и потом естественным образом обобщается на матричный аргумент (см. 2.1). По умолчанию далее так и будет подразумеваться. Однако в силу специфики данной работы, понятие функции устойчивости будет иногда обобщаться, сразу предполагая матричный характер (то есть прообраз среди функций комплексного переменного может отсутствовать).

Приведём несколько примеров линейных численных схем и их функций устойчивости. Эти примеры понадобятся при дальнейшем анализе.

Явный метод Эйлера:	$\dfrac{\mathbf{x}^{n+1} - \mathbf{x}^n}{\Delta t} = f(\mathbf{x}^n)$	$R(z) = 1 + z$
Неявный метод Эйлера:	$\dfrac{\mathbf{x}^{n+1} - \mathbf{x}^n}{\Delta t} = f(\mathbf{x}^{n+1})$	$R(z) = \dfrac{1}{1 - z}$
Метод трапеций:	$\dfrac{\mathbf{x}^{n+1} - \mathbf{x}^n}{\Delta t} = \dfrac{1}{2} f(\mathbf{x}^n) + \dfrac{1}{2} f(\mathbf{x}^{n+1})$	$R(z) = \dfrac{2 + z}{2 - z}$

В общем случае двухшаговой одностадийной схемы:

$$\frac{\mathbf{x}^{n+1} - \mathbf{x}^n}{\Delta t} = (1 - M) f(\mathbf{x}^n) + M f(\mathbf{x}^{n+1}) \qquad R(z) = (1 - Mz)^{-1}(1 + (1 - M)z), \tag{2.6}$$

где M, вообще говоря, может быть матрицей (в таком случае $R(\cdot)$ следует изначально рассматривать как матричную функцию, которая, однако, при определённом выборе M может иметь аналог среди функций комплексного переменного[1]). Если же M — число, то данная схема называется *θ-методом*.

2.1.1 Линейная устойчивость

Рассмотрим формально вопросы затухания аналитического и численного решения с течением времени. Для этого потребуются следующие определения и теоремы:

[1] В частности, далее будет показано, что выбором M можно достичь $R(A) = \exp(A)$, из чего следует $R(z) = e^z$.

Определение 2.3. *Пусть $\sigma(A) \subseteq \mathbb{C}$ — спектр линейного оператора A. Число $r(A) = \sup\{|\lambda| \mid \lambda \in \sigma(A)\}$ называется* спектральным радиусом *линейного оператора A, а $s(A) = \sup\{\operatorname{Re}\lambda \mid \lambda \in \sigma(A)\}$ —* спектральной границей.

Теорема 2.4 (об отображении спектра; [30], утверждение 2.8)**.** *Пусть A — линейный ограниченный оператор, действующий в банаховом пространстве $\mathcal{B}$ над $\mathbb{C}$, f — регулярная в области $U \supset \sigma(A)$ функция. Тогда*

$$\sigma(f(A)) = f(\sigma(A))$$

Здесь и далее под нормой оператора будет подразумеваться норма, подчинённая норме линейного пространства, в котором действует оператор: $\|A\| = \sup\limits_{\|x\|=1} \|Ax\|$.

Теорема 2.5 (формула Бёрлинга-Гельфанда)**.** *Пусть A — линейный ограниченный оператор, действующий в банаховом пространстве $\mathcal{B}$ над $\mathbb{C}$. Тогда $r(A) = \lim\limits_{n\to\infty} \|A^n\|^{\frac{1}{n}} = \inf\limits_{n\in\mathbb{N}} \|A^n\|^{\frac{1}{n}}$.*

Следствие 2.6. *Пусть A — линейный ограниченный оператор, действующий в банаховом пространстве $\mathcal{B}$ над $\mathbb{C}$. Тогда*

$$\forall\varepsilon > 0 \;\; \exists n_0 \in \mathbb{N} : \; \forall n > n_0 \quad (r(A))^n \leqslant \|A^n\| < (r(A) + \varepsilon)^n$$

Доказательство. По теореме 2.5

$$\forall\varepsilon > 0 \;\; \exists n_0 \in \mathbb{N} : \; \forall n > n_0 \quad \left| \|A^n\|^{\frac{1}{n}} - r(A) \right| < \varepsilon,$$

причём $\forall n \in \mathbb{N} \;\; (r(A))^n = r(A^n) \leqslant \|A^n\|$. Тогда

$$\forall\varepsilon > 0 \;\; \exists n_0 \in \mathbb{N} : \; \forall n > n_0 \quad r(A) \leqslant \|A^n\|^{\frac{1}{n}} < r(A) + \varepsilon,$$

откуда получаем искомое неравенство. □

Лемма 2.7. *Пусть A — линейный ограниченный оператор, действующий в банаховом пространстве $\mathcal{B}$ над $\mathbb{C}$. Тогда*

$$s(A) = \inf\left\{\omega \in \mathbb{R} : \lim_{t\to+\infty} e^{-\omega t} \cdot \|\exp(t \cdot A)\| = 0\right\}$$

Доказательство. Обозначим $T(t) = \exp(t \cdot A)$. Тогда

$$T(t) = \exp((n\Delta t + \tau) \cdot A) = (T(\Delta t))^n \cdot T(\tau),$$

где $\Delta t > 0$, $n = \lfloor t/\Delta t \rfloor$, $\tau = t - n\Delta t \in [0; \Delta t)$. В силу ограниченности A имеем $\|T(t)\| \leqslant e^{t \cdot \|A\|}$. Согласно теореме 2.4, $\sigma(T(t)) = \exp(t \cdot \sigma(A))$, из чего следует $r(T(t)) = e^{t \cdot s(A)}$ (так как $|e^z| = e^{\operatorname{Re} z}$, а экспонента — монотонно возрастающая на $\mathbb{R}$ функция). Применяя следствие 2.6 и учитывая, что

$$C^{-1} = e^{-\Delta t \cdot \|A\|} \leqslant \|T(-\tau)\|^{-1} = \|T(\tau)^{-1}\|^{-1} \leqslant \|T(\tau)\| \leqslant e^{\Delta t \cdot \|A\|} = C,$$

получаем, что $\forall \varepsilon > 0 \;\; \exists n_0 \in \mathbb{N} : \;\; \forall n > n_0$

$$C^{-1} \cdot e^{n\Delta t \cdot s(A)} \leqslant \|T(t)\| \leqslant C \cdot e^{n\Delta t \cdot (s(A) + \varepsilon)}$$

Наконец, так как $n\Delta t = t - \tau$, $\forall \varepsilon > 0 \;\; \exists t_0 > 0 : \;\; \forall t > t_0$

$$K^{-1} \cdot e^{t \cdot s(A)} \leqslant \|T(t)\| \leqslant K \cdot e^{t \cdot (s(A) + \varepsilon)},$$

где $K = C \cdot e^{\Delta t \cdot |s(A)|} = e^{\Delta t \cdot (\|A\| + |s(A)|)} > 0$. Из этого следует, что $\forall \varepsilon > 0$

$$\lim_{t \to +\infty} e^{-t \cdot s(A)} \cdot \|T(t)\| \geqslant K^{-1} > 0, \qquad \lim_{t \to +\infty} e^{-t \cdot (s(A) + \varepsilon)} \cdot \|T(t)\| \leqslant \lim_{t \to +\infty} K \cdot e^{-t \cdot \varepsilon / 2} = 0$$

Отсюда по определению точной верхней грани получаем доказываемое утверждение. □

Следствие 2.6 позволяет на основе данных о спектре линейного оператора A оценить асимптотику $\|A^n\|$, а лемма 2.7 — $\|\exp(t \cdot A)\|$.

Вернёмся к проверочному уравнению Далквиста (2.4). Нас интересуют равномерные оценки на норму численного решения, получаемого заданным методом при заданном постоянном шаге интегрирования. Для этого введём следующее определение и утверждение:

Определение 2.8. *Множество* $\mathbf{R} = \{z \in \mathbb{C} \mid |R(z)| < 1\}$ *называется* областью абсолютной устойчивости *численного метода, обладающего функцией устойчивости* $R(z)$. *Множество* $\overline{\mathbf{R}} = \{z \in \mathbb{C} \mid |R(z)| \leqslant 1\}$ *— замыкание области абсолютной устойчивости.*[2]

[2]Здесь неявно предполагается непрерывность $R(z)$. Можно обойтись без этого условия, однако тогда некорректно называть $\overline{\mathbf{R}}$ замыканием $\mathbf{R}$.

Утверждение 2.9. *Пусть численное решение уравнения* (2.4) *ищется интегрированием с постоянным шагом* Δt *при помощи метода, обладающего функцией устойчивости* $R(z)$ *и соответсвующей областью абсолютной устойчивости* $\mathbf{R}$. *Тогда* $\mathbf{x}^n = (R(\Delta t \cdot F_0))^n \cdot \mathbf{x}_0$. *Пусть также* $R(z)$ *регулярна в окрестности* $\Delta t \cdot \sigma(F_0)$. *Тогда выполнено*

$$\Delta t \cdot \sigma(F_0) \subseteq \mathbf{R} \qquad \Longleftrightarrow \qquad \| \left(R(\Delta t \cdot F_0)\right)^n \| \underset{n\to\infty}{\longrightarrow} 0$$

$$\Delta t \cdot \sigma(F_0) \subseteq \mathbb{C} \setminus \overline{\mathbf{R}} \qquad \Longrightarrow \qquad \| \left(R(\Delta t \cdot F_0)\right)^n \| \underset{n\to\infty}{\longrightarrow} \infty$$

Доказательство. В силу (2.5) имеем первое утверждение: $\mathbf{x}^n = (R(\Delta t \cdot F_0))^n \cdot \mathbf{x}_0$. Далее заметим, что по теореме 2.4

$$\sigma\left(R(\Delta t \cdot F_0)\right) = R\left(\Delta t \cdot \sigma(F_0)\right)$$

Отсюда следует, что

$$\Delta t \cdot \sigma(F_0) \subseteq \mathbf{R} \qquad \Longleftrightarrow \qquad r\left(R(\Delta t \cdot F_0)\right) < 1$$

$$\Delta t \cdot \sigma(F_0) \subseteq \mathbb{C} \setminus \overline{\mathbf{R}} \qquad \Longleftrightarrow \qquad r\left(R(\Delta t \cdot F_0)\right) > 1$$

Наконец, применяя следствие 2.6, завершаем доказательство утверждения. □

В теории линейной устойчивости важная роль отводится *A-устойчивости* — свойству численного решения проверочного уравнения Далквиста не возрастать по норме, если не возрастает норма истинного решения. Если к тому же при увеличении шага интегрирования норма численного решения на следующей итерации (или, быть может, через некоторе заранее известное число итераций) также стремится к нулю, то говорят об *L-устойчивости*. Дадим формальное определение.

Определение 2.10. *Численный метод называется* A-устойчивым *в случае, если* $\mathbb{C}^- \equiv \{z \in \mathbb{C} \mid \operatorname{Re} z < 0\} \subseteq \mathbf{R}$.

Определение 2.11. *Численный метод называется* L-устойчивым *в случае, если он A-устойчив и выполнено* $\lim\limits_{\operatorname{Re} z \to -\infty} R(z) = 0$.

Утверждение 2.12. *Пусть* $R(z)$ *регулярна в* $\mathbb{C}^-$. *Соответствующий численный метод A-устойчив тогда и только тогда, когда* $\forall F_0$ *выполнено*

$$\| \exp(t \cdot F_0) \| \underset{t\to+\infty}{\longrightarrow} 0 \qquad \Longrightarrow \qquad \forall \Delta t > 0 \quad \| (R(\Delta t \cdot F_0))^n \| \underset{n\to\infty}{\longrightarrow} 0$$

Доказательство. В силу леммы 2.7 импликацию из утверждения можно переписать в виде

$$\sigma(F_0) \subseteq \mathbb{C}^- \quad \Longrightarrow \quad \forall \Delta t > 0 \quad \|(R(\Delta t \cdot F_0))^n\| \underset{n\to\infty}{\longrightarrow} 0$$

Если также воспользоваться 2.9, получаем

$$\sigma(F_0) \subseteq \mathbb{C}^- \quad \Longrightarrow \quad \forall \Delta t > 0 \quad \Delta t \cdot \sigma(F_0) \subseteq \mathbf{R},$$

что при произвольном F_0 равносильно $\mathbb{C}^- \subseteq \mathbf{R}$. Это даёт определение 2.10. □

Утверждение 2.13. *Пусть $R(z)$ регулярна в $\mathbb{C}^-$. Если соответствующий численный метод L-устойчив, то $\forall F_0$ выполнено*

$$\|\exp(t \cdot F_0)\| \underset{t\to+\infty}{\longrightarrow} 0 \qquad \Longrightarrow \qquad r\left(R(\Delta t \cdot F_0)\right) \underset{\Delta t\to+\infty}{\longrightarrow} 0$$

Доказательство. Аналогично доказательству утверждения 2.12 перепишем импликацию в виде

$$\sigma(F_0) \subseteq \mathbb{C}^- \quad \Longrightarrow \quad r\left(R(\Delta t \cdot F_0)\right) \underset{\Delta t\to+\infty}{\longrightarrow} 0$$

В силу теоремы 2.4 это эквивалентно

$$\sigma(F_0) \subseteq \mathbb{C}^- \quad \Longrightarrow \quad \sup\{|\lambda| \mid \lambda \in R(\Delta t \cdot \sigma(F_0))\} \underset{\Delta t\to+\infty}{\longrightarrow} 0,$$

что при произвольном F_0 равносильно $\forall z \in \mathbb{C}^- \ |R(\Delta t \cdot z)| \underset{\Delta t\to+\infty}{\longrightarrow} 0$. Это верно для L-устойчивых методов по определению 2.11. □

Стоит отметить, что утверждение 2.13, в отличие от 2.12, сформулировано в форме признака, а не критерия. Также в нём получено лишь утверждение о пределе спектрального радиуса, а не нормы. По сути, это означает, что получаемая для L-устойчивого метода матрица перехода с увеличением размера шага становится «почти нильпотентной». Если F_0 диагонализуема, то из стремления к нулю спектрального радиуса матрицы перехода будет автоматически следовать и стремление к нулю её нормы.

2.1.2 Логарифмическая норма

Выше были приведены основные результаты линейной теории устойчивости. Все они в значительной степени опираются на спектральный анализ. С одной стороны, это облегчает исследование свойств численных методов, так как теорема об отображении спектра позволяет по функции устойчивости метода определить асимптотические свойства

матрицы перехода. С другой стороны, спектральный анализ не позволяет получить оценки, справедливые в течение всего рассматриваемого времени, включая начальный этап эволюции линейной системы.

Для получения более строгих в указанном смысле оценок вводится *логарифмическая норма*:

Определение 2.14. *Пусть A — линейный ограниченный оператор, действующий в банаховом пространстве $\mathcal{B}$ над $\mathbb{C}$. Число*

$$\mu[A] = \lim_{h \to +0} \frac{\|I + h \cdot A\| - 1}{h}$$

называется логарифмической нормой *оператора A.*

Далее приведём без доказательства несколько утверждений, связывающих логарифмическую норму с теорией линейной устойчивости. Все утверждения взяты из работ [16; 28].

Утверждение 2.15. *Пусть A, B — линейные операторы, действующие в конечномерном банаховом пространстве $\mathcal{B}$ над $\mathbb{C}$. Тогда верно следующее:*

1. $\mu[A]$ *определена.*
2. $\mu[A] \leqslant \|A\|$.
3. $\mu[\gamma \cdot A] = \gamma \cdot \mu[A]$ *для любого* $\gamma > 0$.
4. $\mu[A + z \cdot I] = \mu[A] + \operatorname{Re} z$.
5. $\mu[A + B] \leqslant \mu[A] + \mu[B]$.
6. $s(A) \leqslant \mu[A]$.
7. $\|\exp(t \cdot A)\| \leqslant e^{t \cdot \mu[A]}$ *для любого* $t \geqslant 0$.

Пункт 6 связывает логарифмическую норму со спектральными свойствами A. Пункт 7 даёт справедливую всё неотрицательное время оценку на оператор эволюции, получаемый из A, но, в силу пункта 6, а также леммы 2.7, данная оценка может не являться асимптотически оптимальной.

2.1.3 Линейная жёсткость

Как видно из приведённых результатов, спектр матрицы F_0 может задавать определённые ограничения на шаг интегрирования. Действительно, если $\mu[F_0] < 0$, но численный метод не A-устойчив, полученное с его помощью решение может вести себя некорректно при некоторых F_0 и Δt: если $\Delta t \cdot \sigma(F_0) \not\subseteq \mathbf{R}$, то численное решение может возрастать, в то время как норма точного решения ограничена сверху убывающей экспонентой. С другой стороны, такая ситуация невозможна независимо от Δt при использовании A-устойчивых методов. Но A-устойчивость не гарантирует соизмеримую с точным решением скорость затухания численного; возможен даже случай $\lim\limits_{\mathrm{Re}\, z \to -\infty} |R(z)| = 1$, что приводит к слабо затухающим осцилляциям численного решения вокруг нуля при сравнительно быстром стремлении к нулю истинного решения. Если требуется рост скорости затухания за конечное число шагов при увеличении Δt, следует пользоваться L-устойчивыми методами.

На рисунках 2.1, 2.2 проиллюстрировано поведение явного метода Эйлера (не A-устойчивый), метода трапеций (A-устойчивый, но не L-устойчивый) и неявного метода Эйлера (L-устойчивый) при разных значениях $\Delta t \cdot F_0$ в одномерной задаче Далквиста.

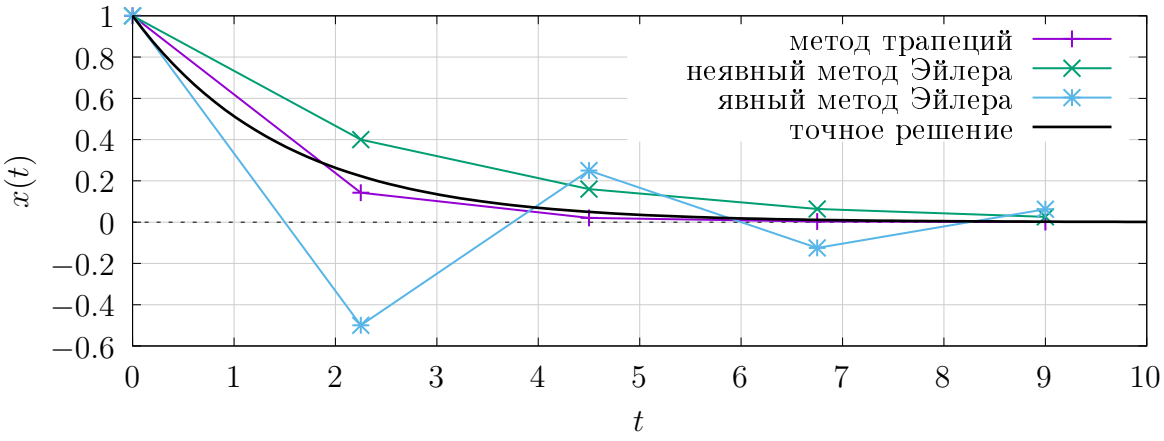

Рис. 2.1: Поведение простейших численных методов при решении одномерного уравнения Далквиста ($\Delta t \cdot F_0 = -1.5$)

Зачастую область устойчивости не A-устойчивых методов ограничена (в частности, это верно для всех явных линейных численных методов) или содержит лишь некоторый подсектор $\mathbb{C}^-$, поэтому ограничение на шаг интегрирования оказывается ограничени-

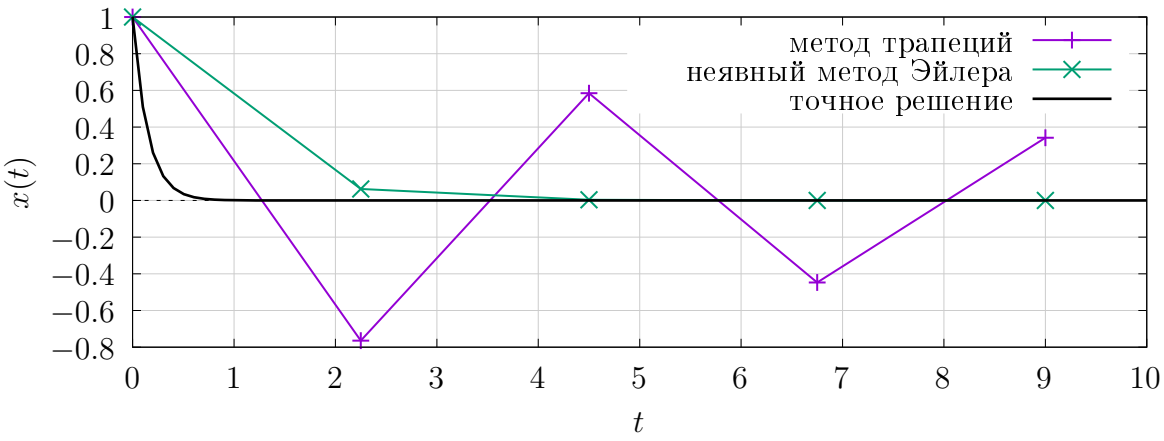

Рис. 2.2: Поведение простейших численных методов при решении одномерного уравнения Далквиста ($\Delta t \cdot F_0 = -15$)

ем сверху. Таким образом, спектральные свойства F_0 обуславливают максимально допустимый шаг численного интегрирования. Ситуацию осложняет следующая теорема, требующая в случае одностадийных схем делать выбор между устойчивостью, высоким порядком аппроксимации и линейностью схемы:

Теорема 2.16 (второй барьер Далквиста)**.** *Не существует A-устойчивых линейных многошаговых одностадийных схем с порядком аппроксимации выше второго.*

Ограничение на Δt может сохранятся даже при решении нелинейных задач вида (2.1). Исходя из вышеизложенного анализа можно ввести следующие определения характерных масштабов времени:

Определение 2.17. *Обозначим характерное время изменения* $F = \frac{\partial f}{\partial \mathbf{x}}(\mathbf{x}(t))$ *как* τ_{nonlin}, *а* $1/r(F)$ *— характерное время реакции линеаризованной системы на небольшие возмущения — как* τ_{lin}.

Рассмотрим случай, когда $\tau_{\text{lin}} \ll \tau_{\text{nonlin}}$. Тогда линеаризация (2.2) остаётся достаточно точной гораздо дольше τ_{lin}. Это автоматически оставляет в силе ограничения на шаг интегрирования, полученные для линейных систем. В частности, если $\mu[F] < 0$, но $\Delta t \cdot \sigma(F) \not\subseteq \mathbf{R}$, численное решение может вести себя неустойчиво к небольшим возмущениям, в то время как точное решение, наоборот, будет обладать эффектом демпфирования.

Приведённые выше рассуждения показывают, что определённые системы дифференциальных уравнений могут обладать свойствами, вынуждающими использовать малый шаг интегрирования при их решении недостаточно устойчивыми в смысле 2.10 и 2.11 методами. Традиционно такие системы называют *жёсткими*. Как указано в [16; 35], существует несколько определений жёсткости, каждое из которых обладает своими достоинствами и недостатками. Приведём здесь одно из них:

Определение 2.18. *Система вида* $\frac{d\mathbf{x}}{dt} = f(t, \mathbf{x})$ *называется* жёсткой *в том случае, если для получения корректного (то есть в заданной степени близкого к точному решению как качественно, так и количественно) численного решения необходимо использовать шаг интегрирования, много меньший характерных масштабов времени, на которых меняется точное решение.*

Данное определение слишком общее и не отвечает на вопросы о природе ограничения на шаг интегрирования. На основе всего вышеизложенного анализа мы дадим более узкое, но в некоторой степени и более информативное определение жёсткости.

Определение 2.19. *Система вида* $\frac{d\mathbf{x}}{dt} = f(\mathbf{x})$ *называется* линейно жёсткой *в том случае, если характерное время реакции линеаризованной системы на небольшие возмущения* τ_{lin} *много меньше характерных масштабов времени, на которых ищется решение.*

Зачастую численное решение ищется на интервалах, сопоставимых с характерными временны́ми масштабами наиболее медленно протекающих процессов, описываемых системой (это действительно так, если в численном решении требуется полноценно отобразить всю динамику системы). Также для численных методов с ограниченной областью устойчивости условие $\Delta t \cdot \sigma(F) \subseteq \mathbf{R}$ влечет $\Delta t \sim \tau_{\text{lin}}$. Тогда 2.19 оказывается частным случаем 2.18, причём необходимость выбора малого шага оказывается обусловленной «жёстким» линейным поведением системы в окрестности истинных решений. В таком случае некорректность численного решения понимается в смысле неустойчивости к малым возмущениям там, где точное решение к ним устойчиво.[3]

[3]Здесь подразумевается устойчивость точного решения к небольшим возмущениям начальных условий в области, где проявляется линейная жёсткость.

Стоит также отметить, что в случае линейных или слабо нелинейных (то есть линеаризацию которых можно долго считать достаточно точной) систем характерные масштабы времени наиболее медленно протекающих процессов задаются наименьшими по модулю элементами спектра матрицы F. Тогда для линейно жёстких систем спектр F распадается на две части: *ведущую* и *паразитическую*.

Определение 2.20. *Будем называть спектр оператора* A распадающимся на ведущую и паразитическую части, *если существует разбиение* $\sigma(A) = \sigma_d(A) \sqcup \sigma_p(A)$ *такое, что*

$$r_d(F) \equiv \sup\{|\lambda| \mid \lambda \in \sigma_d(F)\} \ll \inf\{|\lambda| \mid \lambda \in \sigma_p(F)\} \equiv b_p(F)$$

Более того, в случае седловых задач доля паразитического спектра может иметь положительную действительную часть. Это может вызывать «взрывное» поведение даже A-устойчивых методов, ведь в таком случае попадание $\Delta t \cdot \sigma(F)$ в область устойчивости не гарантируется. В задачах, где известно, что такое поведение точного решение нефизично и отсутствует в виду выбора правильных начальных условий, данное свойство численного решения является нежелательным.

2.1.4 Экспоненциальные интеграторы

Линейный анализ устойчивости известен давно, а потому достаточно подробно разработан [6; 11; 12; 16; 18; 35]. Как следствие, велик и арсенал тех методов, которые способны в той или иной степени решить проблемы, вызываемые линейной жёсткостью. В основном это различные многошаговые или многостадийные линейные схемы, обладающие теми или иными свойствами линейной устойчивости: неявные методы Рунге-Кутты, методы Гира, формулы дифференцирования назад и прочее. Обладая достаточной устойчивостью, они одновременно могут иметь высокий порядок аппроксимации. Тем не менее, их область устойчивости меньше области устойчивости неявного метода Эйлера, что мотивирует разработку новых методов.

В настоящей работе предлагается отойти от стандартных линейных схем с постоянными коэффициентами и обратить внимание на численные методы, утилизирующие информацию о матрице Якоби правой части системы (2.1). Она может предоставить

достаточный объём информации о локальном линейном поведении системы. Эта информация полезна в том числе и для решения проблемы линейной жёсткости. Стоит добавить, что, при решении жёстких систем обычно используются неявные методы. Возникающие при этом алгебраические уравнения зачастую решаются методом Ньютона. Это означает, что матрица Якоби правой части вычисляется в любом случае, поэтому её использование для модификации численной схемы не приносит дополнительной вычислительной сложности (кроме, быть может, последующих операций с матрицей).

Методы, использующие матрицу Якоби правой части интегрируемой системы называются *адаптивными*. Среди них можно перечислить методы Розенброка, Обрешкова, а также *экспоненциальные интеграторы*. В данной работе основное внимание будет уделено именно последним.

Определение 2.21. *Экспоненциальным интегратором называется численный метод, дающий применительно к линейной задаче* (2.2) *точное решение.*

Замечание 2.22. *Любой экспоненциальный интегратор обладает функцией устойчивости* $R(z) = e^z$. *Он также является L-устойчивым.*

Из определения и замечания видно, что экспоненциальный интегратор точно интегрирует линейную часть системы. Это позволяет исключить влияние линейной жёсткости на решение. Преимуществом экспоненциальных интеграторов также является полное отсутствие численной (то есть обусловленной исключительно методом, а не задачей) диссипации в линейных системах. В неоторых задачах, близких к линейным, это качественно отличает данные методы от стандартных A-устойчивых схем с большими областями устойчивости (пример можно увидеть в разделе 4.1).

Некоторые авторы также расширяют определение экспоненциальных интеграторов на все методы, использующие экспоненту матрицы Якоби правой части системы. Обзор существующих экспоненциальных интеграторов можно найти, например, в работах [14; 21]. Мы же остановимся только на простейших из них, необходимых для дальнейшего анализа.

Рассмотрим явный и неявный экспоненциальные методы Эйлера:

$$\textit{Явный:} \qquad \mathbf{x}^{n+1} - \mathbf{x}^n = \left(\exp(\Delta t \cdot F) - I\right) F^{-1} \cdot f(\mathbf{x}^n) \tag{2.7}$$

$$\textit{Неявный:} \qquad \mathbf{x}^{n+1} - \mathbf{x}^n = \left(I - \exp(-\Delta t \cdot F)\right) F^{-1} \cdot f(\mathbf{x}^{n+1}) \tag{2.8}$$

Здесь, как и ранее, $F = \frac{\partial f}{\partial \mathbf{x}}$ — матрица Якоби правой части системы.[4] При этом за пользователем остаётся свобода выбора точки, в которой вычисляются производные. На самом деле, возможно даже использование приближённого значения матрицы.

Зная (2.3), несложно проверить, что при рассмотрении линейной задачи $f(\mathbf{x}) = f_0 + F(\mathbf{x}_0) \cdot (\mathbf{x} - \mathbf{x}_0)$ оба метода дают точное решение. Следует отметить, что коэффициент при f в обоих методах может расти экспоненциально быстро с увеличением Δt. В таком случае с увеличением шага численное решение может быстро выйти за пределы окрестности, в которой применима линеаризация правой части. Это следует отнести к недостаткам данных методов, так как такое поведение потенциально может являться источником неустойчивости в случае сильной нелинейности правой части. Пример такому поведению будет дан в следующих разделах. В качестве одного из результатов работы также будет приведён двухточечный экспоненциальный интегратор, построенный по аналогии с (2.7) и (2.8), но лишённый упомянутого недостатка.

2.2 Нелинейная теория устойчивости

В предыдущем разделе были приведены основные положения линейной теории устойчивости, введено понятие жёсткости и линейной жёсткости, упомянуты адаптивные численные методы, использующие знание матрицы Якоби правой части системы для борьбы с линейной жёсткостью. Вместе с этим, в общих чертах обозначены предполагаемые границы применимости линейной теории устойчивости:

1. Линейный анализ устойчивости некорректно применять при наличии возмущений, величина или характер которых ставит под сомнение точность линеаризации системы.

[4] Отметим, что мы явно не указываем, в какой именно точке вычисляется F. Для линейных систем это не важно, но для нелинейных является определённой степенью свободы, заложенной в методе.

2. Линейный анализ устойчивости некорректно применять, если характерное время линейной реакции на малые возмущения сопоставимо с или больше характерных масштабов времени, на которых меняется матрица Якоби правой части системы.

В главе 7 работы [16] приведены примеры систем, применение линейного анализа к которым даёт некорректные результаты (по крайней мере, если ограничиваться лишь спектральными признаками и не использовать методы из раздела 2.1.2). В этой связи в настоящем разделе рассмотрены основные результаты нелинейной теории устойчивости. Вместе с тем также приведены некоторые сведения, касающиеся поведения неявных численных методов при интегрировании существенно нелинейных систем.

2.2.1 Нелинейная устойчивость

В линейном анализе устойчивости существенная роль отводится линейным системам, решения которых стремятся к нулю при устремлении времени в бесконечность. Такое поведение также означает, что любые два решения с течением времени «сближаются» друг с другом. В секции 2.1.3 этому важному феномену дана интерпретация: при небольших возмущениях начальных условий решение будет стремится к невозмущенному с течением времени. При использовании устойчивых в различных смыслах методов это в той или иной степени гарантирует аналогичное поведение и у численного решения. Поэтому при появлении сомнений в корректности линейного анализа устойчивости логично попытаться обобщить определение данного явления на произвольную систему.

Определение 2.23. *Пусть* $\mathbf{x}(t)$ *и* $\mathbf{y}(t)$ *— решения системы* $\frac{d\mathbf{x}}{dt} = f(t, \mathbf{x})$. *Данные решения называются* сжимающимися *на отрезке* $[a; b]$ *в случае*

$$\forall t_1, t_2 : \ a \leqslant t_1 \leqslant t_2 \leqslant b \qquad \|\mathbf{x}(t_2) - \mathbf{y}(t_2)\| \leqslant \|\mathbf{x}(t_1) - \mathbf{y}(t_2)\|$$

Аналогичное определение можно ввести и для численных решений (и, вообще говоря, для любых последовательностей):

Определение 2.24. *Пусть* $\{\mathbf{x}_n\}_{n\in\mathbb{N}_0}$ *и* $\{\mathbf{y}_n\}_{n\in\mathbb{N}_0}$ *— последовательности элементов некоторого линейного нормированного пространства. Данные последовательности называются* сжимающимися *на отрезке* $[a; b]$ *в случае*

$$\forall n_1, n_2 : \ a \leqslant n_1 \leqslant n_2 \leqslant b \quad \|\mathbf{x}_{n_2} - \mathbf{y}_{n_2}\| \leqslant \|\mathbf{x}_{n_1} - \mathbf{y}_{n_2}\|$$

Для систем вида $\frac{d\mathbf{x}}{dt} = f(t, \mathbf{x})$ известно [6; 7] достаточное условие сжимаемости точных решений.

Определение 2.25. *Пусть $\mathcal{H}$ — гильбертово пространство над $\mathbb{C}$. Функция $f(t, \mathbf{x}) = f : \mathbb{R} \times D \to \mathcal{H}$ (где $D \subseteq \mathcal{H}$ — выпуклая область) и система $\frac{d\mathbf{x}}{dt} = f(t, \mathbf{x})$ называются* односторонне липшициевыми *на отрезке $[a; b]$ в случае*

$$\exists \nu(t) : \ \forall \mathbf{x}, \mathbf{y} \in D, \ \forall t \in [a, b] \quad \operatorname{Re} \langle f(t, \mathbf{x}) - f(t, \mathbf{y}), \mathbf{x} - \mathbf{y} \rangle \leqslant \nu(t) \|\mathbf{x} - \mathbf{y}\|^2$$

Определение 2.26. *Пусть в определении 2.25 $\nu(t) \leqslant 0$. Тогда функция $f(t, \mathbf{x})$ и система $\frac{d\mathbf{x}}{dt} = f(t, \mathbf{x})$ называются* диссипативными.

Утверждение 2.27 (достаточный признак сжимаемости)**.** *Пусть система $\frac{d\mathbf{x}}{dt} = f(t, \mathbf{x})$ диссипативна на $[a; b]$. Тогда все её решения являются сжимающимися на $[a; b]$.*

Замечание 2.28. *Линейная функция $f(t, \mathbf{x}) = A(t) \cdot \mathbf{x} + f_0(t)$ односторонне липшициева на всей области определения $A(t)$ с наименьшим возможным коэффициентом $\nu(t) = \mu[A(t)]$.*

Замечание 2.28 позволяет связать линейную теорию устойчивости с нелинейной. Его естественным продолжением является следующая теорема:

Теорема 2.29 (Далквист, 1959)**.** *Пусть $\mathcal{H}$ — гильбертово пространство над $\mathbb{C}$. Пусть $f(t, \mathbf{x}) = f : \mathbb{R} \times D \to \mathcal{H}$, где $D \subseteq \mathcal{H}$ — выпуклая область. Наконец, пусть существуют $a, b \in \mathbb{R}$ и $\nu(t)$ — такая кусочно-непрерывная функция, что*

$$\forall t \in [a; b], \ \forall \mathbf{x} \in D \quad \mu \left[\frac{\partial f}{\partial \mathbf{x}}(t, \mathbf{x}) \right] \leqslant \nu(t)$$

Тогда, если $\mathbf{x}(t)$ и $\mathbf{y}(t)$ — два решения системы $\frac{d\mathbf{x}}{dt} = f(t, \mathbf{x})$, то для любых $t_1, t_2 \in \mathbb{R}$: $a \leqslant t_1 \leqslant t_2 \leqslant b$ выполнено

$$\|\mathbf{x}(t_2) - \mathbf{y}(t_2)\| \leqslant \exp \left(\int_{t_1}^{t_2} \nu(s) ds \right) \|\mathbf{x}(t_1) - \mathbf{y}(t_1)\|$$

Наконец, приведём достаточное условие сжимаемости численных решений.

Теорема 2.30 (Далквист, 1978)**.** *Пусть рассматривается конечномерное гильбертово пространство над $\mathbb{R}$. Если численный метод A-устойчивый, то существует норма (*G-норма*), в которой любые два численных решения диссипативной на $[a; b]$ системы будут сжимающимися на $[a; b]$.*

Сразу отметим, что известны и другие признаки сжимаемости численных решений, среди которых есть и не требующие введение другой нормы. Однако, в силу специфики численных методов, которые будут предложены в данной работе, упомянутые признаки не являются релевантными.

Приведённые теоремы позволяет обосновать линейный анализ устойчивости даже для нелинейных систем. Стоит, однако, отметить, что это возможно исключительно благодаря ослаблению утверждений до формы достаточных условий, а также благодаря использованию логарифмической нормы вместо спектральной границы. Напомним, что первая даёт справедливую в течение всего времени оценку на норму оператора эволюции линейной системы, а вторая — только асимптотическую оценку (пусть и не менее строгую).

2.2.2 Нелинейная жёсткость

В предыдущей секции были приведены важные утверждения, позволяющие формально распространить определение линейной жёсткости на нелинейные системы, а также показывающие, что A-устойчивые методы позволяют в известной степени устранять эффекты линейной жёсткости даже для нелинейных систем.

Практика, однако, показывает, что при решении систем с сильно нелинейной в том или ином смысле правой частью могут возникать различные нежелательные эффекты неустойчивости, даже если используются A-устойчивые, L-устойчивые или адаптивные методы. Данные эффекты также вынуждают ограничивать величину шага. Таким образом, система проявляет свойства жёсткости, которые невозможно объяснить одной только линейной составляющей.

Для примера рассмотрим задачу Коши

$$\begin{cases} \dfrac{dx}{dt} = \cos\left(\dfrac{\pi}{2} \cdot x\right) \\ x(0) = x_0 = 0 \end{cases} \qquad F(x) = \frac{\partial f}{\partial x} = -\frac{\pi}{2} \sin\left(\frac{\pi}{2} \cdot x\right), \tag{2.9}$$

имеющую точное решение

$$x(t) = \frac{2}{\pi} \arcsin\left(\tanh\left(\frac{\pi}{2} t\right)\right)$$

Рассматриваемая система является автономной и имеет множество положений равновесия $x = 2k + 1,\ k \in \mathbb{Z}$. Из них устойчивые только $x = 4k + 1,\ k \in \mathbb{Z}$. В окрестности каждого из положений равновесия система достаточно хорошо линеаризуема (с кубической точностью). Более того, в любой точке функция $f(x)$ отличается от свой линеаризации в ближайшем положении равновесия не более, чем на $|\cos(\pi/2) - \pi/2| \approx 1.58 \approx 1.58 \cdot \max|f(x)|$. Наконец, в любой момент времени точное решение находится в промежутке $[0; 1)$, с экспоненциальной скоростью стремясь к положению равновесия $x = 1$. Для $x \in [0; 1)$ имеем $-\pi/2 < F(x) \leqslant 0$, то есть $F(x) \in \mathbb{C}^-$. Исходя из этого можно выдвинуть предположение, что численное решение данной системы при помощи A-устойчивого метода не вызовет никаких проблем, даже если взять шаг интегрирования, сравнимый с $\min(1/|F|) = 2/\pi \sim \tau_{\text{lin}}$.

Воспользуемся L-устойчивым неявным методом Эйлера. Шаг интегрирования возьмём $\Delta t = 2$, возникающие при интегрировании нелинейные уравнения будем решать методом Ньютона с начальным приближением в текущей точке. Для сравнения построим также график точного решения. Оба графика можно видеть на рисунке 2.3. Несложно заметить, что полученное методом Эйлера решение некорректно. Использование экспоненциальных интеграторов из раздела 2.1.4 также не приносит положительных результатов (см. рис. 2.4), несмотря на то, что данные методы точно интегрируют линейную составляющую системы.

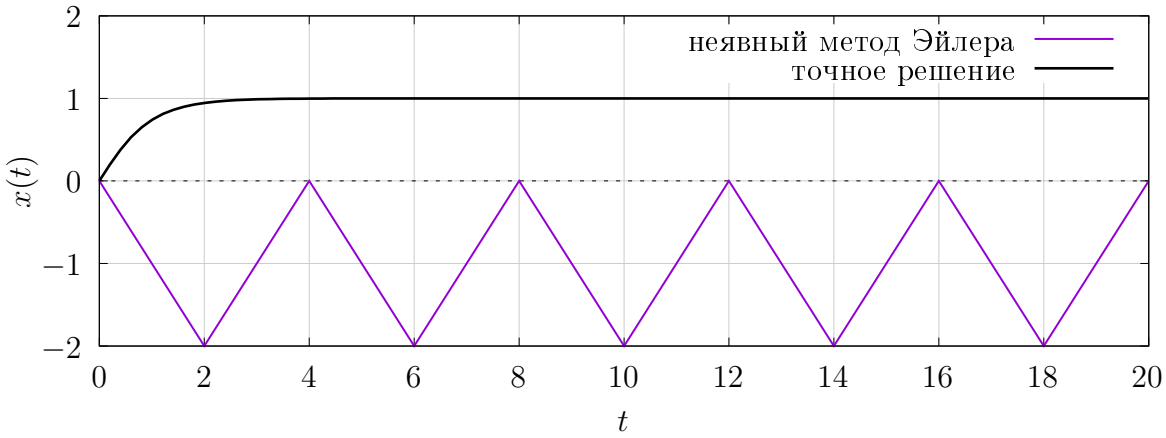

Рис. 2.3: Поведение L-устойчивого метода Эйлера при решении уравнения (2.9)

Данный пример показывает, что явление жёсткости не ограничивается только ли-

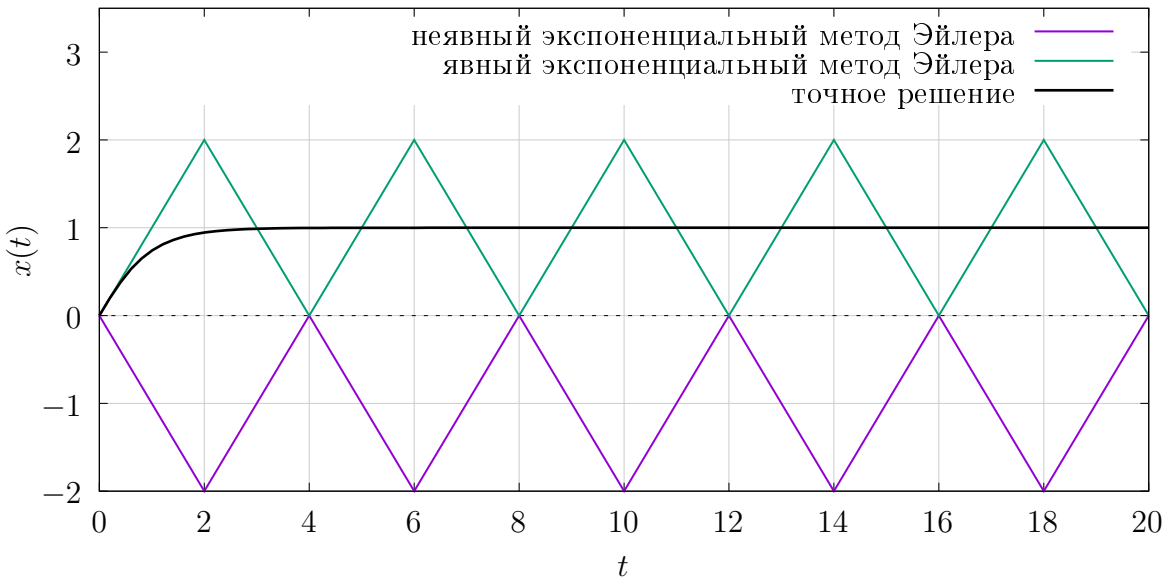

Рис. 2.4: Поведение простейших экспоненциальных интеграторов при решении уравнения (2.9)

нейной её составляющей. Выпишем невязку неявного метода Эйлера и его матрицу Якоби:

$$\mathcal{R}(t^{n+1}, \mathbf{x}^{n+1}) = \mathbf{x}^{n+1} - \mathbf{x}^n - \Delta t \cdot f(t^{n+1}, \mathbf{x}^{n+1}) \tag{2.10}$$

$$\mathcal{J}(t^{n+1}, \mathbf{x}^{n+1}) = \frac{\partial \mathcal{R}}{\partial \mathbf{x}}(t^{n+1}, \mathbf{x}^{n+1}) = I - \Delta t \cdot F(t^{n+1}, \mathbf{x}^{n+1}) \tag{2.11}$$

В нашем случае $f(x)$ и $F(x)$ имеют целое семейство корней. Поэтому при достаточно большом Δt несколько корней также имеет $\mathcal{R}(x)$, а $\mathcal{J}(x)$ перестаёт быть отделимым от нуля. Теорема Канторовича [15; 24] даёт достаточные условия сходимости метода Ньютона к корню уравнения. Одним из условий как раз является отделимость матрицы Якоби от нуля. Таким образом, из-за нелинейности правой части и большого шага интегрирования, с одной стороны, теряются достаточные условия сходимости метода Ньютона, и, с другой стороны, появляются некорректные корни невязки. Приведённый пример показывает, что это может оказаться достаточным для того, чтобы метод выдавал некорректные решения. Как показано в [16], более устойчивый метод простой итерации для поиска корней также может расходиться в случае жёстких систем и большого шага интегрирования. Можно предположить, что это универсальное свойство некоторых систем, заставляющее выбирать между величиной шага и сложностью

алгоритма поиска корней невязок.

Таким образом, мы приходим новому определению жёсткости системы, связанному теперь с её нелинейным характером.

Определение 2.31. *Система вида* $\frac{d\mathbf{x}}{dt} = f(t, \mathbf{x})$ *называется* нелинейно жёсткой *в том случае, если нелинейные свойства правой части существенно влияют на*

- *допустимую величину шага интегрирования для заданного алгоритма поиска корней невязки выбранной численной схемы;*
- *допустимую степень неявности используемой численной схемы для заданного шага интегрирования и алгоритма поиска корней невязки;*
- *сложность алгоритма поиска корней невязки для заданной численной схемы и шага интегрирования,*

при которых возникающая на каждом шаге по времени нелинейная система алгебраических уравнений решается корректно.

Иначе говоря, нелинейная жёсткость проявляется в необходимости выбирать между неявностью численной схемы, большим шагом по времени и простотой метода решения нелинейных уравнений.

Таким образом, борьба с нелинейной жёсткостью переводится в плоскость методов оптимизации. Среди способов улучшения сходимости метода Ньютона можно перечислить линейный поиск [4; 32], метод доверительных областей [29] и различного рода ускорения [3; 9; 23]. Линейный поиск минимизирует невязку вдоль выбранного направления путём подбора оптимального шага. Метод доверительных областей изменяют направление шага, используя информацию о производных высшего порядка. Ускоренные методы используют историю шагов при решении задачи оптимизации. Возможна также комбинация упомянутых методов [10]. Квазиньютоновские методы активно используются для решения уравнений, возникающих при интегрировании жёстких систем [2; 8; 22; 27]. Данная группа методов решает задачу оптимизации или поиска корней уравнения, используя аппроксимации производных, а не их точные значения. Все эти методы отличаются необходимым количеством вычислений невязки, якобиана или гессиана в ходе поиска решения.

Глава 3

Разработка численных методов

В прошлой главе был дан краткий обзор существующих методов оценки устойчивости динамических систем и численных схем. Введены основные элементы линейного и нелинейного анализа устойчивости. Введено понятие жёсткости системы, разделённое в дальнейшем на линейную и нелинейную жёсткость. Было показано, что с линейной жёсткостью эффективно справляются A-устойчивые и L-устойчивые методы, в том числе и экспоненциальные интеграторы, интегрирующие линейную часть системы точно. Наконец, обозначена невозможность решения проблем нелинейной жёсткости в рамках линейной теории устойчивости.

Данный раздел полностью посвящён разработке новых методов численного решения дифференциальных уравнений, потенциально способных показать более устойчивое поведение на нелинейно жёстких системах. В то же время, разрабатываемые методы должны устойчиво интегрировать линейно жёсткие системы.

Первый раздел данной главы посвящён разработке квазиньютоновской модификации стандартной неявной схемы Эйлера. Данная модификация основана на фильтрации спектра матрицы Якоби правой части системы. Во втором разделе рассматривается численная схема, являющаяся взвешенной комбинацией явного и неявного метода Эйлера, и дающая матрицу Якоби, совпадающую с модифицированной матрицей из первого раздела.

3.1 Модифицированный метод Ньютона

Рассмотрим неявный метод Эйлера, невязка и матрица Якоби невязки которого даны в (2.10) и (2.11) соответственно. Для получения значения $\mathbf{x}$ на следующем, $(n+1)$-ом шаге по времени, будем решать в общем случае нелинейное уравнение $\mathcal{R}(\mathbf{x}^{n+1}) = \mathbf{0}$ методом Ньютона:

$$\mathcal{J}_m \cdot \left(\mathbf{x}_{m+1}^{n+1} - \mathbf{x}_m^{n+1}\right) = -\mathcal{R}_m \qquad \Longrightarrow \qquad \mathbf{x}_{m+1}^{n+1} = \mathbf{x}_m^{n+1} - \mathcal{J}_m^{-1} \cdot \mathcal{R}_m,$$

где m — номер нелинейной итерации метода, и используются обозначения

$$\mathcal{R}_m = \mathcal{R}(t^{n+1}, \mathbf{x}_m^{n+1}), \quad \mathcal{J}_m = \mathcal{J}(t^{n+1}, \mathbf{x}_m^{n+1}) = I - \Delta t \cdot F_m,$$

$$F_m = F(t^{n+1}, \mathbf{x}_m^{n+1}) = \frac{\partial f}{\partial \mathbf{x}}(t^{n+1}, \mathbf{x}_m^{n+1})$$

Если F_m имеет седловую структуру, то для достаточно большого Δt матрица $\mathcal{J}_m$ не знакоопределена. Рассмотрим функцию $\theta(z)$ комплексного аргумента, регулярную в области, которая на любой итерации содержит $\Delta t \cdot \sigma(F_m)$. Тогда на каждом шаге определена матрица $M = M_m = \theta(\Delta t \cdot F_m)$. Модифицируем шаг метода Ньютона следующим образом:

$$\mathbf{x}_{m+1}^{n+1} = \mathbf{x}_m^{n+1} - (I - \Delta t \cdot MF_m)^{-1} \cdot \mathcal{R}_m, \tag{3.1}$$

В зависимости от выбора $\theta(z)$ весовая матрица M позволяет «фильтровать» спектр F_m и регулировать его влияние на модифицированную матрицу Якоби $\bar{\mathcal{J}}_m = I - \Delta t \cdot MF_m$. Введём следующий класс функций:

Определение 3.1. *Функцию $\varphi(z)$ назовём* обобщённой функцией знака *в случае, если*

$$\lim_{\operatorname{Re} z \to \pm\infty} \varphi(z) = \pm 1$$

Если рассматривать $\theta(z) = \frac{1}{2}(1 - \varphi(z))$, то с ростом Δt можно ожидать ослабления положительной действительной части спектра $\Delta t \cdot F_m$ при умножении на M. При этом часть спектра с отрицательной действительной частью должна фильтроваться в меньшей степени. Это приближает матрицу $\bar{\mathcal{J}}_m = I - \Delta t \cdot MF_m$ к отрицательно определённой, что потенциально может улучшить глобальную сходимость модифицированного метода Ньютона. Вопрос выбора конкретной функции $\theta(z)$ из представленного семейства будет рассмотрен в следующих секциях.

Затронем вопрос вычисления матрицы M. Диагонализуемые матрицы плотны в множестве квадратных матриц. Соответственно, если $f(t, \mathbf{x})$ не обладает особыми свойствами, логично ожидать, что, как правило, матрица Якоби правой части F будет диагонализуема. Пусть $F = V\Lambda V^{-1}$, $\Lambda = \operatorname{diag}(\lambda_i)$. Тогда, согласно 2.2, $M = \theta(F) = V\theta(\Lambda)V^{-1} = V \operatorname{diag}(\theta(\lambda_i)) V^{-1}$, что уже легко вычисляется, если известны V и Λ. При определённых условиях на θ матрица M всегда будет вещественной.

Утверждение 3.2. *Пусть $f : \mathbb{C} \to \mathbb{C}$ — целая (регулярная во всей $\mathbb{C}$) функция. Она принимает на $\mathbb{R}$ только вещественные значения тогда и только тогда, когда имеет вещественные коэффициенты в разложении в ряд Тейлора в любой точке $\mathbb{R}$.*

Доказательство. Докажем в обе стороны.

$\Rightarrow$ Пусть f принимает только вещественные значения на $\mathbb{R}$. Поскольку она целая, её можно представить в виде сходящегося ряда Тейлора, записанного относительно произвольной точки. Пусть $x_0 \in \mathbb{R}$. Тогда

$$f(x) = \sum_{k=0}^{\infty} \frac{f^{(k)}(x_0)}{k!}(x - x_0)^k$$

Поскольку $f(x)$ принимает только вещественные значения на $\mathbb{R}$, все её производные тоже. Отсюда получаем вещественность коэффициентов в разложении в ряд Тейлора.

$\Leftarrow$ Пусть f имеет вещественные коэффициенты в разложении в ряд Тейлора относительно хотя бы одной точки $x_0 \in \mathbb{R}$. Тогда

$$\forall x \in \mathbb{R} \quad f(x) = \sum_{k=0}^{\infty} c_k (x - x_0)^k,$$

где $c_k \in \mathbb{R}$ для любого k. Значит, значение $f(x)$ совпадает со значением суммы ряда вещественных чисел, а потому вещественно.

□

Лемма 3.3. *Пусть $f : \mathbb{C} \to \mathbb{C}$ — целая функция, принимающая на $\mathbb{R}$ только вещественные значения. Пусть $A \in \mathbb{R}^{d \times d}$. Тогда $f(A)$ — также вещественная матрица.*

Доказательство. Поскольку f — целая,

$$f(A) = \sum_{k=0}^{\infty} c_k \cdot A^k$$

Из утверждения 3.2 имеем, что $\forall k \in \mathbb{N}_0 \ c_k \in \mathbb{R}$. Тогда матрица $f(A)$ равна сумме ряда вещественных матриц, а потому вещественна. □

Следствие 3.4. *Пусть $\theta(z)$ является результатом применения стандартных алгебраических операций (сложение, умножение, обращение) к значениям $f_1(z), \ldots, f_N(z)$, где f_i — целая функция, принимающая на $\mathbb{R}$ только вещественные значения. Пусть $A \in \mathbb{R}^{d\times d}$. Тогда $\theta(A)$ — также вещественная матрица.*

3.2 Взвешенный метод Эйлера

Предложенная в предыдущем разделе модификация является квазиньютоновской. Её побочным эффектом является ухудшение скорости сходимости при поиске корня невязки из-за неточного направления в методе Ньютона. В данной секции вводится численная схема, дающая невязку с той же матрицей Якоби, что получается в предыдущей секции. Рассмотрим модификацию невязки

$$\begin{aligned}\bar{\mathcal{R}}(t^{n+1}, \mathbf{x}^{n+1}) = M \cdot \mathcal{R}(t^{n+1}, \mathbf{x}^{n+1}) + (I - M) \cdot \left(\mathbf{x}^{n+1} - \mathbf{x}^n - \Delta t \cdot f(t^n, \mathbf{x}^n)\right) = \\ = \mathbf{x}^{n+1} - \mathbf{x}^n - \Delta t \cdot \left[M f(t^{n+1}, \mathbf{x}^{n+1}) + (I - M) f(t^n, \mathbf{x}^n)\right],\end{aligned}$$

где $\mathcal{R}$ взята из (2.10). Поскольку суммарный коэффициент при f равен $M+(I-M) = I$, предложенная схема имеет, как минимум, первый порядок аппроксимации.

Несложно заметить, что при фиксированной M итерация метода Ньютона принимает вид, аналогичный (3.1):

$$\mathbf{x}_{m+1}^{n+1} = \mathbf{x}_m^{n+1} - (I - \Delta t \cdot M F_m)^{-1} \cdot \bar{\mathcal{R}}_m, \tag{3.2}$$

Выпишем полностью алгоритм Ньютона для нахождения корней невязки полученной схемы:

- Выбрать начальное приближения $\mathbf{x}_0^{n+1} = \mathbf{x}^n$ и вычислить $f^n = f(t^n, \mathbf{x}^n)$.

- Выполнить m-ую итерацию метода Ньютона:
  - Вычислить значение функции $f_m = f(t^{n+1}, \mathbf{x}_m^{n+1})$ и матрицу производных $F_m = \frac{\partial f}{\partial \mathbf{x}}(t^{n+1}, \mathbf{x}_m^{n+1})$.
  - Вычислить собственное разложение матрицы $F_m = V\Lambda V^{-1}$.
  - Вычислить $\theta(F_m) = V\theta(\Lambda)V^{-1}$ и $\bar{\mathcal{R}}_m = \mathbf{x}_m^{n+1} - \mathbf{x}^n - \Delta t \cdot [M f_m^{n+1} + (I - M) f^n]$.
  - Если $\|\bar{\mathcal{R}}_m\| > \max\left\{\varepsilon_{\text{абс}}, \varepsilon_{\text{отн}} \cdot \|\bar{\mathcal{R}}_0\|\right\}$, то перейти на следующую итерацию по формуле (3.2), иначе $\mathbf{x}^{n+1} = \mathbf{x}_m^{n+1}$ и завершить выполнение.

3.3 Выбор весовой функции

В предыдущих двух разделах были введены два семейства численных методов, основанных на «фильтрации» спектра матрицы Якоби правой части системы посредством её умножения на матрицу M, получаемой применением весовой функции $\theta(z)$ к $\Delta t \cdot F$. В данном разделе будет приведён один из возможных вариантов выбора $\theta(z)$.

Как уже отмечалось в начале настоящей главы, задача построения новых численных методов в данной работе заключается в решении проблемы нелинейной жёсткости. Однако также было упомянуто, что предложенные методы должны удовлетворительно справляться и с эффектами линейной жёсткости. Метод 3.1 обладает той же функцией устойчивости, что и неявный метод Эйлера, а потому уже A-устойчив независимо от выбора $\theta(z)$. С другой стороны, метод 3.2 обладает функцией устойчивости, указанной в (2.6). В общем случае она может не содержать $\mathbb{C}^-$.

Выберем $\theta(z)$ так, чтобы взвешенный метод Эйлера точно интегрировал линейную часть системы, то есть чтобы получился экспоненциальный интегратор. Из (2.6) и замечания 2.22 имеем

$$R(z) = \frac{1 + (1 - \theta(z))z}{1 - \theta(z)z} = e^z$$

Приводя систему, получим

$$e^z(1 - \theta(z)z) = 1 + (1 - \theta(z))z,$$

откуда

$$\theta(z)\,(1 - e^z)\,z = 1 + z - e^z$$

В результате имеем

$$\theta(z) = \begin{cases} \dfrac{1}{z} - \dfrac{1}{e^z - 1}, & z \neq 2\pi i \cdot k,\ k \in \mathbb{Z} \\ \dfrac{1}{2}, & z = 0, \end{cases}$$

где функция сразу доопределена в нуле своим пределом. Она регулярна на всей области определения. Заметим также, что $\theta(z) = \frac{1}{2}(1 - \varphi(z))$, где

$$\varphi(z) = \begin{cases} 1 - \dfrac{2}{z} + \dfrac{2}{e^z - 1}, & z \neq 2\pi i \cdot k,\ k \in \mathbb{Z} \\ 0, & z = 0 \end{cases}$$

— обобщённая функция знака по определению 3.1. Более того,

$$\forall z \in \mathbb{C} \quad 1 - z\theta(z) = \frac{z}{e^z - 1} \neq 0,$$

что гарантирует невырожденность матрицы $\bar{\mathcal{J}}$. Наконец,

$$\theta(z) = \frac{1}{2} - \frac{z}{12} + \frac{z^3}{720} + O(z^4)$$

Таким образом, при $\Delta t \to 0$ взвешенный метод Эйлера с данной весовой функцией переходит в метод трапеций.

Глава 4

Численные эксперименты

В данной главе приведены результаты численных экспериментов, целью которых является проверка работы предложенных методов на примере жёстких систем дифференциальных уравнений. Невязка и матрица Якоби вычисляются при помощи систем автодифференцирования. Решение систем линейных алгебраических уравнений делается при помощи библиотеки INMOST [25]. Диагонализация матрицы производится при помощи библиотеки Eigen [13]. В этой секции референсным решением будет называться численное решение, полученное методом трапеций при использовании малого шага по времени (10^5 точек). Допустимая погрешность в методе Ньютона: $\varepsilon_{\text{абс}} = 10^{-7}$ и $\varepsilon_{\text{отн}} = 10^{-9}$. Максимальное допустимое число ньютоновских итераций: $N = 200$.

4.1 Система Лотки-Вольтерры

Система Лотки-Вольтерры [19] имеет вид

$$\begin{cases} \dfrac{dx}{dt} = (a - by)x \\ \dfrac{dy}{dt} = (-c + dx)y \end{cases}$$

Для численных экспериментов будут использованы параметры $a = 0.3$, $b = 0.01$, $c = 0.3$, $d = 0.3$, и начальные условия $x_0 = 5$, $y_0 = 5$. Время моделирования — $T = 100$. Численное решение искалось для величины шага $\Delta t = 1$ и $\Delta t = 2$. Графики численных решений, а также зависимость числа потребовавшихся итераций метода Ньютона от номера шага приведены на рисунке 4.1.

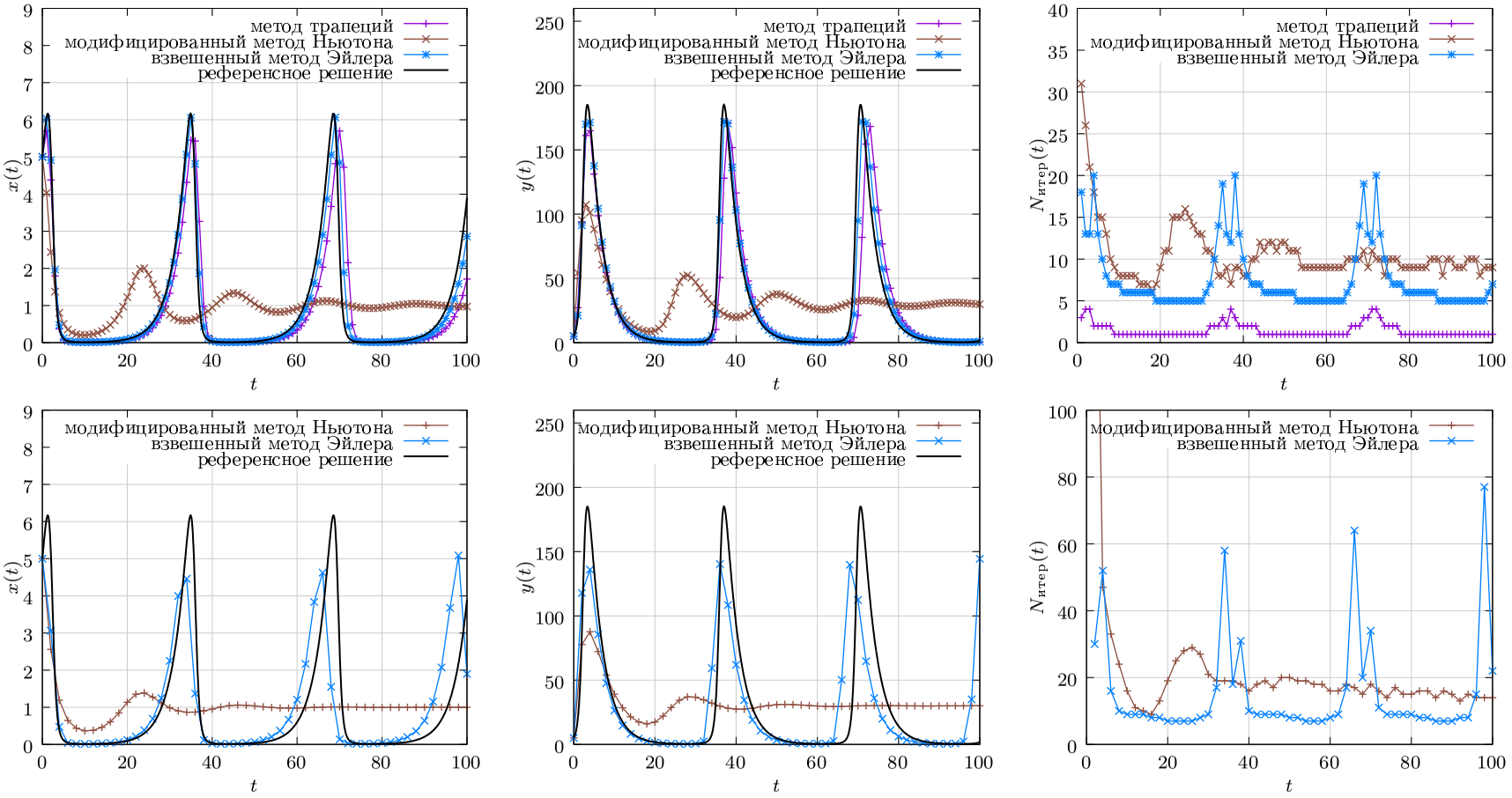

Рис. 4.1: Сравнение методов на примере интегрирования системы Лотки-Вольтерры для шага по времени $\Delta t = 1$ (сверху) и $\Delta t = 2$ (снизу)

Неявный метод Эйлера при использовании стандартного метода Ньютона расходится при $\Delta t \geqslant 1$ из-за отрицательного значения переменной после некоторой итерации. Метод трапеций также расходится при $\Delta t \geqslant 2$ по той же причине. Модифицированный метод Ньютона позволяет убрать проблему сходимости для данных шагов, однако требует для этого значительно большее число итераций. Так как модифицированный метод Ньютона использует ту же невязку, что и численно диссипативный неявный метод Эйлера, решение, полученное с его помощью, сильно затухает и является излишне диссипативным. Взвешенный метод Эйлера позволяет добиться точности, сравнимой с методом трапеций, но за счёт заметно возросшего числа нелинейных итераций.

На рисунке 4.2 приведены графики нормы невязки для метода трапеций, модифицированного метода Ньютона и взвешенного метода Эйлера, возникающие при решении нелинейных систем на некоторых пяти последовательных шагах численного интегрирования системы Лотки-Вольтерры. Хорошо видна седловая структура задачи, которая успешно устраняется взвешенным методом Эйлера.

4.2 Осциллятор Ван дер Поля

Жёсткая система, соответствующая уравнению Ван дер Поля [2], имеет вид

$$\varepsilon \frac{dx}{dt} = y - \left(\frac{1}{3} x^3 - x \right), \qquad \frac{dy}{dt} = -x$$

Параметр ε регулирует жёсткость уравнения. В данном эксперименте он положен равным 10^{-2}. Начальные условия: $x_0 = 0.2$, $y_0 = 0$. Графики численных решений, а также зависимость числа потребовавшихся итераций метода Ньютона от номера шага приведены на рисунке 4.3.

Данная система является значительно жёсткой как в линейном, так и в нелинейном смысле. Неявный метод Эйлера, формула дифференцирования назад второго порядка, а также метод трапеций на некоторой итерации сошлись к физически некорректному корню. Это является признаком нелинейной жёсткости. Также метод трапеций регулярно показывал сильно осциллирующее поведение, что говорит о значительной линейной жёсткости. Оба предложенных метода в то же время дают корректные решения, повторяющие динамику референсного решения, причём без нефизичных осцилляций. Взве-

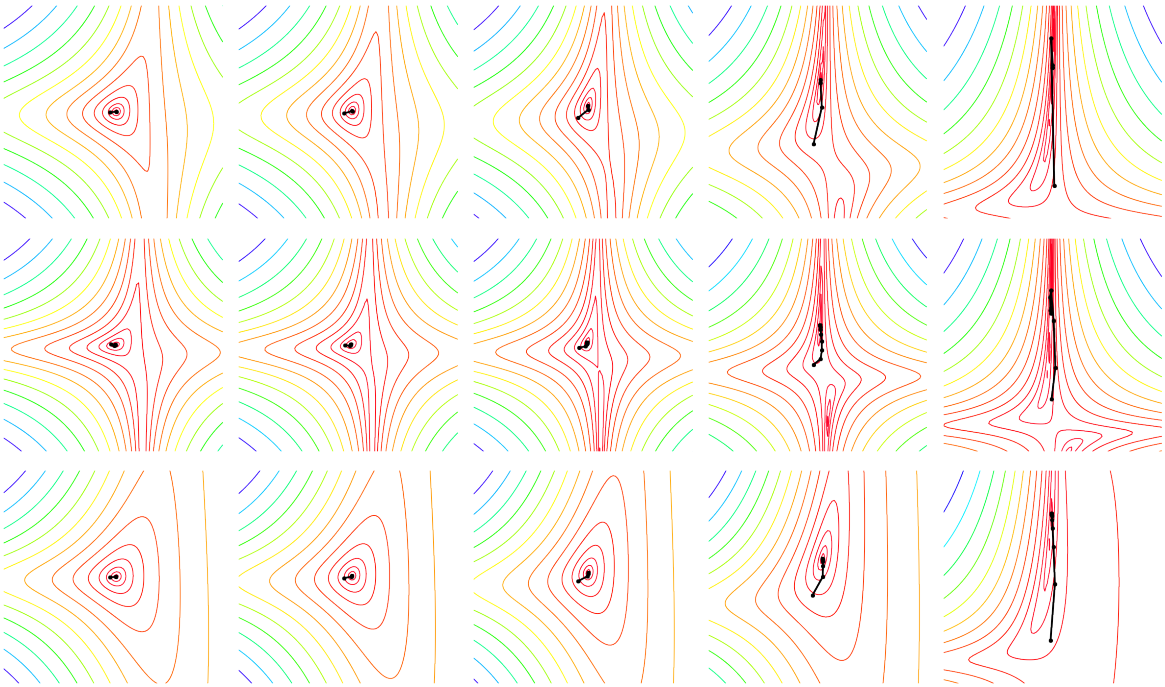

Рис. 4.2: Контуры 2-нормы невязки в зависимости от $x \in [x^* - 15, x^* + 15]$ по горизонтали и $y \in [y^* - 15, y^* + 15]$ по вертикали (где (x^*, y^*) — центр траектории ньютоновских итераций) для метода трапеций (верхний ряд), Эйлера (модифицированный метод Ньютона) (средний ряд) и взвешенного метода Эйлера (нижний ряд) для пяти последовательных шагов по времени ($\Delta t = 1$), система Лотки-Вольтерры. Черная траектория соответствует итерациям Ньютона.

шенный метод Эйлера воспроизводит референсное решение без дисперсионной ошибки, в то время как модифицированный метод Ньютона даёт несколько запаздывающие колебания. Видно, однако, что модифицированный метод Ньютона почти всегда досрочно завершается из-за достижения предельного числа итераций. Взвешенный метод Эйлера также в этом смысле сравнительно затратен, но задачу решает.

В сравнение также были включены *L-устойчивый метод Парески-Руссо* [26] и *метод Кина-Жанга* [20] — два A-устойчивых диагонально-неявных двухстадийных метода Рунге-Кутты второго порядка аппроксимации, задаваемых следующими таблицами

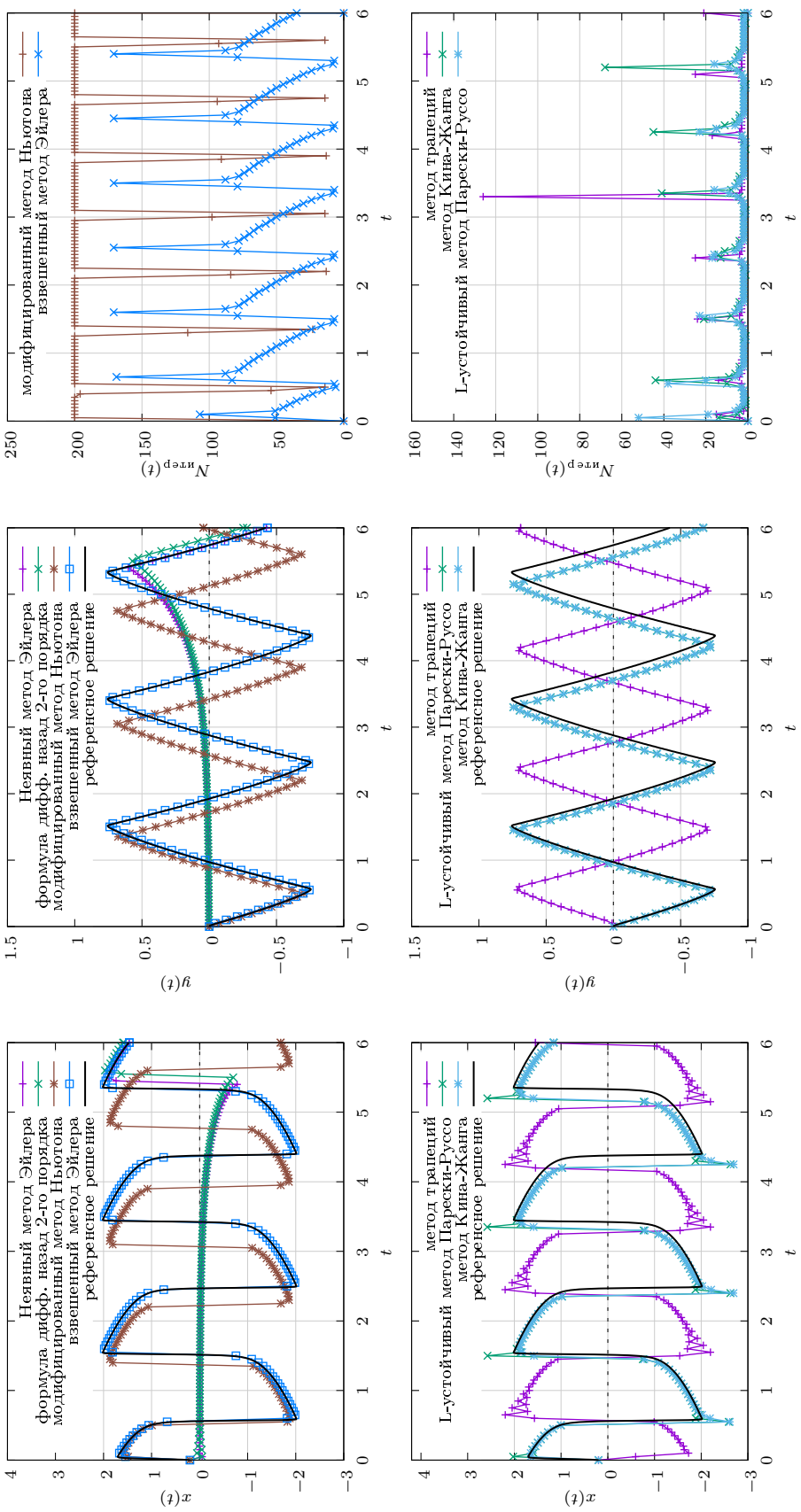

Рис. 4.3: Сравнение методов на примере интегрирования системы Ван дер Поля для шага по времени $\Delta t = 0.05$

Бутчера соответственно:

$$\begin{array}{c|cc} 1-\frac{\sqrt{2}}{2} & 1-\frac{\sqrt{2}}{2} & 0 \\ \frac{\sqrt{2}}{2} & \sqrt{2}-1 & 1-\frac{\sqrt{2}}{2} \\ \hline & \frac{1}{2} & \frac{1}{2} \end{array} \qquad \begin{array}{c|cc} 1/4 & 1/4 & 0 \\ 3/4 & 1/2 & 1/4 \\ \hline & 1/2 & 1/2 \end{array}$$

На графиках видно, что численные решения, полученные при помощи данных методов, имеют нефизичные пики в окресности резких перепадов референсного решения.

На рисунке 4.4 приведены графики нормы невязки для метода трапеций, модифицированного метода Ньютона и взвешенного метода Эйлера, возникающие при решении нелинейных систем на некоторых пяти последовательных шагах численного интегрирования системы Ван дер Поля. Снова хорошо видна седловая структура задачи. В случае метода трапеций и модифицированного метода Ньютона путь, образуемый ньютоновскими итерациями, осциллирует между минимумами (см. третий столбец). С другой стороны, взвешенный метод Эйлера избегает этой проблемы.

4.3 Каскад свёртывания крови

Перейдём, наконец, к системе каскада свёртывания крови. Данная система взята из [1; 25] и является упрощённой моделью более подробного каскада свёртывания крови [33; 34]. Уравнения модели приведены в (4.1), начальные условия и параметры — в таблице 4.1.

$$\begin{gathered} \frac{\partial P}{\partial t} = -\left(k_1\phi_c + k_2 B_\alpha + k_3 T + k_4 T^2 + k_5 T^3\right) P, \\ \frac{\partial T}{\partial t} = \left(k_1\phi_c + k_2 B_\alpha + k_3 T + k_4 T^2 + k_5 T^3\right) P - k_6 A T, \\ \frac{\partial B_\alpha}{\partial t} = \left(k_7\phi_c + k_8 T\right)\left(B^0 - B_\alpha\right) - k_9 A B_\alpha, \quad \frac{\partial A}{\partial t} = -k_6 A T - k_9 A B_\alpha, \\ \frac{\partial F_g}{\partial t} = -\frac{k_{10} T F_g}{K_{10} + F_g}, \quad \frac{\partial F}{\partial t} = \frac{k_{10} T F_g}{K_{10} + F_g} - k_{11} F, \quad \frac{\partial F_p}{\partial t} = k_{11} F, \\ \frac{\partial \phi_c}{\partial t} = -\left(k_{12} T - k_{13}\phi_c\right)\phi_f, \quad \frac{\partial \phi_f}{\partial t} = \left(k_{12} T - k_{13}\phi_c\right)\phi_f. \end{gathered} \tag{4.1}$$

Результаты численного эксперимента можно найти в таблице 4.2. Время моделирования — $T = 100$. Ошибка метода $\mathcal{E}_x$ для переменной $x(t)$ относительно референсного

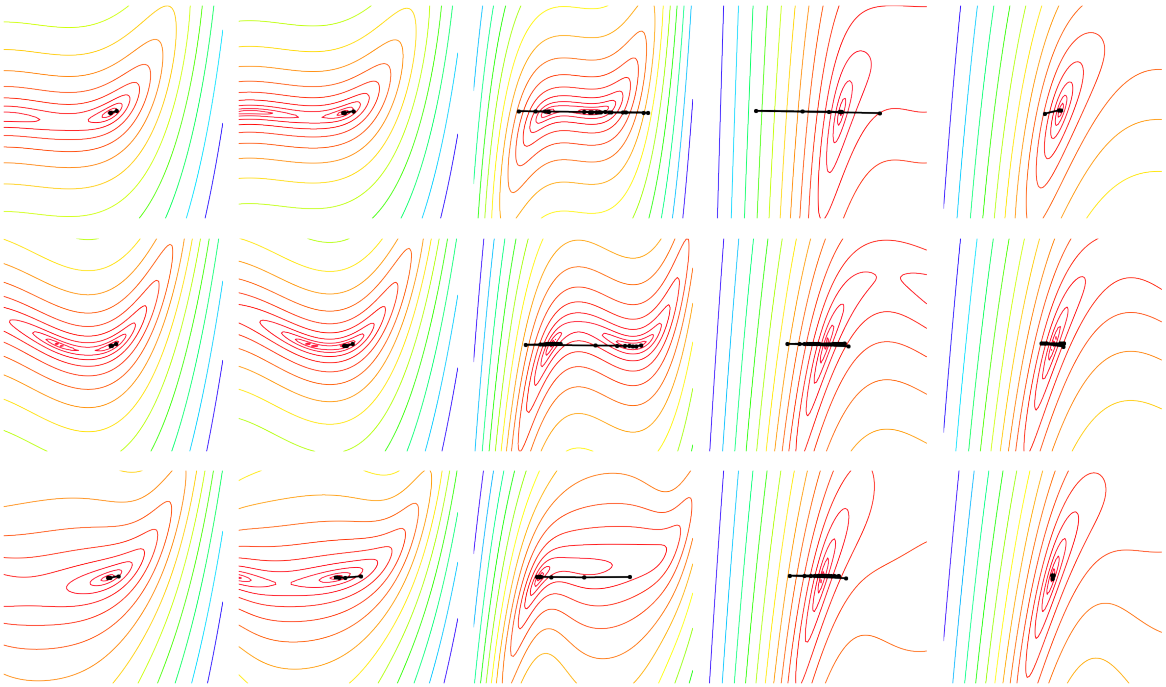

Рис. 4.4: Контуры 2-нормы невязки в зависимости от $x \in [x^* - 1.5, x^* + 1.5]$ по горизонтали и $y \in [y^* - 1.5, y^* + 1.5]$ по вертикали (где (x^*, y^*) — центр траектории ньютоновских итераций) для метода трапеций (верхний ряд), Эйлера (модифицированный метод Ньютона) (средний ряд) и взвешенного метода Эйлера (нижний ряд) для пяти последовательных шагов по времени ($\Delta t = 0.1$), осциллятор Ван дер Поля. Черная траектория соответствует итерациям Ньютона.

P	T	B_α	A	F_g	F	F_p	ϕ_c
1400	0	10	3400	7000	0	0	299
ϕ_f	k_1	k_2	k_3	k_4	k_5	k_6	k_7
1	$1.5 \cdot 10^{-4}$	$7.5 \cdot 10^{-6}$	$1.5 \cdot 10^{-5}$	$8 \cdot 10^{-6}$	10^{-10}	$4.817 \cdot 10^{-6}$	10^{-9}
k_8	k_9	k_{10}	K_{10}	k_{11}	k_{12}	k_{13}	B^0
$5.2173 \cdot 10^{-5}$	$2.223 \cdot 10^{-9}$	0.005	3160	0.1	0.002	$4 \cdot 10^{-9}$	200

Таблица 4.1: Начальные условия и параметры модели (каскад свёртывания крови).

решения $x_{ref}(t)$ вычислялась по формуле

$$\mathcal{E}_x = \sqrt{T \int\limits_{t\in[0,T]} \left(x(t) - x_{ref}(t)\right)^2 \mathrm{d}t} \Big/ \int\limits_{t\in[0,T]} |x_{ref}(t)| \, \mathrm{d}t \, .$$

Полная ошибка метода вычислялась как среднеквадратическая ошибка по всем переменным. В таблице также приведено полное $N_{\text{итер}}^{\text{общ}}$, среднее $N_{\text{итер}}^{\text{сред}}$, минимальное $N_{\text{итер}}^{\text{мин}}$ и максимальное $N_{\text{итер}}^{\text{макс}}$ число потребовавшихся ньютоновских итераций. Дополнительно указано число ньютоновских итераций, на которых решение оказывалось в отрицательной области — $N_{\text{итер}}^{\text{отриц}}$.

Метод	Δt	$N_{\text{итер}}^{\text{общ}}$	$N_{\text{итер}}^{\text{сред}}$	$N_{\text{итер}}^{\text{мин}}$	$N_{\text{итер}}^{\text{макс}}$	$\mathcal{E}$	$N_{\text{итер}}^{\text{отриц}}$
Неявный Эйлера	10^{-2}	10812	1.0812	1	3	$1.7 \cdot 10^{-2}$	0
Неявный Эйлера	10^{-1}	2032	2.032	2	5	0.17	0
Неявный Эйлера	1	233	2.33	2	19	0.79	10
Мод. Ньютона	10^{-2}	22071	2.20	2	7	$1.7 \cdot 10^{-2}$	0
Мод. Ньютона	10^{-1}	3740	3.74	3	20	0.17	0
Мод. Ньютона	1	768	7.68	4	53	0.79	10
Мод. Ньютона	2	502	10.0	5	82	1.14	12
Мод. Ньютона	5	240	12.0	9	26	1.19	13
Мод. Ньютона	10	164	16.4	11	20	1.17	6
Трапеций	10^{-2}	10701	1.0701	1	3	$1.6 \cdot 10^{-3}$	0
Трапеций	10^{-1}	2023	2.023	2	4	$3.8 \cdot 10^{-2}$	0
Взвешенный Эйлера	10^{-2}	10716	1.0716	1	3	$1.6 \cdot 10^{-3}$	0
Взвешенный Эйлера	10^{-1}	2052	2.052	2	7	$3.3 \cdot 10^{-2}$	0
Взвешенный Эйлера	$2.5 \cdot 10^{-1}$	861	2.1525	2	15	$8.2 \cdot 10^{-2}$	3

Таблица 4.2: Результаты для модели каскада свёртывания крови.

Применение неявного метода Эйлера напрямую ведёт к нефизичному отрицательному решению для $\Delta t \geqslant 2$. В то же время, модифицированный метод Ньютона позволил получить корректное решение для шагов вплоть до $\Delta t = 10$ включительно. Использование метода трапеций с тем же шагом приводило к появлению сильно осциллирующих нефизичных осцилляций. Взвешенный метод Эйлера позволил увеличить точность для маленьких шагов по времени ($\Delta t \leqslant 0.25$), но при использовании больших шагов также получались решения с нефизичными осцилляциями (пусть и заметно меньшими по амплитуде, чем у метода трапеций).

Из поведения численных методов снова можно сделать вывод, что модель каскада

свёртывания крови является жёсткой во всех предложенных смыслах. При этом предложенные методы позволяют кратно увеличить шаг интегрирования данной системы без потери устойчивости.

Глава 5

Заключение

Несмотря на бурное развитие теории устойчивости численных методов во второй половине прошлого века, некоторые вопросы данной области и по сей день остаются без ответа. В частности, до сих пор нет удовлетворительного определения жёстких систем, даже несмотря на то, что необходимость их решения возникает повсеместно. Например, уравнения, описывающие химические реакции, могут быть особенно жёсткими в силу разных временных масштабов протекающих процессов, а также в силу значительной нелинейности правой части. Возможность численно решать подобные системы в условиях ограниченных вычислительных ресурсов требует развития теории жёстких систем дифференциальных уравнений, а также соответствующих численных методов.

Таким образом, в ходе исследования способов устойчивого численного интегрирования жёсткой системы каскада свёртывания крови в данной работе были получены следующие результаты:

1. Систематизированы основные положения теории устойчивости численных методов, необходимые для исследования жёсткости систем. Введены два новых понятия: *линейная* и *нелинейная жёсткость*, отражающие разную природу жёсткости различных систем.

2. Показано, что для устранения эффектов линейной жёсткости достаточно использовать устойчивые в том или ином смысле численные методы.

3. Продемонстрировано, что понятие нелинейной жёсткости осмысленно и инфор-

мативно: существуют нелинейно жёсткие системы, которые, тем не менее, некорректно интегрируются A-устойчивыми, L-устойчивыми методами и экспоненциальными интеграторами. Причём сложность интегрирования данных систем обусловлена нелинейностью правых частей.

4. Показано, что проблема нелинейной жёсткости лежит в плоскости методов оптимизации, так как нелинейная жёсткость существенно влияет на поведение метода решения нелинейных систем, приводя к сходимости к нефизичным (паразитическим) корням уравнения. На ряде задач показано, что модифицируя алгоритм решения нелинейных систем можно найти такой метод, который выбирает траекторию к физичному корню.

5. Предложен класс квазиньютоновских модификаций неявного метода Эйлера, а также класс соответствующих двухточечных численных схем, дающих ту же матрицу Якоби при использовании метода Ньютона. Оба класса параметризованы весовой функцией $\theta(z)$, используемой для «фильтрации» спектра матрицы Якоби правой части системы.

6. Предложен вариант весовой функции $\theta(z)$, дающей экспоненциальный интегратор. Обоснован её выбор.

7. Для предложенной функции проведены численные эксперименты на жёстких системах дифференциальных уравнений (в том числе и на системе каскада свёртывания крови), показывающие преимущество предложенных методов в сравнении с другими двухточечными методами в вопросах борьбы с нежелательными эффектами нелинейной жёсткости.

Список литературы

1. A mathematical model to quantify the effects of platelet count, shear rate, and injury size on the initiation of blood coagulation under venous flow conditions / A. Bouchnita [и др.] // PloS one. — 2020. — Т. 15, № 7. — e0235392.
2. *Alexander R.* The modified Newton method in the solution of stiff ordinary differential equations // mathematics of computation. — 1991. — Т. 57, № 196. — С. 673—701.
3. *Anderson D. G.* Iterative procedures for nonlinear integral equations // Journal of the ACM (JACM). — 1965. — Т. 12, № 4. — С. 547—560.
4. *Armijo L.* Minimization of functions having Lipschitz continuous first partial derivatives // Pacific Journal of mathematics. — 1966. — Т. 16, № 1. — С. 1—3.
5. *Auzinger W.*, *Frank R.* Asymptotic error expansions for stiff equations: an analysis for the implicit midpoint and trapezoidal rules in the strongly stiff case // Numerische Mathematik. — 1989. — Т. 56, № 5. — С. 469—499.
6. *Auzinger W.*, *Frank R.*, *Kirlinger G.* Modern convergence theory for stiff initial-value problems // Journal of Computational and Applied Mathematics. — 1993. — Т. 45, № 1/2. — С. 5—16.
7. *Auzinger W.*, *Frank R.*, *Kirlinger G.* A note on convergence concepts for stiff problems // Computing. — 1990. — Т. 44, № 3. — С. 197—208.
8. *Brown P. N.*, *Hindmarsh A. C.*, *Walker H. F.* Experiments with quasi-Newton methods in solving stiff ODE systems // SIAM journal on scientific and statistical computing. — 1985. — Т. 6, № 2. — С. 297—313.
9. *Brown P. N.*, *Saad Y.* Convergence theory of nonlinear Newton–Krylov algorithms // SIAM Journal on Optimization. — 1994. — Т. 4, № 2. — С. 297—330.

10. Composing scalable nonlinear algebraic solvers / P. R. Brune [и др.] // siam REVIEW. — 2015. — Т. 57, № 4. — С. 535—565.

11. *Dahlquist G.* On stability and error analysis for stiff non-linear problems PART I : тех. отч. / CM-P00069396. — 1975.

12. *Dahlquist G. G.* A special stability problem for linear multistep methods // BIT Numerical Mathematics. — 1963. — Т. 3, № 1. — С. 27—43.

13. Eigen v3 / G. Guennebaud, B. Jacob [и др.]. — 2010. — http://eigen.tuxfamily.org.

14. *Hochbruck M., Ostermann A.* Exponential integrators // Acta Numerica. — 2010. — Т. 19. — С. 209—286.

15. *Kantorovich L. V.* On Newton's method // Trudy MIAN SSSR. — 1949. — Т. 28. — С. 104—144.

16. *Lambert J. D.* Numerical Methods for Ordinary Differential Systems: The Initial Value Problem. — 1-е изд. — Wiley, 1991. — ISBN 0471929905; 9780471929901.

17. *Lawson J. D.* Generalized Runge-Kutta Processes for Stable Systems with Large Lipschitz Constants // SIAM Journal on Numerical Analysis. — 1967. — Т. 4, № 3. — С. 372—380. — ISSN 00361429.

18. *Liu M., Zhang L., Zhang C.* Study on Banded Implicit Runge–Kutta Methods for Solving Stiff Differential Equations // Mathematical Problems in Engineering. — 2019. — Т. 2019.

19. *Lotka A. J.* Elements of physical biology. — Williams & Wilkins, 1925.

20. *Meng-zhao Q., Mei-qing Z.* SYMPLECTIC RUNGE-KUTTA ALGORITHMS FOR HAMILTONIAN SYSTEMS // Journal of Computational Mathematics. — 1992. — Т. 10. — С. 205—215.

21. *Minchev B., Wright W.* A review of exponential integrators. — 2005. — Янв.

22. *Moore P. K., Petzold L. R.* A stepsize control strategy for stiff systems of ordinary differential equations // Applied numerical mathematics. — 1994. — Т. 15, № 4. — С. 449—463.

23. *Nesterov Y.* A method of solving a convex programming problem with convergence rate $\mathcal{O}(1/k^2)$ // Dokl. akad. nauk Sssr, т. 27. — 1983. — С. 543—547.

24. *Ortega J. M.*, *Rheinboldt W. C.* Iterative solution of nonlinear equations in several variables. — SIAM, 2000.

25. Parallel Finite Volume Computation on General Meshes / Y. Vassilevski [и др.]. — Springer Nature, 2020.

26. *Pareschi L.*, *Russo G.* Implicit-Explicit Runge-Kutta Schemes and Applications to Hyperbolic Systems with Relaxation // Journal of Scientific Computing. — 2005. — Янв. — Т. 25. — С. 129—155.

27. *Schlenkrich S.*, *Walther A.*, *Griewank A.* Application of AD-based quasi-Newton methods to stiff ODEs // Automatic Differentiation: Applications, Theory, and Implementations. — Springer, 2006. — С. 89—98.

28. *Söderlind G.* The logarithmic norm. History and modern theory // BIT Numerical Mathematics. — 2006. — Сент. — Т. 46. — С. 631—652.

29. *Sorensen D. C.* Newton's method with a model trust region modification // SIAM Journal on Numerical Analysis. — 1982. — Т. 19, № 2. — С. 409—426.

30. *Takesaki M.* Theory of operator algebras 1. — Springer, 2001. — (Encyclopaedia of mathematical sciences, Operator algebras and non-commutative geometry 124-125, 127, 5-6, 8).

31. Threshold response of initiation of blood coagulation by tissue factor in patterned microfluidic capillaries is controlled by shear rate / F. Shen [и др.] // Arteriosclerosis, thrombosis, and vascular biology. — 2008. — Т. 28, № 11. — С. 2035—2041.

32. *Wolfe P.* Convergence conditions for ascent methods // SIAM review. — 1969. — Т. 11, № 2. — С. 226—235.

33. *Пантелеев М. А.*, *Атауллаханов Ф. И.* Свертывание крови: биохимические основы // Клиническая онкогематология. — 2008. — Т. 1, вып. 1. — С. 50—62.

34. Применение проточных систем в лабораторной диагностике для интегральной оценки системы гемостаза / О. Ушакова [и др.] // Вопросы гематологии/онкологии и иммунопатологии в педиатрии. — 2018. — Т. 17, вып. 1. — С. 117—129.

35. *Хайрер Э.*, *Ваннер Г.* Решение обыкновенных дифференциальных уравнений: жёсткие и дифференциально-алгебраические задачи / пер. Е. Л. Старостиной [и др.]. — 2-е изд. — Мир, 1999.