

**Министерство образования и науки Российской Федерации
«Московский физико-технический институт (государственный
университет)»**

**Физтех-школа Прикладной Математики и Информатики (ФПМИ)
Кафедра вычислительных технологий и моделирования в геофизике и
биоматематике**

Выпускная квалификационная работа бакалавра

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

Автор:

Студент 871 группы
Бутаков Иван Дмитриевич

Научный руководитель:

к.ф.-м.н.
Терехов Кирилл Михайлович

Научный консультант:

научная степень
Сергеев Сергей Сергеевич



Москва 2022

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 левого предсердия 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 до некоторой степени решить указанные проблемы. Методы основаны на «сглаживании» спектра матрицы Якоби правой части системы. Сглаживание производится путём комбинирования явного и неявного метода Эйлера с матричным весом. 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 с невзвешенным вариантом. В работе также описаны детали программной реализации предложенных методов. Наконец, дан краткий обзор возможностей по применению предложенной техники для построения других методов интегрирования, обладающих схожими свойствами.

Abstract

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

Оглавление

1	Введение	5
2	Теоретические сведения	7
2.1	Линейная теория устойчивости	8
2.1.1	Линейная устойчивость	9
2.1.2	Логарифмическая норма	12
2.1.3	Линейная жёсткость	12
2.1.4	Экспоненциальные интеграторы	15
2.2	Нелинейная теория устойчивости	16
2.2.1	Нелинейная устойчивость	16
2.2.2	Нелинейная жёсткость	18
3	Разработка численных методов	21
3.1	Модифицированный метод Ньютона	21
3.2	Взвешенный метод Эйлера	23
3.3	Выбор весовой функции	24
4	Численные эксперименты	27

Глава 1

Введение

Написание данной работы мотивировано необходимостью решать жёсткие нелинейные уравнения реакции при моделировании образования тромба в придатке левого предсердия. Моделирование заключается в решении системы переноса-диффузии-реакции в геометрии левого предсердия или только его придатка. 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 [26]. Это вынуждает использовать малый шаг по времени при неявном численном интегрировании. В свою очередь, это приводит к непрактично большому времени расчёта упомянутой модели. С целью увеличения допустимого шага по времени в работе [21] 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 предложенного метода. Предложенный в настоящей работе метод основан на замечании, что матрица Якоби правой части уравнения реакции обладает седловой структурой, которая проявляется себя при численном интегрировании с большим шагом по времени. Мы сосредотачиваем своё внимание на одношаговых схемах интегрирования по времени, чтобы в дальнейшем встроить полученный метод в полностью неявный интегратор уравнения переноса-диффузии.

Согласно теореме Канторовича [14; 20], для липшицевых в окрестности корня функций 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; 27], метод доверительных областей [24] и различного рода ускорения [3; 9; 19]. 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; 18; 22]. 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. Наш 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 задачи, связанной с построением численной схемы специального вида.

Много работ посвящено устойчивости и выбору численных схем [6; 11; 12; 16]. 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.

Глава 2

Теоретические сведения

Прежде чем перейти к описанию численных методов, остановимся подробнее на теории жёстких систем дифференциальных уравнений. В данном разделе приведены необходимые для дальнейшего анализа элементы линейной и нелинейной теории устойчивости численных методов. Отметим, что данная теория гораздо более полно описана в работах [15; 28].

В дальнейшем часто будет рассматриваться задача Коши вида

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

где $D \subseteq \mathbb{R}^d$. Поскольку любую систему обыкновенных дифференциальных уравнений можно свести к автономной, добавив уравнение $dt/dt = 1$, запись задачи можно упростить:

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

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:

$$\begin{cases} \frac{d\mathbf{x}}{dt} = f_0 + F(\mathbf{x}_0) \cdot (\mathbf{x} - \mathbf{x}_0), \\ \mathbf{x}(0) = \mathbf{x}_0, \end{cases} \quad (2.2)$$

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

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

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

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

— регулярная в \mathbb{C} функция.

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 уравнению Даламбера:

$$\begin{cases} \frac{d\mathbf{x}}{dt} = F_0 \cdot \mathbf{x}, \\ \mathbf{x}(0) = \mathbf{x}_0 \end{cases} \iff \mathbf{x}(t) = \exp(t \cdot F_0) \cdot \mathbf{x}_0 \quad (2.4)$$

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

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

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. 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.

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

В общем случае двухшаговой одностадийной схемы:

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

где M , вообще говоря, может быть матрицей (в таком случае $R(z)$ следует изначально рассматривать как матричную функцию). 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:

Определение 2.1. 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 2.2 (об отображении спектра; [25], предложение 2.8). Пусть A — линейный ограниченный оператор, действующий в банаховом пространстве \mathcal{B} над \mathbb{C} , f — регулярная в окрестности $\sigma(A)$ функция. Тогда

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

Теорема 2.3 (формула Бёрлинга-Гельфанда). Пусть A — линейный ограниченный оператор, действующий в банаховом пространстве \mathcal{B} над \mathbb{C} . Тогда $r(A) = \lim_{n \rightarrow \infty} \|A^n\|^{\frac{1}{n}} = \inf_{n \in \mathbb{N}} \|A^n\|^{\frac{1}{n}}$.

Следствие 2.4. Пусть A — линейный ограниченный оператор, действующий в банаховом пространстве \mathcal{B} над \mathbb{C} . Тогда

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

Доказательство. По теореме 2.3

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

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

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

откуда получаем искомое неравенство. □

Лемма 2.5. Пусть A — линейный ограниченный оператор, действующий в банаховом пространстве \mathcal{B} над \mathbb{C} . Тогда

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

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 $\|T(t)\| \leq e^{t \cdot \|A\|}$. Согласно теореме 2.2, $\sigma(T(t)) = \exp(t \cdot \sigma(A))$, из чего следует $r(T(t)) = e^{t \cdot s(A)}$.

Применяя следствие 2.4 и учитывая, что

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

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

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

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

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

где $M = C \cdot e^{\Delta t \cdot |s(A)|} = e^{\Delta t \cdot (\|A\| + |s(A)|)}$. Отсюда получаем доказываемое утверждение. \square

Следствие 2.4 позволяет на основе данных о спектре линейного оператора A оценить асимптотику $\|A^n\|$, а лемма 2.5 — $\|\exp(t \cdot A)\|$.

Вернёмся к 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 2.6. Множество $\mathbf{R} = \{z \in \mathbb{C} \mid |R(z)| < 1\}$ называется областью абсолютной устойчивости численного метода, обладающего функцией устойчивости $R(z)$. Множество $\overline{\mathbf{R}} = \{z \in \mathbb{C} \mid |R(z)| \leq 1\}$ — замыкание области абсолютной устойчивости.¹

Утверждение 2.7. Пусть численное решение уравнения (2.4) ищется интегрированием с постоянным шагом Δt при помощи метода, обладающего функцией устойчивости $R(z)$ и соответствующей 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

$$\begin{aligned} \Delta t \cdot \sigma(F_0) \subseteq \mathbf{R} & \iff \| (R(\Delta t \cdot F_0))^n \| \xrightarrow{n \rightarrow \infty} 0 \\ \Delta t \cdot \sigma(F_0) \subseteq \mathbb{C} \setminus \overline{\mathbf{R}} & \implies \| (R(\Delta t \cdot F_0))^n \| \xrightarrow{n \rightarrow \infty} \infty \end{aligned}$$

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 2.2

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

Отсюда следует, что

$$\begin{aligned} \Delta t \cdot \sigma(F_0) \subseteq \mathbf{R} & \iff r(R(\Delta t \cdot F_0)) < 1 \\ \Delta t \cdot \sigma(F_0) \subseteq \mathbb{C} \setminus \overline{\mathbf{R}} & \iff r(R(\Delta t \cdot F_0)) > 1 \end{aligned}$$

Наконец, применяя следствие 2.4, завершаем доказательство утверждения. \square

¹Здесь неявно предполагается непрерывность $R(z)$. Можно обойтись без этого условия, однако тогда некорректно называть $\overline{\mathbf{R}}$ замыканием \mathbf{R} .

В теории линейной устойчивости важная роль отводится *A-устойчивости* — свойству численного решения 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 некоторое заранее известное число итераций) также стремится к нулю, то говорят об *L-устойчивости*. Дадим формальное определение.

Определение 2.8. Численный метод называется *A-устойчивым* в случае, если $\mathbb{C}^- \equiv \{z \in \mathbb{C} \mid \operatorname{Re} z < 0\} \subseteq \mathbf{R}$.

Определение 2.9. Численный метод называется *L-устойчивым* в случае, если он *A-устойчив* и выполнено $\lim_{\operatorname{Re} z \rightarrow -\infty} R(z) = 0$.

Утверждение 2.10. Пусть $R(z)$ регулярна в \mathbb{C}^- . Соответствующий численный метод *A-устойчив* тогда и только тогда, когда $\forall F_0$ выполнено

$$\|\exp(t \cdot F_0)\| \xrightarrow{t \rightarrow +\infty} 0 \quad \implies \quad \forall \Delta t > 0 \quad \|(R(\Delta t \cdot F_0))^n\| \xrightarrow{n \rightarrow \infty} 0$$

Доказательство. В силу леммы 2.5 импликацию из утверждения можно переписать в виде

$$\sigma(F_0) \subseteq \mathbb{C}^- \quad \implies \quad \forall \Delta t > 0 \quad \|(R(\Delta t \cdot F_0))^n\| \xrightarrow{n \rightarrow \infty} 0$$

Если также воспользоваться 2.7, получаем

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

что при произвольном F_0 равносильно $\mathbb{C}^- \subseteq \mathbf{R}$. Это даёт определение 2.8. \square

Утверждение 2.11. Пусть $R(z)$ регулярна в \mathbb{C}^- . Если соответствующий численный метод *L-устойчив*, то $\forall F_0$ выполнено

$$\|\exp(t \cdot F_0)\| \xrightarrow{t \rightarrow +\infty} 0 \quad \implies \quad r(R(\Delta t \cdot F_0)) \xrightarrow{\Delta t \rightarrow +\infty} 0$$

Доказательство. Аналогично доказательству утверждения 2.10 перепишем импликацию в виде

$$\sigma(F_0) \subseteq \mathbb{C}^- \quad \implies \quad r(R(\Delta t \cdot F_0)) \xrightarrow{\Delta t \rightarrow +\infty} 0$$

В силу теоремы 2.2 это эквивалентно

$$\sigma(F_0) \subseteq \mathbb{C}^- \quad \implies \quad \sup\{|\lambda| \mid \lambda \in R(\Delta t \cdot \sigma(F_0))\} \xrightarrow{\Delta t \rightarrow +\infty} 0,$$

что при произвольном F_0 равносильно $\forall z \in \mathbb{C}^- \quad |R(\Delta t \cdot z)| \xrightarrow{\Delta t \rightarrow +\infty} 0$. Это верно для *L-устойчивых* методов по определению 2.9. \square

Стоит отметить, что утверждение 2.11, в отличие от 2.10, 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 системы.

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

Определение 2.12. Пусть A — линейный ограниченный оператор, действующий в банаховом пространстве \mathcal{B} над \mathbb{C} . Число

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

называется логарифмической нормой оператора A .

Далее приведём без доказательства несколько утверждений, связывающих логарифмическую норму с теорией линейной устойчивости. Все утверждения взяты из работ [15; 23].

Утверждение 2.13. Пусть A, B — линейные операторы, действующие в конечно-мерном банаховом пространстве \mathcal{B} над \mathbb{C} . Тогда верно следующее:

1. $\mu[A]$ определена.
2. $\mu[A] \leq \|A\|$.
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. $\mu[A + B] \leq \mu[A] + \mu[B]$.
6. $s(A) \leq \mu[A]$.
7. $\|\exp(t \cdot A)\| \leq e^{t \cdot \mu[A]}$ для любого $t \geq 0$.

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 2.5, 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 возрасти, в то время как норма точного решения ограничена сверху убывающей экспонентой. С другой стороны, такая ситуация невозможна независимо от Δt при использовании А-устойчивых методов. Но А-устойчивость не гарантирует соизмеримую с точным решением скорость затухания численного; возможен даже случай $\lim_{\operatorname{Re} z \rightarrow -\infty} |R(z)| = 1$, 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 А-устойчивый), метода трапеций (А-устойчивый, но не L-устойчивый) и неявного метода Эйлера (L-устойчивый) при разных значениях $z = \Delta t \cdot F_0$ в одномерной задаче Далквиста.

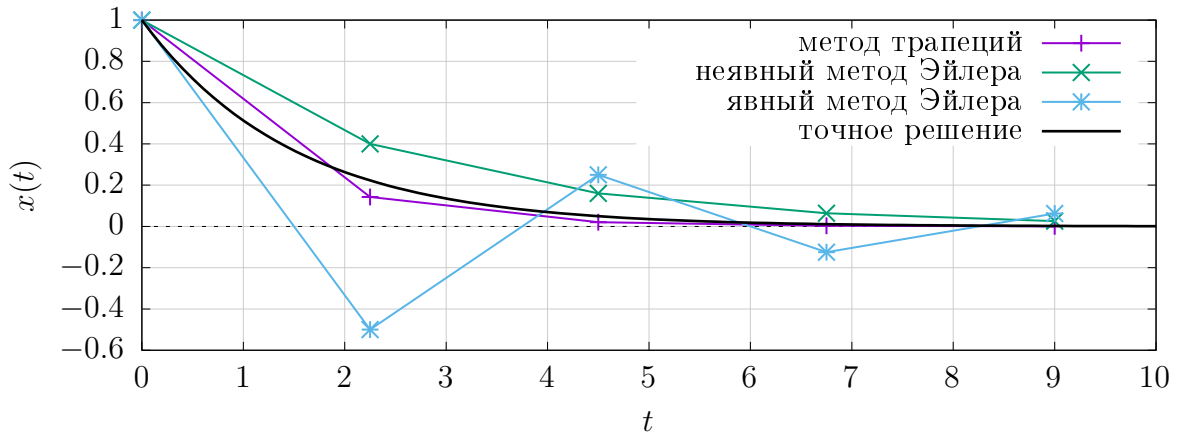


Рис. 2.1: Поведение простейших численных методов при решении одномерного уравнения Далквиста ($z = -1.5$)

Зачастую область устойчивости не А-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 ограничением 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 и высоким порядком аппроксимации:

Теорема 2.14 (второй барьер Далквиста). *Не существует А-устойчивых линейных многошаговых одностадийных схем с порядком аппроксимации выше второго.*

Ограничение на Δt может сохраняться даже при решении нелинейных задач вида (2.1). Рассмотрим случай, когда τ_{nonlin} — характерное время изменения $F = \frac{\partial f}{\partial \mathbf{x}}(\mathbf{x}(t))$ — много больше $\tau_{\text{lin}} = 1/r(F)$ — характерного времени реакции системы на небольшие возмущения. В таком случае линеаризация (2.2) остаётся достаточно точной дольше характерного времени τ . Это автоматически оставляет в силе ограничения на шаг интегрирования, полученные для линейных систем. В частности, если

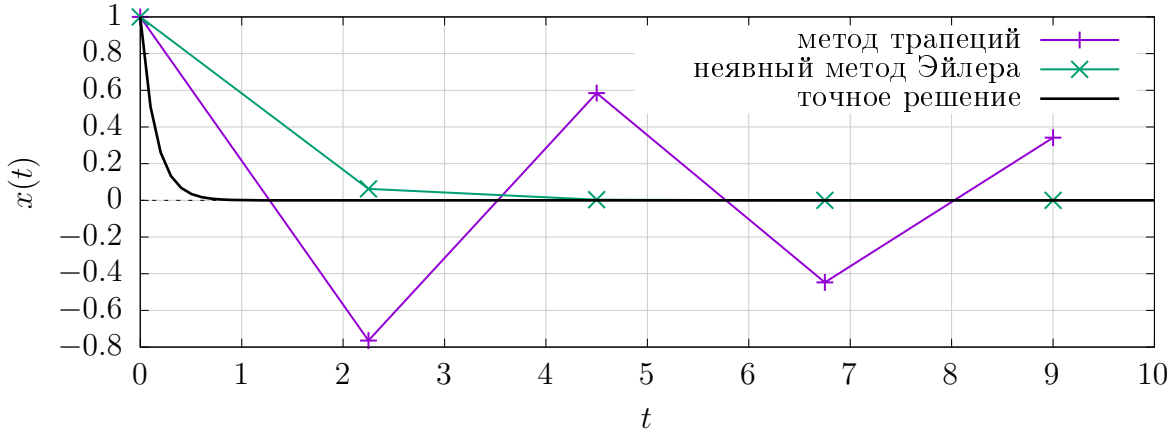


Рис. 2.2: Поведение простейших численных методов при решении одномерного уравнения Далквиста ($z = -15$)

$\mu[F] < 0$, 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 2.8 и 2.9 методами. Традиционно такие системы называют *жесткими*. Как указано в [15; 28], существует несколько определений жёсткости, каждое из которых обладает своими достоинствами и недостатками. Приведём здесь одно из них:

Определение 2.15. Система вида $\frac{dx}{dt} = f(t, x)$ 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 2.16. Система вида $\frac{dx}{dt} = f(x)$ называется линейно жёсткой в том случае, если характерное время (линейной) реакции системы на небольшие возмущения $\tau_{\text{lin}} = 1/r(\frac{\partial f}{\partial x}) = 1/r(F)$ много меньше характерных масштабов времени, на которых меняется точное решение.

Для численных методов с ограниченной областью устойчивости условие $\Delta t \cdot \sigma(F) \subseteq \mathbf{R}$ влечет $\Delta t \sim \tau_{\text{lin}}$. Тогда 2.16 оказывается частным случаем 2.15, 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.

2.1.4 Экспоненциальные интеграторы

Линейный анализ устойчивости известен давно, а потому достаточно подробно разработан [6; 11; 12; 15; 16; 28]. 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. Все они обладают достаточно большой областью устойчивости, что, как видно из приведённого ранее анализа, делает их полезными для интегрирования линейно жёстких систем.

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

Замечание 2.18. *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. Некоторые авторы, тем не менее, расширяют определение экспоненциальных интеграторов на все методы, использующие экспоненту матрицы Якоби правой части системы. Обзор существующих экспоненциальных интеграторов можно найти, например, в работах [13; 17]. Мы же остановимся только на простейших из них, необходимых для дальнейшего анализа.

Рассмотрим две схемы, полученных из явного и неявного методов Эйлера соответственно (так называемые «exponential time differencing»-методы Рунге-Кутты первого порядка):

$$\text{Явный ETDRK1:} \quad \mathbf{x}^{n+1} - \mathbf{x}^n = (\exp(\Delta t \cdot F) - I) F^{-1} \cdot f(\mathbf{x}^n) \quad (2.7)$$

$$\text{Неявный ETDRK1:} \quad \mathbf{x}^{n+1} - \mathbf{x}^n = (I - \exp(-\Delta t \cdot F)) F^{-1} \cdot f(\mathbf{x}^{n+1}) \quad (2.8)$$

Здесь, как и ранее, $F = \frac{\partial f}{\partial \mathbf{x}}$ — матрица Якоби правой части системы. 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 $f(\mathbf{x}) = f_0 + F(\mathbf{x}_0) \cdot (\mathbf{x} - \mathbf{x}_0)$ 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.
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 [15] дано множество примеров систем, применение линейного анализа к которым даёт некорректные результаты (по крайней мере, если ограничиться лишь спектральными признаками). В этой связи в настоящем разделе решено привести также основные результаты нелинейной теории устойчивости. Вместе с тем будут 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 стремиться 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 2.19. Пусть $\mathbf{x}(t)$ и $\mathbf{y}(t)$ — решения системы $\frac{d\mathbf{x}}{dt} = f(t, \mathbf{x})$. Данные решения называются сжимающимися («contractive») на отрезке $[a; b]$ в случае

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

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

Определение 2.20. 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 («contractive») на отрезке $[a; b]$ в случае

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

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

Определение 2.21. Пусть \mathcal{H} — гильбертово пространство над \mathbb{C} . Функция $f(t, \mathbf{x}) = f : \mathbb{R} \times D \rightarrow \mathcal{H}$ (где $D \subseteq \mathcal{H}$ — выпуклая область) и система $\frac{d\mathbf{x}}{dt} = f(t, \mathbf{x})$ называются односторонне липшицевыми на отрезке $[a; b]$ в случае

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

Определение 2.22. Пусть в определении 2.21 $\nu(t) \leq 0$. Тогда функция $f(t, \mathbf{x})$ и система $\frac{d\mathbf{x}}{dt} = f(t, \mathbf{x})$ называются диссипативными.

Утверждение 2.23 (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 2.24. Линейная функция $f(t, \mathbf{x}) = A(t) \cdot \mathbf{x} + f_0(t)$ односторонне липшицева на всей области определения $A(t)$ с наименьшим возможным коэффициентом $\nu(t) = \mu[A(t)]$.

Замечание 2.24 позволяет связать линейную теорию устойчивости с нелинейной. Его естественным продолжением является следующая теорема:

Теорема 2.25 (Далквист, 1959). Пусть \mathcal{H} — гильбертово пространство над \mathbb{C} . Пусть $f(t, \mathbf{x}) = f : \mathbb{R} \times D \rightarrow \mathcal{H}$, где $D \subseteq \mathcal{H}$ — выпуклая область. Наконец, пусть существуют $a, b \in \mathbb{R}$ и $\nu(t)$ — такая кусочно-непрерывная функция, что

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

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

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

Наконец, приведём достаточное условие сжимаемости численных решений.

Теорема 2.26 (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 систему

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

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 своей линеаризации в ближайшем положении равновесия не более, чем на $|\cos(\pi/2) - \pi/2| \approx 1.58 \approx 1.58 \cdot \max |f(x)|$. Можно выдвинуть предположение, что решение данной системы при помощи А-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 референсное решение методом трапеций с шагом $\Delta t = 0.02$. 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.

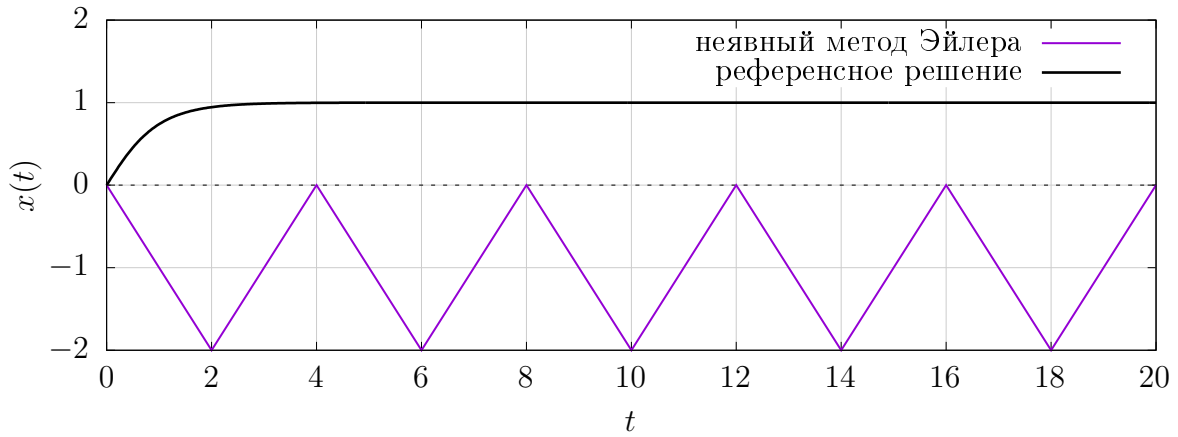


Рис. 2.3: Поведение L-устойчивого метода Эйлера при решении уравнения (2.9)

Данный пример показывает, что явление жёсткости не ограничивается только линейной её составляющей. Выпишем невязку неявного метода Эйлера и его матрицу Якоби:

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

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

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 [14; 20] 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.

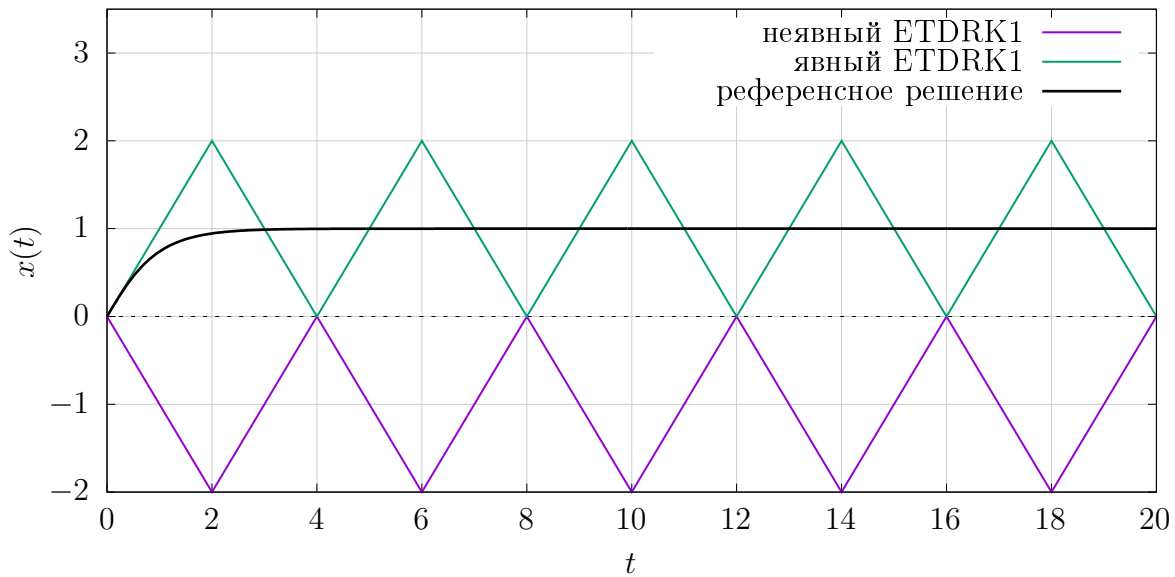


Рис. 2.4: Поведение простейших экспоненциальных интеграторов при решении уравнения (2.9)

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

Определение 2.27. Система вида $\frac{dx}{dt} = f(t, \mathbf{x})$ называется нелинейно жёсткой в том случае, если нелинейные свойства правой части существенно влияют на допустимую величину шага интегрирования, при которой возможна сходимость используемого метода поиска корней невязки выбранной численной схемы.

Таким образом, борьба с нелинейной жёсткостью переводится в плоскость методов оптимизации. Среди способов улучшить сходимость метода Ньютона можно перечислить линейный поиск [4; 27], метод доверительных областей [24] и различного рода ускорения [3; 9; 19]. 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; 18; 22]. 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 А-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:

$$\mathcal{J}_m \cdot (\mathbf{x}_{m+1}^{n+1} - \mathbf{x}_m^{n+1}) = -\mathcal{R}_m \quad \implies \quad \mathbf{x}_{m+1}^{n+1} = \mathbf{x}_m^{n+1} - \mathcal{J}_m^{-1} \cdot \mathcal{R}_m,$$

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:

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

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

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

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

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 $\bar{\mathcal{J}}_m = I - \Delta t \cdot M F_m$ 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 $M = \theta(F) = V \theta(\Lambda) V^{-1} = V \operatorname{diag}(\theta(\lambda_i)) V^{-1}$, что уже легко вычисляется, если известны V и Λ . При определённых условиях на θ матрица M всегда будет вещественной.

Утверждение 3.2. Пусть $f : \mathbb{C} \rightarrow \mathbb{C}$ — 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. \square

Лемма 3.3. Пусть $f : \mathbb{C} \rightarrow \mathbb{C}$ — 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 $\forall k \in \mathbb{N}_0 \quad c_k \in \mathbb{R}$. Тогда матрица $f(A)$ равна сумме ряда вещественных матриц, а потому вещественна. \square

Следствие 3.4. Пусть $\theta(z)$ является результатом применения стандартных алгебраических операций (сложение, умножение, обращение) к значениям $f_1(z), \dots, f_N(z)$, 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. В данной секции вводится численная схема, дающая невязку с той же матрицей Якоби, что получается в предыдущей секции. Рассмотрим

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

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 замороженной M итерация метода Ньютона тогда принимает вид, аналогичный (3.1):

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

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 производную $F_m = \frac{\partial f}{\partial \mathbf{x}}(t^{n+1}, \mathbf{x}_m^{n+1})$.
 - Вычислить собственное разложение матрицы Якоби правой части $F_m = V \Lambda V^{-1}$.
 - 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 $\|\bar{\mathcal{R}}_m\| > \max \{ \varepsilon_{\text{абс}}, \varepsilon_{\text{отн}} \cdot \|\bar{\mathcal{R}}_0\| \}$, то перейти на следующую итерацию по формуле (3.2), иначе $\mathbf{x}^{n+1} = \mathbf{x}_m^{n+1}$.

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 А-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 2.18 имеем

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

Данную функцию можно сразу доопределить в нуле по непрерывности. Она регулярна на всей области определения. Заметим также, что $\theta(z) = \frac{1}{2}(1 - \varphi(z))$, где

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

— 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 $\Delta t \rightarrow 0$ взвешенный метод Эйлера с данной весовой функцией переходит в метод трапеций.

Глава 4

Численные эксперименты

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. *Hochbruck M., Ostermann A.* Exponential integrators // Acta Numerica. — 2010. — Т. 19. — С. 209—286. — DOI: 10.1017/S0962492910000048.
14. *Kantorovich L. V.* On Newton’s method // Trudy MIAN SSSR. — 1949. — Т. 28. — С. 104—144.

15. *Lambert J. D.* Numerical Methods for Ordinary Differential Systems: The Initial Value Problem. — 1-е изд. — Wiley, 1991. — ISBN 0471929905; 9780471929901.
16. *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.
17. *Minchev B., Wright W.* A review of exponential integrators. — 2005. — ЯНВ.
18. *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.
19. *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.
20. *Ortega J. M., Rheinboldt W. C.* Iterative solution of nonlinear equations in several variables. — SIAM, 2000.
21. Parallel Finite Volume Computation on General Meshes / Y. Vassilevski [и др.]. — Springer Nature, 2020.
22. *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.
23. *Söderlind G.* The logarithmic norm. History and modern theory // BIT Numerical Mathematics. — 2006. — Сент. — Т. 46. — С. 631–652. — DOI: 10.1007/s10543-006-0069-9.
24. *Sorensen D. C.* Newton’s method with a model trust region modification // SIAM Journal on Numerical Analysis. — 1982. — Т. 19, № 2. — С. 409–426.
25. *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). — ISBN 354042248X; 9783540422488; 354042914X; 9783540429142; 9783540429135; 3540429131.
26. 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.
27. *Wolfe P.* Convergence conditions for ascent methods // SIAM review. — 1969. — Т. 11, № 2. — С. 226–235.
28. *Хайрер Э., Ваннер Г.* Решение обыкновенных дифференциальных уравнений: жёсткие и дифференциально-алгебраические задачи / пер. Е. Л. Старостиной [и др.]. — 2-е изд. — Мир, 1999.