ЖУРНАЛ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И МАТЕМАТИЧЕСКОЙ ФИЗИКИ, 2021, том 61, № 5, с. 706–722

ОБЩИЕ ЧИСЛЕННЫЕ МЕТОДЫ

УДК 519.6

ТОЧНЫЙ ПЕРЕЗАПУСК МЕТОДА ПОДПРОСТРАНСТВА КРЫЛОВА “СДВИГ–ОБРАЩЕНИЕ” ДЛЯ ВЫЧИСЛЕНИЯ ДЕЙСТВИЯ ЭКСПОНЕНТЫ НЕСИММЕТРИЧНЫХ МАТРИЦ[1)]

© 2021 г. М. А. Бочев[1,2]

[1] *125047 Москва, Миусская пл., 4, ИПМ РАН им. М.В. Келдыша, Россия*

[2] *119333 Москва, ул. Губкина, 8, ИВМ РАН им. Г.И. Марчука, Россия*

e-mail: botchev@ya.ru

Поступила в редакцию 24.12.2020 г.
Переработанный вариант 24.12.2020 г.
Принята к публикации 14.01.2021 г.

Предложен алгоритм перезапуска метода подпространства Крылова “сдвиг–обращение” для вычисления действия матричной экспоненты несимметричных матриц. Представленный метод является развитием недавно предложенного невязочно-временного перезапуска и разработан, чтобы предотвратить потерю точности, возможную в неувязочно-временном перезапуске. Наиболее затратная по вычислениям часть метода подпространства Крылова “сдвиг–обращение” – решение линейных систем со сдвинутой матрицей. Поскольку наш алгоритм перезапуска подразумевает изменение величины сдвига, мы показываем, что можно реализовать перезапуск так, чтобы единственного построения предобусловливателя (или LU разложения) было достаточно. Вычислительные эксперименты демонстрируют улучшенную точность и эффективность подхода. Библ. 44. Фиг. 6. Табл. 2.

Ключевые слова: метод подпространства Крылова “сдвиг–обращение”, экспоненциальное интегрирование по времени, процесс Арнольди, перезапуск методов подпространства Крылова.

DOI: 10.31857/S0044466921050033

1. ВВЕДЕНИЕ

Вычисление действия матричной экспоненты на заданный вектор – задача, часто возникающая в различных приложениях, таких как интегрирование по времени (см. [1]), анализ сетей (см. [2]) или редукция моделей (см. [3]). Для больших матриц методы подпространства Крылова являются важной группой методов, хорошо подходящих для этой задачи (см., например, [4]). Среди прочих методов для вычисления действия матричной экспоненты больших матриц можно выделить, например, методы, основанные на полиномах Чебышёва (см. [5]), метод масштабирования и квадратирования на основе ряда Тейлора (см. [6]) и другие. Методы подпространства Крылова для вычисления действия матричной экспоненты и других матричных функций представляют собой область активных исследований, среди недавних результатов и направлений которых отметим методы на основе рациональных подпространств Крылова (см. [7]–[11]), методы перезапуска (см. [12]–[17]) и эффективное решение задач большой размерности в разнообразных приложениях (см. [9], [18]–[21]).

Методы перезапуска позволяют ограничить число базисных векторов (размерность) подпространства Крылова, сохраняя при этом сходимость метода. Недавно предложенный невязочно-временной (НВ) перезапуск для вычисления матричной экспоненты представляется привлекательным инструментом (см. [22]) для этой цели. Одним из преимуществ этого метода перезапуска является то, что полиномиальные методы подпространства Крылова с НВ перезапуском гарантированно сходятся с требуемой точностью для любой длины перезапуска (т.е. для любой размерности подпространства) (см. [22]).

Другим свойством НВ перезапуска является то, что размерность малой спроецированной задачи, возникающей в ходе итераций, не растет с числом перезапусков, как в некоторых других

[1)]Работа выполнена при финансовой поддержке РНФ, проект № 19-11-00338.

методах перезапуска (см., например, [23, гл. 3]). Это означает, что, например, при длине перезапуска 10 размерность подпространства и размер малой спроецированной задачи не превосходят 10. Кроме того, спроецированная задача в методе подпространства Крылова с HB перезапуском представляет собой вычисление действия матричной экспоненты спроецированной матрицы. Это относительно простая задача, которая может быть эффективно решена стандартными методами линейной алгебры (см. [24]–[26]). Напротив, спроецированная задача в так называемых невязочных перезапусках (см. [12], [27]) – это система неавтономных дифференциальных уравнений (см. [12, соотношение (3)]). Хотя такая система, как правило, имеет небольшой размер, ее решение обычно более затратно по вычислениям и требует внимания, в частности, правильного подбора решателя систем дифференциальных уравнений и его параметров.

Важным классом рациональных методов подпространства Крылова (см. [11]) являются методы типа “сдвиг–обращение” (СО) (см. [7], [8]). Эти методы часто оказываются эффективными в разных приложениях, в частности, потому что они требуют решения линейных систем с одной и той же сдвинутой матрицей – в методах используется единственный сдвиг. Один из недостатков HB перезапуска, предложенного в [22], состоит в том, что точность метода подпространства Крылова СО с HB перезапуском не всегда может быть гарантирована. В данной статье предлагается перезапуск для метода подпространства Крылова СО, который является развитием HB перезапуска и позволяет получать требуемую точность вычислений. Предложенный подход работает для несимметричных матриц. Кроме того, в данной статье показано, как реализовать предложенный алгоритм эффективно, так чтобы достаточно было выполнить LU разложение или построение предобусловливателя только один раз.

Статья организована следующим образом. Предлагаемый алгоритм перезапуска, который мы называем THB (точный невязочно-временной) перезапуск, описан в разд. 2. Здесь сначала приводятся основные необходимые для изложения факты по методам подпространства Крылова (п. 2.1), затем описывается и обсуждается THB перезапуск (п. 2.2) и рассматривается, как организовать решение сдвинутых линейных систем в методе с THB перезапуском эффективно (п. 2.3). В разд. 3 представлены вычислительные эксперименты для двух тестовых задач. Заключительные выводы представлены в разд. 4.

2. ТОЧНЫЙ НЕВЯЗОЧНО-ВРЕМЕННОЙ ПЕРЕЗАПУСК

Условимся, что, если не оговорено иначе, $\|\cdot\|$ обозначает стандартную евклидову норму $\|\cdot\|_2$. По всей статье предполагаем также, что для матрицы $A \in \mathbb{R}^{n\times n}$ выполняется

$$\mathrm{Re}(x^*Ax) \geq 0 \quad \forall x \in \mathbb{C}^n, \tag{1}$$

где $\mathrm{Re}(z)$ обозначает вещественную часть $z \in \mathbb{C}$.

2.1. Методы подпространства Крылова и HB перезапуск

Предположим, что по данным $A \in \mathbb{R}^{n\times n}$, $v \in \mathbb{R}^n$ и $t > 0$ нужно вычислить действие матричной экспоненты матрицы $-tA$ на вектор v, т.е.

$$\text{вычислить} \quad y := \exp(-tA)v. \tag{2}$$

Эта задача эквивалентна решению задачи Коши

$$y'(t) = -Ay(t), \quad y(0) = v, \tag{3}$$

где, слегка пренебрегая точностью обозначений, мы используем t и как независимую переменную в (3), и для обозначения длины интервала в (2). Метод подпространства Крылова для вычисления действия матричной экспоненты можно рассматривать как галеркинскую проекцию задачи Коши (3) на подпространство Крылова

$$\mathcal{K}_k(A, v) = \mathrm{span}(v, Av, A^2v, \ldots, A^{k-1}v). \tag{4}$$

Сначала ортонормальный базис подпространства $\mathcal{K}_k(A, v)$ вычисляется обычным процессом Арнольди (или, если $A = A^{\mathrm{T}}$, процессом Ланцоша) (см. [24], [28]–[30]) и сохраняется в виде

столбцов $v_1, \ldots, v_k$ матрицы $V_k = [v_1 \ \ldots \ v_k] \in \mathbb{R}^{n\times k}$, так что выполняется так называемое разложение Арнольди:

$$AV_k = V_{k+1}\underline{H}_k = V_k H_k + h_{k+1,k}v_{k+1}e_k^{\mathrm{T}}, \tag{5}$$

где $e_k = (0, \ldots, 0, 1)^{\mathrm{T}} \in \mathbb{R}^k$, $\underline{H}_k \in \mathbb{R}^{(k+1)\times k}$ – верхняя хессенбергова матрица, а матрица $H_k \in R^{k\times k}$ состоит из первых k строк матрицы $\underline{H}_k$. После этого крыловская аппроксимация $y_k(t) \approx \exp(-tA)v$ определяется как (см. [31]–[33])

$$y_k(t) = V_k u(t), \tag{6}$$

где $u(t) : \mathbb{R} \to \mathbb{R}^k$ решает задачу Коши с проецированной матрицей $H_k = V_k^{\mathrm{T}} A V_k$:

$$u'(t) = -H_k u(t), \quad u(0) = \beta e_1. \tag{7}$$

Здесь $\beta = \|v\|$ и $e_1 = (1, 0, \ldots, 0)^{\mathrm{T}} \in \mathbb{R}^k$. Заметим, что задача Коши (7) – это галеркинская проекция задачи Коши (3) на подпространство Крылова и что $u(t)$ может быть вычислена как

$$u(t) = \exp(-tH_k)\beta e_1. \tag{8}$$

Если k не слишком велико, то (8) – предпочтительный способ вычисления u, который может быть реализован многими стандартными методами линейной алгебры (см. [24]–[26]). Вычисление $\exp(-tH_k)$ обычно является более простой операцией, чем решение системы дифференциальных уравнений в (7), что требует выбора подходящего решателя (в частности, жесткого или нежесткого) и его параметров. Метод подпространства Крылова (5)–(8) иногда называют полиномиальным методом подпространства Крылова, чтобы подчеркнуть тот факт, что вектора подпространства (4) – многочлены от A.

Естественным способом контроля (неизвестной) ошибки крыловского приближения (6) является отслеживание невязки $r_k(t)$ этого приближения $y_k(t)$ по отношению к системе дифференциальных уравнений $y' = -Ay$, а именно (см. [12], [27], [34]),

$$r_k(t) = -Ay_k(t) - y_k'(t). \tag{9}$$

Невязка $r_k(t)$ доступна в ходе процесса Арнольди и может быть вычислена по формуле (см. [12], [27], [34])

$$r_k(t) = -h_{k+1,k}(e_k^{\mathrm{T}}u(t))v_{k+1}. \tag{10}$$

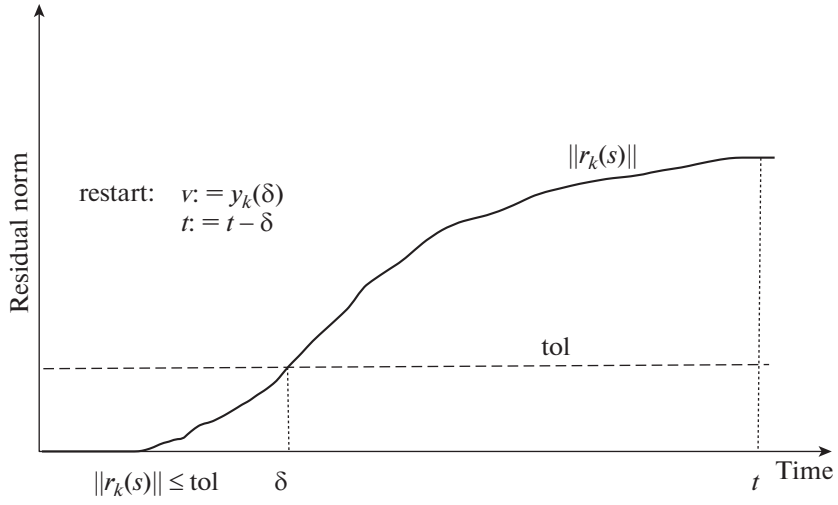
Как видим, $r_k(t)$ – скалярная функция, помноженная на v_{k+1}. Следовательно, $V_k^{\mathrm{T}} r_k(t) = 0$ для всех $t > 0$, а (6) – действительно галеркинская проекция на $\mathcal{K}_k(A, v)$. Некоторые результаты по сходимости невязки и ее связи с ошибкой могут быть найдены, например, в [22], [27].

Метод подпространства Крылова CO (“сдвиг–обращение”) (см. [7], [8]) для вычисления (2) отличается от стандартного полиномиального метода подпространства Крылова, описанного выше, тем, что подпространство Крылова строится для сдвинутой и обращенной матрицы $(I + \gamma A)^{-1}$, а не для A, где параметр $\gamma > 0$ фиксирован. Это делается для ускорения сходимости: процесс Арнольди имеет тенденцию лучше приближать наибольшие по модулю собственные числа, а для матрицы $(I + \gamma A)^{-1}$ они соответствуют малым собственным значениям матрицы A. Именно эти собственные числа важны для матричной экспоненты (компоненты, соответствующие большим собственным числам, не столь важны, они угасают экспоненциально быстро). Цена этого ускорения в сходимости – необходимость решать линейные системы с матрицей $I + \gamma A$ на каждой крыловской итерации. Разложение Арнольди (5) для сдвинутой и обращенной (CO) матрицы $(I + \gamma A)^{-1}$ принимает вид

$$(I + \gamma A)^{-1}V_k = V_{k+1}\underline{\tilde{H}}_k = V_k\tilde{H}_k + \tilde{h}_{k+1,k}v_{k+1}e_k^{\mathrm{T}}.$$

Это соотношение удобнее использовать после преобразования

$$AV_k = V_k H_k - \frac{\tilde{h}_{k+1,k}}{\gamma}(I + \gamma A)v_{k+1}e_k^{\mathrm{T}}\tilde{H}_k^{-1}, \tag{11}$$



Фиг. 1. Схема НВ перезапуска, взятая из работы [22].

где обозначение $\tilde{}$ указывает, что проекция построена для СО матрицы $(I + \gamma A)^{-1}$, а матрица H_k определяется как преобразование, обратное к сдвигу и обращению:

$$H_k = \frac{1}{\gamma}(\tilde{H}_k^{-1} - I). \tag{12}$$

Подчеркнем, что матрицы V_{k+1} и H_k здесь отличаются от матриц V_{k+1} и H_k из соотношения (5). Подробный анализ метода подпространства Крылова СО и родственных методов можно найти в (см. [7], [8], [10]).

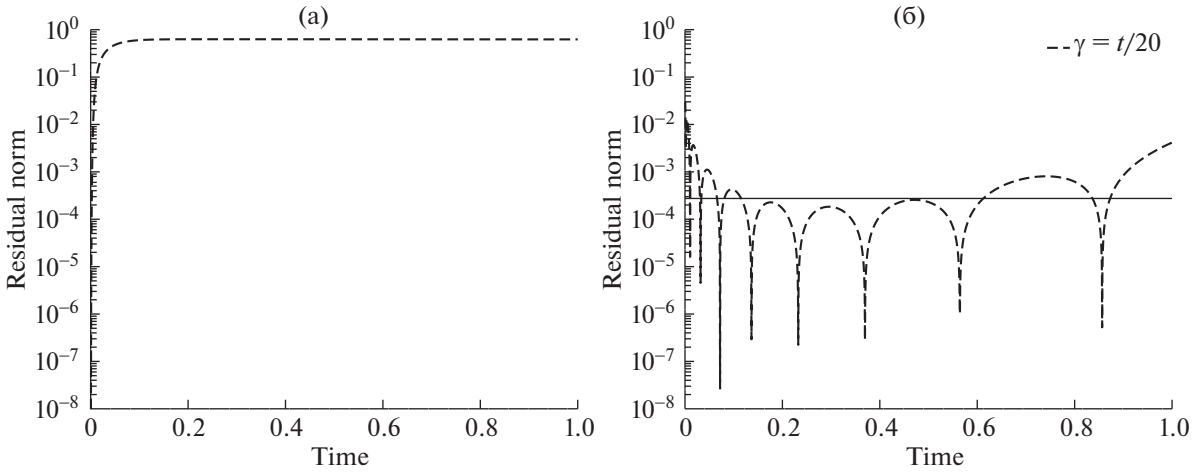
В методе подпространства Крылова СО невязка легко может быть вычислена следующим образом (см. [27]):

$$r_k(t) = \frac{\tilde{h}_{k+1,k}}{\gamma}(e_k^{\mathrm{T}}\tilde{H}_k^{-1}u(t))(I + \gamma A)v_{k+1}. \tag{13}$$

Здесь $u(t)$ – решение спроецированной задачи Коши (7), где H_k – матрица, получающаяся обратным преобразованием (12).

Величина сдвига γ обычно выбирается в соответствии с длиной t временного интервала $[0, t]$ (см. [8]), при этом часто используется величина $\gamma = t/10$. Таким образом, изменение параметра γ означает изменение t. Стандартный полиномиальный метод подпространства Крылова (5)–(8) имеет привлекательное свойство инвариантности относительно t: коль скоро V_{k+1} и $\underline{H}_k$ вычислены, они могут быть использованы для любых времен t (хотя качество аппроксимации $y_k(t) \approx y(t)$, вообще говоря, ухудшается с ростом t). К сожалению, это свойство полностью не распространяется на метод подпространства Крылова СО: в этом методе матрицы V_{k+1} и $\underline{\tilde{H}}_k$ зависят от величины γ, которая, в свою очередь, зависит от t. Тем не менее на практике можно использовать вычисленные матрицы Арнольди V_{k+1} и $\underline{\tilde{H}}_k$ для определенного диапазона t, не вычисляя их заново.

Недавно предложенный НВ перезапуск основан на том факте, что невязка как функция от t обычно является для обычного метода подпространства Крылова неубывающей функцией. Следовательно, когда выполнено максимально допустимое число $k_{\max}$ крыловских итераций (так что хранение большего числа базисных векторов Крылова и работа с ними слишком затратны), мы можем найти подынтервал $[0, \delta]$, $\delta < t$, на котором невязка уже достаточно мала по норме. Тогда мы можем перезапустить метод, полагая $v := y_{k_{\max}}(\delta)$, уменьшая временной интервал $t := t - \delta$ и выполняя следующие $k_{\max}$ крыловских итераций для задачи (2) с обновленными v и t. Схема НВ перезапуска представлена на фиг. 1.



Фиг. 2. Норма невязки $\|r_k(s)\|$ как функция времени $s \in [0,t]$, $t = 1$, для обычного полиномиального (а) и СО (б) методов подпространства Крылова после $k = 10$ крыловских итераций. Величина сдвига $\gamma = t/20$. Матрица A – дискретизированный оператор конвекции-диффузии для числа Пекле $\mathrm{Pe} = 1000$ на сетке 402×402 (см. п. 3.2). Для обоих графиков норма невязки вычислена в 2000 равноотстоящих точках интервала [0, 1]. Горизонтальная сплошная линия на графике (б) показывает геометрическое среднее $\overline{r_k} = 2.75\mathrm{e}{-04}$ значений нормы невязки $\|r_k(s_i)\|$ в 2000 точках.

2.2. THB перезапуск: идеи и алгоритм

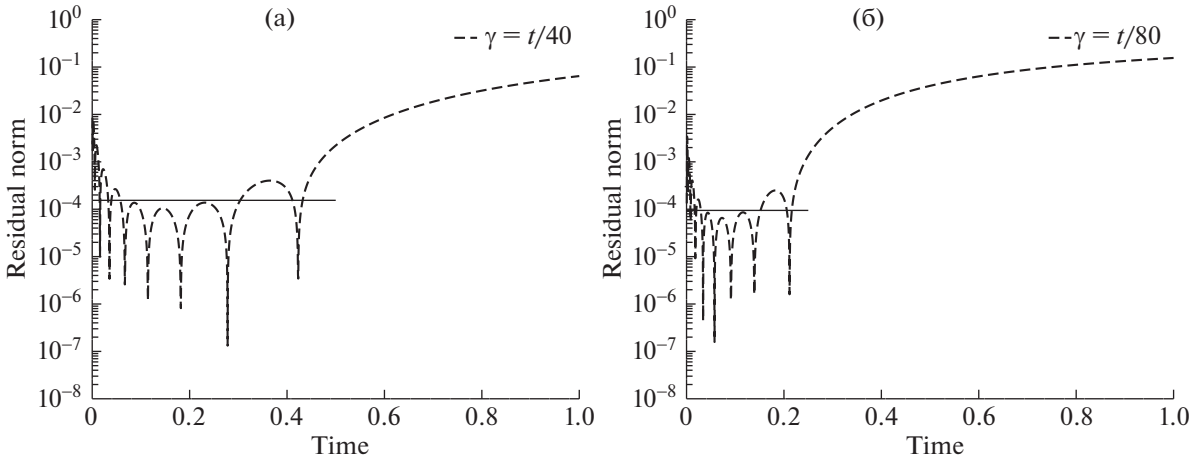
В методе подпространства Крылова СО (5)–(8) невязка как функция от t проявляет гораздо более нерегулярное поведение, чем в обычном полиномиальном методе подпространства Крылова (фиг. 2). Если для метода подпространства Крылова СО применяется НВ перезапуск, то может случиться, что такое δ, что $\|r_k(s)\|$ не превышает заданной точности для $s \in [0,\delta]$, не может быть найдено или слишком мало для практически эффективного перезапуска. Разумеется, можно устроить перезапуск, положив δ равной любой точке s, где $\|r_k(s)\|$ достаточно мала (фиг. 2). Однако гарантии, что $\min_{s\in[0,t]} \|r_k(s)\|$ не превосходит заданной точности, нет, и в таком случае перезапуск с любым $\delta \in [0,t]$ неизбежно приводит к потере точности.

В данной работе предлагается подход, позволяющий устранить этот недостаток НВ перезапуска для метода подпространства Крылова СО. Подход наш основан на следующих двух наблюдениях.

1. Поскольку γ обычно выбирается пропорционально t, меньшая величина сдвига γ означает более короткий временной интервал $[0,t]$. Для несимметричных матриц A невязка $r_k(s)$ в методе подпространства Крылова СО становится меньше по норме с уменьшением γ на некотором подынтервале $s \in [0,\delta]$, $0 < \delta < t$ (фиг. 2, 3).

2. Как уже обсуждалось выше, чтобы изменить γ в подпространстве Крылова СО, необходимо заново выполнить весь процесс Арнольди. Однако если линейная система с матрицей $I + \gamma A$ решена для определенного значения сдвига γ, то часть уже проделанной вычислительной работы может быть использована для решения линейных систем с меньшим значением сдвига $\tilde{\gamma} \le \gamma$. В частности, если для некоторого сдвига γ вычислено (разреженное) LU разложение, то оно может быть успешно использовано как предобусловливатель для решения линейных систем с новым значением сдвига, с матрицами $I + \tilde{\gamma}A$, $\tilde{\gamma} \le \gamma$ (см. утверждение 2).

На основе этих наблюдений для метода подпространства Крылова СО предлагается организовать НВ перезапуск без потери точности следующим образом. Предположим, что может быть выполнено не более $k_{\max}$ шагов процесса Арнольди или Ланцоша, так как хранение и использование более $k_{\max}$ крыловских векторов слишком затратны. Тогда выполняются итерации $k = 1, 2, \ldots, k_{\max}$ с контролем на каждой итерации нормы невязки $\|r_k(t)\|$ (см. (13)). Если норма невязки меньше заданной точности, процесс успешно оканчивается. Иначе после выполнения шага $k = k_{\max}$, значения функции $\|r_k(s)\|$ анализируются на отрезке $s \in [0,t]$. Если не может быть най-



Фиг. 3. Норма невязки $\|r_k(s)\|$ как функция времени $s \in [0, t]$, $t = 1$, для методов подпространства Крылова СО со сдвигом $\gamma = t/40$ (а) и $\gamma = t/80$ (б) после $k = 10$ крыловских итераций. Матрица A – дискретизированный оператор конвекции-диффузии для числа Пекле Pe = 1000 на сетке 402×402 (см. п. 3.2). Для обоих графиков норма невязки вычислена в 2000 равноотстоящих точках интервала [0, 1]. Горизонтальные сплошные линии показывают геометрические средние $\overline{r}_k$ значений нормы невязки, вычисленные для $s \in [0, 0.5]$ с $\overline{r}_k$ = 1.54e-04 (а) и $s \in [0, 0.25]$ с $\overline{r}_k$ = 9.44e-05 (б).

дено точки s такой, что норма $\|r_k(s)\|$ достаточно мала, то мы уменьшаем γ в два раза и повторяем шаги метода $k = 1, 2, \ldots, k_{\max}$. Затем процедура перезапуска (подробно описанная на фиг. 4) повторяется до тех пор, пока норма невязки не будет достаточно малой в конечной точке заданного временного интервала.

Если обычное или разреженное LU разложение слишком затратно, то для решения линейных систем СО может быть использован какой-либо предобусловленный итерационный метод. В этом случае в алгоритме перезапуска, представленном на фиг. 4, мы заменяем вычисление LU разложения построением предобусловливателя. Построенный предобусловливатель может быть использован для всех величин сдвига, т.е. достаточно построить предобусловливатель один раз.

Отметим, что для симметричных матриц A описанное выше поведение невязки в методе подпространства Крылова СО с уменьшением γ не наблюдается. Это подтверждается довольно точной оценкой леммы 3.1 из [8] (где полагаем $\mu = 0$ и выбираем γ пропорционально $t = \tau$).

Следующие лемма и утверждение показывают, что норма невязки $r_k(s)$ метода подпространства Крылова СО является функцией, ограниченной по времени. Оценка, дающая ограничение, зависит от γ.

Лемма 1. *Пусть для* $A \in \mathbb{R}^{n\times n}$ *выполняется соотношение* (1) *и пусть* H_k *– матрица, полученная в методе подпространства Крылова СО* (*см.* (11), (12)). *Тогда существует такая константа* $\omega_k \geq 0$, *что*

$$\|\exp(-tH_k)\| \leq e^{-t\omega_k}. \tag{14}$$

Доказательство. Пусть $\omega = \min_{x\in\mathbb{C}^n, \|x\|=1} \mathrm{Re}(x^*Ax)$. Известно (см., например, [35, Теорема 2.4]), что

$$\mathrm{Re}(x^*Ax) \geq \omega \quad \forall x \in \mathbb{C}^n \Leftrightarrow \|\exp(-tA)\| \leq e^{-t\omega}.$$

В силу (1) эти два эквивалентные соотношения выполняются для некоторой константы $\omega \geq 0$. Кроме того, эти соотношения эквивалентны неравенству (см. [35, теорема 2.13])

$$\left\|(I + \gamma A)^{-1}\right\| \leq \frac{1}{1 + \gamma\omega},$$

```
% Даны: A ∈ ℝ^{n×n}, v ∈ ℝ^n, t > 0, k_max и tol > 0
convergence := false
γ_changed := false
вычислить LU разложение LU := I + γA
while (not(convergence) and t > 0)
    β := ||v||, v_1 := v/β
    for k = 1,..., k_max
        if γ_changed
            решить (I + γA)w = v_k итерационно,
                с предобусловливателем LU
        else
            решить (I + γA)w = v_k LU разложением
        end
        for i = 1,..., k
            h̃_{i,k} := w^T v_i, w := w − h̃_{i,k} v_i
        end
        h_{k+1,k} := ||w||
        H_k := (1/γ)(H̃_k^{-1} − I)
        вычислить u(s_j), ||r_k(s_j)||, s_j = jt/3, j = 1, 2, 3
        resnorm := max_j ||r_k(s_j)||
        if resnorm ≤ tol and k > 1
            convergence := true
            прервать цикл for k = ...
        elseif k = k_max
            % -- перезапуск на шаге k_max
            вычислить ||r_k(s_j)||, s_j = jt/500, j = 1,..., 500
            r_min := min_j ||r_k(s_j)||
            if r_min > tol
                δ := 0
                γ := γ/2
                γ_changed := true
            else
                δ := max{s_j | ||r_k(s_j)|| ≤ tol}
            end
            u := exp(−δH_k)e_1, v := V_k(βu)
            t := t − δ
        end
        v_k+1 := w/h_{k+1,j}
    end
end
y_k := V_k(βu(s_3))
```

Фиг. 4. Алгоритм THB перезапуска метода подпространства Крылова CO. Вычисляется приближение $y_k(t) \approx \exp(-tA)v$, для невязки $r_k(t)$ которого $\|r_k(s)\| \le \text{tol}$ для $s = t/3$, $s = 2t/3$ и $s = t$.

которое выполняется для всех $\gamma > 0$ и всех $\omega \in \mathbb{R}$ при условии, что $1 + \gamma\omega > 0$. Пусть $\gamma > 0$ так же, как и в методе подпространства Крылова CO. Определим

$$\omega_k := \frac{1}{\gamma}\left(\left\|\tilde{H}_k\right\|^{-1} - 1\right)$$

так, чтобы $\left\|\tilde{H}_k\right\| = 1/(1 + \gamma\omega_k)$. Получаем

$$\frac{1}{1+\gamma\omega_k} = \left\|\tilde{H}_k\right\| = \left\|V_k^{\mathrm{T}}(I+\gamma A)^{-1}V_k\right\| \le \left\|(I+\gamma A)^{-1}\right\| \le \frac{1}{1+\gamma\omega},$$

откуда следует, что $0 \le \omega \le \omega_k$. Поскольку $\tilde{H}_k = (I + \gamma H_k)^{-1}$ (см. (12)), видим, снова используя (см. [35, теорема 2.13]), что

$$\left\|(I+\gamma H_k)\right\|^{-1} = \frac{1}{1+\gamma\omega_k} \le \frac{1}{1+\gamma\omega} \Leftrightarrow \mathrm{Re}(x^* H_k x) \ge \omega_k \quad \forall x \in \mathbb{C}^k. \tag{15}$$

Это равносильно искомому неравенству (14). Доказательство завершено.

Определим функцию $\varphi(z)$ (см., например, [1]):

$$\varphi(z) = (e^z - 1)/z. \tag{16}$$

Утверждение 1. *Пусть для* $A \in \mathbb{R}^{n\times n}$ *выполняется соотношение* (1) *и пусть* $r_k(t)$ *— невязка метода подпространства Крылова СО* (13) *при решении задачи* (2). *Тогда для всех* $t \geq 0$ *справедливо*

$$\begin{gathered} r_k(t) = \beta_k(t) w_{k+1}, \quad \beta_k(t) = \frac{\tilde{h}_{k+1,k}}{\gamma} e_k^{\mathrm{T}}(I + \gamma H_k)u(t), \quad w_{k+1} = (I + \gamma A)v_{k+1}, \\ \|r_k(t)\| = |\beta_k(t)|\|w_{k+1}\|, \end{gathered} \tag{17}$$

$$|\beta_k(t)| \leq \beta\tilde{h}_{k+1,k}\left(\frac{1}{\gamma}\min\left\{t\|(I + \gamma H_k)H_k\|\varphi(-t\omega_k), \|I + \gamma H_k\|(1 + e^{-t\omega_k})\right\} + |h_{k,1}|\right), \tag{18}$$

где функция $u(t)$ *определена в* (8), (12), $\omega_k \geq 0$ *— константа из* (14), *а* $\tilde{h}_{k+1,k}$ *и* $h_{k,1}$ *— элементы матриц* $\underline{\tilde{H}}_k \in \mathbb{R}^{(k+1)\times k}$ *и* $H_k \in \mathbb{R}^{k\times k}$ *соответственно* (см. (11), (12)). *Минимум в соотношении* (18) *берется по двум элементам множества, обозначенного* $\{\ldots\}$. *Подчеркнем, что* $\underline{\tilde{H}}_k$ *в оценке выше зависит от* γ (*а следовательно, также и* H_k, $u(t)$, ω_k).

Доказательство. Соотношение (17) идентично (13) (см. (12)), а доказательство (13) можно найти в [27]. В силу (7), (12) имеем

$$|\beta_k(0)| = \frac{\tilde{h}_{k+1,k}}{\gamma}\left|e_k^{\mathrm{T}}\tilde{H}_k^{-1}u(0)\right| = \frac{\tilde{h}_{k+1,k}}{\gamma}\left|e_k^{\mathrm{T}}(I + \gamma H_k)\beta e_1\right| = \tilde{h}_{k+1,k}\left|e_k^{\mathrm{T}}H_k\beta e_1\right| = \beta\tilde{h}_{k+1,k}|h_{k,1}|.$$

Далее нетрудно проверить, что (см. (16))

$$u(t) - u(0) = (\exp(-tH_k) - I)u(0) = -tH_k\varphi(-tH_k)u(0).$$

Поэтому, учитывая $u(0) = \beta e_1$ и $\tilde{H}_k^{-1} = I + \gamma H_k$, можно оценить

$$\begin{gathered} \|(I + \gamma H_k)(u(t) - u(0)\| = t\|(I + \gamma H_k)H_k\varphi(-tH_k)u(0)\| \leq \\ \leq t\|(I + \gamma H_k)H_k\|\|\varphi(-tH_k)\|\|u(0)\| \leq \beta t\|(I + \gamma H_k)H_k\|\varphi(-t\omega_k). \end{gathered} \tag{19}$$

Здесь используется неравенство

$$\|\varphi(-tH_k)\| \leq \varphi(-t\omega_k),$$

справедливое в силу (14), (15) (см., например, [1, доказательство леммы 2.4]). Оценка (19) особенно полезна для малых t. Получим теперь альтернативную оценку, которая может быть точнее для больших t:

$$\begin{gathered} \|(I + \gamma H_k)(u(t) - u(0)\| = \|(I + \gamma H_k)(\exp(-tH_k) - I)u(0)\| \leq \\ \leq \|I + \gamma H_k\|\|\exp(-tH_k) - I\|\|u(0)\| \leq \beta\|I + \gamma H_k\|(1 + \|\exp(-tH_k)\|) \leq \\ \leq \beta\|I + \gamma H_k\|(1 + e^{-t\omega_k}), \end{gathered} \tag{20}$$

где используется неравенство (14). Из (19), (20) следует, что

$$\|(I + \gamma H_k)(u(t) - u(0)\| \leq \beta\min\left\{t\|(I + \gamma H_k)H_k\|\varphi(-t\omega_k), \|I + \gamma H_k\|(1 + e^{-t\omega_k})\right\}.$$

Это позволяет оценить

$$|\beta_k(t)| \leq |\beta_k(t) - \beta_k(0)| + |\beta_k(0)| = \frac{\tilde{h}_{k+1,k}}{\gamma}\left|e_k^{\mathrm{T}}(I + \gamma H_k)(u(t) - u(0))\right| + \beta\tilde{h}_{k+1,k}|h_{k,1}| \leq$$

$$\leq \frac{\tilde{h}_{k+1,k}}{\gamma}\|(I + \gamma H_k)(u(t) - u(0)\| + \beta\tilde{h}_{k+1,k}|h_{k,1}| \leq$$

$$\leq \frac{\tilde{h}_{k+1,k}}{\gamma}\beta\min\left\{t\|(I + \gamma H_k)H_k\|\varphi(-t\omega_k), \|I + \gamma H_k\|(1 + e^{-t\omega_k})\right\} + \beta\tilde{h}_{k+1,k}|h_{k,1}| =$$

$$= \beta\tilde{h}_{k+1,k}\left(\frac{1}{\gamma}\min\left\{t\left\|(I+\gamma H_k)H_k\right\|\varphi(-t\omega_k), \left\|I+\gamma H_k\right\|(1+e^{-t\omega_k})\right\} + \left|h_{k,1}\right|\right),$$

что и дает (18). Доказательство завершено.

Подчеркнем, что оценка (18), к сожалению, не настолько точна, чтобы отразить зависимость $\|r_k(t)\|$ от γ (см. фиг. 2 и 3). Однако следует сделать следующее

Замечание 1. Численные эксперименты показывают, что величина $|h_{k,1}|$ (вспомним, что $|\beta_k(0)| = \beta\tilde{h}_{k+1,k}\,|h_{k,1}|$), появляющаяся в (18), обычно мала, много порядков меньше, чем другой член $\frac{1}{\gamma}\min\{\ldots\}$, участвующий в правой части (18). Если $|h_{k,1}| = 0$, то $\beta_k(0) = 0$. Таким образом, оценка (18) показывает, что для любого k и для любой точности $\varepsilon > 0$ можно найти такой временной интервал $[0, \delta]$, что $\|r_k(s)\| \le \varepsilon$ для $s \in [0, \delta]$. В этом случае временной интервал можно сократить (фиг. 1), поэтому как HB, так и THB перезапуски гарантируют сходимость метода подпространства Крылова CO для любой длины перезапуска. Разумеется, указанное δ может быть слишком мало, чтобы использовать его на практике. Поэтому подстройка параметра γ, как это делается в THB перезапуске, может быть необходима для успешной работы.

2.3. Решение сдвинутых линейных систем

Покажем теперь, что LU разложение, вычисленное для матрицы $I + \gamma A$, может быть успешно использовано как предобусловливатель для сдвинутой матрицы $I + \tilde{\gamma}A$ с уменьшенным значением сдвига $\tilde{\gamma}$, $0 < \tilde{\gamma} \le \gamma$. Говоря точнее, для решений сдвинутой линейной системы

$$\mathscr{A}x = b, \quad \mathscr{A} = I + \tilde{\gamma}A,$$

будем применять предобусловливатель

$$\mathscr{M}^{-1}\mathscr{A}x = \mathscr{M}^{-1}b, \quad \mathscr{M} = I + \gamma A. \tag{21}$$

Тогда нетрудно показать (см. утверждение 2 ниже), что даже метод простых итераций с таким предобусловливанием, точнее

$$x_{m+1} = \tilde{G}x_m + \mathscr{M}^{-1}b, \quad \tilde{G} = I - \mathscr{M}^{-1}\mathscr{A}, \tag{22}$$

сходится безусловно, т.е. для спектрального радиуса $\rho(\tilde{G})$ матрицы перехода $\tilde{G}$ выполняется $\rho(\tilde{G}) < 1$. Следовательно, собственные числа предобусловленной матрицы $\mathscr{M}^{-1}\mathscr{A}$ расположены на комплексной плоскости внутри круга единичного радиуса с центром в точке $1 + 0i$, $i^2 = -1$. Это означает, что современные итерационные методы подпространства Крылова, такие как GMRES, BiCGSTAB, QMR или аналогичные им (см. [29], [30], [36]), будут успешно решать предобусловленную линейную систему (21).

Однако чем меньше значение сдвига $\tilde{\gamma}$, тем лучше обусловлена сдвинутая матрица $I + \tilde{\gamma}A$. Следовательно, для малого $\tilde{\gamma}$ может оказаться, что непредобусловленный итерационный метод сходится достаточно быстро. Поэтому в утверждении 2 нами дается условие, достаточное для того, чтобы предобусловленный метод простых итераций сходился быстрее, чем непредобусловленный метод.

Утверждение 2. *Пусть для $A \in \mathbb{R}^{n\times n}$ выполняется соотношение* (1)*, $0 < \tilde{\gamma} \le \gamma$ и пусть линейная система с матрицей $I + \tilde{\gamma}A$ решается итерационно. Тогда предобусловленный метод простых итераций* (22) *с матрицей предобусловливателя $\mathscr{M} = I + \gamma A$ сходится.*

Кроме того, пусть метод простых итераций без предобусловливания также сходится. Тогда предобусловленный метод простых итераций (22) *с матрицей предобусловливателя $\mathscr{M} = I + \gamma A$ сходится быстрее, чем метод простых итераций без предобусловливания при условии, что*

$$\frac{1}{1+\gamma\rho(A)} < \frac{\tilde{\gamma}}{\gamma}, \tag{23}$$

где $\rho(A)$ – спектральный радиус матрицы A.

Доказательство. Пусть λ – некоторое собственное число матрицы A. Тогда собственные числа предобусловленной матрицы $(I + \gamma A)^{-1}(I + \tilde{\gamma}A)$ имеют вид

$$\frac{1 + \tilde{\gamma}\lambda}{1 + \gamma\lambda} = 1 - \left(1 - \frac{\tilde{\gamma}}{\gamma}\right)\frac{\gamma\lambda}{1 + \gamma\lambda}.$$

Предобусловленный метод простых итераций сходится тогда и только тогда, когда собственные числа матрицы перехода $\tilde{G} = I - (I + \gamma A)^{-1}(I + \tilde{\gamma}A)$ по модулю меньше единицы, т.е. если

$$\left|\left(1 - \frac{\tilde{\gamma}}{\gamma}\right)\frac{\gamma\lambda}{1 + \gamma\lambda}\right| < 1.$$

Левую часть этого неравенства можно оценить сверху:

$$\left|\left(1 - \frac{\tilde{\gamma}}{\gamma}\right)\frac{\gamma\lambda}{1 + \gamma\lambda}\right| \le \frac{|\gamma\lambda|}{|1 + \gamma\lambda|} < 1,$$

где второе неравенство выполняется потому, что собственные числа матрицы A имеют неотрицательную вещественную часть (см. (1)). Следовательно, предобусловленный метод простых итераций сходится.

Далее заметим, что матрица перехода G метода простых итераций без предобусловливания – это матрица $G = I - (I + \tilde{\gamma}A) = -\tilde{\gamma}A$. Предобусловленный метод сходится быстрее, чем непредобусловленный, при условии $\rho(\tilde{G}) < \rho(G)$, т.е. если

$$\left(1 - \frac{\tilde{\gamma}}{\gamma}\right)\max_{\lambda}\left|\frac{\gamma\lambda}{1 + \gamma\lambda}\right| < \tilde{\gamma}\max_{\lambda}|\lambda| = \tilde{\gamma}\rho(A).$$

Левую часть здесь можно оценить величиной $1 - \dfrac{\tilde{\gamma}}{\gamma}$, так что неравенство выполняется, если

$$\left(1 - \frac{\tilde{\gamma}}{\gamma}\right) < \tilde{\gamma}\rho(A).$$

Легко проверить, что это неравенство равносильно (23). Доказательство завершено.

3. ЧИСЛЕННЫЕ ЭКСПЕРИМЕНТЫ

3.1. Подробности экспериментов

Мы реализовали ТНВ перезапуск так, как показано на фиг. 4, со следующими модификациями.

1. Находимое алгоритмом приближенное наименьшее значение нормы невязки зависит от числа точек, в которых вычисляется норма невязки. Поэтому это число точек, обозначенное в описании алгоритма (фиг. 4) n_{steps}, задается в соответствии с требуемой точностью `tol`:

tol	1e–06	1e–07	≤1e–08
n_{steps}	500	1000	2000

2. Если величина δ (т.е. такая величина, что $\|r_k(\delta)\| \le \text{tol}$, см. фиг. 4) остается равной нулю после двух последовательных уменьшений γ, то γ получает свое изначальное значение 4γ, помноженное на 0.8, т.е. $\gamma := 0.8 \times 4 \times \gamma$, а число тестовых точек n_{steps} удваивается. После этого обычная работа алгоритма продолжается. Это означает, что если изначально $\gamma = 1$, γ в алгоритме последовательно принимает значения 1, 0.5, 0.25, 0.8, 0.4, 0.2, 0.64, 0.32, ... Как только найдено положительное δ, т.е. перезапуск успешен, число тестовых точек n_{steps} уменьшается до 500.

3. Как только в ходе выполнения алгоритма γ меняется, пропорционально меняется и длина временного интервала, на котором по n_{steps} тестовым точкам ищется приближенное минимальное значение нормы невязки. Например, если значение γ уменьшается вдвое, то временной интервал поиска меняется от $[0, t]$ к $[0, t/2]$. Это делается потому, что $\|r_k(s)\|$ навряд ли будет мала для

$s > t/2$ (фиг. 3). Как только перезапуск оказался успешным, т.е. $\delta > 0$ и временной интервал уменьшен (строка алгоритма $t := t - \delta$), мы увеличиваем интервал поиска до $[0, t]$.

4. Последняя модификация состоит в том, что в нашей реализации итерационный метод GMRES(10) (см. [37]) может быть использован вместо LU разложения не только тогда, когда сдвиг γ уменьшен. В качестве предобусловливателя с GMRES(10) может быть использовано неполное LU разложение ILUT(ε) (см. [38, гл. 10]. Оно вычисляется один раз и используется для всех значений γ. Мы используем реализацию GMRES, доступную на сайте www.netlib.org/templates/, из [36]. Предобусловливатель применяется справа, а в качестве остановочного критерия итераций GMRES(10) при решении линейных систем "сдвиг–обращение" $(I + \gamma A)x = b$ используется такое условие на невязку res_i линейной системы:

$$\|\mathrm{res}_i\| \le \mathrm{tol}_{\mathrm{gmres}}\|b\|, \quad \text{with} \quad \mathrm{tol}_{\mathrm{gmres}} = \min\{1\mathrm{e}{-}08,\ \mathrm{tol}/10\},$$

где tol – требуемая заданная точность вычисления действия матричной экспоненты. При малых значениях γ может быть разумным отключить предобусловливатель (см. утверждение 2).

Начальное значение γ может быть задано как необязательный параметр нашей процедуры THB перезапуска. По умолчанию, если начальное значение γ не задано пользователем, оно устанавливается равным $t/20$. Заметим, что значение $\gamma = t/10$ рекомендуется в [8] для умеренных величин допустимой точности $\mathrm{tol} \approx 10^{-6}$ для симметричных матриц. Выбор начального значения $\gamma := t/20$ в THB перезапуске представляется разумным выбором, поскольку, согласно нашему опыту, оптимальное значение γ для несимметричных матриц обычно меньше, чем $t/10$.

В экспериментах, представленных ниже, в рамках метода подпространства Крылова СО мы сравниваем THB перезапуск с HB перезапуском. В работе [22] HB перезапуск был всесторонне протестирован и сравнен с другими тремя методами перезапуска: перезапуском пакета EXPOKIT (см. [39]), перезапуском Нихоффа–Хохбрук (см. [23, гл. 3]) и невязочным перезапуском (см. [12], [27]). Значения ошибок, представленные в этом разделе, получены для численного решения $y_k(t)$, как

$$\frac{\|y_k(t) - y_{\mathrm{ref}}(t)\|}{\|y_{\mathrm{ref}}(t)\|},$$

где $y_{\mathrm{ref}}(t)$ – референтное решение, вычисленное с высокой точностью функцией phiv пакета EXPOKIT (см. [39]). Численные тесты были проведены в Матлабе на линукс-компьютере с 8 процессорами Intel Xeon E5504 2.00GHz.

3.2. Задача конвекции-диффузии

В этой задаче матрица A получается стандартной пятиточечной конечно-разностной аппроксимацией оператора конвекции-диффузии, определенного для функций $u(x, y)$ с $(x, y) \in \Omega = [0,1]\times[0,1]$ и $u|_{\partial\Omega} = 0$. Оператор имеет вид

$$L[u] = -(D_1 u_x)_x - (D_2 u_y)_y + \mathrm{Pe}\left(\frac{1}{2}(v_1 u_x + v_2 u_y) + \frac{1}{2}((v_1 u)_x + (v_2 u_y)_y)\right),$$

$$D_1(x,y) = \begin{cases} 10^3 & (x,y) \in [0.25,\ 0.75]^2, \\ 1 & \text{иначе}, \end{cases} \qquad D_2(x,y) = \frac{1}{2} D_1(x,y),$$

$$v_1(x,y) = x + y, \quad v_2(x,y) = x - y,$$

где Pe – число Пекле. Здесь конвективные члены (первые производные) записаны в специальном виде так, чтобы их вклады в матрицу A представляли собой кососимметричную матрицу (см. [40]). В экспериментах использовалась равномерная сетка 802×802 и значения числа Пекле $\mathrm{Pe} = 200$ и $\mathrm{Pe} = 1000$. Размер задачи для этой сетки – $n = 800^2 = 640\,000$. Для обоих значений числа Пекле $\left\|\frac{1}{2}\left(A + A^{\mathrm{T}}\right)\right\|_2 \approx 6000$, в то время как $\left\|\frac{1}{2}\left(A - A^{\mathrm{T}}\right)\right\|_2 \approx 0.5$ для $\mathrm{Pe} = 200$ и $\left\|\frac{1}{2}\left(A - A^{\mathrm{T}}\right)\right\|_2 \approx 2.5$ для $\mathrm{Pe} = 1000$. Следовательно, в обоих случаях матрицы можно считать слабо несимметричными. Значения функции $\sin(\pi x)\sin(\pi y)$ на конечно-разностной сетке присваивались начальному вектору v, который затем нормализовывался $v := v/\|v\|$. Задавалось конечное время $t = 1$.

Таблица 1. Результаты для тестовой задачи конвекции-диффузии, сетка 802×802, конечное время $t = 1$

Метод	Точности заданная, полученная	Процессорное время, с	Число итераций (итераций GMRES(10))
Pe = 200, длина перезапуска 10			
HB, разреж. LU	1e–06, 2.50e–07	46.2	20 (–)
HB, разреж. LU	1e–08, 2.51e–07	48.0	27 (–)
THB, разреж. LU, GMRES(10)	1e–08, 1.60e–08	484	73 (962)
THB, разреж. LU, GMRES(10), найденное γ	1e–08, 1.65e–08	56.2	53 (–)
HB, GMRES(10)/ILUT	1e–08, 2.54e–07	90.5	36 (440)
THB, GMRES(10)/ILUT	1e–08, 1.85e–08	191	77 (1258)
THB, GMRES(10)/ILUT, найденное γ	1e–08, 1.51e–08	68.2	57 (342)
Pe = 1000, длина перезапуска 10			
HB, GMRES(10)/ILUT	1e–06, 3.36e–07	49.2	14 (154)
HB, GMRES(10)/ILUT	1e–08, 3.52e–07	72.9	27 (325)
THB, GMRES(10)/ILUT	1e–06, 2.43e–07	44.5	17 (136)
THB, GMRES(10)/ILUT	1e–08, 7.55e–08	110	55 (630)
THB, GMRES(10)/ILUT, найденное γ	1e–08, 7.41e–08	52.8	25 (205)

В экспериментах начальное значение сдвига γ в THB перезапуске не задавалось и было по умолчанию $t/20$, а в HB перезапуске использовалось обычное значение $\gamma = t/10$. Это не обязательно дает преимущество THB перезапуску, потому что оптимальные значения γ, находимые в THB перезапуске, все равно были меньше, чем $t/20$.

Результаты для этой тестовой задачи представлены в табл. 1. Как можно увидеть в первых двух строках таблицы, HB перезапуск не в состоянии дать меньшую ошибку при уменьшенном значении допустимой точности, несмотря на возросший объем вычислений (27 вместо 20 итераций). В то же время THB перезапуск, где используется тот же самый линейный решатель – разреженное LU разложение – справляется с поставленной задачей, хотя процессорное время и увеличивается в 10 раз (см. строку 3 табл. 1). Отметим, что процессорное время, измеренное средствами Матлаба, не всегда правильно отражает вычислительную работу. В данном случае оно не соответствует числу крыловских шагов метода (73 шага с THB перезапуском вместо 27 шагов с HB перезапуском). Причина здесь в том, что прямые методы решения линейных систем (LU разложение и оператор действия обратной матрицы "обратная дробная черта" \) реализованы в среде Матлаб весьма эффективно, чего нельзя сказать об итерационных решателях. Из-за этого процессорное время вычисления действий предобусловливателя внутри GMRES(10) оказывается значительным. Однако, как только алгоритм определил подходящее значение γ, это значение успешно может быть использовано для повторных вычислений действия матричной экспоненты: в этом случае, как видим в строке 4 табл. 1, мы получаем требуемую высокую точность при небольшом увеличении расчетного времени.

Мы также тестируем наш подход на этой задаче с итерационным линейным решателем – итерационным методом GMRES(10) с предобусловливателем ILUT($\varepsilon = 10^{-3}$). В строке 5 табл. 1 показано, что HB перезапуск в комбинации с предобусловленным GMRES требует 36 шагов Арнольди (вместо 27 шагов для HB перезапуска в комбинации с прямым методом решения, см. табл. 1, строка 2), потому что невязки в этих двух реализациях метода (с LU разложением и с методом GMRES(10)) слегка отличаются. THB перезапуск с тем же самым предобусловленным итерационным методом требует в два раза больше процессорного времени, чем HB перезапуск, но дает меньшую ошибку (см. табл. 1, строка 6). Наконец, в последней строке таблицы мы видим, что коль скоро правильное значение сдвига определено THB алгоритмом, THB перезапуск позволяет получать лучшую точность при сравнимых вычислительных затратах.

В нижней части табл. 1 представлены результаты для большего числа Пекле. При требуемой точности `tol = 1e-06` перезапуск HB дает результат с точностью `3.36e-07`, что вполне удовлетворительно. Однако при требуемой точности `tol = 1e-08` метод оказывается не в состоя-

нии получить меньшую ошибку, хотя вычислительные затраты выросли с 14 до 27 шагов. В следующих двух строках таблицы показаны результаты для THB перезапуска. THB перезапуск позволяет получить более точный результат при возросшем процессорном времени. Из последней строки табл. 1 следует, что как только оптимальное значение γ найдено, аналогичная точность может быть получена за примерно то же процессорное время.

3.3. Уравнения Максвелла в среде без потерь

Рассмотрим уравнения Максвелла в трехмерной области, заполненной непроводящей средой без источников электромагнитного поля:

$$\frac{\partial \mathbf{H}}{\partial t} = -\frac{1}{\mu}\nabla \times \mathbf{E}, \qquad \frac{\partial \mathbf{E}}{\partial t} = \frac{1}{\varepsilon}\nabla \times \mathbf{H}. \tag{24}$$

Здесь ε и μ – относительные диэлектрическая и магнитная проницаемости соответственно, являющиеся скалярными функциями переменных (x, y, z), а магнитное поле $\mathbf{H}$ и электрическое поле $\mathbf{E}$ – неизвестные вектор-функции переменных (x, y, z, t). Краевые условия задают на границе области нулевые тангенциальные компоненты электрического поля, что физически означает идеально проводящую границу области или так называемые краевые условия “большого бака” (см. [41], [42]). Данная модельная задача взята из [42]: в пространственной области $[-6.05, 6.05]\times[-6.05, 6.05]\times[-6.05, 6.05]$, наполненной воздухом (относительная диэлектрическая проницаемость $\varepsilon_r = 1$), помещен образец из материала с относительной диэлектрической проницаемостью $\varepsilon_r = 5.0$, занимающий подобласть $[-4.55, 4.55]\times[-4.55, 4.55]\times[-4.55, 4.55]$. В образце имеются 27 сферических отверстий ($\varepsilon_r = 1$) радиуса 1.4, центры отверстий расположены в точках $(x_i, y_j, z_k) = (3.03i, 3.03j, 3.03k)$, $i, j, k = -1, 0, 1$. Задаются нулевые начальные значения для всех компонент обоих полей $\mathbf{H}$ и $\mathbf{E}$, кроме компонент x и y поля $\mathbf{E}$. Последние не равны нулю в центре области и представляют собой световой импульс.

Дискретизация по пространству центральными, разнесенными по сетке, конечными разностями (схема Йи (Yee)) приводит в системе дифференциальных уравнений вида (3). Пространственная сетка в этом тесте состоит из $40\times40\times40$ или $80\times80\times80$ ячеек Йи, так что размер задачи $n = 413\,526$ или $n = 3\,188\,646$ соответственно. После дискретизации вектор начальных значений $v \in \mathbb{R}^n$ нормализуется $v := v/\|v\|$. Сравнение результатов, полученных на этих двух пространственных сетках, показывает, что разрешение сетки достаточно для этого теста. Конечное время $t = 1$.

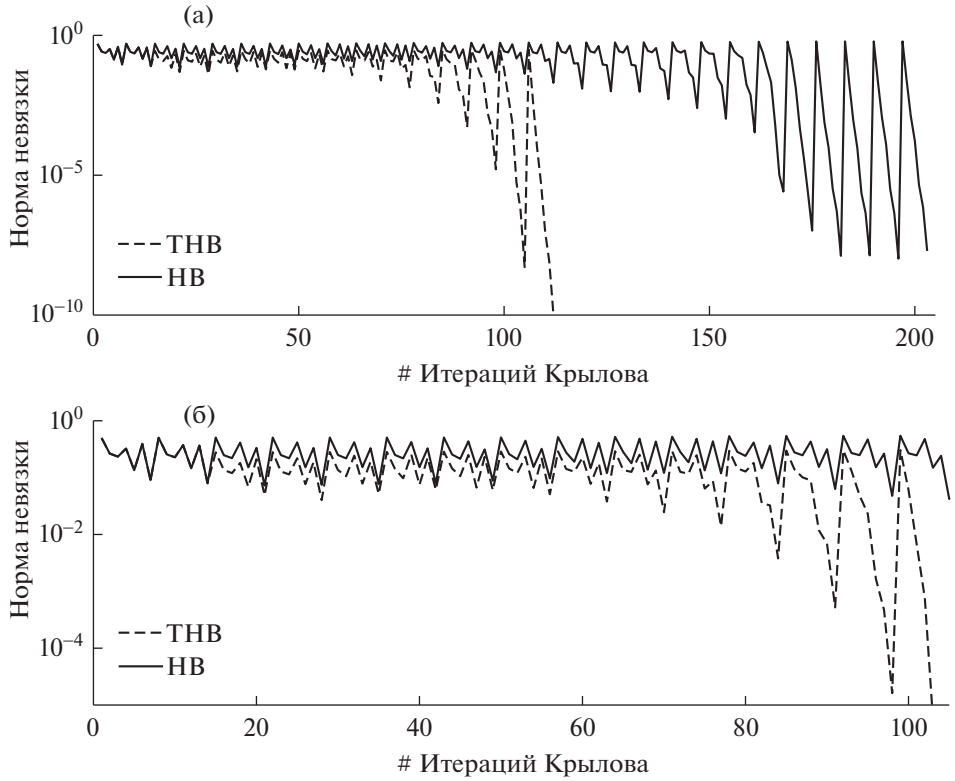
Этот тест является сложным для метода подпространства Крылова CO, потому что матрица A сильно несимметрична (можно выбрать такую диагональную матрицу D, что $D^{-1}AD$ – кососимметричная). Для сильно несимметричных задач, таких как дискретизированные уравнения Максвелла в непроводящей среде, методы подпространства Крылова CO зачастую не являются эффективными (см. [43]). Действительно, другие методы подпространства Крылова с перезапуском оказываются более эффективными в этой тестовой задаче (см. [22]). Кроме того, это – трехмерная векторная задача, где 6 переменных (компоненты x, y и z обоих полей) соответствуют каждой ячейке вычислительной сетки. Поэтому в зависимости от конкретных значений параметров задачи решать линейные системы со сдвинутой матрицей $I + \gamma A$ может быть весьма непросто. Тем не менее для данной конкретной задачи оказывается, что величина $\gamma\|A\|$ достаточно мала, так что условие (23) не выполняется и даже непредобусловленный метод простых итераций может быть успешно использован для решения сдвинутых линейных систем. В представленных здесь экспериментах для этой цели был использован итерационный метод GMRES(10). Таким образом, мы рассматриваем эту тестовую задачу, чтобы показать возможности предложенного THB перезапуска.

Опыт показывает, что для успешной работы с сильно несимметричными матрицами A методы подпространства Крылова CO должны использовать гораздо меньшие величины сдвига γ, чем обычно используемое значение $t/10$ (см. [44]). Поэтому мы задаем для γ значение $t/80 = 1/80$ в обоих методах перезапуска (при этом THB перезапуск при необходимости может уменьшить это значение). Результаты представлены в табл. 2 и на фиг. 5 и 6. THB перезапуск очевидно пре-

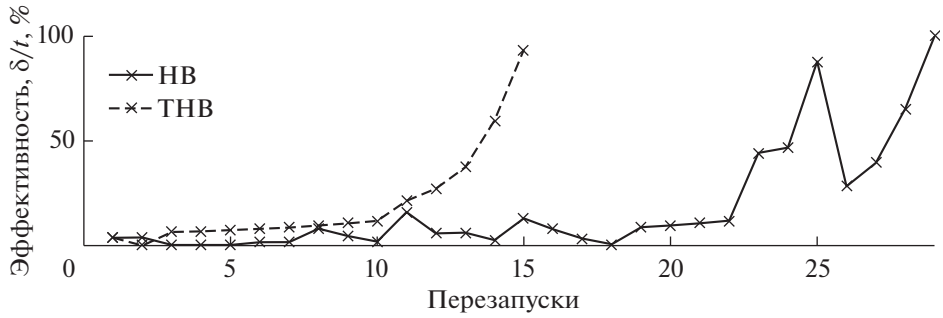
Таблица 2. Результаты для уравнений Максвелла в непроводящей среде

Метод	Точности заданная, полученная	Процессорное время, с	Число итераций (итераций GMRES(10))
Сетка $40 \times 40 \times 40$, длина перезапуска 7			
HB, GMRES(10)	1e–09, 2.96e–07	113.7	203 (1827)
THB, GMRES(10)	1e–09, 9.51e–08	48.8	112 (812)
THB, GMRES(10)	1e–10, 1.22e–08	78.1	161 (1302)
Сетка $80 \times 80 \times 80$, длина перезапуска 8			
HB, GMRES(10)	1e–09, 5.33e–07	2067	352 (4263)
THB, GMRES(10)	1e–09, 3.90e–08	785	191 (1410)

восходит HB перезапуск как по вычислительным затратам, так и по полученной точности. Потеря точности в HB перезапуске происходит на первых перезапусках, когда норма невязки оказывается больше требуемой точности по всему временному интервалу. Во избежание потери точности THB перезапуск уменьшает γ, что не только восстанавливает точность, но и приводит к более быстрому решению сдвинутых линейных систем (напомним, что при меньших значениях γ непредобусловленный GMRES сходится быстрее). Более того, уменьшенные значения γ при этом приводят к дальнейшему выигрышу эффективности в THB перезапуске. Как видно на фиг. 6, этот выигрыш получается из-за больших временных интервалов $[0, \delta]$, на которых норма невязки оказывается меньше заданной точности (напомним, что временной интервал на каждом перезапуске сокращается с $[0, t]$ до $[0, t - \delta]$).



Фиг. 5. (а) – Сходимость метода подпространства Крылова CO с HB перезапуском (сплошная линия) и THB перезапуском (штриховая линия) для уравнений Максвелла в непроводящей среде, сетка $40 \times 40 \times 40$, длина перезапуска – 7. (б) – Увеличение верхнего графика. Каждый зигзаг соответствует перезапуску.



Фиг. 6. Эффективность перезапусков, представленная как отношение сокращаемой части временного интервала δ к оставшемуся временному интервалу t. Тестовая задача: уравнения Максвелла в непроводящей среде, сетка $40 \times 40 \times 40$, длина перезапуска – 7. Эффективность 0% на втором ТНВ перезапуске означает уменьшение параметра γ.

4. ВЫВОДЫ

Предлагаемый ТНВ (точный невязочно-временной) перезапуск представляется полезным подходом для повышения эффективности методов подпространства Крылова СО ("сдвиг–обращение") при вычислении действий матричной экспоненты несимметричных матриц. Данный подход обладает всеми свойствами обычного НВ (невязочно-временного) перезапуска, а также позволяет предотвратить потерю точности при сохранении эффективности НВ перезапуска.

Можно обозначить несколько направлений дальнейших исследований. Во-первых, поиск минимума нормы невязки сейчас выполняется на равномерной сетке точек временного интервала. Этот поиск может быть организован более эффективно на неравномерной сетке, сгущающейся в областях, где норма невязки имеет локальные минимумы. Может быть разработана адаптивная процедура для построения такой сетки. Кроме того, следует изучить вопрос о том, как обобщить данный подход на симметричные матрицы. Мы надеемся заняться этими вопросами в будущем.

СПИСОК ЛИТЕРАТУРЫ

1. *Hochbruck M., Ostermann A.* Exponential integrators // Acta Numer. 2010. V. 19. P. 209–286.
2. *De la Cruz Cabrera O., Matar M., Reichel L.* Analysis of directed networks via the matrix exponential // J. of Comput. and Appl. Math. 2019. V. 355. P. 182–192. Access mode: https://doi.org/10.1016/j.cam.2019.01.015
3. *Kürschner P.* Balanced truncation model order reduction in limited time intervals for large systems // Adv. Comput. Math. 2018. V. 44. https://doi.org/10.1007/s10444-018-9608-6
4. *Frommer A., Simoncini V.* Matrix functions // Model Order Reduction: Theory, Research Aspects and Applications / Ed. by Wil H. A. Schilders, Henk A. van der Vorst, Joost Rommes. Springer, 2008. P. 275–304.
5. *Bergamaschi L., Vianello M.* Efficient computation of the exponential operator for large, sparse, symmetric matrices // Numer. Linear Algebra with Appl. 2000. V. 7. № 1. P. 27–45.
6. *Al-Mohy A.H., Higham N.J.* Computing the action of the matrix exponential, with an application to exponential integrators // SIAM J. Sci. Comput. 2011. V. 33. № 2. P. 488–511. https://doi.org/10.1137/100788860
7. *Moret I., Novati P.* RD rational approximations of the matrix exponential // BIT. 2004. V. 44. P. 595–615.
8. *van den Eshof J., Hochbruck M.* Preconditioning Lanczos approximations to the matrix exponential // SIAM J. Sci. Comput. 2006. V. 27. № 4. P. 1438–1457.
9. *Druskin V., Knizhnerman L., Zaslavsky M.* Solution of large scale evolutionary problems using rational Krylov subspaces with optimized shifts // SIAM J. on Sci. Comput. 2009. V. 31. № 5. P. 3760–3780.
10. *Güttel S.* Rational Krylov Methods for Operator Functions: Ph. D. thesis / Stefan Güttel; Technischen Universität Bergakademie Freiberg. 2010. March. www.guettel.com.
11. *Güttel S.* Rational Krylov approximation of matrix functions: Numerical methods and optimal pole selection // GAMM Mitteilungen. 2013. V. 36. № 1. P. 8–31. www.guettel.com.

12. *Celledoni E., Moret I.* A Krylov projection method for systems of ODEs // Appl. Numer. Math. 1997. V. 24. № 2–3. P. 365–378.

13. *Tal-Ezer H.* On restart and error estimation for Krylov approximation of $w = f(A)v$ // SIAM J. Sci. Comput. 2007. V. 29. № 6. P. 2426–2441. Access mode: https://doi.org/10.1137/ 040617868

14. *Afanasjew M., Eiermann M., Ernst O.G., Güttel S.* Implementation of a restarted Krylov subspace method for the evaluation of matrix functions // Linear Algebra Appl. 2008. V. 429. P. 2293–2314.

15. *Eiermann M., Ernst O.G., Güttel S.* Deflated restarting for matrix functions // SIAM J. Matrix Anal. Appl. 2011. V. 32. № 2. P. 621–641.

16. *Güttel S., Frommer A., Schweitzer M.* Efficient and stable Arnoldi restarts for matrix functions based on quadrature // SIAM J. Matrix Anal. Appl. 2014. V. 35. № 2. P. 661–683.

17. *Jawecki T., Auzinger W., Koch O.* Computable strict upper bounds for Krylov approximations to a class of matrix exponentials and ϕ-functions // arXiv preprint arXiv:1809.03369. 2018. https://arxiv.org/pdf/1809.03369.

18. *Hochbruck M., Lubich C.* Exponential integrators for quantum-classical molecular dynamics // BIT. 1999. V. 39. № 4. P. 620–645.

19. *Hochbruck M., Pažur T., Schulz A. et al.* Efficient time integration for discontinuous Galerkin approximations of linear wave equations // ZAMM. 2015. V. 95. № 3. P. 237–259. Access mode: https://doi.org/10.1002/zamm.201300306

20. *Börner R.-U., Ernst O.G., Güttel S.* Three-dimensional transient electromagnetic modeling using rational Krylov methods // Geophys. J. Internat. 2015. V. 202. № 3. P. 2025–2043.

21. *Botchev M.A., Hanse A.M., Uppu R.* Exponential Krylov time integration for modeling multifrequency optical response with monochromatic sources // J. Comput. Appl. Math. 2018. V. 340. P. 474–485. https://doi.org/10.1016/j.cam.2017.12.014

22. *Botchev M.A., Knizhnerman L.A.* ART: Adaptive residual-time restarting for Krylov subspace matrix exponential evaluations // J. Comput. Appl. Math. 2020. V. 364. № 112311. https://doi.org/10.1016/j.cam.2019.06.027

23. *Niehoff J.* Projektionsverfahren zur Approximation von Matrixfunktionen mit Anwendungen auf die Implementierung exponentieller Integratoren: Ph.D. thesis / Jörg Niehoff; Mathematisch-Naturwissenschaftlichen Fakultät der Heinrich-Heine-Universität Düusseldorf. 2006. December.

24. *Golub G.H., Van Loan C.F.* Matrix Computations. Third edition. Baltimore and London: The Johns Hopkins Univ. Press, 1996. P. 694.

25. *Moler C.B., Van Loan C.F.* Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later // SIAM Rev. 2003. V. 45. № 1. P. 3–49.

26. *Higham N.J.* Functions of Matrices: Theory and Computation. Philadelphia, PA, USA: Soc. for Industrial and Appl. Math., 2008.

27. *Botchev M.A., Grimm V., Hochbruck M.* Residual, restarting and Richardson iteration for the matrix exponential // SIAM J. Sci. Comput. 2013. V. 35. № 3. P. A1376–A1397. https://doi.org/10.1137/110820191

28. *Parlett B.N.* The Symmetric Eigenvalue Problem. SIAM, 1998.

29. *van der Vorst H.A.* Iterative Krylov methods for large linear systems. Cambridge Univ. Press, 2003.

30. *Saad Y.* Iterative Methods for Sparse Linear Systems. 2d ed. SIAM, 2003. Available from http://www-users.cs.umn.edu/~saad/books.html.

31. *Druskin V.L., Knizhnerman L.A.* Two polynomial methods of calculating functions of symmetric matrices // U.S.S.R. Comput. Maths. Math. Phys. 1989. V. 29. № 6. P. 112–121.

32. *Knizhnerman L.A.* Calculation of functions of unsymmetric matrices using Arnoldi's method // U.S.S.R. Comput. Math. Math. Phys. 1991. V. 31. № 1. P. 1–9.

33. *Hochbruck M., Lubich C.* On Krylov subspace approximations to the matrix exponential operator // SIAM J. Numer. Anal. 1997. V. 34. № 5. P. 1911–1925.

34. *Druskin V.L., Greenbaum A., Knizhnerman L.A.* Using nonorthogonal Lanczos vectors in the computation of matrix functions // SIAM J. Sci. Comput. 1998. V. 19. № 1. P. 38–54.

35. *Hundsdorfer W., Verwer J.G.* Numerical Solution of Time-Dependent Advection-Diffusion-Reaction Equations. Springer Verlag, 2003.

36. *Barrett R., Berry M., Chan T.F. et al.* Templates for the Solution of Linear Systems: Building Blocks for Iterative Methods / Philadelphia, PA: SIAM, 1994. Available at www.netlib.org/templates/.

37. *Saad Y., Schultz M.H.* GMRES: a generalized minimal residual algorithm for solving nonsymmetric linear systems // SIAM J. Sci. Stat. Comput. 1986. V. 7. № 3. P. 856–869.

38. *Saad Y.* Iterative Methods for Sparse Linear Systems. Book out of print, 2000. www-users.cs.umn.edu/~saad/books.html.

39. *Sidje R.B.* Expokit. A software package for computing matrix exponentials // ACM Trans. Math. Softw. 1998. V. 24. № 1. P. 130–156. www.maths.uq.edu.au/expokit/.

40. *Krukier L.A.* Implicit difference schemes and an iterative method for solving them for a certain class of systems of quasi-linear equations // Sov. Math. 1979. V. 23. № 7. P. 43–55. Translation from Izv. Vyssh. Uchebn. Zaved., Mat. 1979. № 7(206). P. 41–52.

41. *Taflove A., Hagness S.C.* Computational electrodynamics: the finite-difference time-domain method. 3d ed. Boston, MA: Artech House Inc., 2005.

42. *Kole J.S., Figge M.T., De Raedt H.* Unconditionally stable algorithms to solve the timedependent Maxwell equations // Phys. Rev. E. 2001. V. 64. P. 066705.

43. *Verwer J.G., Botchev M.A.* Unconditionally stable integration of Maxwell's equations // Linear Algebra and its Applications. 2009. V. 431. № 3–4. P. 300–317.

44. *Botchev M.A.* Krylov subspace exponential time domain solution of Maxwell's equations in photonic crystal modeling // J. Comput. Appl. Math. 2016. V. 293. P. 24–30. https://doi.org/10.1016/j.cam.2015.04.022