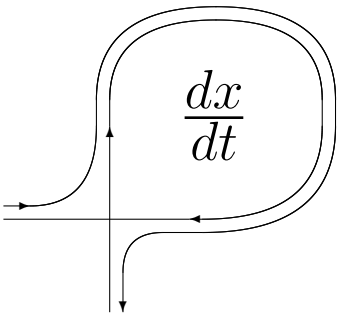

ДИФФЕРЕНЦИАЛЬНЫЕ УРАВНЕНИЯ
И
ПРОЦЕССЫ УПРАВЛЕНИЯ
N. 4, 2024
Электронный журнал,
рег. Эл. N ФС77-39410 от 15.04.2010
ISSN 1817-2172

http://diffjournal.spbu.ru/
e-mail: jodiff@mail.ru

Численные методы

Применение модифицированного метода Рунге — Кутты к построению метода спуска для решения краевых задач

Г. В. Кривовичев, Н. В. Егоров

Факультет прикладной математики - процессов управления
Санкт-Петербургского государственного университета
E-mail: g.krivovichev@spbu.ru, n.v.egorov@spbu.ru

Аннотация. Работа посвящена построению и анализу градиентного метода, основанного на модифицированном явном методе Рунге — Кутты второго порядка, построенном с использованием разложения Лагранжа — Бюрмана. Предложен двухшаговый метод с инерцией, основанный на методе тяжелого шарика. Доказаны теоремы о сходимости для сильно выпуклой квадратичной и возмущенной квадратичной функции. Получены аналитические выражения для параметров метода, обеспечивающие оптимальную скорость сходимости. В случае квадратичной функции показано, что предложенный метод сходится быстрее, чем другие ускоренные методы.

Представлены результаты применения метода к численному решению линейных и нелинейных краевых задач: задачи Дирихле для трехмерного уравнения Пуассона, задачи вариационного исчисления, задач для интегро-дифференциальных уравнений. Показано, что по сравнению с известными методами, предложенный метод позволяет получать численное решение с нужной точностью при разных разбиениях сетки за меньшее число итераций и время.

Ключевые слова: методы Рунге — Кутты, выпуклая оптимизация, градиентный спуск, краевые задачи.

1 Введение

В настоящее время для решения задач оптимизации активно используются методы, разработанные для решения задачи Коши для обыкновенных дифференциальных уравнений (ОДУ) [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11]. В частности, ряд градиентных методов построен при применении методов Рунге — Кутты (МРК).

Применению МРК к решению разных задач оптимизации посвящено много работ. В [12] описано применение явных МРК к решению задачи минимизации, эквивалентной решению системы нелинейных алгебраических уравнений. В [13] показано, что использование МРК при решении таких задач приводит к ускорению сходимости по сравнению с методом градиентного спуска. В [14] на примерах разных задач безусловной оптимизации проведен сравнительный анализ градиентных методов на основе явных и неявных МРК. В [15] неявные МРК низких порядков точности используются для построения квазиньютоновских методов. Работа [2] посвящена построению предобусловленных градиентных методов на основе явных и неявных МРК. Авторами показано, что ограничения на шаг, обеспечивающие сходимость, связаны с условиями устойчивости используемых МРК. Также показано, что такие методы можно использовать для построения преобуславливателей для метода сопряженных градиентов. В [16] авторами предложено ОДУ второго порядка, эквивалентное методу Нестерова и проведено исследование свойств его решений. В [5, 6] построена модификация этого ОДУ и с помощью метода функций Ляпунова показано, что применение явных МРК к решению задачи Коши для такого уравнения приводит к ускорению сходимости при применении метода к специальному классу функций. В [8] представлена динамическая система (т.н. Nesterov gradient flow), построенная как база для ускоренных методов на основе A-устойчивых неявных МРК. В [9, 17] для построения ускоренных методов используются симплектические интеграторы. В [3] предложены градиентные методы на основе методов РК – Чебышева, ориентированных на решение жестких задач Коши. Методы стохастического градиентного спуска на основе таких схем предложены в [4]. В [11] предложены градиентные методы на основе явной схемы Ньюмарка, широко используемой для интегрирования по времени уравнений механики деформируемого твердого тела.

В работе показано, что полученный метод эквивалентен методу Нестерова.

В последние десятилетия для улучшения свойств МРК (устойчивости, монотонности, сокращения числа стадий)предложен ряд подходов (например, см. [3, 18, 19, 20, 21, 22]). Один из наиболее интересных подходов связан с построением методов на основе разложений локальной погрешности по формулам, отличным от формулы Тейлора. В [23, 24] Е.В. Ворожцовым были построены явные МРК на основе разложения Лагранжа — Бюрмана (ЛБ), использующего степени нелинейной функции от шага интегрирования. В [23] были построены одно- и двухстадийные явные схемы и было показано, что области устойчивости таких методов могут иметь бóльшие площади по сравнению с стандартными методами. В [25, 26] показано, что с помощью таких разложений можно строить монотонные схемы высокого порядка аппроксимации для гиперболических систем уравнений в частных производных. Не зависимо от этих исследований, МРК на основе разложений ЛБ были предложены в [27].

С учетом описанной в [2] близости между условиями сходимости градиентных методов и условиями устойчивости МРК, использование МРК с возможностью увеличения площади области устойчивости для построения градиентных методов является весьма перспективным.

Настоящая работа посвящена построению и анализу сходимости градиентного метода на основе МРК, использующего разложения ЛБ. Предложен предобусловленный метод с инерцией, основанный на комбинации идей, предложенных в работах [2, 28]. Доказана теорема о сходимости и получены формулы для оптимального шага и скорости сходимости в случае квадратичной и возмущенной квадратичной функции. Метод применяется к численному решению краевых задач для дифференциальных и интегро-дифференциальных уравнений. Показано, что метод позволяет находить решение за меньшее число итераций и время по сравнению с известными градиентными методами.

2 Метод спуска на основе метода Рунге — Кутты второго порядка с разложением Лагранжа — Бюрмана

Рассмотрим функцию $f : \mathbb{R}^d \to \mathbb{R}$, являющуся сильно выпуклой с константой выпуклости $l > 0$, градиент которой удовлетворяет условию Липшица с константой $L > 0$. Рассмотрим задачу безусловной оптимизации:

$$f(x) \to \min_{x \in \mathbb{R}^d}, \quad x^* = \arg\min_{x \in \mathbb{R}^d} f(x). \tag{1}$$

Одним из наиболее известных методов решения такой задачи является метод градиентного спуска (gradient descent, GD):

$$x^{k+1} = x^k - h\nabla f(x^k), \quad k = 0, 1, \ldots, \tag{2}$$

где $h > 0$. Непрерывным аналогом этого метода является задача Коши для системы вида:

$$\dot{x} = -\nabla f(x). \tag{3}$$

Как известно [1, 2], положение равновесия этой системы совпадает с x^*. С учетом этого, методы решения начальной задачи для (3) можно трактовать как методы решения задачи (1), что и используется при построении новых градиентных методов.

2.1 Явные методы Рунге — Кутты на основе разложений Лагранжа — Бюрмана

Рассмотрим задачу Коши для автономной системы ОДУ:

$$\dot{x} = g(x), \quad x(t_0) = x_0, \tag{4}$$

где $g : \mathbb{R}^d \to \mathbb{R}$ будем считать достаточно гладкой.

Рассмотрим бесконечно дифференцируемую функию $y(t)$ и пусть функция $z = s(t)$, $z_0 = s(t_0)$, $\dot{s}(t_0) \neq 0$ является достаточно гладкой. Формула разложения ЛБ имеет следующий вид [23, 29]:

$$y(t) = y(t_0) + \sum_{k=1}^{\infty} \frac{(z - z_0)^k}{k!} \left(\frac{d^{k-1}}{dt^{k-1}} \left[\dot{y}(t) \left(\frac{t - t_0}{s(t) - z_0} \right)^k \right] \right)_{t=t_0}. \tag{5}$$

Как можно видеть, при $s(t) = t$, $z_0 = t_0$ формула (5) приводит к разложению в ряд Тейлора. Вводя функию φ, такую что $\varphi(t - t_0) = s(t) - z_0$ и требуя, чтобы $\varphi(0) = 0$, $\dot{\varphi}(0) \neq 0$, получим, что (5) может быть переписана в виде:

$$y(t) = y(t_0) + \sum_{k=1}^{\infty} b_k [\varphi(t - t_0)]^k, \tag{6}$$

где коэффициенты b_k вычисляются как [23]:

$$b_k = \lim_{t \to t_0} \frac{1}{k!} \frac{d^{k-1}}{dt^{k-1}} \left[\dot{y}(t) \left(\frac{t - t_0}{\varphi(t - t_0)} \right)^k \right], \quad k = 1, 2, \ldots$$

Выражения для b_i при $i = \overline{1,4}$ представлены в [23]. Как можно видеть, формула (6) представляет собой разложение $y(t)$ в ряд по степеням $\varphi(t - t_0)$.

Явный МРК, основанный на разложении ЛБ, имеет следующий вид:

$$x^{k+1} = x^k + \sum_{i=1}^{q} p_i K_i(h),$$

где $x^k \approx x(t_k)$, t_k — узлы сетки, построенной с шагом h,

$$K_1(h) = \varphi(h)g(x^k), \quad K_i(h) = \varphi(h)g\left(x^k + \sum_{j=1}^{i-1} \beta_{ij} K_j(h)\right), \quad i = \overline{2,q},$$

где p_i, β_{ij} являются параметрами метода. Эти параметры находятся из условий порядка, как в случае обычных МРК, но при этом используется разложение (6) локальной погрешности $x(t_{k+1}) - x^{k+1}$ в окрестности точки $h = 0$.

В [23] построен явный метод второго порядка следующего вида:

$$x^{k+1} = x^k + \frac{K_1 + 3K_2}{4\dot{\varphi}(0)}, \quad K_1 = \varphi(h)g(x^k), \quad K_2 = \varphi(h)g\left(x^k + \frac{2K_1}{3\dot{\varphi}(0)}\right). \quad (7)$$

Как показано в [23, 25, 26], посредством выбора функции $\varphi(h)$ можно построить монотонные схемы для численного решения (4), что важно при решении жестких задач Коши и задач с разрывными решениями. Метод (7) имеет следующий полином устойчивости:

$$R(z) = 1 + \gamma z + \frac{\gamma^2 z^2}{2}, \quad (8)$$

где $\gamma = \dfrac{\varphi(h)}{h\dot{\varphi}(0)}$. Отметим, что обычный МРК второго порядка соответствует значению $\gamma = 1$. Как можно видеть из (8), на область устойчивости такого метода можно влиять посредством варьирования значений $\gamma > 0$.

Для улучшения точности (7) в [23] предложено использовать модификацию этого метода. При нечетной φ получим, что $\varphi(h) = \dot{\varphi}(0)h + O(h^3)$. Тогда $\dot{\varphi}(0) = \dfrac{\varphi(h)}{h} + O(h^2)$. Используя приближение $\dot{\varphi}(0) \approx \dfrac{\varphi(h)}{h}$ только в формуле для расчета x^{k+1}, получим метод вида:

$$x^{k+1} = x^k + \frac{h(K_1 + 3K_2)}{4\varphi(h)}, \quad (9)$$

$$K_1 = \varphi(h)g(x^k), \quad K_2 = \varphi(h)g\left(x^k + \frac{2K_1}{3\dot{\varphi}(0)}\right).$$

Как отмечается в [23], этот метод тоже имеет второй порядок точности. При этом показано, что при решении задач с известным точным решением он дает меньшую погрешность, чем исходный метод. Полином устойчивости метода (9) имеет вид:

$$R(z) = 1 + z + \frac{\gamma z^2}{2}.$$

При применении метода (9) к задаче для линейной системы с $g(x) = Px + c$, где $\dim(P) = d \times d$, $\dim(c) = d$, получается следующий итерационный метод:

$$x^{k+1} = \left(E + hP + \frac{\gamma h^2}{2}P^2\right)x^k + h\left(E + \frac{h\gamma}{2}P\right)c,$$

где E есть единичная матрица. Таким образом, в случае линейной системы нет необходимости указывать конкретный вид функции $\varphi(h)$ и метод определяется значением параметра γ, при котором он должен быть устойчив.

2.2 Случай квадратичной функции

Рассмотрим функцию следующего вида:

$$f(x) = \frac{1}{2}(x, Ax) - (b, x),$$

где $b \in \mathbb{R}^d$, A — положительно определенная симметричная матрица с собственными знаечниями $0 < l = \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_d = L$. Градиент этой функции записывается как $\nabla f(x) = Ax - b$. Система (3) примет следующий вид:

$$\dot{x} = -(Ax - b), \tag{10}$$

и точка минимума $f(x)$ является решением линейной системы $Ax = b$.

Применяя (9) к задаче Коши для (10), получим метод следующего вида:

$$x^{k+1} = x^k - hD\nabla f(x^k), \tag{11}$$

где матрица $D = E - \frac{\gamma h}{2}A$ может рассматриваться как предобуславливатель для метода градиентного спуска.

Один из подходов к ускорению сходимости метода (2) заключается в добавлении инерционного слагаемого, что приводит к т.н. методу тяжелого шарика (heavy ball method, HB):

$$x^{k+1} = x^k - h\nabla f(x^k) + \beta(x^k - x^{k-1}), \tag{12}$$

где $\beta \in [0, 1)$.

По аналогии с (12), добавим в (11) инерционный член:

$$x^{k+1} = x^k - hD\nabla f(x^k) + \beta(x^k - x^{k-1}). \tag{13}$$

Для обозначения метода (13) в дальнешем будем использовать аббревиатуру LBHB (Lagrange — Burmann heavy ball).

Можно сформулировать следующую теорему о сходимости метода (13):

Теорема 1. *Пусть шаг $h > 0$, параметры $\gamma > 0$, $\beta \in [0, 1)$ и число обусловленности $\kappa = \dfrac{L}{l} \geq 14$ удовлетворяют следующим условиям:*

$$h < \frac{2}{\gamma L}; \tag{14}$$

$$\beta \geq \left(1 - \sqrt{\psi(h, \lambda; \gamma)}\right)^2, \quad \forall \lambda \in [l, L], \tag{15}$$

где $\psi(h, \lambda; \gamma) = h\lambda\left(1 - \frac{\gamma}{2}h\lambda\right)$;

$$\gamma > \frac{1}{4}\left(\frac{\sqrt{2\kappa}}{1+\kappa} + \frac{1}{\sqrt{2}}\right)^2. \tag{16}$$

Тогда: метод (13) *линейно сходится к x^*, при этом оптимальная скорость сходимости имеет вид:*

$$\rho_{opt} = 1 - \sqrt{\frac{2}{\gamma}}\frac{\sqrt{\kappa}}{1+\kappa}, \tag{17}$$

и достигается при следующих значениях h и β:

$$h_{opt} = \frac{2}{\gamma(l+L)}, \quad \beta_{opt} = \left(1 - \sqrt{\frac{2}{\gamma}}\frac{\sqrt{\kappa}}{1+\kappa}\right)^2. \tag{18}$$

Доказательство. Отметим, что выполнение (14) гарантирует, что $\psi > 0$ и допустимо рассматривать квадратный корень от этой величины.

1) Получим ограничение на γ, которое обеспечивает выполнение следующего неравенства:

$$-1 < 1 - \sqrt{\psi} < 1, \quad \forall \lambda. \tag{19}$$

Выполнение правой части этого неравенства очевидно. Левая часть приводит к неравенству $2 - \sqrt{\psi} > 0$, которое эквивалентно $\psi < 4$. Последнее неравенство можно переписать в виде:

$$\frac{\gamma}{2}t^2 - t + 4 > 0, \tag{20}$$

где $t = h\lambda$. Дискриминант соответствующего уравнения равен $1 - 8\gamma$. Таким образом, при выполнении условия

$$\gamma > \frac{1}{8}, \tag{21}$$

неравенство (20) будет верно для всех $\lambda \in [l, L]$ и h, определяемого по (14).

Отметим, что условие (16) согласуется с (21). Это очевидно из неравенства:

$$\frac{1}{4}\left(\frac{\sqrt{2\kappa}}{1+\kappa} + \frac{1}{\sqrt{2}}\right)^2 > \frac{1}{8}. \tag{22}$$

Таким образом, выбирая γ по (16), получим, что будет справедливо неравенство (19).

2) С использованием вектора $z^k = (x^k - x^*, x^{k-1} - x^*)^T$ метод (13) может быть быть представлен в виде:

$$z^{k+1} = Tz^k,$$

где матрица T имеет вид:

$$T = \begin{pmatrix} (1+\beta)E - hDA & -\beta E \\ E & 0_{d\times d} \end{pmatrix}.$$

Как известно [30], необходимое и достаточное условие сходимости такого метода имеет вид $r(T) < 1$, где $r(T)$ есть спектральный радиус матрицы T.

Представим A через спектральное разложение: $A = S\Lambda S^T$, где Λ есть диагональная матрица собственных значений A, а S — матрица собственных векторов, $SS^T = S^TS = E$. Используя S, построим матрицу $\overline{T} = \Sigma^T T \Sigma$, где

$$\Sigma = \begin{pmatrix} S & 0_{d\times d} \\ 0_{d\times d} & S \end{pmatrix}, \quad \overline{T} = \begin{pmatrix} (1+\beta)E - \Psi(h,\Lambda;\gamma) & -\beta E \\ E & 0_{d\times d} \end{pmatrix},$$

где $\Psi(h, \Lambda; \gamma) = h(E - \frac{\gamma h}{2}\Lambda)\Lambda$. Отметим, что матрицы $\overline{T}$ и T имеют одинаковые собственные значения.

Покажем, что собственные значения $\overline{T}$ совпадают с собственными значениями матрицы вида:

$$\tilde{T} = \begin{pmatrix} T_1 & 0_{2\times 2} & \dots & 0_{2\times 2} \\ 0_{2\times 2} & T_2 & \dots & 0_{2\times 2} \\ \dots & \dots & \dots & \dots \\ 0_{2\times 2} & 0_{2\times 2} & \dots & T_d \end{pmatrix},$$

где T_i есть матрицы размерности 2×2, имеющие вид

$$T_i = \begin{pmatrix} 1 + \beta - \psi(h, \lambda_i; \gamma) & -\beta \\ 1 & 0 \end{pmatrix}.$$

Матрица $\overline{T} - \zeta E$ имеет следующий вид:

$$\overline{T} - \zeta E = \begin{pmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{pmatrix},$$

где $T_{11} = (1 + \beta)E - \Psi(h, \Lambda; \gamma) - \zeta E$, $T_{12} = -\beta E$, $T_{21} = E$, $T_{22} = -\zeta E$. Ее определитель вычисляется по следующей формуле [31]:

$$\det(\overline{T} - \zeta E) = \det(T_{11}) \det(T_{22} - T_{21} T_{11}^{-1} T_{12}) =$$

$$= \det(T_{11}) \det \begin{pmatrix} -\zeta + \dfrac{\beta}{\eta_1} & 0 & \dots & 0 \\ 0 & -\zeta + \dfrac{\beta}{\eta_2} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & -\zeta + \dfrac{\beta}{\eta_d} \end{pmatrix} =$$

$$= (\beta - \zeta\eta_1)(\beta - \zeta\eta_2) \dots (\beta - \zeta\eta_d),$$

где $\eta_i = 1 + \beta - \psi(h, \lambda_i; \gamma) - \zeta$, $i = \overline{1, d}$.

В свою очередь, определитель блочно-диагональной матрицы $\tilde{T} - \zeta E$ вычисляется как:

$$\det(\tilde{T} - \zeta E) = \det(T_1 - \zeta E_{2\times 2}) \det(T_2 - \zeta E_{2\times 2}) \dots \det(T_d - \zeta E_{2\times 2}),$$

и совпадает с $\det(\overline{T} - \zeta E)$. Таким образом, матрицы $\overline{T}$ и $\tilde{T}$ имеют одинаковые собственные значения ζ_k, $k = \overline{1, 2d}$, которые вычисляются как собственные значения матриц T_i.

3) При выполнении условия (15) собственные значения T лежат внутри единичного круга. Собственные значения являются корнями уравнения:

$$\zeta^2 - (1 + \beta - \psi)\zeta + \beta = 0,$$

которые имеют вид

$$\zeta_{1,2} = \frac{1}{2}(1 + \beta - \psi \pm \sqrt{\mathcal{D}}),$$

где $\mathcal{D} = (1+\beta-\psi)^2 - 4\beta$. Нетрудно показать, что при выполнении условия (15) $\mathcal{D} \leq 0$. Модули корней имеют вид $|\zeta_{1,2}| = \sqrt{\beta}$, и в соответствии с тем, что $\beta < 1$, получаем, что (13) сходится к x^*.

Линейная сходимость метода (13) следует из теоремы 1 §1 Главы 2 из [32]. Скорость сходимости определяется спектральным радиусом $r(T) = \sqrt{\beta}$:

$$||x^k - x^*|| \leq (\sqrt{\beta} + \varepsilon_k)^k ||x^0 - x^*||,$$

где $\varepsilon_k \to 0$, $k \to \infty$.

4) Рассмотрим функцию $G(t) = \left(1 - \sqrt{t\left(1 - \frac{\gamma t}{2}\right)}\right)^2$, $t = \lambda h > 0$, где h считается фиксированным. С учетом условия (14) и положительности λ она определена на интервале $\left(0, \frac{2}{\gamma}\right)$. Знак ее первой производной

$$G'(t) = -\left(1 - \sqrt{t\left(1 - \frac{\gamma t}{2}\right)}\right) \frac{1 - \gamma t}{\sqrt{t\left(1 - \frac{\gamma t}{2}\right)}},$$

определяется знаками функций $1 - \gamma t$ и $\eta(t) = 1 - \sqrt{t\left(1 - \frac{\gamma t}{2}\right)}$.

Исследуем поведение функции $\eta(t)$: с учетом знака $\eta'(t)$ она убывает при $t < \frac{1}{\gamma}$ и возрастает при $t > \frac{1}{\gamma}$. Интервал, в котором $\eta(t) > 0$, определяется неравенством следующего вида:

$$\sqrt{t - \frac{\gamma t^2}{2}} < 1 \Leftrightarrow \gamma t^2 - 2t + 2 > 0.$$

Дискриминант соответствующего уравнения равен $4 - 8\gamma$, так что при $\gamma > \frac{1}{2}$ $\eta(t)$ строго положительна и при выполнении этого ограничения на γ производная $G'(t)$ меняет знак только в точке $t = \frac{1}{\gamma}$.

При $\gamma < \frac{1}{2}$ $\eta'(t)$ обращается в нуль в двух точках:

$$t_{1,2} = \frac{1 \pm \sqrt{1 - 2\gamma}}{\gamma}.$$

Несложно увидеть, что $t_{1,2} \in \left(0, \frac{2}{\gamma}\right)$.

Таким образом, при $\gamma \geq \frac{1}{2}$ $G(t)$ убывает при $t < \frac{1}{\gamma}$ и возрастает при $t > \frac{1}{\gamma}$, ее минимум достигается в точке $t = \frac{1}{\gamma}$ а максимальное значение достигается на границе интервала определения. При $\gamma \in \left(\frac{1}{8}, \frac{1}{2}\right)$ функция $G(t)$ убывает

при $t < t_1$ и $t \in \left(\frac{1}{\gamma}, t_2\right)$, а возрастает при $t \in \left(t_1, \frac{1}{\gamma}\right)$ и $t > t_2$. В этом случае максимальные значения могут достигаться при $t = \frac{1}{\gamma}$ и в граничных точках промежутка. Очевидно, что подобное поведение будет место и на промежутке $[lh, Lh] \subset \left(0, \frac{2}{\gamma}\right)$.

5) Предположим, что $G(t)$ имеет наибольшее значение в граничной точке $t = lh$ или $t = Lh$, при этом значение $\eta(t)$ строго положительно (как будет показано ниже, последнее будет справедливо при выполнении (16)).

С учетом этого предположения, оптимальное значение h можно получить как решение следующей задачи минимизации функции вида:

$$\beta_{opt} = \min_h \max(\chi_1(h), \chi_2(h)),$$

где $\chi_1(h) = \left(1 - \sqrt{\psi(h, l; \gamma)}\right)^2$, $\chi_2(h) = \left(1 - \sqrt{\psi(h, L; \gamma)}\right)^2$. С учетом отмеченных выше свойств, оптимальное значение h соответствует точке пересечения графиков этих функций:

$$\left(1 - \sqrt{hl\left(1 - \frac{\gamma(lh)^2}{2}\right)}\right)^2 = \left(1 - \sqrt{hL\left(1 - \frac{\gamma(Lh)^2}{2}\right)}\right)^2. \tag{23}$$

Решение этого уравнения имеет вид:

$$h_{opt} = \frac{2}{\gamma(l + L)}.$$

Оптимальное значение β имеет вид:

$$\beta_{opt} = \left(1 - \sqrt{\frac{2}{\gamma}}\frac{\sqrt{\kappa}}{1 + \kappa}\right)^2,$$

а оптимальная скорость сходимости определяется как $\rho_{opt} = \sqrt{\beta_{opt}}$.

6) Покажем, что при $\kappa \geq 14$ и выполнении (16) значение h, определяемое из (23), отвечает строго положительному значению $1 - \sqrt{\frac{2}{\gamma}}\frac{\sqrt{\kappa}}{1 + \kappa}$. Неравенство

$$1 - \sqrt{\frac{2}{\gamma}}\frac{\sqrt{\kappa}}{1 + \kappa} > 0$$

эквивалентно следующему ограничению на γ:

$$\gamma > \frac{2\kappa}{(1 + \kappa)^2}. \tag{24}$$

Покажем, что

$$\frac{2\kappa}{(1+\kappa)^2} < \frac{1}{8},$$

при $\kappa \geq 14$. Это неравенство эквивалентно $\kappa^2 - 14\kappa + 1 > 0$, и как несложно увидеть, выполняется при $\kappa > 7+\sqrt{45} \approx 13.71$. Таким образом, с учетом (22) условие (24) справедливо для значений γ, определяемых из (16) и при таком условии оптимальное значение h получается как решение уравнения (23).

7) Локальный максимум $G(t)$ в точке $t = \frac{1}{\gamma}$ равен $\left(1 - \frac{1}{\sqrt{2\gamma}}\right)^2$. Покажем, что при выполнении (16) справедливо следующее неравенство:

$$\max\left(\chi_1(h), \chi_2(h)\right) > \left(1 - \frac{1}{\sqrt{2\gamma}}\right)^2, \quad \forall h \in \left(0, \frac{2}{\gamma L}\right). \tag{25}$$

Как было показано ранее, (25) при выполнении (16) сводится к неравенству:

$$\left(1 - \sqrt{\frac{2}{\gamma}}\frac{\sqrt{\kappa}}{1+\kappa}\right)^2 > \left(1 - \frac{1}{\sqrt{2\gamma}}\right)^2,$$

которое может быть переписано в виде:

$$(1 - a\tau)^2 > (1 - b\tau)^2,$$

где $\tau = \frac{1}{\sqrt{\gamma}}$, $a = \frac{\sqrt{2\kappa}}{1+\kappa}$, $b = \frac{1}{\sqrt{2}}$. Последнее неравенство эквивалентно неравенству:

$$(a^2 - b^2)\tau > 2(a - b). \tag{26}$$

В соответствии с

$$2\sqrt{\kappa} < 1 + \kappa \Leftrightarrow (1 - \sqrt{\kappa})^2 > 0,$$

можно видеть, что при $\kappa > 1$ в (26) $a < b$. Таким образом, (26) принимает вид:

$$\frac{a+b}{2} < \frac{1}{\tau}.$$

Это неравенство следует из (16).

□

Замечание. При практическом применении метода (13) γ рассматривается как входной параметр, на значения которого наложено условие вида (16).

Рассмотрим функцию следующего вида:

$$c(\kappa) = \frac{1}{4}\left(\frac{\sqrt{2\kappa}}{1+\kappa} + \frac{1}{\sqrt{2}}\right)^2.$$

Как было показано ранее, значения этой функции больше $\frac{1}{8}$ (см. (22)). Производная этой функции имеет вид:

$$c'(\kappa) = \frac{1}{4}\frac{(1+\sqrt{k})^2(1-\kappa)}{\sqrt{\kappa}(1+\kappa)^3}.$$

Как можно видеть, она является отрицательной при $\kappa \geq 14$. Таким образом, $c(\kappa)$ убывает и ее наибольшее значение $c(14) \approx 0.281$. При этом $c(\kappa) \to \frac{1}{8}$ при $\kappa \to \infty$.

Как можно видеть из выражения для h_{opt} (см. (18)), большие значения шага h получаются при уменьшении значений γ. Из (17) в этом случае получим малые значения ρ_{opt}, что приводит к ускорению сходимости. Таким образом, можно ожидать, что при бóльшем шаге будем получать меньшее число итераций и время расчетов, необходимое для получения решения с заданной точностью. Таким образом, при проведении расчетов при фиксированном κ можно предложить следующий способ выбора γ:

$$\gamma = c(\kappa) + eps, \tag{27}$$

где $eps > 0$ является достаточно малым для того, чтобы избежать влияния погрешностей округления.

2.3 Случай возмущенной квадратичной функции

Рассмотрим функцию вида [3]:

$$f(x) = \frac{1}{2}(x, Ax) + g(x), \tag{28}$$

где A есть положительно определенная симметричная матрица, собственные значения которой принадлежат промежутку $[l, L]$, градиент $g(x)$ удовлетворяет условию Липшица:

$$\exists M > 0, \quad \forall x, y \in \mathbb{R}^d : ||\nabla g(x) - \nabla g(y)|| \leq M||x - y||,$$

а ее старшие производные являются малыми:

$$||\nabla^{(i)} g(x)|| \ll 1, \quad i = 2, 3, \ldots$$

В силу последнего условия, матрица A доминирует в матрице Гессе для $f(x)$: $\nabla^2 f(x) \approx A$. По аналогии с [3], построим метод, в рамках которого можно линеаризовать $g(x)$ в окрестности x^k: $g(x) \approx g(x^k) + (\nabla g(x^k), x - x^k)$. При использовании такого приближения градиент $f(x)$ для всех x из этой окрестности вычисляется как:

$$\nabla f(x) = Ax + \nabla g(x^k). \tag{29}$$

Применяя метод второго порядка (9) к задаче (3) в случае, когда $f(x)$ представима в виде (28), получим:

$$x^{k+1} = x^k - \frac{h}{4\varphi(h)}(K_1 + 3K_2),$$

где

$$K_1 = -\varphi(h)\nabla f(x^k) = -\varphi(h)\left(Ax^k + \nabla g(x^k)\right),$$

$$K_2 = -\varphi(h)\nabla f\left(x^k + \frac{2K_1}{3\dot{\varphi}(0)}\right) = -\varphi(h)A\left(x^k + \frac{2K_1}{3\dot{\varphi}(0)}\right) - \varphi(h)\nabla g(x^k).$$

Таким образом, получим метод:

$$x^{k+1} = \left(E - hA + \frac{h\varphi(h)}{2\dot{\varphi}(0)}A^2\right)x^k + h\left(\frac{\varphi(h)}{2\dot{\varphi}(0)}A - E\right)\nabla g(x^k),$$

который с использованием выражения для γ примет вид:

$$x^{k+1} = Bx^k - hD\nabla g(x^k), \tag{30}$$

где матрицы B и D определяются аналогично случаю квадратичной $f(x)$. Сформулируем следующую теорему о сходимости метода (30):

Теорема 2. *Пусть $M > 0$ удовлетворяет неравенству:*

$$\frac{8M}{L} < 1,$$

где $h > 0$, $\gamma > 0$ такие, что:

$$h < \frac{2}{\gamma L}, \quad \gamma > \frac{1}{4}, \quad \rho \in \left(0, 1 - \frac{8M}{L}\right), \tag{31}$$

где $\rho = ||B|| = \max\limits_i \left|1 - h\lambda_i + \frac{\gamma h^2}{2}\lambda_i^2\right|$. Тогда: метод (30) сходится при любом x^0 и справедливо следующее неравенство:

$$||x^k - x^*|| \leq \left(\rho + \frac{8M}{L}\right)^k ||x^0 - x^*||. \tag{32}$$

Доказательство. Учитывая, что $\nabla f(x^*) = 0 \Leftrightarrow Ax^* + \nabla g(x^*) = 0$, метод (30) можно переписать в виде:

$$x^{k+1} - x^* = Bx^k - hD\nabla g(x^k) - x^* \pm Bx^* + hD(Ax^* + \nabla g(x^*)), \qquad (33)$$

Используя выражения

$$B - E = -hA + \frac{\gamma h^2}{2}A^2, \quad hDA = hA - \frac{\gamma h^2}{2}A^2,$$

из (33) получим:

$$||x^{k+1} - x^*|| \leq ||B(x^k - x^*)|| + h||D(\nabla g(x^k) - \nabla g(x^*))||.$$

Покажем, что при выполнении (31) матрица B такая, что $||B|| < 1$. В силу симметричности этой матрицы, ее собственные значения имеют вид $1 - h\lambda_i + \frac{\gamma h^2}{2}\lambda_i^2$, в связи с чем это условие можно переписать в виде:

$$-1 < 1 - \lambda_i h + \frac{\gamma}{2}\lambda_i^2 h^2 < 1, \quad \forall i. \qquad (34)$$

Его правая часть имеет вид: $-h\lambda_i\left(1 - \frac{\gamma h}{2}\lambda_i\right) < 0$, и в связи с положительностью λ_i выполняется при всех i в силу $h < \dfrac{2}{\gamma L}$. Левая часть (34) принимает вид:

$$2 - \lambda_i h + \frac{\gamma}{2}\lambda_i^2 h^2 > 0.$$

Соответствующий дискриминант равен $1 - 4\gamma$, в связи с чем оно выполняется для всех i в силу (31). Кроме того, в силу $1 - \dfrac{\gamma h}{2}\lambda_i > 0$ получим, что $||D|| < 1$. Учитывая отмеченные свойства матриц B и D, получим:

$$||x^{k+1} - x^*|| \leq (\rho + hM)||x^k - x^*|| < \left(\rho + \frac{2}{\gamma L}M\right)||x^k - x^*|| < \left(\rho + \frac{8M}{L}\right)||x^k - x^*||.$$

С учетом (31) $\rho + \frac{8M}{L} < 1$ и неравенство (32) получается итерированием последнего неравенства.

□

Замечание. С учетом приближения $\nabla^2 f(x) \approx A$, метод (13) при практическом применении к минимизации (28) можно применять при параметрах, вычисляемых по L и l (по аналогии с методом тяжелого шарика [32]), но в этом случае можно говорить только о локальной сходимости метода.

3 Вычислительные эксперименты

Рассмотрим применение разработанного метода к численному решению задач оптимизации, к которым сводятся краевые задачи из различных приложений. Численные расчеты проводятся при использовании оптимальных параметров β_{opt} и h_{opt}. Производится сравнение метода (13) со следующими известными методами, которые тоже будут применяться при значениях параметров, обеспечивающих наилучшую скорость сходимости:

1) Метод градиентного спуска (2) с оптимальным шагом для сильно выпуклой функции с константой выпуклости $l > 0$ и липшицевым градиентом с константой $L > 0$ [32]:

$$h_{opt} = \frac{2}{L + l}, \quad \rho_{opt} = \frac{\kappa - 1}{\kappa + 1}.$$

2) Метод тяжелого шарика (12) с оптимальными параметрами, полученными для сильно выпуклой квадратичной функции [32]:

$$h_{opt} = \frac{4}{(\sqrt{L} + \sqrt{l})^2}, \quad \beta_{opt} = \left(\frac{\sqrt{\kappa} - 1}{\sqrt{\kappa} + 1}\right)^2, \quad \rho_{opt} = \frac{\sqrt{\kappa} - 1}{\sqrt{\kappa} + 1}.$$

3) Метод Нестерова с оптимальными параметрами, полученными для класса выпуклых функций с липшицевыми градиентами (метод Nesterov1) [33]:

$$x^{k+1} = y^k - h\nabla f(y^k), \quad y^k = x^k + \beta(x^k - x^{k-1}),$$

$$h_{opt} = \frac{1}{L}, \quad \beta_{opt} = \frac{\sqrt{\kappa} - 1}{\sqrt{\kappa} + 1}, \quad \rho_{opt} = 1 - \frac{1}{\sqrt{\kappa}}.$$

4) Метод Нестерова с оптимальными параметрами, полученными для квадратичной сильно выпуклой функции (метод Nesterov2) [34]:

$$h_{opt} = \frac{4}{3L + l}, \quad \beta_{opt} = \frac{\sqrt{3\kappa + 1} - 2}{\sqrt{3\kappa + 1} + 2}, \quad \rho_{opt} = 1 - \frac{2}{\sqrt{3\kappa + 1}}.$$

На рис. 1 представлены графики зависимости ρ_{opt} от логарифма κ при $\kappa \geq 14$ в случае $\gamma = 0.29$ (при таком значении (16) выполняется для всех рассматриваемых κ). Как можно видеть, значения ρ_{opt} в случае метода (13) меньше, чем для остальных методов, что должно приводить к его более быстрой сходимости на практике. На рис. 2 для сравнения представлены графики оптимальной скорости сходимости при фиксированном γ и для значений

γ, получаемых при каждом κ из условия (27) при eps $= 0.001$. Как можно видеть, при выборе γ в зависимости от κ получается наилучшая скорость сходимости среди всех рассмотренных методов.

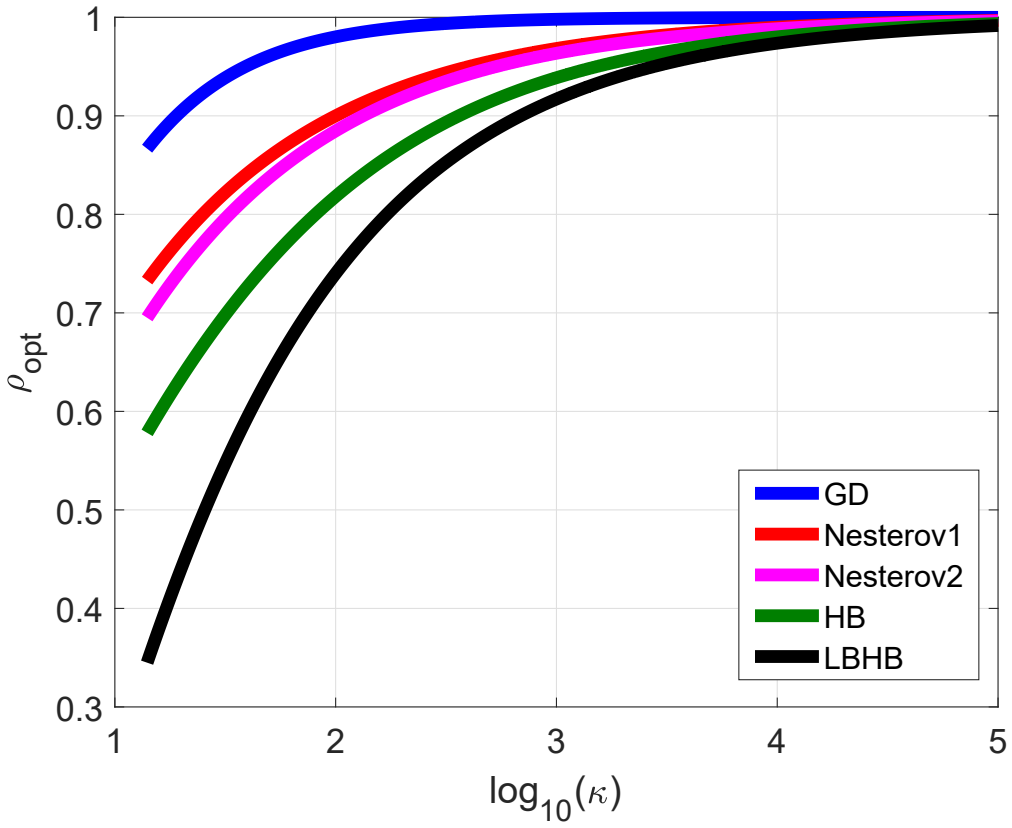

Рис. 1: Графики зависимости оптимальной скорости сходимости от логарифма κ для различных методов

При решении рассматриваемых ниже задач значение γ выбиралось из условия (27) при $eps = 0.001$. Вычисления проводились на ПК следующей конфигурации: Intel(R) Core(TM) i7-12700k F 3.60 GHz 32 Gb RAM, программная реализация алгоритмов проводилась в пакете Matlab R2021a. В качестве тестовых задач были выбраны краевые задачи, возникающие в математической физике и вариационном исчислении. Методы спуска являются одними из наиболее широко используемых и эффективных итерационных методов решения систем алгебраических уравнений, возникающих при дискретизации таких задач [35, 36].

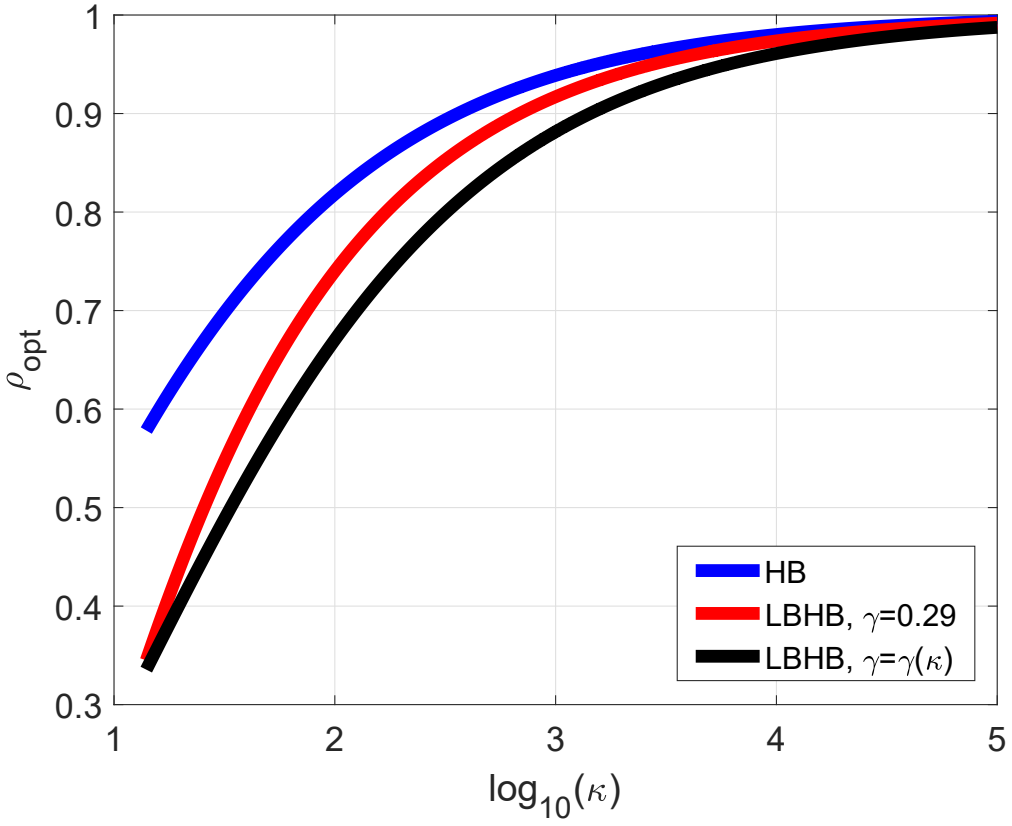

Рис. 2: Графики зависимости оптимальной скорости сходимости от логарифма κ

3.1 Задача Дирихле для уравнения Пуассона

Рассмотрим уравнение Пуассона в декартовых координатах в области, представляющей собой единичный куб:

$$\frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2} + \frac{\partial^2 v}{\partial z^2} = -\sin(\pi y)\sin(\pi z), \quad x, y, z \in (0, 1). \tag{35}$$

Будем считать, что на границе области Γ поставлены нулевые условия первого рода: $U|_\Gamma = 0$. Точное решение этой задачи имеет вид:

$$v(x, y, z) = \frac{\sin(\pi y)\sin(\pi z)}{2\pi^2}\left(1 - \frac{\sinh(\sqrt{2}\pi x) + \sinh(\sqrt{2}\pi(1-x))}{\sinh(\sqrt{2}\pi)}\right).$$

Построим в области равномерную сетку из N^3 внутренних узлов с шагом Δh по всем переменным. Заменим вторые производные в (35) с помощью симметричных конечных разностей:

$$\frac{\partial^2 v}{\partial x^2}(x_i, y_j, z_k) \approx \frac{v_{i+1jk} - 2v_{ijk} + v_{i-1jk}}{\Delta h^2}, \quad \frac{\partial^2 v}{\partial y^2}(x_i, y_j, z_k) \approx \frac{v_{ij+1k} - 2v_{ijk} + v_{ij-1k}}{\Delta h^2},$$

$$\frac{\partial^2 v}{\partial z^2}(x_i, y_j, z_k) \approx \frac{v_{ijk+1} - 2v_{ijk} + v_{ijk-1}}{\Delta h^2},$$

где $v_{ijk} \approx v(x_i, y_j, z_k)$. Подставляя эти приближения в (35), учитывая граничные условия и производя специальную нумерацию злов (см. [37]), получим систему линейных алгебраических уравнений:

$$Au = F,$$

где F ($\dim(F) = N^3$) — вектор, получаемый по правой части уравнения (35), u — вектор приближенных значений v в узлах сетки, A — симметричная положительно определенная блочно-диагональная матрица. Ее максимальное и минимальное собственные значения вычисляются как [38]:

$$l = \frac{12}{\Delta h^2} \sin^2\left(\frac{\pi \Delta h}{2}\right), \quad L = \frac{12}{\Delta h^2} \cos^2\left(\frac{\pi \Delta h}{2}\right).$$

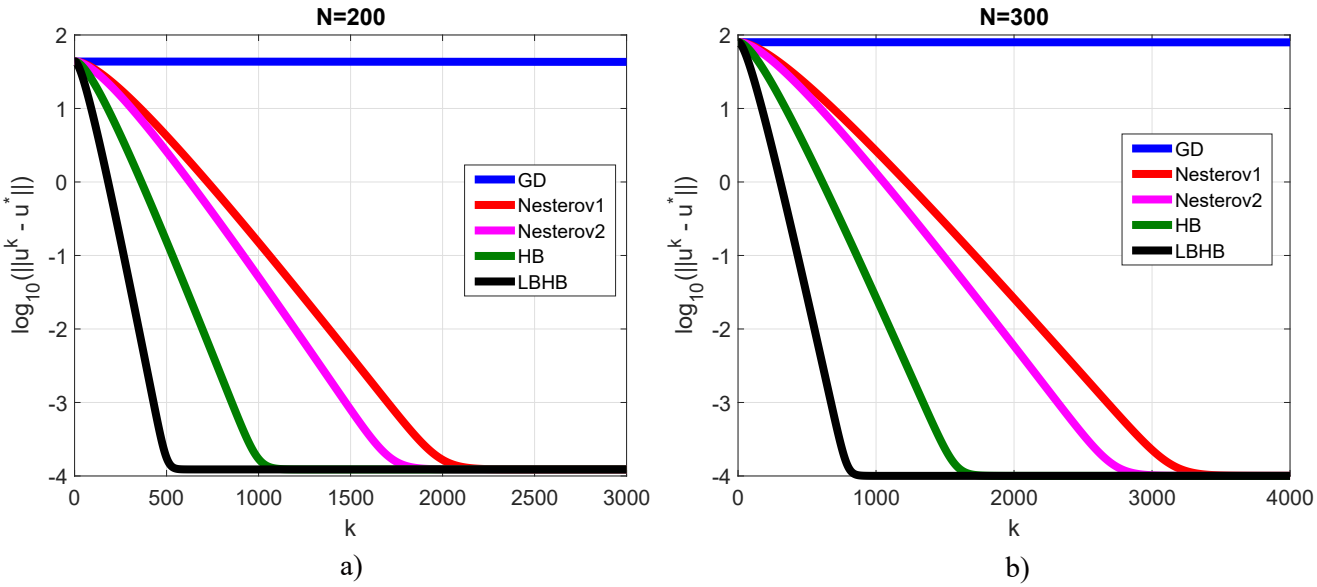

Рис. 3: Графики зависимости логарифма погрешности от номера итерации при численном решении задачи для (35): a) случай сетки с $N = 200$; b) случай сетки с $N = 300$

При проведении численных расчетов проводилась минимизация квадратичной функции с матрицей A. Для сравнения методов использовалась абсолютная погрешность $||u^k - u^*||_2$, где u^* есть точное решение задачи. На рис. 3 представлены графики зависимости логарифма нормы погрешности от номера итерации в случаях сеток с $N = 200$ ($\kappa \approx 1.637 \cdot 10^4$) и $N = 300$ ($\kappa \approx 3.671 \cdot 10^4$). Как можно видеть, метод (13) сходится быстрее всех рассматриваемых методов. В табл. 1 представлено число итераций и время работы в

секундах, нужных для достижения точности $5 \cdot 10^{-4}$ ускоренными методами спуска. Отметим, что наименьшее число итераций и время тоже имеет место в случае метода (13).

Таблица 1. Число шагов и время в секундах (в скобках), требуемые для достижения точности $5 \cdot 10^{-4}$ при применении ускоренных методов к численному решению задачи для (35)

N	κ	HB	Nesterov1	Nesterov2	LBHB
200	$1.637 \cdot 10^4$	904 (89.3)	1800 (142)	1558 (123)	454 (71.2)
250	$2.553 \cdot 10^4$	1157 (175)	2305 (363)	1996 (314)	581 (151)
300	$3.671 \cdot 10^4$	1414 (370)	2821 (772)	2442 (665)	711 (325)
350	$4.993 \cdot 10^4$	1676 (697)	3344 (1447)	2895 (1250)	842 (635)
400	$6.517 \cdot 10^4$	1942 (1221)	3857 (2540)	3355 (2179)	975 (1119)

3.2 Задача для линейного интегро-дифференциального уравнения

Рассмотрим линейное интегро-дифференциальное уравнение следующего вида:

$$z''(x) - z'(x) - 6z(x) + \varepsilon \int_0^1 z(t)dt = -2\pi \cos(2\pi x) - (6 + 4\pi^2) \sin(2\pi x), \quad (36)$$

с граничными условиями $z(0) = z(1) = 0$, при $0 < \varepsilon \ll 1$. Точное решение этой задачи имеет вид:

$$z^*(x) = C_1 \left(e^{3x} + \frac{\varepsilon(e^3 - 1)}{3(6 - \varepsilon)} \right) + C_2 \left(e^{-2x} + \frac{\varepsilon(1 - e^{-2})}{2(6 - \varepsilon)} \right) + \sin(2\pi x),$$

где

$$C_1 = \frac{3((1 + e^{-2})\varepsilon - 12)}{\sigma(\varepsilon)}, \quad C_2 = \frac{2(18 + \varepsilon(e^3 - 4))}{\sigma(\varepsilon)},$$

$$\sigma(\varepsilon) = \varepsilon e^3 + 5\varepsilon - 11\varepsilon e^{-2} + 5\varepsilon e + 36e^{-2} - 36e^3.$$

Для получения численного решения задачи для (36) построим на промежутке $[0,1]$ сетку с шагом Δh и узлами $0 = x_0 < x_1 < \ldots < x_N < x_{N+1} = 1$. Аппроксимируем производные, входящие в (36), с помощью разностных производных второго порядка аппроксимации:

$$z''(x_i) \approx \frac{z_{i+1} - 2z_i + z_{i-1}}{\Delta h^2}, \quad z'(x_i) \approx \frac{z_{i+1} - z_{i-1}}{2\Delta h}, \quad i = \overline{2, N},$$

$$z''(x_1) \approx \frac{z_2 - 2z_1}{\Delta h^2}, \quad z''(x_N) \approx \frac{-2z_N + z_{N-1}}{\Delta h^2}, \quad z'(x_1) \approx \frac{z_2}{2\Delta h}, \quad z'(x_N) \approx \frac{-z_{N-1}}{2\Delta h},$$

где $z_i \approx z(x_i)$, а при аппроксимации производных в приграничных узлах использованы граничные условия.

Для приближения интеграла воспользуемся составной квадратурной формулой трапеций:

$$\int_0^1 z(t)dt \approx \frac{\Delta h}{2} z(x_0) + \Delta h \sum_{i=1}^{N} z(x_i) + \frac{\Delta h}{2} z(x_{N+1}) \approx \Delta h \sum_{i=1}^{N} z_i.$$

Подставляя все приближения в (36), получим линейную систему относительно z_i:

$$(A + B + \varepsilon \Delta h^3 I)z = \Delta h^2 b, \tag{37}$$

где

$$A = \begin{pmatrix} 2 & -1 & 0 & 0 & \ldots & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & \ldots & 0 & 0 & 0 \\ 0 & -1 & 2 & -1 & \ldots & 0 & 0 & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & 0 & \ldots & -1 & 2 & -1 \\ 0 & 0 & 0 & 0 & \ldots & 0 & -1 & 2 \end{pmatrix},$$

$$B = \begin{pmatrix} -6\Delta h^2 & -\frac{\Delta h}{2} & 0 & 0 & \ldots & 0 & 0 & 0 \\ \frac{\Delta h}{2} & -6\Delta h^2 & -\frac{\Delta h}{2} & 0 & \ldots & 0 & 0 & 0 \\ 0 & \frac{\Delta h}{2} & -6\Delta h^2 & -\frac{\Delta h}{2} & \ldots & 0 & 0 & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & 0 & \ldots & \frac{\Delta h}{2} & -6\Delta h^2 & -\frac{\Delta h}{2} \\ 0 & 0 & 0 & 0 & \ldots & 0 & \frac{\Delta h}{2} & -6\Delta h^2 \end{pmatrix},$$

где I — матрица из единиц, а вектор b вычисляется через значения правой части уравнения (37).

При малых значениях ε и Δh можно считать, что $A \approx A + B + \varepsilon \Delta h^3 I$, в связи с чем в качестве L и l будем рассматривать максимальное и минимальное собственные значения матрицы A, которые вычисляются следующим образом [38]:

$$l = 4\sin^2\left(\frac{\pi\Delta h}{2}\right), \quad L = 4\cos^2\left(\frac{\pi\Delta h}{2}\right).$$

Численные расчеты производились при значении $\varepsilon = 0.01$. Погрешность вычислялась как $||z^k - z^*||_2$. В качестве начального приближения была выбрана функция $z^0(x) = x(1-x)$.

На рис. 4 представлены графики зависимости логарифма погрешности от номера итерации для сеток с $N = 500$ ($\kappa \approx 1.017 \cdot 10^5$) и $N = 1000$ ($\kappa \approx 4.061 \cdot 10^5$). Как можно заметить, при применении метода (13) нужная точность достигается быстрее, чем в случаях остальных методов. В табл. 2 представлено число итераций и время работы, необходимые для получения точности 10^{-6} ускоренными методами в случае разных разбиений сетки. Наилучшие результаты имеют место для метода (13).

Таблица 2. Число шагов и время в секундах (в скобках), требуемые для достижения точности 10^{-6} при применении ускоренных методов к численному решению задачи для (36)

N	κ	HB	Nesterov1	Nesterov2	LBHB
1000	$4.061 \cdot 10^5$	5024 (0.173)	10043 (0.296)	8697 (0.234)	2522 (0.0672)
1500	$9.131 \cdot 10^5$	7548 (0.703)	15092 (0.904)	13070 (0.845)	3789 (0.253)
2000	$1.622 \cdot 10^6$	10077 (2.17)	20149 (3.69)	17449 (3.36)	5058 (0.959)
2500	$2.535 \cdot 10^6$	12609 (7.97)	25214 (16.2)	21836 (13.9)	6330 (4.04)
5000	$1.013 \cdot 10^7$	27524 (123)	55038 (247)	47664 (213)	13817 (62.1)
10000	$4.053 \cdot 10^7$	55566 (1111)	111128 (2229)	96240 (1928)	27896 (557)

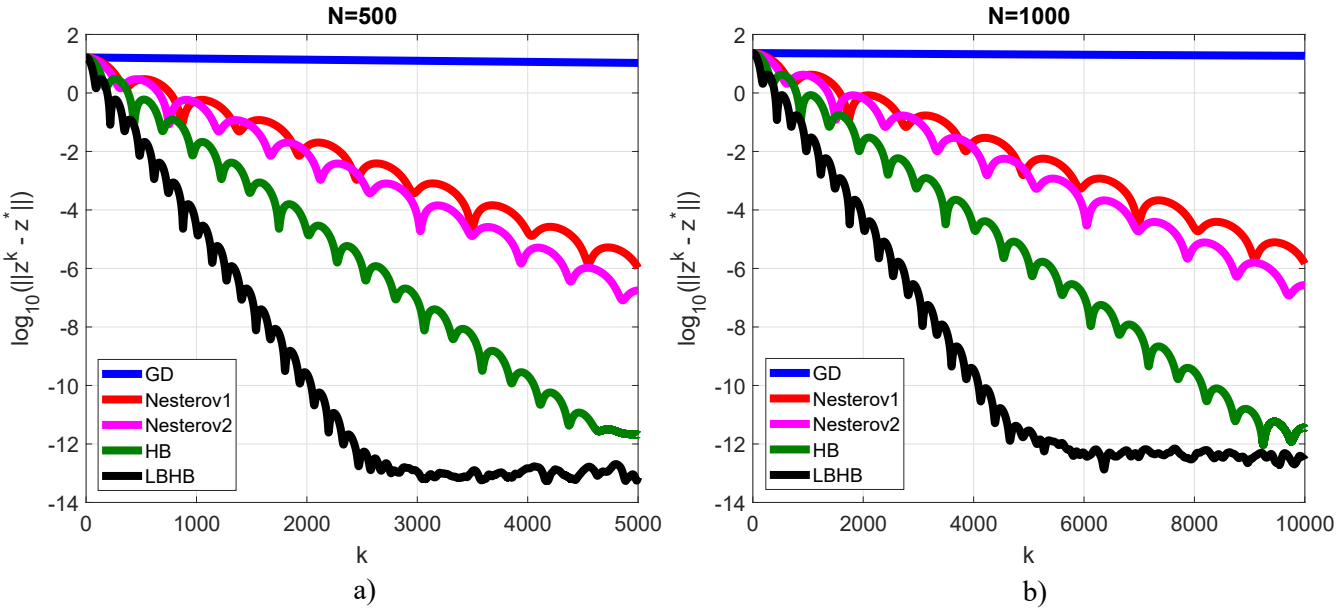

Рис. 4: Графики зависимости логарифма погрешности от номера итерации при численном решении задачи для (36): a) случай сетки с $N = 500$; b) случай сетки с $N = 1000$

3.3 Задача минимизации интегрального функционала

В качестве примера задачи о минимизации функции вида (28) рассмотрим задачу, возникающую при дискретизации следующей задачи вариационного исчисления:

$$I(y) = \int_0^1 ((y'(x))^2 - \varepsilon(y'(x))^4)dx, \quad \varepsilon \ll 1. \tag{38}$$

где $y(x) \in C^1[0,1]$, $y(0) = y(1) = 0$.

С учетом того, что подынтегральная функция зависит только от y', уравнение Эйлера для этого функционала имеет общее решение $y(x) = C_1x + C_2$ [39]. С учетом поставленных граничных условий, экстремаль задачи имеет вид $y^*(x) \equiv 0$.

Для перехода к конечномерной задаче построим на промежутке $[0,1]$ сетку с шагом Δh из N внутренних узлов. Интеграл из (38) вычислим с использованием составной квадратурной формулы трапеций:

$$I(y) \approx \frac{\Delta h}{2}((y'(x_0))^2 - \varepsilon(y'(x_0))^4) + \Delta h \sum_{i=1}^{N}((y'(x_i))^2 - \varepsilon(y'(x_i))^4) +$$

$$\frac{\Delta h}{2}((y'(x_{N+1}))^2 - \varepsilon(y'(x_{N+1}))^4). \tag{39}$$

Аппроксимируем входящие в эту сумму производные с помощью конечных разностей первого порядка (для аппроксимации $y'(x_0)$ и $y'(x_{N+1})$ используем граничные условия):

$$y'(x_i) \approx \frac{y_{i+1} - y_i}{\Delta h}, \quad i = \overline{1, N}, \quad y'(x_0) \approx \frac{y_1}{\Delta h}, \quad y'(x_{N+1}) \approx -\frac{y_N}{\Delta h}, \tag{40}$$

где $y_i \approx y(x_i)$.

Подставляя (40) в (39), получим приближение функционала функцией конечного числа переменных: $I(y) \approx f(y_1, \ldots, y_N)$. Где f представляется как (28), где g отвечает конечномерному приближению члена $-\varepsilon \int\limits_0^1 (y'(x))^4 dx$. Матрица A имеет вид:

$$A = \frac{1}{\Delta h} \begin{pmatrix} 3 & -2 & 0 & 0 & \ldots & 0 & 0 & 0 \\ -2 & 4 & -2 & 0 & \ldots & 0 & 0 & 0 \\ 0 & -2 & 4 & -2 & \ldots & 0 & 0 & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & 0 & \ldots & -2 & 4 & -2 \\ 0 & 0 & 0 & 0 & \ldots & 0 & -2 & 5 \end{pmatrix}.$$

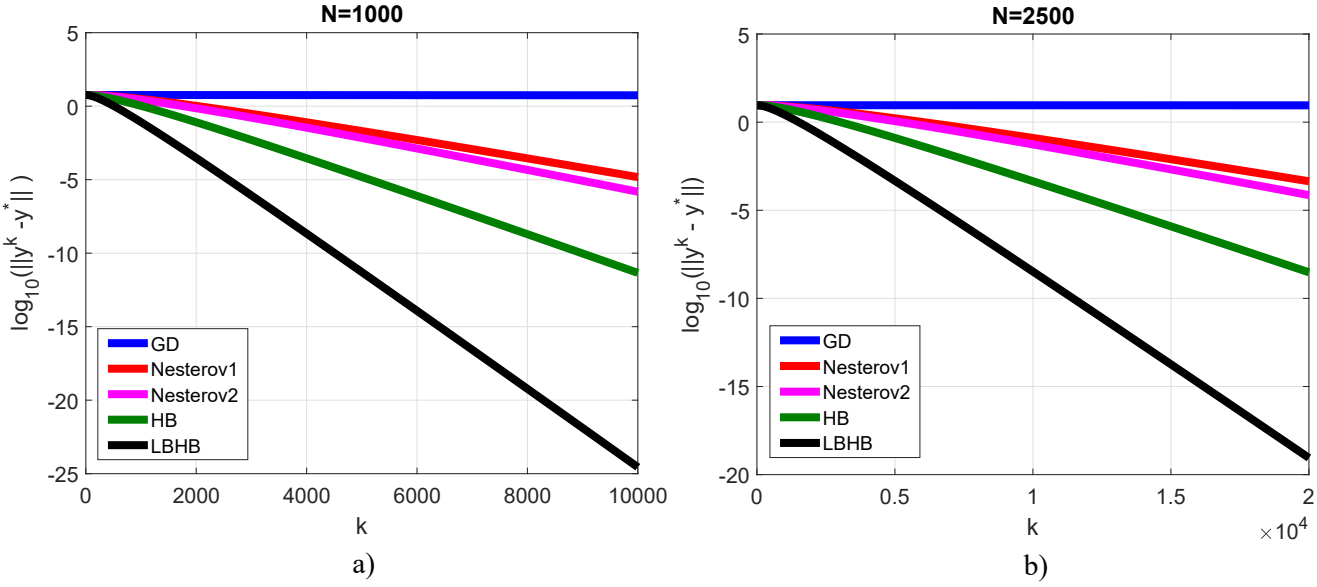

Рис. 5: Графики зависимости логарифма погрешности от номера итерации при численном решении задачи о минимизации (38): a) случай сетки с $N = 1000$; b)) случай сетки с $N = 2500$

Эффект малого возмущения реализуется за чем малого параметра ε и шага сетки. При проведении численных расчетов на различных сетках ис-

пользовалось значение $\varepsilon = 0.01$. При таком выборе собственные значения матрицы A доминируют в спектре $\nabla^2 f(y)$. Начальное приближение вычислялось как $y^0(x) = x(1-x)$, l и L находились численно с использованием Matlab. Погрешность вычислялась как $||y^k - y^*||_2$.

На рис. 5 представлены графики зависимости логарифма погрешности от номера итерации для сеток с $N = 1000$ ($\kappa \approx 4.066 \cdot 10^5$) и $N = 2500$ ($\kappa \approx 2.536 \cdot 10^6$). Как можно заметить, метод (13) сходится быстрее, чем остальные методы. В табл. 3 представлено число итераций и время работы, необходимые для получения точности 10^{-6} ускоренными методами в случае разных разбиений сетки. Как можно видеть, наилучшие результаты имеют место для метода (13).

Таблица 3. Число шагов и время в секундах (в скобках), требуемые для достижения точности 10^{-6} при применении ускоренных методов к численному решению задачи о минимизации (38)

N	κ	HB	Nesterov1	Nesterov2	LBHB
500	$1.017 \cdot 10^5$	2262 (14.1)	4523 (27.9)	3917 (25.3)	1137 (7.21)
1000	$4.061 \cdot 10^5$	4546 (74.4)	9090 (149)	7872 (131)	2238 (39.1)
1500	$9.131 \cdot 10^5$	6836 (209)	13670 (420)	11838 (362)	3433 (108)
2000	$1.622 \cdot 10^6$	9129 (463)	18257 (922)	15811 (803)	4584 (206)
2500	$2.535 \cdot 10^6$	11425 (858)	22849 (1706)	19788 (1478)	5737 (445)

3.4 Задача для нелинейного интегро-дифференциального уравнения

Рассмотрим нелинейное интегро-дифференциальное уравнение [3]:

$$u''(x) = \int_0^1 \frac{u^4(s)ds}{(1+|x-s|)^2}, \quad x \in (0,1), \tag{41}$$

с граничными условиями $u(0) = 1$, $u(1) = 0$. Такая модель описывает стационарное распределение температуры с учетом нелокальных эффектов [40].

При численном решении задачи построим на $[0,1]$ равномерную сетку, как в предыдущих примерах. Конечномерная задача о нахождении $u_i \approx u(x_i)$ получается следующим образом: вторая производная в (41) аппроксимируется с помощью второй разностной производной, а для вычисления интеграла при каждом x_i используем составную формулу трапеций с учетом граничных условий:

$$\int_0^1 \frac{u^4(s)ds}{(1+|x-s|)^2} \approx \frac{\Delta h}{2(1+i\Delta h)^2} + \sum_{j=1}^{N} \frac{u_j^4 \Delta h}{(1+\Delta h|i-j|)^2}. \tag{42}$$

После дискретизации (41) получим систему вида:

$$Au = -\nabla g(u), \tag{43}$$

где ∇g соответствует правой части (42), а A есть симметричная положительная определенная матрица, отвечающая дискретизации второй производной. Ее минимальное и максимальное собственные значения вычисляются как [38]:

$$l = \frac{4}{\Delta h^2}\sin^2\left(\frac{\pi\Delta h}{2}\right), \quad L = \frac{4}{\Delta h^2}\cos^2\left(\frac{\pi\Delta h}{2}\right).$$

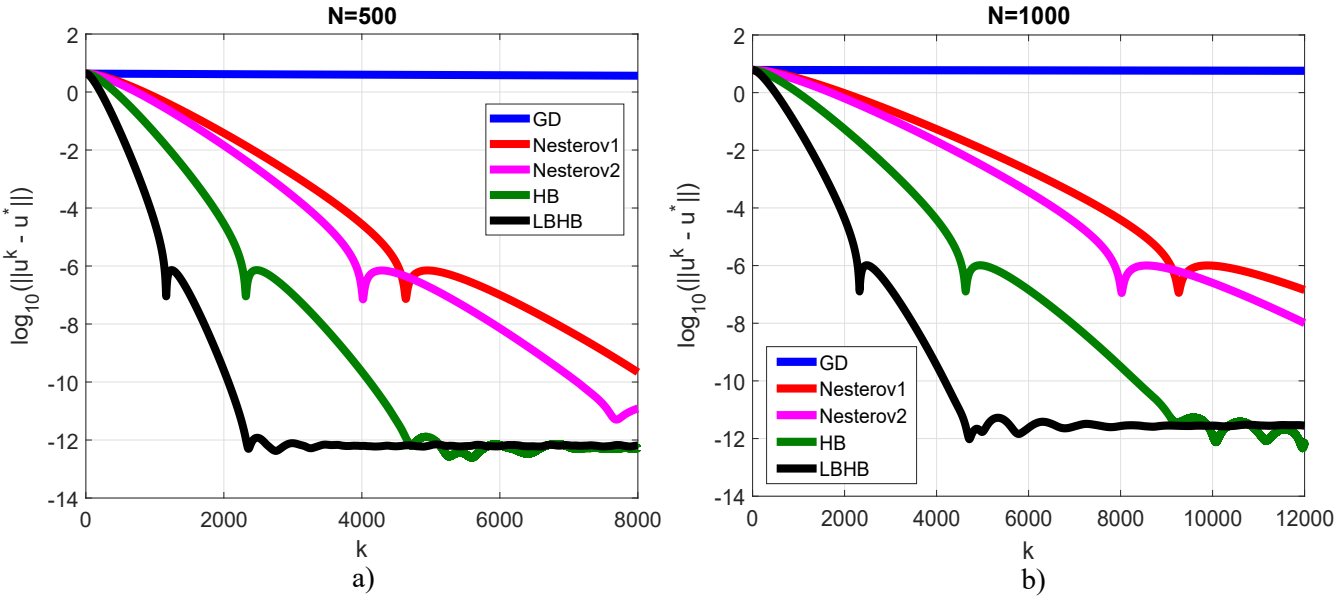

Рис. 6: Графики зависимости логарифма погрешности от номера итерации при численном решении краевой задачи для (41): a) случай сетки с $N = 500$; b)) случай сетки с $N = 1000$

Систему (43) можно трактовать как необходимое условие минимума возмущенной квадратичной функции $f(u)$:

$$Au + \nabla g(u) = 0 \Leftrightarrow \nabla f(u) = 0.$$

Таким образом, градиент f известен, и для численного решения (43) можно применять градиентные методы. Погрешность вычислялась следующим образом: $||u^k - u^*||_2$, где решение u^* для каждого разбиения сетки находилось посредством применения метода (13) при $5 \cdot 10^4$ итерациях. Начальное приближение вычислялось следующим образом: $u^0(x) = 1 - x^2$.

На рис. 6 представлены графики зависимости логарифма погрешности от номера итерации в случае сеток с $N = 500$ ($\kappa \approx 1.017 \cdot 10^5$) и $N = 1000$ ($\kappa \approx 4.061 \cdot 10^5$). Как можно видеть, для этой задачи метод (13) тоже демонстрирует лучшую скорость сходимости по сравнению с другими методам. В табл. 4 представлено число итераций и время, необходимое для достижения точности 10^{-6} при применении ускоренных методов. Наименьшее число шагов и время характерны для метода (13).

Таблица 4. Число шагов и время в секундах (в скобках), требуемые для достижения точности 10^{-6} при применении ускоренных методов к численному решению задачи для (41)

N	κ	HB	Nesterov1	Nesterov2	LBHB
200	$1.637 \cdot 10^4$	904 (89.3)	1800 (142)	1558 (123)	454 (71.2)
250	$2.553 \cdot 10^4$	1157 (175)	2305 (363)	1996 (314)	581 (151)
300	$3.671 \cdot 10^4$	1414 (370)	2821 (772)	2442 (665)	711 (325)
350	$4.993 \cdot 10^4$	1676 (697)	3344 (1447)	2895 (1250)	842 (635)
400	$6.517 \cdot 10^4$	1942 (1221)	3857 (2540)	3355 (2179)	975 (1119)

4 Заключение

В настоящей работе предложен метод градиентного спуска с инерцией для решения задачи минимизации выпуклой функции, основанный на применении модифицированного МРК второго порядка, при построении которого используется разложение ЛБ. Для сильно выпуклой квадратичной и возмущенной квадратичной функций доказаны теоремы о сходимости и получены аналитические выражения для оптимальных параметров, при которых скорость

сходимости является наилучшей. В квадратичном случае для широкого диапазона значений числа обусловленности κ показано, что скорость сходимости метода лучше, чем для ряда других известных методов. Рассмотрено применение метода к численному решению задачи для трехмерного уравнения Пуассона, задач для интегро-дифференциальных уравнений, задачи вариационного исчисления. Показано, что предложенный метод сходится за меньшее число итераций и время по сравнению с другими известными градиентными методами.

Список литературы

[1] Ascher U. M., van den Doel K., Huang H., Svaiter B. F. Gradient descent and fast artificial time integration// ESAIM: M2AN. 2009, vol. 43. P. 689–708.

[2] Porta F., Cornelio A., Ruggiero V. Runge–Kutta-like scaling techniques for first-order methods in convex optimization // Applied Numerical Mathematics. 2017, vol. 116. P. 256–272.

[3] Eftekhari A. , Vandereycken B., Vilmart G., Zygalakis K. C. Explicit stabilised gradient descent for faster strongly convex optimisation // BIT Numerical Mathematics. 2021, vol. 61. P. 119–139.

[4] Stillfjord T. , Williamson M. SRKCD: A stabilized Runge–Kutta method for stochastic optimization // Journal of Computational and Applied Mathematics. 2023, vol. 417. Art. no. 114575.

[5] Zhang J., Mokhtari A., Sra S., Jadbabaie A. Direct Runge-Kutta discretisation achieves acceleration // Advances in Neural Information Processing Systems. 2018, vol. 31.

[6] Zhang J., Sra S., Jadbabaie A. Acceleration in first order quasi-strongly convex optimization by ode discretization // 2019 IEEE 58th Conference on Decision and Control. 2019.

[7] Shi B., Du S.S., Jordan M.I., Su W.J. Understanding the acceleration phenomenon via high-resolution differential equations // Mathematical Programming. 2022, vol. 195. P. 79–148.

[8] Luo H., Chen L. From differential equation solvers to accelerated first-order methods for convex optimization// Mathematical Programming. 2022, vol. 195. P. 735–781.

[9] Duruisseaux V., Leok M. Practical perspectives on symplectic accelerated optimization // Optimization Methods and Software. 2023, vol. 38, no. 6. P. 1230–1268.

[10] Chen R., Li X. Implicit Runge-Kutta methods for accelerated unconstrained convex optimization // IEEE Access. 2020, vol. 8. P. 28624–28634.

[11] Areias P., Rabczuk T. An engineering interpretation of Nesterov's convex minimization algorithm and time integration: application to optimal fiber orientation// Computational Mechanics. 2021, vol. 68, no. 1. P. 211–227.

[12] Альбер С. И., Альбер Я. И. Применение метода дифференциального спуска для решения нелинейных систем // Журнал вычислительной математики и математической физики. 1967, т. 68, № 1. С. 14–32.

[13] Abbott J. P., Brent R. P. Fast local convergence with single and multistep methods for nonlinear equations // The Journal of the Australian Mathematical Society. Series B. Applied Mathematics. 1977, vol. 20, no. 2. P. 173–199.

[14] Brown A. A., Bartholomew-Biggs M. C. Some effective methods for unconstrained optimization based on the solution of systems of ordinary differential equations // Journal of Optimization Theory and Applications. 1989, vol. 62, no. 2. P. 211–224.

[15] Khiyabani F. M., Leong W. J. Quasi-Newton methods based on ordinary differential equation approach for unconstrained nonlinear optimization // Applied Mathematics and Computation. 2014, vol. 233. P. 272–291.

[16] Su W., Boyd S., Candes E. J. A differential equation for modeling Nesterov's accelerated gradient method: Theory and insights // Journal of Machine Learning Research. 2016, vol. 17, no. 53. P. 1–43.

[17] Shi B., Du S. S., Su W., Jordan M. I. Acceleration via symplectic discretization of high-resolution differential equations // Advances in Neural Information Processing Systems. 2019, vol. 17, no. 32.

[18] Chan R. P. K., Tsai A. Y. J. On explicit two-derivative Runge-Kutta methods // Numerical Algorithms. 2010, vol. 53, no. 1. P. 171–194.

[19] Turaci M. O., Ozis T. Derivation of three-derivative Runge-Kutta methods // Numerical Algorithms. 2016, vol. 74. P. 247–265.

[20] Qin X., Yu J., Yan C. Derivation of three-derivative two-step Runge–Kutta methods // Mathematics. 2024, vol. 12, no. 5. Art. no. 711.

[21] Dang Q. A., Hoang M. T. Positive and elementary stable explicit nonstandard Runge–Kutta methods for a class of autonomous dynamical systems // International Journal of Computer Mathematics. 2020, vol. 97, no. 10. P. 2036–2054.

[22] Арушанян О. Б. , Залеткин С. Ф. Приближенное решение задачи Коши для обыкновенных дифференциальных уравнений методом рядов Чебышева // Вычислительные методы и программирование. 2016, т. 17, № 2. С. 121–131.

[23] Ворожцов Е. В. Построение явных разностных схем для обыкновенных дифференциальных уравнений с помощью разложений Лагранжа — Бюрмана // Вычислительные методы и программирование. 2010, т. 11, № 2. С. 198–209.

[24] Vorozhtsov E. V. Derivation of explicit difference schemes for ordinary differential equations with the aid of Lagrange–Burmann expansions // Lecture Notes in Computer Science. 2010, vol. 6244. P. 250–266.

[25] Ворожцов Е. В. Применение разложений Лагранжа – Бюрмана для численного интегрирования уравнений невязкого газа // Вычислительные методы и программирование. 2011, т. 12, № 3. С. 348–361.

[26] Ворожцов Е. В. Конструирование схем третьего порядка точности с помощью разложений Лагранжа – Бюрмана для численного интегрирование уравнений невязкого газа // Вычислительные методы и программирование. 2016, т. 17, № 1. С. 21–43.

[27] Jerez S. Non-standard Lagrange–Burman methods for the numerical integration of differential equations // Journal of Difference Equations and Applications. 2012, vol. 18, no. 11. P. 1899–1912.

[28] Поляк Б. Т. О некоторых способах ускорения сходимости итерационных методов // Журнал вычислительной математики и математической физики. 1964, т. 4, № 5. С. 791–803.

[29] Абрамовиц М., Стиган И. Справочник по специальным функциям. М.: Наука, 1979. 832 с.

[30] Бахвалов Н. С., Жидков Н. П., Кобельков Г. М. Численные методы. М.: Лаборатория знаний, 2020. 636 с.

[31] Гантмахер Ф. Р. Теория матриц. М.: ФИЗМАТЛИТ, 2010. 560 с.

[32] Поляк Б. Т. Введение в оптимизацию. М.: Наука, 1983. 384 с.

[33] Нестеров Ю. Е. Введение в выпуклую оптимизацию. М.: МЦНМО, 2010. 280 с.

[34] Lessard L., Recht B., Packard A. Analysis and design of optimization algorithms via integral quadratic constraints // SIAM Journal on Optimization. 2016, vol. 26, no. 1. P. 57–95.

[35] Самарский А. А., Николаев Е. С. Методы решения сеточных уравнений. М.: Наука, 1978. 592 с.

[36] Николаев Е. С. Методы решения сеточных уравнений. М.: Изд-во МГУ, 2023. 404 с.

[37] Ортега Д., Пул У. Введение в численные методы решения дифференциальных уравнений. М.: Наука, 1986. 288 с.

[38] Самарский А. А., Гулин А. В. Устойчивость разностных схем. М.: Наука, 1973. 415 с.

[39] Эльсгольц Л. Э. Вариационное исчисление. М.: УРСС, 2023. 208 с.

[40] Vasudeva M. A., Verwer J. Solving parabolic integro-differential equations by an explicit integration method // Journal of Computational and Applied Mathematics. 1992, vol. 39, no. 1. P. 121–132.

Application of the modified Runge–Kutta method to the construction of the descent method for solving boundary value problems

G.V. Krivovichev, N.V. Egorov

Saint-Petersburg State University, Faculty of Applied Mathematics and Control Processes

E-mail: g.krivovichev@spbu.ru, n.v.egorov@spbu.ru

Abstract. The work is devoted to the construction and analysis of the gradient method based on a modified explicit second-order Runge–Kutta method, constructed using the Lagrange–Burmann expansion. A two-step method with inertia based on the heavy ball method is proposed. Convergence theorems are proven for strongly convex quadratic and perturbed quadratic functions. Analytical expressions for the optimal parameters of the method are obtained. For the quadratic function it is demonstrated, that the proposed method converges faster, than other well-known accelerated methods.

The results of the application of the method to the numerical solution of linear and nonlinear boundary value problems (Dirichlet problem for a 3D Poisson equation, problems from the calculus of variations, problems for integro-differential equations) are presented. It is demonstrated that, in comparison with well-known methods, the proposed method allows one to obtain a numerical solution with the required accuracy for different grid resolutions in a smaller number of iterations and time.

Key words: Runge–Kutta methods, convex optimization, gradient descent, boundary value problems