

УДК 519.632.4, 519.653

Об одном подходе к численному решению задач Дирихле произвольной размерности*

Б.В. Семисалов^{1,2}

¹Новосибирский национальный исследовательский государственный университет (НГУ), ул. Пирогова, 2, Новосибирск, 630090

²Институт математики им. С.Л. Соболева Сибирского отделения Российской академии наук, просп. Акад. Коптюга, 4, Новосибирск, 630090

E-mail: vibis@ngs.ru

Английская версия этой статьи печатается в журнале “Numerical Analysis and Applications” № 1, Vol. 15, 2022.

Семисалов Б.В. Об одном подходе к численному решению задач Дирихле произвольной размерности // Сиб. журн. вычисл. математики / РАН. Сиб. отд-ние. — Новосибирск, 2022. — Т. 25, № 1. — С. 77–95.

Разработан метод численного решения краевых задач Дирихле для нелинейных дифференциальных уравнений эллиптического типа произвольной размерности, обеспечивающий на гладких решениях низкий расход памяти и машинного времени. Метод основан на применении модифицированных интерполяционных полиномов с узлами Чебышёва для приближения искомой функции и нового подхода к формированию и решению задач линейной алгебры, соответствующих исходным дифференциальным уравнениям. С применением интервальных методов проведён анализ спектра и чисел обусловленности матриц, формируемых алгоритмом. Доказаны теоремы об аппроксимации и устойчивости предложенного алгоритма в линейном случае. Установлено, что на решениях, имеющих высокий порядок гладкости, метод обеспечивает многократное снижение вычислительных затрат по сравнению с классическими схемами методов коллокаций и конечных разностей.

DOI: 10.15372/SJNM20220106

Ключевые слова: *краевая задача Дирихле, снижение вычислительных затрат, псевдоспектральный метод, метод коллокаций, метод установления.*

Semisalov B.V. On an approach to numerical solutions of the Dirichlet problem of an arbitrary dimension // Siberian J. Num. Math. / Sib. Branch of Russ. Acad. of Sci.—Novosibirsk, 2022. — Vol. 25, № 1. — P. 77–95.

A method for the search for numerical solutions to the Dirichlet boundary value problems for nonlinear partial differential equations of the elliptic type and of an arbitrary dimension is proposed. It ensures low consumptions of memory and computer time for the problems with smooth solutions. The method is based on the modified interpolation polynomials with the Chebyshev nodes for approximation of the sought for function and on the new approach to constructing and solving the problems of linear algebra corresponding to the given differential equations. The analysis of spectra and condition numbers of matrices of the designed algorithm is made by applying the interval methods. The theorems on approximation and stability of the algorithm proposed for the linear case are proved. It is shown that the algorithm ensures an essential decrease in computational costs as compared to the classical collocation methods and to finite difference schemes.

Keywords: *Dirichlet boundary value problem, decrease of computational costs, pseudospectral method, collocation method, relaxation method.*

*Работа выполнена за счет гранта РНФ (проект № 20-71-00071).

Введение

При решении краевых задач высокой вычислительной сложности на ЭВМ специалисты неизбежно сталкиваются с проблемой минимизации вычислительных затрат при сохранении необходимой точности решения. Актуальность этой проблемы особенно остро проявляется в приложениях, где требуется проведение расчётов с высокой точностью (квантовая физика, оптика, химия, проблемы создания высокоточных систем навигации, метеорология и др.), при решении нелинейных задач с малыми параметрами и большими градиентами, задач большой размерности, в которых ограниченность объёма памяти ЭВМ становится наиболее критичной, а также при оптимизации сложных систем, сводящейся к решению большого количества прямых задач.

Настоящая работа посвящена решению указанной проблемы. В ней разработан и апробирован новый алгоритм решения нелинейных краевых задач Дирихле для дифференциальных уравнений эллиптического типа, гладкость решений которых зачастую достаточно высока. Известно, что при численном анализе одномерных дифференциальных уравнений с высокой гладкостью коэффициентов и правых частей спектральные методы на основе приближений Фурье и Чебышёва демонстрируют значительную эффективность [1, 2]. Однако в случае задач высокой размерности применение таких приближений требует обращения матриц большого размера и сложной структуры, что влечёт существенный рост вычислительных затрат и погрешностей. В настоящей работе предложен подход к аппроксимации и приближённому решению дифференциальных уравнений в областях канонических форм, свободный от указанных недостатков. Конкретно в работе рассматриваются d -мерные краевые задачи Дирихле в областях $D = [-1, 1]^d$ для уравнений эллиптического типа второго порядка вида

$$\begin{aligned} d = 1 : & \quad \mathbf{u}_{xx} = f(\mathbf{u}, x), \\ d = 2 : & \quad \mathbf{u}_{xx} + \mathbf{u}_{yy} = f(\mathbf{u}, x, y), \\ d - \text{произвольное} : & \quad \frac{\partial^2 \mathbf{u}}{\partial (x^1)^2} + \frac{\partial^2 \mathbf{u}}{\partial (x^2)^2} + \dots + \frac{\partial^2 \mathbf{u}}{\partial (x^d)^2} = f(\mathbf{u}, x^1, \dots, x^d), \end{aligned} \quad (1)$$

где (x, y) и (x^1, \dots, x^d) — прямоугольные декартовы системы координат, $\mathbf{u} \in C^2(D)$ — неизвестная функция, f — правая часть, зависящая нелинейным образом от решения. Разрабатываемый метод несложно модифицировать для решения уравнений с переменными коэффициентами и квазилинейных уравнений. Однако для краткости в этой работе такое обобщение опущено. Граничные условия для уравнений (1) выглядят следующим образом:

$$\begin{aligned} d = 1 : & \quad \mathbf{u}(-1) = a, \quad \mathbf{u}(1) = b, \quad a, b \in \mathbb{R}, \\ d = 2 : & \quad \mathbf{u}(\pm 1, y) = \varphi_{\pm 1}^{\pm}(y), \quad \mathbf{u}(x, \pm 1) = \varphi_{\pm 2}^{\pm}(x), \\ d - \text{произвольное} : & \quad \mathbf{u}(x^1, \dots, x^{\delta-1}, \pm 1, x^{\delta+1}, \dots, x^d) \\ & \quad = \psi_{\delta}^{\pm}(x^1, \dots, x^{\delta-1}, x^{\delta+1}, \dots, x^d), \quad \delta = 1, \dots, d, \end{aligned} \quad (2)$$

причём значения функций $\varphi_{1,2}^{\pm}$ и $\psi_{1,\dots,d}^{\pm}$ согласованы в угловых точках области $[-1, 1]^2$ и на рёбрах области $[-1, 1]^d$ соответственно. Для аппроксимации функции в работе будет использовано представление

$$\mathbf{u}(\mathbf{x}) = u(\mathbf{x}) + v(\mathbf{x}), \quad \mathbf{x} \in D, \quad (3)$$

где функция $u(\mathbf{x})$ принимает нулевые значения на границе D , а функция $v(\mathbf{x})$ реализует заданные граничные условия.

В пункте 1 описан удобный способ аппроксимации неизвестных функций и их производных на основе интерполяционных полиномов с узлами Чебышёва, позволяющий автоматически учесть тип граничных условий. Получены оценки погрешности аппроксимации решения и его производных. Спектр матричных операторов, аппроксимирующих производные, вычислен с гарантией точности на основе интервальных методов. Пункт 2 посвящён разработке методов решения задачи линейной алгебры, соответствующей исходной дифференциальной постановке. Здесь на основе метода коллокаций, итерационного метода установления с нестационарными регуляризациями и методов, описанных в [3], получены системы линейных алгебраических уравнений (СЛАУ) в форме уравнений Сильвестра и их тензорных обобщений. Далее предложен аппарат для работы с этими уравнениями, позволивший на несколько порядков снизить количество операций по сравнению с классической схемой метода коллокаций (КСМК). Под последней здесь и далее понимается алгоритм, работающий со СЛАУ вида $\tilde{A}\vec{x} = \vec{b}$ с матрицей \tilde{A} большого размера и сложной структуры.

Сравнивая разработанный метод с методами конечных разностей (МКР), отметим, что для достижения заданной точности в задачах с гладкими решениями созданному методу требуется значительно меньше узлов пространственной сетки, чем, например, схеме переменных направлений на основе МКР. Это определяет значительный выигрыш предложенного метода в количестве операций и затратах памяти. Тестовые расчёты, демонстрирующие указанные преимущества, приводятся в п. 3.

Важной особенностью работы является применение единого (нелокального) приближения неизвестной функции полиномом (1D-случай) либо прямым произведением полиномов (случай произвольной размерности). В связи с этим разработанный метод будем именовать далее **нелокальным методом** (НМ).

1. Аппроксимация неизвестных функций и их производных в задачах Дирихле

Следуя [4, гл. 9, § 5], для приближения решений будем использовать специальные модификации интерполяционных полиномов с узлами в корнях многочленов Чебышёва. Опишем ниже метод аппроксимации дифференциальных операторов, основанный на таких приближениях, и получим оценки погрешности. Обозначим $x_j = \cos((2j-1)\pi/2N)$ — узлы многочлена Чебышёва $T_N(x)$ степени N , $j = 1, \dots, N$. Положим

$$P_{1D}(u, x) = \sum_{j=1}^N \frac{\zeta(x, x_j) T_N(x)}{(x - x_j) T'_N(x_j)} u(x_j), \quad \zeta(x, x_j) = \frac{1 - x^2}{1 - x_j^2}, \quad x \notin \{x_1, \dots, x_N, \pm 1\}. \quad (4)$$

Здесь и в формулах (14), (18) значения полинома в узлах интерполяции доопределяются по непрерывности: $P_{1D}(u, x_j) = u(x_j)$, $P_{1D}(u, \pm 1) = 0$. Множитель $\zeta(x, x_j)$ позволяет автоматически учесть, что функция $u(x)$ из (3) удовлетворяет однородным условиям Дирихле. Оценка погрешности для модифицированных интерполяций дана ниже.

Для приближения решения будем использовать выражение $u(x) \approx P_{1D}(u, x) + v(x)$, где $P_{1D}(u, \pm 1) = 0$, $v(x)$ — линейная функция, удовлетворяющая неоднородным условиям (см. (2) случай $d = 1$).

Для приближения первых и вторых производных $u(x)$ с условиями $u(-1) = a$, $u(1) = b$ положим $v(x) = ((b - a)x + (a + b))/2$,

$$u'(x) \approx P'_{1D}(u, x) + (b - a)/2, \quad u''(x) \approx P''_{1D}(u, x)$$

и используем матричную аппроксимацию операторов дифференцирования (в [5] подобная аппроксимация названа “matrix differentiation technique”). При построении аппроксимации для упрощения выкладок будем полагать, что краевые условия являются однородными, т. е. $a = b = 0$, $v(x) \equiv 0$.

Положим $s = \sqrt{1 - x^2}$, $s_j = \sqrt{1 - x_j^2}$, $u_j = u(x_j)$. Дифференцируя полином (4) по x , получаем

$$P'_{1D}(u, x) = \frac{1}{N} \sum_{j=1}^N (-1)^{j-1} \left\{ \frac{2xx_j - x^2 - 1}{(x - x_j)^2 s_j} T_N(x) + \frac{s^2}{s_j(x - x_j)} T'_N(x) \right\} u_j, \quad (5)$$

$$P''_{1D}(u, x) = \frac{1}{N} \sum_{j=1}^N (-1)^{j-1} \left\{ \frac{2s_j^2 - N^2(x - x_j)^2}{(x - x_j)^3 s_j} T_N(x) - \frac{(x^2 - 3xx_j + 2)}{s_j(x - x_j)^2} T'_N(x) \right\} u_j. \quad (6)$$

Переходя в (5), (6) к пределу при $x \rightarrow x_i$ и используя правило Лопиталья, находим

$$P'_{1D}(u, x_i) = \sum_{j=1, j \neq i}^N (-1)^{i+j} \frac{s_i}{s_j(x_i - x_j)} u_j - \frac{3x_i}{2s_i^2} u_i, \quad (7)$$

$$P''_{1D}(u, x_i) = \sum_{j=1, j \neq i}^N (-1)^{i+j-1} \frac{2s_i^2 + 3x_i(x_i - x_j)}{s_j s_i(x_i - x_j)^2} u_j - \frac{(N^2 + 5)s_i^2 + 3x_i^2}{3s_i^4} u_i. \quad (8)$$

Введём обозначения:

$$\mathbf{a}_{ij} = \frac{(-1)^{i+j} s_i}{s_j(x_i - x_j)}, \quad a_{ij} = (-1)^{i+j} \frac{2s_i^2 + 3x_i(x_i - x_j)}{s_j s_i(x_i - x_j)^2}, \quad i, j = 1, \dots, N, \quad i \neq j, \quad (9)$$

$$\mathbf{n}_i = -\frac{3x_i}{2s_i^2}, \quad \nu_i = -\frac{(N^2 + 5)s_i^2 + 3x_i^2}{3s_i^4}, \quad i = 1, \dots, N, \quad (10)$$

$$U = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_N \end{pmatrix}, \quad \mathbf{u}_x = \begin{pmatrix} (\mathbf{u}_x)_1 \\ (\mathbf{u}_x)_2 \\ \vdots \\ (\mathbf{u}_x)_N \end{pmatrix} = \begin{pmatrix} \mathbf{u}_x(x_1) \\ \mathbf{u}_x(x_2) \\ \vdots \\ \mathbf{u}_x(x_N) \end{pmatrix}, \quad \mathbf{u}_{xx} = \begin{pmatrix} (\mathbf{u}_{xx})_1 \\ (\mathbf{u}_{xx})_2 \\ \vdots \\ (\mathbf{u}_{xx})_N \end{pmatrix} = \begin{pmatrix} \mathbf{u}_{xx}(x_1) \\ \mathbf{u}_{xx}(x_2) \\ \vdots \\ \mathbf{u}_{xx}(x_N) \end{pmatrix},$$

и сформируем $N \times N$ -матрицы:

$$\mathfrak{A} = \begin{pmatrix} \mathbf{n}_1 & \mathbf{a}_{12} & \dots & \mathbf{a}_{1N} \\ \mathbf{a}_{21} & \mathbf{n}_2 & \dots & \mathbf{a}_{2N} \\ \vdots & \vdots & \dots & \vdots \\ \mathbf{a}_{N1} & \mathbf{a}_{N2} & \dots & \mathbf{n}_N \end{pmatrix}, \quad A = \begin{pmatrix} \nu_1 & a_{12} & \dots & a_{1N} \\ a_{21} & \nu_2 & \dots & a_{2N} \\ \vdots & \vdots & \dots & \vdots \\ a_{N1} & a_{N2} & \dots & \nu_N \end{pmatrix}.$$

Для аппроксимации производных в уравнениях (1) (при $d = 1$) в случае однородных краевых условий получим формулы

$$\mathbf{u}_x \approx \mathfrak{A}U, \quad \mathbf{u}_{xx} \approx AU, \quad (11)$$

где V_x — вектор с компонентами $(v_x)_i = (b - a)/2 \quad \forall i = 1, \dots, N$.

Теорема 1. Пусть $\mathbf{u}(x) \in C^{N+2}([-1, 1])$ и $\mathbf{u}(-1) = a$, $\mathbf{u}(1) = b$. Тогда модифицированный полином $\mathbf{P}_N(\mathbf{u}, x) = P_{1D}(u, x) + v(x)$, где $u = \mathbf{u} - v$, $v = ((b-a)x + (a+b))/2$, удовлетворяет оценке погрешности приближения функции $\mathbf{u}(x)$ интерполяционным полиномом с узлами Чебышёва:

$$\|\mathbf{P}_N(\mathbf{u}, x) - \mathbf{u}(x)\| \leq \frac{\|\mathbf{u}^{(N+2)}(x)\|}{(N+2)!} 2^{1-N}, \quad (12)$$

где $\|\cdot\|$ – супремум норма непрерывной функции.

Доказательство. Для начала отметим, что $\mathbf{P}_N(\mathbf{u}, x) - \mathbf{u}(x) = P_{1D}(u, x) - u(x)$.

Пусть $\mathcal{P}_{1D}(u, x)$ – интерполяционный полином для функции u с узлами $-1, x_1, \dots, x_N, 1$, где x_1, \dots, x_N – корни многочлена Чебышёва $T_N(x)$; $\omega(x) = (x^2 - 1)T_N(x)$. Заметим, что в силу того, что $u(\pm 1) = 0$, имеют место соотношения

$$\begin{aligned} \mathcal{P}_{1D}(u, x) &= \frac{\omega(x)u(-1)}{(x+1)\omega'(-1)} + \sum_{k=1}^N \frac{T_N(x)(x^2-1)u(x_k)}{(x-x_k)[(T_N(x)(x^2-1))]'|_{x=x_k}} + \frac{\omega(x)u(1)}{(x-1)\omega'(1)} \\ &= \sum_{k=1}^N \frac{(1-x^2)T_N(x)}{(x-x_k)(1-x_k^2)T'_N(x_k)} u(x_k) = P_{1D}(u, x). \end{aligned}$$

Здесь использованы равенства $[(T_N(x)(x^2-1))]'|_{x=x_k} = [(T'_N(x)(x^2-1)) + 2xT_N(x)]|_{x=x_k} = T'_N(x_k)(x_k^2-1) + 2x_kT_N(x_k) = T'_N(x_k)(x_k^2-1)$.

С другой стороны, применяя классическую теорему об оценке погрешности интерполяции функции $u(x)$ [6], получаем:

$$\begin{aligned} \|\mathcal{P}_{1D}(u, x) - u(x)\| &\leq \frac{\|u^{(N+2)}(x)\|}{(N+2)!} \|\omega(x)\| = \frac{\|u^{(N+2)}(x)\|}{(N+2)!} \|(x^2-1)T_N(x)\| \\ &\leq \frac{\|u^{(N+2)}(x)\|}{(N+2)!} \|(x^2-1)\| \|T_N(x)\| = \frac{\|u^{(N+2)}(x)\|}{(N+2)!} \times 1 \times 2^{N-1}. \end{aligned}$$

Объединяя приведённые выкладки и учитывая, что при любых N $u^{(N+2)}(x) = \mathbf{u}^{(N+2)}(x)$, получаем, что для функции $\mathbf{u}(x) \in C^{N+2}([-1, 1])$ с условиями $\mathbf{u}(-1) = a$, $\mathbf{u}(1) = b$ выполнена оценка (12). \square

Теорема 2. Обозначим $\text{Lip}_M(D)$ – класс липшицевых функций на отрезке $D = [-1, 1]$ с константой Липшица M , $W_\infty^N(M, D)$ – замыкание класса функций $\{f \in C^{N-1}(D) : f^{(N-1)} \in \text{Lip}_M(D)\}$ в супремум-норме. Пусть $\mathbf{u}(x) \in W_\infty^{N+2}(M, D)$, $1 \leq s < N+2$, тогда

$$\|\mathbf{P}_N^{(s)}(\mathbf{u}, x) - \mathbf{u}^{(s)}(x)\| \leq M \frac{2^{N+2-s}}{(N+2-s)!}.$$

Заключение теоремы вытекает из соотношений $\|\mathbf{P}_N^{(s)}(\mathbf{u}, x) - \mathbf{u}^{(s)}(x)\| = \|\mathcal{P}_{1D}^{(s)}(u, x) - u^{(s)}(x)\|$, которые имеют место в силу рассуждений, проведённых выше, а также из [4, гл. 3, § 3, теорема 4], применённой для оценки $\|\mathcal{P}_{1D}^{(s)}(u, x) - u^{(s)}(x)\|$.

Из теоремы 2 следует, что при высоком порядке гладкости функции $\mathbf{u}(x)$ погрешность предложенных аппроксимаций операторов дифференцирования стремится к нулю со скоростью не ниже $O(2^N/N!)$ при увеличении числа узлов N .

Для разработки быстрого метода решения краевых задач нам потребуется спектральное разложение матрицы A , аппроксимирующей вторые производные:

$$A = R_A D_A R_A^{-1}, \quad (13)$$

где R_A — матрица собственных векторов A , D_A — диагональная матрица, содержащая собственные значения d_A^j матрицы A на диагонали ($j = 1, \dots, N$).

Замечание 1. При численной реализации алгоритма решения задач Дирихле (1), (2) матрицы \mathfrak{A} , A , R_A , D_A , R_A^{-1} для разных значений N целесообразно занести в базу данных, чтобы не тратить время на их расчёт.

Анализ собственных чисел матрицы A и чисел обусловленности матрицы R_A был проведён с помощью интервальных методов [7, 8].

Пакет позволяет проводить расчёты с гарантией точности, иными словами, вычислять радиусы интервалов, в которые гарантированно попадает искомое значение собственного числа или значения числа обусловленности матрицы. В таблице 1 приведены значения центров и радиусов интервалов, в которые гарантированно попадают значения первых 10-ти собственных чисел матрицы A .

Таблица 1. Значения собственных чисел d_A^j матрицы A при $N = 10$ и радиусы интервалов rad_j , гарантированно содержащих собственные числа d_A^j

j	d_A^j	rad_j	j	d_A^j	rad_j
1	$-3.169920593590947e + 03$	$9.09e - 13$	6	-62.595104941014185	$2.13e - 14$
2	$-3.034196464055946e + 03$	$4.55e - 13$	7	-39.453871795378625	$2.84e - 14$
3	$-1.964915419758917e + 02$	$5.68e - 14$	8	-22.206898938237902	$2.13e - 14$
4	$-1.785341309646627e + 02$	$5.68e - 14$	9	-9.869604101936476	$2.66e - 14$
5	-84.264388535869273	$2.84e - 14$	10	-2.467401100161208	$2.44e - 14$

На рис. 1 и 2 приведены значения верхних границ обусловленностей $\text{cond}_{R_A}^{2,\infty}$ матриц R_A , посчитанных во второй и бесконечной нормах, и относительные радиусы интервалов $\text{rad}_{R_A}^{2,\infty}$ в логарифмической шкале, в которых гарантированно лежат эти значения.

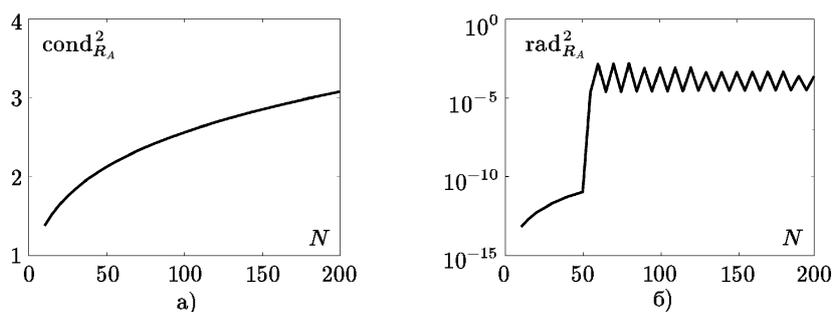


Рис. 1. Зависимости $\text{cond}_{R_A}^2$ (а) и $\text{rad}_{R_A}^2$ (б) от количества узлов N

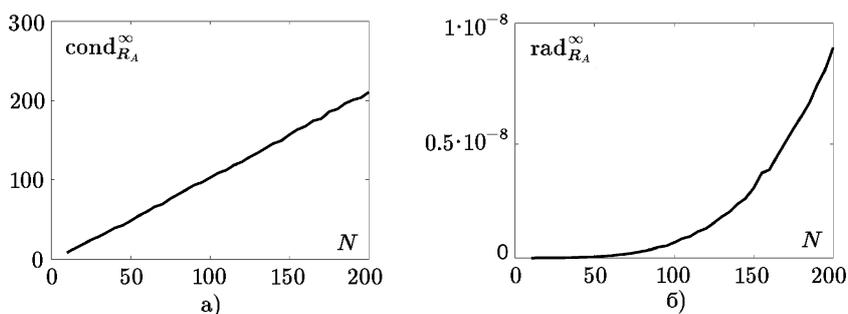


Рис. 2. Зависимости $\text{cond}_{R_A}^\infty$ (а) и $\text{rad}_{R_A}^\infty$ (б) от N

Благодаря проведённому анализу удаётся сделать следующие выводы, имеющие важное значение для обоснования предложенного метода:

1. Собственные значения матрицы A являются отрицательными действительными числами и достаточно далеко отстоят друг от друга на действительной оси.
2. Обусловленность матрицы R_A с ростом числа узлов растёт достаточно медленно и принимает малые значения. Этот факт обеспечивает высокую устойчивость предложенного алгоритма к погрешностям округления.

В случае двумерной задачи ($d = 2$, $D = [-1, 1]^2$, $u(x, y) \in C^2(D)$ — искомая функция) в области D вводится сетка с узлами (x_j, y_k) , где $x_j = \cos \frac{2j-1}{2N} \pi$, $y_k = \cos \frac{2k-1}{2K} \pi$, $j = 1, \dots, N$, $k = 1, \dots, K$. Далее используем обозначения: $u_{jk} = u(x_j, y_k)$, $(u_\mu)_{jk} = u_\mu(x_j, y_k)$, где μ обозначает операцию дифференцирования ($\mu \in \{x, y, xx, yy, xy\}$, $u = u - v$, см. (3)). Пусть $U = (u_{jk})$, $\mathfrak{U}_\mu = ((u_\mu)_{jk})$ — $N \times K$ -матрицы. Аппроксимация $u(x, y)$ с применением прямого произведения интерполяционных полиномов (4) имеет вид

$$u(x, y) \approx P_{2D}(u, x, y) + v(x, y),$$

$$P_{2D}(u, x, y) = \sum_{j=1}^N \sum_{k=1}^K \frac{\zeta(x, x_j) T_N(x)}{(x - x_j) T'_N(x_j)} \frac{\zeta(y, y_k) T_K(y)}{(y - y_k) T'_K(y_k)} u(x_j, y_k). \quad (14)$$

Здесь P_{2D} удовлетворяет нулевым граничным условиям, а $v(x, y)$ реализует заданные ненулевые граничные значения $v = v^x + v^y$, где $v^x(x, y) = \alpha_x(x)y + \beta_x(y)$, $v^y(x, y) = \alpha_y(x)y + \beta_y(x)$,

$$v^x|_{y=\pm 1} = v^y|_{x=\pm 1} = 0. \quad (15)$$

Величины (α_x, β_x) и (α_y, β_y) рассчитываются, исходя из условий на границах $x = \pm 1$, $y = \pm 1$ соответственно. Равенство (15) обеспечивает для функции $v(x, y)$ выполнение необходимых условий на границе D при суммировании $v = v^x + v^y$.

Реализуя вычисления, аналогичные (5)–(11) (см. также [3]), получаем матрицы, служащие для аппроксимации производных $u(x, y)$ по x (\mathfrak{A} , A размера $N \times N$) и по y (\mathfrak{B} , B размера $K \times K$):

$$\mathfrak{U}_x \approx \mathfrak{A}U + V_x, \quad \mathfrak{U}_y \approx U\mathfrak{B}^\top + V_y, \quad \mathfrak{U}_{xx} \approx AU, \quad \mathfrak{U}_{yy} \approx UB^\top, \quad \mathfrak{U}_{xy} \approx \mathfrak{A}U\mathfrak{B}^\top + V_{xy}. \quad (16)$$

Здесь V_μ представляют матрицы значений производных $v_\mu(x, y)$ на сетке (x_j, y_k) . Эти значения рассчитываются следующим образом:

$$V_x = V_x^x + \mathfrak{A}V^y, \quad V_y = V^x\mathfrak{B}^\top + V_y^y, \quad V_{xy} = V_x^x\mathfrak{B}^\top + \mathfrak{A}V_y^y, \quad (17)$$

где $V_\mu^{x,y}$ — матрицы значений $v_\mu^{x,y}(x_j, y_k)$.

Замечание 2. Функции α_x , β_x , α_y , β_y за счёт условия (15) могут иметь разрывы в точках $y = \pm 1$, $x = \pm 1$ соответственно. В экспериментах установлено, что осцилляции, возникающие из-за этого при аппроксимации производных функций $v(x, y)$ по формулам (17), подавляются в итерационном процессе метода установления (см. п. 2) и не увеличивают порядок погрешности полученных решений.

В случае произвольной размерности задачи d ($D = [-1, 1]^d$, $\mathbf{u}(\mathbf{x}) \in C^2(D)$) — иско-мая функция, $\mathbf{x} = (x^1, x^2, \dots, x^d)$ введём в области D сетку с узлами $(x_{j_1}^1, \dots, x_{j_d}^d)$, где $x_{j_\delta}^\delta = \cos \frac{2j_\delta - 1}{2N_\delta} \pi$, $j_\delta = 1, \dots, N_\delta$, N_δ — число узлов по направлению δ ($\delta = 1, \dots, d$). Обозначим $u\langle j_1 j_2, \dots, j_d \rangle = u(x_{j_1}^1, x_{j_2}^2, \dots, x_{j_d}^d)$, $\mathbf{u}_\mu\langle j_1 j_2, \dots, j_d \rangle = \mathbf{u}_\mu(x_{j_1}^1, x_{j_2}^2, \dots, x_{j_d}^d)$, $\mu \in \{x^1, \dots, x^d, x^1 x^1, x^1 x^2, \dots, x^d x^d\}$ — одна из производных, $u = \mathbf{u} - v$, см. (3). Пусть $U = (u\langle j_1 j_2, \dots, j_d \rangle)$, $\mathfrak{U}_\mu = (\mathbf{u}_\mu\langle j_1 j_2, \dots, j_d \rangle)$ — d -мерные массивы размера $N_1 \times N_2 \times \dots \times N_d$. Для аппроксимации решения $\mathbf{u}(\mathbf{x})$ используем прямое произведение полиномов (4):

$$P_d(u, \mathbf{x}) = P_d(u, x^1, \dots, x^d) = \sum_{j_1=1}^{N_1} \dots \sum_{j_d=1}^{N_d} u\langle j_1 j_2, \dots, j_d \rangle \prod_{k=1}^d \frac{\zeta(x^k, x_{j_k}^k) T_{N_k}(x^k)}{(x^k - x_{j_k}^k) T'_{N_k}(x_{j_k}^k)}. \quad (18)$$

Тогда $\mathbf{u}(\mathbf{x}) \approx P_d(u, \mathbf{x}) + v(\mathbf{x})$, P_d удовлетворяет однородным, а $v(\mathbf{x})$ — неоднородным граничным условиям и строится аналогично 2D-случаю. По аналогии с (5)–(11) получаем матрицы размеров $N_1 \times N_1; \dots; N_d \times N_d$, служащие для аппроксимации производных $\mathbf{u}(\mathbf{x})$ по переменным x^1, \dots, x^d . Обозначим эти матрицы $\mathfrak{A}_1, A_1; \dots; \mathfrak{A}_d, A_d$.

Определение. δ -произведением аппроксимирующей матрицы $A_\delta = (a_{j_s})_{N_\delta \times N_\delta}$ и массива U будем называть массив $\alpha_\delta = A_\delta \times U$, элементы которого определяются по формуле $\alpha_\delta\langle j_1, \dots, j_{\delta-1}, k, j_{\delta+1}, \dots, j_d \rangle = \sum_{s=1}^{N_\delta} a_{ks} u\langle j_1, \dots, j_{\delta-1}, s, j_{\delta+1}, \dots, j_d \rangle$, $k = 1, \dots, N_\delta$, $\delta = 1, \dots, d$.

Графическая интерпретация введённого определения в частном трёхмерном случае на основе операций со слоистыми 3D-массивами дана в [3].

Введённая операция δ -произведения позволяет представить аппроксимацию операций дифференцирования в следующем виде:

$$\begin{aligned} \mathfrak{U}_{x^1} &\approx \mathfrak{A}_1 \times^1 U + V_{x^1}, \quad \dots, \quad \mathfrak{U}_{x^d} \approx \mathfrak{A}_d \times^d U + V_{x^d}, \\ \mathfrak{U}_{x^1 x^1} &\approx A_1 \times^1 U + V_{x^1 x^1}, \quad \dots, \quad \mathfrak{U}_{x^d x^d} \approx A_d \times^d U + V_{x^d x^d}, \\ \mathfrak{U}_{x^1 x^2} &\approx \mathfrak{A}_1 \times^1 \mathfrak{A}_2 \times^2 U + V_{x^1 x^2}, \quad \dots, \quad \mathfrak{U}_{x^{d-1} x^d} \approx \mathfrak{A}_{d-1} \times^{d-1} \mathfrak{A}_d \times^d U + V_{x^{d-1} x^d}. \end{aligned} \quad (19)$$

Несложно проверить (см. [3] для случая $d = 3$), что введённые операции δ -произведения являются ассоциативными в том смысле, что, например,

$$\mathfrak{A}_1 \times^1 \mathfrak{A}_2 \times^2 U = \mathfrak{A}_2 \times^2 \mathfrak{A}_1 \times^1 U.$$

Кроме того, эти произведения можно комбинировать со спектральными разложениями аппроксимирующих матриц:

$$A_1 \times^1 U = R_{A_1} \times^1 [D_{A_1} \times^1 (R_{A_1}^{-1} \times^1 U)].$$

Отметим, что если выражение содержит несколько операций δ -произведения, то для корректности эти операции должны выполняться справа налево.

Описанные приближения неизвестной функции и её производных можно обобщить на случай производных произвольного порядка.

2. Разработка метода итерационного решения задач (1)

После аппроксимации решения и его производных в уравнениях (1) приходим к нелинейным соотношениям, записанным для векторов ($d = 1$), матриц ($d = 2$) или d -мерных массивов. Будем искать решения этих нелинейных соотношений с помощью **метода установления**. Для этого обозначим оператор Лапласа символом Δ и представим (1) и (2) в общем виде:

$$\Delta u = f(u, \mathbf{x}), \quad u|_{\partial D} = g(\mathbf{x}). \quad (20)$$

Метод установления использует дополнительную временную переменную t и нестационарный оператор — **регуляризацию** B_t — для организации итерационного процесса. Далее осуществляется переход от задачи (1), (2) к начально-краевой задаче с неизвестной функцией $u(t, \mathbf{x})$:

$$B_t u = \Delta u - f(u, \mathbf{x}), \quad u(t, \mathbf{x})|_{x \in \partial D} = g(\mathbf{x}), \quad u(0, \mathbf{x}) = u_0(\mathbf{x}). \quad (21)$$

Решение задачи (20) ищется как предел решений (21) при $t \rightarrow \infty$. Для построения вычислительных схем рассмотрим простую регуляризацию $B_t = \frac{\partial}{\partial t}$ и регуляризацию Соболева $B_t = (k_1 - k_2 \Delta) \frac{\partial}{\partial t}$, где $k_1, k_2 > 0$ — параметры. Введём дискретизацию по переменной t с постоянным шагом τ и узлами $t_s = \tau s$, $s = 1, 2, \dots$. Обозначим \hat{u} , u — компоненты решений, получаемых на s -м и $(s - 1)$ -м шагах по времени, удовлетворяющие однородным граничным условиям, см. (3). Используя разностное представление производной $\frac{\partial}{\partial t} u \approx \frac{\hat{u} - u}{\tau}$, выведем соотношения для расчёта значений \hat{u} по значениям u для случая простой регуляризации:

$$\hat{u} - \tau \Delta \hat{u} = u - \tau f(u, \mathbf{x}) + \tau \Delta v, \quad (22)$$

а также для случая регуляризации Соболева:

$$k_1 \hat{u} - (k_2 + \tau) \Delta \hat{u} = (k_1 - k_2 \Delta) u - \tau f(u, \mathbf{x}) + \tau \Delta v. \quad (23)$$

Отметим, что выражения для добавочной функции $v(\mathbf{x})$ не зависят от номера шага по времени, и в силу линейности v выражения $\tau \Delta v$, стоящие в правых частях (21), (22), равны нулю. Сходимость полученного итерационного процесса для каждой конкретной задачи может быть исследована на основе теорем о неподвижных точках нелинейных операторов [9, гл. VIII]. Условие останова процесса

$$\frac{\|B_t u\|}{\|f(u, \mathbf{x})\|} = \frac{\|\Delta u - f(u, \mathbf{x})\|}{\|f(u, \mathbf{x})\|} \leq \varepsilon_S \quad (24)$$

даёт приближённое решение задачи (20) с невязкой установления ε_S . В качестве начальных данных во многих задачах можно использовать нулевые значения.

Далее рассмотрим пример построения алгоритма с использованием простой регуляризации. В соответствии с (22), для решения (1) получим:

$$\begin{aligned} d = 1 : & \quad \left(1 - \tau \frac{\partial^2}{\partial x^2}\right) \hat{u} = u - \tau f(u, x), \\ d = 2 : & \quad \left(1 - \tau \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right)\right) \hat{u} = u - \tau f(u, x, y), \\ d - \text{произвольное} : & \quad \left(1 - \tau \left(\frac{\partial^2}{\partial (x^1)^2} + \dots + \frac{\partial^2}{\partial (x^d)^2}\right)\right) \hat{u} = u - \tau f(u, \mathbf{x}). \end{aligned} \quad (25)$$

Используя представления (11), (16), (19) для приближения вторых производных в уравнениях (25), приходим к задачам линейной алгебры на каждом шаге метода установления:

$$(E - \tau A)\widehat{U} = U - \tau F(U + \mathcal{V}) = H(U), \quad (26)$$

$$\widehat{U} - \tau(A\widehat{U} + \widehat{U}B^\top) = U - \tau F(U + \mathcal{V}) = H(U), \quad (27)$$

$$\widehat{U} - \tau\left(A_1 \overset{1}{\times} \widehat{U} + A_2 \overset{2}{\times} \widehat{U} + \dots + A_d \overset{d}{\times} \widehat{U}\right) = U - \tau F(U + \mathcal{V}) = H(U), \quad (28)$$

где E — единичная матрица, \widehat{U} и U — вектора, матрицы или массивы (в зависимости от размерности задачи), содержащие значения функций $u = \mathbf{u} - v$ в узлах интерполяции на текущем и предыдущем временных слоях, \mathcal{V} содержит значения добавочной функции, а F — значения правой части в узлах интерполяции. Уравнение (27) представляет собой матричное уравнение Сильвестра, а (28) — его тензорное обобщение (см. рис. 3 для случая $d = 3$).

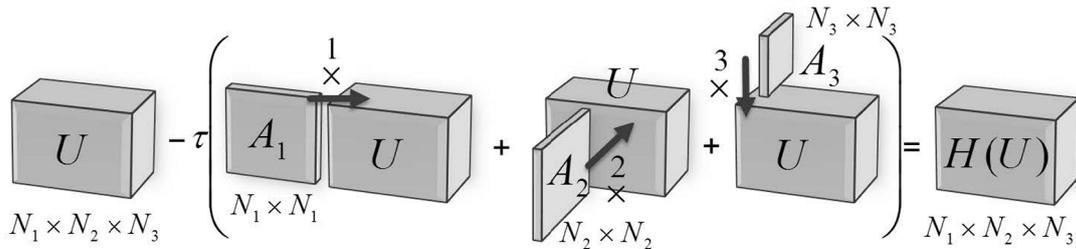


Рис. 3. Схема δ -произведений матриц и массивов в уравнении (28) при $d = 3$

Умножим теперь (26) слева на матрицу R_A^{-1} , затем умножим (27) слева на матрицу R_A^{-1} и справа на матрицу R_B^{-1} , а также запишем δ -произведения матриц $R_{A_d}^{-1}$ и выражений, стоящих в левых и правых частях (28). В итоге получим следующие выражения:

$$\begin{aligned} (E - \tau D_A)\widehat{V} &= G(U), \\ \widehat{V} - \tau(D_A\widehat{V} + \widehat{V}D_B) &= G(U), \\ \widehat{V} - \tau(D_{A_1} \overset{1}{\times} \widehat{V} + D_{A_2} \overset{2}{\times} \widehat{V} + \dots + D_{A_d} \overset{d}{\times} \widehat{V}) &= G(U), \end{aligned} \quad (29)$$

где

$$\begin{aligned} \text{при } d = 1 : \quad & \widehat{V} = R_A^{-1}\widehat{U}, \quad G(U) = R_A^{-1}H(U), \\ \text{при } d = 2 : \quad & \widehat{V} = R_A^{-1}\widehat{U}R_B^{-1}, \quad G(U) = R_A^{-1}H(U)R_B^{-1}, \\ \text{при произвольных } d : \quad & \widehat{V} = R_{A_1}^{-1} \overset{1}{\times} \dots \overset{d-1}{\times} R_{A_d}^{-1} \overset{d}{\times} \widehat{U}, \\ & G(U) = R_{A_1}^{-1} \overset{1}{\times} \dots \overset{d-1}{\times} R_{A_d}^{-1} \overset{d}{\times} H(U). \end{aligned} \quad (30)$$

Отметим, что в последнем выражении операции нужно выполнять справа налево. Решение (29) даётся простыми формулами:

- при $d = 1$ компоненты вектора \widehat{V} выражаются равенствами $\hat{v}_j = \frac{g_j}{1 - \tau d_A^j}$;

- при $d = 2$ элементы матрицы \widehat{V} определяются равенствами $\widehat{v}_{jk} = \frac{g_{jk}}{1 - \tau(d_A^j + d_B^k)}$;
- при $d = 3$ элементы d -мерного массива \widehat{V} суть $\widehat{v}\langle j_1, \dots, j_d \rangle = \frac{g\langle j_1, \dots, j_d \rangle}{1 - \tau(d_{A_1}^{j_1} + d_{A_2}^{j_2} + \dots + d_{A_d}^{j_d})}$.

Здесь $g_j, g_{jk}, g\langle j_1, \dots, j_d \rangle$ — компоненты вектора, элементы матрицы или массива G (в зависимости от размерности), $d_A^j, d_B^k, d_{A_\delta}^{j_\delta}$ — собственные числа матриц A, B, A_δ , $j = 1, \dots, N, k = 1, \dots, K, j_\delta = 1, \dots, N_\delta, \delta = 1, \dots, d$. После вычисления \widehat{V} выполняем операции:

$$d = 1: \widehat{U} = R_A \widehat{V}; \quad d = 2: \widehat{U} = R_A \widehat{V} R_B; \quad d = 3: \widehat{U} = R_{A_1} \times \dots \times^{d-1} R_{A_d} \times^d \widehat{V} \quad (31)$$

и находим решения на текущем шаге метода установления, прибавляя к значениям \widehat{U} значения добавочной функции $v(x^1, \dots, x^d)$ в соответствующих узлах. При решении необходимо следить, чтобы выполнялись условия:

$$\tau \neq \frac{1}{d_A^j}, \quad \tau \neq \frac{1}{d_A^j + d_B^k}, \quad \tau \neq \left(\sum_{\delta=1}^d d_{A_\delta}^{j_\delta} \right)^{-1} \quad \forall j, k, j_\delta. \quad (32)$$

Теорема 3. При решении линейных задач, когда правые части уравнений (1) не зависят от u , разработанный метод является устойчивым для любых τ при использовании простой регуляризации и для $\tau \geq k_2 \left(\frac{M}{m} - 1 \right)$ в случае регуляризации Соболева, где $M = \sum_{\delta=1}^d \max_{j_\delta=1, \dots, N_\delta} |d_{A_\delta}^{j_\delta}|$, $m = \sum_{\delta=1}^d \min_{j_\delta=1, \dots, N_\delta} |d_{A_\delta}^{j_\delta}|$, A_δ — матрицы, аппроксимирующие вторые производные по переменным x_1, \dots, x_d , $d_{A_\delta}^{j_\delta}$ — их собственные числа.

Доказательство. Итерацию метода установления представим как $\widehat{U} = S(U)$, где $S: \mathbb{R}^{\mathcal{N}} \rightarrow \mathbb{R}^{\mathcal{N}}$ — линейный оператор с максимальным собственным значением λ_{\max} , $\mathcal{N} = N_1 \times \dots \times N_d$. Отметим, что проведённый анализ значений спектра матриц A_δ позволяет гарантировать, что все собственные числа являются строго отрицательными, по крайней мере при $N_\delta = 2, \dots, 200$. Из этого обстоятельства с учётом (22), (23) несложно сделать следующие выводы:

- для простой регуляризации $\lambda_{\max} = (1 + \tau m)^{-1} \implies \forall \tau |\lambda_{\max}| < 1$, поскольку все матрицы A_δ невырождены ($m \neq 0$);
- для регуляризации Соболева $\lambda_{\max} = \frac{k_1 + k_2 M}{k_1 + (k_2 + \tau)m}$; $|\lambda_{\max}| \leq 1$ при $\tau \geq k_2(M/m - 1)$.

Таким образом, спектральный критерий устойчивости (критерий Неймана) при указанных ограничениях на шаг τ обеспечивает устойчивость предложенного метода в линейных задачах. \square

Схема поиска решения краевой задачи (1), (2).

1. **Инициализация.** Определяем начальные значения решения и его производных, присутствующих в задаче. В рамках этой работы указанные значения задавались нулевыми. Задаём количество узлов N_δ и считываем из базы данных (см. замечание 1) матрицы $A_\delta, R_{A_\delta}, R_{A_\delta}^{-1}, D_{A_\delta}$. В соответствии с (32) и теоремами о сходимости итерационного процесса (если таковые имеются) задаём параметры τ, k_1, k_2 и невязку ε_S .

2. **Шаг метода установления.** Используя выражения (26)–(28), (30), рассчитываем элементы массивов $H(U)$ и $G(U)$. По формулам (31) определяем значения элементов \widehat{V} и \widehat{U} и вычисляем приближённые значения производных решения, присутствующие в правой части $f(\mathbf{u}, \mathbf{x})$, по формулам вида (11), (16), (19).
3. **Проверка условий сходимости.** Если $\frac{\|\widehat{U} - U\|}{\|F(\widehat{U} + \mathcal{V})\|_\tau} \leq \varepsilon_S$, то $\widehat{U} + \mathcal{V}$ — массив, содержащий значения решения. В противном случае задаём $U = \widehat{U}$ и переходим к шагу 2. Здесь $\|\widehat{U}\|$ — максимальное по модулю значение элементов \widehat{U} .

Заметим, что в случае линейной задачи доказательство сходимости такого процесса удастся провести на основе теорем об аппроксимации и устойчивости (см. теоремы 1–3). Для каждого нелинейного уравнения обоснование сходимости также связано с анализом зависимости спектрального радиуса оператора перехода (см. теорему 3) от итерационных параметров. В большинстве случаев для прикладных задач строгие оценки получить не удастся, однако, как показывает вычислительный опыт автора, удачный выбор шага τ и параметров k_1, k_2 позволяет не только добиться сходимости при решении нелинейных задач с малыми параметрами, но и уменьшить число итераций в десятки и даже сотни раз. Спектральный портрет аппроксимирующих матриц при этом играет ключевую роль [10, 11]. Интересные соображения по этому поводу читатель может найти также в [12].

Теорема 4. *Положим для простоты, что $N = K$ при $d = 2$, а для произвольных размерностей d задачи $N_1 = N_2 = \dots = N_d = N$. Описанный метод требует порядка $p((2d + \sigma(d + 1))N^{d+1} + O(N^d))$ операций и $s(n_1N^2 + N + \{\sigma(d + 1) + 5\}N^d + O(N^{d-1}))$ байт оперативной памяти, где σ — количество различных производных от неизвестной функции, присутствующих в уравнении (кратные производные считаются за одну); p — количество итераций метода установления; s — объём памяти в байтах, занимаемый одним действительным числом; $n_1 = 4$, если задача включает смешанные или первые производные, $n_1 = 3$ в противном случае.*

Доказательство. Поскольку арифметические операции, стоящие в правой части $f(\mathbf{u}, \mathbf{x})$, в рамках данного метода “превращаются” в соответствующие операции с N^d элементами массива решения и его производных, то таких операций насчитывается $O(N^d)$. Основные затраты времени при вычислении $H(U)$ состоят в операциях по расчёту производных. Общее количество таких операций равно σN^{d+1} . Остальные операции включают вычисление $G(U)$ по формулам (30) (dN^{d+1} операций), расчёт элементов массивов \widehat{V} ($O(N^d)$ операций) и \widehat{U} по формулам (31) (dN^{d+1} операций), а также добавочных массивов и их произведений на аппроксимирующие матрицы по формулам вида (19) ($dN^{d+1} + O(N^d)$ операций для каждой производной в уравнении). Суммируя все затраты, получаем $(2d + \sigma(d + 1))N^{d+1} + O(N^d)$ операций на каждой итерации.

В ходе расчётов в оперативной памяти должны храниться:

- 1) аппроксимирующие матрицы \mathfrak{A}_δ и A_δ , матрица спектрального разложения R_{A_δ} и обратная к ней (в сумме 3 или 4 матрицы размера $N \times N$ в зависимости от того, есть ли необходимость аппроксимировать первые производные массивом \mathfrak{A}_δ);
- 2) диагональная матрица D_A (суммарным объёмом N чисел);
- 3) массивы значений производных в выражениях правой части (всего таких производных $\sigma - d$), массивы решения на текущем и предыдущем временных слоях, массив правой части и массивы \widehat{V} и G , см. (30), (суммарный объём $s(\sigma - d + 5)N^d$);

- 4) добавочные массивы для неизвестной функции и её производных, реализующие граничные условия (в сумме $d(\sigma + 1)$ массивов);
- 5) прочие данные малого объёма (параметры метода, коэффициенты уравнения).

Суммируя все затраты памяти, получаем выражение из условия теоремы. \square

Замечание 3. Для любой размерности разработанный алгоритм требует приблизительно в N раз больше операций, чем экономичные схемы МКР. Тем не менее подчеркнём, что, благодаря высокой скорости сходимости используемых приближений на гладких решениях (см. теоремы 1, 2), число узлов N в алгоритме может быть значительно уменьшено, по сравнению с МКР, без потери точности. Это обеспечивает преимущества метода в задачах с гладкими решениями (см. пункт 3).

Подчеркнём также преимущества разработанного метода перед КСМК, которую традиционно используют при реализации спектральных методов [2]. Такая схема приводит к системе линейных уравнений $\tilde{A}\vec{x} = \vec{b}$ с разреженной матрицей, где \vec{x}, \vec{b} — векторы размера $N_1 \times \dots \times N_d$, содержащие значения решения $\hat{u}(\mathbf{x})$ и значения правой части $u(\mathbf{x}) - \tau f(\mathbf{u}, \mathbf{x})$ в узлах коллокации. Для решения такой системы потребуется $O(N^{3d})$ операций при $N_1 = N_2 = \dots = N_d = N$. В соответствии с теоремой 4, предложенная схема построения и решения матричных и тензорных уравнений Сильвестра (см. рис. 3) позволяет снизить количество операций в $O(N^{3d-d-1})$ раз, что уже в случае $d = 2$ даёт существенное сокращение числа операций.

3. Тестовые численные эксперименты

Разработанный метод был реализован на языке программирования JAVA с использованием современных концепций объектно-ориентированного анализа проектирования и программирования [13]. Все расчёты проводились на ЭВМ Intel Core i5-3330 CPU 3.00 GHz, DIMM DDR3 1600 MHz 8 Gb.

Рассмотрим тестовую задачу

$$\frac{d^2 \mathbf{u}}{dx^2} = \mathbf{u}^2 e^{-x}, \quad \mathbf{u}(-1) = e^{-1}, \quad \mathbf{u}(1) = e. \quad (33)$$

Её точное решение — целая функция $\mathbf{u}_{\text{ex}}(x) = e^x$. С помощью описанного метода найдены приближённые значения решения \mathbf{u}_j в узлах интерполяции $x_j, j = 1, \dots, N$. Пусть

$$\mathbf{u}_{\text{ex},j} = \mathbf{u}_{\text{ex}}(x_j), \quad \varepsilon_{\text{spec}} = \mathcal{M}^{-1} \max_{j=1, \dots, N} |\mathbf{u}_{\text{ex},j} - \mathbf{u}_j|, \quad \mathcal{M} = \max_{j=1, \dots, N} \mathbf{u}_{\text{ex},j}. \quad (34)$$

На рис. 4а сплошной линией изображена зависимость $\log_{10} \varepsilon_{\text{spec}}$ от количества узлов коллокации N . Из графика видно, что порядок аппроксимации решения является переменным, он стремительно растёт с ростом N . Данное обстоятельство полностью согласуется с оценкой погрешности использованной полиномиальной интерполяции (см. теорему 1):

$$\|\mathbf{u}_{\text{ex}}(x) - \mathbf{P}_N(\mathbf{u}, x)\| \leq \frac{\|\mathbf{u}_{\text{ex}}^{(N+2)}\|}{(N+2)!} 2^{1-N} = \frac{2^{1-N} e}{(N+2)!} = \varepsilon_{\text{est}}. \quad (35)$$

На рис. 4а пунктирной линией показана зависимость от N логарифма правой части оценки (35). Отметим, что сплошная и пунктирная линии имеют одинаковый угол наклона

и выпуклость, что свидетельствует о том, что асимптотика погрешности построенного метода с высокой точностью совпадает с асимптотикой полиномиального приближения функции u_{ex} . Данное обстоятельство свидетельствует о высокой вычислительной устойчивости предложенной схемы метода коллокаций.

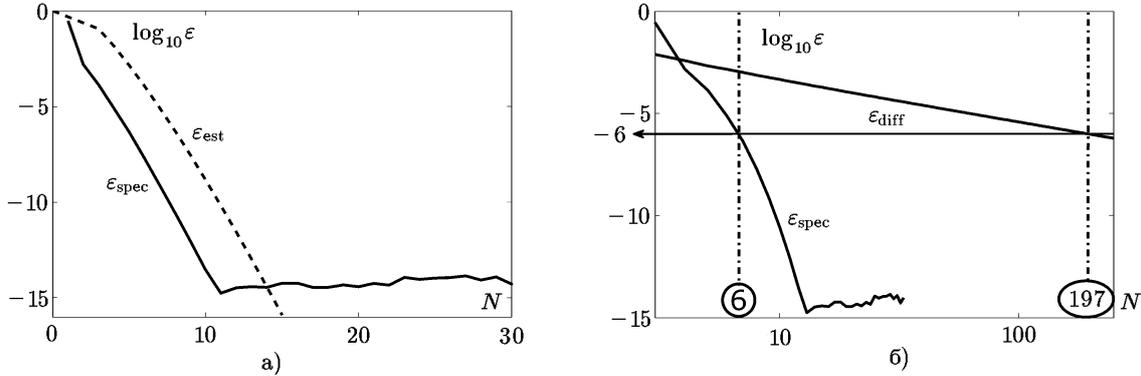


Рис. 4. Зависимость погрешности численных решений (33) от N в логарифмической шкале: а) погрешность ϵ_{spec} ; б) сравнение погрешностей ϵ_{diff} (МКР) и ϵ_{spec} (НМ)

Решение задачи (33) найдено также с помощью МКР с использованием трёхточечной аппроксимации второй производной и метода прогонки. Пусть по условию задачи требуется найти решение с точностью до шести значащих цифр (не менее). Для достижения точности $\epsilon = 10^{-6}$ (рис. 4б) МКР требуется построить сетку из $N = 197$ узлов, в то время как разработанному методу достаточно $N = 6$ узлов. Аналогичные результаты имеют место в случае задач более высоких размерностей:

$$\frac{d^2 u}{dx^2} + \frac{d^2 u}{dy^2} = 2u^2 e^{-x-y}, \quad u(\pm 1, y) = e^{y \pm 1}, \quad u(x, \pm 1) = e^{x \pm 1}, \quad u_{ex}(x, y) = e^{x+y}; \quad (36)$$

$$\frac{d^2 u}{dx^2} + \frac{d^2 u}{dy^2} + \frac{d^2 u}{dz^2} = 3u^2 e^{-x-y-z}, \quad u(\pm 1, y, z) = e^{y+z \pm 1}, \quad u(x, \pm 1, z) = e^{x+z \pm 1}, \\ u(x, y, \pm 1) = e^{x+y \pm 1}, \quad u_{ex}(x, y, z) = e^{x+y+z}. \quad (37)$$

С целью сравнения вычислительных затрат найдём решение нелинейных задач (36), (37) с помощью хорошо известного экономичного метода переменных направлений на основе МКР. Опишем кратко его суть на примере двумерного уравнения из (1). Трёхмерное уравнение решается аналогично с добавлением ещё одного (3-го) направления. Введём в области $D = [-1, 1]^2$ равномерную сетку с узлами (x_j, y_k) , $x_j = 2j/N - 1$, $y_k = 2k/K - 1$, $j = 0, \dots, N$, $k = 0, \dots, K$, с шагами $h_x = 2/N$, $h_y = 2/K$. Обозначим $u_{jk} = u(x_j, y_k)$, разностные аппроксимации вторых производных u_{xx} , u_{yy} запишем как

$$\Lambda_x u_{jk} = \frac{u_{j+1k} - 2u_{jk} + u_{j-1k}}{h_x^2}, \quad \Lambda_y u_{jk} = \frac{u_{jk+1} - 2u_{jk} + u_{jk-1}}{h_y^2}. \quad (38)$$

Для сохранения второго порядка аппроксимации производных на границе дополним (38) выражениями

$$\Lambda_x u_{0k} = \frac{2u_{0k} - 5u_{1k} + 4u_{2k} - u_{3k}}{h_x^2}, \quad \Lambda_x u_{Nk} = \frac{2u_{Nk} - 5u_{N-1k} + 4u_{N-2k} - u_{N-3k}}{h_x^2}$$

и аналогичными выражениями для $\Lambda_y u_{j0}$, $\Lambda_y u_{jK}$.

Применяя к двумерному уравнению из (1) простую регуляризацию B_t с шагом τ и описанную аппроксимацию, получаем

$$\frac{\bar{\mathbf{u}}_{jk} - \mathbf{u}_{jk}}{\tau} = \Lambda_x \bar{\mathbf{u}}_{jk} + \Lambda_y \mathbf{u}_{jk} + f(\mathbf{u}_{jk}, x_j, y_k), \quad \frac{\hat{\mathbf{u}}_{jk} - \bar{\mathbf{u}}_{jk}}{\tau} = \Lambda_x \bar{\mathbf{u}}_{jk} + \Lambda_y \hat{\mathbf{u}}_{jk} + f(\bar{\mathbf{u}}_{jk}, x_j, y_k),$$

где \mathbf{u}_{jk} , $\bar{\mathbf{u}}_{jk}$, $\hat{\mathbf{u}}_{jk}$ — решения на предыдущем, промежуточном и следующем временных слоях схемы соответственно. Критерий остановки итераций по времени с невязкой ε_S выглядит аналогично (24). Пусть I — тождественный сеточный оператор. Получаем двухшаговую схему

$$\begin{aligned} (I - \tau \Lambda_x) \bar{\mathbf{u}}_{jk} &= (I + \tau \Lambda_y) \mathbf{u}_{jk} + \tau f(\mathbf{u}_{jk}, x_j, y_k), \\ (I - \tau \Lambda_y) \hat{\mathbf{u}}_{jk} &= (I + \tau \Lambda_x) \bar{\mathbf{u}}_{jk} + \tau f(\bar{\mathbf{u}}_{jk}, x_j, y_k), \end{aligned} \quad (39)$$

соотношения которой решаются последовательно методом прогонки, требующим $8N$ и $8K$ операций для реализации каждого из двух шагов. Для описанного алгоритма в линейном случае доказаны теоремы аппроксимации, устойчивости и сходимости, а также известны оптимальные величины шага τ [14, с. 391–394]. При решении нелинейных задач (36), (37) шаг, обеспечивающий скорейшую сходимость, был подобран в экспериментах. Для реализации алгоритма и проведения расчётов использованы те же язык программирования и ЭВМ, что и для НМ.

В таблицах 2 и 3, в соответствии с теоремой 4, приведены оценки вычислительных затрат МКР и НМ на одном шаге по времени в задачах (33), (36), (37) при $N = K = M$, необходимые для вычисления решения с погрешностью не более 10^{-6} ($N = 6$ для НМ, $N = 197$, $\sigma = d$ для МКР). При оценке объёмов памяти МКР учитывались размеры массивов для хранения решения на всех промежуточных шагах схемы переменных направлений и массивов для хранения значений вторых производных, стоящих в правых частях (39). Отметим, что конкретные числа, приведённые в этих таблицах, являются приближёнными и дают лишь качественные характеристики. Более точные данные, полученные при проведении расчётов, даны в табл. 3. В ней приведены параметры и результаты тестовых решений задач (33), (36), (37): шаг τ , количество узлов по одному направлению $N = (K = M)$, количество итераций N_{it} , время выполнения программы T , приблизительный объём использованной памяти N_M , значение относительной погрешности ε_{ex} . Подчеркнём, что оба алгоритма были запущены с параметрами, обеспечивающими минимум затрат памяти и машинного времени (значения параметров были найдены экспериментально). Конкретно, невязка установления НМ для решения 1D–3D-задач $\varepsilon_S = 10^{-6}$, этого было достаточно, чтобы получить приближённое решение с погрешностью $\varepsilon_{spec} < 10^{-6}$. МКР для этого потребовалось задать $\varepsilon_S = 10^{-8}$. Шаги по времени НМ в 1D–3D-случаях были заданы как $\tau = 0.5, 0.24, 0.172$; шаги по времени МКР — $\tau = 0.5, 0.006, 0.00013$ соответственно.

Таблица 2. Оценка числа операций на каждой итерации, необходимых МКР и НМ для достижения точности 10^{-6}

Размерность задачи	Оценка числа операций		Число операций		Преимущество НМ
	МКР	НМ	МКР	НМ	
$d = 1$	$8N$	$4N^2$	1576	144	в 11 раз
$d = 2$	$16N^2$	$10N^3$	620944	2160	в 287 раз
$d = 3$	$24N^3$	$18N^4$	$1.835 \cdot 10^8$	23328	в 7866 раз

Таблица 3. Оценка объёмов памяти, необходимых МКР и НМ для достижения точности 10^{-6}

Размерность задачи	Оценка объёма памяти	Объём памяти, необходимый		Преимущество НМ
		МКР	НМ	
$d = 1$	Ns	$3 \cdot 197s$	$(3 \cdot 36 + 6 + 7 \cdot 6)s$	в 4 раза
$d = 2$	N^2s	$4 \cdot 197^2s$	$(3 \cdot 36 + 6 + 11 \cdot 36)s$	в 304 раза
$d = 3$	N^3s	$5 \cdot 197^3s$	$(3 \cdot 36 + 6 + 17 \cdot 216)s$	в 10097 раз

Таблица 4. Параметры и результаты МКР и НМ, полученные при решении задач (33), (36), (37) с относительной погрешностью не выше 10^{-6}

d	Схема	τ	N	N_{it}	$T, \text{мс}$	$N_M, \text{Мб}$	ε_{ex}
1	МКР	1.0	197	5	10	0.01	$9.98 \cdot 10^{-7}$
	НМ	0.5	6	5	4	0.00137	$4.54 \cdot 10^{-7}$
	Преимущества НМ			—	в 2.5 раза	в 7.3 раза	в 2.2 раза
2	МКР	0.006	149	310	1914	0.179	$9.98 \cdot 10^{-7}$
	НМ	0.24	6	8	13	0.00179	$1.73 \cdot 10^{-7}$
	Преимущества НМ			в 39 раз	в 147 раз	в 100 раз	в 5.8 раза
3	МКР	0.00013	119	3 913	3 009 965	141.4	$9.9 \cdot 10^{-7}$
	НМ	0.172	6	14	37	0.019	$1.52 \cdot 10^{-7}$
	Преимущества НМ			в 280 раз	в 81 350 раз	в 7 442 раз	в 6.5 раза

Отметим, что оценки, приведённые в табл. 2, 3, отличаются от данных реальных экспериментов (табл. 4): количество узлов по одному направлению, необходимых МКР для достижения точности 10^{-6} , падает с ростом размерности, в то время как количество итераций стремительно растёт. Кроме того, оценки в табл. 2, 3 не учитывают работу вспомогательных служб программной среды, что особенно отразилось на характеристиках НМ (в случае, когда T и N_M близки к нулю, вклад таких служб превышает вклад от работы алгоритма).

Результаты работы алгоритма в более сложных задачах, имеющих непосредственное отношение к приложениям, читатель может найти в статьях [15–17].

Заключение

В заключение подчеркнём, что разработанный алгоритм применим к решению задач сколь угодно высокой размерности d только в теории. На практике даже при использовании современного суперкомпьютера решать задачи размерности порядка 10-ти и выше не представляется возможным из-за экспоненциального роста затрат памяти с ростом размерности. Данный феномен, известный также как “проклятие размерности”, является фундаментальной проблемой теории приближения, см. [4, гл. 4, § 4, п. 2].

Заметим также, что в соответствии с оценками скорости сходимости спектральных методов на основе полиномов Чебышёва [1, 2] преимущества разработанного метода, которые обсуждались выше, имеют место лишь для функций высокой гладкости. Для функций с ограниченным порядком гладкости и функций, имеющих особенности в комплексной плоскости, расположенные близко к области задачи, скорость сходимости разработанного алгоритма снижается. Пути решения данной проблемы автор видит в использовании современных подходов построения дробно-рациональных аппроксимаций [18].

Литература

1. **Trefethen L.N.** Spectral Methods in Matlab. — Philadelphia: SIAM, 2000.
2. **Boyd J.** Chebyshev and Fourier Spectral Methods. Second ed. — Mineola, New York: DOVER Publication Inc., 2000.
3. **Семисалов Б.В.** Нелокальный алгоритм поиска решений уравнения Пуассона и его приложения // Журн. вычисл. матем. и мат. физики. — 2014. — Т. 54, № 7. — С. 1110–1135.
4. **Бабенко К.И.** Основы численного анализа. — М.: Наука. Гл. ред. физ.-мат. лит., 1986.
5. **Gottlieb D., Hussaini M.Y., Orszag S.A.** Theory and applications of spectral methods // Spectral Methods for Partial Differential Equations / R.G. Voigt, D. Gottlieb and M.Y. Hussaini. — Philadelphia: SIAM, 1984. — P. 1–54.
6. **Бахвалов Н.С., Жидков Н.П., Кобельков Г.М.** Численные методы. 6-е изд. — М.: БИНОМ. Лаборатория знаний, 2008.
7. **Rump S.M.** Verification methods: Rigorous results using floating-point arithmetic // Acta Numerica. — 2010. — Vol. 19. — P. 287–449.
8. **Rump S.M.** INTLAB — INTerval LABoratory // Developments in Reliable Computing / T. Csendes. — Kluwer Academic Publishers, 1999. — P. 77–104.
9. **Треногин В.А.** Функциональный анализ. — М.: Наука, 1980.
10. **Белов А.А., Калиткин Н.Н.** Эволюционная факторизация и сверхбыстрый счёт на установление // Мат. моделирование. — 2014. — Т. 26, № 9. — С. 47–64. Перевод: Belov A.A., Kalitkin N.N. Evolutionary factorization and superfast relaxation count // Mathematical Models and Computer Simulations. — 2015. — Vol. 7, № 2. — P. 103–116.
11. **Коновалов А.Н.** Введение в вычислительные методы линейной алгебры. — Новосибирск: Наука. Сиб. отд-ние, 1993.
12. **Белых В.Н.** Особенности реализации ненасыщаемого численного метода для внешней осесимметричной задачи Неймана // Сиб. мат. журнал. — 2013. — Т. 54, № 6. — С. 1237–1249. Перевод: Belykh V.N. Particular features of implementation of an unsaturated numerical method for the exterior axisymmetric Neumann problem // Siberian Mathematical Journal. — 2013. — Vol. 54. — P. 984–993.
13. **Семисалов Б.В.** Программа поиска решений краевых задач для уравнений в частных производных с высокой точностью и малыми вычислительными затратами “Нелокальный метод без насыщения”. — Свидетельство о государственной регистрации программы для ЭВМ № 2015615527 от 20 мая 2015 г. — URL: <http://www1.fips.ru/Archive/EVM/2015/2015.06.20/DOC/RUNW/000/002/015/615/527/document.pdf>.
14. **Калиткин Н.Н.** Численные методы. — М.: Наука, 1978.
15. **Блохин А.М., Семисалов Б.В.** Расчет стационарных неизотермических МГД течений полимерной жидкости в каналах с внутренними нагревательными элементами // Сиб. журн. индустр. математики. — 2020. — Т. 23, № 2. — С. 17–40. Перевод: Blokhin A.M., Semisalov B.V. Simulation of the stationary nonisothermal MHD flows of polymeric fluids in channels with interior heating elements // Journal of Applied and Industrial Mathematics. — 2020. — Vol. 14. — P. 222–241.
16. **Golushko S.K., Semisalov B.V.** Calculation and design of lattice cylindrical shells manufactured of unidirectional CFRPs // J. Phys.: Conf. Series. — 2017. — Vol. 894. — Art. 012022. — URL: <http://iopscience.iop.org/article/10.1088/1742-6596/894/1/012022>.
17. **Блохин А.М., Круглова Е.А., Семисалов Б.В.** Оценка двух компонент погрешности численного решения задачи о неизотермическом течении полимерных растворов между двумя соосными цилиндрами // Журн. вычисл. матем. и мат. физики. — 2018. — Т. 58,

- № 7. — С. 1099–1115. Перевод: Blokhin A.M., Kruglova E.A., Semisalov B.V. Estimation of two error components in the numerical solution to the problem of nonisothermal flow of polymer fluid between two coaxial cylinders // Computational Mathematics and Mathematical Physics. — 2018. — Vol. 58. — P. 1099–1115.
18. **Tee T.W., Trefethen L.N.** A rational spectral collocation method with adaptively transformed Chebyshev grid points // SIAM J. Scientific Computing. — 2006. — Vol. 28, № 5. — P. 1798–1811.

Поступила в редакцию 23 ноября 2020 г.

После исправления 8 апреля 2021 г.

Принята к печати 5 октября 2021 г.

Литература в транслитерации

1. **Trefethen L.N.** Spectral Methods in Matlab. — Philadelphia: SIAM, 2000.
2. **Boyd J.** Chebyshev and Fourier Spectral Methods. Second ed. — Mineola, New York: DOVER Publication Inc., 2000.
3. **Semisalov B.V.** Nelokal'nyi algoritm poiska reshenii uravneniya Puassona i ego prilozheniya // Zhurn. vychisl. matem. i mat. fiziki. — 2014. — Т. 54, № 7. — С. 1110–1135.
4. **Babenko K.I.** Osnovy chislennogo analiza. — М.: Nauka. Gl. red. fiz.-mat. lit., 1986.
5. **Gottlieb D., Hussaini M.Y., Orszag S.A.** Theory and applications of spectral methods // Spectral Methods for Partial Differential Equations / R.G. Voigt, D. Gottlieb and M.Y. Hussaini. — Philadelphia: SIAM, 1984. — P. 1–54.
6. **Bakhvalov N.S., Zhidkov N.P., Kobel'kov G.M.** Chislennyye metody. 6-e izd. — М.: BINOM. Laboratoriya znaniy, 2008.
7. **Rump S.M.** Verification methods: Rigorous results using floating-point arithmetic // Acta Numerica. — 2010. — Vol. 19. — P. 287–449.
8. **Rump S.M.** INTLAB — INTerval LABoratory // Developments in Reliable Computing / T. Csendes. — Kluwer Academic Publishers, 1999. — P. 77–104.
9. **Trenogin V.A.** Funkcional'nyi analiz. — М.: Nauka, 1980.
10. **Belov A.A., Kalitkin N.N.** Evolyucionnaya faktorizatsiya i sverkhbystryi schet na ustanovlenie // Mat. modelirovanie. — 2014. — Т. 26, № 9. — С. 47–64. Перевод: Belov A.A., Kalitkin N.N. Evolutionary factorization and superfast relaxation count // Mathematical Models and Computer Simulations. — 2015. — Vol. 7, № 2. — P. 103–116.
11. **Konovalov A.N.** Vvedenie v vychislitel'nye metody lineinoi algebry. — Novosibirsk: Nauka. Sib. otd-nie, 1993.
12. **Belykh V.N.** Osobennosti realizatsii nenasyschaemogo chislennogo metoda dlya vneshnei osesimmetrichnoi zadachi Neimana // Sib. mat. zhurnal. — 2013. — Т. 54, № 6. — С. 1237–1249. Перевод: Belykh V.N. Particular features of implementation of an unsaturated numerical method for the exterior axisymmetric Neumann problem // Siberian Mathematical Journal. — 2013. — Vol. 54. — P. 984–993.
13. **Semisalov B.V.** Programma poiska reshenii kraevykh zadach dlya uravnenii v chastnykh proizvodnykh s vysokoi tochnost'yu i malymi vychislitel'nymi zatratami "Nelokal'nyi metod bez nasysheniya". — Svidetel'stvo o gosudarstvennoy registratsii programmy dlya EVM № 2015615527 ot 20 maya 2015 g. — URL: <http://www1.fips.ru/Archive/EVM/2015/2015.06.20/DOC/RUNW/000/002/015/615/527/document.pdf>.
14. **Kalitkin N.N.** Chislennyye metody. — М.: Nauka, 1978.

15. **Blokhin A.M., Semisalov B.V.** Raschet stacionarnykh neizotermicheskikh MGD techenii polimernoi zhidkosti v kanalah s vnutrennimi nagrevatel'nymi elementami // Sib. zhurn. industr. matematiki. — 2020. — T. 23, № 2. — S. 17–40.. Perevod: Blokhin A.M., Semisalov B.V. Simulation of the stationary nonisothermal MHD flows of polymeric fluids in channels with interior heating elements // Journal of Applied and Industrial Mathematics. — 2020. — Vol. 14. — P. 222–241.
16. **Golushko S.K., Semisalov B.V.** Calculation and design of lattice cylindrical shells manufactured of unidirectional CFRPs // J. Phys.: Conf. Series. — 2017. — Vol. 894. — Art. 012022. — URL: <http://iopscience.iop.org/article/10.1088/1742-6596/894/1/012022>.
17. **Blokhin A.M., Kruglova E.A., Semisalov B.V.** Ocenka dvukh komponent pogreshnosti chislennogo resheniya zadachi o neizotermicheskom techenii polimernykh rastvorov mezhdu dvumya soosnymi cilindrami // Zhurn. vychisl. matem. i mat. fiziki. — 2018. — T. 58, № 7. — S. 1099–1115. Perevod: Blokhin A.M., Kruglova E.A., Semisalov B.V. Estimation of two error components in the numerical solution to the problem of nonisothermal flow of polymer fluid between two coaxial cylinders // Computational Mathematics and Mathematical Physics. — 2018. — Vol. 58. — P. 1099–1115.
18. **Tee T.W., Trefethen L.N.** A rational spectral collocation method with adaptively transformed Chebyshev grid points // SIAM J. Scientific Computing. — 2006. — Vol. 28, № 5. — P. 1798–1811.

