

Метод переменных направлений для построения малорангового поэлементного приближения тензоров в каноническом формате

С. В. Морозов

Институт вычислительной математики имени Г. И. Марчука РАН,
Москва, Российская Федерация

ORCID: 0000-0002-5801-2130, e-mail: stanis-morozov@yandex.ru

Аннотация: Приближение тензоров в малопараметрическом формате — важная составляющая при решении многих задач математического моделирования и анализа данных. Одним из наиболее популярных форматов представлений тензоров является каноническое тензорное разложение. На сегодняшний день большинство алгоритмов осуществляет приближение тензоров в норме Фробениуса, в то время как для некоторых приложений могут быть полезны поэлементные приближения. В данной статье предлагается метод переменных направлений для получения малорангового приближения тензоров в каноническом формате в чебышевской норме. В результате экспериментального исследования в статье демонстрируется эффективность предложенной процедуры.

Ключевые слова: норма Чебышева, каноническое тензорное разложение, метод переменных направлений.

Благодарности: Работа выполнена за счет средств гранта Российского Научного Фонда 21–71–10072.

Для цитирования: Морозов С.В. Метод переменных направлений для построения малорангового поэлементного приближения тензоров в каноническом формате // Вычислительные методы и программирование. 2024. 25, № 3. 302–314. doi 10.26089/NumMet.v25r323.

Alternating minimization method for low-rank entrywise approximation of tensors in canonical polyadic format

Stanislav V. Morozov

Marchuk Institute of Numerical Mathematics of RAS,
Moscow, Russia

ORCID: 0000-0002-5801-2130, e-mail: stanis-morozov@yandex.ru

Abstract: The approximation of tensors in low-parametric format is an important component in many mathematical modelling and data analysis tasks. One of the most popular low-parametric representations for tensors is the canonical polyadic (CP) decomposition. Nowadays, most of the algorithms for CP approximation aim to construct the approximation in Frobenius norm, however, some applications require entrywise approximation. In this paper, we propose an alternating



minimization method to obtain low-rank approximation of tensors in the canonical polyadic format in the Chebyshev norm. Through an extensive evaluation, we demonstrate the effectiveness of the proposed algorithm.

Keywords: Chebyshev norm, canonical polyadic, alternating minimization.

Acknowledgements: This work was supported by the Russian Science Foundation, project 21–71–10072.

For citation: S. V. Morozov, “Alternating minimization method for low-rank entrywise approximation of tensors in canonical polyadic format,” Numerical Methods and Programming. 25 (3), 302–314 (2024). doi 10.26089/NumMet.v25r323.

1. Введение. Задачи обработки данных, представимых в тензорных форматах, встречаются повсеместно в современной науке. Примеры таких задач можно встретить в различных областях, например теории аппроксимации [1], механике сплошных сред [2], интегро-дифференциальных уравнениях [3], анализе данных [4] и многих других областях. Число элементов, которое содержит d -мерный тензор $T \in \mathbb{R}^{n_1 \times \dots \times n_d}$, равно $n_1 n_2 \dots n_d$, что делает невозможным их явное хранение и вычисления с ними даже для небольших значений размерности d . Для решения этой проблемы обычно используются малопараметрические представления тензоров. Наиболее популярными среди них являются каноническое тензорное разложение (canonical polyadic decomposition, CP) [5, 6], разложение Таккера [7, 8] и разложение в формате тензорного поезда (tensor-train decomposition, ТТ) [9].

Каноническое тензорное разложение определяется следующим образом. Пусть дан $T \in \mathbb{R}^{n_1 \times \dots \times n_d}$ — d -мерный массив. Тогда каноническое тензорное разложение — это представление тензора T в виде

$$T = \sum_{t=1}^r u_t^{(1)} \otimes \dots \otimes u_t^{(d)}, \quad (1)$$

где $u_t^{(j)} \in \mathbb{R}^{n_j}$ и \otimes обозначает операцию тензорного произведения. Каждое слагаемое вида $u^{(1)} \otimes \dots \otimes u^{(d)}$ называется *тензором ранга 1*. Когда r минимально в (1), это представление называется *каноническим тензорным разложением*, а число r называется *каноническим рангом* тензора T .

Каноническое тензорное разложение, по сравнению с другими форматами, обычно требует меньшего числа параметров для достижения требуемой точности приближения [10] и естественным образом возникает в различных приложениях, например при быстром умножении матриц [11], обработке сигналов [12] и машинном обучении [13]. Задача построения канонического тензорного разложения является NP-сложной [14], что делает исключительно трудной задачу построения аппроксимации тензоров в этом формате. Обычно для построения приближений в каноническом формате используются итерационные процедуры, например метод переменных наименьших квадратов (alternating least squares, ALS) [15].

На сегодняшний день большинство аппроксимационных алгоритмов приближают тензоры в норме Фробениуса, однако в некоторых приложениях требуется поэлементное приближение, т.е. такое, что ошибка аппроксимации для каждого элемента ограничена и мала. Целью данной статьи является построение алгоритма для нахождения малорангового приближения тензоров в чебышевской норме в каноническом формате. А именно, пусть $T \in \mathbb{R}^{n_1 \times \dots \times n_d}$ является тензором. Требуется решить задачу

$$\left\| T - \sum_{t=1}^r u_t^{(1)} \otimes \dots \otimes u_t^{(d)} \right\|_C \rightarrow \min_{u_t^{(1)}, \dots, u_t^{(d)}}, \quad (2)$$

где чебышевская норма тензора $X \in \mathbb{R}^{n_1 \times \dots \times n_d}$ определяется как

$$\|X\|_C = \max_{i_1, \dots, i_d} |x_{i_1, \dots, i_d}|.$$

Для решения задачи (2) в статье предлагается *метод переменных направлений*, который, подобно алгоритму ALS, фиксирует все факторы кроме одного и находит оптимальное значение незафиксированного фактора, минимизирующее чебышевскую норму ошибки. Эффективность предложенной процедуры демонстрируется при помощи обширного экспериментального исследования. Для простоты изложения все

результаты данной статьи приведены для трехмерных тензоров, хотя они могут быть естественным образом обобщены на d -мерный случай.

Оставшаяся часть статьи организована следующим образом. В разделе 2 приводится обзор недавних результатов о приближениях матриц и тензоров в норме Чебышева. В разделе 3 дается описание основных результатов для задачи о наилучшем равномерном приближении для векторов и приводится алгоритм решения этой задачи. В разделе 4 предлагается процедура метода переменных направлений и анализируются базовые свойства предложенного алгоритма. В разделе 5 проводится обширное численное исследование предложенной процедуры.

2. Обзор литературы. Аппроксимации в чебышевской норме обладают существенно иными свойствами в сравнении с унитарно-инвариантными нормами (такими как спектральная норма или норма Фробениуса). Хорошо известно, что точность малоранговой аппроксимации в унитарно-инвариантных нормах связана со скоростью убывания сингулярных чисел. Однако можно доказать [16], что для любой матрицы $X \in \mathbb{R}^{m \times n}$, где $m \geq n$, и любого $\varepsilon > 0$ существует матрица $Y \in \mathbb{R}^{m \times n}$ ранга r , не превосходящего $r = \lceil 72 \log(2n + 1)/\varepsilon^2 \rceil$, такая, что $\|X - Y\|_C \leq \varepsilon \|X\|_2$. Таким образом, для любой последовательности матриц с ограниченной спектральной нормой и возрастающими размерами ранг, требуемый для достижения заданной точности в чебышевской норме, растет логарифмически с размером матриц.

Похожие результаты известны для приближений в формате тензорного произведения. В [17] доказывается, что для любой последовательности d -мерных тензоров с растущими размерами и такими, что их ТТ-ядра в точном разложении ограничены, ТТ-ранги, требуемые для достижения заданной точности приближения, растут логарифмически с размером тензоров. Стоит отметить, что упомянутые результаты являются асимптотическими и могут оказаться бесполезными для матриц и тензоров небольшого размера. Кроме того, нам неизвестно о каких-либо результатах, касающихся точности приближения тензоров в каноническом формате в чебышевской норме.

Задача построения малоранговых приближений является достаточно трудной. Например, известно, что уже задача построения оптимального приближения ранга 1 для матриц является NP-сложной [18]. На сегодняшний день существуют два основных метода для решения этой задачи: метод переменных направлений и метод переменных проекций. Первый метод впервые был опубликован в [19], где рассматривается задача построения аппроксимации ранга 1 для матриц. Пусть $A \in \mathbb{R}^{m \times n}$. Задача построения приближения ранга 1 в чебышевской норме ставится следующим образом:

$$\|A - uv^T\|_C \rightarrow \min_{u \in \mathbb{R}^m, v \in \mathbb{R}^n} . \tag{3}$$

Для решения (3) выбирается случайный вектор $v^{(0)} \in \mathbb{R}^n$ и попеременно решаются задачи

$$u^{(i+1)} = \arg \min_{u \in \mathbb{R}^m} \|A - u(v^{(i)})^T\|_C ,$$

$$v^{(i+1)} = \arg \min_{v \in \mathbb{R}^n} \|A - u^{(i+1)}v^T\|_C$$

при $i = 0, 1, 2, \dots$. В [19] авторы исследуют свойства предельных точек метода переменных направлений и показывают, что все предельные точки являются локальными минимумами функционала $c(u, v) = \|A - uv^T\|_C$. В работе [20] метод переменных направлений обобщается на случай произвольного ранга, а в [21] авторы продолжают исследование свойств метода переменных направлений и предлагают алгоритм, способный гарантированно строить оптимальные приближения ранга 1 для матриц. В данной статье теория, изложенная в [19–21], расширяется на случай построения приближений тензоров в каноническом формате в чебышевской норме.

Другая техника построения малоранговых чебышевских аппроксимаций называется методом переменных проекций [17, 22, 23]. Идея метода состоит в попеременном проектировании на множество матриц или тензоров малого ранга (например, при помощи SVD для матриц или ТТ-SVD для тензоров в ТТ-формате) и на ε -шар в чебышевской норме с центром в приближаемых матрице или тензоре. Таким образом, метод “пытается” найти общую точку двух описанных множеств, если она существует. Величина ε может быть оценена при помощи бинарного поиска. Стоит отметить, что эта техника не гарантирует оптимальность приближения и не находит оптимальные приближения на практике.



3. О задаче наилучшего равномерного приближения для векторов. Для построения метода переменных направлений нам понадобится алгоритм решения задачи

$$\|Vu - a\|_\infty \rightarrow \min_{u \in \mathbb{R}^r}, \quad (4)$$

где $V \in \mathbb{R}^{n \times r}$ и $a \in \mathbb{R}^n$. В [20] был предложен обобщенный алгоритм Ремеза для решения задачи (4). В данной статье мы приведем этот алгоритм для полноты изложения. Кроме того, важным компонентом метода, предложенного в [20], является решение задачи (4) при $n = r + 1$, однако в [20] эта процедура не описана подробно, поэтому мы также приведем ее здесь. Для этого нам понадобятся некоторые предварительные сведения.

3.1. Предварительные сведения. Доказательства приведенных ниже утверждений могут быть найдены в [20], поэтому здесь мы приведем только необходимые определения и формулировки.

Пусть $V \in \mathbb{R}^{n \times r}$ является матрицей, где $n \geq r$, и S является упорядоченным множеством целых чисел $1 \leq i_1, i_2, \dots, i_k \leq n$. Будем обозначать через $V(S)$ подматрицу матрицы V , содержащую строки i_1, i_2, \dots, i_k . Аналогично, если $a \in \mathbb{R}^n$, будем обозначать через $a(S)$ подвектор вектора a , содержащий элементы i_1, i_2, \dots, i_k .

Ключевым понятием, связанным с решением задачи (4), являются *чебышевские матрицы*.

Определение 1. Матрица $V \in \mathbb{R}^{n \times r}$ при $n \geq r$ называется *чебышевской*, если все ее $r \times r$ подматрицы невырождены.

Важность этого определения состоит в том, что только для таких матриц V можно гарантировать единственность и непрерывность решения задачи (4) для любой правой части a (см. [20, Теорема 2, Теорема 3]).

Теорема 1. Пусть $V \in \mathbb{R}^{n \times r}$ является чебышевской матрицей и $a \in \mathbb{R}^n$. Тогда решение задачи (4) существует, единственно и непрерывно зависит от матрицы V и правой части a .

Одно из важных свойств решения задачи (4) может быть сформулировано следующим образом.

Лемма 1. Пусть $V \in \mathbb{R}^{n \times r}$, где $n > r$, является чебышевской, $a \in \mathbb{R}^n$ и $\hat{u} \in \mathbb{R}^r$ является решением (4). Пусть $w = a - V\hat{u}$. Тогда существуют по крайней мере $r + 1$ различных индексов $1 \leq i_1 < \dots < i_{r+1} \leq n$ таких, что

$$|w_{i_j}| = \|w\|_\infty, \quad j = 1, \dots, r + 1.$$

Доказательство. См. [20, Утверждение 2].

Одно из важных свойств задачи (4) связано с понятием *характеристического множества*. Пусть $J = (1, 2, \dots, n)$ и $J' \subset J$. Тогда обозначим

$$\mu(J') = \min_{u \in \mathbb{R}^r} \|a(J') - V(J')u\|_\infty.$$

Определение 2. Множество J' называется *характеристическим*, если $\mu(J) = \mu(J')$ и для любого собственного подмножества $J'' \subsetneq J'$ выполнено $\mu(J'') < \mu(J)$.

Теорема 2. Пусть $V \in \mathbb{R}^{n \times r}$, где $n > r$, и $a \in \mathbb{R}^n$ не принадлежит образу матрицы V . Тогда существует по крайней мере одно характеристическое множество, состоящее из не более чем $r + 1$ элемента. Более того, если матрица V чебышевская, то любое характеристическое множество состоит по крайней мере из $r + 1$ элемента.

Доказательство. См. [20, Теорема 6].

Следующая теорема дает удобный критерий того, что вектор является решением задачи (4).

Теорема 3. Пусть $V \in \mathbb{R}^{n \times r}$ и $a \in \mathbb{R}^n$. Пусть $\hat{u} \in \mathbb{R}^r$ и обозначим $w = a - V\hat{u}$ и

$$E = \{j \in \{1, 2, \dots, n\} : |w_j| = \|w\|_\infty\}.$$

Тогда \hat{u} является решением задачи (4) тогда и только тогда, когда существует ненулевой вектор $\delta \in \mathbb{R}^{|E|}$ с неотрицательными компонентами такой, что

$$V(E)^T \text{diag}(\text{sign } w(E))\delta = 0.$$

Доказательство. См. [20, Теорема 10].

3.2. О решении задачи размера $(r + 1) \times r$. Опишем алгоритм решения задачи (4) при $n = r + 1$. Рассмотрим следующую систему:

$$V^T s = 0.$$

Это система с $r + 1$ переменной и r уравнениями, которая имеет нетривиальное решение. Если матрица V является чебышевской, то размерность $\ker V^T$ равна 1. Базисный вектор ядра может быть найден при помощи QR-разложения для матрицы $V^T = QR$, где $Q \in \mathbb{R}^{(r+1) \times (r+1)}$, $R \in \mathbb{R}^{(r+1) \times r}$. Тогда последний столбец матрицы Q соответствует базисному вектору $\ker V^T$. Пусть $s = q_{r+1}$. Можно показать, что если матрица V чебышевская, то все компоненты вектора s не равны нулю. По теореме 3 вектор $\hat{u} \in \mathbb{R}^r$ является решением (4) тогда и только тогда, когда существует ненулевой вектор $\delta \in \mathbb{R}^{r+1}$ с неотрицательными компонентами такой, что

$$V^T \text{diag}(\text{sign } w)\delta = 0,$$

где $w = a - V\hat{u}$. Тогда

$$\delta_k = |s_k|, \quad \text{sign } w_k = e^{i\phi} \text{sign } s_k, \quad k = 1, 2, \dots, r + 1,$$

где $\phi \in \{0, \pi\}$. Таким образом, можно вычислить знаки невязки для оптимального приближения. Заметим также, что для оптимального решения должно быть выполнено $|w_k| = |w_j|$ при всех k и j (лемма 1). Тогда

$$w_k e^{i\phi} \text{sign } s_k = w_j e^{i\phi} \text{sign } s_j, \quad k = 1, 2, \dots, r + 1,$$

откуда

$$\hat{u}^T (v^j \text{sign } s_j \text{sign } s_k - v^k) = a_j \text{sign } s_j \text{sign } s_k - a_k, \quad k = 1, 2, \dots, r + 1. \quad (5)$$

При $k = j$ уравнение тривиально, поэтому мы получаем систему линейных уравнений размера $r \times r$. Решение системы (5) является решением (4).

Таким образом, мы получаем алгоритм решения задачи (4) с матрицей размера $(r + 1) \times r$ за $O(r^3)$ операций. Общая схема метода приведена в Алгоритме 1.

Алгоритм 1. Решение задачи наилучшего равномерного приближения размера $(r + 1) \times r$
 Algorithm 1. Solving the problem of the best uniform approximation of size $(r + 1) \times r$

- 1: **input** чебышевская матрица $V \in \mathbb{R}^{(r+1) \times r}$, вектор $a \in \mathbb{R}^{r+1}$
 - 2: **output** решение задачи наилучшего равномерного приближения $\hat{u} \in \mathbb{R}^r$
 - 3: $Q, R = \text{qr_decomposition}(V^T)$ // $Q \in \mathbb{R}^{(r+1) \times (r+1)}$, $R \in \mathbb{R}^{(r+1) \times r}$
 - 4: $s = \text{sign}(Q[:, r + 1])$
 - 5: $c = s_{r+1}$
 - 6: **for** $k = 1, 2, \dots, r + 1$ **do**
 - 7: $s_k = s_k / c$
 - 8: **end for**
 - 9: Построить матрицу $G \in \mathbb{R}^{r \times r}$ такую, что $g^j = v^{r+1} - s_j v^j$ // верхний индекс обозначает строку матрицы
 - 10: Построить вектор $b \in \mathbb{R}^r$ такой, что $b_j = a_{r+1} - s_j b_j$
 - 11: $\hat{u} = G^{-1}b$
-

3.3. О решении задачи произвольного размера. Алгоритм решения задачи (4) для произвольных n и r был впервые предложен в [20]. Его идея похожа на классический алгоритм Ремеза и состоит в том, чтобы итерационно построить характеристическое множество задачи. Алгоритм состоит из следующих шагов.

1. Выберем произвольное множество из $r + 1$ различных индексов строк матрицы V . Обозначим это множество через J_1 и положим $t = 1$.
2. Решим задачу наилучшего равномерного приближения для матрицы $V(J_t)$ и вектора $a(J_t)$. Обозначим решение через u_t .
3. Вычислим невязку $w_t = a - V u_t$ и найдем позицию \hat{j}_t максимального по модулю элемента в векторе w_t . Если $|(w_t)_{\hat{j}_t}| = \|w_t(J_t)\|_\infty$, то u_t является решением задачи по теореме 3.
4. Если $|(w_t)_{\hat{j}_t}| > \|w_t(J_t)\|_\infty$, то попробуем заменить каждый из элементов множества J_t на \hat{j}_t . Пусть $J_t = (j_1^t, j_2^t, \dots, j_{r+1}^t)$, и обозначим $J_t^k = (j_1^t, \dots, j_{k-1}^t, \hat{j}_t, j_{k+1}^t, \dots, j_{r+1}^t)$, где круглые скобки обозначают упорядоченное множество. Пусть u_t^k является решением задачи наилучшего равномерного при-



ближения для матрицы $V(J_t^k)$ и вектора $a(J_t^k)$. Обозначим $w_t^k = a(J_t^k) - V(J_t^k)u_t^k$. Можно показать (см. [20, Теорема 12]), что существует такое \hat{k} , что $\|w_t^{\hat{k}}\|_\infty > \|w_t(J_t)\|_\infty$.

5. $J_{t+1} = J_t^{\hat{k}}$, $t = t + 1$ и перейдем к шагу 2.

Описанный алгоритм строит решение задачи о наилучшем равномерном приближении (4) для произвольной чебышевской матрицы $V \in \mathbb{R}^{n \times r}$ и произвольного вектора $a \in \mathbb{R}^n$ за $O(I(nr + r^4))$ операций, где I — число итераций алгоритма [20].

4. Метод переменных направлений. Пусть $T \in \mathbb{R}^{m \times n \times k}$ является трехмерным тензором. Здесь и далее будем считать, что m , n и k строго больше 1. Наша цель состоит в том, чтобы построить мало-ранговое поэлементное приближение тензора T с каноническим рангом r , а именно:

$$T \approx \sum_{t=1}^r u_t \otimes v_t \otimes w_t. \quad (6)$$

Обозначим через $U \in \mathbb{R}^{m \times r}$, $V \in \mathbb{R}^{n \times r}$ и $W \in \mathbb{R}^{k \times r}$ матрицы, составленные из векторов u_t , v_t и w_t соответственно. Будем сокращенно записывать (6) как

$$T \approx U \otimes V \otimes W.$$

Пусть матрицы U и V известны и требуется найти матрицу W , минимизирующую чебышевскую норму ошибки:

$$W = \arg \min_{X \in \mathbb{R}^{k \times r}} \|T - U \otimes V \otimes X\|_C. \quad (7)$$

Заметим, что (7) может быть записано как

$$W = \arg \min_{X \in \mathbb{R}^{k \times r}} \|(U \odot V)X^T - T^{(1,2)}\|_C, \quad (8)$$

где $U \odot V = \begin{bmatrix} u_1 \otimes v_1 & \dots & u_r \otimes v_r \end{bmatrix} \in \mathbb{R}^{mn \times r}$ обозначает произведение Хатри–Рао матриц U и V , а матрица $T^{(1,2)} \in \mathbb{R}^{mn \times k}$ получена из тензора T путем склеивания первых двух размерностей. Обозначим через $T[:, :, l] \in \mathbb{R}^{m \times n}$ l -ю срезку вдоль третьей координаты тензора T . Тогда легко видеть, что задача (8) распадается на набор независимых подзадач

$$w^l = \arg \min_{x \in \mathbb{R}^r} \|(U \odot V)x - \text{vec}(T[:, :, l])\|_\infty, \quad l = 1, 2, \dots, k, \quad (9)$$

где w^l обозначает l -ю строку матрицы W , а $\text{vec}(\cdot)$ обозначает операцию вытягивания матрицы в вектор. Если матрица $U \odot V$ является чебышевской, то по теореме 1 можно однозначно определить функцию $\chi : \mathbb{R}^{m \times n \times k} \times \mathbb{R}^{m \times r} \times \mathbb{R}^{n \times r} \rightarrow \mathbb{R}^{k \times r}$ такую, что l -я строка $\chi(T, U, V)$ определяется как

$$\chi(T, U, V)^l = \arg \min_{x \in \mathbb{R}^r} \|(U \odot V)x - \text{vec}(T[:, :, l])\|_\infty, \quad l = 1, 2, \dots, k.$$

Функция χ задает оптимальное значение матрицы W при известных U и V . Аналогично обозначим через $T[i, :, :] \in \mathbb{R}^{n \times k}$ i -ую срезку тензора T вдоль первой координаты и через $T[:, j, :] \in \mathbb{R}^{m \times k}$ j -ю срезку вдоль второй координаты. Если матрица $U \odot W$ является чебышевской, мы можем определить функцию $\psi : \mathbb{R}^{m \times n \times k} \times \mathbb{R}^{m \times r} \times \mathbb{R}^{k \times r} \rightarrow \mathbb{R}^{n \times r}$ такую, что j -я строка $\psi(T, U, W)$ определяется как

$$\psi(T, U, W)^j = \arg \min_{x \in \mathbb{R}^r} \|(U \odot W)x - \text{vec}(T[:, j, :])\|_\infty, \quad j = 1, 2, \dots, n.$$

Наконец, если $V \odot W$ является чебышевской, мы можем определить функцию $\phi : \mathbb{R}^{m \times n \times k} \times \mathbb{R}^{n \times r} \times \mathbb{R}^{k \times r} \rightarrow \mathbb{R}^{m \times r}$ такую, что i -я строка $\phi(T, V, W)$ определяется как

$$\phi(T, V, W)^i = \arg \min_{x \in \mathbb{R}^r} \|(V \odot W)x - \text{vec}(T[i, :, :])\|_\infty, \quad i = 1, 2, \dots, m.$$

Заметим, что отображения ϕ , ψ и χ непрерывны в точках, где $V \odot W$, $U \odot W$ и $U \odot V$ соответственно являются чебышевскими.

Определение 3. Пусть $T \in \mathbb{R}^{m \times n \times k}$. Будем говорить, что тройка последовательностей чебышевских матриц $\{U^{(t)} \in \mathbb{R}^{m \times r}\}_{t \in \mathbb{N}}$, $\{V^{(t)} \in \mathbb{R}^{n \times r}\}_{t \in \mathbb{N}}$ и $\{W^{(t)} \in \mathbb{R}^{k \times r}\}_{t \in \mathbb{N}}$ получена методом переменных направлений для тензора T и пары начальных точек $(V^{(0)}, W^{(0)})$, где $V^{(0)} \in \mathbb{R}^{n \times r}$ и $W^{(0)} \in \mathbb{R}^{k \times r}$ являются чебышевскими матрицами, если

$$\begin{cases} U^{(t)} = \phi(T, V^{(t-1)}, W^{(t-1)}), \\ V^{(t)} = \psi(T, U^{(t)}, W^{(t-1)}), \\ W^{(t)} = \chi(T, U^{(t)}, V^{(t)}) \end{cases}$$

для любого $t \in \mathbb{N}$.

Заметим, что если матрица $U \odot V$ чебышевская, из этого не следует, что $\chi(T, U, V)$ также является чебышевской. Однако в проведенных численных экспериментах ситуации, когда матрицы перестают быть чебышевскими, достаточно редки и для большинства тензоров T и почти всех начальных точек $(V^{(0)}, W^{(0)})$ процедура, описанная в определении 3, порождает тройку последовательностей чебышевских матриц. Тем не менее при теоретических построениях нам требуется явно предполагать, что все построенные матрицы оказываются чебышевскими.

Сформулируем несколько базовых свойств, касающихся метода переменных направлений.

Лемма 2. Пусть $T \in \mathbb{R}^{m \times n \times k}$ и матрицы $V^{(0)} \in \mathbb{R}^{n \times r}$ и $W^{(0)} \in \mathbb{R}^{k \times r}$ являются чебышевскими. Пусть тройка последовательностей $\{U^{(t)} \in \mathbb{R}^{m \times r}\}_{t \in \mathbb{N}}$, $\{V^{(t)} \in \mathbb{R}^{n \times r}\}_{t \in \mathbb{N}}$ и $\{W^{(t)} \in \mathbb{R}^{k \times r}\}_{t \in \mathbb{N}}$ порождена методом переменных направлений для тензора T и пары начальных точек $(V^{(0)}, W^{(0)})$. Тогда выполнены следующие утверждения.

a) Верны неравенства

$$\begin{aligned} \|T - U^{(t)} \otimes V^{(t-1)} \otimes W^{(t-1)}\|_C &\geq \|T - U^{(t)} \otimes V^{(t)} \otimes W^{(t-1)}\|_C \geq \\ &\|T - U^{(t)} \otimes V^{(t)} \otimes W^{(t)}\|_C \geq \|T - U^{(t+1)} \otimes V^{(t)} \otimes W^{(t)}\|_C \end{aligned}$$

для всех $t \in \mathbb{N}$.

b) Если тройка последовательностей $\{\tilde{U}^{(t)}\}_{t \in \mathbb{N}}$, $\{\tilde{V}^{(t)}\}_{t \in \mathbb{N}}$ и $\{\tilde{W}^{(t)}\}_{t \in \mathbb{N}}$ получена методом переменных направлений для тензора T и пары начальных точек $(\alpha V^{(0)}, \beta W^{(0)})$, где $\alpha, \beta \neq 0$, то $\tilde{U}^{(t)} = 1/(\alpha\beta) U^{(t)}$, $\tilde{V}^{(t)} = \alpha V^{(t)}$, $\tilde{W}^{(t)} = \beta W^{(t)}$.

Доказательство. По построению ϕ

$$\inf_{U \in \mathbb{R}^{m \times r}} \|T - U \otimes V^{(t)} \otimes W^{(t)}\|_C = \|T - \phi(T, V^{(t)}, W^{(t)}) \otimes V^{(t)} \otimes W^{(t)}\|_C,$$

откуда имеем

$$\|T - U^{(t)} \otimes V^{(t)} \otimes W^{(t)}\|_C \geq \|T - U^{(t+1)} \otimes V^{(t)} \otimes W^{(t)}\|_C,$$

поскольку $U^{(t+1)} = \phi(T, V^{(t)}, W^{(t)})$. Оставшиеся неравенства в утверждении a) доказываются аналогично.

Утверждение b) следует из единственности решения задачи (9).

Пусть $V \in \mathbb{R}^{n \times r}$ и $W \in \mathbb{R}^{k \times r}$ являются чебышевскими матрицами и тройка последовательностей $\{U^{(t)}\}_{t \in \mathbb{N}}$, $\{V^{(t)}\}_{t \in \mathbb{N}}$ и $\{W^{(t)}\}_{t \in \mathbb{N}}$ получена методом переменных направлений для тензора T и пары начальных точек $V^{(0)} = V$ и $W^{(0)} = W$. Из леммы 2 a) следует, что последовательность $\|T - U^{(t)} \otimes V^{(t)} \otimes W^{(t)}\|_C$ не возрастает и, поскольку она состоит из неотрицательных чисел, сходится. Обозначим предел этой последовательности через $E(T, U, V)$. Следующая лемма содержит элементарные свойства этой функции.

Лемма 3. Пусть $T \in \mathbb{R}^{m \times n \times k}$ и матрицы $V \in \mathbb{R}^{n \times r}$ и $W \in \mathbb{R}^{k \times r}$ являются чебышевскими. Пусть также метод переменных направлений для тензора T и пары начальных точек $V^{(0)} = V$ и $W^{(0)} = W$ определен корректно (т.е. порождаемые матрицы являются чебышевскими). Тогда выполнены следующие утверждения.

a) $E(T, V, W) \geq 0$ и $E(T, V, W) = E(T, \alpha V, \beta W)$ при $\alpha, \beta \neq 0$.

b) $E(T, V, W) = E(T, \tilde{V}, W) = E(T, \tilde{V}, \tilde{W})$, где $\tilde{V} = \psi(T, \phi(T, V, W), W)$ и $\tilde{W} = \chi(T, \phi(T, V, W), \tilde{V})$.

c) Функция $E(T, V, W)$ полунепрерывна сверху по паре (V, W) .



Доказательство. Утверждения а) и б) сразу следуют из определения $E(T, V, W)$ и леммы 2. Полунепрерывность сверху имеет место, поскольку $E(T, U, V)$ является пределом невозрастающей последовательности непрерывных функций.

Итоговая процедура метода переменных направлений представлена в алгоритме 2. Заметим, что в алгоритме также используются перенормировки матриц, поскольку, согласно лемме 3, они не влияют на решение, но улучшают численную устойчивость.

Алгоритм 2. Метод переменных направлений
 Algorithm 2. Alternating minimization method

```

1: input тензор  $T \in \mathbb{R}^{m \times n \times k}$ , ранг  $r \geq 1$ , начальные точки  $V^{(0)} \in \mathbb{R}^{n \times r}$ ,  $W^{(0)} \in \mathbb{R}^{k \times r}$ 
2: output факторы приближения в каноническом формате  $\hat{U} \in \mathbb{R}^{m \times r}$ ,  $\hat{V} \in \mathbb{R}^{n \times r}$ ,  $\hat{W} \in \mathbb{R}^{k \times r}$ 
3:  $t = 1$ 
4: repeat
5:    $U^{(t)} = \phi(T, V^{(t-1)}, W^{(t-1)})$ 
6:    $V^{(t)} = \psi(T, U^{(t)}, W^{(t-1)})$ 
7:    $W^{(t)} = \chi(T, U^{(t)}, V^{(t)})$ 
8:    $C = \|U^{(t)}\|_C \|V^{(t)}\|_C \|W^{(t)}\|_C$ 
9:    $U^{(t)} = U^{(t)} / \|U^{(t)}\|_C \cdot C^{1/3}$ 
10:   $V^{(t)} = V^{(t)} / \|V^{(t)}\|_C \cdot C^{1/3}$ 
11:   $W^{(t)} = W^{(t)} / \|W^{(t)}\|_C \cdot C^{1/3}$ 
12:   $t = t + 1$ 
13: until не достигнута сходимость
14:  $\hat{U} = U^{(t-1)}$ ,  $\hat{V} = V^{(t-1)}$ ,  $\hat{W} = W^{(t-1)}$ 
    
```

5. Численные эксперименты. Построенный алгоритм не имеет гарантий оптимальности. Более того, так же как для известного алгоритма ALS [15], неизвестно даже о гарантиях сходимости предложенной итерационной процедуры. Однако во всех проведенных численных экспериментах метод переменных направлений сходится. Для того чтобы оценить эффективность предложенного метода, была проведена серия численных экспериментов.

5.1. Тензор Гильберта. Рассмотрим тензор

$$T = \left[\frac{1}{i + j + l} \right]_{i,j,l=1}^{256}.$$

В данном эксперименте строилась малоранговая аппроксимация тензора T в каноническом формате, используя метод переменных направлений (алгоритм 2) и метод переменных наименьших квадратов (ALS). В случае алгоритма ALS итерационная процедура запускалась с 20 случайных матриц $V^{(0)}$ и $W^{(0)}$ из стандартного нормального распределения. Для того чтобы сделать время работы метода приемлемым, число итераций было ограничено 1000. В случае метода переменных направлений итерационная процедура запускалась с тех же начальных точек, что и алгоритм ALS, а также с наилучшей точки, полученной в результате работы алгоритма ALS. Среди всех начальных точек для обоих методов была выбрана наилучшая. На рис. 1 приведена ошибка аппроксимации в чебышевской норме для различных рангов. На рис. 2 приведена ошибка аппроксимации в норме Фробениуса. Нетрудно заметить, что используемые методы качественно различны, а именно: метод переменных направлений значительно лучше приближает тензор в чебышевской норме, в то время как для алгоритма ALS значение ошибки меньше в норме Фробениуса.

5.2. Тензор, порожденный гладкой функцией. Рассмотрим тензор, порожденный значениями функции

$$f(x, y, z) = \cos(2\pi xyz)$$

на равномерной сетке в кубе $[0, 1]^3$, а именно:

$$T = \left[\cos \left(2\pi \cdot \frac{i-1}{n} \cdot \frac{j-1}{n} \cdot \frac{l-1}{n} \right) \right]_{i,j,l=1}^n.$$

В данном эксперименте было выбрано $n = 256$. Вся остальная конфигурация была выбрана так же, как в предыдущем эксперименте. На рис. 3 изображена ошибка в чебышевской норме для метода переменных направлений.

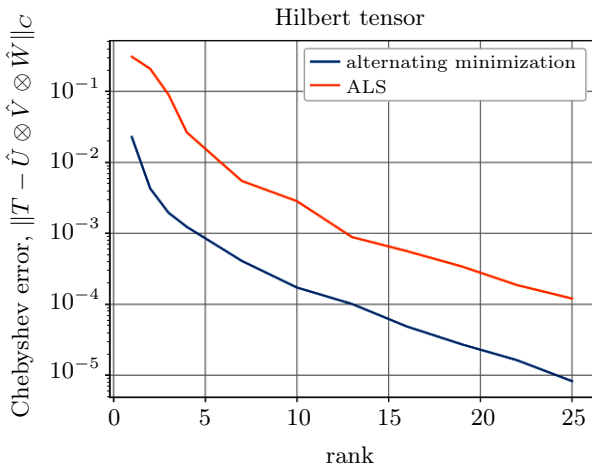


Рис. 1. Ошибка аппроксимации в чебышевской норме для тензора Гильберта

Fig. 1. Chebyshev approximation error for Hilbert tensor

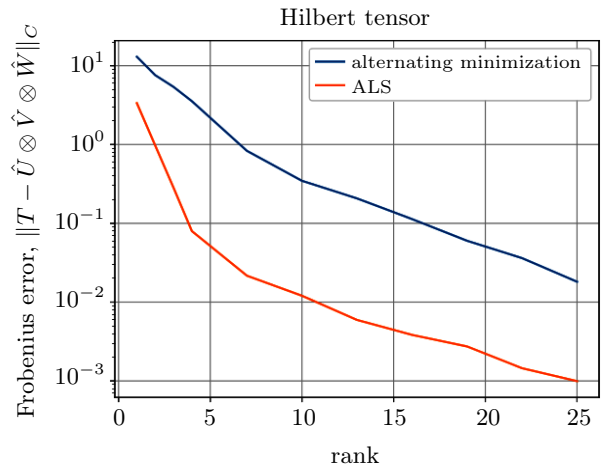


Рис. 2. Ошибка аппроксимации во фробениусовой норме для тензора Гильберта

Fig. 2. Frobenius approximation error for Hilbert tensor

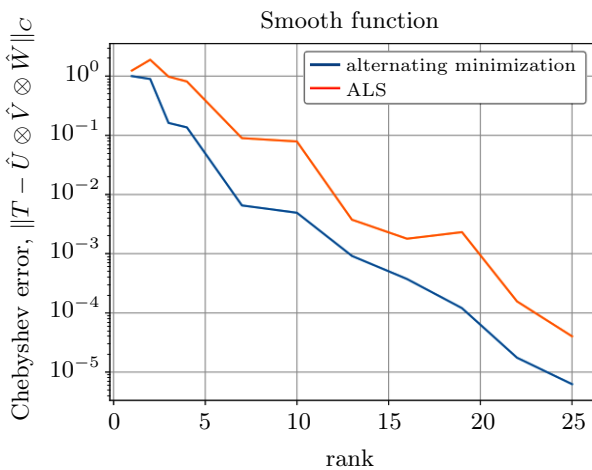


Рис. 3. Ошибка аппроксимации в чебышевской норме для тензора, порожденного гладкой функцией

Fig. 3. Chebyshev approximation error for tensor generated by smooth function

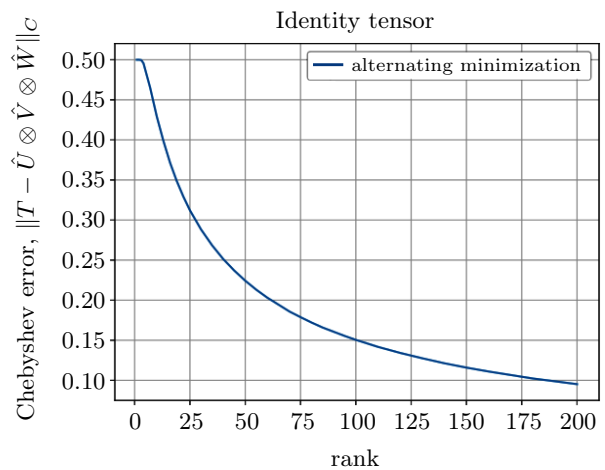


Рис. 4. Ошибка аппроксимации в чебышевской норме для единичного тензора

Fig. 4. Chebyshev approximation error for identity tensor

ных направлений и алгоритма ALS для различных рангов. Из эксперимента легко видеть, что ошибка в чебышевской норме снова значительно меньше для метода переменных направлений.

5.3. Единичный тензор. Рассмотрим тензор $T \in \mathbb{R}^{n \times n \times n}$, заданный формулой

$$t_{ijl} = \begin{cases} 1, & i = j = l, \\ 0, & \text{иначе.} \end{cases}$$

В данном эксперименте было выбрано $n = 256$. На рис. 4 приведена ошибка в чебышевской норме для метода переменных направлений. Стоит отметить, что для метода ALS для всех приведенных рангов ошибка равна 1. Проведенный эксперимент демонстрирует качественную разницу между чебышевской нормой и нормой Фробениуса, а именно: приведенный “единичный” тензор не имеет разумного малорангового приближения в норме Фробениуса, но может быть приближен поэлементно.



5.4. Восстановление тензора из равномерного шума. В данном эксперименте рассматривается задача восстановления малорангового тензора из равномерного шума. Пусть $U \in \mathbb{R}^{m \times r}$, $V \in \mathbb{R}^{n \times r}$ и $W \in \mathbb{R}^{k \times r}$. Рассмотрим тензор

$$T = U \otimes V \otimes W + N,$$

где $N \in \mathbb{R}^{m \times n \times k}$ является случайным тензором с независимыми элементами из равномерного распределения $\mathcal{U}(-\varepsilon, \varepsilon)$. В данном эксперименте варьировался уровень шума ε и для каждого ε было сгенерировано 10 случайных матриц U , V и W из стандартного нормального распределения и тензор N из равномерного распределения. Для каждого полученного тензора была запущена процедура восстановления малорангового тензора $U \otimes V \otimes W$ из тензора T при помощи метода переменных направлений и алгоритма ALS. Для каждой процедуры восстановления методы были запущены из 20 случайных точек и среди них был выбран результат, наилучшим образом приближающий тензор T . Пусть какой-либо из методов вернул факторы \hat{U} , \hat{V} и \hat{W} . Тогда ошибка восстановления определяется как

$$\|U \otimes V \otimes W - \hat{U} \otimes \hat{V} \otimes \hat{W}\|,$$

где норма выбиралась чебышевской или фробениусовой. Для каждого уровня шума результаты были усреднены по различным выборам тензора T . В данном эксперименте было выбрано $m = n = k = 128$ и $r = 12$. На рис. 5 приведена ошибка восстановления малорангового тензора для различных уровней шума в чебышевской норме, а на рис. 6 — в норме Фробениуса. Легко заметить, что метод переменных направлений восстанавливает малоранговый тензор из шума значительно лучше, чем алгоритм ALS, причем как в чебышевской норме, так и в норме Фробениуса.

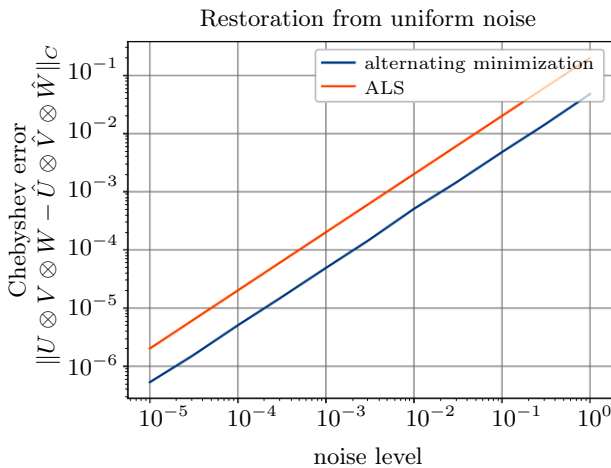


Рис. 5. Ошибка восстановления малорангового тензора из равномерного шума в чебышевской норме

Fig. 5. Chebyshev approximation error for recovery of low-rank tensor from uniform noise

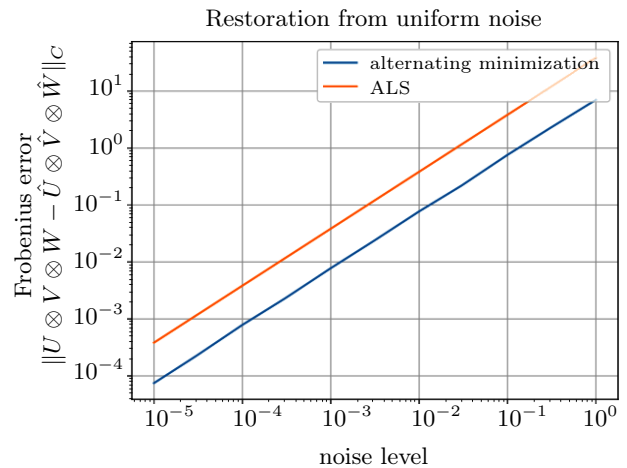


Рис. 6. Ошибка восстановления малорангового тензора из равномерного шума в фробениусовой норме

Fig. 6. Frobenius approximation error for recovery of low-rank tensor from uniform noise

5.5. Цветные изображения. Для визуализации разницы между малоранговыми аппроксимациями в чебышевской норме и норме Фробениуса был проведен эксперимент по построению малоранговых приближений для цветных картинок. Цветное изображение может быть представлено как тензор $T \in \mathbb{R}^{3 \times w \times h}$, где w и h соответствуют геометрическим размерам картинки. Значения в тензоре соответствуют вещественным числам от 0 до 1. В данном эксперименте были рассмотрены изображения с фиксированными геометрическими размерами 512×512 . На рис. 7 изображены примеры малорангового приближения цветных картинок в каноническом формате при помощи метода переменных направлений и алгоритма ALS. Каждая картинка на рис. 7 содержит ошибки приближения в чебышевской норме и норме Фробениуса. Можно заметить, что приближения, построенные при помощи алгоритма ALS, оказываются более размытыми, а приближения, построенные методом переменных направлений, являются более четкими, но имеют “дрожащую” структуру.

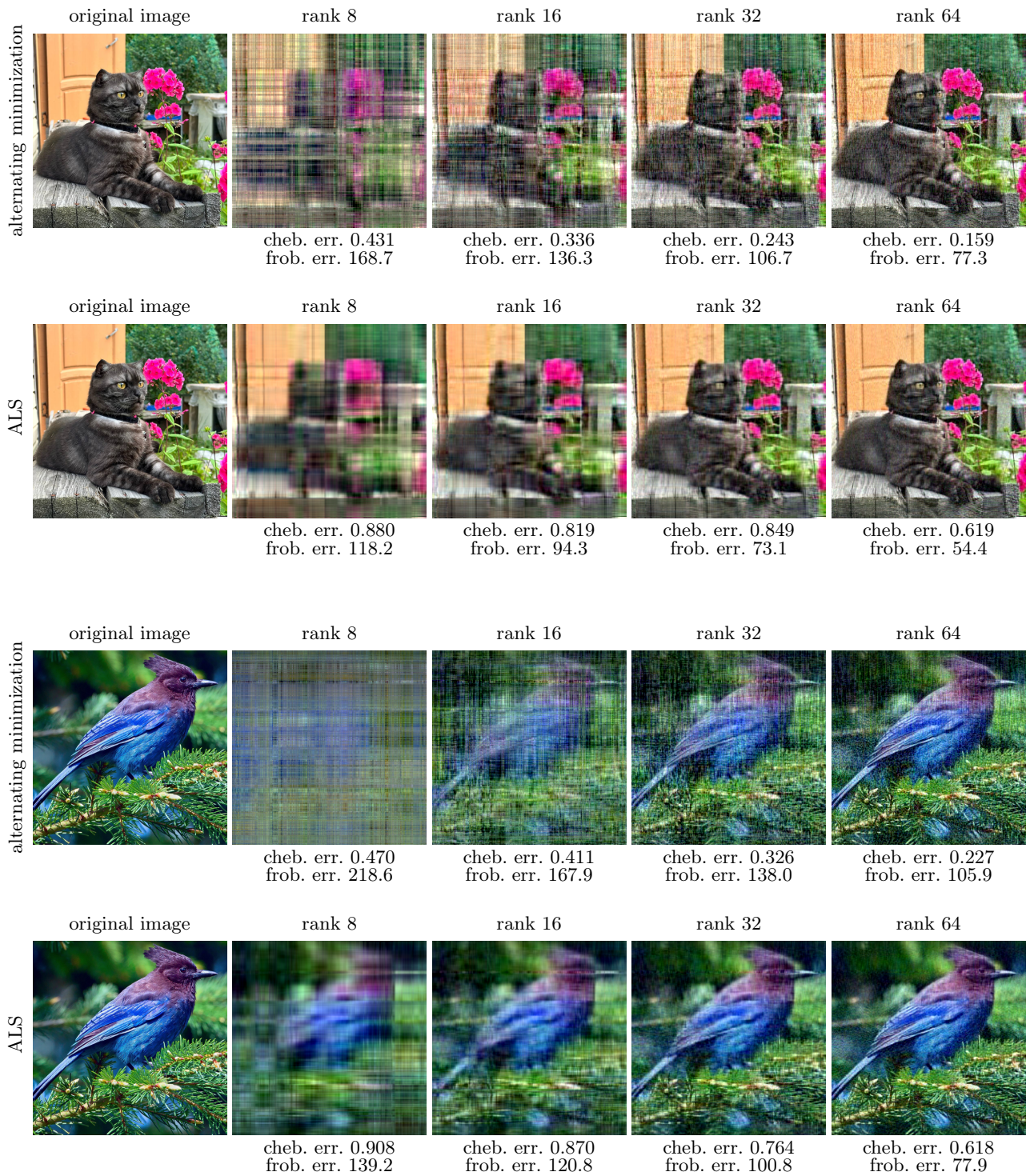


Рис. 7. Аппроксимация цветных картинок при помощи метода переменных направлений и алгоритма ALS

Fig. 7. Approximation of colour images with the alternating minimization method and ALS

6. Заключение. В данной статье был предложен метод переменных направлений для построения малоранговых приближений тензоров в каноническом формате в норме Чебышева. При помощи большого количества численных экспериментов была продемонстрирована эффективность предложенной процедуры, а также качественное различие между чебышевской нормой и нормой Фробениуса.



Список литературы

1. *Khoromskij B.N., Schwab C.* Tensor-structured Galerkin approximation of parametric and stochastic elliptic PDEs // *SIAM Journal on Scientific Computing*. 2011. **33**, N 1. 364–385. doi [10.1137/100785715](https://doi.org/10.1137/100785715).
2. *Eshelby J.D.* Energy relations and the energy-momentum tensor in continuum mechanics // *Fundamental contributions to the continuum theory of evolving phase interfaces in solids*. Berlin: Springer, 1999. 82–119.
3. *Matveev S.A., Zheltkov D.A., Tyrtysnikov E.E., Smirnov A.P.* Tensor train versus Monte Carlo for the multicomponent Smoluchowski coagulation equation // *Journal of Computational Physics*. 2016. **316**. 164–179. doi [10.1016/j.jcp.2016.04.025](https://doi.org/10.1016/j.jcp.2016.04.025).
4. *Lu H., Plataniotis K.N., Venetsanopoulos A.N.* A survey of multilinear subspace learning for tensor data // *Pattern Recognition*. 2011. **44**, N 7. 1540–1551. doi [10.1016/j.patcog.2011.01.004](https://doi.org/10.1016/j.patcog.2011.01.004).
5. *Hitchcock F.L.* Multiple invariants and generalized rank of a P-way matrix or tensor // *Journal of Mathematics and Physics*. 1928. **7**, N 1–4. 39–79. doi [10.1002/sapm19287139](https://doi.org/10.1002/sapm19287139).
6. *Harshman R.A.* Foundations of the PARAFAC procedure: models and conditions for an “explanatory” multimodal factor analysis. Vol. 16. Ann Arbor: UCLA Working Papers in Phonetics, 1970.
7. *Tucker L.R.* Some mathematical notes on three-mode factor analysis // *Psychometrika*. 1966. **31**, N 3. 279–311. doi [10.1007/BF02289464](https://doi.org/10.1007/BF02289464).
8. *De Lathauwer L., De Moor B., Vandewalle J.* A multilinear singular value decomposition // *SIAM Journal on Matrix Analysis and Applications*. 2000. **21**, N 4. 1253–1278. doi [10.1137/S0895479896305696](https://doi.org/10.1137/S0895479896305696).
9. *Oseledets I.V.* Tensor-train decomposition // *SIAM Journal on Scientific Computing*. 2011. **33**, N 5. 2295–2317. doi [10.1137/090752286](https://doi.org/10.1137/090752286).
10. *Shi T., Townsend A.* On the compressibility of tensors // *SIAM Journal on Matrix Analysis and Applications*. 2021. **42**, N 1. 275–298. doi [10.1137/20M1316639](https://doi.org/10.1137/20M1316639).
11. *Strassen V.* Gaussian elimination is not optimal // *Numerische Mathematik*. 1969. **13**, N 4. 354–356. doi [10.1007/BF02165411](https://doi.org/10.1007/BF02165411).
12. *Cichocki A., Mandic D., De Lathauwer L., et al.* Tensor decompositions for signal processing applications: from two-way to multiway component analysis // *IEEE Signal Processing Magazine*. 2015. **32**, N 2. 145–163. doi [10.1109/MSP.2013.2297439](https://doi.org/10.1109/MSP.2013.2297439).
13. *Lebedev V., Ganin Y., Rakhuba M., et al.* Speeding-up convolutional neural networks using fine-tuned CP-decomposition. <https://arxiv.org/pdf/1412.6553>. Cited August 29, 2024.
14. *Håstad J.* Tensor rank is NP-complete // *Journal of Algorithms*. 1990. **11**, N 4. 644–654. doi [10.1016/0196-6774\(90\)90014-6](https://doi.org/10.1016/0196-6774(90)90014-6).
15. *Mohlenkamp M.J.* Musings on multilinear fitting // *Linear Algebra and its Applications*. 2013. **438**, N 2. 834–852. doi [10.1016/j.laa.2011.04.019](https://doi.org/10.1016/j.laa.2011.04.019).
16. *Udell M., Townsend A.* Why are big data matrices approximately low rank? // *SIAM Journal on Mathematics of Data Science*. 2019. **1**, N 1. 144–160. doi [10.1137/18M1183480](https://doi.org/10.1137/18M1183480).
17. *Budzinskiy S.* Entrywise tensor-train approximation of large tensors via random embeddings. <https://arxiv.org/pdf/2403.11768>. Cited August 29, 2024.
18. *Gillis N., Shitov Y.* Low-rank matrix approximation in the infinity norm // *Linear Algebra and its Applications*. 2019. **581**. 367–382. doi [10.1016/j.laa.2019.07.017](https://doi.org/10.1016/j.laa.2019.07.017).
19. *Даугавет В.* О равномерном приближении функции двух переменных, заданной таблично, произведением функций одной переменной // *Журн. вычисл. матем. и матем. физ.* 1971. **11**, № 2. 289–303.
20. *Замарашкин Н.Л., Морозов С.В., Тыртышников Е.Е.* Об алгоритме наилучшего приближения матрицами малого ранга в норме Чебышёва // *Журн. вычисл. матем. и матем. физ.* 2022. **62**, № 5. 723–741. <https://elibrary.ru/item.asp?id=48506047>. Cited August 29, 2024.
21. *Morozov S., Smirnov M., Zamarashkin N.* On the optimal rank-1 approximation of matrices in the Chebyshev norm // *Linear Algebra and its Applications*. 2023. **679**. 4–29. doi [10.1016/j.laa.2023.09.007](https://doi.org/10.1016/j.laa.2023.09.007).
22. *Budzinskiy S.* Quasioptimal alternating projections and their use in low-rank approximation of matrices and tensors. <https://arxiv.org/pdf/2308.16097>. Cited August 29, 2024.
23. *Budzinskiy S.* On the distance to low-rank matrices in the maximum norm // *Linear Algebra and its Applications*. 2024. **688**. 44–58. doi [10.1016/j.laa.2024.02.012](https://doi.org/10.1016/j.laa.2024.02.012).

Поступила в редакцию
12 августа 2024 г.

Принята к публикации
18 августа 2024 г.

Информация об авторе

Станислав Викторович Морозов — магистр, мл. науч. сотр.; Институт вычислительной математики имени Г. И. Марчука РАН, улица Губкина, 8, 119333, Москва, Российская Федерация.

References

1. B. N. Khoromskij and C. Schwab, “Tensor-Structured Galerkin Approximation of Parametric and Stochastic Elliptic PDEs,” *SIAM J. Sci. Comput.* **33** (1), 364–385 (2011). doi [10.1137/100785715](https://doi.org/10.1137/100785715).
2. J. D. Eshelby, “Energy Relations and the Energy-Momentum Tensor in Continuum Mechanics,” in *Fundamental Contributions to the Continuum Theory of Evolving Phase Interfaces in Solids* (Springer, Berlin, 1999), pp. 82–119.
3. S. A. Matveev, D. A. Zheltkov, E. E. Tyrtysnikov, and A. P. Smirnov, “Tensor Train Versus Monte Carlo for the Multicomponent Smoluchowski Coagulation Equation,” *J. Comput. Phys.* **316**, 164–179 (2016). doi [10.1016/j.jcp.2016.04.025](https://doi.org/10.1016/j.jcp.2016.04.025).
4. H. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, “A Survey of Multilinear Subspace Learning for Tensor Data,” *Pattern Recognit.* **44** (7), 1540–1551 (2011). doi [10.1016/j.patcog.2011.01.004](https://doi.org/10.1016/j.patcog.2011.01.004).
5. F. L. Hitchcock, “Multiple Invariants and Generalized Rank of a P-Way Matrix or Tensor,” *J. Math. Phys.* **7** (1–4), 39–79 (1928). doi [10.1002/sapm19287139](https://doi.org/10.1002/sapm19287139).
6. R. A. Harshman, *Foundations of the PARAFAC Procedure: Models and Conditions for an “Explanatory” Multimodal Factor Analysis* (UCLA Working Papers in Phonetics, Ann Arbor, 1970), Vol. 16.
7. L. R. Tucker, “Some Mathematical Notes on Three-Mode Factor Analysis,” *Psychometrika* **31** (3), 279–311 (1966). doi [10.1007/BF02289464](https://doi.org/10.1007/BF02289464).
8. L. De Lathauwer, B. De Moor, and J. Vandewalle, “A Multilinear Singular Value Decomposition,” *SIAM J. Matrix Anal. Appl.* **21** (4), 1253–1278 (2000). doi [10.1137/S0895479896305696](https://doi.org/10.1137/S0895479896305696).
9. I. V. Oseledets, “Tensor-Train Decomposition,” *SIAM J. Sci. Comput.* **33** (5) 2295–2317 (2011). doi [10.1137/090752286](https://doi.org/10.1137/090752286).
10. T. Shi and A. Townsend, “On the Compressibility of Tensors,” *SIAM J. Matrix Anal. Appl.* **42** (1), 275–298 (2021). doi [10.1137/20M1316639](https://doi.org/10.1137/20M1316639).
11. V. Strassen, “Gaussian Elimination Is Not Optimal,” *Numerische Mathematik* **13** (4), 354–356 (1969). doi [10.1007/BF02165411](https://doi.org/10.1007/BF02165411).
12. A. Cichocki, D. Mandic, L. De Lathauwer, et al., “Tensor Decompositions for Signal Processing Applications: from Two-Way to Multiway Component Analysis,” *IEEE Signal Process. Mag.* **32** (2), 145–163 (2015). doi [10.1109/MSIP.2013.2297439](https://doi.org/10.1109/MSIP.2013.2297439).
13. V. Lebedev, Y. Ganin, M. Rakhuba, et al., “Speeding-up Convolutional Neural Networks Using Fine-tuned CP-Decomposition,” <https://arxiv.org/pdf/1412.6553>. Cited August 29, 2024.
14. J. Håstad, “Tensor Rank Is NP-Complete,” *J. Algo.* **11** (4), 644–654 (1990). doi [10.1016/0196-6774\(90\)90014-6](https://doi.org/10.1016/0196-6774(90)90014-6).
15. M. J. Mohlenkamp, “Musings on Multilinear Fitting,” *Linear Algebra Its Appl.* **438** (2), 834–852 (2013). doi [10.1016/j.laa.2011.04.019](https://doi.org/10.1016/j.laa.2011.04.019).
16. M. Udell and A. Townsend, “Why Are Big Data Matrices Approximately Low Rank?” *SIAM J. Math. Data Sci.* **1** (1), 144–160 (2019). doi [10.1137/18M1183480](https://doi.org/10.1137/18M1183480).
17. S. Budzinskiy, “Entrywise Tensor-Train Approximation of Large Tensors Via Random Embeddings,” <https://arxiv.org/pdf/2403.11768>. Cited August 29, 2024.
18. N. Gillis and Y. Shitov, “Low-Rank Matrix Approximation in the Infinity Norm,” *Linear Algebra Its Appl.* **581**, 367–382 (2019). doi [10.1016/j.laa.2019.07.017](https://doi.org/10.1016/j.laa.2019.07.017).
19. V. Daugavet, “Uniform Approximation of a Function of Two Variables, Tabulated as the Product of Functions of a Single Variable,” *Zh. Vychisl. Mat. Mat. Fiz.* **11** (2), 289–303 (1971) [*USSR Comput. Math. Math. Phys.* **11** (2), 1–16 (1971)]. doi [10.1016/0041-5553\(71\)90160-1](https://doi.org/10.1016/0041-5553(71)90160-1).
20. N. Zamarashkin, S. Morozov, and E. Tyrtysnikov, “On the Best Approximation Algorithm by Low-Rank Matrices in Chebyshev’s Norm,” *Zh. Vychisl. Mat. Mat. Fiz.* **62** (5), 723–741 (2022) [*USSR Comput. Math. Math. Phys.* **62** (5), 701–718 (2022)]. doi [10.1134/S0965542522050141](https://doi.org/10.1134/S0965542522050141).
21. S. Morozov, M. Smirnov, and N. Zamarashkin, “On the Optimal Rank-1 Approximation of Matrices in the Chebyshev Norm,” *Linear Algebra Its Appl.* **679**, 4–29 (2023). doi [10.1016/j.laa.2023.09.007](https://doi.org/10.1016/j.laa.2023.09.007).
22. S. Budzinskiy, “Quasioptimal Alternating Projections and Their Use in Low-Rank Approximation of Matrices and Tensors,” <https://arxiv.org/pdf/2308.16097>. Cited August 29, 2024.
23. S. Budzinskiy, “On the Distance to Low-Rank Matrices in the Maximum Norm,” *Linear Algebra Its Appl.* **688**, 44–58 (2024). doi [10.1016/j.laa.2024.02.012](https://doi.org/10.1016/j.laa.2024.02.012).

Received

August 12, 2024

Accepted for publication

August 18, 2024

Information about the author

Stanislav V. Morozov — MSc, Junior Scientist; Marchuk Institute of Numerical Mathematics of RAS, ulitsa Gubkina 8, 119333, Moscow, Russia.