

УДК 517.977

ОЦЕНКИ АППРОКСИМАЦИИ ТЕНЗОРНЫХ ПОЕЗДОВ ПО НОРМЕ ЧЕБЫШЁВА¹⁾

© 2019 г. А. И. Осинский

(119333 Москва, ул. Губкина, 8, Ин-т вычисл. матем. РАН, Россия)

e-mail: sasha_o@list.ru

Поступила в редакцию 23.05.2018 г.

Получена новая поэлементная оценка точности крестовой аппроксимации, применяемой для приближения многоиндексного массива (тензора) в формате тензорного поезда. Новая оценка является первой известной оценкой точности, отличающейся от наилучшей на множитель, зависящий лишь от ранга приближения r и размерности тензора d , причем зависимость от размерности при фиксированном ранге лишь порядка d^{const} , а не const^d . Тем самым она обосновывает применение крестового метода даже для больших величин размерности тензора. Библ. 10.

Ключевые слова: многомерные массивы, нелинейные приближения, принцип максимального объема.

DOI: 10.1134/S0044466919020121

Тензорные поезда, впервые предложенные в [1], стали популярным инструментом приближения тензоров благодаря небольшим затратам памяти $O(dnr + dr^3)$, высокой скорости выполнения операций (для малых рангов такая же асимптотика, как для канонического разложения) и быстрому алгоритму построения (см. [2]) с помощью крестовых аппроксимаций матриц развертки, для которых применяется алгоритм `maxvol` [3]. Данный алгоритм позволяет найти подматрицу \hat{A} , объем которой $V_2(\hat{A}) = \text{vol}(\hat{A}) = |\det \hat{A}|$ нельзя уменьшить заменой лишь одной строки или столбца. В том случае, когда объем найденной матрицы оказывается максимальным или почти максимальным (что часто выполняется на практике) среди подматриц, отличающихся не более, чем одной строкой и одним столбцом, можно гарантировать высокую точность построенной на основе \hat{A} крестовой аппроксимации по норме Чебышёва.

Теорема 1 (см. [4]). Пусть матрица $A \in \mathbb{C}^{M \times N}$ имеет следующую блочную форму:

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix},$$

где подматрица $A_{11} \in \mathbb{C}^{r \times r}$, $\text{rank } A_{11} = r$ обладает максимальным объемом среди всех подматриц размера $r \times r$. Рассмотрим псевдоскелетную аппроксимацию матрицы A , построенную на столбцах

$C = \begin{bmatrix} A_{11} \\ A_{21} \end{bmatrix}$ и строках $R = [A_{11} \ A_{12}]$, используя в качестве ядра G матрицу A_{11}^{-1} . Тогда выполняется следующая оценка:

$$\|A - CA_{11}^{-1}R\|_C \leq (r+1)^2 \|E\|_C, \quad (1)$$

где $\|E\|_C$ — ошибка наилучшего приближения A по C -норме (норме Чебышёва),

$$\|A\|_C = \max_{i,j} |A_{ij}|.$$

¹⁾Работа выполнена при финансовой поддержке РФФИ (проект № 14-11-00806).

В случае, когда объем A_{i_1} отличается от максимального в $\alpha \geq 1$ раз, получим

$$\|A - CA_{i_1}^{-1}R\|_C \leq \alpha(r + 1)^2 \|E\|_C. \tag{2}$$

Попытки прямого обобщения данной и похожих оценок на тензоры вызывают часто встречающуюся в таких случаях трудность — появление в коэффициенте множителя, экспоненциально зависящего от размерности тензора d . Такие оценки для разложения Таккера были впервые получены в [5] (случай $d = 3$) и обобщены и улучшены в [6].

В данной статье будут доказаны оценки погрешности аппроксимации тензоров по норме Чебышёва с коэффициентом, пропорциональным $r^{\log_2 d}$ вместо r^d , что гарантирует высокую точность аппроксимаций даже при большом числе размерностей тензора.

Стоит отметить, что в [7] уже была предложена похожая оценка, однако с коэффициентом, зависящим от максимального числа обусловленности $\kappa = \|\hat{A}\|_C \|\hat{A}^{-1}\|_C$ среди всех используемых при построении подматриц \hat{A} , который ничем не ограничен. Например, если

$$A = \begin{bmatrix} 1 & 0 \\ 0 & \varepsilon I \end{bmatrix},$$

то матрица максимального объема размера 2×2 равна:

$$\hat{A} = \begin{bmatrix} 1 & 0 \\ 0 & \varepsilon \end{bmatrix},$$

и $\kappa = 1/\varepsilon$. В результате оценка ошибки не будет убывать при уменьшении ε , хотя ошибка наилучшего приближения стремится к нулю.

Итак, пусть нам дан d -мерный тензор A с индексами столбцов i_1, \dots, i_d . Его мы будем обозначать через $A(i_1, \dots, i_d)$. Буквой J далее будем обозначать мультииндексы, содержащие в себе значения одновременно нескольких i_k . При записи в скобках это означает, что несколько индексов объединены в один, что уменьшает размерность тензора. Например, матрица $A(i_k, J^{\neq k})$ является k -й матрицей развертки тензора A . Ее строки соответствуют индексу i_k , а столбцы, наоборот, соответствуют столбцам тензора A , когда все индексы, кроме i_k , фиксированы и их значения хранятся в мультииндексе $J^{\neq k}$.

Мы будем приближать заданный нам тензор способом, аналогичным предложенному в [2]:

$$\begin{aligned} A(i_1, \dots, i_d) &\approx \tilde{A} = \sum_{\tilde{s}, \tilde{t}} A(i_1, J_{i_1}^{>1}) [A(J_{s_1}^{\leq 1}, J_{t_1}^{>1})]_{\tau}^+ \dots A(J_{s_{d-1}}^{\leq d-1}, i_d) = \\ &= \sum_{\tilde{s}, \tilde{t}} \prod_{k=1}^d A(J_{s_{k-1}}^{\leq k-1}, i_k, J_{t_k}^{>k}) [A(J_{s_k}^{\leq k}, J_{t_k}^{>k})]_{\tau}^+, \end{aligned} \tag{3}$$

где d — число измерений тензора, $J_{t_k}^{>k}$ — часть номеров столбцов, выбранных среди мультииндекса $J^{>k} = \overline{i_{k+1} \dots i_d}$. Аналогично, $J_{s_k}^{\leq k}$ представляет из себя выбор части номеров строк из мультииндекса $J^{\leq k} = \overline{i_1 \dots i_k}$. Если все подматрицы вида

$$A_k = A(J_{s_k}^{\leq k}, J_{t_k}^{>k}) \tag{4}$$

состоят из r строк и столбцов, то мы получаем некоторое приближение ранга r .

Для случая $d = 2$ мы получаем известное псевдоскелетное разложение матрицы:

$$A(i_1, i_2) = A(i_1, J_t) [A(J_s, J_t)]_{\tau}^+ A(J_s, i_2) + E(i_1, i_2) = C\hat{A}_{\tau}^+ R + E. \tag{5}$$

Здесь и далее индекс τ обозначает τ -псевдообращение: перед взятием псевдообратной матрицы все сингулярные числа, меньшие τ , зануляются. Это позволяет ограничить 2-норму τ -псевдообратной матрицы значением $1/\tau$, так как при псевдообращении нулевые сингулярные числа остаются нулевыми. Индекс r будет обозначать, что у матрицы все сингулярные числа после r -го (с нумерацией по убыванию) приравниваются нулю. Когда используются сразу два индекса, τ и r , это означает, что считаются нулевыми все сингулярные числа, меньшие τ или s с номером,

большим r . Индекс k при этом, как и ранее, будет использоваться только для обозначения того, что это матрица вида (4).

Если $A_k \in \mathbb{C}^{r \times r}$ обладает максимальным объемом, а $\tau = 0$, то из теоремы 1 известна оценка на C – норму приближения (5):

$$\|E_{\tau=0}\|_C = \|A - C\hat{A}^{-1}R\|_C \leq (r + 1)^2 \|F\|_C,$$

где F – наилучшее по C -норме приближение ранга r .

Вместо i_1 и i_2 в (5) можно подставить мультииндексы. В частности,

$$A(J^{\leq k-1}, i_k, i_{k+1}, J^{> k+1}) = A(J^{\leq k-1}, i_k, J_{i_k}^{> k}) A_{k\tau}^+ A(J_{s_k}^{\leq k}, i_{k+1}, J^{> k+1}) + E(J^{\leq k-1}, i_k, i_{k+1}, J^{> k+1}),$$

$$\|E_{\tau=0}\|_C \leq (r + 1)^2 \|F\|_C.$$

Далее можно рассмотреть уже тензоры столбцов $C = A(J^{\leq k-1}, i_k, J_{i_k}^{> k})$ и строк $R = A(J_{s_k}^{\leq k}, i_{k+1}, J^{> k+1})$ по отдельности, перегруппировать в них индексы и снова применить псевдоскелетное разложение. В конечном итоге, продолжая эту процедуру, получаем структуру вида (3). Оценим C -норму погрешности аппроксимации такого представления тензора.

Теорема 2. Для любого тензора $A = A(i_1, \dots, i_d)$ существует его приближение \tilde{A} вида (3) такое, что

$$\|A - \tilde{A}\|_C \leq \frac{(4r)^{\lceil \log_2 d \rceil} - 1}{4r - 1} (r + 1)^2 \|F\|_C, \tag{6}$$

где F – наилучшее по C -норме приближение тензора A ранга r .

Доказательство. Итак, выбрав в матрице-развертке тензора подматрицу максимального объема \hat{A} , на первом шаге мы получим разложение следующего вида:

$$A = CBR + E,$$

$$B = \hat{A}_\tau^+,$$

где через A будем уже обозначать матрицу развертки, а \hat{A} – матрицу на пересечении столбцов C и строк R , E – погрешность такой аппроксимации.

Далее мы строим разложение для тензоров, соответствующих матрицам C и R . В конечном итоге, получим некоторые приближения этих тензоров:

$$C = \tilde{C} + E_C, \quad R = \tilde{R} + E_R.$$

На каждом шаге мы будем использовать разложение ранга r , то есть r строк и столбцов. Если каждый раз при построении матриц развертки делить индексы примерно поровну, то конечный результат будет достигнут не более, чем за $\lceil \log_2 d \rceil$ шагов. При этом погрешность, которая окажется на k -м слое (шаге), будет оцениваться следующим образом (для простоты индекс k , который здесь является верхним, чтобы не путать с подматрицей A_k , далее будет опущен):

$$\|A^k - \tilde{A}^k\|_C = \|CBR + E - \tilde{C}\tilde{R}\tilde{B}\|_C \leq \|E_C B R\|_C + \|C B E_R\|_C + \|E_C B E_R\|_C + \|E\|_C,$$

$$\|E_C\|_C \leq \epsilon_{k+1}, \quad \|E_R\|_C \leq \epsilon_{k+1},$$

где ϵ_{k+1} – оценка для наибольшей погрешности на $k + 1$ -м слое. Таким образом, нужно выбрать такую оценку ϵ_k , чтобы гарантировать неравенство

$$\|A^k - \tilde{A}^k\|_C \leq \epsilon_k.$$

При этом при $k = \lceil \log_2 d \rceil$ можно считать значение ϵ_k равным оценке для $\|E\|_C$ для всего тензора.

Для начала оценим погрешность E :

$$E = E_{\tau=0} + C(\hat{A}^{-1} - \hat{A}_\tau^+)R = E_{\tau=0} + (C\hat{A}^{-1})(\hat{A} - \hat{A}_\tau)(\hat{A}^{-1}R).$$

Одним из свойств подматрицы максимального объема является то, что 2-норма строк $C\hat{A}^{-1}$ не превосходит \sqrt{r} . Оно следует напрямую из соотношения

$$\|C\hat{A}^{-1}\|_C \leq 1,$$

которое доказано в [8]. То же верно и для 2-нормы столбцов $\hat{A}^{-1}R$. C -норма произведения трех матриц далее оценивается через произведение максимальной 2-нормы строк первой матрицы, 2-нормы второй матрицы и максимальной 2-нормы столбцов третьей матрицы:

$$\|ABC\|_C = \max_{i,j} |A_{i,:} B C_{:,j}| \leq \max_{i,j} \|A_{i,:}\|_2 \|B\|_2 \|C_{:,j}\|_2. \tag{7}$$

Поэтому для C -нормы погрешности будет верно следующее соотношение с учетом (7):

$$\begin{aligned} \|E\|_C &\leq \|E_{\tau=0}\|_C + \|(C\hat{A}^{-1})(\hat{A} - \hat{A}_\tau)(\hat{A}^{-1}R)\|_C \leq (r+1)^2 \|F\|_C + \max_{i,j} \|(C\hat{A}^{-1})_{i,:}\|_2 \|\hat{A} - \hat{A}_\tau\|_2 \|(A^{-1}R)_{:,j}\|_2 \leq \\ &\leq (r+1)^2 \|F\|_C + r \|\hat{A} - \hat{A}_\tau\|_2 \leq (r+1)^2 \|F\|_C + r\tau. \end{aligned}$$

Теперь оценим остальные слагаемые, применяя, где необходимо, выражение (7) и оценивая 2-норму строк и столбцов через C -норму:

$$\begin{aligned} \|E_C B R\|_C &= \|E_C \hat{A}_\tau^+ R\|_C = \|E_C (\hat{A} \hat{A}_\tau^+) (\hat{A}^{-1} R)\|_C \leq r \|E_C\|_C \|\hat{A} \hat{A}_\tau^+\|_2 \leq r \varepsilon_{k+1}, \\ \|C B E_R\|_C &\leq r \varepsilon_{k+1}, \quad \|E_C B E_R\|_C \leq r \|E_C\|_C \|\hat{A}_\tau^+\|_2 \|E_R\|_C \leq r \frac{(\varepsilon_{k+1})^2}{\tau}. \end{aligned} \tag{8}$$

Взяв $\tau = \varepsilon_{k+1}$, получим следующую оценку для погрешности на k -м слое:

$$\begin{aligned} \|A^k - \tilde{A}^k\|_C &\leq r \varepsilon_{k+1} + r \varepsilon_{k+1} + r \frac{(\varepsilon_{k+1})^2}{\tau} + r\tau + (r+1)^2 \|F\|_C = \\ &= 4r \varepsilon_{k+1} + (r+1)^2 \|F\|_C. \end{aligned}$$

Таким образом, достаточно взять

$$\begin{aligned} \varepsilon_k &= 4r \varepsilon_{k+1} + (r+1)^2 \|F\|_C, \\ k &= 1, \lceil \log_2 d \rceil, \\ \varepsilon_{\lceil \log_2 d \rceil} &= (r+1)^2 \|F\|_C, \\ \|A - \tilde{A}\| &\leq \varepsilon_1. \end{aligned}$$

Суммируя полученную в результате раскрытия рекурренты геометрическую прогрессию, получаем оценку (6).

Оценку (6) можно улучшить, если использовать $n > r$ строк и столбцов. В этом случае в качестве \hat{A} нужно использовать подматрицу максимального r -проективного объема, т.е. обладающую максимальным произведением первых r сингулярных чисел:

$$V_2^r(\hat{A}) = \prod_{i=1}^r \sigma_i(\hat{A}).$$

Крестовая аппроксимация на основе максимального проективного объема была подробно исследована в [9], где, в частности, были получены следующие оценки:

$$\|A - C \hat{A}_r^+ R\|_C \leq \frac{(n+1)^2}{n-r+1} \|F\|_C, \tag{9}$$

$$\|\hat{A}_r^+ R_i\|_2 \leq \sqrt{\frac{r}{n-r+1}}, \quad \|C_j \hat{A}_r^+\|_2 \leq \sqrt{\frac{r}{n-r+1}}, \tag{10}$$

где R_i – произвольная строка матрицы R , а C_j – произвольный столбец C , не содержащиеся в \hat{A} . Для столбцов и строк из \hat{A} в правой части неравенств будет 1, а потому они остаются верными, пока $n \leq 2r - 1$.

Теорема 3. Для любого тензора $A = A(i_1, \dots, i_d)$ существует его приближение \tilde{A} вида (3) такое, что

$$\|A - \tilde{A}\|_C \leq \frac{\left(4\sqrt{\frac{rn}{n-r+1}}\right)^{\lceil \log_2 d \rceil} - 1}{4\sqrt{\frac{rn}{n-r+1}} - 1} \frac{(n+1)^2}{n-r+1} \|F\|_C, \tag{11}$$

где F – наилучшее по C -норме приближение тензора A ранга r , а $n \leq 2r - 1$.

Доказательство проводится полностью аналогично случаю $n = r$. В качестве матрицы B будем использовать \hat{A}_r^+ . Нам снова нужно оценить каждое из слагаемых погрешности на k -м слое:

$$\begin{aligned} \|A^k - \tilde{A}^k\|_C &\leq \|E_C BR\|_C + \|CBE_R\|_C + \|E_C BE_R\|_C + \|E\|_C, \\ \|E_C\|_C &\leq \varepsilon_{k+1}, \quad \|E_R\|_C \leq \varepsilon_{k+1}. \end{aligned}$$

Для этого воспользуемся выражениями (9) и (10):

$$\begin{aligned} \|E_C\|_C &\leq \|E_{\tau=0}\|_C + \|(C\hat{A}_r^+)(\hat{A} - \hat{A}_\tau)(\hat{A}_r^+ R)\|_C \leq \frac{(n+1)^2}{n-r+1} \|F\|_C + \frac{r}{n-r+1} \|\hat{A} - \hat{A}_\tau\|_2 \leq \\ &\leq \frac{(n+1)^2}{n-r+1} \|F\|_C + \frac{r}{n-r+1} \tau, \\ \|E_C BR\|_C &= \|E_C \hat{A}_r^+ R\|_C = \|E_C (\hat{A} \hat{A}_\tau^+) (\hat{A}_r^+ R)\|_C \leq \frac{\sqrt{rn}}{\sqrt{n-r+1}} \|E_C\|_C \|\hat{A} \hat{A}_\tau^+\|_2 \leq \frac{\sqrt{rn}}{\sqrt{n-r+1}} \varepsilon_{k+1}, \\ \|CBE_R\|_C &\leq \frac{\sqrt{rn}}{\sqrt{n-r+1}} \varepsilon_{k+1}, \\ \|E_C BE_R\|_C &\leq n \|E_C\|_C \|\hat{A}_r^+\|_2 \|E_R\|_C \leq n \frac{(\varepsilon_{k+1})^2}{\tau}. \end{aligned}$$

Полагая $\tau = \sqrt{\frac{n(n-r+1)}{r}} \varepsilon_{k+1}$, получаем, что можно выбрать

$$\varepsilon_k = 4 \sqrt{\frac{rn}{n-r+1}} \varepsilon_{k+1} + \frac{(n+1)^2}{n-r+1} \|F\|_C.$$

Здесь снова можно просуммировать геометрическую прогрессию, что даст в итоге (11).

Следствие. Выбрав $n = 2r - 1$, получим оценку

$$\|A - \tilde{A}\|_C \leq \frac{(4\sqrt{2r-1})^{\lceil \log_2 d \rceil} - 1}{4\sqrt{2r-1} - 1} 4r \|F\|_C.$$

Как видим, в случае тензоров, как и в случае матриц, дополнительные $r - 1$ строк и столбцов позволяют уменьшить в два раза степенную зависимость коэффициента от r .

Насколько данные оценки могут быть близки к реальному положению дел? Рассмотрим для простоты оценку (1) для матриц. В ней присутствует коэффициент $(r + 1)^2$. Численные эксперименты со случайными матрицами в [9] показали коэффициент вида $(\alpha r + 1)^2$ с достаточно малым α вместо коэффициента $(r + 1)^2$. В результате предсказанная оценкой квадратичная зависимость наблюдалась лишь на очень больших рангах, когда $\alpha^2 r^2 \gg 1$. Что касается коэффициента, который возводится в степень $\log_2 d$, то для него также разумно предположить, что, например, 2-норма строк E_C изменяется как $\sqrt{q + \alpha n} \|E_C\|_C$. Это означает, что на практике экспоненциальная зависимость от $\log_2 d$ будет заметна лишь на очень больших рангах, в то время как при достаточно малых рангах зависимость будет не более, чем степенной или даже линейной: в случае малости погрешности даже при больших τ верно равенство $\hat{A}_\tau = \hat{A}$, которое исключает одно из слагаемых и почти сводит к нулю $\|E_C BE_R\|_C$, поэтому в итоге в степень $\log_2 d$ будет возводиться число, близкое к 2.

Например, для $r = 1$ можно выбрать $\tau = \sigma_1(\hat{A}) \geq \|A\|_C \geq \|F\|_C$, что уже исключает необходимость τ -псевдообращения. И тогда

$$\|A - \tilde{A}\|_C \leq 4(d-1) \|F\|_C + O(\|F\|_C^2 / \|A\|_C). \tag{12}$$

Более того, $\|E_C BE_R\|_C$ может уменьшать погрешность, а не увеличивать, и тогда $O(\|F\|_C^2 / \|A\|_C)$ также исчезнет. Более глубокий анализ может позволить полностью избавиться от двух слагаемых даже при больших рангах.

Ранг 1 также можно рассматривать как частный случай канонического разложения, что при малых значениях погрешности приводит к оценке вида

$$\|A - \tilde{A}\|_C \leq 2d \|F\|_C + O(\|F\|_C^2 / \|A\|_C). \quad (13)$$

При $d = 2$ данная оценка совпадает с (12). При этом легко привести пример последовательности тензоров, для которой оценка (13) достигается, когда величина погрешности стремится к нулю. Например, для

$$A = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \times \dots \times \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \varepsilon \begin{bmatrix} 1 \\ -1 \end{bmatrix} \times \dots \times \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

$\|F\|_C = \varepsilon$, а погрешность построенной на основе максимального элемента (что соответствует максимальному объему для $r = 1$) аппроксимации равна

$$\left\| A - \frac{1}{(1 + \varepsilon)^{d-1}} \begin{bmatrix} 1 + \varepsilon \\ 1 - \varepsilon \end{bmatrix} \times \dots \times \begin{bmatrix} 1 + \varepsilon \\ 1 - \varepsilon \end{bmatrix} \right\|_C = 2d\varepsilon - O(\varepsilon^2).$$

Тем не менее, как уже было сказано, как минимум для матриц полученная оценка достигается крайне редко.

В целом точность полученной формулы по некоторым аспектам остается под вопросом и требует проведения соответствующих вычислительных экспериментов с тензорами.

Естественно, рассуждения для среднего не отменяют возможность существования худшего случая, на котором полученная оценка будет достигаться для рангов, больших 1, хотя на данный момент такие примеры неизвестны.

Конечно, возможно также существование оценок по норме Чебышёва, никак не связанных с максимальным объемом или крестовыми аппроксимациями, однако, в матричном случае никаких аналогов, дающих схожие оценки по C -норме, автору неизвестны и мало какие алгоритмы способны соревноваться по простоте и скорости построения аппроксимации с крестовым методом.

Также крайне интересны были бы оценки крестовых аппроксимаций, например, по норме Фробениуса, так как в матричном случае известны оценки, сколь угодно близкие к точности сингулярного разложения [10]. Однако за неимением полного аналога SVD для тензоров получение таких оценок до сих пор является крайне сложной задачей.

Вышесказанное еще сильнее подчеркивает важность полученного результата — доказаны такие оценки, аналоги которых пока остаются практически недостижимыми для всех прочих норм и методов.

СПИСОК ЛИТЕРАТУРЫ

1. *Oseledets I.V.* Tensor-train decomposition // *SIAM J. Scientific Comput.* 2011. V. 33. № 5. P. 2295–2317.
2. *Oseledets I.V., Tyrtshnikov E.E.* TT-cross approximation for multidimensional arrays // *Linear Algebra and its Appl.* 2010. V. 432. P. 70–88.
3. *Goreinov S.A., Oseledets I.V., Savostyanov D.V. et al.* How to find a good submatrix // *Matrix Methods: Theory, Algorithms, Applications.* 2010. P. 247–256.
4. *Горейнов С.А., Тиртышников Е.Е.* Квазиоптимальность скелетного приближения матрицы в чебышевской норме // *Докл. АН.* 2011. М. 438. № 5. С. 593–594.
5. *Oseledets I.V., Savostyanov D.V., Tyrtshnikov E.E.* Tucker dimensionality reduction of the three dimensional arrays in linear time // *SIAM J. of Matrix Anal. Appl.* 2008. V. 30. № 3. P. 939–956.
6. *Горейнов С.А.* О крестовой аппроксимации многоиндексного массива // *Докл. АН.* 2008. Т. 420. № 4. С. 439–441.
7. *Savostyanov D.V.* Quasioptimality of maximum-volume cross interpolation of tensors // *Linear Algebra and its Applications.* 2014. V. 458. P. 217–244.
8. *Goreinov S.A., Tyrtshnikov E.E.* The maximal-volume concept in approximation by low-rank matrices // *Contemporary Math.* 2001. V. 268. P. 47–51.
9. *Osinsky A.I., Zamrashkin N.L.* Pseudo-skeleton approximations with better accuracy estimates // *Linear Algebra and its Appl.* 2018. V. 537. P. 221–249.
10. *Boutsidis C., Woodruff D.P.* Optimal cur matrix decompositions // *Proc. of the 46th Annual ACM Symposium on Theory of Comput.* 2014. P. 353–362.