УДК 519.61

Вероятностные оценки точности крестового приближения ранга 1

А. И. Осинский[1,2]

[1]Институт вычислительной математики Российской Академии наук

[2]Московский физико-тахнический институт (государственный университет)

Среди методов малорангового приближения матриц важное место занимает так называемая крестовая или скелетная аппроксимация, в которой исходная матрица A приближается с помощью произведения $C\hat{A}^{-1}R$, где $\hat{A} \in \mathbb{C}^{r \times r}$ - подматрица матрицы A, стоящая на пересечении столбцов C и строк R.

Далее полученное разложение можно использовать как для приближения исходной матрицы, так и для поиска в ней максимального элемента. Кроме того, этот же подход можно использовать для аппроксимации тензоров [1].

Для скелетного и связанного с ним псевдоскелетного CGR разложения известны оценки, гарантирующие высокую точность приближения, если подматрица на пересечении C и R обладает близким к максимальному объемом [2, 3] или, в более общем случае, проективным объемом [4]. Однако, в общем случае поиск подматрицы максимального объема является NP-сложной задачей, и напрямую эти оценки не применимы.

Одним из наиболее популярных методов построения крестового малорангового приближения является алгоритм maxvol [5]. В рассматриваемом нами случае ранга 1 он сводится к поиску максимального по модулю элемента в случайно выбранном столбце, затем в строке с найденным максимальным элементом и так далее. В итоге полученный в результате элемент матрицы окажется максимальным в своей строке и своем столбце. К сожалению, это не гарантирует, что он будет максимальным или даже хоть в какой-то мере близок к максимальному по модулю элементу во всей матрице, а значит нельзя гарантировать, что полученное приближение будет достаточно точным.

Тем не менее, на практике в большинстве случаев получаемая крестовым методом матрица малого ранга является хорошим приближением для оригинальной матрицы, что говорит о том, что если внести случайность в элементы матрицы, найденный с помощью maxvol элемент с большой вероятностью действительно будет близок к максимальному.

Оценки даже для этого частного случая очень важны, так как, например, применив алгоритм k раз, можно построить приближение ранга k, причем итоговая матрица погрешности будет иметь ранг на k меньше, чем исходная матрица.

Прежде всего, нам понадобятся несколько общих утверждений о случайных величинах. Первое из них проверяется приближенным интегрированием плотности распределения.

Утверждение 1. *Пусть случайная величина x имеет распределение χ^2 с $n > 2$ степенями свободы. Тогда*

$$\mathbb{P}(x > n - 2 + 2\sqrt{c(n-2)\ln n}) \leqslant \alpha n^{-c},$$

$$\alpha = \left(\frac{1}{\sqrt{\pi(n-2)}} + \frac{1}{2\sqrt{c\pi \ln n}} \right) e^{\frac{4\sqrt{\frac{c^3 \ln^3 n}{n-2}}}{3}}.$$

Лемма 1. *Пусть случайный вектор v обладает равномерным распределением на сфере в пространстве $\mathbb{C}^n$. Тогда среди первых k его элементов с вероятностью $1 - \alpha n^{-c} - \beta^k$ найдется хотя бы один, который по модулю не меньше μ,*

$$\alpha = \left(\frac{1}{\sqrt{\pi(n-2)}} + \frac{1}{2\sqrt{c\pi \ln n}} \right) e^{\frac{4}{3}\sqrt{\frac{c^3 \ln^3 n}{n-2}}},$$

$$\beta = \sqrt{\frac{2\mu^2 \left(n - 2 + 2\sqrt{c(n-2)\ln n}\right)}{\pi}}.$$

Доказательство. Такой вектор можно получить, взяв в качестве компонент нормально распределенные случайные величины, отнормировав его, и выбрав случайный поворот каждой компоненты в $\mathbb{C}$. Таким образом, если за основу взять величины x_i, то $|x_i|^2 \sim \chi^2(1)$, а

$$|v_i|^2 = \frac{|x_i|^2}{\sum\limits_{j=1}^{n} |x_j|^2}.$$

Из Утверждения 1 мы получаем, что

$$\mathbb{P}(\sum_{i=1}^{n} |x_i|^2 > n - 2 + 2\sqrt{c(n-2)\ln n}) \leqslant \alpha n^{-c},$$

Кроме того,

$$\mathbb{P}\left(|x_i|^2 < t^2\right) = \mathbb{P}\left(|x_i| < t\right) = 2\int\limits_0^t \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} dx \leqslant \sqrt{\frac{2t^2}{\pi}}.$$

$$\mathbb{P}\left(|x_i|^2 < t^2, i = \overline{1,k}\right) \leqslant \left(\frac{2t^2}{\pi}\right)^{\frac{k}{2}}.$$

В итоге, выбрав $t = \mu\sqrt{n - 2 + 2\sqrt{c(n-2)\ln n}}$, мы получим, что

$$\mathbb{P}\left(|v_i| < \mu,\ i = \overline{1,k}\right) \leqslant \alpha n^{-c} + \left(\frac{2\mu^2\left(n - 2 + 2\sqrt{c(n-2)\ln n}\right)}{\pi}\right)^{\frac{k}{2}}.$$

□

Теорема 1. *Пусть* $A = \sigma u v^* + E$, $\sigma > 0$, $\|u\|_2 = \|v\|_2 = 1$, $A \in \mathbb{C}^{m \times n}$. *Пусть вектор* v *имеет равномерное распределение на сфере в* $\mathbb{C}^n$, $n > 2$. *Обозначим*

$$\delta = \|E\|_C.$$

Пусть

$$\varepsilon = \frac{\|E\|_C}{\|A - E\|_C} = \frac{\|E\|_C}{\sigma \|u\|_\infty \|v\|_\infty} \leqslant \frac{1}{8}. \tag{1}$$

Пусть

$$\alpha = \left(\frac{1}{\sqrt{\pi(n-2)}} + \frac{1}{2\sqrt{c\pi \ln n}}\right) e^{\frac{4}{3}\sqrt{\frac{c^3 \ln^3 n}{n-2}}},$$

$$\beta = \frac{\left(1 - \sqrt{1 - 8\varepsilon}\right)\|v\|_\infty \sqrt{n - 2 + 2\sqrt{c(n-2)\ln n}}}{\sqrt{\pi}}.$$

Пусть алгоритм maxvol[5], *на первом шаге которого выбирается максимальный элемент среди первых* k *столбцов матрицы, возвратил элемент* a, *находящийся на пересечении строки* r *и столбца* c. *Тогда с вероятностью* $1 - \alpha n^{-c} - \beta^k$

$$\left\|A - c a^{-1} r\right\|_C \leqslant 8\delta \frac{1 + \varepsilon}{1 + \sqrt{1 - 8\varepsilon} - 2\varepsilon} \leqslant 4\delta(1 + 4\epsilon). \tag{2}$$

Доказательство целиком достаточно громоздко и не будет здесь приведено, однако стоит упомянуть об основной идее: благодаря лемме 1 мы можем утверждать, что с достаточно большой вероятностью для выбранного столбца j верно неравенство

$$|v_j| \geqslant \mu_1 \|v\|_\infty, \mu_1 \sim \epsilon.$$

Причем далее этот коэффициент будет только расти: если мы начали с $\mu = 4\epsilon$, то у следующего столбца будет $\mu \geqslant 1 - 4\epsilon$. В итоге полученный элемент окажется близок к максимальному, что и приведет нас к оценке (2).

Кроме того, свойство быстрого роста μ можно учесть для того, чтобы оценить количество шагов, необходимое для получения достаточно точной оценки. Итоговый результат дает следующая теорема.

Теорема 2. *Пусть в условиях Теоремы 1*

$$\beta = \frac{8\varepsilon \left\|v\right\|_\infty \sqrt{n - 2 + 2\sqrt{c(n-2)\ln n}}}{\sqrt{\pi}}.$$

Пусть процедура maxvol *возвращает найденный после 4-го шага элемент вне зависимости от того, является ли он максимальным в своем столбце. Тогда*

$$\left\|A - ca^{-1}r\right\|_C \leqslant 4\delta\left(1 + 4\varepsilon\right).$$

Таким образом, чтобы с указанной вероятностью найти элемент матрицы, по модулю близкий к максимальному, достаточно сравнить лишь $(k+1)m + 2n$ ее элементов.

Что же можно сказать о погрешности аппроксимации, если неравенство (1) не выполнено? В этом случае погрешность окажется порядка C-нормы всей матрицы. В худшем случае это

$$|d - ba^{-1}c| \leqslant |d| + |a| \leqslant \sigma \left\|u\right\|_\infty \left\|v\right\|_\infty + \delta + |a|.$$

Если же a достаточно велико, то, как мы уже знаем, погрешность не может отличаться от 4δ больше, чем в $\frac{|d|}{|a|}$.

Взяв минимум по a среди $4\delta\frac{|d|}{|a|}$ и $|d| + |a|$ и подставив оценку для d, мы получим, что погрешность не будет превосходить

$$\frac{1 + \delta + \sqrt{(1+\delta)(1+17\delta)}}{2}. \tag{3}$$

Чтобы проверить точность полученных оценок, были проведены вычисления для случайных матриц. А именно, матрица задавалась своим сингулярным разложением. Левые и правые сингулярные векторы выбирались случайно, сингулярные числа после первого приравнивались 1, а первое сингулярное число подбиралось из соотношения

$$x = \frac{\sigma \left\|u\right\|_\infty \left\|v\right\|_\infty}{\delta}.$$

Значение x откладывалось по оси абсцисс. При $x \geqslant 8$ перед применением алгоритма maxvol проверялось, что столбец является “хорошим”. На рисунке 1 показано отношение найденного элемента матрицы к ее максимальному элементу. На рисунке 2 показана погрешность аппроксимации. На рисунке 3 показана вероятность попадания в “плохой” столбец.

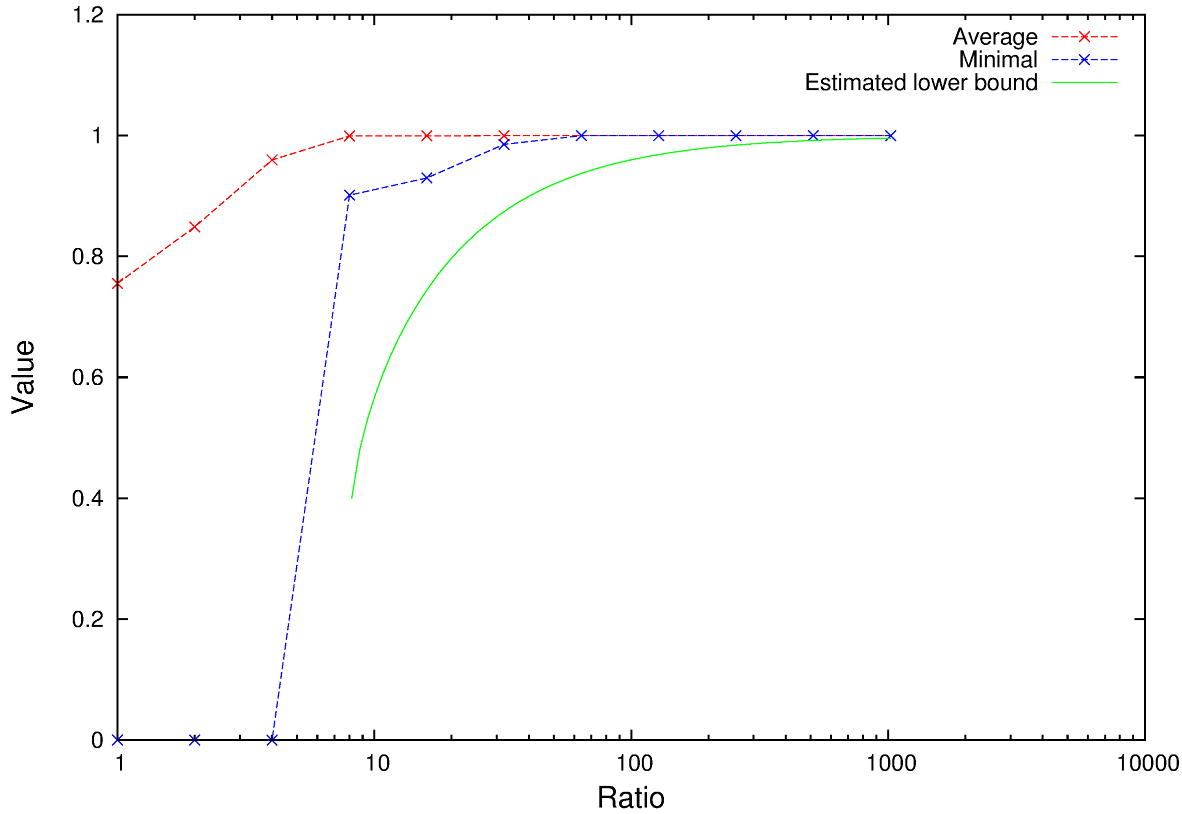

Рис. 1: Зависимость отношения модуля найденного элемента матрицы к максимальному от соотношения $\frac{\sigma\|u\|_\infty\|v\|_\infty}{\delta}$. Показано среднее и минимальное за 1000 генераций матрицы значение, а также оценка, равная $\sigma\mu_2^2\|u\|_\infty\|v\|_\infty+\delta$.

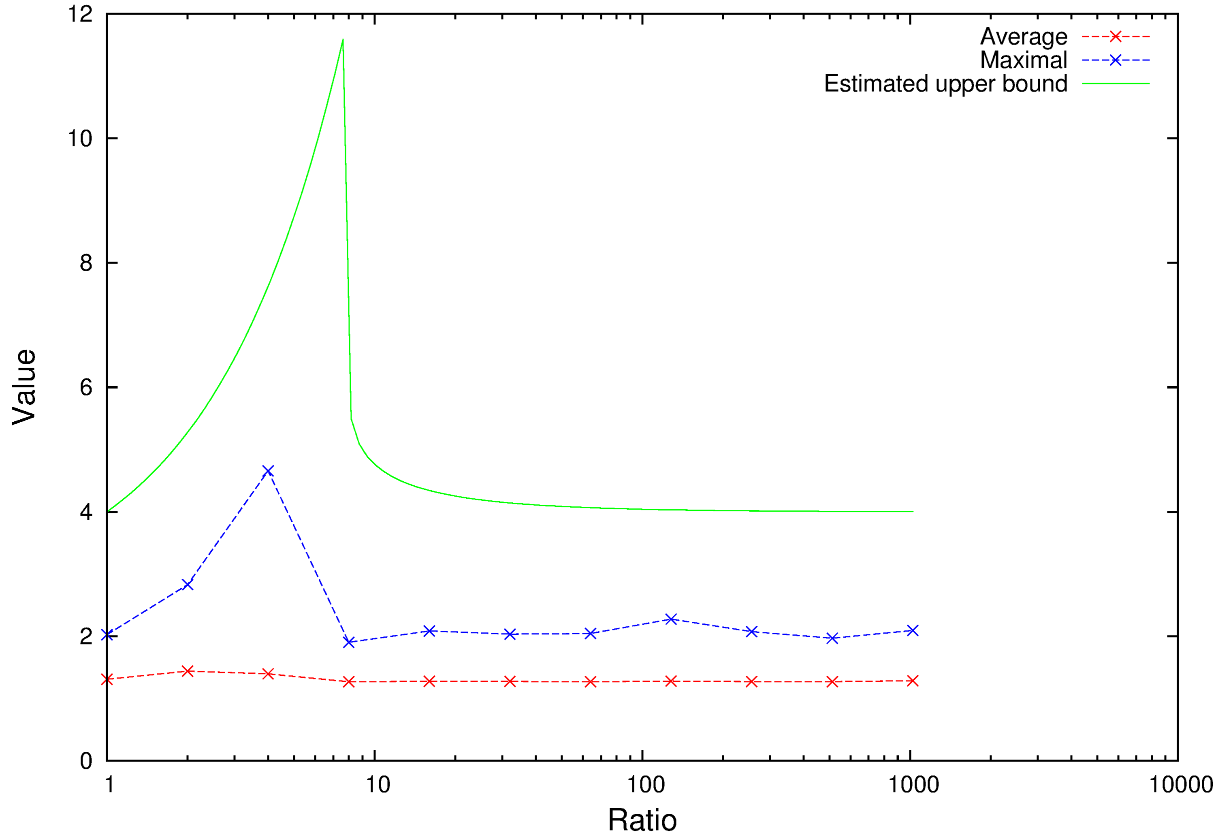

Рис. 2: Зависимость погрешности аппроксимации по отношению к δ от отношения $\frac{\sigma\|u\|_\infty\|v\|_\infty}{\delta}$. Показано среднее и максимальное за 1000 генераций матрицы значение, а также оценка погрешности. Для случая $\frac{1}{\varepsilon} < 8$ использовалось выражение (3).

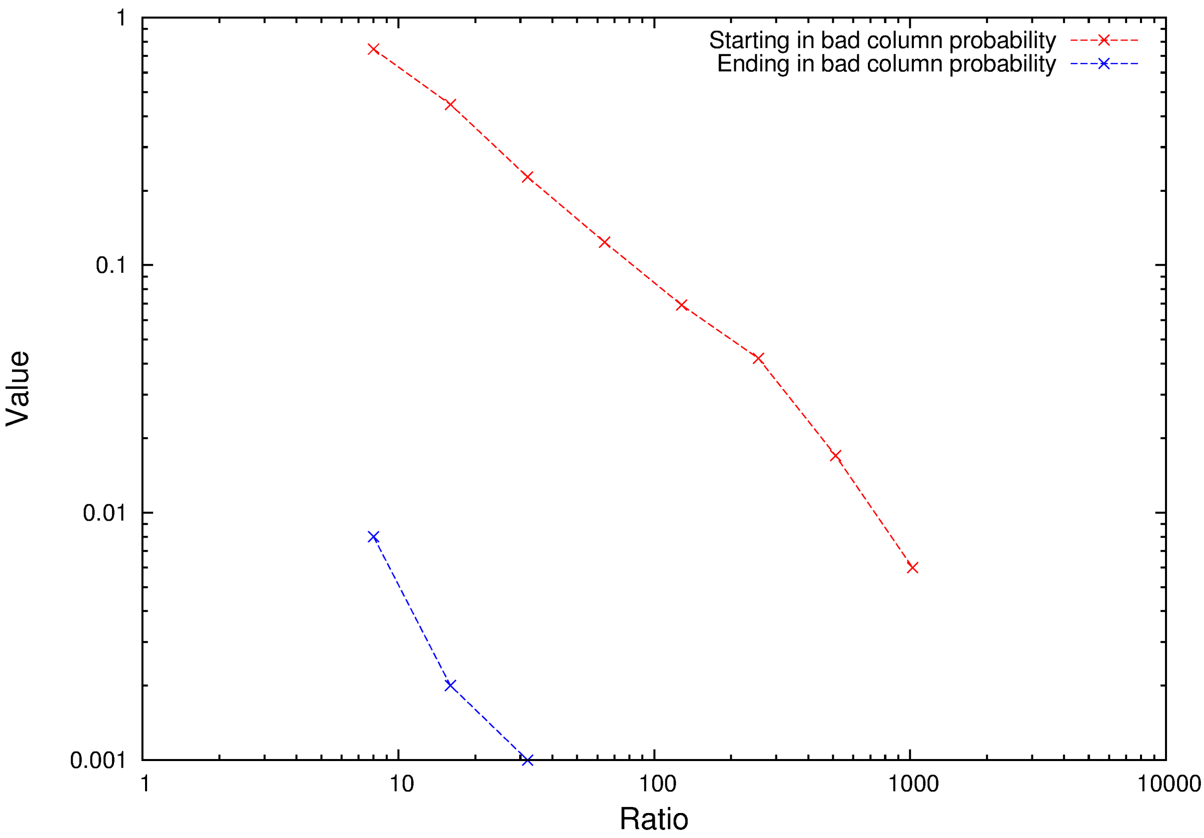

Рис. 3: Зависимость вероятности попасть в "плохой" столбец от отношения $\frac{\sigma\|u\|_\infty\|v\|_\infty}{\delta}$. Вероятность считалась на основе 1000 генераций матрицы. Вероятности показаны для случайно выбранного столбца и столбца, полученного после применения алгоритма.

Видно, что максимальная погрешность отличается от полученной нами оценки примерно в 2 раза. Это, возможно, связано с тем, что выбранная матрица E является наилучшим приближением по 2-норме, а не по C-норме.

На последнем рисунке видно, что вероятность не попасть в "хороший" столбец почти зануляется после применения алгоритма. Это связано, во-первых, с тем, что каждый шаг, по сути, примерно то же, что увеличение k на 1. Кроме того, при этом вероятность выбора элемента, соответствующего большому значению в $\sigma\|u\|_\infty\|v\|_\infty$ оказывается выше, чем элемента, соответствующего меньшему значению. Однако, анализ подобных вероятностей оказывается гораздо более сложной задачей. Тем не менее, его можно провести, наложив дополнительные ограничения на матрицу E.

Теорема 3. *Пусть в условиях Теоремы 2 вектор u также равномерно распределен на сфере в $\mathbb{C}^n$, $m = n$,*

$$\beta < 1/2,$$

$$\beta_u = \frac{\left(1-\sqrt{1-8\varepsilon}\right)\|u\|_\infty\sqrt{n-2+2\sqrt{c(n-2)\ln n}}}{\sqrt{\pi}} < 1/2,$$

матрица E состоит из независимых (в том числе от u и v) и равномерно распределенных на отрезке $[-\delta;\delta]$ случайных величин.

Пусть вместо просмотра k случайных столбцов, в начале будет применена процедура maxvol, *сделано минимум k шагов, и выбран максимальный элемент среди просмотренных (если шагов меньше k, процедура продолжается с максимального из непросмотренных элементов).*

Тогда оценка (2) будет верна с вероятностью

$$1 - 2\alpha n^{-k} - (4\gamma(1-\gamma))^{\frac{n}{2}} - \left(\frac{c_0 \ln n + 2}{n}\right)^k,$$

где

$$\gamma = \max(\beta, \beta_u) + \frac{8\varepsilon}{c_0} \cdot \frac{2\left(n - 2 - 2\sqrt{k\,(n-2)\ln n}\right)}{\pi n},$$

c_0 *- произвольная константа.*

Таким образом, каждый шаг алгоритма будет уменьшать вероятность ошибки почти в n раз, если константу c_0 выбрать достаточно большой (порядка $\frac{\varepsilon}{1/2-\beta}$).

Мы для простоты взяли квадратную матрицу, однако утверждение легко обобщается и на случай $m \neq n$: для этого нужно вместо просто умножения на 2 взять отдельно слагаемые для u и v. Шаги тоже нужно будет рассматривать отдельно для строк и столбцов из-за их разных размеров.

Полученные оценки показывают, что процедура maxvol в важном частном случае одноранговoго приближения часто находит элемент матрицы, близкий к максимальному, что гарантирует высокую точность крестового приближения. И чем более случайной является матрица, тем более вероятным по сравнению с худшим случаем становится положительный исход алгоритма.

Литература

1. *Oseledets I.V., Tyrtyshnikov E.E.* TT-cross approximation for multidimensional arrays // Linear Algebra and Its Applications. — 2010. — Vol. 432. — P. 70-78

2. *Goreinov S. A., Tyrtyshnikov E. E.* Quasioptimality of skeleton approximation of a matrix in the chebyshev norm // Doklady Mathematics. — 2011. — V. 83, N 3. — P. 1–2.

3. *Goreinov S. A., Tyrtyshnikov E. E.* The maximal-volume concept in approximation by low-rank matrices // Contemporary Mathematics. — 2001. — V. 268. — P. 47–51.

4. *Замарашкин Н. Л., Осинский А. И.* Новые оценки точности псевдоскелетных аппроксимаций матриц // Доклады академии наук. Математика. — 2016. — Т. 471, № 3. — С. 1-4.

5. *Goreinov S.A.* [*et al.*] How to find a good submatrix // Matrix Methods: Theory, Algorithms, Applications. — Word Scientific. — 2010. — P. 247–256.