

**ОБЩИЕ  
ЧИСЛЕННЫЕ МЕТОДЫ**

УДК 519.6

**НОВЫЕ ПРИЛОЖЕНИЯ МАТРИЧНЫХ МЕТОДОВ<sup>1)</sup>**

© 2021 г. Н. Л. Замарашкин<sup>1,\*</sup>, И. В. Оселедец<sup>1,2,\*\*</sup>, Е. Е. Тыртышников<sup>1,\*\*\*</sup>

<sup>1</sup> 119333 Москва, ул. Губкина, 8, Институт вычислительной математики  
им. Г.И. Марчука РАН, Россия

<sup>2</sup> 121205 Москва, Большой бульвар, 30, с. 1, Сколтех, Россия

\*e-mail: nikolai.zamarashkin@gmail.com

\*\*e-mail: ivan.oseledets@gmail.com

\*\*\*e-mail: eugene.tyrtysnikov@gmail.com

Поступила в редакцию 24.11.2020 г.  
Переработанный вариант 24.11.2020 г.  
Принята к публикации 14.01.2021 г.

Представлен очерк современных направлений развития матричных методов и их приложений, отраженных в работах данного тематического выпуска журнала. Особое внимание уделяется методам, связанным с идеей разделения переменных, реализующим ее специальным разложением матриц и тензоров, основанным на них алгоритмам и их применениям при решении многомерных задач вычислительной математики, анализа данных и машинного обучения. Библ. 32.

**Ключевые слова:** матрицы малого ранга, тензорные разложения, машинное обучение.

**DOI:** 10.31857/S0044466921050197

**1. МАЛОПАРАМЕТРИЧЕСКИЕ ПРЕДСТАВЛЕНИЯ И ПРИБЛИЖЕНИЯ**

Основное направление развития вычислительной математики и прикладного анализа данных в последние десятилетия определяется все увеличивающимся размером данных, с которыми работают алгоритмы. Сама возможность использования таких данных в численных расчетах напрямую зависит от того, найдется ли в них удобная для вычислений структура. С математической точки зрения наличие в данных скрытой структуры означает, что данные описываются моделью с относительно небольшим, приемлемым числом параметров. Естественно предположить, что для оценки параметров в малопараметрическом представлении достаточно знать не все данные, а их незначительную часть. Эта общая идея лежит в основе многих современных алгоритмов.

Анализ существующих приложений показывает, что малопараметрические приближения для матриц больших размеров основаны главным образом на том, что эти матрицы оказываются близкими к матрицам малого ранга. Недавний цикл работ (см. [1]–[3]) можно интерпретировать как попытку осмысления этого весьма общего явления.

Интерес к эффективным алгоритмам представления или приближения матриц матрицами малого ранга с использованием всех или лишь небольшого числа элементов не ослабевает вот уже более 30 лет. Среди разных подходов заметное место занимают *крестовые* приближения, которые строятся по небольшому числу столбцов и строк матрицы. Пусть  $m \times n$ -матрица

$$A = R + F$$

является суммой матрицы  $R$  ранга  $r$  и матрицы возмущения  $F$ , которое по спектральной норме не превосходит  $\epsilon$ . В [4], [5] было показано, что если столбцы  $m \times r$ -матрицы  $C$  и строки  $r \times n$ -матрицы  $R$  выбраны из  $A$  таким образом, что  $r \times r$ -подматрица  $\hat{A}$  на их пересечении имеет наибольший объем (модуль определителя) среди всех подматриц порядка  $r$ , то выполняется неравенство

$$\|A - C\hat{A}^{-1}R\|_C \leq (r + 1)\sigma_{r+1}(A) \leq (r + 1)\epsilon, \quad (1)$$

<sup>1)</sup>Работа выполнена при поддержке Московского центра фундаментальной и прикладной математики (соглашение 075-15-2019-1624 с Минобрнауки РФ).

где  $\|\cdot\|_C$  – поэлементная (чебышёвская) норма, а  $\sigma_{r+1}(A)$  – сингулярное число матрицы  $A$ , которое при их упорядочивании по невозрастанию от старшего к младшему имеет номер  $r + 1$ . Приближения вида  $CGR$  (иногда пишут  $CUR$ ) принято называть *крестовыми*, так как они строятся по элементам некоторого креста, составленного из столбцов и строк заданной матрицы.

Крестовые методы для получения тензорных разложений в формате тензорного произведения впервые были предложены в [6]. В работе [7] данного номера получены новые оценки ТТ-рангов приближенных тензоризаций массивов значений некоторых гладких функций.

Достоинство крестовых приближений заключается в том, что для их построения нужно знать лишь малое число элементов матрицы. Но есть и недостаток, связанный непосредственно с оценкой (1): ее прямое использование в матричных нормах приводит к множителям, зависящим от размеров матрицы. Это обстоятельство несколько ограничивало область применения данного подхода в алгоритмах анализа данных. В то же время активно применялись вероятностные методы построения весьма эффективных алгоритмов приближений вида  $CGR$  (см. [8]–[10]). Во многих интересных случаях рандомизированные алгоритмы обладали существенно лучшими оценками точности, но в них, однако, требовалось знать все элементы приближаемой матрицы.

Статья [11] в данном номере рассматривает оценки точности крестовых приближений на основе принципа обобщенного максимального объема в среднем. В ней рассматриваются наиболее употребительные матричные нормы, а полученные результаты подтверждают экспериментально наблюдаемый факт, что точность крестового метода сравнима с точностью рандомизированных алгоритмов. Это означает, что крестовый метод сохраняет свое главное преимущество – возможность строить высокоточные приближения на основе весьма лапидарной информации о матрице.

## 2. ЗАДАЧИ ВОСПОЛНЕНИЯ

Задача восполнения малоранговых матриц отличается от задачи приближения только в способе выбора элементов. Пусть матрица  $A \in \mathbb{R}^{n \times n}$  (для упрощения обозначений рассмотрим случай квадратных матриц) допускает представление  $A = R + F$  с матрицей  $R$  “малого” ранга  $r$  и “малым” по норме возмущением  $F$ . Допустим, что множество известных элементов  $A$  мало и “равномерно” распределяется в  $A$ . Можно ли восстановить матрицу  $A$  по некоторому заданному набору ее элементов?

Первоначальной мотивацией для задачи восполнения матриц послужили приложения, связанные с анализом данных: рекомендательные системы, конкурсы типа “Netflix prize”, совместная фильтрация (collabrative filtering) и др. К настоящему времени спектр приложений стал практически необозримым. Более того, восполнение матриц малого ранга является замечательным примером вычислительной задачи, решение которой опирается на глубокие теоретические результаты из разных областей математики.

В [12], [13] доказано, что для успешного восполнения число необходимых элементов не может быть меньше  $\mathcal{O}(n \log n)$ . Этими же авторами доказано, что при необременительных ограничениях  $\mathcal{O}(n \log^2(n))$  элементов достаточно для восполнения. Предложенная конструкция использует идеи релаксации задачи к выпуклой постановке с последующим применением известных методов оптимизации. Алгоритмическая сложность методов выпуклой оптимизации мотивировала дальнейшие исследования по поиску более быстрых вычислительных процедур. В результате появились итерационные алгоритмы SVT (Singular Value Thresholding) (см. [14]) и SVP (Singular Value Projection) (см. [15]). Каждая итерация этих алгоритмов состоит из двух шагов. На первом шаге применяется безусловный градиентный метод, а на втором новое приближение проектируется на многообразии матриц ранга  $r$ . Практическая применимость этих методов ограничивается только сложностью шага проектирования, который использует сингулярное разложение матриц.

В работе [16], представленной в данном номере, предлагается на каждом шаге проектирования заменить наилучшую проекцию на приближенную, сложность вычисления которой существенно меньше сложности сингулярного разложения. Для этого применяются крестовые приближения на основе принципа максимального объема. Важно заметить, что для SVP-алгоритма высокая точность приближений малого ранга во фробениусовой норме является критически важным свойством, определяющим сходимость метода. В теоретической части работы показано, что свойства сходимости методов изменятся незначительно. В качестве быстрых приближений

использовались крестовое приближение на основе принципа обобщенного максимального объема (см. [17]) и вероятностные алгоритмы малоранговых приближений (см. [18]). В вычислительных экспериментах на матрицах порядка 1000 достигалось ускорение в сотни раз. Особую ценность данной работе дает описание важных деталей программной реализации алгоритма, таких как адаптивный выбор шага в градиентном методе и адаптивная процедура набора ранга. Без них SVP-алгоритм существенно проигрывает как в скорости вычислений, так и в устойчивости.

В приложениях часто есть дополнительная информация о выполняемой малоранговой матрице (см. [19], [20]), которую было бы полезно учесть при разработке методов выполнения. Например, могут быть известны подпространства, в которых лежат линейные оболочки строк и столбцов. Такие задачи называют задачами *выполнения со сторонней информацией* (side information). Пусть  $d_1$  и  $d_2$  – размерности подпространств, которые содержат пространства строк и столбцов соответственно (очевидно, что  $r \leq \min\{d_1, d_2\}$ ). Будем считать, что базисы пространств задаются матрицами  $X \in \mathbb{R}^{m \times d_1}$  и  $Y \in \mathbb{R}^{d_2 \times n}$ . В этом случае число элементов, необходимое для восстановления, имеет вид  $\mathcal{O}(\log(N))$  (см. [21]). Запишем задачу выполнения в виде  $A = XYU$ , где  $A$  – выполняемая матрица, а  $U \in \mathbb{R}^{d_1 \times d_2}$  – искомая матрица предсказательной модели. В работе [22] данного номера матрица  $U$  представляется в факторизованном виде  $U = UV$  с матрицами  $U \in \mathbb{R}^{d_1 \times r}$  и  $V \in \mathbb{R}^{r \times d_2}$ , разреженными по строкам и столбцам соответственно. Предложена модификация алгоритма выполнения, учитывающая разреженную структуру факторов. При определенных условиях новый алгоритм позволяет дополнительно снизить требования к числу известных элементов. Последнее является критически важным для приложений, где получение необходимой информации затруднено.

### 3. МАТРИЧНЫЕ И ТЕНЗОРНЫЕ МЕТОДЫ В МАШИННОМ ОБУЧЕНИИ

Подавляющее большинство современных моделей машинного обучения основано на глубоких искусственных нейросетях, которые представляют собой композицию линейных преобразований и поточечных нелинейных преобразований. Линейные преобразования параметризуются матрицами (“весами”), что дает прямую связь с матричным анализом. В частности, можно использовать методы малоранговой аппроксимации матриц и тензоров для сжатия моделей машинного обучения: по заданным весам строятся их аппроксимации, и возникает новая модель, которая содержит меньшее число параметров, но при этом приближает предыдущую. Даже если точности такой сжатой модели недостаточно, то ее можно *дообучить*, используя полученные с помощью матричных и/или тензорных разложений представления. Такой подход стал очень популярным. В данном выпуске тематике сжатия посвящена статья [23], в ней предложена идея использования аппроксимации не параметров модели, а промежуточных значений – активаций. Оказывается, что используя скелетное разложение матриц, составленное из векторов активаций, можно существенно сжать различные модели машинного обучения. Следует отметить, что матричные и тензорные методы сжатия нейросетевых моделей (они называются “факторизационными”) стали отдельным направлением в построении быстрых и компактных моделей машинного обучения и уже используются в коммерческих программных пакетах.

Важным направлением развития матричных и тензорных методов является решение задач оптимизации, где неизвестные естественным образом представляются в виде двумерных или многомерных массивов. Такие оптимизационные задачи не всегда сводятся к классическим, но могут быть решены с использованием специальных оптимизационных методов, где на каждом шаге необходимо использовать эффективные и устойчивые матричные алгоритмы. В частности, подобные постановки возникают в рекомендательных системах, анализе социальных сетей и анализе естественных языков. Иногда даже возникает необходимость по-новому взглянуть на классические алгоритмы матричного анализа с точки зрения методов оптимизации. В данном выпуске этой тематике посвящена статья [22]. В [24] тензорные разложения используются для постановки модели о разделении смесей. Классических тензорных разложений в этом случае оказывается недостаточно, и вводится новое представление. В этом случае задача исследователя состоит в подборе правильного представления, или разработке новой модели факторизации входного тензора, и решении возникающей задачи оптимизации.

Таким образом, матричные методы активно развиваются, применяются в машинном обучении как для создания новых моделей, так и для сжатия существующих. Устойчивые матричные алгоритмы дают основу для более устойчивых методов машинного обучения (в частности, ис-

пользование ортогональных матриц позволяет в ряде случаев решить проблему “затухающего градиента”). При этом высокая надежность и скорость существующих матричных и тензорных разложений позволяют успешно использовать их для решения целого ряда многомерных задач. Интересным направлением видится также развитие новых оптимизационных методов для решения задач с ограничениями на свойства неизвестных матриц и тензоров, в том числе, с использованием методов стохастической оптимизации и теории игр.

#### 4. ДРУГИЕ ПРИЛОЖЕНИЯ И ЗАКЛЮЧЕНИЕ

Традиционные применения матричных методов связаны с решением уравнений математической физики. Это могут быть как сложные многомерные уравнения, такие как уравнение Беллмана (тензорный метод для его решения рассмотрен в [31]), так и решение классических задач, например, электростатики многочастичных систем, с существенно более высокой скоростью (см. [32]). Статья [25] данного номера посвящена весьма актуальной проблеме численного решения объемных интегральных уравнений на неравномерных сетках. При дискретизации таких задач возникают плотные матрицы огромных размеров, которые, однако, обладают скрытой структурой — показано, что они допускают приближенную факторизацию в виде произведения разреженных матриц и многоуровневой тёплицевой матрицы. Этот факт позволяет эффективно выполнять умножение матрицы на вектор и применять итерационные методы. В [26] изучаются алгебры, появившиеся в результате исследования различных обобщений специфики тёплицевых матриц.

В целом ряде задач требуется вычисление матричной экспоненты. В статье [27] данного тематического номера рассматриваются некоторые вопросы, связанные с применением подпространств Крылова при вычислении экспоненты от несимметричной матрицы. В [28] предлагается новый алгоритм вычисления собственных векторов несимметричных трехдиагональных матриц. В [29] предлагаются и исследуются новые алгоритмы для решения нелинейной проблемы собственных значений. В [30] дается полезный обзор методов экстраполяции Шэнкса и их приложений к ускорению итерационных процессов.

Методы матричного анализа и прикладной линейной алгебры имеют огромное значение для развития наук и технологий. Они обсуждаются на многочисленных семинарах и являются основной темой некоторых серийных конференций. В России — это, прежде всего, регулярная международная конференция “Матричные методы в математике и приложениях”, которая обычно проводится в Москве на базе Института вычислительной математики им. Г.И. Марчука РАН и Сколковского университета науки и технологий. Часть результатов, представленных в данном тематическом выпуске, были анонсированы в докладах 5-й конференции этой серии, состоявшейся в августе 2019 г.

Авторы этого очерка и одновременно редакторы данного тематического выпуска выражают особую благодарность Сергею Александровичу Матвееву, взявшему на себя основную часть труда по его подготовке.

#### СПИСОК ЛИТЕРАТУРЫ

1. *Udell M., Townsend A.* Why are big data matrices approximately low rank? // *SIAM J. Math. Data Sci.* 2019. V. 1. No. 1. P. 144–160.
2. *Beckermann B., Townsend A.* On the singular values of matrices with displacement structure // *SIAM J. Matrix Anal. Appl.* 2017. V. 38. No. 4. P. 1227–1248.
3. *Townsend A., Wilber H.* Near-Optimal Column-Based Matrix Reconstruction // *Linear Algebra Appl.* 2018. V. 548. P. 19–41.
4. *Goreinov S., Tyrtyshnikov E., Zamarashkin N.* A theory of pseudoskeleton approximations // *Linear Algebra Appl.* 1997. V. 261. No. 1–3. P. 19–41.
5. *Goreinov S., Tyrtyshnikov E.* The maximal-volume concept in approximation by low-rank matrices // *Contemporary Math.* 2001. V. 280. P. 47–52.
6. *Oseledets I., Tyrtyshnikov E.* TT-cross approximation for multidimensional arrays // *Linear Algebra Appl.* 2010. V. 432. P. 70–88.
7. *Высоцкий Л.И.* О ТТ-рангах приближенных тензоризаций некоторых гладких функций // *Ж. вычисл. матем. и матем. физ.* 2021. Т. 61. № 5.
8. *Boutsidis C., Drienas P., Magdon-Ismail M.*, Near-Optimal Column-Based Matrix Reconstruction // *SIAM J. Comput.* 2014. V. 43. No. 2. P. 183–202.

9. *Boutsidis C., Woodruff D.P.* Optimal CUR matrix decompositions // *Proceed. 46th Ann. ACM Symp. Theory Comput.* 2014. P. 353–362.
10. *Deshpande A., Rademacher L.* Efficient volume sampling for row/column subset selection // *51st Ann. Symp. Foundat. Comput. Sci.* 2010. P. 329–338.
11. *Замарашкин Н.Л., Осинский А.И.* О точности крестовых и столбцовых малоранговых MAXVOL-приближений в среднем // *Ж. вычисл. матем. и матем. физ.* 2021. Т. 61. № 5. 09. V. 2. No. 1. P. 183–202.
12. *Candes E.J., Tao T.* The Power of Convex Relaxation: Near-Optimal Matrix Completion // *IEEE Transact. Inform. Theory.* 2009. V. 56. No. 5. P. 2053–2080.
13. *Recht B.* A Simpler Approach to Matrix Completion // *J. Machine Learn. Res.* 2011. V. 12. P. 3413–3430.
14. *Cai J.-F., Candes E.J., Z. Shen Z.* A Singular Value Thresholding Algorithm for Matrix Completion // *SIAM J. Optimizat.* 2010. V. 20. No. 4. P. 1956–1982.
15. *Meka R., Jain P., Dhillon I.S.,* Guaranteed Rank Minimization via Singular Value Projection // *Proceed. 23rd Inter. Conf. Neural Informat. Proc. Syst.* 2010. V. 1. No. 1–3. P. 937–945.
16. *Лебедева О.С., Осинский А.И., Петров С.В.* Приближенные алгоритмы малоранговой аппроксимации в задаче выполнения матрицы на случайном шаблоне // *Ж. вычисл. матем. и матем. физ.* 2021. Т. 61. № 5.
17. *Osinsky A., Zamarashkin N.* Pseudoskeleton approximations with better accuracy estimates // *Linear Algebra Applicat.* 2018. V. 537. P. 221–249.
18. *Tropp J.A., Halko N., Martinsson P.G.* Finding structures with randomness: probabilistic algorithms for constructing approximate matrix decompositions // *SIAM Rev.* 2011. V. 53. No. 2. P. 217–288.
19. *Guo Y.* Convex Co-Embedding for Matrix Completion with Predictive side information // *Proceed. 31th AAAI Conf. Artific. Intelligence Symp. Theory Comput.* 2017.
20. *Wang H., Wei Y., Cao M., Xu M., Wu W., Xing E.P.* Deep Inductive Matrix Completion for Biomedical Interaction Prediction // *IEEE Inter. Conf. Bioinformatics and Biomedicine (BIBM).* 2019. P. 520–527.
21. *Xu M., Jin R., Zhou Z.-H.* Speedup matrix completion with side information: Application to multi-label learning // *Adv. Neural Informat. Proc. Syst.* 2013. P. 2301–2309.
22. *Буркина М., Назаров И., Панов М., Федонин Г., Широких Б.* Индуктивное восстановление матриц с отбором признаков // *Ж. вычисл. матем. и матем. физ.* 2021. Т. 61. № 5.
23. *Гусак Ю.В., Даулбаев Т.К., Оселедец И.В., Пономарев Е.С., Чихоцкий А.С.* Малоранговое представление нейронных сетей // *Ж. вычисл. матем. и матем. физ.* 2021. Т. 61. № 5.
24. *Оселедец И.В., Харюк П.В.* Моделирование структуры данных с помощью блочного тензорного разложения: разложение объединенных тензоров и вариационное блочное тензорное разложение как параметризованная модель смесей // *Ж. вычисл. матем. и матем. физ.* 2021. Т. 61. № 5.
25. *Самохин А.Б., Тыртышников Е.Е.* Численный метод решения объемных интегральных уравнений на неравномерной сетке // *Ж. вычисл. матем. и матем. физ.* 2021. Т. 61. № 5.
26. *Боццо Э., Диедда П., ди Фиоре К.* Замкнутые относительно  $J$ -сопряжения алгебры и формулы смещения // *Ж. вычисл. матем. и матем. физ.* 2021. Т. 61. № 5.
27. *Бочев М.А.* Точный перезапуск метода подпространства Крылова “сдвиг–обращение” для вычисления действия экспоненты несимметричных матриц // *Ж. вычисл. матем. и матем. физ.* 2021. Т. 61. № 5.
28. *Ван Дорен П., Лаудаддио Т., Матронарди Н.* Вычисление собственных векторов несимметричных трехдиагональных матриц // *Ж. вычисл. матем. и матем. физ.* 2021. Т. 61. № 5.
29. *Гандер В.* Новые алгоритмы для решения нелинейной проблемы собственных значений // *Ж. вычисл. матем. и матем. физ.* 2021. Т. 61. № 5.
30. *Брезински К., Редиво-Дзалья М.* Методы экстраполяции Шэнкса и их приложения // *Ж. вычисл. матем. и матем. физ.* 2021. Т. 61. № 5.
31. *Бойко А.И., Оселедец И.В., Феррер Г.* T-QI: ускоренная итерация функции ценности в формате тензорного поезда для задач стохастического оптимального управления // *Ж. вычисл. матем. и матем. физ.* 2021. Т. 61. № 5.
32. *Хоромская В.Х., Хоромский Б.Н.* Перспективы численного моделирования с использованием тензорных разложений для моделирования электростатики в многочастичных системах // *Ж. вычисл. матем. и матем. физ.* 2021. Т. 61. № 5.