

---

GENERAL  
NUMERICAL METHODS

---

## New Applications of Matrix Methods

N. L. Zamarashkin<sup>a,\*</sup>, I. V. Oseledets<sup>a,b,\*\*</sup>, and E. E. Tyrtyshnikov<sup>a,\*\*\*</sup>

<sup>a</sup> Marchuk Institute of Numerical Mathematics, Russian Academy of Sciences, Moscow, 119333 Russia

<sup>b</sup> Skolkovo Institute of Science and Technology (Skoltech), Moscow, 121205 Russia

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

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

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

Received November 24, 2020; revised November 24, 2020; accepted January 14, 2021

**Abstract**—Modern directions in the development of matrix methods and their applications described in the present issue are overviewed. Special attention is given to methods associated with separation of variables, special decompositions of matrices and tensors implementing this technique, related algorithms, and their applications to multidimensional problems in computational mathematics, data analysis, and machine learning.

**Keywords:** low-rank matrices, tensor decompositions, machine learning

**DOI:** 10.1134/S0965542521050183

### 1. LOW-PARAMETRIC REPRESENTATIONS AND APPROXIMATIONS

The basic direction of the development of computational mathematics and applied data analysis over last decades is associated with progressively increasing sizes of data operated by algorithms. The possibility of using such data in numerical computations depends directly on whether they have a structure convenient for computations. Mathematically, the presence of a hidden structure in data means that they are described by a model with a relatively small acceptable number of parameters. It is natural to assume that knowledge of a small portion of data, rather than all of them, is sufficient to estimate parameters in a low-parametric representation. This general idea underlies numerous modern algorithms.

An analysis of available applications shows that low-parametric approximations for large-sized matrices are based primarily on the fact that they are close to low-rank matrices. A recent series of works (see [1–3]) can be regarded as an attempt of interpreting this general phenomenon.

Interest in efficient algorithms for representing or approximating matrices by low-rank ones with the use of all or only a few entries has been persisted for more than 30 years. Among the variety of approaches, a noticeable place is occupied by *cross* approximations, which are constructed using a small number of matrix columns and rows. Suppose that an  $m \times n$  matrix  $A$  can be represented in the form

$$A = R + F,$$

where  $R$  is a matrix of rank  $r$  and  $F$  is a perturbation matrix that does not exceed  $\varepsilon$  in the spectral norm. In [4, 5] it was shown that, if the columns of an  $m \times r$  matrix  $C$  and the rows of an  $r \times n$  matrix  $R$  are chosen from  $A$  in such a way that the  $r \times r$  submatrix  $\hat{A}$  obtained at their intersection has the largest volume (the absolute value of the determinant) among all submatrices of order  $r$ , then

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

where  $\|\cdot\|_C$  is a step-by-step (Chebyshev) norm and  $\sigma_{r+1}(A)$  is a singular value of  $A$  indexed by  $r + 1$  when all singular values are arranged in nonincreasing order. Approximations of the form  $CGR$  (sometimes written as  $CUR$ ) are known as *cross* approximations, since they are constructed using entries of a cross consisting of columns and rows of a given matrix.

Cross approximation methods for obtaining tensor decompositions in the form of tensor trains were first proposed in [6]. New estimates for TT ranks of approximate tensorizations of some smooth functions were obtained in [7] (see the present issue).

An advantage of cross approximations is that they can be constructed using a small number of matrix entries. However, there is a disadvantage associated directly with estimate (1): its direct use in matrix norms leads to multipliers depending on the matrix size. To a certain degree, this shortcoming limited the application of this approach in data analysis algorithms. At the same time, probabilistic methods were extensively used to construct fairly efficient *CGR* approximation algorithms (see [8–10]). In many interesting, randomized algorithms had much better accuracy estimates, but they required knowledge of all entries of the approximated matrix.

Accuracy bounds for cross approximations are examined using the generalized maximum volume principle on average are addressed in [11]. Specifically, the most common matrix norms are considered, and the results confirm the experimentally observed fact that the accuracy of the cross approximation method is comparable with the accuracy of randomized algorithms. This means that the cross approximation method preserves its main advantage, i.e., the capability of constructing high-order accurate approximations based on rather scarce information on a matrix.

## 2. MATRIX COMPLETION PROBLEMS

The low-rank matrix completion (MC) problem differs from the approximation problem only in the entry-choosing method. Suppose that a matrix  $A \in \mathbb{R}^{n \times n}$  (the case of square matrices is considered for notational simplicity) can be represented in the form  $A = R + F$ , where  $R$  is a matrix of “low” rank  $r$  and  $F$  is a perturbation “small” in the norm. Assume that the known entries of  $A$  make up a small set and are “uniformly” distributed in  $A$ . Can  $A$  be recovered from a given set of its entries?

Initially, interest in MC was motivated by applications associated with data analysis: recommender systems, Netflix prize competitions, collaborative filtering, etc. To date, the range of applications has become immense. Moreover, low-rank MC is a remarkable example of a computational problem whose solution relies on profound theoretical results from different areas of mathematics.

It is proved in [12, 13] that, for successful MC, the number of known entries has to be at least  $\mathcal{O}(n \log n)$ . Additionally, it is proved that, under rather mild constraints,  $\mathcal{O}(n \log^2(n))$  entries are sufficient for MC. The proposed construction relies on the idea of problem relaxation to a convex formulation with the subsequent application of well-known optimization methods. The high algorithmic complexity of convex optimization methods motivated further investigations concerning the search for faster numerical procedures. As a result, iterative algorithms, such as singular value thresholding (SVT) [14] and singular value projection (SVP) [15] were developed. Every iteration of these algorithms consists of two steps. An unconstrained gradient method is applied at the first step, and the new approximation is projected onto the manifold of rank- $r$  matrices at the second step. The practical applicability of these methods is limited only by the complexity of the projection step, which involves singular value decompositions of matrices.

In [16], it is suggested that, at every projection step, the best projection be replaced by an approximate one, whose computational complexity is much lower than the complexity of singular value decomposition. Cross approximations based on the maximum volume principle are applied for this purpose. Importantly, for the SVP algorithm, the high accuracy of low-rank approximations in the Frobenius norm is a critically important property determining the convergence of the method. In the theoretical part of [16], it is shown that the convergence properties of the methods change insignificantly. As fast approximations, one can use a cross approximation based on the generalized maximum volume principle (see [17]) and probabilistic low-rank approximation algorithms (see [18]). A speedup of several hundred times is achieved in numerical experiments on matrices of order 1000. A point of special value is that [16] provides important details concerning a software implementation of the algorithm, such as adaptive stepsize selection in the gradient method and an adaptive procedure for rank increase. Without these procedures, the SVP algorithm is significantly outperformed in terms of the speed of computations and stability.

In applications, additional information on the low-rank matrix to be recovered is often available (see [19, 20]), which would be useful in developing MC methods. For example, subspaces containing the linear spans of rows and columns can be known. Such problems are called *matrix completion with side information*. Let  $d_1$  and  $d_2$  be the dimensions of the subspaces containing the row and column spaces, respectively (obviously,  $r \leq \min\{d_1, d_2\}$ ). Assume that the bases of the spaces are specified by matrices  $X \in \mathbb{R}^{m \times d_1}$  and  $Y \in \mathbb{R}^{d_2 \times n}$ . In this case, the number of entries required for MC is  $\mathcal{O}(\log(N))$  (see [21]). The MC problem is written as  $A = XWY$ , where  $A$  is the matrix to be recovered and  $W \in \mathbb{R}^{d_1 \times d_2}$  is the sought matrix of the predictive model. In [22]  $W$  is factorized in the form  $W = UV$  with matrices  $U \in \mathbb{R}^{d_1 \times r}$  and  $V \in \mathbb{R}^{r \times d_2}$ .

sparsified in rows and columns, respectively. A modification of the MC algorithm is proposed that takes into account the sparse structure of the factors. Under certain conditions, the new algorithm additionally reduces the requirement for the number of known entries, which is of critical importance in applications where the necessary information is difficult to obtain.

### 3. MATRIX AND TENSOR METHODS IN MACHINE LEARNING

The overwhelming majority of modern machine-learning models are based on deep artificial neural networks, which represent a composition of linear transformations and pointwise nonlinear transformations. Linear transformations are parametrized by matrices (weights), which are a direct link with matrix analysis. In particular, low-rank approximation methods for matrices and tensors can be used to compress machine learning models. Namely, given weights, the construction of their approximations yields a new model with fewer parameters that approximates the previous one. Even if the accuracy of the compressed model is insufficient, we can apply *transfer learning* using representations obtained with the help of matrix and/or tensor decompositions. This approach has become very popular. In this issue, compression is addressed in [23], where the idea of using approximations of intermediate values (i.e., activations) rather than model parameters is proposed. It turns out that various machine learning models can be significantly compressed by applying a rank factorization of the matrix made up of activation vectors. It should be noted that matrix and tensor methods for compression of neural network models (they are called “factorizational”) have become a separate direction in the design of fast and compact machine learning models and have been used in commercial software packages.

An important direction in the development of matrix and tensor methods is the solution of optimization problems in which the unknowns are naturally represented in the form of two- or multidimensional arrays. Such optimization problems sometimes fail to be reduced to classical problems, but they can be solved using special optimization methods with every step involving efficient and stable matrix algorithms. Specifically, such formulations arise in recommender systems, social network analysis, and natural language analysis. Sometimes, it becomes necessary to take a fresh look at classical matrix analysis algorithms in terms of optimization methods. In the present issue, this subject is considered in [22]. In [24] term decompositions are used to formulate a mixture separation model. Classical tensor decompositions are then insufficient, so a new representation is introduced. In this case, researcher’s task is to choose an appropriate representation, or to develop a new model of input tensor factorization and to solve the arising optimization problem.

Thus, matrix methods are under active development and are applied in machine learning for the design of new models and compression of available ones. Stable matrix algorithms provide a basis for more robust machine learning methods (specifically, in some cases, the use of orthogonal matrices makes it possible to cope with the vanishing gradient problem). Moreover, due to their high reliability and speed, existing matrix and tensor decompositions can be successfully applied to a number of multidimensional problems. An interesting direction seems to be the development of new optimization methods for solving problems with constraints on the properties of unknown matrices and tensors, including methods of stochastic optimization and game theory.

### 4. OTHER APPLICATIONS AND CONCLUSIONS

Traditional applications of matrix methods are associated with solving equations of mathematical physics. They include complicated multidimensional equations, such as the Bellman equation (a tensor method for its solution is considered in [31]) and classical problems, for example, electrostatics of multi-particle systems with a significantly higher velocity (see [32]). The important problem of the numerical solution of volume integral equations on nonuniform grids is considered in the present issue (see [25]). Discretization of such problems gives rise to dense matrices of huge size, which, however, have a hidden structure. It is shown that they admit an approximate factorization in the form of the product of sparse matrices and a multilevel Toeplitz matrix. As a result, efficient implementations of matrix-vector multiplication and iterative methods can be designed. Algebras appearing in the study of various generalizations of specific features of Toeplitz matrices are addressed in [26].

In a number of problems, it is necessary to compute matrix exponentials. Some issues concerning the application of Krylov subspaces in calculating the exponential of a nonsymmetric matrix are considered in [27]. A new algorithm for computing eigenvectors of asymmetric tridiagonal matrices is proposed in [28]. New algorithms for solving nonlinear eigenvalue problems are suggested and investigated in [29]. A useful survey of Shanks’ extrapolation methods and their applications to iteration acceleration are presented in [30].

Methods of matrix analysis and applied linear algebra play an important role in science and engineering. They are discussed at numerous seminars and some serial conferences. In Russia, they are addressed primarily at regular international conferences *Matrix Methods in Mathematics and Applications*, which are usually held in Moscow based on the Marchuk Institute of Numerical Mathematics of the Russian Academy of Sciences and the Skolkovo Institute of Science and Technology. Some of the results published in this issue were reported at the 5th conference of this series held in August 2019.

#### ACKNOWLEDGMENTS

This paper's authors, who are simultaneously the editors of this themed issue, are sincerely grateful to Sergey Aleksandrovich Matveev, who did most of the work in preparing this issue.

#### FUNDING

This work was supported by the Moscow Center for Fundamental and Applied Mathematics (contract no. 075-15-2019-1624 with the Ministry of Education and Science of the Russian Federation).

#### REFERENCES

1. M. Udell and A. Townsend, "Why are big data matrices approximately low rank?" *SIAM J. Math. Data Sci.* **1** (1), 144–160 (2019).
2. B. Beckermann and A. Townsend, "On the singular values of matrices with displacement structure," *SIAM J. Matrix Anal. Appl.* **38** (4), 1227–1248 (2017).
3. A. Townsend and H. Wilber, "Near-optimal column-based matrix reconstruction," *Linear Algebra Appl.* **548**, 19–41 (2018).
4. S. Goreinov, E. Tyrtyshnikov, and N. Zamarashkin, "A theory of pseudoskeleton approximations," *Linear Algebra Appl.* **261** (1–3), 19–41 (1997).
5. S. Goreinov and E. Tyrtyshnikov, "The maximal-volume concept in approximation by low-rank matrices," *Contemp. Math.* **280**, 47–52 (2001).
6. I. Oseledets and E. Tyrtyshnikov, "TT-cross approximation for multidimensional arrays," *Linear Algebra Appl.* **432**, 70–88 (2010).
7. L. I. Vysotsky, "TT ranks of approximate tensorizations of some smooth functions," *Comput. Math. Math. Phys.* **61** (5), 750–760 (2021).
8. C. Boutsidis, P. Drienas, and M. Magdon-Ismail, "Near-optimal column-based matrix reconstruction," *SIAM J. Comput.* **43** (2), 183–202 (2014).
9. C. Boutsidis and D. P. Woodruff, "Optimal CUR matrix decompositions," *Proceedings of the 46th Annual ACM Symposium on Theory of Computing* (2014), pp. 353–362.
10. A. Deshpande and L. Rademacher, "Efficient volume sampling for row/column subset selection," *IEEE 51st Annual Symposium on Foundations of Computer Science* (2010), pp. 329–338.
11. N. L. Zamarashkin and A. I. Osinsky, "On the accuracy of cross and column low-rank Maxvol approximations in average," *Comput. Math. Math. Phys.* **61** (5), 786–798 (2021).
12. E. J. Candes and T. Tao, "The power of convex relaxation: Near-optimal matrix completion," *IEEE Trans. Inf. Theory* **56** (5), 2053–2080 (2009).
13. B. Recht, "A simpler approach to matrix completion," *J. Mach. Learn. Res.* **12**, 3413–3430 (2011).
14. J.-F. Cai, E. J. Candes, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM J. Optim.* **20** (4), 1956–1982 (2010).
15. R. Meka, P. Jain, and I. S. Dhillon, "Guaranteed rank minimization via singular value projection," *Proceedings of the 23rd International Conference on Neural Information Processing Systems* (2010), Vol. 1, No. 1–3, pp. 937–945.
16. O. S. Lebedeva, A. I. Osinsky, and S. V. Petrov, "Low-rank approximation algorithms in the matrix completion problem on a random stencil," *Comput. Math. Math. Phys.* **61** (5), 799–815 (2021).
17. A. Osinsky and N. Zamarashkin, "Pseudoskeleton approximations with better accuracy estimates," *Linear Algebra Appl.* **537**, 221–249 (2018).
18. J. A. Tropp, N. Halko, and P. G. Martinsson, "Finding structures with randomness: Probabilistic algorithms for constructing approximate matrix decompositions," *SIAM Rev.* **53** (2), 217–288 (2011).
19. Y. Guo, "Convex co-embedding for matrix completion with predictive side information," *Proceedings of the 31th AAAI Conference on Artificial Intelligence* (2017).

20. H. Wang, Y. Wei, M. Cao, M. Xu, W. Wu, and E. P. Xing, “Deep inductive matrix completion for biomedical interaction prediction,” *IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* (2019), pp. 520–527.
21. M. Xu, R. Jin, and Z.-H. Zhou, “Speedup matrix completion with side information: Application to multi-label learning,” *Advances in Neural Information Processing Systems* (2013), pp. 2301–2309.
22. M. Burkina, I. Nazarov, M. Panov, G. Fedonin, and B. Shirokikh, “Inductive matrix completion with selection of features,” *Comput. Math. Math. Phys.* **61** (5), 719–732 (2021).
23. Yu. V. Gusak, T. K. Daulbaev, E. S. Ponomarev, A. S. Cichocki, and I. V. Oseledets, “Reduced-order modeling of deep neural networks,” *Comput. Math. Math. Phys.* **61** (5), 774–785 (2021).
24. I. V. Oseledets and P. V. Kharyuk, “Structuring data with block term decomposition: Decomposition of joint tensors and variational block term decomposition as a parametrized mixture distribution model,” *Comput. Math. Math. Phys.* **61** (5), 816–835 (2021).
25. A. B. Samokhin and E. E. Tyrtshnikov, “Numerical method for solving volume integral equations on a non-uniform grid,” *Comput. Math. Math. Phys.* **61** (5), 847–853 (2021).
26. E. Bozzo, P. Deidda, and C. Di Fiore, “Algebras closed by J-hermitianity in displacement formulas,” *Comput. Math. Math. Phys.* **61** (5), 674–683 (2021).
27. M. A. Botchev, “An accurate restarting for shift-and-invert Krylov subspaces computing matrix exponential actions of nonsymmetric matrices,” *Comput. Math. Math. Phys.* **61** (5), 684–698 (2021).
28. P. Van Dooren, T. Laudadio, and N. Mastronardi, “Computing the eigenvectors of nonsymmetric tridiagonal matrices,” *Comput. Math. Math. Phys.* **61** (5), 733–749 (2021).
29. W. Gander, “New algorithms for solving nonlinear eigenvalue problems,” *Comput. Math. Math. Phys.* **61** (5), 761–773 (2021).
30. C. Brezinski and M. Redivo-Zaglia, “A survey of Shanks’ extrapolation methods and their applications,” *Comput. Math. Math. Phys.* **61** (5), 699–718 (2021).
31. A. I. Boyko, I. V. Oseledets, and G. Ferrer, “TT-QI: Faster value iteration in tensor train format for stochastic optimal control,” *Comput. Math. Math. Phys.* **61** (5), 836–846 (2021).
32. V. Kh. Khoromskaia and B. N. Khoromskij, “Prospects of tensor-based numerical modeling of collective electrostatics in many-particle systems,” *Comput. Math. Math. Phys.* **61** (5), 864–886 (2021).

*Translated by I. Ruzanova*