A fast algorithm for low-rank approximation with error control

Yuji Nakatsukasa

Abstract

Computing a low-rank approximation to a large $m \times n$ matrix A is a ubiquitous task in Numerical Linear Algebra (NLA), and possibly the single topic that contributed the most to making randomized NLA algorithms popular, trusted, and widely used. Typically [1, 5], the first step is to compute a random sketch of the form AS (or $\hat{S}A$, or both [12]), where the size of the sketch is at least the target rank, which is often unknown. Extensive theory is now available [5, 8, 11] that gives strong guarantees for the quality of the resulting approximation that hold with extremely high probability.

In this work we develop an algorithm for low-rank approximation that (i) requires only an $O(1)$ sketch size, (ii) comes with high-probability error control to achieve a user-defined error tolerance, without requiring the knowledge of the rank, (iii) avoids computing orthogonal projections, and (iv) is based on the CUR decomposition [6] and its stable implementation [10], so inherits properties of A such as sparsity and nonnegativity, if present. These are achieved by bringing together techniques in randomized NLA algorithms, including CUR, subset selection methods [2, 9] based on a sketch-and-pivot strategy [3, 4], and error estimation via trace estimation [7].

The algorithm finds a near-optimal (up to a modest polynomial in r) rank-r approximation in $O(N + (m + n)r^2)$ operations, where N is the cost of a matrix-vector multiplication with A. Advantages over the MATLAB routine `svdsketch` [13] include faster runtime and the ability to set the error tolerance to be smaller than $\sqrt{u}$, where u is the unit roundoff.

This talk is based on joint projects with the following collaborators: Per-Gunnar Martinsson and Nathaniel Pritchard; Anjali Narendran; and Taejun Park.

References

[1] K. L. Clarkson and D. P. Woodruff. Low-rank approximation and regression in input sparsity time. *Journal of the ACM*, 63(6):54, 2017.

[2] A. Cortinovis and D. Kressner. Low-rank approximation in the frobenius norm by column and row subset selection. *SIAM J. Matrix Anal. Appl.*, 41(4):1651–1673, 2020.

[3] Y. Dong and P.-G. Martinsson. Simpler is better: a comparative study of randomized pivoting algorithms for CUR and interpolative decompositions. *Adv. in Comput. Math.*, 49(4):66, 2023.

[4] J. A. Duersch and M. Gu. Randomized projection for rank-revealing matrix factorizations and low-rank approximations. *SIAM Rev.*, 62(3):661–682, 2020.

[5] N. Halko, P.-G. Martinsson, and J. A. Tropp. Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions. *SIAM Rev.*, 53(2):217–288, 2011.

[6] M. W. Mahoney and P. Drineas. CUR matrix decompositions for improved data analysis. *Proc. Natl. Acad. Sci.*, 106(3):697–702, 2009.

[7] P.-G. Martinsson and J. A. Tropp. Randomized numerical linear algebra: Foundations and algorithms. *Acta Numer.*, pages 403—572, 2020.

[8] Y. Nakatsukasa. Fast and stable randomized low-rank matrix approximation. *arXiv preprint arXiv:2009.11392*, 2020.

[9] A. Osinsky. Close to optimal column approximations with a single SVD. *arXiv preprint arXiv:2308.09068*, 2023.

[10] T. Park and Y. Nakatsukasa. Accuracy and stability of CUR decompositions with oversampling. *arXiv preprint arXiv:2405.06375*, 2024.

[11] J. A. Tropp, A. Yurtsever, M. Udell, and V. Cevher. Practical sketching algorithms for low-rank matrix approximation. *SIAM J. Matrix Anal. Appl.*, 38(4):1454–1485, 2017.

[12] F. Woolfe, E. Liberty, V. Rokhlin, and M. Tygert. A fast randomized algorithm for the approximation of matrices. *Appl. Comput. Harmon. Anal.*, 25(3):335–366, 2008.

[13] W. Yu, Y. Gu, and Y. Li. Efficient randomized algorithms for the fixed-precision low-rank matrix approximation. *SIAM J. Matrix Anal. Appl.*, 39(3):1339–1359, 2018.