Randomized Nyström approximation of non-negative self-adjoint operators

David Persson, Nicolas Boullé, Daniel Kressner

Abstract

A ubiquitous task in numerical linear algebra is to compute a low-rank approximation to a matrix $\boldsymbol{A}$. Randomized techniques [8, 9, 10, 12] are becoming increasingly popular for computing cheap, yet accurate, low-rank approximations to matrices. Most notably, the *randomized singular value decomposition* (SVD) [9] has evolved into one of the primary choices, due to its simplicity, performance, and reliability. In its most basic form, the randomized SVD performs the approximation $\boldsymbol{Q}\boldsymbol{Q}^*\boldsymbol{A} \approx \boldsymbol{A}$, where $\boldsymbol{Q}$ is an orthonormal basis for the range of $\boldsymbol{A}\boldsymbol{\Omega}$, with $\boldsymbol{\Omega}$ being a tall and skinny random sketch matrix. In many applications of low-rank approximation, such as k-means clustering [13], PCA [14], and Gaussian process regression [7], it is known that $\boldsymbol{A}$ is symmetric positive semi-definite. In this case, one usually prefers the so-called *randomized Nyström approximation* [8]

$$\widehat{\boldsymbol{A}} := \boldsymbol{A}\boldsymbol{\Omega}(\boldsymbol{\Omega}^*\boldsymbol{A}\boldsymbol{\Omega})^\dagger\boldsymbol{\Omega}^*\boldsymbol{A} \approx \boldsymbol{A}, \tag{1}$$

where $\boldsymbol{\Omega}$ is, again, a random sketch matrix. This approximation has received significant attention in the literature [8, 11, 12] and, like the randomized SVD, it enjoys strong theoretical guarantees. With the same number of matrix-vector products, the randomized Nyström approximation is typically significantly more accurate than the randomized SVD when the matrix has rapidly decaying singular values. Additionally, the Nyström method requires only a single pass over the matrix, compared to two passes for the randomized SVD, enabling all matrix-vector products to be performed in parallel.

Recently, Boullé and Townsend [4, 5] generalized the randomized SVD from matrices to Hilbert-Schmidt operators. Subsequent works [3, 6] employed this infinite-dimensional generalization of the randomized SVD to learn Green's functions associated with an elliptic or parabolic partial differential equations (PDE) from a few solutions of the PDE. This approach uses hierarchical low-rank techniques and exploits the fact that Green's functions are smooth away from the diagonal and therefore admit accurate off-diagonal low-rank approximations [1, 2]. Other applications, like Gaussian process regression and Support Vector Machines, involve integral operators that feature positive and *globally* smooth kernels. In turn, the operator is not only self-adjoint and positive but it also allows for directly applying low-rank approximation, without the need to resort to hierarchical techniques. Given existing results on matrices, it would be sensible to use an infinite-dimensional extension of the randomized Nyström approximation in such situations.

In this work, we present and analyze an infinite-dimensional extension of the randomized Nyström approximation for computing low-rank approximations to self-adjoint, positive, trace class operators. A significant advantage of the proposed framework is that once a low-rank approximation of the operator is computed, one can use this approximation to compute a low-rank approximation to *any* discretization of the operator.

References

[1] Mario Bebendorf. *Hierarchical matrices*. Springer, 2008.

[2] Mario Bebendorf and Wolfgang Hackbusch. Existence of $\mathcal{H}$-matrix approximants to the inverse FE-matrix of elliptic operators with L^∞-coefficients. *Numer. Math.*, 95:1–28, 2003.

[3] Nicolas Boullé, Seick Kim, Tianyi Shi, and Alex Townsend. Learning Green's functions associated with time-dependent partial differential equations. *J. Mach. Learn. Res.*, 23(1):9797–9830, 2022.

[4] Nicolas Boullé and Alex Townsend. A generalization of the randomized singular value decomposition. In *International Conference on Learning Representations*, 2022.

[5] Nicolas Boullé and Alex Townsend. Learning elliptic partial differential equations with randomized linear algebra. *Found. Comput. Math.*, 23(2):709–739, 2023.

[6] Nicolas Boullé, Diana Halikias, and Alex Townsend. Elliptic PDE learning is provably data-efficient. *Proc. Natl. Acad. Sci. U.S.A.*, 120(39):e2303904120, 2023.

[7] Jacob Gardner, Geoff Pleiss, Kilian Q Weinberger, David Bindel, and Andrew G Wilson. GPyTorch: Blackbox matrix-matrix Gaussian process inference with GPU acceleration. In *Advances in Neural Information Processing Systems*, volume 31, 2018.

[8] Alex Gittens and Michael W. Mahoney. Revisiting the Nyström method for improved large-scale machine learning. *J. Mach. Learn. Res.*, 17:Paper No. 117, 65, 2016.

[9] 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.

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

[11] David Persson and Daniel Kressner. Randomized low-rank approximation of monotone matrix functions. *SIAM J. Matrix Anal. Appl.*, 44(2):894–918, 2023.

[12] Joel A Tropp, Alp Yurtsever, Madeleine Udell, and Volkan Cevher. Fixed-rank approximation of a positive-semidefinite matrix from streaming data. In *Advances in Neural Information Processing Systems*, volume 30, 2017.

[13] Shusen Wang, Alex Gittens, and Michael W Mahoney. Scalable kernel K-means clustering with Nyström approximation: relative-error bounds. *J. Mach. Learn. Res.*, 20(1):431–479, 2019.

[14] Kai Zhang, Ivor W. Tsang, and James T. Kwok. Improved Nyström low-rank approximation and error analysis. In *Proc. International Conference on Machine Learning*, pages 1232–1239, 2008.