Randomly Pivoted Cholesky: Near-Optimal Positive Semidefinite Low-Rank Approximation from a Small Number of Entry Evaluations

Ethan N. Epperly, Yifan Chen, Joel A. Tropp, Robert J. Webber

Abstract

This talk describes randomly pivoted Cholesky (RPCholesky), a randomized algorithm for computing a low-rank approximation to a Hermitian postive semidefinite (psd) matrix. To compute a rank-k approximation to an $N \times N$ matrix, RPCholesky performs a k-step partial Cholesky decomposition with a pivot entry randomly chosen with probabilities proportional to diagonal entries of the current residual matrix (i.e., Schur complement). The algorithm requires $\mathcal{O}(k^2N)$ operations and reads only $(k+1)N$ entries of the input matrix.

The RPCholesky method has an interesting history. The existence of the method was briefly noted in a 2017 paper of Musco and Woodruff [9], and it is algebraically related to an earlier "randomly pivoted QR" algorithm of Desphande, Rademacher, Vempala, and Wang (2006, [3]). Our paper [2], originally released in 2022, reintroduced the algorithm, described its connection to Cholesky decomposition, evaluated the method numerically, and provided new theoretical results.

Surprisingly, this simple algorithm is guaranteed to produce a near-optimal low-rank approximation. The output of RPCholesky, and any other partial Cholesky decomposition, is low-rank approximation of the form

$$\widehat{\boldsymbol{A}} = \boldsymbol{A}(:, \mathsf{S})\boldsymbol{A}(\mathsf{S}, \mathsf{S})^{\dagger}\boldsymbol{A}(\mathsf{S}, :),$$

where S denotes the set of pivots selected by the algorithm and † denotes the Moore–Penrose pseudoinverse. This type of low-rank approximation is known as a *(column) Nyström approximation* and is used widely to accelerate kernel machine learning methods. It is known [7] that $k \geq r/\varepsilon$ columns S are needed to produce a Nyström approximation $\widehat{\boldsymbol{A}}$ within a $1+\varepsilon$ multiplicative factor of the best rank-r approximation $[\![\boldsymbol{A}]\!]_r$, i.e.,

$$\|\boldsymbol{A} - \widehat{\boldsymbol{A}}\|_* \leq (1+\varepsilon)\,\|\boldsymbol{A} - [\![\boldsymbol{A}]\!]_r\|_*\,.$$

Here, $\|\cdot\|_*$ denotes the trace norm. In [2], we showed that RPCholesky achieves the guarantee:

$$\mathbb{E}\left[\|\boldsymbol{A} - \widehat{\boldsymbol{A}}\|_*\right] \leq (1+\varepsilon)\,\|\boldsymbol{A} - [\![\boldsymbol{A}]\!]_r\|_* \quad \text{when } k \geq \frac{r}{\varepsilon} + r\log\left(\frac{1}{\varepsilon\eta}\right).$$

Here, $\widehat{\boldsymbol{A}}$ is the approximation produced by k steps of RPCholesky and $\eta = \|\boldsymbol{A} - [\![\boldsymbol{A}]\!]_r\|_* / \|\boldsymbol{A}\|_*$ denotes the relative error of the best rank-r approximation. In expectation, RPCholesky achieves the optimal scaling $k = r/\varepsilon$ up to an additive term that is logarithmic in the relative error η.

RPCholesky has proven effective at accelerating kernel machine learning methods. Given a data set $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N$, kernel methods perform machine learning tasks such as regression and clustering by manipulating a psd kernel matrix $\boldsymbol{A} = (\kappa(\boldsymbol{x}_i, \boldsymbol{x}_j))_{1\leq i,j\leq N}$, where κ is a given positive definite kernel function. When implemented directly, kernel methods require $\mathcal{O}(N^3)$ operations and $\mathcal{O}(N^2)$ storage. By replacing $\boldsymbol{A}$ with a low-rank approximation $\widehat{\boldsymbol{A}}$ (say, of rank $k = \mathcal{O}(1)$), the storage and runtime costs of these methods are reduced to $\mathcal{O}(N)$. This talk will present numerical experiments from [2], which show that an RPCholesky-accelerated clustering method can be 9× to 14× more accurate than accelerated clustering methods using other low-rank approximation techniques. Subsequent papers have applied RPCholesky to accelerate learning of committer functions in

biochemistry [1], as a preconditioner for conjugate gradient [4], for quadrature in reproducing kernel Hilbert spaces [5], and compression of data sets [8].

While the standard version of RPCholesky is already fast, it is slower than it could be because it processes the columns of the input matrix one-by-one. A blocked version of the method is faster, but can produce approximations of lower accuracy. This talk will conclude by discussing the recently introduced *accelerated RPCholesky method* [6], which simulates the performance of original RPCholesky using a combination of rejection sampling and block-wise computations. The accelerated RPCholesky method can be up to 40× faster than the original method while producing the same random output (in exact arithmetic).

References

[1] David Aristoff, Mats Johnson, Gideon Simpson, and Robert J. Webber. The fast committor machine: Interpretable prediction with kernels. *The Journal of Chemical Physics*, 161(8):084113, 2024.

[2] Yifan Chen, Ethan N. Epperly, Joel A. Tropp, and Robert J. Webber. Randomly pivoted Cholesky: Practical approximation of a kernel matrix with few entry evaluations. *Communications on Pure and Applied Mathematics, accepted*, 2024.

[3] Amit Deshpande, Luis Rademacher, Santosh Vempala, and Grant Wang. Matrix approximation and projective clustering via volume sampling. In *Proceedings of the 2006 Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1117–1126, 2006.

[4] Mateo Díaz, Ethan N. Epperly, Zachary Frangella, Joel A. Tropp, and Robert J. Webber. Robust, randomized preconditioning for kernel ridge regression. *arXiv preprint arXiv:2304.12465*, 2024.

[5] Ethan N. Epperly and Elvira Moreno. Kernel quadrature with randomly pivoted Cholesky. In *Advances in Neural Information Processing Systems 36*, 2023.

[6] Ethan N. Epperly, Joel A. Tropp, and Robert J. Webber. Embrace rejection: Kernel matrix approximation by accelerated randomly pivoted Cholesky. *arXiv preprint arXiv:2410.03969*, 2024.

[7] Venkatesan Guruswami and Ali Kemal Sinop. Optimal column-based low-rank matrix reconstruction. In *Proceedings of the 2012 Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1207–1214. 2012.

[8] Lingxiao Li, Raaz Dwivedi, and Lester Mackey. Debiased distribution compression. *arXiv preprint arXiv:2404.12290*, 2024.

[9] C. Musco and D. P. Woodruff. Sublinear Time Low-Rank Approximation of Positive Semidefinite Matrices. In *2017 IEEE Annual Symposium on Foundations of Computer Science*, pages 672–683, 2017.