

Least squares solvers based on randomized normal equations

Ilse C.F. Ipsen

Abstract

For the better part of my life I have taught that least squares problems are to be solved with a QR decomposition or SVD, cautioning that formation of the normal equations is to be avoided if possible.

Now I am re-thinking this advice, in light of developments underlying the Blendenpik least squares solver [1, 2], and our version of the randomized preconditioned Cholesky-QR algorithm [3].

Proposed Algorithm. Given a real $m \times n$ matrix \mathbf{A} with $\text{rank}(\mathbf{A}) = n$, we investigate the solution of the least squares problems $\min_{\mathbf{x}} \|\mathbf{Ax} - \mathbf{b}\|_2$ by solving the normal equation of a randomized preconditioned problem. In the spirit of the original Householder meetings, this is work in progress. Specifically, we right-precondition \mathbf{A} with a randomized preconditioner \mathbf{R}_s , to obtain $\mathbf{A}_1 \equiv \mathbf{AR}_s^{-1}$. Instead of taking the Blendenpik route and solving $\min_{\mathbf{y}} \|\mathbf{A}_1\mathbf{y} - \mathbf{b}\|_2$ via the iterative method LSQR, we solve the normal equations. That is, the Gram matrix $\mathbf{G} \equiv \mathbf{A}_1^T \mathbf{A}_1$ is formed explicitly, followed by solution of the normal equations $\mathbf{G}\mathbf{y} = \mathbf{A}_1^T \mathbf{b}$. This can be done with a Cholesky factorization of \mathbf{G} or of the bordered matrix $[\mathbf{A}_1 \ \mathbf{b}]$ [4, §2.2]. At last, one recovers the solution of the original problem via the triangular solve $\mathbf{R}_s \mathbf{x} = \mathbf{y}$.

To compute the randomized preconditioner \mathbf{R}_s , first improve the coherence of \mathbf{A} with a random orthogonal matrix \mathbf{FA} , where \mathbf{F} is the product of a fast transform (FFT, Walsh-Hadamard, DCT, Hartley) and a random diagonal matrix with independent Rademacher variables on the diagonal. Then sample c rows, uniformly and independently with replacement from \mathbf{FA} to obtain the sampled matrix $\mathbf{A}_s = \mathbf{SFA}$. At last compute the thin QR decomposition $\mathbf{A}_s = \mathbf{Q}_s \mathbf{R}_s$.

Advantages. Unlike Blendenpik [1] which solves a $m \times n$ least squares problem with an iterative method, we solve a small $n \times n$ problem with a direct method. Direct methods, and Cholesky decompositions in particular, tend to perform well on cache-based and parallel architectures, where data movement and synchronization can dominate arithmetic. This is in contrast to the normal equations like approach with iterative methods in [5, 6], which also requires an initial guess.

The simplicity of our approach, in contrast to the involved multi-stage [5, Algorithm 4], will lead to a rigorous and informative perturbation analysis for the accuracy of the computed solution. The potential backward stability issues due to the formation of the Gram matrix can be handled in the same way as for the randomized Cholesky-QR algorithm in [3].

The preconditioner \mathbf{R}_s needs to be applied only once and is applied explicitly, thereby improving the backward stability issues discussed in [7]. Even for matrices \mathbf{A} with worst case coherence and a condition number $\kappa(\mathbf{A}) \approx 10^{15}$, a sampling amount of $c = 3n$ suffices to produce preconditioned matrices \mathbf{A}_1 that are very well conditioned, with condition numbers $\kappa(\mathbf{A}_1) \approx 10$.

Preliminary numerical results suggest that for matrices with condition number $\kappa(\mathbf{A}) \leq 10^9$, the preconditioner \mathbf{R}_s can be computed faster, in single precision, without loss of accuracy in the preconditioned problem. We will show that solving the normal equations via a Cholesky decomposition represents an efficient least squares solver on NVIDIA RTX 2080 GPUs.

- [1] H. Avron, P. Maymounkov, and S. Toledo. Blendenpik: Supercharging Lapack’s Least-Squares Solver, *SIAM J. Sci. Comput.*, vol. 32, no. 3, pp 1217–1236 (2010)
- [2] I.C.F. Ipsen and T. Wentworth. The Effect of Coherence on Sampling from Matrices with Orthonormal Columns, and Preconditioned Least Squares Problems, *SIAM J. Matrix Anal. Appl.*, vol. 35, no. 4, pp 149–1520 (2014)
- [3] J.E. Garrison, and I.C.F. Ipsen. A Randomized preconditioned Cholesky-QR Algorithm, *SIAM J. Matrix Anal. Appl.*, under review (2024)
- [4] Å. Björck. *Numerical Methods for Least Squares Problems*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA (1996)
- [5] E.N. Epperly, M. Meier and Y. Nakatsukasa. Fast Randomized Least-Squares Solvers can be just as Accurate and Stable as Classical Direct Solvers, arXiv:2406.03468 (2024)
- [6] R. Xu and Y. Lu. Randomized Iterative Solver as Iterative Refinement: A Simple Fix Towards Backward Stability, arXiv:2410.11115 (2024)
- [7] M. Meier, Y. Nakatsukasa, A. Townsend, and M. Webb. Are sketch-and-precondition least squares solvers numerically stable? *SIAM J. Matrix Anal. Appl.*, vol. 45, no. 2, pp 905–929 (2024)