When does the randomized SVD actually compute an SVD?

Randomized subspace approximation beyond singular value gaps

Christopher Wang, Alex Townsend

Abstract

The randomized SVD (rSVD) is excellent at constructing a low-rank approximation to a matrix with rapidly decaying singular values, and its theoretical behavior as such was thoroughly explained in [6]. However, the singular values and singular subspaces of a good low-rank approximation may not accurately approximate the true singular values and singular subspaces, even with oversampling. The following example illustrates the problem.

Let A be a 1000×1000 matrix, where five of its eigenvalues are 1 and the remaining 995 eigenvalues are 0.05. We aim to estimate the 5-dimensional dominant eigenspace corresponding to the 1's using the rSVD with standard Gaussian test vectors and oversampling by 10. A simple numerical experiment returns the following results, over 1000 iterations of the rSVD:

- 1. Average relative low-rank error in the Frobenius norm $\frac{\|A-\tilde{A}\|_{\rm F}}{\|A-A_5\|_{\rm F}}$: 1.176969
- 2. Average relative low-rank error in the spectral norm $\frac{\|A-\tilde{A}\|_2}{\|A-A_5\|_2}$: 12.152447
- 3. Average maximum relative error for eigenvalues $\max_{j=1,...,5} \frac{|\sigma_j \tilde{\sigma}_j|}{|\sigma_j|}:$ 0.209691
- 4. Average principal angle error for dominant eigenspace $\|\Theta(\mathcal{U}_5, \mathcal{U}_5)\|_2$: 0.654832

Here, A is the rank-5 truncated approximant generated by the rSVD, A_5 is the best rank-5 approximant to A (that is, the diagonal matrix with five 1's in the top left corner), σ_j , $\tilde{\sigma}_j$ are the eigenvalues of A, \tilde{A} respectively, and \mathcal{U}_5 , \mathcal{U}_5 are the 5-dimensional dominant eigenspaces of A, \tilde{A} respectively. Observe that \tilde{A} is only a small factor away from being the optimal rank-5 approximant to A in the Frobenius norm, but it is much further from optimality in the spectral norm. Additionally, the rSVD has a hard time distinguishing between the eigenvalue 1 and the eigenvalue 0.1, and the rSVD's estimate for the dominant eigenspace is off by nearly 40 degrees, for the largest principal angle between the subspaces, on average. In short, the rSVD fails to be an SVD, even though it gives a good low-rank approximation in the Frobenius norm.

Therefore, we ask: for what matrices A does the rSVD actually compute an accurate SVD? We approach the question by considering how well the rSVD approximates the kth dominant singular subspaces (indeed, if the rSVD manages to accurately capture the dominant singular subspaces, then it must be accurate in the singular values as well [7]). This problem has received considerable attention over the past decade, especially through the lens of perturbation theory. The Davis– Kahan theorem, and its generalization by Wedin to nonsymmetric matrices, provides bounds, dependent on the size of the singular value gap between σ_k and σ_{k+1} , on the angular change of eigenvectors under a deterministic perturbation of a matrix. Building on these two early results, gap-dependent bounds on the accuracy of singular subspace approximations by projection-based methods, such as the rSVD, were derived for both deterministic and random settings by [5, 13, 14, 15]. Gap-independent bounds have also appeared in the works of [2, 9, 12]. While such bounds tell us when we expect to have a hard time determining the singular subspaces of a given matrix, they may fail to tell us whether a matrix is in fact conducive to singular subspace approximation. In fact, we observe in practice matrices with small or nonexistent singular value gaps whose singular subspaces are nevertheless well-approximated by the rSVD with high probability.

Our contribution is an exact, relatively straightforward formula for the cumulative distribution function of the largest principal angle between the true and the approximate dominant singular subspace, when using the rSVD with Gaussian test vectors. Our formula encapsulates the advantages of previous works in that (a) it is computable, interpretable, and a priori in the sense that it is space-agnostic, meaning it does not depend on prior knowledge of the singular subspaces; (b) it applies for any power of subspace iteration, with any amount of oversampling (including none at all); and (c) it can be used to derive existing bounds for the largest principal angle. Since our result is exact, it is certainly gap-independent. More importantly, it helps explain why a large singular value gap improves subspace estimates, but it also explains when and why the rSVD succeeds at singular subspace approximation even when the singular value gap is small. We show that the gap-dependent bounds of [15] assume the worst-case scenario given a gap, and we show that that worst-case scenario is when the singular values of A are $\sigma_1 = \cdots = \sigma_k > \sigma_{k+1} = \cdots = \sigma_n$ —the dominant singular values are as small as possible, while the tail is as large as possible.

To be precise, let $N \ge n$ and fix the target rank $k \ge 1$ and oversampling $p \ge 0$ such that $k + p < \frac{n}{2}$. Let $M \in \mathbb{R}^{N \times n}$ and let Σ_1, Σ_2 be $k \times k$ and $(n - k) \times (n - k)$ diagonal matrices of the dominant and tail singular values of M, respectively. If θ_1 denotes the largest principal angle between the kth dominant left singular subspace of M and the (k + p)-dimensional column space of $M\Omega$, where Ω is an $n \times (k + p)$ standard Gaussian matrix, then the cumulative distribution function of θ_1 is given, for $0 \le \theta \le \frac{\pi}{2}$, by

$$\mathbb{P}(\theta_1 < \theta) = \mathbb{E}\left[\det(S(\theta, \Sigma, Q, H_1, Q_1))^{\frac{n-k-p}{2}} {}_2F_1(\frac{-p+1}{2}, \frac{n-k-p}{2}; \frac{-p+1}{2}; I_k - S(\theta, \Sigma, Q, H_1, Q_1))\right]$$

where

1. $S(\theta, \Sigma, Q, H_1, Q_1)$ is the $k \times k$ matrix

$$\sin^{2}(\theta)\Sigma_{1}\left(\sin^{2}(\theta)\Sigma_{1}^{2}+\cos^{2}(\theta)QQ_{1}'(H_{1}'\Sigma_{2}^{-2}H_{1})^{-1}Q_{1}Q'\right)^{-1}\Sigma_{1}$$

- 2. Q, H_1 , and Q_1 are random matrices which are respectively the Q factors in the QR decomposition of a $k \times k$ standard Gaussian matrix, an $(n-k) \times (k+p)$ matrix whose columns are independently sampled from the multivariate Gaussian $\mathcal{N}(0, \Sigma_2^2)$, and a $(k+p) \times k$ matrix whose columns are independently sampled from $\mathcal{N}(0, H_1^\top \Sigma_2^{-2} H_1)$; and
- 3. $_2F_1(a, b; c; X)$ is the Gaussian hypergeometric function of matrix argument.

All of these quantities are computable given the singular values of M; the expectation can be computed by Monte–Carlo simulation, while the hypergeometric function can be evaluated extremely quickly via [8]. Numerical experiments demonstrate excellent agreement between our formula and empirical observations.

The primary tools used to derive our formula for the cumulative distribution function come from the statistical side of random matrix theory [4, 11]. Our result and proof generalizes those of [1, 3], the latter of which plays a major role in the theoretical guarantees of [6, 15]. We expect that our formula and techniques can be used to explain, in more detail, empirically observed phenomena in randomized methods for computing SVDs, including randomized Krylov iteration. We also expect that our formula can be used to analyze the possibility of using the rSVD as a test for low-rank structure, which has come up in [10, 16], or as a rank revealer. Finally, we hope to use our result to distinguish the classes of matrices for which randomized subspace methods succeed or fail at different tasks, including singular value and singular subspace approximation.

References

- ABSIL, P.-A., EDELMAN, A., AND KOEV, P. On the largest principal angle between random subspaces. *Linear Algebra its Appl.* 414 (2006), 288–294.
- [2] ALLEN-ZHU, Z., AND LI, Y. First efficient convergence for streaming k-PCA: A global, gapfree, and near-optimal rate. In *IEEE 58th Annual Symposium on Foundations of Computer Science* (2017), pp. 487–492.
- [3] CHEN, Z., AND DONGARRA, J. Condition numbers of Gaussian random matrices. SIAM J. Matrix Anal. Appl. 26, 3 (2005), 1389–1404.
- [4] CHIKUSE, Y. Statistics on Special Manifolds. Springer, 2003.
- [5] DONG, Y., MARTINSSON, P.-G., AND NAKATSUKASA, Y. Efficient bounds and estimates for canonical angles in randomized subspace approximations. SIAM J. Matrix Anal. Appl. 45, 4 (2024), 1978–2006.
- [6] HALKO, N., MARTINSSON, P.-G., AND TROPP, J. A. Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions. SIAM Rev. 53, 2 (2011), 217–288.
- [7] KNYAZEV, A. V. Sharp a priori error estimates of the Rayleigh–Ritz method without assumptions of fixed sign or compactness. *Math. Notes* 38 (1985), 998–1002.
- [8] KOEV, P., AND EDELMAN, A. The efficient evaluation of the hypergeometric function of a matrix argument. *Math. Comput.* 75, 254 (2006), 833–846.
- [9] MASSEY, P. Admissible subspaces and the subspace iteration method. *BIT Numer. Math.* 64, 1 (2024).
- [10] MEIER, M., AND NAKATSUKASA, Y. Fast randomized numerical rank estimation for numerically low-rank matrices. *Linear Algebra its Appl. 686* (2024), 1–32.
- [11] MUIRHEAD, R. J. Aspects of Multivariate Statistical Theory. John Wiley & Sons, 1982.
- [12] MUSCO, C., AND MUSCO, C. Randomized block Krylov methods for stronger and faster approximate singular value decomposition. In Advances in Neural Information Processing Systems (2015), vol. 28.
- [13] NAKATSUKASA, Y. Accuracy of singular vectors obtained by projection-based SVD methods. BIT Numer. Math. 57 (2017), 1137–1152.
- [14] NAKATSUKASA, Y. Sharp error bounds for Ritz vectors and approximate singular vectors. Math. Comput. 89, 324 (2020), 1843–1866.
- [15] SAIBABA, A. K. Randomized subspace iteration: Analysis of canonical angles and unitarily invariant norms. SIAM J. Matrix Anal. Appl. 40, 1 (2019), 23–48.
- [16] WANG, C., AND TOWNSEND, A. Operator learning for hyperbolic partial differential equations. Preprint arXiv:2312.17489 (2023).