Randomized SVD

Brady Metherall

25 November 2019

1 Introduction

singular value decomposition (SVD)

-svd good

-svd slow

-compression

-signposting

2 Truncated SVD

Often times the SVD sheds light on the underlying properties of its corresponding matrix. For example if several of the singular values are very small we may omit these to further compress the matrix. Such a process is called the truncated SVD where only the r largest singular values are kept and the rest are discarded along with the associated columns of U and V. This provides us with a rank-r approximation to the original matrix. In fact, truncated SVD yields the best rank-r approximation, this can be proved by the following two theorems.

Theorem 1. *If AB^T and B^TA are both symmetric matrices, then and only then can two orthogonal matrices U and V be found such that $\Sigma_A = U^TAV$, and $\Sigma_B = U^TBV$ are both diagonal matrices.*

Proof. The proof follows from the spectral theorem [1]. □

Theorem 2 (Eckart–Young Theorem)**.** *The best (in the Frobenius norm) rank-r approximation to a matrix, A, is obtained by the truncated SVD, A_r.*

Proof. The following proof has been adapted from [2]. The best approximation can be found by

$$M = \underset{X\in\bowtie_r}{\arg\min} \|A - X\|_F^2, \tag{1}$$

where $\bowtie_r$ is the set of all rank-r $m \times n$ matrices. The minimum error is then

$$\begin{aligned}\|A - M\|_F^2 &= \langle A, A\rangle - 2\langle A, M\rangle + \langle M, M\rangle, &(2)\\ &= \langle A, A\rangle - 2\big\langle A, U\Sigma_M V^T\big\rangle &(3)\\ &\qquad + \langle \Sigma_M, \Sigma_M\rangle.\end{aligned}$$

At the minimum, the change in $\|A - X\|_F^2$ is zero for some change in X. This change in X can be encapsulated as $U \mapsto sU$ where s is infinitesimal and antisymmetric to maintain orthogonality. Thus, at the minimum

$$0 = \langle A, sM\rangle = \big\langle AM^T, s\big\rangle. \tag{4}$$

Therefore, it is the case that AM^T is symmetric. By following a similar procedure, we find that M^TA must be symmetric as well.

By Theorem 1 A and M exhibit the same U and V in their SVDs. Now (2) can be simplified to

$$\begin{aligned}\|A - M\|_F^2 &= \|\Sigma_A - \Sigma_M\|_F^2, &(5)\\ &= \sum_{i=1}^{n} (\sigma_i(A) - \sigma_i(M))^2, &(6)\\ &= \sum_{i=r+1}^{n} \sigma_i(A)^2. &(7)\end{aligned}$$

This minimum is indeed achieved by the truncated SVD, since $\sigma_i(A_r) = \sigma_i(A)H(r - i)$. □

Additionally, the analysis in Section 3 will be dependent on:

Remark 1. *Theorem 2 was extended to all unitarily invariant norms in 1960 by Mirsky [3].*

Although, truncated SVD can provide a more compressed version of a matrix, it can be rather expensive to compute as the full SVD needs to be computed. To find a similar decomposition in faster time we turn our attention to the randomized SVD.

Algorithm 1 Randomized SVD.

1: Input: Matrix A of size $m \times n$, Int r, Int l.
2: Output: Matrix of size $m \times n$ of rank-r.
3: $\quad \Omega \leftarrow \mathcal{N}(0,1)^{n\times(r+l)}$
4: $\quad Q,\ \_ \leftarrow qr(A * \Omega)$
5: $\quad U,\ \Sigma,\ V \leftarrow svd(Q^T * A)$
6: **return** $(Q * U)[:, 1{:}r] * \Sigma[1{:}r, 1{:}r] * V[1{:}r, :]$

3 Randomized SVD

Randomized SVD attempts to obtain an approximation to the truncated SVD in a fraction of the time at the expensive of the accuracy. We shall investigate the expected value of the relative error for a variety of matrices. If the relative error is $\mathcal{O}(1)$ then the randomized SVD would be much more effective than truncated SVD because of the reduced time.

We find the randomized SVD using the method outlined in Algorithm 1. Effectively, our matrix, $A \in \mathbb{R}^{m\times n}$, is first multiplied by a random matrix of rank-$(r+l)$, where r is the rank we wish to approximate, and l is the buffer size. This extracts $r+l$ random directions of A, we then take the QR decomposition of this product to orthonormalize the projections. Finally, we compute the thin SVD on this much smaller matrix.

To have a better understanding of the expected speed up of randomized over truncated SVD we proceed by finding the computational complexity. The complexity of Algorithm 1 can be broken down as follows:

- Line 3 is $\mathcal{O}(n(r+l))$ for creating the random matrix Ω.
- Line 4 is $\mathcal{O}(mn(r+l))$ for the product $A\Omega$ and $\mathcal{O}\big(m(r+l)^2\big)$ for the QR decomposition[1].
- Line 6 is $\mathcal{O}\big(n(r+l)^2\big)$ for the SVD.

Therefore, the process is bottlenecked by the QR decomposition, and will scale as $\mathcal{O}(mnr)$, as opposed to $\mathcal{O}\big(mn^2\big)$ as in truncated SVD which first calculates the full SVD. Thus, we would expect randomized SVD to be $\mathcal{O}(n/r)$ faster, which for a large, redundant matrix can be quite large.

We implement Algorithm 1 in Julia[2], and compare the relative error in the spectral, Frobenius, and trace norms for matrices of full rank, rank-r, algebraically decaying singular values, and geometrically decaying singular values[3]. The results of the computations can be seen in Figure 1.

We can see from Figure 1a that the relative error monotonically increases for all three norms—however, once $r+l$ approaches n the relative error quickly approaches one as the approximation becomes exact. Furthermore, even for a somewhat sizeable matrix the error is indeed $\mathcal{O}(1)$, and the standard deviation quite small. We find very different behaviour for rank-r matrices (Figure 1b). In this case we are approximating a rank-r matrix with a rank-r matrix, and so we expect to—mathematically—have a zero error. Of course computationally we are unlikely to obtain zero, and instead we find errors of $\mathcal{O}\big(10^{-11}\big)$. Randomized SVD performs considerably worse here (relatively) than with a full rank matrix, but, we are still able to approximate the matrix almost exactly.

-accuracy (good for full rank and algebraic decay. unstable for rank r since equal, unstable for geo since decays very quick)

-bounded above by an order of magnitude, pretty good much quicker

4 Conclusion

References

[1] S. Axler, *Linear Algebra Done Right*. Undergraduate Texts in Mathematics, Springer, 2 ed., 1997.

[2] C. Eckart and G. Young, "The Approximation of One Matrix by Another of Lower Rank," *Psychometrika*, vol. 1, pp. 211–218, Sept 1936.

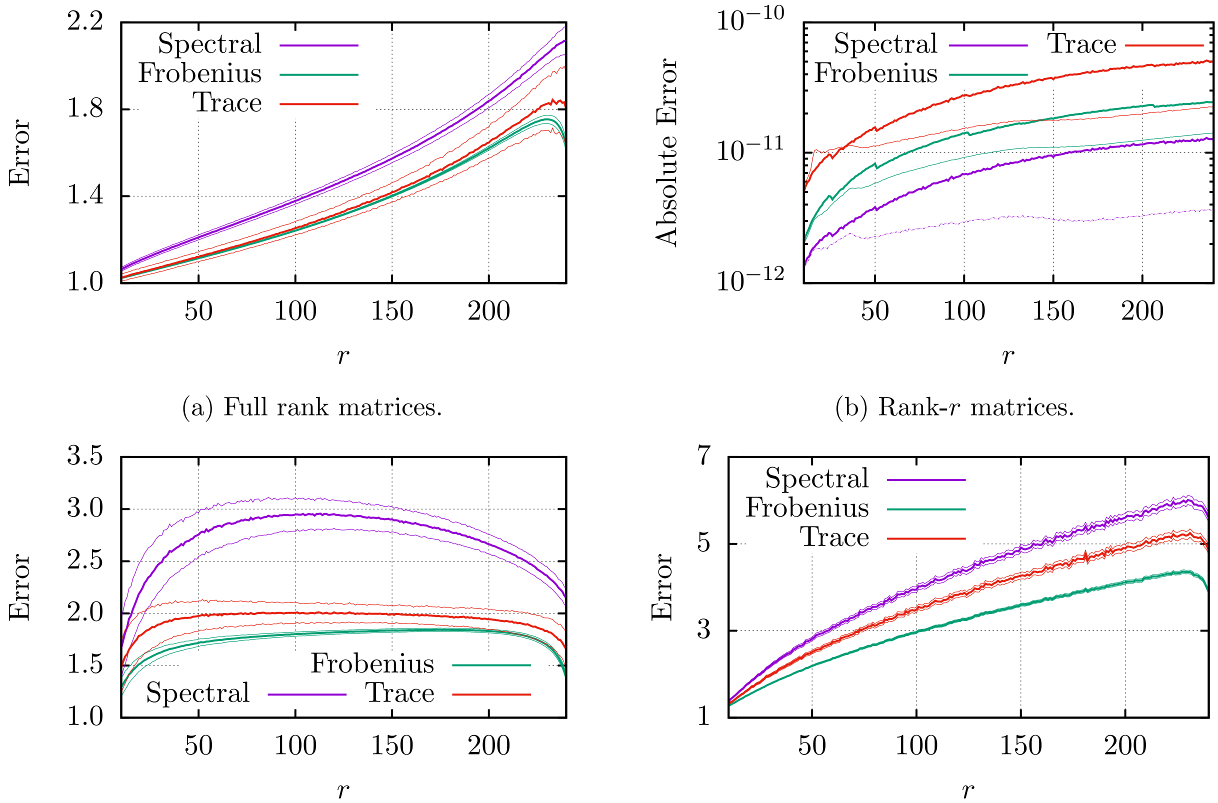
[3] L. Mirsky, "Symmetric Gauge Functions and Unitarily Invariant Norms," *The Quarterly Journal of Mathematics*, vol. 11, pp. 50–59, Jan 1960.

[4] J. Bezanson, A. Edelman, S. Karpinski, and V. B. Shah, "Julia Linear Algebra Documentation." Online, Retrieved Nov 2019. url: `docs.julialang.org/en/v1/stdlib/LinearAlgebra/`.

[1] As we shall see, Algorithm 1 is implemented in Julia which calls the C Householder QR decomposition [4].

[2] The code can be found at `github.com/bmetherall/Oxford/blob/master/Courses/Computational_Techniques/RandSVD.jl`.

[3] In the case of rank-r matrices we compare the absolute error instead since the approximation should be exact.



(a) Full rank matrices.

(b) Rank-r matrices.

(c) Algebraically decaying singular values ($\sigma_i = 10 \times i^{-1.5}$).

(d) Geometrically decaying singular values ($\sigma_i = 10 \times 0.9^{1-i}$).

Figure 1: Relative error of four types of matrix for the spectral, Frobenius, and trace norms. The thin lines represent the standard deviation in Figures 1a and 1c, the standard error in Figure 1d since the singular values decay quickly, and the error of the truncated SVD in Figure 1b since the division becomes numerically unstable. The parameters of the computations were $m = 500$, $n = 250$, $l = 5$, and $N = 10^3$.