

An Accelerated Accuracy-enhanced Randomized Singular Value Decomposition for Factorizing Matrices with Low-rank Structure

Joseph Roger Arhin^{1,*}, Francis Sam¹, Kenneth Coker², Toufic Seini¹

¹School of Mathematical Sciences, University of Electronic Science and Technology of China, Chengdu, China

²School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu, China

Email address:

mailroger2day@gmail.com (J. R. Arhin), fransamang@gmail.com (F. Sam), kennethcoker25@gmail.com (K. Coker),

seinitouficgh@gmail.com (T. Seini)

*Corresponding author

To cite this article:

Joseph Roger Arhin, Francis Sam, Kenneth Coker, Toufic Seini. An Accelerated Accuracy-enhanced Randomized Singular Value Decomposition for Factorizing Matrices with Low-rank Structure. *Pure and Applied Mathematics Journal*. Vol. 9, No. 4, 2020, pp. 64-73. doi: 10.11648/j.pamj.20200904.11

Received: June 24, 2020; **Accepted:** July 13, 2020; **Published:** July 28, 2020

Abstract: Big data has in recent years gained ground in many scientific and engineering problems. It seems to some extent prohibitive for traditional matrix decomposition methods (i.e. QR, SVD, EVD, etc.) to handle such large-scale problems involving data matrix. Many researchers have developed several algorithms to decompose such big data matrices. An accuracy-enhanced randomized singular value decomposition method (referred to as AE-RSVD) with orthonormalization recently becomes the state-of-the-art to factorize large data matrices with satisfactory speed and accuracy. In our paper, low-rank matrix approximations based on randomization are studied, with emphasis on accelerating the computational efficiency on large data matrices. By this, we accelerate the AE-RSVD with modified normalized power iteration to result in an accelerated version. The accelerated version is grounded on a two-stage scheme. The first stage seeks to find the range of a sketch matrix which involves a Gaussian random matrix. A low-dimensional space is then created from the high-dimensional data matrix via power iteration. Numerical experiments on matrices of different sizes demonstrate that our accelerated variant achieves speedups while attaining the same reconstruction error as the AE-RSVD with orthonormalization. And with data from Google art project, we have made known the computational speed-up of the accelerated variant over the AE-RSVD algorithm for decomposing large data matrices with low-rank form.

Keywords: Low-rank, Singular Value Decomposition (SVD), Accelerated AE-RSVD, Orthonormalization, Power Iteration

1. Introduction

Matrices in many applications appear with a low-rank structure and matrix factorization approaches are often employed to develop compacted and informative observations to make data calculation and interpretation easier. A long-established method is the almighty singular value decomposition (SVD) which discovers the finest low-rank approximation of a data matrix. Various SVD algorithms in scientific computing depend on low-rank approximation via matrix decomposition of the form

$$C_{m \times n} \approx E_{m \times k} \times F_{k \times n}, \quad (1)$$

where $k \ll \min\{m, n\}$. This illustration permits data analysts to analyze or work with the matrix D using the factor matrices E and F as an alternative of the full matrix, which is more efficient both computationally and memory usage. Besides, these smaller factor matrices can offer definite structures to achieve the desired result by analyzing a data matrix [1]. SVD provides a numerically robust matrix factorization which can facilitate to calculate the pseudo-inverses of general matrices [2], get low-rank approximations [3] and evaluate the least-squares and minimum-norm outcomes for a linear model [4]. Besides, the SVD is the algorithm behind several machine learning concepts, including sparse coding, matrix completion, dictionary learning, PCA and robust PCA. For a

comprehensive technical outline of SVD, one can refer to the work of Trefethen *et al.* [5].

1.1. Some Applications of Low-rank Approximation

Many applications in scientific computing encompass matrices with low-rank structure. Singular value decomposition is employed in the following applications:

A basic technique in statistics is to calculate directions of highest variance in vector-valued data by performing PCA on an input matrix [6].

Singular value decomposition also plays an essential part in tensor decomposition [7].

Another on the list of applications includes image compression [8].

1.2. Early Works of Randomized Algorithm

Recent algorithms for performing singular value decomposition are grounded on Krylov approaches, like Lanczos algorithm [9]. These approaches are precise and are particularly prevailing for approximating structured, and sparse matrices.

In the last 20 years, random matrix factorization algorithms have become more common to measure low-rank matrix approximation. Introducing the “Monte Carlo” SVD, a strenuous approach was used in the paper of Frieze, Kannan and Vempala to competently measure the low-rank SVD based on a uniform column sample row [10]. Several works used a random projection based on a robust method [11, 12]. Specifically, the features of random vectors are dominated so that the range of a matrix can be captured effectively. Woolfe *et al.* further improved computational performance by using the characteristics of very structured matrices, that allow matrix multiplication speed [13]. The significant effort by Halko *et al.* united and extended previous work on randomized SVD and presented advanced prototype algorithms to calculate low-rank singular value decomposition which is closer to the optimum value [3]. The algorithmic work of Halko *et al.* is summarized in algorithm 1 below:

Algorithm 1: A Prototype Randomized SVD

Inputs: An $m \times n$ matrix A , a target rank- k , an oversampling parameter p say ($p=5,10$)

Outputs: Matrices U , Σ , V in an approximate rank- $(k+p)$ SVD of A (i.e. U and V are orthonormal and Σ is diagonal)

Stage A:

- (1) $\Omega = \text{randn}(n, k+p)$;
- (2) $Y = A\Omega$;
- (3) $[Q, \sim] = \text{qr}(Y, 0)$;

Stage B:

- (4) $B = Q^* A$;
- (5) $[\tilde{U}, \Sigma, \tilde{V}] = \text{svd}(B, 'econ')$;
- (6) $U = Q\tilde{U}$;

To improve on accuracy, Martinsson introduced the “Accuracy-enhanced randomized SVD with

orthonormalization (AE-RSVD)” [14] which is displayed in algorithm 2.

Algorithm 2: AE-RSVD (with orthonormalization)

Inputs: An $m \times n$ matrix A , a rank parameter $k \leq \min\{m, n\}$, p is an oversampling parameter ($p=5,10$) and q is an index ($q=1,2,\dots$).

Outputs: Two orthonormal matrices U and V and a diagonal matrix Σ in an approximate rank- $(k+p)$ SVD of the input matrix A .

- (1) $\Omega = \text{randn}(n, k+p)$;
- (2) $Q = \text{orth}(A\Omega)$;
- (3) for $j = 1 : q$ do
- (4) $W = \text{orth}(A^* Q)$;
- (5) $Q = \text{orth}(AW)$;
- (6) end for
- (7) $B = Q^* A$;
- (8) $[\tilde{U}, \Sigma, \tilde{V}] = \text{svd}(B, 'econ')$;
- (9) $U = Q\tilde{U}$;

1.3. Main Contribution

Inspired by the work of Martinsson [14], we found it attractive to improve upon the speed of the AE-RSVD (with orthonormalization). This manuscript focuses on a conceptually workable algorithm for accelerating the computational performance of the numerical low-rank approximation of singular value decomposition (SVD) via random projection. The algorithms presented in this manuscript are obtained by a significant disparity on the normalized power iteration scheme [15] and are easy to implement. We present the so-called “accelerated AE-RSVD with modified normalized power iteration” algorithm which is an extension of Martinsson's work. Section 4 of this work shows several numerical experiments on both ‘fat’ and ‘thin’ matrices and comparing the performance with the AE-RSVD (with orthonormalization) [14]. With the accelerated version of the AE-RSVD, efficiency is improved by the use of LU factorization (with partial pivoting) in step 2 of the modified normalized power iteration instead of QR decomposition to re-normalize after each step of the first q applications of Y , the sketch matrix. The precision is also observed because the reconstruction errors are not significantly different.

1.4. Notation

Throughout this manuscript, all vectors are measured in \mathbb{R}^n using their ℓ_2 norm (Euclidean norm), $\|u\| = \sqrt{\sum_{j=1}^n |u(j)|^2}$.

Matrices are denoted with capital letters such as A, Q . For simplicity, we prefer to work with only real matrices in this manuscript e.g., $A \in \mathbb{R}^{m \times n}$, but most of the proofs and reasoning works as well for $A \in \mathbb{C}^{m \times n}$. We measure matrices by the Frobenius and the spectral norms defined by

$$\|A\|_F = \left(\sum_{i,j=1}^{m,n} |A(i,j)|^2 \right)^{\frac{1}{2}} \text{ and } \|A\| = \max_{\|x\|=1} \|Ax\|, \quad (2)$$

respectively [14].

1.5. Orthonormalization

Given an $m \times \ell$ matrix B , with $m \geq \ell$, we introduce a function

$$Q = \text{orth}(B),$$

to signify orthonormalization of the columns of B . In other words, Q will be an $m \times \ell$ orthonormal matrix whose columns make a basis for the range of B [14]. In practice, the keyword 'qr' means to use a householder transformation to perform a QR factorization; e.g., in MATLAB, we would write $[Q, \sim] = \text{qr}(B, 0)$.

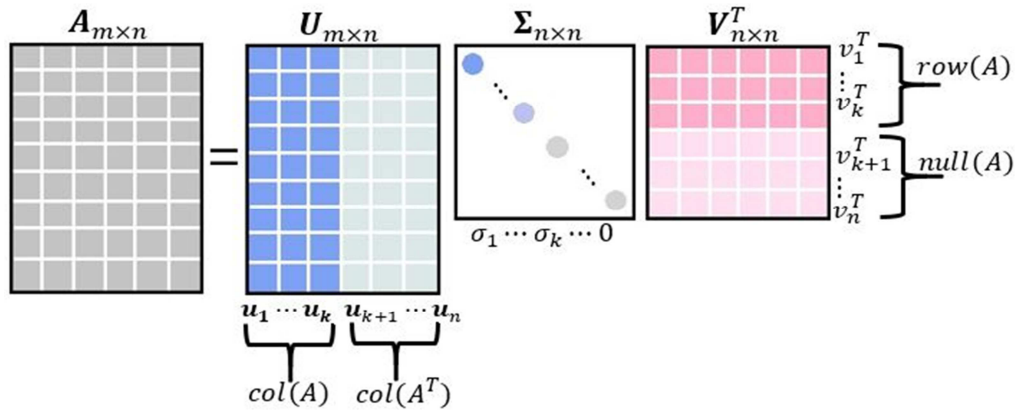


Figure 1. An illustration of the 'thin' SVD for a rank- k with $m \geq n$ [16].

2.2. Geometrical Interpretation of SVD

Let \mathbb{S}^{n-1} be the unit sphere in \mathbb{R}^n :

$$\mathbb{S}^{n-1} = \{x \in \mathbb{R}^n : \|x\|_2 = 1\} \quad (4)$$

Let $A \cdot \mathbb{S}^{n-1}$ be the image of \mathbb{S}^{n-1} under A :

$$A \cdot \mathbb{S}^{n-1} = \{Ax : x \in \mathbb{R}^n \text{ and } \|x\|_2 = 1\} \quad (5)$$

Then $A \cdot \mathbb{S}^{n-1}$ is an ellipsoid centered at the origin of \mathbb{R}^n , with principal axes $\sigma_i u_i$.

Proof:

We know that $\|x\|_2 = 1$ and replacing x with $A^{-1}Ax$, we define $y \square Ax$, then $\|x\|_2 = 1$ becomes

$$1 = \|A^{-1}y\|_2^2 = \|V\Sigma^{-1}U^T y\|_2^2, \quad (6)$$

$$\text{Let } w = U^T \text{ and } y = UV^T w$$

2. Preliminaries in Linear Algebra

2.1. Conceptual Overview of Singular Value Decomposition

Every real matrix of arbitrary rectangular dimensions can be factored into the form

$$A \approx U \Sigma V^T, \quad (3)$$

where U and V , both have orthonormal columns, and whose columns hold the left and right singular vectors, respectively. The matrix Σ is diagonal and contains singular values which match the singular vectors. By convention, we denote the i^{th} singular value as $\sigma_i(A)$, and arrange them and their corresponding singular vectors into U and V such that $\Sigma = \text{diag}[\sigma_1(A), \sigma_2(A), \dots]$ and $\sigma_i(A) \geq \sigma_{i+1}(A) \forall i$.

$$1 = \|\Sigma^{-1} w\|_2^2 = \frac{w_1^2}{\sigma_1^2} + \frac{w_2^2}{\sigma_2^2} + \dots + \frac{w_n^2}{\sigma_n^2} \quad (7)$$

We know that $A = U \Sigma V^T$. So,

$$AV = U \Sigma = [u_1, \dots, u_n] \begin{pmatrix} a_1 & & \\ & \ddots & \\ & & a_n \end{pmatrix} \quad (8)$$

$$Av_j = \sigma_j u_j. \quad (9)$$

Geometrically, U and V can be assumed as rotation matrices $\in SO(m)$ and $SO(n)$, and thus we can view the matrix-vector multiplication of A acting on vector x as:

Takes the vector $x \in \mathbb{R}^n$ and rotates it by V^T .

Scale each dimension of the result independently via multiplication with Σ . When $m < n$, Σ also discards the last $m-n$ dimensions to map to $x \in \mathbb{R}^m$; when $m > n$, it maps it to $x \in \mathbb{R}^n$ by zero-padding the new $m-n$ dimensions.

Rotates a second time by U .

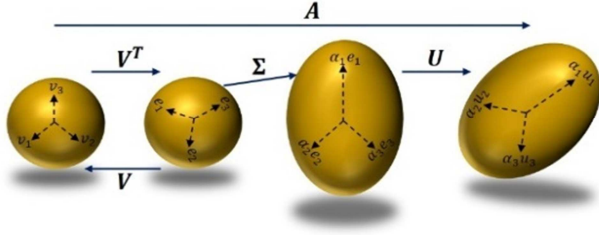


Figure 2. A graphical illustration of the SVD of $A \in \mathbb{R}^n$.

3. Randomized Algorithm Enhanced

The basis matrix Q usually fails to deliver a good estimation for the range of the input matrix. The reason being that, most real-world data matrices do not have an exact rank- k , but instead have a range of singular values that gradually decay [14]. Efficiency can be significantly improved with the concept of over-sampling [3] and power repetition as suggested in Martinsson's paper [14].

3.1. Oversampling

Majority of data matrices do not have an exact rank, which implies that the singular values $\{\sigma_j\}_{j=k+1}^n$ of the unique input matrix A are all non-zero. The sketch matrix Y , therefore, does not exactly span the range of the input matrix. This problem can be fixed via oversampling by setting $\ell = k + p$ random projections to form the sketch, in place of only k . In this discussion, p refers to the number of extra projections and a small number, say $\{p=5,10\}$ is normally enough to achieve a strong basis similar to the best possible one [3].

3.2. The Idea of Power Iteration

The idea of power sample iterations is another technique for enhancing the value of the basis Q . The method outlined in algorithm 3 works well for matrices with some kind of decay in singular values but they can generate a weak basis for a flat singular spectrum or a very big input matrix. The main clue about power iteration is for enhancing the exactness of random algorithms, which were originally proposed in the manuscript of Rokhlin *et al.* [15]. To review, presumably, we have calculated a rank- k approach to an $m \times n$ matrix A with singular values $\{\sigma_j\}_{j=1}^{\min(m,n)}$. The concept demonstrates that the spectral norm error is restricted to a factor that measures $(\sum_{j>k} \sigma_j^2)^{\frac{1}{2}}$ [3]. Such quantity can be much larger when the singular values decay slowly than the theoretically minimal estimated error (i.e. σ_{k+1}) [14].

3.3. The Modified Normalized Power Iteration Algorithm

Martinsson's algorithm tends to be accurate but time-wasting to compute the approximated rank- k SVD. We therefore, found it necessary to improve upon speed and also maintain the accuracy of the RSVD via the modified

normalized power iteration. We therefore, introduce the modified normalized power iteration algorithm.

Algorithm 3: Modified Normalized Power Iteration

Inputs: An input matrix $A \in \mathbb{R}^{m \times n}$ and the sketch matrix $Y \in \mathbb{R}^{m \times \ell}$.

Outputs: An orthonormal matrix $Y \in \mathbb{R}^{m \times \ell}$

- (1) for $j = 1 : q$ do
- (2) $[L, \sim] = lu(Y)$; $O(qm\ell^2)$
- (3) $[Q, \sim] = qr(A^T L)$; $O(qmn\ell + qm\ell^2)$
- (4) $Y = orth(AQ)$; $O(qmn\ell + qm\ell^2)$
- (5) end for

3.4. The Accelerated AE-RSVD Algorithm

The prototype randomized scheme [3] gives accurate results for matrices which rapidly decay with respect to its singular value but tends to produce sub-optimal results when they do not [14]. We therefore, introduce the accelerated AE-RSVD with modified normalized power iteration algorithm. We give a quick overview of the accelerated variant algorithm. As usual, the two-stage scheme was used:

Stage A-The range finder: For us to obtain a low-rank SVD approximation of $A \in \mathbb{R}^{m \times n}$ with target rank- k , where $k \leq \min\{m, n\}$ and $m \geq n$, we require a random test matrix $\Omega \in \mathbb{R}^{n \times \ell}$, where $\ell = k + p$. By this, p denotes a parameter for oversampling, and we will see that a small amount of oversampling is often sufficient in practice.

The definition of a random projection is widely used to sample the range of the data matrix A to construct such an orthogonal projection scheme efficiently. Random projections are orthodox data and generated from the standard normal distribution, for example, by initially drawing a set of k random vectors $\{a_i\}_{i=1}^k$. Random vectors remain linearly independent with high probability; the theory of probability ensures this. The mapping of A to low-dimensional space then calculates a number of random $\{y_i\}_{i=1}^k$ projections

$$y_i = Aa_i \text{ for } i = 1, 2, \dots, k. \quad (10)$$

Equation (10) can be efficiently executed in parallel. Therefore, let us generate the random test matrix $\Omega \in \mathbb{R}^{n \times \ell}$, which is indeed generated from the standard normal distribution, whose columns are the vectors $\{a_i\}_{i=1}^k$. The sketch matrix $Y \in \mathbb{R}^{m \times \ell}$, is then acquired by the random test matrix after the input matrix has been multiplied

$$Y := A\Omega. \quad (11)$$

Once Y is found, it only remains to use the modified normalized power iteration in algorithm 3 to renormalize and orthonormalize the sketch (sample) matrix to update the sketch matrix such that $Y \in \mathbb{R}^{m \times \ell}$.

Stage B-Computation of smaller matrix: After

orthonormalizing the columns of Y , we construct an orthonormal basis $Q \in \mathbb{R}^{m \times \ell}$ to span the columns of $Y \in \mathbb{R}^{m \times \ell}$. We now aim to find a smaller matrix $B \in \mathbb{R}^{\ell \times n}$. This means we are therefore constructing a low-dimensional space from a high-dimensional data matrix such that

$$B := Q^T A. \quad (12)$$

We now compute the SVD of the smaller matrix B to obtain two orthonormal matrices $\bar{U} \in \mathbb{R}^{\ell \times \ell}$, $V \in \mathbb{R}^{n \times n}$ and a diagonal matrix $\Sigma \in \mathbb{R}^{\ell \times n}$ as shown

$$[\bar{U}, \Sigma, V] = \text{svd}(B). \quad (13)$$

The left singular vectors are then updated by multiplying the orthonormal basis Q and the truncated matrix \bar{U}

$$U = Q\bar{U}. \quad (14)$$

Algorithm 4: Accelerated AE-RSVD with Modified Normalized Power Iteration

Inputs: An $m \times n$ matrix A , a rank parameter $k \leq \min\{m, n\}$, p is an oversampling parameter ($p=5, 10$) and q is an index ($q=1, 2, \dots$).

Outputs: Two orthonormal matrices U and V and a diagonal matrix Σ in an approximate rank- $(k+p)$ SVD of the input

Table 1. Computational Complexities of the Two Algorithms.

Scheme	Computational Complexities
AE-RSVD (with orthonormalization)	$C_{mm}(2+2q)mn\ell + C_{qr}(1+2q)m\ell^2$
Accelerated AE-RSVD with modified normalized power iteration	$C_{mm}[(2+2q)mn\ell + m\ell^2] + C_{qr}(qm+qn+m)\ell^2 + C_{lu}qm\ell^2$

3.6. Theoretical Error Bound of the Accelerated AE-RSVD with Modified Normalized Power Iteration

To state the theorem of the error bound of algorithm 4, we make reference to two most important propositions about random measurement matrix [17-19].

Proposition 1: Let G be a Gaussian matrix. Then

$$\left(\mathbb{E} \left[\|SGT\|_F^2 \right] \right)^{\frac{1}{2}} \leq \|S\|_F \|T\|_F \quad (15)$$

and

$$\mathbb{E}[\|SGT\|] \leq \mathbb{E}[\|S\|_F \|T\|_F + \|S\|_F \|T\|] \quad (16)$$

Proposition 2: Let G be a Gaussian matrix of size $k \times k+p$, $p \geq 2$. Then

$$\left\{ \left(\mathbb{E} \left[\|SGT\|_F^2 \right] \right)^{\frac{1}{2}} \leq \sqrt{\frac{k}{p-1}} \text{ and } \left[\|G^{\dagger}\| \right] \leq \frac{e\sqrt{k+p}}{p} \right\} \quad (17)$$

Theorem 1: Let $A \in \mathbb{R}^{m \times n}$ with singular values

matrix A .

- (1) $\ell = k + p$;
- (2) $\Omega = \text{randn}(n, \ell)$;
- (3) $Y = A\Omega$; $O(mn\ell)$
- (4) **Power iteration.** alg. 3; $O(2qm\ell^2 + 2qmn\ell + qn\ell^2)$
- (5) $[Q, \sim] = \text{qr}(Y, 0)$; $O(m\ell^2)$
- (6) $B = Q^T A$; $O(mn\ell)$
- (7) $[\bar{U}, \Sigma, V] = \text{svd}(B, 'econ')$;
- (8) $U = Q\bar{U}$;

Efficiency is enhanced by the use of LU factorization (with partial pivoting) in step 2 of algorithm 3 instead of QR decomposition to re-normalize after each step of the first q applications of Y , the sketch matrix. This normalization ensures speedup and maintains accuracy as compared to the AE-RSVD (with orthonormalization) as proposed by Martinsson [14]. This is summarized in algorithm 3.

3.5. Analysis of Computational Complexity

We analyze the significance of each step to contextualize the computational complexity of both the accelerated AE-RSVD with modified normalized power iteration and the AE-RSVD (with orthonormalization) in floating-point operation. The computational complexities are summarized in Table 1. To sum it up, both algorithms are dominated by the complexity $O(mn\ell)$.

$\{\sigma_j\}_{j=1}^{\min(m,n)}$, $k \leq \min\{m, n\}$ is the target rank, let q denote a small positive integer (for the power iteration) and we set $p \geq 2$ to be an oversampling parameter. Letting $\ell = k + p$, draw a random Gaussian test matrix Ω of size $n \times \ell$, set $Y = (AA^T)^q A\Omega$. Setting $Q \in \mathbb{R}^{m \times \ell}$, the orthonormal matrix as a result of orthonormalizing the columns of Y . Then

$$\mathbb{E} \left[\|A - QQ^T A\| \right] \leq \left[\sqrt{\frac{k}{p-1}} \sigma_{k+1}^{2q+1} + \frac{e\sqrt{\ell}}{p} \left(\sum_{j=k+1}^{\min\{m,n\}} \sigma_j^{2(2q+1)} \right)^{\frac{1}{2}} \right]^{\frac{1}{2q+1}}. \quad (18)$$

Here, the operator \mathbb{E} refers to the expectation in relation to the Gaussian measurement matrix Ω , and e is known to be Euler's number. Furthermore, p the over-sampling parameter is presumed to be larger than or equal to 2. From the error bound in (18), it follows that the approximation error can be controlled both by the over-sampling parameter, p and the power iteration scheme parameter, q . The power iteration scheme parameter q speeds up the decay rate in the singular values of the sampled matrix by keeping the same eigenvectors [14].

Proof:

We proceed with the proof. Partition the SVD of A as follows:

$$A = U \begin{pmatrix} \Sigma_1 & \Sigma_2 \\ \Sigma_1^T & \Sigma_2^T \end{pmatrix} \begin{pmatrix} V_1^T \\ V_2^T \end{pmatrix} \begin{matrix} k \\ n-k \end{matrix}. \quad (19)$$

$V_1^T \in k \times n$ and $\Omega \in n \times (k+p)$, let $\Omega_1 = V_1^T \Omega$. Therefore, $\Omega_1 \in k \times (k+p)$, $V_2^T \in (n-k) \times n$, let $\Omega_2 = V_2^T \Omega$. This implies that $\Omega_2 \in (n-k) \times (k+p)$

$$\begin{bmatrix} \Omega_1 \\ \Omega_2 \end{bmatrix} = \begin{bmatrix} V_1^T \Omega \\ V_2^T \Omega \end{bmatrix}. \quad (20)$$

From theorem (2.9.1) of [3], assuming Ω_1 is not singular, it holds that

$$\|A - QQ^T A\| \leq \|\Sigma_2\|^2 + \|\Sigma_2 \Omega_2 \Omega_1^\dagger\|^2. \quad (21)$$

Here, $\|\cdot\|$ represents either the ℓ_2 -operator norm, or the Frobenius norm. Because Ω is obtained from a Gaussian distribution, the matrices Ω_1 and Ω_2 also exhibits the features of a Gaussian distribution. As a consequence, the matrices U and V are not involved in the analysis and we can simply assume that A is diagonal, $A = \text{diag}(\sigma_1, \sigma_2, \dots)$.

If $p = 0$, then $\|\Omega_1^\dagger\|$ is typically large, and very unstable [3]. First observe that the expectation of the third term of (21) and making use of proposition 1,

$$\|\Sigma_2 \Omega_2 \Omega_1^\dagger\| \leq \left[\|\Sigma_2\| \|\Omega_1^\dagger\|_F + \|\Sigma_2\|_F \|\Omega_1^\dagger\| \right] \quad (22)$$

$$\leq \{\text{Hölder}\} \leq \|\Sigma_2\| \left(\|\Omega_1^\dagger\|_F \right)^{\frac{1}{2}} + \|\Sigma_2\|_F \|\Omega_1^\dagger\|. \quad (23)$$

Proposition 2 now provides bounds for $\mathbb{E} \|\Omega_1^\dagger\|_F^2$ and $\mathbb{E} \|\Omega_1^\dagger\|$ so we obtain

$$\mathbb{E} \left[\|\Sigma_2 \Omega_2 \Omega_1^\dagger\| \right] \leq \sqrt{\frac{k}{p-1}} \|\Sigma_2\| + \frac{e\sqrt{k+p}}{p} \|\Sigma_2\|_F \quad (24)$$

$$= \left[\sqrt{\frac{k}{p-1}} \sigma_{k+1}^{2q+1} + \frac{e\sqrt{\ell}}{p} \left(\sum_{j=k+1}^{\min\{m,n\}} \sigma_j^{2(2q+1)} \right)^{\frac{1}{2}} \right]^{\frac{1}{2q+1}}. \quad (25)$$

4. Numerical Experiments

In this section, we include numerical instances to demonstrate the computational efficiency of the accelerated AE-RSVD with modified normalized power iteration scheme and to compare its efficiency with the AE-RSVD with orthonormalization found in the manuscript of Martinsson [14]. Both algorithms are executed in MATLAB R2018a software. All computations were done on a machine with the following specifications: A desktop computer with a 4-core (8 logical processors) Intel Core i7-7700 CPU (3.60GHz), and 32GB DDR4 RAM. The relative reconstruction error is calculated using the Frobenius norm as

$$\frac{\|M - M_k\|_F}{\|M\|_F}, \quad (26)$$

where k is the target rank and M_k is the approximated matrix. The spectral norm was also used.

4.1. Comparison of Execution Speed

We compare the run time and the reconstruction errors of the accelerated variant to that of the AE-RSVD with orthonormalization. A synthetic ‘thin’ matrix whose singular values do not decay easily of size 10000×8500 was used to test the efficiency of the two algorithms. We controlled the efficiency of the algorithms with varying q . A target rank $\ell = 3000$ was used. The results are shown in Table 2. Figure 3 also shows the rate of decay of the singular values of both the 10000×8500 and the 2048×4096 synthetic matrices.

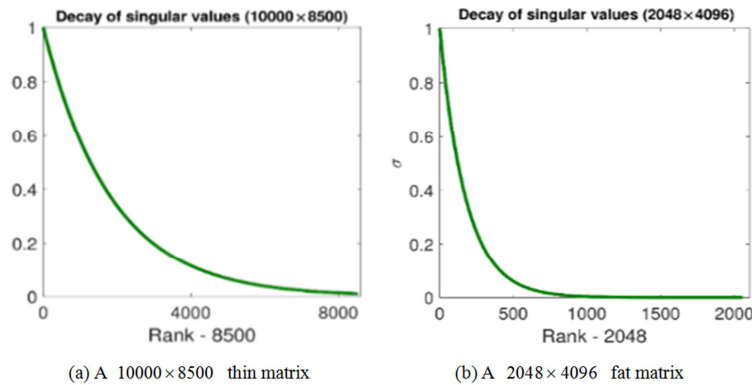
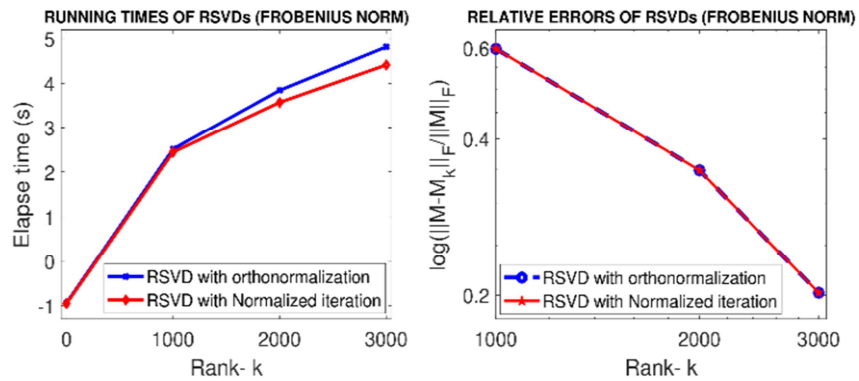


Figure 3. Decay of singular values of two distinct synthetic matrices of different sizes.

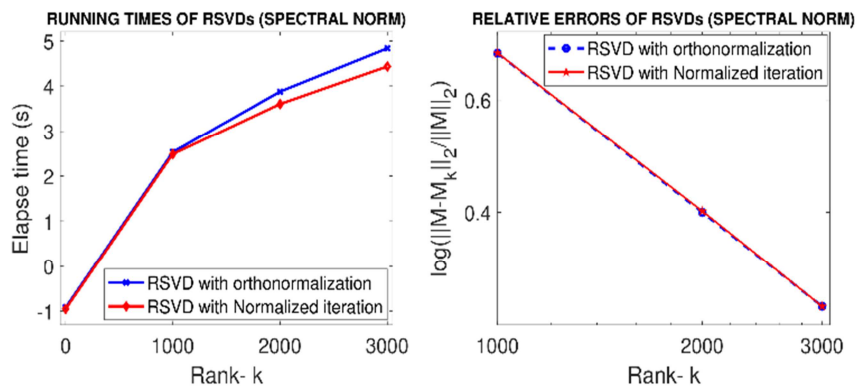
Table 2. Computational results for a 10000×8500 'thin' synthetic matrix, $\ell = 3000$.

Scheme	q	Time (s)	Error (%)
AE-RSVD (with orthonormalization)	1	60.02	21.6159
Accelerated AE-RSVD with modified normalized power iteration	1	41.40	21.6163
AE-RSVD (with orthonormalization)	2	86.23	20.3272
Accelerated AE-RSVD with modified normalized power iteration	2	61.72	20.3262
AE-RSVD (with orthonormalization)	3	112.71	19.9933
Accelerated AE-RSVD with modified normalized power iteration	3	84.36	19.9934

Figure 4 shows the graphical representation of the run times and relative reconstruction errors of both algorithms on the 10000×8500 synthetic matrix.



(a) Runtime and errors of the two RSVD in Frobenius norm

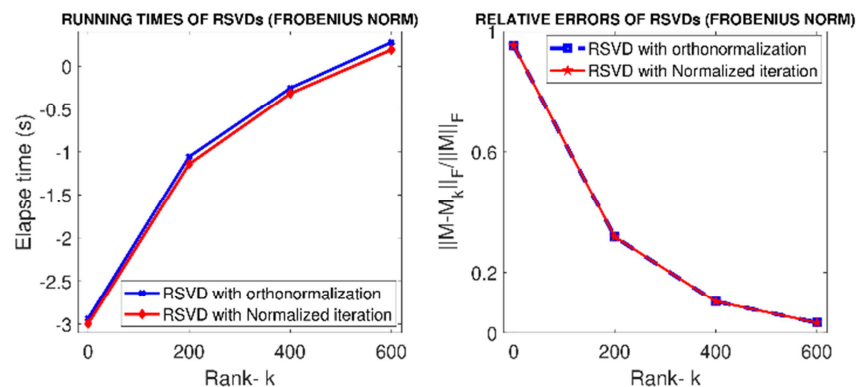


(b) Runtime and errors of the two RSVD in spectral norm

Figure 4. Runtime and relative approximate errors of both RSVDs on the 10000×8500 matrix (measured in Frobenius and spectral norms).

From Table 2 and Table 3, the accelerated AE-RSVD with modified normalized power iteration achieves speed-ups over the AE-RSVD (with orthonormalization) over

different q , while attaining nearly the same reconstruction errors on both matrices used in the numerical computations. Refer to Figure 4 and Figure 5.



(a) Runtime and errors of the two RSVD in Frobenius norm

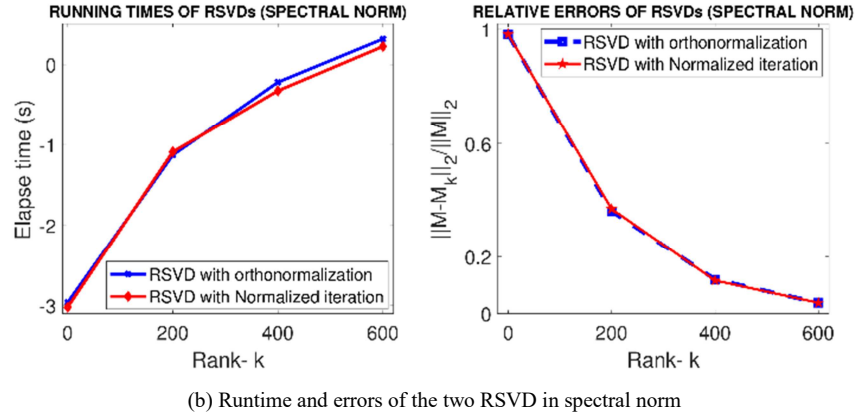


Figure 5. Runtime and relative approximate errors of both RSVDs on the 2048×4096 matrix (measured in Frobenius and spectral norms).

Table 3. Computational results for a 2048×4096 ‘fat’ synthetic matrix, whose singular values decay rapidly, $\ell = 600$.

Scheme	q	Time (s)	Error (%)
AE-RSVD (with orthonormalization)	1	0.76	35.2937
Accelerated AE-RSVD with modified normalized power iteration	1	0.68	35.2727
AE-RSVD (with orthonormalization)	2	1.05	31.1577
Accelerated AE-RSVD with modified normalized power iteration	2	0.93	35.2820
AE-RSVD (with orthonormalization)	3	1.46	29.1836
Accelerated AE-RSVD with modified normalized power iteration	3	1.25	29.0803

4.2. Image Compression

SVD is a common method for image compression. An image can be regenerated from a subset of the principal singular values and singular vectors. We use Benvenuto Tisi’s well-known painting “Garofalo Annunciation” which was obtained from the Google Art Project. This image has a high resolution of 30000×22882 . We seek to compute the low-rank

approximation of this high-resolution image via the two randomized algorithms to ascertain the time each algorithm takes.

To begin, we reduced the resolution to 5120×3072 pixels in order for the computer to bear the computational load. We seek to show the reconstruction of the data image with different $\ell = k + p$. Figure 6 depicts the original input of the “Garofalo Annunciation” image.

Table 4. Computational results for a 5120×3072 image of the ‘Garofalo Annunciation’ with different ℓ .

Scheme	ℓ	Time (s)	Error (%)
AE-RSVD (with orthonormalization)	20	0.15	16.97
Accelerated AE-RSVD	20	0.05	16.53
AE-RSVD (with orthonormalization)	100	0.39	9.83
Accelerated AE-RSVD	100	0.23	9.41
AE-RSVD (with orthonormalization)	1000	3.54	3.45
Accelerated AE-RSVD	1000	2.84	3.45



Figure 6. A 5120×3072 scaled image of the “Garofalo Annunciation” painted by Benvenuto Tisi.

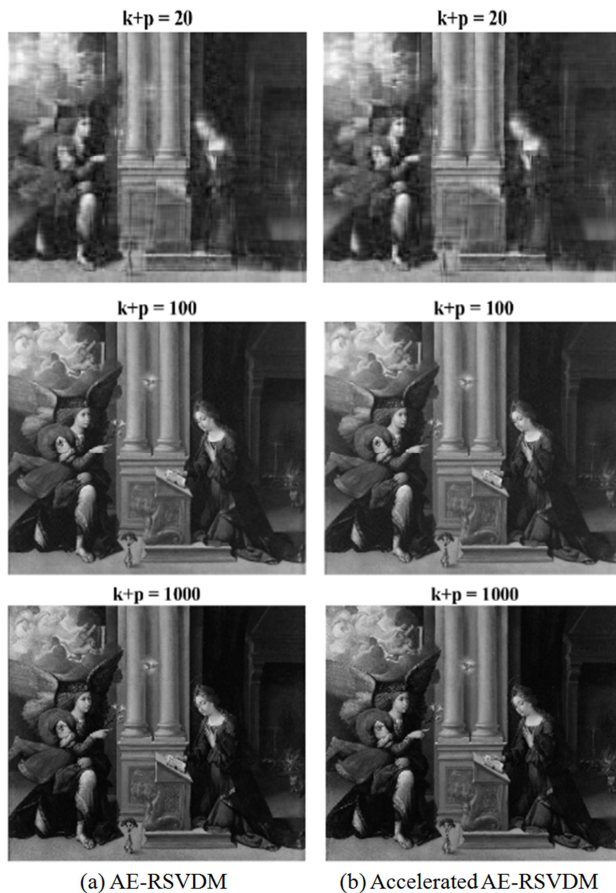


Figure 7. Reconstructed low-rank images with different rank- ℓ .

5. Conclusion

Reduced dimensionality and the associated theory of low-rank matrix approximations are crucial algorithmic techniques in many fields of study. Big data are, however, increasingly difficult to compute in classical matrix algorithms. The randomized SVD is undoubtedly the greatest influential and ubiquitous method for randomized algorithms. This work has examined both the runtime and the relative reconstruction errors of the accelerated AE-RSVDM with modified normalized power iteration against the AE-RSVDM (with orthonormalization) as formulated by [15]. Inspired by randomization, the computation of the approximate low-rank SVD of a big data matrix from a small compressed matrix is based on an efficient two-pass algorithm. The results show that the accelerated AE-RSVDM variant achieves a faster runtime than the AE-RSVDM (with orthonormalization) but does not achieve any significant reconstruction errors. Future research focuses on improving on the relative approximate error as well as implementing the accelerated variant in facial recognition.

Acknowledgements

The authors acknowledge the following persons for their constant review of the manuscript; Kwame Obour Opuni-Boachie Agyemang, Dr. Jerry John Kponyo, Elijah

Asante Boakye, Justice Owusu Agyemang and Fred Appiah-Twum.

References

- [1] Chan T and Hansen PC (1992) Some applications of the rank revealing QR factorization. *SIAM Journal on Scientific and Statistical Computing*, 13 (3), 727-741.
- [2] Martinsson P-G and Voronin S (2016) A randomized blocked algorithm for efficiently computing rank-revealing factorizations of matrices. *SIAM Journal on Scientific Computing*, 38 (5), S485-S507.
- [3] Halko N, Martinsson P-G and Tropp JA (2011) Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions. *SIAM review*, 53 (2), 217-288.
- [4] G. H. Golub and C. F. V Loan, *Linear Algebra and its Applications*. 1996.
- [5] L. N. Trefethen and D. Bau III, *Numerical Linear Algebra*. 50, SIAM, 1997.
- [6] Hastie T, Tibshirani R, Friedman J and Franklin J (2005) The elements of statistical learning: data mining, inference and prediction. *The Mathematical Intelligencer*, 27 (2), 83-85.
- [7] Khoromskij BN (2011) O ($d \log N$)-quantics approximation of N-d tensors in high-dimensional numerical modeling. *Constructive Approximation*, 34 (2), 257-280.
- [8] B. N. Erichson, S. L. Brunton and N. J. Kutz, "Compressed singular value decomposition for image and video processing," In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pp. 1880-1888, 2017.
- [9] Calvetti D, Reichel L and Sorensen DC (1994) An implicitly restarted Lanczos method for large symmetric eigenvalue problems. *Electronic Transactions on Numerical Analysis*, 2 (1), 21.
- [10] Frieze A, Kannan R and Vempala S (2004) Fast Monte-Carlo algorithms for finding low-rank approximations. *Journal of the ACM (JACM)*, 51 (6), 1025-1041.
- [11] T. Sarlos, "Improved approximation algorithms for large matrices via random projections," In *2006 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS'06)*, pp. 143-152, 2006.
- [12] Martinsson, P-G, Rokhlin V and Tygert M (2011) A randomized algorithm for the decomposition of matrices. *Applied and Computational Harmonic Analysis*, 30 (1), 47-68.
- [13] Woolfe F, Liberty E, Rokhlin V and Tygert M (2008) A fast randomized algorithm for the approximation of matrices. *Applied and Computational Harmonic Analysis*, 25 (3), 335-366.
- [14] Martinsson, P-G (2019) Randomized methods for matrix computations. *The Mathematics of Data*, 25, 187-231.
- [15] Rokhlin V, Szlam A and Tygert M (2010) A randomized algorithm for principal component analysis. *SIAM Journal on Matrix Analysis and Applications*, 31 (3), 1100-1124.

- [16] Erichson NB, Voronin S, Brunton, SL and Kutz, JN (2016) Randomized matrix decompositions using R. arXiv preprint arXiv: 1608.02148.
- [17] Gordon, Y (1985) Some inequalities for Gaussian processes and applications. Israel Journal of Mathematics, 50 (4), 265-289.
- [18] Balkema, A A, and De Haan L (1974) Residual life time at great age. The Annals of probability, 792-804.
- [19] Chen Z and Dongarra JJ (2005) Condition numbers of Gaussian random matrices. SIAM Journal on Matrix Analysis and Applications, 27 (3), 603-620.

Biography



Joseph Roger Arhin obtained his BSc degree in Mathematics from the University of Education, Winneba, Ghana, in 2017 and received the MSc degree in Mathematics with an orientation, Numerical Linear Algebra and Scientific Computing with Applications from University of Electronic Science and Technology of China, Chengdu, China, in

2020. He will be starting his PhD course in Information and Communication Engineering at the latter university in September 2020. His PhD research will focus on machine learning and artificial intelligence.



Francis Sam received the BSc degree in Mathematics from the University of Education Winneba, Ghana, in 2017. He is currently pursuing the MSc. degree in Mathematics with the School of Mathematical Sciences from the University of Electronic Science and Technology of China. He has a long-term interest in tensor

decompositions and their practical applications in massive MIMO. His broad research interests include Tensors and wireless communication.



Kenneth Coker had his BSc degree in Telecommunication Engineering from the Kwame Nkrumah University of Science and Technology, Ghana, in 2017 and obtained his MSc degree in Information and Communication Engineering from University of Electronic Science and Technology of China, Chengdu, China, in

2020. He is planning to commence his PhD course in Information and Communication Engineering at University of Electronic Science and Technology of China, Chengdu, China, in September 2020.



Toufic Seini attained his Bachelor's degree in Mathematics from the University for Development Studies, Ghana in 2016. He had his MSc degree in Mathematics with a research area of Modeling for Image and Visual Processing and High-Performance Numerical Algorithm

from the University of electronic science and technology of China, 2020.