A Randomized Algorithm for the Approximation of Matrices

Per-Gunnar Martinsson, Vladimir Rokhlin*, and Mark Tygert

Technical Report YALEU/DCS/TR-1352
June 7, 2006

Abstract

Given an $m \times n$ matrix A and a positive integer k, we introduce a randomized procedure for the approximation of A with a matrix B of rank k. The procedure relies on applying A^{T} to a collection of l random vectors, where l is an integer equal to or slightly greater than k; the scheme is efficient whenever A and A^{T} can be applied rapidly to arbitrary vectors. The discrepancy $\|B - A\|$ is of the same order as the $(k+1)^{\mathrm{st}}$ greatest singular value σ_{k+1} of A, with negligible probability of even moderately large deviations. The actual estimates derived in the paper are fairly complicated, but simplify when $l-k$ is fixed at a small nonnegative integer. For example, according to one of our estimates for $l - k = 20$, the probability that $\|B - A\|$ is greater than $10\sqrt{k\,(k+20)\,m\,n}\;\sigma_{k+1}$ is less than 10^{-17}. The paper contains a number of estimates for $\|B - A\|$, including several that are stronger (but more detailed) than the preceding example. The scheme provides a simple, efficient means for constructing an accurate approximation to the Singular Value Decomposition of A, and operates reliably independently of the structure of the matrix A. The results are illustrated via several numerical examples.

1 Introduction

In many practical circumstances, it is desirable to approximate a matrix A with a sum of rank-1 matrices. Such an approximation of A often facilitates understanding of the properties of A. Moreover, if the approximation involves only a small number of rank-1 matrices, then the approximation also facilitates rapid calculations involving A.

There are at least two classical forms of such matrix approximations. One is an approximation to a Singular Value Decomposition (SVD), which is known in the statistical literature as a Principal Component Analysis. The other is an approximation obtained via subset selection; we will refer to the matrix representation obtained via subset selection as

*Partially supported by the U.S. DoD under Air Force STTR Grant #F49620-03-C-0031 and DARPA Grant #HR0011-05-1-0002.

an interpolative decomposition. These two types of matrix approximations are defined as follows.

An approximation to an SVD of a real $m \times n$ matrix A consists of nonnegative real numbers σ_1, σ_2, ..., σ_{k-1}, σ_k known as singular values, orthonormal real $m \times 1$ column vectors u^1, u^2, ..., u^{k-1}, u^k known as left singular vectors, and orthonormal real $n \times 1$ column vectors v^1, v^2, ..., v^{k-1}, v^k known as right singular vectors, such that

$$\left\| A - \sum_{j=1}^{k} u^j \, \sigma_j \, (v^j)^{\mathrm{T}} \right\| \le \delta, \tag{1}$$

where k, m, and n are positive integers, δ is a positive real number specifying the precision of the approximation, and, for any matrix B, $\|B\|$ denotes the spectral (l^2-operator) norm of B, that is, $\|B\|$ is the greatest singular value of B. An approximation to an SVD of A is often written in the equivalent form

$$\left\| A - U \, \Sigma \, V^{\mathrm{T}} \right\| \le \delta, \tag{2}$$

where U is a real $m \times k$ matrix whose columns are orthonormal, V is a real $n \times k$ matrix whose columns are orthonormal, and Σ is a real $k \times k$ matrix whose entries are all nonnegative and whose entries off of the main diagonal are zero. See, for example, [18] for a discussion of SVDs.

An interpolative decomposition of a real $m \times n$ matrix A consists of a real $m \times k$ matrix B whose columns constitute a subset of the columns of A, and a real $k \times n$ matrix P, such that

1. some subset of the columns of P makes up the $k \times k$ identity matrix,

2. no entry of P has an absolute value greater than 2, and

3. $A = B\,P$.

See, for example, [17], [6], or Sections 4 and 5 of [4] for a discussion of interpolative decompositions.

Given an algorithm permitting the fast application of a numerically low-rank matrix A, and an algorithm permitting the fast application of A^{T}, the algorithm of the present paper provides a simple, efficient way for computing an accurate approximation to an SVD of A. Moreover, the algorithm provides a similar method for computing an accurate approximation to an interpolative decomposition of A under the same circumstances.

Our scheme also provides an efficient, robust means for approximating the k greatest singular values and corresponding singular vectors of any matrix A for which a representation enabling the fast application of both A and A^{T} is available. The precision δ of the resulting approximation given by formula (2) is at most a reasonably small multiple of the $(k+1)^{\mathrm{st}}$ greatest singular value of A. In this regard, the algorithm described below should be compared to the classical Lanczos method (for a description of the Lanczos method, see, for example, Chapter 9 in [15]).

Unlike the deterministic Lanczos scheme, the algorithm of the present paper is a randomized one, and fails with a rather negligible probability. Examples of the probabilities involved can be found in (92) and (103) in Section 4 below.

Some potential applications of the algorithm include finding the eigenmodes of certain networks, mining digital documents for information via latent semantic analysis, computing electron densities within the density functional theory of quantum chemistry, simplifying the implementation of algorithms for fast matrix inversion that are based on the compression of blocks within matrices, and improving condition number estimation and subspace determination algorithms that are based on inverse iteration.

We should point out that [16] and [4] motivated many aspects of the algorithm and analysis of the present paper. Moreover, a number of recent publications address issues similar to those addressed by the present paper; we refer the reader to [19] and [2], which describe deterministic methods, and to [1], [7], [8], [9], [10], [11], [12], [13], and the extensive references contained therein, all of which describe probabilistic Monte Carlo methods.

We do not analyze in detail the effects of round-off upon the algorithm of the present paper. However, most of the bounds that we discuss have finite-precision analogues. This is confirmed by both our preliminary analysis and our numerical experiments (some of which are described in Section 5 below). For simplicity, we discuss only real matrices; the analysis below extends easily to the complex case.

The present paper has the following structure: Section 2 collects together various known facts which later sections utilize, Section 3 provides the principal lemmas which Section 4 uses to construct algorithms, Section 4 describes the algorithm of the present paper, providing details about its accuracy and computational costs, and Section 5 illustrates the algorithm via several numerical examples.

2 Preliminaries from linear algebra and the theory of probability

In this section, we summarize various facts about matrices. Subsection 2.1 discusses the approximation of arbitrary matrices. Subsection 2.2 discusses the singular values of arbitrary matrices. Subsection 2.3 discusses the singular values of certain random matrices.

In the present section and throughout the rest of the paper, we employ the following notation. In accordance with the standard practice, we will denote the base of the natural logarithm by e. For any matrix A, we define the norm $\|A\|$ of A to be the spectral (l^2-operator) norm of A, that is, $\|A\|$ is the greatest singular value of A. For any positive integer n, and real $n \times 1$ column vector $v \in \mathbb{R}^n$, we define the norm $\|v\|$ of v to be the root-sum-square (l^2 norm) of the entries of v, that is,

$$\|v\| = \sqrt{\sum_{k=1}^{n} (v_k)^2}, \tag{3}$$

where v_k is the k^{th} entry of v. (Of course, the norm of v as viewed as a real $n \times 1$ matrix is equal to the norm of v as viewed as a real $n \times 1$ column vector.)

2.1 Approximation of general matrices

The following lemma states that, for any $m \times n$ matrix A whose rank is k, where k, m, and n are positive integers, there exist an $m \times k$ matrix B whose columns constitute a subset of the columns of A, and a $k \times n$ matrix P, such that

1. some subset of the columns of P makes up the $k \times k$ identity matrix,
2. P is not too large, and
3. $B\,P = A$.

Moreover, the lemma provides an analogous approximation $B\,P$ to A when the exact rank of A is not k, but the $(k+1)^{\text{st}}$ singular value of A is nevertheless small. The lemma is a reformulation of Theorem 3.2 in [17] and Theorem 3 in [6].

Lemma 1 *Suppose that m and n are positive integers, and A is a real $m \times n$ matrix.*

Then, for any positive integer k with $k \le m$ and $k \le n$, there exist a real $k \times n$ matrix P, and a real $m \times k$ matrix B whose columns constitute a subset of the columns of A, such that

1. *some subset of the columns of P makes up the $k \times k$ identity matrix,*
2. *no entry of P has an absolute value greater than* 1,
3. $\|P\| \le \sqrt{k\,(n-k)+1}$,
4. *the least (that is, the k^{th} greatest) singular value of P is at least* 1,
5. $B\,P = A$ *when* $k = m$ *or* $k = n$, *and*
6. $\|B\,P - A\| \le \sqrt{k\,(n-k)+1}\;\sigma_{k+1}$ *when $k < m$ and $k < n$, where σ_{k+1} is the $(k+1)^{st}$ greatest singular value of A.*

Remark 2 Properties 1, 2, 3, and 4 in Lemma 1 ensure that the interpolative decomposition $B\,P$ of A is numerically stable. Also, Property 3 follows directly from Properties 1 and 2, and Property 4 follows directly from Property 1.

Observation 3 There exists an algorithm which computes B and P in Lemma 1 from A, provided that we require only that

1. some subset of the columns of P makes up the $k \times k$ identity matrix,
2. no entry of P has an absolute value greater than 2,
3. $\|P\| \le \sqrt{4k\,(n-k)+1}$,
4. the least (that is, the k^{th} greatest) singular value of P is at least 1,
5. $B\,P = A$ when $k = m$ or $k = n$, and

6. $\|B\,P-A\| \le \sqrt{4k\,(n-k)+1}\;\sigma_{k+1}$ when $k < m$ and $k < n$, where σ_{k+1} is the $(k+1)^{\mathrm{st}}$ greatest singular value of A.

For any positive real number ε, the algorithm can identify the least k such that $\|B\,P-A\| \approx \varepsilon$. Furthermore, there exists a real number C such that the algorithm computes both B and P using at most $Ckmn\,\log(\min(m,n))$ floating-point operations and $Ckmn$ floating-point words of memory. The algorithm is based upon the Cramer rule and the ability to obtain the minimal-norm (or at least roughly minimal-norm) solutions to linear algebraic systems of equations (see [17], [6], and [16]).

2.2 Singular values of general matrices

The following technical lemma will be needed in Section 3.

Lemma 4 *Suppose that m and n are positive integers with $m \ge n$. Suppose further that A is a real $m \times n$ matrix such that $A^{\mathrm{T}}\,A$ is invertible.*

Then,

$$\left\|(A^{\mathrm{T}}\,A)^{-1}\,A^{\mathrm{T}}\right\| = \frac{1}{\sigma_n}, \tag{4}$$

where σ_n is the least (that is, the n^{th} greatest) singular value of A.

Proof. We form an SVD of A,

$$A = U\,\Sigma\,V^{\mathrm{T}}, \tag{5}$$

where U is a real unitary $m\times m$ matrix, Σ is a real $m\times n$ matrix whose entries are nonnegative everywhere and zero off of the main diagonal, and V is a real unitary $n\times n$ matrix. It follows from (5) that

$$(A^{\mathrm{T}}\,A)^{-1}\,A^{\mathrm{T}} = V\,(\Sigma^{\mathrm{T}}\,\Sigma)^{-1}\,\Sigma\,U^{\mathrm{T}}. \tag{6}$$

Combining (6) and the fact that U and V are unitary yields

$$\left\|(A^{\mathrm{T}}\,A)^{-1}\,A^{\mathrm{T}}\right\| = \left\|(\Sigma^{\mathrm{T}}\,\Sigma)^{-1}\,\Sigma^{\mathrm{T}}\right\|. \tag{7}$$

Combining (7) and the fact that Σ is zero off of the main diagonal yields (4). $\square$

The following lemma provides what is known as the Courant-Fischer maximin characterization of singular values; Theorem 8.1.2 in [15] provides an equivalent formulation of (8).

Lemma 5 *Suppose that m and n are positive integers, and A is a real $m \times n$ matrix.*

Then, the k^{th} greatest singular value σ_k of A is given by the formula

$$\sigma_k = \max_{S\subseteq\mathbb{R}^n:\,\dim S=k}\;\min_{v\in S:\,\|v\|\neq 0}\;\frac{\|A\,v\|}{\|v\|} \tag{8}$$

for all $k = 1,\,2,\,\ldots,\,\min(m,n)-1,\,\min(m,n)$, where the maximum is taken over all k-dimensional subspaces of $\mathbb{R}^n$, and the minimum is taken over all vectors in S that have nonzero norms.

The following lemma states that the singular values of the product $G\,A$ of matrices G and A are at most $\|G\|$ times greater than the corresponding singular values of A.

Lemma 6 *Suppose that l, m, and n are positive integers, A is a real $m \times n$ matrix, and G is a real $l \times m$ matrix.*

Then, the k^{th} greatest singular value ρ_k of the product $G\,A$ is at most a factor of $\|G\|$ times the k^{th} greatest singular value σ_k of A, that is,

$$\rho_k \le \|G\|\,\sigma_k \tag{9}$$

for all $k = 1,\, 2,\, \ldots,\, \min(l,m,n) - 1,\, \min(l,m,n)$.

Proof. For any vector $v \in \mathbb{R}^n$ with $\|v\| \neq 0$,

$$\frac{\|G\,A\,v\|}{\|v\|} \le \|G\|\,\frac{\|A\,v\|}{\|v\|}. \tag{10}$$

Combining (8) and (10) yields (9). □

The following lemma states that the greatest singular value of a matrix A is at least as large as the greatest singular value of any rectangular block of entries in A; the lemma is a straightforward consequence of the minimax properties of singular values (see, for example, Section 47 of Chapter 2 in [20]).

Lemma 7 *Suppose that k, l, m, and n are positive integers with $k \le m$ and $l \le n$. Suppose further that A is a real $m \times n$ matrix, and B is a $k \times l$ rectangular block of entries in A.*

Then, the greatest singular value of B is at most the greatest singular value of A.

The following lemma states that if the norm of the difference of two matrices is small, then their corresponding singular values are close; Corollary 8.6.2 in [15] provides an equivalent formulation of (11).

Lemma 8 *Suppose that m and n are positive integers, and A and R are real $m \times n$ matrices.*

Then, the k^{th} greatest singular value τ_k of the sum $A + R$ and the k^{th} greatest singular value σ_k of A differ by at most $\|R\|$, that is,

$$|\tau_k - \sigma_k| \le \|R\| \tag{11}$$

for all $k = 1,\, 2,\, \ldots,\, \min(m,n) - 1,\, \min(m,n)$.

2.3 Singular values of random matrices

The following lemma provides a highly probable upper bound on the greatest singular value of a square matrix whose entries are independent, identically distributed (i.i.d.) Gaussian random variables of zero mean and unit variance; Formula 8.8 of [14] provides an equivalent formulation of the lemma.

Lemma 9 *Suppose that n is a positive integer, G is a real $n \times n$ matrix whose entries are i.i.d. Gaussian random variables of zero mean and unit variance, and γ is a positive real number, such that $\gamma > 1$ and*

$$1 - \frac{1}{4\left(\gamma^2 - 1\right)\sqrt{\pi n \gamma^2}} \left(\frac{2\gamma^2}{e^{\gamma^2 - 1}}\right)^n \tag{12}$$

is nonnegative.

Then, the greatest singular value of G is at most $\sqrt{2n}\,\gamma$ with probability not less than the amount in (12).

Combining Lemmas 7 and 9 yields the following lemma, providing a highly probable upper bound on the greatest singular value of a rectangular matrix whose entries are i.i.d. Gaussian random variables of zero mean and unit variance.

Lemma 10 *Suppose that l, m, and n are positive integers with $n \geq l$ and $n \geq m$. Suppose further that G is a real $l \times m$ matrix whose entries are i.i.d. Gaussian random variables of zero mean and unit variance, and γ is a positive real number, such that $\gamma > 1$ and (12) is nonnegative.*

Then, the greatest singular value of G is at most $\sqrt{2n}\,\gamma$ with probability not less than the amount in (12).

The following lemma provides a highly probable lower bound on the least singular value of a rectangular matrix whose entries are i.i.d. Gaussian random variables of zero mean and unit variance; Formula 2.5 in [5] and the proof of Lemma 4.1 in [5] together provide an equivalent formulation of Lemma 11.

Lemma 11 *Suppose that k and l are positive integers with $k \leq l$. Suppose further that G is a real $l \times k$ matrix whose entries are i.i.d. Gaussian random variables of zero mean and unit variance, and β is a positive real number, such that*

$$1 - \frac{1}{\sqrt{2\pi\,(l - k + 1)}} \left(\frac{e}{(l - k + 1)\,\beta}\right)^{l-k+1} \tag{13}$$

is nonnegative.

Then, the least (that is, the k^{th} greatest) singular value of G is at least $1/(\sqrt{l}\;\beta)$ with probability not less than the amount in (13).

3 Mathematical apparatus

In this section, we describe the principal tools used in Section 4.

The following lemma states that the product $B\,P$ of matrices B and P is a good approximation to a matrix A, provided that there exists a matrix G such that

1. the columns of B constitute a subset of the columns of A,

2. $\|P\|$ is not too large,

3. $G\,B\,P$ is a good approximation to $G\,A$, and

4. there exists a matrix F such that $\|F\|$ is not too large, and $F\,G\,A$ is a good approximation to A.

Lemma 12 *Suppose that k, l, m, and n are positive integers with $k \le n$. Suppose further that A is a real $m \times n$ matrix, B is a real $m \times k$ matrix whose columns constitute a subset of the columns of A, P is a real $k \times n$ matrix, F is a real $m \times l$ matrix, and G is a real $l \times m$ matrix.*

Then,

$$\|B\,P - A\| \le \|F\,G\,A - A\|\,(\|P\| + 1) + \|F\|\,\|G\,B\,P - G\,A\|. \tag{14}$$

Proof. We observe that

$$\|B\,P - A\| \le \|B\,P - F\,G\,B\,P\| + \|F\,G\,B\,P - F\,G\,A\| + \|F\,G\,A - A\|, \tag{15}$$

$$\|B\,P - F\,G\,B\,P\| \le \|B - F\,G\,B\|\,\|P\|, \tag{16}$$

and

$$\|F\,G\,B\,P - F\,G\,A\| \le \|F\|\,\|G\,B\,P - G\,A\|. \tag{17}$$

Since the columns of B constitute a subset of the columns of A, it follows that the columns of $B - F\,G\,B$ constitute a subset of the columns of $A - F\,G\,A$, and therefore,

$$\|B - F\,G\,B\| \le \|A - F\,G\,A\|. \tag{18}$$

Combining (15), (16), (17), and (18) yields (14). $\square$

Remark 13 Since the columns of B constitute a subset of the columns of A in Lemma 12, it follows that the columns of $G\,B$ constitute a subset of the columns of $G\,A$. Conversely, whenever a matrix S is formed by gathering distinct columns of $G\,A$ together into S, then clearly $S = G\,B$ for some matrix B whose columns constitute a subset of the columns of A.

The following lemma states that, for any matrix A, and matrix G whose entries are i.i.d. Gaussian random variables of zero mean and unit variance, with very high probability there exists a matrix F with a reasonably small norm, such that $F\,G\,A$ is a good approximation to A.

Lemma 14 *Suppose that k, l, m, and n are positive integers with $k \le l$, such that $l < m$ and $l < n$. Suppose further that A is a real $m \times n$ matrix, G is a real $l \times m$ matrix whose entries are i.i.d. Gaussian random variables of zero mean and unit variance, and β and γ are positive real numbers, such that $\gamma > 1$ and*

$$1 - \frac{1}{\sqrt{2\pi\,(l-k+1)}} \left(\frac{e}{(l-k+1)\,\beta} \right)^{l-k+1} - \frac{1}{4\,(\gamma^2 - 1)\,\sqrt{\pi m \gamma^2}} \left(\frac{2\gamma^2}{e^{\gamma^2 - 1}} \right)^{m} \tag{19}$$

is nonnegative.

Then, there exists a real $m \times l$ matrix F such that

$$\|F\,G\,A - A\| \le \sqrt{2lm\beta^2\,\gamma^2 + 1}\;\sigma_{k+1} \tag{20}$$

and

$$\|F\| \le \sqrt{l}\;\beta \tag{21}$$

with probability not less than the amount in (19), where σ_{k+1} is the $(k+1)^{st}$ greatest singular value of A.

Proof. We prove the existence of a matrix F satisfying (20) and (21) by constructing one.

We start by forming an SVD of A,

$$A = U\,\Sigma\,V^{\mathrm{T}}, \tag{22}$$

where U is a real unitary $m \times m$ matrix, Σ is a real $m \times n$ matrix whose entries are nonnegative everywhere and zero off of the main diagonal, and V is a real unitary $n \times n$ matrix, such that

$$\Sigma_{i,i} = \sigma_i \tag{23}$$

for all $i = 1, 2, \ldots, \min(m,n) - 1, \min(m,n)$, where $\Sigma_{i,i}$ is the entry in row i and column i of Σ, and σ_i is the i^{th} greatest singular value of A.

Next, we define auxiliary matrices H, R, and P. We define H to be the leftmost $l \times k$ block of the $l \times m$ matrix $G\,U$, and R to be the rightmost $l \times (m-k)$ block of $G\,U$, so that

$$G\,U = \left(\begin{array}{c|c} H & R \end{array}\right). \tag{24}$$

Combining the facts that U is real and unitary, and that the entries of G are i.i.d. Gaussian random variables of zero mean and unit variance, we see that the entries of H are also i.i.d. Gaussian random variables of zero mean and unit variance, as are the entries of R. We define $H^{(-1)}$ to be the real $k \times l$ matrix given by the formula

$$H^{(-1)} = (H^{\mathrm{T}}\,H)^{-1}\,H^{\mathrm{T}}. \tag{25}$$

We define P to be the $m \times l$ matrix whose uppermost $k \times l$ block is $H^{(-1)}$, and whose entries in the lowermost $(m-k) \times l$ block are all zeros, so that

$$P = \left(\begin{array}{c} H^{(-1)} \\ \hline \mathbf{0} \end{array}\right). \tag{26}$$

Finally, we define F to be the $m \times l$ matrix given by

$$F = U\,P = U\,\left(\begin{array}{c} H^{(-1)} \\ \hline \mathbf{0} \end{array}\right). \tag{27}$$

Combining (25), (4), the fact that the entries of H are i.i.d. Gaussian random variables of zero mean and unit variance, and Lemma 11 yields

$$\left\|H^{(-1)}\right\| \le \sqrt{l}\;\beta \tag{28}$$

with probability not less than

$$1 - \frac{1}{\sqrt{2\pi\,(l-k+1)}} \left(\frac{e}{(l-k+1)\,\beta} \right)^{l-k+1}. \tag{29}$$

Combining (27), (28), and the fact that U is unitary yields (21).

We now show that F defined in (27) satisfies (20).

We define S to be the leftmost uppermost $k \times k$ block of Σ, and T to be the rightmost lowermost $(m-k) \times (n-k)$ block of Σ, so that

$$\Sigma = \left(\begin{array}{c|c} S & \mathbf{0} \\ \hline \mathbf{0} & T \end{array} \right). \tag{30}$$

Combining (22), (24), and (27) yields

$$F\,G\,A - A = U \left(\left(\begin{array}{c} H^{(-1)} \\ \hline \mathbf{0} \end{array} \right) \left(\begin{array}{c|c} H & R \end{array} \right) - \mathbf{1} \right) \Sigma\,V^{\mathrm{T}}. \tag{31}$$

Combining (25) and (30) yields

$$\left(\left(\begin{array}{c} H^{(-1)} \\ \hline \mathbf{0} \end{array} \right) \left(\begin{array}{c|c} H & R \end{array} \right) - \mathbf{1} \right) \Sigma = \left(\begin{array}{c|c} \mathbf{0} & H^{(-1)}\,R\,T \\ \hline \mathbf{0} & -T \end{array} \right). \tag{32}$$

Furthermore,

$$\left\| \left(\begin{array}{c|c} \mathbf{0} & H^{(-1)}\,R\,T \\ \hline \mathbf{0} & -T \end{array} \right) \right\|^2 \le \left\| H^{(-1)}\,R\,T \right\|^2 + \|T\|^2. \tag{33}$$

Moreover,

$$\left\| H^{(-1)}\,R\,T \right\| \le \left\| H^{(-1)} \right\| \, \|R\| \, \|T\|. \tag{34}$$

Combining (30) and (23) yields

$$\|T\| \le \sigma_{k+1}. \tag{35}$$

Combining (31), (32), (33), (34), (35), and the fact that U and V are unitary yields

$$\|F\,G\,A - A\| \le \sqrt{\left\|H^{(-1)}\right\|^2 \, \|R\|^2 + 1}\; \sigma_{k+1}. \tag{36}$$

Combining Lemma 10 and the fact that the entries of R are i.i.d. Gaussian random variables of zero mean and unit variance shows that

$$\|R\| \le \sqrt{2m}\,\gamma \tag{37}$$

with probability not less than

$$1 - \frac{1}{4\,(\gamma^2 - 1)\,\sqrt{\pi m \gamma^2}} \left(\frac{2\gamma^2}{e^{\gamma^2-1}} \right)^m. \tag{38}$$

Combining (36), (28), and (37) yields (20). □

The following lemma is very similar to Lemma 14. Lemma 15 is tighter than Lemma 14 when the singular values of the matrix A decay sufficiently fast, and the numbers j and l in the lemma are both much less than m.

Lemma 15 *Suppose that j, k, l, m, and n are positive integers with $k \le l$, such that $k+j < m$ and $k+j < n$, as well as $l < m$ and $l < n$. Suppose further that A is a real $m \times n$ matrix, G is a real $l \times m$ matrix whose entries are i.i.d. Gaussian random variables of zero mean and unit variance, and β and γ are positive real numbers, such that $\gamma > 1$ and*

$$\Phi = 1 - \frac{1}{\sqrt{2\pi\,(l-k+1)}} \left(\frac{e}{(l-k+1)\,\beta} \right)^{l-k+1} - \frac{1}{4\,(\gamma^2-1)\,\sqrt{\pi\,\max(m-k-j,l)\,\gamma^2}} \left(\frac{2\gamma^2}{e^{\gamma^2-1}} \right)^{\max(m-k-j,l)} - \frac{1}{4\,(\gamma^2-1)\,\sqrt{\pi\,\max(j,l)\,\gamma^2}} \left(\frac{2\gamma^2}{e^{\gamma^2-1}} \right)^{\max(j,l)} \tag{39}$$

is nonnegative.

Then, there exists a real $m \times l$ matrix F such that

$$\|F\,G\,A - A\| \le \sqrt{2l\,\max(j,l)\,\beta^2\,\gamma^2+1}\;\sigma_{k+1} + \sqrt{2l\,\max(m-k-j,l)\,\beta^2\,\gamma^2+1}\;\sigma_{k+j+1} \tag{40}$$

and

$$\|F\| \le \sqrt{l}\,\beta \tag{41}$$

with probability not less than the amount in (39), where σ_{k+1} is the $(k+1)^{st}$ greatest singular value of A, and σ_{k+j+1} is the $(k+j+1)^{st}$ greatest singular value of A.

Proof. We prove the existence of a matrix F satisfying (40) and (41) by constructing one.

We start by forming an SVD of A,

$$A = U\,\Sigma\,V^{\mathrm{T}}, \tag{42}$$

where U is a real unitary $m \times m$ matrix, Σ is a real $m \times n$ matrix whose entries are nonnegative everywhere and zero off of the main diagonal, and V is a real unitary $n \times n$ matrix, such that

$$\Sigma_{i,i} = \sigma_i \tag{43}$$

for all $i = 1, 2, \ldots, \min(m,n)-1, \min(m,n)$, where $\Sigma_{i,i}$ is the entry in row i and column i of Σ, and σ_i is the i^{th} greatest singular value of A.

Next, we define auxiliary matrices H, R, Γ, and P. We define H to be the leftmost $l \times k$ block of the $l \times m$ matrix $G\,U$, R to be the $l \times j$ block of $G\,U$ whose first column is the $(k+1)^{\text{st}}$ column of $G\,U$, and Γ to be the rightmost $l \times (m-j-k)$ block of $G\,U$, so that

$$G\,U = \left(\begin{array}{c|c|c} H & R & \Gamma \end{array} \right). \tag{44}$$

Combining the facts that U is real and unitary, and that the entries of G are i.i.d. Gaussian random variables of zero mean and unit variance, we see that the entries of H are also i.i.d. Gaussian random variables of zero mean and unit variance, as are the entries of R, and as are the entries of Γ. We define $H^{(-1)}$ to be the real $k \times l$ matrix given by the formula

$$H^{(-1)} = (H^{\mathrm{T}}\,H)^{-1}\,H^{\mathrm{T}}. \tag{45}$$

We define P to be the real $m \times l$ matrix whose uppermost $k \times l$ block is $H^{(-1)}$, whose entries in the $j \times l$ block whose first row is the $(k+1)^{\text{st}}$ row of P are all zeros, and whose entries in the lowermost $(m-k-j) \times l$ block are all zeros, so that

$$P = \left(\begin{array}{c} H^{(-1)} \\ \hline \mathbf{0} \\ \hline \mathbf{0} \end{array} \right). \tag{46}$$

Finally, we define F to be the $m \times l$ matrix given by

$$F = U\,P = U \left(\begin{array}{c} H^{(-1)} \\ \hline \mathbf{0} \\ \hline \mathbf{0} \end{array} \right). \tag{47}$$

Combining (45), (4), the fact that the entries of H are i.i.d. Gaussian random variables of zero mean and unit variance, and Lemma 11 yields

$$\left\| H^{(-1)} \right\| \leq \sqrt{l}\,\beta \tag{48}$$

with probability not less than

$$1 - \frac{1}{\sqrt{2\pi\,(l-k+1)}} \left(\frac{e}{(l-k+1)\,\beta} \right)^{l-k+1}. \tag{49}$$

Combining (47), (48), and the fact that U is unitary yields (41).

We now show that F defined in (47) satisfies (40).

We define S to be the leftmost uppermost $k \times k$ block of Σ, T to be the $j \times j$ block of Σ whose leftmost uppermost entry is the entry in the $(k+1)^{\text{st}}$ row and $(k+1)^{\text{st}}$ column of Σ, and Θ to be the rightmost lowermost $(m-k-j) \times (n-k-j)$ block of Σ, so that

$$\Sigma = \left(\begin{array}{c|c|c} S & \mathbf{0} & \mathbf{0} \\ \hline \mathbf{0} & T & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} & \Theta \end{array} \right). \tag{50}$$

Combining (42), (44), and (47) yields

$$F\,G\,A - A = U \left(\left(\begin{array}{c} H^{(-1)} \\ \hline \mathbf{0} \\ \hline \mathbf{0} \end{array} \right) \left(\begin{array}{c|c|c} H & R & \Gamma \end{array} \right) - \mathbf{1} \right) \Sigma\,V^{\mathrm{T}}. \tag{51}$$

Combining (45) and (50) yields

$$\left(\left(\begin{array}{c} H^{(-1)} \\ \hline \mathbf{0} \\ \hline \mathbf{0} \end{array} \right) \left(\begin{array}{c|c|c} H & R & \Gamma \end{array} \right) - \mathbf{1} \right) \Sigma = \left(\begin{array}{c|c|c} \mathbf{0} & H^{(-1)}\,R\,T & H^{(-1)}\,\Gamma\,\Theta \\ \hline \mathbf{0} & -T & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} & -\Theta \end{array} \right). \tag{52}$$

Furthermore,

$$\left\| \left(\begin{array}{c|c|c} \mathbf{0} & H^{(-1)}\,R\,T & H^{(-1)}\,\Gamma\,\Theta \\ \hline \mathbf{0} & -T & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} & -\Theta \end{array} \right) \right\|^2 \leq \left\| H^{(-1)}\,R\,T \right\|^2 + \left\| H^{(-1)}\,\Gamma\,\Theta \right\|^2 + \|T\|^2 + \|\Theta\|^2. \tag{53}$$

Moreover,

$$\left\|H^{(-1)}\,R\,T\right\| \le \left\|H^{(-1)}\right\|\;\|R\|\,\|T\| \tag{54}$$

and

$$\left\|H^{(-1)}\,\Gamma\,\Theta\right\| \le \left\|H^{(-1)}\right\|\;\|\Gamma\|\,\|\Theta\|. \tag{55}$$

Combining (50) and (43) yields

$$\|T\| \le \sigma_{k+1} \tag{56}$$

and

$$\|\Theta\| \le \sigma_{k+j+1}. \tag{57}$$

Combining (51), (52), (53), (54), (55), (56), (57), and the fact that U and V are unitary yields

$$\|F\,G\,A - A\|^2 \le \left(\left\|H^{(-1)}\right\|^2\;\|R\|^2 + 1\right)\;(\sigma_{k+1})^2 + \left(\left\|H^{(-1)}\right\|^2\;\|\Gamma\|^2 + 1\right)\;(\sigma_{k+j+1})^2. \tag{58}$$

Combining Lemma 10 and the fact that the entries of R are i.i.d. Gaussian random variables of zero mean and unit variance, as are the entries of Γ, yields

$$\|R\| \le \sqrt{2\,\max(j,l)}\;\gamma \tag{59}$$

and

$$\|\Gamma\| \le \sqrt{2\,\max(m-k-j,l)}\;\gamma, \tag{60}$$

with probability not less than

$$1 - \frac{1}{4\,(\gamma^2-1)\,\sqrt{\pi\,\max(m-k-j,l)\,\gamma^2}}\left(\frac{2\gamma^2}{e^{\gamma^2-1}}\right)^{\max(m-k-j,\,l)} - \frac{1}{4\,(\gamma^2-1)\,\sqrt{\pi\,\max(j,l)\,\gamma^2}}\left(\frac{2\gamma^2}{e^{\gamma^2-1}}\right)^{\max(j,\,l)}. \tag{61}$$

Combining (58), (48), (59), and (60) shows that

$$\|F\,G\,A - A\|^2 \le \left(2l\,\max(j,l)\,\beta^2\,\gamma^2 + 1\right)\;(\sigma_{k+1})^2 + \left(2l\,\max(m-k-j,l)\,\beta^2\,\gamma^2 + 1\right)\;(\sigma_{k+j+1})^2 \tag{62}$$

with probability not less than the amount in (39). Combining (62) and the fact that

$$\sqrt{x+y} \le \sqrt{x} + \sqrt{y} \tag{63}$$

for any nonnegative real numbers x and y yields (40). □

Given an $m \times n$ matrix A, and a matrix G whose entries are i.i.d. Gaussian random variables of zero mean and unit variance, the following lemma provides an upper bound on the singular values of the product $G\,A$ in terms of the singular values of A; the lemma is most useful when the singular values of A decay sufficiently fast, and the numbers j and l in the lemma are both much less than m.

Lemma 16 *Suppose that j, k, l, m, and n are positive integers with $k < l$, such that $k+j < m$ and $k+j < n$. Suppose further that A is a real $m \times n$ matrix, G is a real $l \times m$ matrix whose entries are i.i.d. Gaussian random variables of zero mean and unit variance, and γ is a positive real number, such that $\gamma > 1$ and*

$$\Psi = 1 - \frac{1}{4\,(\gamma^2-1)\,\sqrt{\pi\,\max(m-k-j,l)\,\gamma^2}} \left(\frac{2\gamma^2}{e^{\gamma^2-1}}\right)^{\max(m-k-j,l)} - \frac{1}{4\,(\gamma^2-1)\,\sqrt{\pi\,\max(k+j,l)\,\gamma^2}} \left(\frac{2\gamma^2}{e^{\gamma^2-1}}\right)^{\max(k+j,l)} \tag{64}$$

is nonnegative.

Then, the $(k+1)^{st}$ greatest singular value ρ_{k+1} of $G\,A$ is at most a certain linear combination of the $(k+1)^{st}$ greatest singular value σ_{k+1} of A and the $(k+j+1)^{st}$ greatest singular value σ_{k+j+1} of A, namely,

$$\rho_{k+1} \le \sqrt{2\,\max(k+j,l)}\;\gamma\,\sigma_{k+1} + \sqrt{2\,\max(m-k-j,l)}\;\gamma\,\sigma_{k+j+1}, \tag{65}$$

with probability not less than the amount in (64).

Proof. We start by forming an SVD of A,

$$A = U\,\Sigma\,V^{\mathrm{T}}, \tag{66}$$

where U is a real unitary $m \times m$ matrix, Σ is a real $m \times n$ matrix whose entries are nonnegative everywhere and zero off of the main diagonal, and V is a real unitary $n \times n$ matrix, such that

$$\Sigma_{i,i} = \sigma_i \tag{67}$$

for all $i = 1, 2, \ldots, \min(m,n)-1, \min(m,n)$, where $\Sigma_{i,i}$ is the entry in row i and column i of Σ, and σ_i is the i^{th} greatest singular value of A.

Combining (66) and the fact that V is unitary yields that $G\,A$ has the same singular values as $G\,U\,\Sigma$.

Next, we define auxiliary matrices H and R. We define H to be the leftmost $l \times (k+j)$ block of the $l \times m$ matrix $G\,U$, and R to be the rightmost $l \times (m-k-j)$ block of $G\,U$, so that

$$G\,U = \left(\begin{array}{c|c} H & R \end{array}\right). \tag{68}$$

Combining the facts that U is real and unitary, and that the entries of G are i.i.d. Gaussian random variables of zero mean and unit variance, we see that the entries of H are also i.i.d. Gaussian random variables of zero mean and unit variance, as are the entries of R.

Combining (68) and the fact that $G\,A$ has the same singular values as $G\,U\,\Sigma$ yields that $G\,A$ has the same singular values as $\left(\begin{array}{c|c} H & \mathbf{0} \end{array}\right)\Sigma + \left(\begin{array}{c|c} \mathbf{0} & R \end{array}\right)\Sigma$.

It follows from (67) that

$$\left\|\left(\begin{array}{c|c} \mathbf{0} & R \end{array}\right)\Sigma\right\| \le \|R\|\,\sigma_{k+j+1}. \tag{69}$$

Combining (11) and (69) yields

$$\rho_{k+1} \le \tau_{k+1} + \|R\|\, \sigma_{k+j+1}, \tag{70}$$

where ρ_{k+1} is the $(k+1)^{\text{st}}$ greatest singular value of $\left(\, H \,\middle|\, \mathbf{0} \, \right) \Sigma + \left(\, \mathbf{0} \,\middle|\, R \, \right) \Sigma$, and τ_{k+1} is the $(k+1)^{\text{st}}$ greatest singular value of $\left(\, H \,\middle|\, \mathbf{0} \, \right) \Sigma$; ρ_{k+1} is also the $(k+1)^{\text{st}}$ greatest singular value of $G\, A$, since $G\, A$ has the same singular values as $\left(\, H \,\middle|\, \mathbf{0} \, \right) \Sigma + \left(\, \mathbf{0} \,\middle|\, R \, \right) \Sigma$.

Furthermore,

$$\left\| \left(\, H \,\middle|\, \mathbf{0} \, \right) \right\| \le \|H\|. \tag{71}$$

Combining (9), (71), and (67) yields

$$\tau_{k+1} \le \|H\|\, \sigma_{k+1}. \tag{72}$$

Combining Lemma 10 and the fact that the entries of H are i.i.d. Gaussian random variables of zero mean and unit variance, as are the entries of R, shows that

$$\|H\| \le \sqrt{2\, \max(k+j,l)}\, \gamma \tag{73}$$

and

$$\|R\| \le \sqrt{2\, \max(m-k-j,l)}\, \gamma, \tag{74}$$

with probability not less than the amount in (64).

Combining (70), (72), (73), and (74) yields (65). □

The following lemma provides an efficient means of computing an SVD of a real $m \times n$ matrix A from a real $m \times k$ matrix B and a real $k \times n$ matrix P such that $A = B\, P$ and k is much less than both m and n. If, in addition, $\|B\| \le \|A\|$ and $\|P\|$ is not too large, then the scheme described by the lemma is numerically stable.

Lemma 17 *Suppose that k, m, and n are positive integers with $k \le n$. Suppose further that A is a real $m \times n$ matrix, B is a real $m \times k$ matrix, and P is a real $k \times n$ matrix, such that*

$$A = B\, P. \tag{75}$$

Suppose in addition that L is a real $k \times k$ matrix, and Q is a real $n \times k$ matrix whose columns are orthonormal, such that

$$P = L\, Q^{\mathrm{T}}. \tag{76}$$

Suppose finally that C is a real $m \times k$ matrix, U is a real $m \times k$ matrix whose columns are orthonormal, Σ is a real $k \times k$ matrix, and W is a real $k \times k$ matrix whose columns are orthonormal, such that

$$C = B\, L \tag{77}$$

and

$$C = U\, \Sigma\, W^{\mathrm{T}}. \tag{78}$$

Then,

$$A = U\, \Sigma\, V^{\mathrm{T}}, \tag{79}$$

where V is the real $n \times k$ matrix given by the formula

$$V = Q\,W. \tag{80}$$

Moreover, the columns of V are orthonormal (as are the columns of U), and

$$\|L\| = \|P\|. \tag{81}$$

Proof. Combining (75), (76), (77), (78), and (80) yields (79). Combining (80) and the facts that W is unitary and the columns of Q are orthonormal yields that the columns of V are orthonormal. Combining (76) and the fact that the columns of Q are orthonormal yields (81). $\square$

Remark 18 The matrices L and Q in (76) can be computed from P as follows. Using the algorithms described, for example, in Chapter 5 of [15], we construct an upper triangular real $k \times k$ matrix R, and a real $n \times k$ matrix Q whose columns are orthonormal, such that

$$P^{\mathrm{T}} = Q\,R. \tag{82}$$

We thus obtain Q. We then define L to be the transpose of R, that is,

$$L = R^{\mathrm{T}}. \tag{83}$$

4 Description of the algorithm

In this section, we describe the algorithm of the present paper. In Subsection 4.1, we discuss approximations to interpolative decompositions. In Subsection 4.2, we discuss approximations to SVDs. In Subsection 4.3, we tabulate the computational costs of various parts of the algorithm. In Subsection 4.4, we describe Table 1.

4.1 Interpolative decomposition

Suppose that k, m, and n are positive integers with $k < m$ and $k < n$, and A is a real $m \times n$ matrix. In this subsection, we will collect together k appropriately chosen columns of A into a real $m \times k$ matrix B, and construct a real $k \times n$ matrix P, such that

$$\|P\| \le \sqrt{4k\,(n-k)+1} \tag{84}$$

and

$$\|B\,P - A\| \lesssim \sigma_{k+1}, \tag{85}$$

where σ_{k+1} is the $(k+1)^{\mathrm{st}}$ greatest singular value of A. To do so, we select an integer l with $l > k$, such that $l < m$ and $l < n$ ($l = k + 20$ is often a suitable choice), and make the following three steps:

1. Using a random number generator, form a real $l \times m$ matrix G whose entries are i.i.d. Gaussian random variables of zero mean and unit variance, and compute the $l \times n$ product matrix
$$R = G\,A. \tag{86}$$

2. Using the algorithm of [6], form a real $l \times k$ matrix S whose columns constitute a subset of the columns of R, and a real $k \times n$ matrix P satisfying (84), such that
$$\|S\,P - R\| \le \sqrt{4k\,(n-k)+1}\;\rho_{k+1}, \tag{87}$$
where ρ_{k+1} is the $(k+1)^{\text{st}}$ greatest singular value of R. (See Observation 3 for a brief discussion of the properties of the algorithm of [6].)

3. Using the fact that the columns of S constitute a subset of the columns of R, for any $j = 1, 2, \ldots, k-1, k$, let i_j denote an integer such that the j^{th} column of S is the ${i_j}^{\text{th}}$ column of R. Form the real $m \times k$ matrix B whose j^{th} column is the ${i_j}^{\text{th}}$ column of A for all $j = 1, 2, \ldots, k-1, k$.

The matrices B and P obtained via the preceding three steps satisfy (84) and (85); see the following remark.

Remark 19 In this remark, we demonstrate that the matrices B and P satisfy (85). Indeed, combining (86) and Remark 13 yields
$$S = G\,B. \tag{88}$$
Combining (87), (86), and (88) yields
$$\|G\,B\,P - G\,A\| \le \sqrt{4k\,(n-k)+1}\;\rho_{k+1}, \tag{89}$$
where ρ_{k+1} is the $(k+1)^{\text{st}}$ greatest singular value of R. Suppose that β and γ are positive real numbers such that $\gamma > 1$ and
$$\chi = 1 - \frac{1}{\sqrt{2\pi\,(l-k+1)}}\left(\frac{e}{(l-k+1)\,\beta}\right)^{l-k+1} - \frac{1}{2\,(\gamma^2-1)\,\sqrt{\pi m \gamma^2}}\left(\frac{2\gamma^2}{e^{\gamma^2-1}}\right)^{m} \tag{90}$$
is nonnegative. Then, combining (14), (20), (21), (84), (89), (9), (86), and Lemma 10 yields
$$\begin{aligned}&\|B\,P - A\| \\ &\quad\le \left(\sqrt{2lm\beta^2\,\gamma^2+1}\;\left(\sqrt{4k\,(n-k)+1}+1\right) + \beta\,\gamma\,\sqrt{2lm}\,\sqrt{4k\,(n-k)+1}\right)\sigma_{k+1}\end{aligned} \tag{91}$$
with probability not less than χ defined in (90), where σ_{k+1} is the $(k+1)^{\text{st}}$ greatest singular value of A. The bound (91) is a precise version of (85). For example, choosing $\beta = 3/4$, $\gamma^2 = 5$, and $l = k+20$, and combining (91) and (90), we obtain
$$\|B\,P - A\| \le 10\,\sqrt{k\,(k+20)\,mn}\;\sigma_{k+1} \tag{92}$$
with probability not less than $1 - 10^{-17}$. Table 1 contains similar results obtained by taking other values for $l-k$, β, and γ.

Remark 20 When the singular values of A decay sufficiently fast, and l is much less than m, the factors $\sqrt{2lm\beta^2\gamma^2+1}$ and $\sqrt{2lm}$ in (91) are much larger than necessary. Indeed, suppose that j is a positive integer with $k+j<m$ and $k+j<n$, and β and γ are positive real numbers, such that $\gamma>1$ and $\Phi+\Psi>1$, where Φ is defined in (39), and Ψ is defined in (64). Then, combining (14), (40), (41), (84), (89), (65), and (86) yields

$$\|B\,P - A\| \le \xi\,\sigma_{k+1} + \eta\,\sigma_{k+j+1} \tag{93}$$

with probability not less than $\Phi+\Psi-1$, where Φ is defined in (39), Ψ is defined in (64), σ_{k+1} is the $(k+1)^{\mathrm{st}}$ greatest singular value of A, and σ_{k+j+1} is the $(k+j+1)^{\mathrm{st}}$ greatest singular value of A, and where

$$\xi = \sqrt{2l\,\max(j,l)\,\beta^2\,\gamma^2+1}\,\left(\sqrt{4k\,(n-k)+1}+1\right) + \beta\,\gamma\,\sqrt{2l\,\max(k+j,l)}\,\sqrt{4k\,(n-k)+1} \tag{94}$$

and

$$\eta = \sqrt{2l\,\max(m-k-j,l)\,\beta^2\,\gamma^2+1}\,\left(\sqrt{4k\,(n-k)+1}+1\right) + \beta\,\gamma\,\sqrt{2l\,\max(m-k-j,l)}\,\sqrt{4k\,(n-k)+1}. \tag{95}$$

When j, k, and l are all much less than m, clearly ξ is much less than η.

Remark 21 If we choose $l=k$ in the algorithm of the present subsection (instead of choosing $l>k$), then we must replace (87) with the formula

$$\|S\,P - R\| = 0. \tag{96}$$

All other aspects of the algorithm stay the same in the case that $l=k$. In particular, (91) and (93) hold in the case that $l=k$, too.

4.2 Singular Value Decomposition

Suppose that k, m, and n are positive integers with $k<m$ and $k<n$, and A is a real $m\times n$ matrix. In this subsection, we will compute an approximation to an SVD of A such that

$$\|U\,\Sigma\,V^{\mathrm{T}} - A\| \lesssim \sigma_{k+1}, \tag{97}$$

where U is a real $m\times k$ matrix whose columns are orthonormal, V is a real $n\times k$ matrix whose columns are orthonormal, Σ is a diagonal real $k\times k$ matrix whose entries are all nonnegative, and σ_{k+1} is the $(k+1)^{\mathrm{st}}$ greatest singular value of A. To do so, we first use the algorithm of Subsection 4.1 to construct the matrices B and P in (84), (91), and (93). Then, we make the following four steps:

1. Construct a lower triangular real $k\times k$ matrix L, and a real $n\times k$ matrix Q whose columns are orthonormal, such that

$$P = L\,Q^{\mathrm{T}} \tag{98}$$

(see Remark 18 for details concerning the construction of such matrices L and Q).

2. Form the product

$$C = B\,L. \tag{99}$$

3. Construct an SVD of C, that is,

$$C = U\,\Sigma\,W^{\mathrm{T}}, \tag{100}$$

   where U is a real $m \times k$ matrix whose columns are orthonormal, Σ is a diagonal $k \times k$ matrix whose entries are all nonnegative, and W is a real $k \times k$ matrix whose columns are orthonormal (see, for example, Chapter 8 in [15] for details concerning the construction of such an SVD).

4. Form the product

$$V = Q\,W. \tag{101}$$

The matrices U, Σ, and V obtained via the preceding four steps (after first using the algorithm of Subsection 4.1 to construct the matrices B and P) satisfy (97); see Remark 23 below.

Remark 22 Steps 2 and 4 in the procedure of the present subsection are somewhat subtle numerically. Both Steps 2 and 4 involve constructing products of matrices, and in general constructing the product $S\,T$ of matrices S and T can be numerically unstable. Indeed, in general some entries of S or T can have unmanageably large absolute values, while in exact arithmetic no entry of the product $S\,T$ has an unmanageably large absolute value; in such circumstances, constructing the product $S\,T$ can be unstable in finite-precision arithmetic. However, this problem does not arise in Steps 2 and 4 above, due to (84), (81), the fact that the columns of B constitute a subset of the columns of A (so that $\|B\| \le \|A\|$), and the fact that the columns of Q are orthonormal, as are the columns of W.

Remark 23 In this remark, we demonstrate that the matrices U, Σ, and V satisfy (97). Indeed, suppose that β and γ are positive real numbers such that $\gamma > 1$ and χ defined in (90) is nonnegative. Then, combining (91), (75), (98), (99), (100), (101), and (79) yields

$$\begin{aligned} &\|U\,\Sigma\,V^{\mathrm{T}} - A\| \\ &\quad\le \left(\sqrt{2lm\beta^2\,\gamma^2+1}\;\left(\sqrt{4k\,(n-k)+1}+1\right) + \beta\,\gamma\,\sqrt{2lm}\,\sqrt{4k\,(n-k)+1}\right)\sigma_{k+1} \end{aligned} \tag{102}$$

with probability not less than χ defined in (90), where σ_{k+1} is the $(k+1)^{\mathrm{st}}$ greatest singular value of A. The bound (102) is a precise version of (97). As in Subsection 4.1, if we choose $\beta = 3/4$, $\gamma^2 = 5$, and $l = k + 20$, and combine (102) and (90), then we obtain

$$\|U\,\Sigma\,V^{\mathrm{T}} - A\| \le 10\,\sqrt{k\,(k+20)\,mn}\;\sigma_{k+1} \tag{103}$$

with probability not less than $1 - 10^{-17}$. Table 1 contains similar results obtained by taking other values for $l - k$, β, and γ.

Remark 24 When the singular values of A decay sufficiently fast, and l is much less than m, the factors $\sqrt{2lm\beta^2\gamma^2+1}$ and $\sqrt{2lm}$ in (102) are much larger than necessary. Indeed, suppose that j is a positive integer with $k+j<m$ and $k+j<n$, and β and γ are positive real numbers, such that $\gamma>1$ and $\Phi+\Psi>1$, where Φ is defined in (39), and Ψ is defined in (64). Then, combining (93), (75), (98), (99), (100), (101), and (79) yields

$$\|U\,\Sigma\,V^{\mathrm{T}}-A\| \le \xi\,\sigma_{k+1}+\eta\,\sigma_{k+j+1} \tag{104}$$

with probability not less than $\Phi+\Psi-1$, where Φ is defined in (39), Ψ is defined in (64), σ_{k+1} is the $(k+1)^{\mathrm{st}}$ greatest singular value of A, σ_{k+j+1} is the $(k+j+1)^{\mathrm{st}}$ greatest singular value of A, ξ is defined in (94), and η is defined in (95). When j, k, and l are all much less than m, clearly ξ is much less than η.

4.3 CPU time and memory requirements

In this subsection, we tabulate the numbers of floating-point operations and words of memory required by the algorithms described in Subsections 4.1 and 4.2, as applied once to a matrix A.

4.3.1 Interpolative decomposition

The algorithm described in Subsection 4.1 incurs the following costs in order to compute an approximation to an interpolative decomposition of A:

1. Forming R in (86) requires applying A^{T} to l vectors.
2. Computing S and P in (87) or (96) costs $\mathcal{O}(lkn\,\log(l))$.
3. Forming B in (88) requires applying A to k vectors, where each vector has a single entry of 1 and $n-1$ entries of 0.

Summing up the costs in Steps 1–3 above, we conclude that the algorithm of Subsection 4.1 costs

$$C_{\mathrm{ID}} = k\cdot C_A + l\cdot C_{A^{\mathrm{T}}} + \mathcal{O}(lkn\,\log(l)), \tag{105}$$

where C_A is the cost of applying A to a real $n\times 1$ column vector, and $C_{A^{\mathrm{T}}}$ is the cost of applying A^{T} to a real $m\times 1$ column vector.

4.3.2 Singular Value Decomposition

The algorithm described in Subsection 4.2 incurs the following costs in order to compute an approximation to an SVD of A from the matrices B and P obtained via the algorithm described in Subsection 4.1:

1. Computing the matrices L and Q in (98) costs $\mathcal{O}(k^2\,n)$.
2. Forming C in (99) costs $\mathcal{O}(k^2\,m)$.
3. Computing the SVD (100) of C costs $\mathcal{O}(k^2\,m)$.

4. Forming V defined in (101) costs $\mathcal{O}(k^2\, n)$.

Summing up the costs in Steps 1–4 above, we conclude that the algorithm of Subsection 4.2 costs

$$C_{\mathrm{SVD}} = C_{\mathrm{ID}} + \mathcal{O}(k^2\,(m+n)), \tag{106}$$

where C_{ID} is the cost in (105) of applying the algorithm described in Subsection 4.1 in order to obtain B and P.

Remark 25 We observe that in order to form the matrix $G\,A$ in (86), we need only to take each column of A and form its inner products with the l rows of G. In order to form B in (88), we need only to collect together certain columns of A. Thus, we only ever need to access columns of A, and never need to transpose A explicitly. The data movement required by the algorithm of the present paper is hence fairly simple.

4.4 Description of Table 1

Tables 1.1–1.6 provide an upper bound $\Pi_{l-k,\beta,\gamma}$ on the probability that

$$\|U\,\Sigma\,V^{\mathrm{T}} - A\| > \zeta\,\sqrt{klmn}\,\sigma_{k+1}, \tag{107}$$

where U, Σ, and V are the matrices in the approximation to an SVD of the $m \times n$ matrix A in (102). In (107), k and l are any positive integers with $k \le l$, such that $l < m$ and $l < n$, σ_{k+1} is the $(k+1)^{\mathrm{st}}$ greatest singular value of A, and ζ takes on the values specified by the penultimate columns of the tables. The quantity $\Pi_{l-k,\beta,\gamma}$ is defined by the formula

$$\Pi_{l-k,\beta,\gamma} = 1 - \chi, \tag{108}$$

where χ is defined in (90), and $l-k$, β, and γ take on the values specified by the first, second, and third columns of the tables. The quantity ζ is specified by β and γ via (102). Please note that $\Pi_{l-k,\beta,\gamma}$ depends only on $l-k$, β, and γ, and provides an upper bound that is otherwise independent of k, m, n, and A; (103) provides a similar bound. When the singular values of A decay sufficiently fast, and l is much less than m, the factor of $\sqrt{m}$ in the right-hand side of (107) is larger than necessary; see Remark 24 above.

Remark 26 Due to (91), the quantity $\Pi_{l-k,\beta,\gamma}$ defined in (108) also provides an upper bound on the probability that

$$\|B\,P - A\| > \zeta\,\sqrt{klmn}\,\sigma_{k+1}, \tag{109}$$

where B and P are the matrices in the approximation to an interpolative decomposition of the $m \times n$ matrix A in (91).

5 Numerical results

In this section, we describe the results of four numerical tests of the algorithm of the present paper. Table 2 summarizes the numerical output of the examples described in the present section.

Tables 2.1–2.4 display the results of applying the algorithm of the present paper once to a real $n \times n$ matrix A, for the indicated values of n. The matrix A is defined at the end of the present section. The numbers k and l are those from Section 4; k is the rank of the approximations to A, and l is the number of rows in the matrix G whose entries are i.i.d. Gaussian random variables of zero mean and unit variance (the algorithm uses the product $G\,A$). The displayed times refer to the seconds of CPU time used by the algorithm to compute both the approximation to an interpolative decomposition and the approximation to an SVD of A. (Please note that our implementation is optimized for accuracy and for analyzing the numerical properties of the algorithm, and is probably not very efficient.) The numbers σ_k and σ_{k+1} are those from the definition of A below; furthermore, σ_{k+1} appears in the bounds (91) and (102) on the errors of the approximations. The number δ_{ID} is the norm of the difference between A and an approximation $B\,P$ to an interpolative decomposition of A, that is,

$$\delta_{\mathrm{ID}} = \|B\,P - A\|, \tag{110}$$

where the matrices B and P are those from (91). The number δ_{SVD} is the norm of the difference between A and the approximation $U\,\Sigma\,V^{\mathrm{T}}$ to an SVD of A, that is,

$$\delta_{\mathrm{SVD}} = \|U\,\Sigma\,V^{\mathrm{T}} - A\|, \tag{111}$$

where the matrices U, Σ, and V are those from (102).

We define $\delta_{\text{rel. max.}}$ as follows. First, we define auxiliary vectors t^1, t^2, …, t^{j-1}, t^j, and τ^1, τ^2, …, τ^{j-1}, τ^j, where $j = 30$ in Examples 1 and 2, $j = 110$ in Example 3, and $j = 6$ in Example 4. We choose the test vectors t^1, t^2, …, t^{j-1}, t^j to include a variety of deterministic and random vectors (specifically, we set every entry of t^1 to be 1, and use a random number generator to generate t^2, t^3, …, t^{j-1}, t^j so that their entries are i.i.d. Gaussian random variables of zero mean and unit variance). For any $i = 1, 2, \ldots, j-1, j$, we define τ^i to be the vector resulting from the application of $B\,P - A$ to t^i, that is,

$$\tau^i = (B\,P - A)\,t^i, \tag{112}$$

where the matrices B and P are those from (91). Then, we define $\delta_{\text{rel.max.}}$ via the formula

$$\delta_{\text{rel. max.}} = \max_{i=1,2,\ldots,j-1,j} \left(\frac{\max_{p=1,2,\ldots,m-1,m} |(\tau^i)_p|}{\max_{q=1,2,\ldots,n-1,n} |(t^i)_q|} \right), \tag{113}$$

where $(\tau^i)_p$ is the p^{th} entry of τ^i, and $(t^i)_q$ is the q^{th} entry of t^i.

All estimates displayed in Table 2 are the maximum values obtained from three independent realizations of the random variables involved.

The values of δ_{ID} and δ_{SVD} displayed in Tables 2.2, 2.3, and 2.4 are those obtained via the power method for estimating the norm of a matrix, after the estimates stabilized to three significant figures. The values of δ_{ID} and δ_{SVD} displayed in Table 2.1 are those obtained after 100 iterations of the power method. The estimates of δ_{ID} and δ_{SVD} summarized in Table 2.1 did not stabilize to three significant figures after 100 (or any other number of) iterations, undoubtedly due to round-off.

We performed all computations using IEEE standard double-precision variables, whose mantissas have approximately one bit of precision more than 16 digits (so that the relative precision of the variables is approximately .2e-15). We ran all computations on a 2.8 GHz Pentium Xeon microprocessor with 512 KB of L2 cache and 2 GB of RAM. We compiled the Fortran 77 code using the Lahey-Fujitsu compiler, with the optimization flag `--o2` enabled.

In our implementation, we computed SVDs using 2-sided plane (Jacobi/Givens) rotations (see, for example, Chapter 8 in [15]). We used an algorithm based upon pivoted "$Q\,R$" decompositions to compute the matrices S and P in (87) and (96) (see, for example, Chapter 5 in [15] for a description of "$Q\,R$" decompositions, and [6] for further details regarding our particular implementation).

In Examples 1, 2, and 3, we use a pseudorandom number generator to construct real $n \times 1$ vectors μ^1, μ^2, …, μ^{j-1}, μ^j, and ν^1, ν^2, …, ν^{j-1}, ν^j, such that their entries are a realization of i.i.d. Gaussian random variables of zero mean and unit variance, with $j = 20$ in Examples 1 and 2, and $j = 60$ in Example 3. We orthonormalize μ^1, μ^2, …, μ^{j-1}, μ^j via the Gram-Schmidt process with reorthogonalization (see, for example, [3]) to obtain real $n \times 1$ vectors u^1, u^2, …, u^{j-1}, u^j, and do the same with ν^1, ν^2, …, ν^{j-1}, ν^j to obtain real $n \times 1$ vectors v^1, v^2, …, v^{j-1}, v^j. We denote by σ_1, σ_2, …, σ_{j-1}, σ_j the positive real numbers displayed in Figure 1 (we use the numbers in Figure 1.1 for Example 1, the numbers in Figure 1.2 for Example 2, and those in Figure 1.3 for Example 3). We define A to be the $n \times n$ matrix given by the formula

$$A = \sum_{i=1}^{j} u^i\, \sigma_i\, (v^i)^{\mathrm{T}}. \tag{114}$$

Clearly, the rank of A is j. Since u^1, u^2, …, u^{j-1}, u^j are orthonormal, as are v^1, v^2, …, v^{j-1}, v^j, the i^{th} singular value of A is σ_i, for all $i = 1, 2, \ldots, j-1, j$. Table 2 displays the results of applying the algorithm of the present paper to A, for various values of n (Table 2.1 displays the results for Example 1, Table 2.2 displays the results for Example 2, and Table 2.3 displays those for Example 3).

Example 4 is designed to illustrate that factors of the order of $\sqrt{4k\,(n-k)+1}$ are necessary in bounds such as (91), (93), (102), and (104). This example is identical to Examples 1, 2, and 3, using the same matrix A defined in (114), but with n assumed to be divisible by 8, with $j = 4$, and using the numbers σ_1, σ_2, σ_3, and σ_4 displayed in Figure 1.4 ($\sigma_1 = 1$, $\sigma_2 = 1$, $\sigma_3 = .1\text{e-}7$, and $\sigma_4 = .1\text{e-}7$), and with the following vectors u^1, u^2, u^3, u^4, and v^1, v^2, v^3, v^4:

$$(u^1)^{\mathrm{T}} = \frac{1}{\sqrt{n}} \begin{pmatrix} 1 & 1 & \ldots & 1 & 1 \end{pmatrix}, \tag{115}$$

$$(u^2)^{\mathrm{T}} = \frac{1}{\sqrt{n}} \begin{pmatrix} 1 & -1 & 1 & -1 & \ldots & 1 & -1 & 1 & -1 \end{pmatrix}, \tag{116}$$

$$(u^3)^{\mathrm{T}} = \frac{1}{\sqrt{n}} \begin{pmatrix} 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 & \ldots & 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \end{pmatrix}, \tag{117}$$

$$(u^4)^{\mathrm{T}} = \frac{1}{\sqrt{n}} \begin{pmatrix} 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & \ldots & 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \end{pmatrix}, \tag{118}$$

$$(v^1)^{\mathrm{T}} = \frac{1}{\sqrt{n-1}} \begin{pmatrix} 1 & 1 & \ldots & 1 & 1 & 1 & 0 \end{pmatrix}, \tag{119}$$

$$(v^2)^{\mathrm{T}} = \begin{pmatrix} 0 & 0 & \dots & 0 & 0 & 0 & 1 \end{pmatrix}, \tag{120}$$

$$(v^3)^{\mathrm{T}} = \frac{1}{\sqrt{n-2}} \begin{pmatrix} 1 & -1 & 1 & -1 & \dots & 1 & -1 & 1 & -1 & 0 & 0 \end{pmatrix}, \tag{121}$$

$$(v^4)^{\mathrm{T}} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 0 & -1 & 0 & 0 & 0 & \dots & 0 & 0 \end{pmatrix}, \tag{122}$$

that is,

1. (a) every entry of u^1 is $1/\sqrt{n}$,
   (b) entries 1, 2, ..., $n-2$, $n-1$ of v^1 are $1/\sqrt{n-1}$, and entry n of v^1 is 0,
2. (a) every even entry of u^2 is $-1/\sqrt{n}$, and every odd entry of u^2 is $1/\sqrt{n}$,
   (b) entries 1, 2, ..., $n-2$, $n-1$ of v^2 are 0, and entry n of v^2 is 1,
3. (a) the first pair of entries of u^3 is $1/\sqrt{n}$, the second pair of entries is $-1/\sqrt{n}$, the third pair of entries is $1/\sqrt{n}$, the fourth pair of entries is $-1/\sqrt{n}$, and so on (with each successive pair of entries alternating sign),
   (b) every even entry of v^3 except for entry n is $-1/\sqrt{n-2}$, every odd entry of v^3 except for entry $n-1$ is $1/\sqrt{n-2}$, and entries $n-1$ and n of v^3 are 0,
4. (a) the first quadruplet of entries of u^4 is $1/\sqrt{n}$, the second quadruplet of entries is $-1/\sqrt{n}$, the third quadruplet of entries is $1/\sqrt{n}$, the fourth quadruplet of entries is $-1/\sqrt{n}$, and so on (with each successive quadruplet of entries alternating sign), and
   (b) entry 1 of v^4 is $1/\sqrt{2}$, entry 2 of v^4 is 0, entry 3 is $-1/\sqrt{2}$, and entries 4, 5, ..., $n-1$, n are 0.

For this example (Example 4),

$$u^1 \; \sigma_1 \; (v^1)^{\mathrm{T}} + u^2 \; \sigma_2 \; (v^2)^{\mathrm{T}} = \frac{1}{\sqrt{n\,(n-1)}} \begin{pmatrix} 1 & 1 & \cdots & 1 & 1 & \sqrt{n-1} \\ 1 & 1 & \cdots & 1 & 1 & -\sqrt{n-1} \\ 1 & 1 & \cdots & 1 & 1 & \sqrt{n-1} \\ 1 & 1 & \cdots & 1 & 1 & -\sqrt{n-1} \\ \vdots & \vdots & \vdots\vdots\vdots & \vdots & \vdots & \vdots \\ 1 & 1 & \cdots & 1 & 1 & \sqrt{n-1} \\ 1 & 1 & \cdots & 1 & 1 & -\sqrt{n-1} \end{pmatrix}. \tag{123}$$

Clearly, u^1, u^2, u^3, and u^4 are orthonormal, as are v^1, v^2, v^3, and v^4. Therefore, the i^{th} singular value of A is σ_i, for all $i = 1, 2, 3, 4$. Table 2.4 displays the results of applying the algorithm of the present paper to A, for various values of n.

Remark 27 All numerical data that we have examined — including the data displayed in Table 2, as well as the data from further experiments — appear to satisfy the bounds (91), (93), (102), and (104). However, we are not entirely certain that bounds such as (91), (93), (102), and (104) must necessarily contain factors of the order of $\sqrt{m}$ and $\sqrt{4k\,(n-k)+1}$. We would not be surprised if tighter bounds were possible.

The loss of precision displayed in Table 2.1 as n increases is probably largely due to round-off (compare Table 2.2), whereas the loss of precision displayed in Table 2.4 as n increases suggests that any bounds such as (91), (93), (102), and (104) must contain a factor of the order of $\sqrt{4k\,(n-k)+1}$. In contrast, in many practical situations the bounds mentioned in Remarks 20 and 24 are effectively independent of m.

Acknowledgements

We would like to thank R. R. Coifman, R. Kannan, and G. C. Papanicolaou for useful discussions.

$l-k$	β	γ^2	ζ	$\Pi_{l-k,\beta,\gamma}$
0	10	3	10^2	.15e-0
0	88	4	10^3	.18e-1
0	790	5	10^4	.21e-2
0	6,600	7	10^5	.18e-3
0	58,000	9	10^6	.19e-4

(1.1)

$l-k$	β	γ^2	ζ	$\Pi_{l-k,\beta,\gamma}$
1	7.9	5	10^2	.85e-2
1	72	6	10^3	.11e-3
1	620	8	10^4	.15e-5
1	5,500	10	10^5	.17e-7
1	53,000	11	10^6	.19e-9

(1.2)

$l-k$	β	γ^2	ζ	$\Pi_{l-k,\beta,\gamma}$
2	7.9	5	10^2	.37e-3
2	66	7	10^3	.61e-6
2	589	9	10^4	.85e-9
2	5,300	11	10^5	.12e-11
2	51,000	12	10^6	.14e-14

(1.3)

$l-k$	β	γ^2	ζ	$\Pi_{l-k,\beta,\gamma}$
2	1.02	3	10	.18e-0
4	0.88	4	10	.16e-1
8	0.88	4	10	.10e-4
16	0.79	5	10	.16e-12
32	0.72	6	10	.59e-32

(1.4)

$l-k$	β	γ^2	ζ	$\Pi_{l-k,\beta,\gamma}$
0	10	3	100	.15e-0
1	7.9	5	100	.85e-2
2	7.9	5	100	.37e-3
4	7.2	6	100	.44e-6
8	6.6	7	100	.12e-12
16	6.6	7	100	.55e-28
32	6.2	8	100	.83e-63

(1.5)

$l-k$	β	γ^2	ζ	$\Pi_{l-k,\beta,\gamma}$
0	10	3	1,000	.18e-1
1	72	6	1,000	.11e-3
2	66	7	1,000	.61e-6
4	62	8	1,000	.93e-11
8	58	9	1,000	.37e-21
16	55	10	1,000	.73e-44
32	53	11	1,000	.14e-93

(1.6)

Table 1 (See Subsection 4.4.)

k	l	n	time (sec.)	σ_k	σ_{k+1}	$\delta_{\text{rel. max.}}$	δ_{ID}	δ_{SVD}
10	10	10^2	.29e-03	.232e-15	.200e-15	.162e-14	.533e-14	.534e-14
10	10	10^3	.27e-02	.232e-15	.200e-15	.211e-14	.286e-14	.934e-14
10	10	10^4	.42e-01	.232e-15	.200e-15	.844e-14	.187e-13	.187e-13
10	10	10^5	.16e+01	.232e-15	.200e-15	.181e-13	.530e-13	.189e-12
10	10	10^6	.12e+02	.232e-15	.200e-15	.856e-13	.121e-12	.683e-11

(2.1)

k	l	n	time (sec.)	σ_k	σ_{k+1}	$\delta_{\text{rel. max.}}$	δ_{ID}	δ_{SVD}
10	10	10^2	.28e-03	.152e-07	.100e-07	.139e-06	.156e-06	.156e-06
10	10	10^3	.28e-02	.152e-07	.100e-07	.953e-07	.806e-07	.806e-07
10	10	10^4	.40e-01	.152e-07	.100e-07	.104e-06	.242e-06	.242e-06
10	10	10^5	.16e+01	.152e-07	.100e-07	.165e-06	.182e-06	.182e-06
10	10	10^6	.12e+02	.152e-07	.100e-07	.102e-06	.134e-06	.134e-06

(2.2)

k	l	n	time (sec.)	σ_k	σ_{k+1}	$\delta_{\text{rel. max.}}$	δ_{ID}	δ_{SVD}
30	30	50,000	.41e+01	.206e-08	.100e-08	.179e-07	.482e-07	.482e-07
30	31	50,000	.42e+01	.206e-08	.100e-08	.213e-07	.348e-07	.348e-07
30	32	50,000	.41e+01	.206e-08	.100e-08	.114e-07	.304e-07	.304e-07
30	34	50,000	.43e+01	.206e-08	.100e-08	.139e-07	.201e-07	.201e-07
30	38	50,000	.45e+01	.206e-08	.100e-08	.104e-07	.144e-07	.144e-07
30	46	50,000	.48e+01	.206e-08	.100e-08	.805e-08	.863e-08	.868e-08

(2.3)

k	l	n	time (sec.)	σ_k	σ_{k+1}	$\delta_{\text{rel. max.}}$	δ_{ID}	δ_{SVD}
2	2	$.48 \cdot 10^2$	.19e-04	.100e+01	.100e-07	.128e-06	.757e-07	.757e-07
2	2	$.48 \cdot 10^3$	.13e-03	.100e+01	.100e-07	.374e-06	.260e-06	.260e-06
2	2	$.48 \cdot 10^4$	.22e-02	.100e+01	.100e-07	.728e-05	.663e-05	.663e-05
2	2	$.48 \cdot 10^5$	.16e-01	.100e+01	.100e-07	.853e-05	.714e-05	.714e-05
2	2	$.48 \cdot 10^6$	.18e-00	.100e+01	.100e-07	.152e-04	.114e-04	.114e-04
2	2	$.48 \cdot 10^7$	.28e+01	.100e+01	.100e-07	.306e-04	.216e-04	.216e-04

(2.4)

Table 2 (See Section 5.)

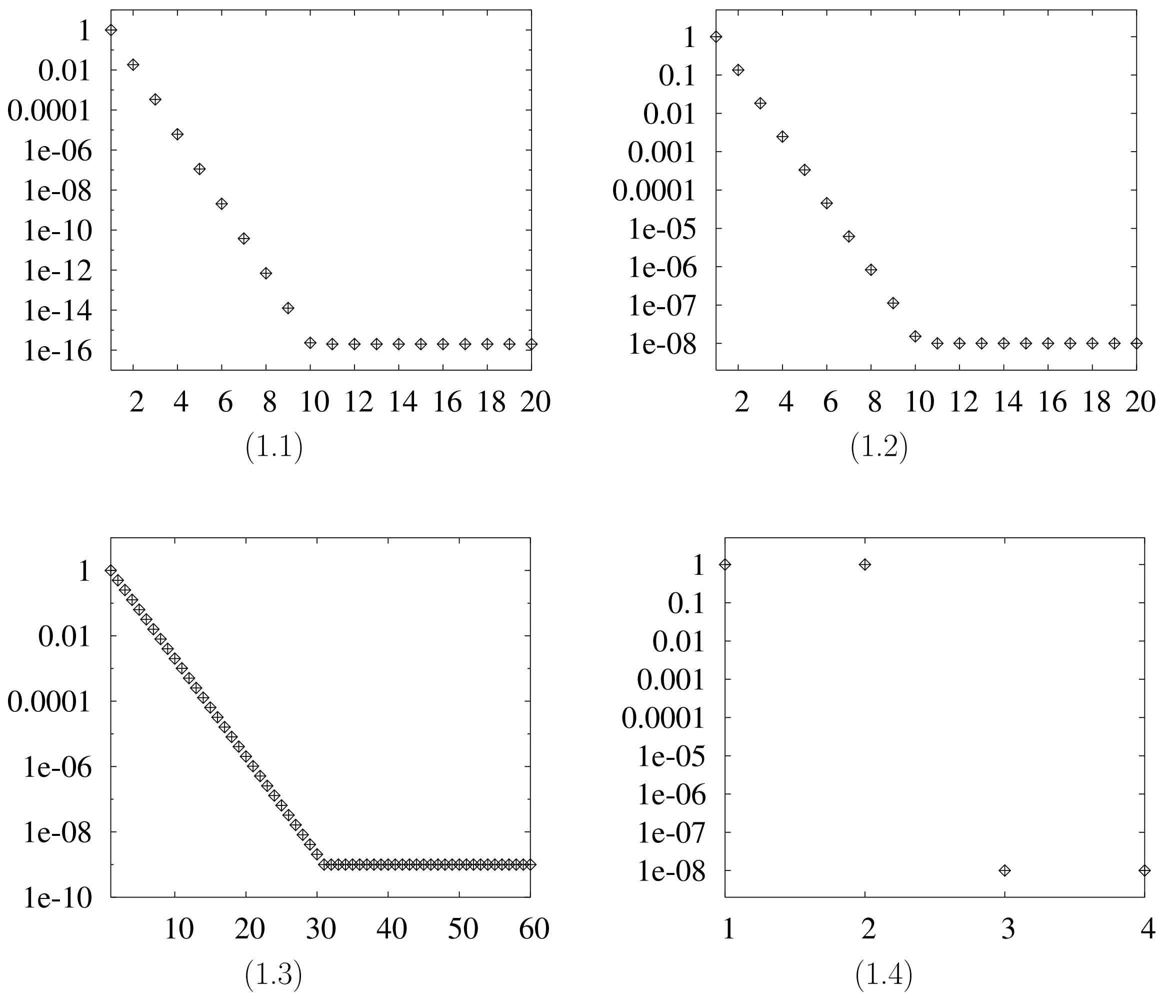

Figure 1 (See Section 5.)

References

[1] D. Achlioptas and F. McSherry, *Fast computation of low-rank matrix approximation*, J. ACM, (2006). To appear.

[2] M. W. Berry, S. A. Pulatova, and G. W. Stewart, *Algorithm 844: Computing sparse reduced-rank approximations to sparse matrices*, ACM Trans. Math. Software, 31 (2005), pp. 252–269.

[3] Å. Björck, *Numerics of Gram-Schmidt orthogonalization*, Linear Algebra Appl., 197–198 (1994), pp. 297–316.

[4] T. F. Chan and P. C. Hansen, *Some applications of the rank-revealing QR factorization*, SIAM J. Sci. Stat. Comput., 13 (1992), pp. 727–741.

[5] Z. Chen and J. J. Dongarra, *Condition numbers of Gaussian random matrices*, SIAM J. Matrix Anal. Appl., 27 (2005), pp. 603–620.

[6] H. Cheng, Z. Gimbutas, P.-G. Martinsson, and V. Rokhlin, *On the compression of low rank matrices*, SIAM J. Sci. Comput., 26 (2005), pp. 1389–1404.

[7] A. Deshpande, L. Rademacher, S. Vempala, and G. Wang, *Matrix approximation and projective clustering via volume sampling*, in Proceedings of the 17th ACM-SIAM Symposium on Discrete Algorithms, 2006.

[8] P. Drineas, R. Kannan, and M. W. Mahoney, *Fast Monte Carlo algorithms for matrices I: Approximating matrix multiplication*, SIAM J. Comput., (2006). To appear.

[9] ——, *Fast Monte Carlo algorithms for matrices II: Computing low-rank approximations to a matrix*, SIAM J. Comput., (2006). To appear.

[10] ——, *Fast Monte Carlo algorithms for matrices III: Computing an efficient approximate decomposition of a matrix*, SIAM J. Comput., (2006). To appear.

[11] P. Drineas, M. W. Mahoney, and S. Muthukrishnan, *Polynomial time algorithm for column-row-based relative-error low-rank matrix approximation*, Tech. Rep. 2006-04, DIMACS, March 2006.

[12] A. Frieze and R. Kannan, *Quick approximation to matrices and applications*, Combinatoria, 19 (1999), pp. 175–220.

[13] A. Frieze, R. Kannan, and S. Vempala, *Fast Monte-Carlo algorithms for finding low-rank approximations*, J. ACM, 51 (2004), pp. 1025–1041.

[14] H. H. Goldstine and J. von Neumann, *Numerical inverting of matrices of high order, II*, Amer. Math. Soc. Proc., 2 (1951), pp. 188–202.

[15] G. H. Golub and C. F. Van Loan, *Matrix Computations*, Johns Hopkins University Press, Baltimore, Maryland, third ed., 1996.

[16] M. Gu and S. C. Eisenstat, *Efficient algorithms for computing a strong rank-revealing QR factorization*, SIAM J. Sci. Comput., 17 (1996), pp. 848–869.

[17] P.-G. Martinsson, V. Rokhlin, and M. Tygert, *On interpolation and integration in finite-dimensional spaces of bounded functions*, Comm. Appl. Math. Comput. Sci., (2006). To appear.

[18] W. Press, S. Teukolsky, W. Vetterling, and B. Flannery, *Numerical Recipes*, Cambridge University Press, Cambridge, UK, second ed., 1992.

[19] G. W. Stewart, *Four algorithms for the efficient computation of truncated pivoted QR approximations to a sparse matrix*, Numer. Math., 83 (1999), pp. 313–323.

[20] J. H. Wilkinson, *The Algebraic Eigenvalue Problem*, Oxford University Press, Oxford, 1965.