ANALYSIS OF RANDOMIZED SUBSPACE ITERATION: CANONICAL ANGLES AND UNITARILY INVARIANT NORMS

ARVIND K. SAIBABA

Abstract. This paper is concerned with the analysis of the randomized subspace iteration for the computation of low-rank approximations. We present three different kinds of bounds. First, we derive both bounds for the canonical angles between the exact and the approximate singular subspaces. Second, we derive bounds for the low-rank approximation in any unitarily invariant norm (including the Schatten-p norm). This generalizes the bounds for Spectral and Frobenius norms found in the literature. Third, we present bounds for the accuracy of the singular values. The bounds are structural in that they are applicable to any starting guess, be it random or deterministic, that satisfies some minimal assumptions. Specialized bounds are provided when a Gaussian random matrix is used as the starting guess. Numerical experiments demonstrate the effectiveness of the proposed bounds.

Key words. Singular Value Decomposition, Randomized algorithms, Canonical angles, Low-rank approximation.

1. Introduction. The computation of low-rank approximations of large-scale matrices is a vital step in many applications in data analysis and scientific computing, such as principal component analysis, facial recognition, spectral clustering, model reduction techniques such as proper orthogonal decomposition (POD) and discrete empirical interpolation method (DEIM), approximation algorithms for PDEs and integral equations. The celebrated Eckart-Young theorem \cite{5} says that the optimal low-rank approximation can be obtained by means of the Singular Value Decomposition (SVD); however, computing the full or truncated SVD can be computationally challenging, or even prohibitively expensive for many applications of interest.

Randomized algorithms for computing low-rank approximations have become increasingly popular in the last two decades, see the survey papers \cite{2, 7}. Randomized methods have gained in popularity since they are easy to implement, computationally efficient, and numerically robust. While randomized algorithms tend to have the same asymptotic cost comparable to classical methods, randomized algorithms are able to exploit parallel computing efficiently. Additionally, for large datasets that are too large to fit in memory, the randomized algorithms are efficient in the number of times they access the data, compared to classical algorithms. We focus on a particular method known as randomized subspace iteration. The main idea of this method is to identify a subspace of the matrix by randomly mixing its columns, i.e., by multiplying the matrix by a random matrix of appropriate dimensions. A low-rank approximation is obtained by projecting the matrix onto this subspace. A post-processing step here involves compressing the matrix by a deterministic method, and a conversion step to obtain an equivalent representation in the desired format (typically, a truncated SVD representation). In practice, to obtain a higher quality approximation, the mixing step requires repeated applications of the action of a matrix.

Randomized algorithms also have excellent numerical robustness and the resulting approximation come with guarantees of very high probability. Many advances have been made in the analysis of randomized algorithms for low-rank approximations. The analysis typically has two stages: the structural, or deterministic stage,
in which minimal assumption about the distribution of the random matrix is made, and probabilistic stage, in which the distribution of the random matrix is taken into account to derive bounds for expected and tail bounds for the error distribution. As mentioned earlier, existing literature only targets the error in the low-rank representation [11, 12]. When the low-rank representation is in the SVD format, it is desirable to understand the quality of the approximate subspaces and the individual singular triplets. This paper aims to fill in some missing gaps in the literature by a rigorous analysis of the accuracy of approximate singular values, vectors and subspaces obtained using randomized subspace iteration. This analysis will be beneficial in applications where an analysis beyond the low-rank approximation is desired. Examples include Model Reduction techniques [9, 2], Leverage Score computation [14], Spectral Clustering [6], FEAST eigensolvers [22], Canonical Correlation Analysis [1].

1.1. Contributions and overview of paper. We survey the contents and the main contributions of this paper.

**Canonical angles** We have developed bounds for all the canonical angles between the spaces spanned by the exact and the approximate singular vectors. Several different flavors of bounds are provided:

1. The bounds in § 3.1 relate the canonical angles between the exact and the approximate singular subspaces. Analysis is also provided for unitarily invariant norms of the canonical angles.
2. In applications where lower dimensional subspaces are extracted from the approximate singular subspaces, the bounds in § 3.2 quantifies the accuracy in the extraction process.
3. Also in § 3.2 are bounds for the angles between the individual exact and approximate singular vectors, extracted from the appropriate subspaces.

Our bounds suggest that the accuracy of the singular values and vectors, in addition to the low-rank approximations, is high provided (1) singular values decay rapidly beyond the target rank $k$, and (2) the larger the singular value gaps, the higher is the accuracy to be expected. Furthermore, the truncation step to extract the $k$ dimensional subspaces does not significantly increase the canonical angles.

**Low-rank approximation** This paper provides the first known analysis of the randomized subspace iteration for an arbitrary unitarily invariant norm, with stronger, specialized results for Schatten-p norms. Bounds for the special cases of the Schatten-p norm, namely the spectral and Frobenius norms, have already appeared in the literature—our result for the Schatten-p norm recovers these results as special cases.

**Singular values** We derive upper and lower bounds on the approximate singular values obtained by the randomized subspace iteration. Similar bounds also appear in [11]; however, our proof technique is different. We also present Hoffman-Weilandt type bounds for the accuracy of the singular values.

The conclusion of the bounds for the low-rank approximations and the singular values are similar to those of the conclusions for the canonical angles. We also make the following contribution, maybe of independent interest beyond the study of randomized algorithms.

**Generalization of sin theta theorem** The sin theta theorem [24] is a well known result in numerical analysis and relates the canonical angles between the true and approximate singular subspaces in the unitarily invariant norms. We derive a generalization of the sin theta theorem that derives bounds for the individual canonical angles between the two subspaces. The sin theta theorem is recovered as a special case.
2. Background and preliminaries.

2.1. Notation. Denote the target rank by \( k \). Let the matrix \( A \in \mathbb{C}^{m \times n} \), with \( m \geq n \), have the singular value decomposition

\[
A = [U_1 \quad U_2] \begin{bmatrix}
\Sigma_1 & \\
\Sigma_2 & \\
\end{bmatrix} \begin{bmatrix}
V_1^* \\
V_2^* \\
\end{bmatrix}.
\]

Here \( \Sigma_1 \in \mathbb{C}^{k \times k} \) and \( \Sigma_2 \in \mathbb{C}^{(n-k) \times (n-k)} \); the columns of \( U_1 \) and \( U_2 \) are the corresponding left singular vectors, and columns of \( V_1 \) and \( V_2 \) are the corresponding right singular vectors. We also denote by \( \llbracket A \rrbracket_k = U_1 \Sigma_1 V_1^* \) as the best rank-\( k \) approximation to the matrix \( A \), in any unitarily invariant norm (for a definition, see below).

**Singular values and ratios.** Let \( \| \cdot \|_2 \) denote the spectral norm, so that \( \| \Sigma_2 \|_2 = \sigma_{k+1} \) and \( \| \Sigma_2^{-1} \|_2 = \frac{1}{\sigma_k} \). The singular values of \( A \) can be arranged in decreasing order as

\[
\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_k \geq \sigma_{k+1} \geq \cdots \geq \sigma_n.
\]

For later use, we define the singular value ratios

\[
(1) \quad \gamma_j = \frac{\sigma_{k+1}}{\sigma_j}, \quad j = 1, \ldots, k.
\]

Since the singular values are monotonically decreasing, the singular value ratios are monotonically increasing, i.e., \( \gamma_1 \leq \cdots \leq \gamma_k \).

**Norms.** We have already defined the spectral norm. The Frobenius norm of a matrix is \( \| A \|_F = \sqrt{\text{trace}(A^*A)} \). We use the symbol \( \| \cdot \| \) to denote any unitarily invariant norm. An example of the unitarily invariant norms is Schatten-\( p \) class of norms, defined as the vector \( \ell_p \) norm of the singular values of \( A \), i.e.,

\[
\| A \|_p = \left( \sum_{j=1}^{\min\{m,n\}} \sigma_j^p \right)^{1/p}.
\]

With this definition it can be readily seen that \( \| A \|_2 = \| A \|_\infty \) and \( \| A \|_F = \| A \|_2 \). Another example is the Ky-Fan-\( k \) class of norms defined \( \| A \|_k = \sum_{j=1}^{k} \sigma_j \) for every \( k = 1, \ldots, \min\{m, n\} \). Associated with every unitarily invariant norm is a symmetric gauge function acting on the singular values of the matrix that it acts on.

**Projection matrices.** Suppose the matrix \( Z \) has full column rank \( Z^\dagger \) is a left multiplicative inverse and where \( ^\dagger \) represents the Moore-Penrose inverse. We define the projection matrix \( P_Z = ZZ^\dagger \). For a matrix \( Q \) with orthonormal columns, the formula simplifies and \( P_Q = QQ^* \).

**Canonical angles.** The separation between subspaces can be measured by the principal or canonical angles. Let \( \mathcal{M} \) and \( \mathcal{N} \) be two subspaces of \( \mathbb{C}^n \), such that \( \dim \mathcal{M} = k \), \( \dim \mathcal{N} = \ell \) and \( \ell \geq k \). Then the principal angles between the subspaces \( \mathcal{M} \) and \( \mathcal{N} \) are recursively defined to be the numbers \( 0 \leq \theta_i \leq \pi/2 \) such that

\[
\cos \theta_i = \max_{u \in \mathcal{M}, v \in \mathcal{N} \|u\|_2 = \|v\|_2 = 1} v^* u = v_i^* u_i, \quad i = 1, \ldots, k
\]

subject to the constraints \( \|u_i\|_2 = \|v_i\|_2 = 1 \), and

\[
u_i^* u, \quad v_i^* v = 0, \quad j = 1, \ldots, i - 1.
\]
The canonical angles are arranged in increasing order as
\[ 0 \leq \theta_1 \leq \cdots \leq \theta_k \leq \pi/2. \]

It can also be shown that \( \sin \theta_i \) are also the singular values of \( P_M - P_N \).

We denote \( \angle(M, N) \) to be the canonical angles between subspaces \( M \) and \( N \). Let \( M \) and \( N \) be matrices with orthonormal columns, which form bases for subspaces \( M \) and \( N \) respectively. Then, \( \sin \angle(M, N) \) are obtained as singular values of \( (I - MM^*)N \) and \( \cos \angle(M, N) \) are the singular value of \( M^*N \) [5]. For ease of notation, in the rest of this paper, we write \( \angle(M,N) \) instead of \( \angle(M, N) \).

### 2.2. Randomized subspace iteration.

The basic version of the randomized subspace iteration is summarized in Algorithm 1. Given a starting guess, denoted by \( \Omega \in \mathbb{C}^{n \times (k+\rho)} \), the algorithm performs \( q \) steps of the randomized subspace iteration to obtain the matrix \( Y \), also known as the “sketch.” The algorithm to compute

\begin{algorithm}
\caption{Idealized version of Subspace iteration for Singular Value Decomposition}
\begin{algorithmic}
\Require Matrix \( A \), Starting guess \( \Omega \in \mathbb{C}^{n \times (k+\rho)} \), an integer \( q \geq 0 \).
\State Compute \( Y = (AA^*)^{q/2}A \Omega \).
\State Compute thin QR factorization of \( Y \), so that \( Y = QR \).
\State Compute \( T = Q^*A \) and its SVD \( T = U\hat{\Sigma}\hat{V}^* \).
\State Compute \( \hat{U} = QU \).
\State return Matrices \( \hat{U}, \hat{\Sigma}, \hat{V} \) that define \( \hat{A} \equiv \hat{U}\hat{\Sigma}\hat{V}^* \).
\end{algorithmic}
\end{algorithm}

an approximate singular value decomposition, given starting guess \( \Omega \in \mathbb{C}^{n \times (k+\rho)} \) is summarized in Algorithm 1. We say this is an idealized version, since the algorithm can behave poorly in the presence of round-off errors. A practical implementation of this algorithm alternates the QR factorization with matvecs involving \( A \); the reader is referred to [20, 12].

In Algorithm 1 the output \( \hat{A} \equiv QQ^*A = \hat{U}\hat{\Sigma}\hat{V}^* \) has rank \( k + \rho \). If a rank-\( k \) approximation to \( A \) is desired, then it can be obtained by discarding the \( p \) smallest singular values of \( \hat{A} \). We denote this low-rank representation by
\[ [\hat{A}]_k = \hat{U}_1\hat{\Sigma}_1\hat{V}_1^*. \]

This is summarized in Algorithm 2.

\begin{algorithm}
\caption{Truncated SVD of \( \hat{A} = QQ^*A \)}
\begin{algorithmic}
\Require Matrix \( A \in \mathbb{C}^{m \times n} \) and \( Q \in \mathbb{C}^{m \times (k+\rho)} \). Target rank \( k \).
\State Form matrix \( B = Q^*A \).
\State Compute the truncated SVD representation \( [B]_k = \hat{U}_B\hat{\Sigma}_B\hat{V}_B^* \).
\State Form \( \hat{U}_1 = Q\hat{U}_B \)
\State return Matrices \( \hat{U}_1, \hat{\Sigma}_1, \hat{V}_1 \) such that \( [\hat{A}]_k = \hat{U}_1\hat{\Sigma}_1\hat{V}_1^* \).
\end{algorithmic}
\end{algorithm}

Partition the matrix \( V^*\Omega \), which captures the influence of the starting guess on the right singular matrix, as
\[ V^*\Omega = \begin{bmatrix} V_1^*\Omega \\ V_2^*\Omega \end{bmatrix} = \begin{bmatrix} \Omega_1 \\ \Omega_2 \end{bmatrix}. \]
where \( \Omega_1 \in \mathbb{C}^{k \times (k+\rho)} \) and \( \Omega_2 \in \mathbb{C}^{(n-k) \times (k+\rho)} \). The following assumptions will be required for our analysis.

**Assumption 1.** Let \( \Omega_1 \) be defined as above. We assume that

\[
\text{rank} (\Omega_1) = k, \tag{2}
\]

and the singular value gap is inversely proportional to

\[
\gamma_k = \|\Sigma_2\|_2 \left\| \Sigma_1^{-1} \right\|_2 = \frac{\sigma_{k+1}}{\sigma_k} < 1. \tag{3}
\]

The first assumption guarantees that the starting guess \( \Omega \) has a significant influence over the right singular vectors, whereas the second assumption ensures that the \( k \) dimensional subspace \( \mathcal{R}(U_1) \) is well defined.

**3. Accuracy of singular vectors.** We want to understand how well \( \mathcal{R}(\hat{U}) \) approximates the subspace spanned by the first \( k \) singular vectors \( U_1 \), measured in terms of the canonical angles between the subspaces.

Abbreviate the subspace angles between \( \hat{U} \in \mathbb{C}^{m \times \ell} \) and \( U_1 \in \mathbb{C}^{m \times k} \) as \( \theta_1, \ldots, \theta_k \). Similarly, denote the angles between \( \hat{V} \in \mathbb{C}^{n \times \ell} \) and \( V_1 \in \mathbb{C}^{n \times k} \) by \( \nu_1, \ldots, \nu_k \). We are also interested in obtaining bounds for the canonical angles \( \angle(U_1, \hat{U}) \) and \( \angle(V_1, \hat{V}) \). To distinguish these angles from \( \angle(U_1, \hat{U}) \) and \( \angle(V_1, \hat{V}) \), we call them \( \theta'_j \) and \( \nu'_j \) for \( j = 1, \ldots, k \).

**3.1. Bounds for canonical angles.** Our first result derives bounds for the canonical angles \( \angle(U_1, \hat{U}) \). The analysis is based on the perturbation of projectors and the tools used here are similar to [12].

**Theorem 1.** Let \( \hat{U} \) and \( \hat{V} \) be obtained from Algorithm 1. With Assumption 1, the canonical angles \( \theta_j \) and \( \nu_j \) satisfy for \( j = 1, \ldots, k \):

\[
\sin \theta_j \leq \frac{\gamma_j^{2q+1} \left\| \Omega_2 \Omega_1^\dagger \right\|_2}{\sqrt{1 + \gamma_j^{4q+2} \left\| \Omega_2 \Omega_1^\dagger \right\|_2^2}}, \quad \sin \nu_j \leq \frac{\gamma_j^{2q+2} \left\| \Omega_2 \Omega_1^\dagger \right\|_2}{\sqrt{1 + \gamma_j^{4q+4} \left\| \Omega_2 \Omega_1^\dagger \right\|_2^2}}.
\]

This theorem has several interesting features worth pointing out. First, if the matrix has exact rank \( k \), then all of the canonical angles are uniformly zero; that is, the randomized subspace iteration identifies the subspace exactly. Suppose the rank \( \rho \geq k \). The bounds for the canonical angles show explicit dependence on the singular value ratios \( \gamma_j \). In particular, the smaller the singular value ratio, smaller the canonical angles. Furthermore, the canonical angles converge to zero at different rates.

Second, the term \( \left\| \Omega_2 \Omega_1^\dagger \right\|_2 \) can be written in terms of the singular vector matrix \( V \) and the starting guess \( \Omega \) as

\[
\left\| \Omega_2 \Omega_1^\dagger \right\|_2 = \left\| (V_2^* \Omega)(V_1^* \Omega)^\dagger \right\|_2.
\]

When the columns of \( \Omega \) is linearly independent, this quantity is nothing but the tangent of the largest canonical angle between \( \mathcal{R}(V_1) \) and \( \mathcal{R}(\Omega) \). This term appears frequently in randomized linear algebra and can be interpreted as a measure of the subspace overlap between the starting guess and the right singular vectors. In the
ideal case, $\Omega$ contains the singular vectors in $V_1$. A discussion of the meaning and interpretation of this term, is provided in [8, Section 2.5]. In particular, when $\Omega$ is a Gaussian random matrix $\|\Omega_2\Omega_1^\dagger\|_2$ is roughly on the order of $\sqrt{(n-k)k}$.

Third, the influence of $\|\Omega_2\Omega_1^\dagger\|_2$ is subdued by the singular value ratios $\gamma_j^{2q+1}$. The bounds are informative in the sense that the sine of the canonical angles are at most 1, which is a desirable feature. With sufficiently large number of iterations, the denominator is $O(1)$, the numerator is a sufficiently good approximation.

Lastly, it is worth mentioning here that the bounds for the canonical angles are smaller than $\nu_j$ because of the presence of an additional power of $\gamma_j$. The reason for this higher accuracy is because the columns of $\hat{V}$ are the right singular vectors of $Q^*A$. The multiplication step with $Q$ amounts to an additional step of subspace iteration and gives the extra factor. Theorem 1 gives the sine of the canonical angles; these bounds can also be used to obtain upper bounds for the tangents and lower bounds for the cosines.

**Remark 1.** For $j = 1, \ldots, k$, the relationship between the tangent and sine implies
\[
\tan \theta_j \leq \gamma_j^{2q+1} \|\Omega_2\Omega_1^\dagger\|_2, \quad \tan \nu_j \leq \gamma_j^{2q+2} \|\Omega_2\Omega_1^\dagger\|_2.
\]
Lower bounds for cosine of the canonical angles follow similarly.

**Unitarily invariant norms.** The following result derives bounds for the canonical angles in any unitarily invariant norm, in contrast to Theorem 1 which bounds the individual canonical angles.

**Theorem 2.** Let $\Omega \in \mathbb{R}^{n \times (k+p)}$. Let the approximate singular vectors $\hat{U}$ and $\hat{V}$ for a matrix $A$ be computed according to Algorithm 1. Under Assumption 1, for every unitarily invariant norm,
\[
\left\| \sin \angle(U_1, \hat{U}) \right\| \leq \gamma_k^{2q} \frac{\|\Sigma\|}{\sigma_k} \|\Omega_2\Omega_1^\dagger\|_2, \quad \left\| \sin \angle(V_1, \hat{V}) \right\| \leq \gamma_k^{2q+1} \frac{\|\Sigma\|}{\sigma_k} \|\Omega_2\Omega_1^\dagger\|_2.
\]

The interpretation of this theorem is similar to that of Theorem 1. The connection between the two theorems follows from the identity $\sin \theta_k = \|\sin \angle(U_1, \hat{U})\|_2$. It is immediately seen that the result in the spectral norm is weaker.

### 3.2. Extraction of $k$-dimensional subspaces.** In the previous subsection, the columns of $\hat{U}$ and $\hat{V}$ spanned $\ell = k + \rho$ dimensional subspaces. Many applications, however, require the extraction of $k$ dimensional singular subspaces from the low-rank approximation $\hat{A} \equiv QQ^*A$. One way to extract the appropriate subspaces is to first compute the optimal rank-$k$ truncation of $\hat{A}$, denoted by $[[\hat{A}]]_k$. The singular vectors of $[[\hat{A}]]_k$ denoted by $\hat{U}_1$ and $\hat{V}_1$ are then used instead of $\hat{U}$ and $\hat{V}$. See Algorithm 2.

The bounds derived in the previous subsection are not directly applicable here since [25, Corollary 10] says
\[
\theta_j \leq \theta_j' \quad \nu_j \leq \nu_j' \quad j = 1, \ldots, k.
\]

We would like to understand how much additional error is incurred during this extraction process, for which we present several results. The important conclusion is that the accuracy of the extracted subspaces of dimension $k$ is comparable to the accuracy of the $k + \rho$ dimensional subspace up to a modest constant.
The approach we take is different from that of the previous section. The starting point of our analysis is the well-known sin-theta theorem for singular subspaces [24]. Let $A, \hat{A}$ be two matrices of conformal dimensions. Assuming

$$\text{gap} \equiv \sigma_k(A) - \sigma_{k+1}(\hat{A}) > 0,$$

we have

$$\max \left\{ \| \sin \angle(U_1, \hat{U}_1) \|, \| \sin \angle(V_1, \hat{V}_1) \| \right\} \leq \frac{\max\{\|E_{12}\|, \|E_{21}\|\}}{\text{gap}},$$

where the two matrices $E_{12}$ and $E_{21}$ are

$$E_{12} = (I - P_{\hat{U}_1})(A - \hat{A})P_{V_1}$$
$$E_{21} = P_{U_1}(A - \hat{A})(I - P_{\hat{V}_1}).$$

However, this version of the sin theta theorem does not provide us with a way to obtain bounds for the individual canonical angles. To this end, we first present a new generalization of the sin theta theorem.

**Theorem 3.** Let $A \in \mathbb{C}^{m \times n}$ with rank $(A) \geq k$ and let $\hat{A}$ be the perturbed matrix with same dimensions. Suppose the singular value gap satisfies (4) Let $\hat{U}_1$ and $\hat{V}_1$ be the left and right singular vectors of $\hat{A}$ corresponding to the top $k$ singular values. Then

$$\max\{\sin \theta_j', \sin \nu_j'\} \leq \frac{\sigma_k(A)}{\sigma_j(A)} \max\{\sin \theta_j', \sin \nu_j'\} \quad j = 1, \ldots, k.$$

This theorem states that the sine of the canonical angles $\sin \theta_j'$ are bounded by $\sin \theta_j'$ up to a multiplicative factor, which is at most 1.

Our main result provides the following bounds canonical angle when both the subspaces have the same dimension. The proof involves simplifying every term in equations (6).

**Theorem 4.** Let $\hat{U}$ and $\hat{V}$ be obtained from Algorithm 1, and matrices $\hat{U}_1$ and $\hat{V}_1$ from Algorithm 2. Under assumption 1,

- for every unitarily invariant norm

$$\max \left\{ \| \sin \angle(U_1, \hat{U}_1) \|, \| \sin \angle(V_1, \hat{V}_1) \| \right\} \leq \phi \frac{\gamma_k}{\sqrt{1 - \gamma_k}} \frac{\| \Sigma_2 \|}{\sigma_k} \left\| \Omega_2 \Omega_1^\dagger \right\|_2.$$

The factor $\phi = \sqrt{2}$ for an arbitrary unitarily invariant norm, and $\phi = 1$ for the spectral and Frobenius norms.

- canonical angles $\theta_j'$ and $\nu_j'$ satisfy

$$\max\{\sin \theta_j', \sin \nu_j'\} \leq \gamma_j \frac{\gamma_k}{1 - \gamma_k} \left\| \Omega_2 \Omega_1^\dagger \right\|_2 \quad j = 1, \ldots, k.$$

The interpretation of this theorem is: as the number of iterations $q$ increase, the largest canonical angle converges to 0 quadratically and a larger singular value gap means that the subspace is computed more accurately. Compared to Theorem 2, for any unitarily invariant norm, we have gained the factor $\max\{1, \sqrt{2}\gamma_k\}/(1 - \gamma_k)$ in the denominator. For the spectral and Frobenius norms, this bound is only $1/(1 - \gamma_k)$. 

7
Both factors are greater than 1, and independent of the number of iterations $q$. This additional factor is the (one-time) price to be paid by restricting the subspace $\mathcal{R}(\hat{U})$ which is $\ell = k + \rho$ dimensional, to $\mathcal{R}(\hat{U}_1)$ which is $k$ dimensional. The bound is devastating when $\gamma_k \approx 1$; however, in this case, we are in deep trouble to begin with, since the subspaces may not be well defined.

Individual singular vectors. The previous results give insight into the accuracy measured using the canonical angles between the exact and approximate singular subspaces. When individual singular vectors need to be extracted, does the extraction process introduce additional error? The following result quantifies the accuracy of the extraction process.

**Theorem 5.** Let the approximate singular vectors $\hat{U}$ and $\hat{V}$ be computed according to Algorithm 1. With assumption 1, we have the following inequalities for $j = 1, \ldots, k$

$$\sin \angle(u_j, \hat{U}) \leq \gamma_j^{2q+1} \|\Omega_2 \Omega_1^\dagger\|_2 \quad \sin \angle(v_j, \hat{V}) \leq \gamma_j^{2q+2} \|\Omega_2 \Omega_1^\dagger\|_2.$$  

Denote the approximate singular triplets $(\hat{\sigma}_j, \hat{u}_j, \hat{v}_j)$ for $j = 1, \ldots, k$. Under assumption 1

$$\max \{\sin \angle(u_j, \hat{u}_j), \sin \angle(v_j, \hat{v}_j)\} \leq \sqrt{1 + 2 \frac{\tilde{\delta}^2}{\delta^2} \gamma_j^{2q+1} \|\Omega_2 \Omega_1^\dagger\|_2}.$$  

Here $\tilde{\delta}^2 = \|\Sigma_2\|^2 + \|\Sigma_2 \Omega_2^\dagger\|_2^2$ and $\delta = \min\{\min_{\hat{\sigma}_i \neq \sigma_j} |\sigma_j - \hat{\sigma}_i|, \sigma_j\}$.

The first result bounds the angles between the exact singular vector and the corresponding approximate singular subspaces. The second result compares the angles of the exact and the approximate singular vectors. This result also says that the extraction process does not adversely increase the error in the singular subspaces, provided the singular values are well-separated.

The convergence of the individual singular vectors tell a similar story to that of Theorem 1. The singular vectors corresponding to the largest singular values converge earlier than the singular vectors corresponding to the smaller singular vectors. This is a consequence of the fact that the singular value ratios are non-decreasing.

### 3.3. Comparison with other bounds.

The subspace iteration dates to a 1957 paper by Bauer [3] for eigenvalue problems. The analysis of the subspace iteration has also been well-established, for we refer to [19, Chapter 14]. The effect of randomization has been discussed in [6] which has similar results as ours. One major exception in previous works, however, is that they assumed that the oversampling parameter $\rho = 0$. Since $\Omega_1 \in \mathbb{C}^{k \times k}$, the assumption $\text{rank}(\Omega_1) = k$, requires that $\Omega_1$ is invertible, which is a rather strong assumption. They were able to show (in our notation)

$$\|\sin \angle(U_1, \hat{U}_1)\|_2 \leq \frac{\gamma_k^{2q+1} \|\Omega_2 \Omega_1^{-1}\|_2}{\sqrt{1 + \gamma_k^{2q+2} \|\Omega_2 \Omega_1^{-1}\|_2^2}}.$$  

Returning to this assumption that $\Omega_1$ has full row rank, when $\Omega$ is standard Gaussian matrix,

$$\|\Omega_1^{-1}\|_2 \leq k^{1/2}/(2.35\delta)$$
with probability at least $1 - \delta$. By contrast, when $Ω_1 \in \mathbb{C}^{k \times (k + ρ)}$ and $p > 0$, then when $Ω$ is a Gaussian random matrix, with probability at least $1 - \delta$

$$\|Ω_1\|_2 \leq e\sqrt{\frac{k + ρ}{ρ}} \left(\frac{1}{δ}\right)^{1/(ρ+1)}.$$ 

It is clear that oversampling has an impact on the accuracy of the randomized subspace iteration. When $Ω$ is a random matrix that arises from the SRHT, or Rademacher distributions, a more aggressive form of oversampling $ℓ \sim k \log k$ is necessary to ensure that $\text{rank}(Ω_1) = k$. Therefore, by allowing for oversampling, our bounds are applicable to starting guesses that are not just Gaussian random matrices. Not only that, our bounds are also informative for matrices with decaying singular values and significant singular value gap.

A recent paper by Nakatsukasa [18] considered the issue of accuracy of extracting singular subspaces for general projection-based approximation methods. In our notation, these refer to relating bounds for $\angle(U_1, \hat{U})$ to $\angle(U_1, \hat{U}_1)$. Our bounds for the canonical angles appear to be tighter than the result [18, Corollary 1]; here we do not go into a detailed comparison. This may be because the analysis was applicable to arbitrary subspace projections, whereas ours is specialized to randomized subspace iteration. Furthermore, our analysis is able to bound the individual canonical angles which is missing in [18].

### 3.4. Probabilistic bounds

Thus far, we have not made specific assumptions on the matrix $Ω$, as long as it satisfies $\text{rank}(Ω_1) = k$. In particular, $Ω$ need not be even be random, and may be deterministic. However, more can be said about the bounds when $Ω$ is random is drawn from a specific distribution.

In many applications, the matrix $Ω \in \mathbb{R}^{n \times (k + ρ)}$ is taken to be the standard Gaussian random matrix. That is, the entries of $Ω$ are i.i.d. $\mathcal{N}(0, 1)$ random variables.

Here we give a taste of the probabilistic results that can be derived. Let $ρ \geq 2$ and define the constant

$$C_e = \sqrt{\frac{k}{ρ - 1}} + \frac{e\sqrt{(k + ρ)(n - r)}}{ρ}$$

and for $0 < δ < 1$ define the constant

$$C_d = \frac{e\sqrt{k + ρ}}{ρ + 1} \left(\frac{2}{δ}\right)^{1/(ρ+1)} \left(\sqrt{n - k} + \sqrt{k + ρ} + \sqrt{2 \log \frac{2}{δ}}\right).$$

**Theorem 6 (Probabilistic bounds).** Let $Ω \in \mathbb{R}^{n \times (k + ρ)}$ be a standard Gaussian random matrix with $ρ \geq 2$. Assume that the singular value ratio $γ_k < 1$. For $j = 1, \ldots, k$, the expected value of the canonical angles satisfy

$$E[\sin \theta_j] \leq \frac{γ_j^{2q+1}C_e}{\sqrt{1 + γ_j^{4q+2}C_e^2}} \quad E[\sin \nu_j] \leq \frac{γ_j^{2q+2}C_e}{\sqrt{1 + γ_j^{4q+4}C_e^2}}.$$ 

Let $0 < δ < 1$ be a user defined failure tolerance. With probability, at least $1 - δ$, the following inequalities hold independently for $j = 1, \ldots, k$

$$\sin \theta_j \leq \frac{γ_j^{2q+1}C_d}{\sqrt{1 + γ_j^{4q+2}C_d^2}} \quad \sin \nu_j \leq \frac{γ_j^{2q+2}C_d}{\sqrt{1 + γ_j^{4q+4}C_d^2}}.$$
Probabilistic results of this form can be derived for other distributions, following the strategy in [12]. We will not pursue them here.

4. Low-rank approximation and Singular values.

4.1. Low-rank approximation. Several results are available for estimating the error in the low-rank approximation $A \approx QQ^* A$ in the spectral and Frobenius norms, when the matrix $Q$ is obtained from the randomized subspace iteration [12, 11, 26]. As was mentioned earlier, the spectral and Frobenius norms are special cases of the Schatten-p norm, which are examples of unitarily invariant norms.

Here we present the first known analysis of randomized subspace iteration in a unitarily invariant norm.

Theorem 7. Let $\hat{A} \in \mathbb{C}^{m \times n}$ be computed using Algorithm 1. Under assumption 1, the following inequalities hold in every unitarily invariant norm

\begin{align}
\| (I - QQ^*) A \| &\leq \| \Sigma_2 \| + \gamma_k^2 \left\| \Sigma_2 \Omega_2 \Omega_1^\dagger \right\| \\
\| (I - QQ^*) [A]_k \| &\leq \gamma_k^{2 q} \left\| \Sigma_2 \Omega_2 \Omega_1^\dagger \right\|.
\end{align}

(11) (12)

If a rank-$k$ approximation is desired, then the error in the low-rank approximation is

\begin{equation}
\| A - Q [Q^* A]_k \| \leq \left( 1 + \frac{\sigma_1 \phi \gamma_k^{2 q}}{\sigma_k (1 - \gamma_k)} \left\| \Omega_2 \Omega_1^\dagger \right\|_2 \right) \| \Sigma_2 \|.
\end{equation}

(13)

As in Theorem 4, $\phi = 1$ for spectral and Frobenius norms, and $\sqrt{2}$ for an arbitrary unitarily invariant norm.

In this theorem, as the number of iterations $q \to \infty$, the error in the low-rank approximation goes to zero.

We present a variant of the error in the low-rank approximation for the special case that a Schatten-p norm is used. The proof for the special case of the Frobenius norm was provided in [26].

Theorem 8. Let $\hat{A}$ be computed using Algorithm 1. Under assumption 1, we have

\begin{equation}
\| (I - QQ^*) A \|_p^2 \leq \| \Sigma_2 \|_p^2 + \gamma_k^{4 q} \left\| \Sigma_2 \Omega_2 \Omega_1^\dagger \right\|_p^2.
\end{equation}

(14)

The error bound in Theorem 7 is weaker than Theorem 8 for the Schatten-p norm since for $\alpha, \beta \geq 0$, we have $\sqrt{\alpha^2 + \beta^2} \leq \alpha + \beta$. More generally, Theorem 8 is applicable to any unitarily invariant norm that is also a Q-norm [4, Definition IV.2.9]. A unitarily invariant norm $\| \cdot \|_Q$ is a Q-norm, if there exists another unitarily invariant norm $\| \cdot \|_a$ such that $\| A \|_Q^2 = \| A^* A \|_a$. Note that the Schatten-p norms satisfy this property for $p \geq 2$, since $\| A \|_p^2 = \| A^* A \|_{p/2}$.

4.2. Accuracy of singular values. How are the singular values of $A$ related to the singular values of $\hat{A}$? We now present a result that quantifies the accuracy of the individual singular values. This result is similar to [11, Theorem 4.3]. Our proof techniques are substantially different. We make extensive use of the Cauchy interlacing theorem and the multiplicative singular value inequalities.
Theorem 9. Let $\hat{A} = \hat{U}\hat{\Sigma}\hat{V}^*$ be computed using Algorithm 1. Under Assumption 1, the approximate singular values $\sigma_j(\hat{A})$ satisfy for $j = 1, \ldots, k$

$$\sigma_j(A) \geq \sigma_j(\hat{A}) \geq \frac{\sigma_j(A)}{\sqrt{1 + \gamma_4 2 q k \|\Omega_2 \Omega_1^\dagger\|^2}}.$$ 

It can be readily seen that the large singular values are computed more accurately since the singular value ratio corresponding to larger singular values is smaller. This result also shows that in the worst case scenario the singular values are a factor $1/\sqrt{2}$ smaller.

Rather than quantify the accuracy of the individual singular values, the next results are of the Hoffman-Weilandt type and account for all the singular values together.

Define the two matrices of conformal sizes

$$\Sigma = \begin{bmatrix} \Sigma_1 \\ \Sigma_2 \end{bmatrix}, \quad \Sigma' = \begin{bmatrix} \hat{\Sigma} & 0 \end{bmatrix}.$$ 

Under Assumption 1, the error in the singular values satisfies

$$(15) \quad \|\Sigma - \Sigma'\| \leq \|\Sigma_2\| + \gamma 2 q \|\Sigma_2 \Omega_2 \Omega_1^\dagger\|.$$ 

The proof combines [4, III.6.13] with Theorem 7. For the Schatten-p norm, with $p \geq 2$, we can derive the bound

$$(16) \quad \|\Sigma - \Sigma'\|_p \leq \sqrt{\|\Sigma_2\|_p^2 + \gamma 2 q \|\Sigma_2 \Omega_2 \Omega_1^\dagger\|_p^2}.$$ 

The proof is similar, and is therefore omitted.

5. Proofs. We recall some results here that will be useful in our analysis, see [15, Section 7.7] for proofs. Let $M, N$ be hermitian positive definite. The notation $M \preceq N$ means $N - M$ is positive semi-definite and it defines a partial ordering on the set of hermitian matrices. Clearly, this also implies $I - N \preceq I - M$. The partial order is preserved under the conjugation rule. That is

$$S M S^* \preceq S N S^* \quad \forall S \in \mathbb{C}^{n \times n}.$$ 

Weyl’s theorem implies that the eigenvalues satisfy $\lambda_j(M) \leq \lambda_j(N)$ for all $j = 1, \ldots, n$. If additionally, $M, N$ are both positive semi-definite then $M^{1/2} \preceq N^{1/2}$ [4, Proposition V.1.8].

Singular value inequalities. Let $A, B \in \mathbb{C}^{m \times n}$. For all $i, j$ such that $1 \leq i, j \leq \min\{m, n\}$ and $i + j - 1 \leq \min\{m, n\}$, the following singular value inequalities hold for the sum $A + B$ [15, Equation 7.3.13]

$$(17) \quad \sigma_{i+j-1}(A + B) \leq \sigma_i(A) + \sigma_j(B),$$ 

and product $AB^*$ [15, Equation (7.3.14)]

$$(18) \quad \sigma_{i+j-1}(AB^*) \leq \sigma_i(A)\sigma_j(B).$$

A useful corollary of these results is that $\sigma_i(A + B) \leq \sigma_i(A) + \sigma_1(B)$ and $\sigma_i(AB^*) \leq \sigma_i(A)\sigma_1(B)$ for $i = 1, \ldots, \min\{m, n\}$. 

11
5.1. Proofs of Section 3.1 Theorems.

Proof (Theorem 1). We tackle each case separately.

Bounds for $\sin \theta_j$. The proof is lengthy and proceeds in four steps. We give a great level of detail here, since the proof technique will be applicable to the subsequent proofs.

1. **Converting an SVD to an EVD.** We compute the thin SVD of $(I - \mathcal{P}_U)U_1 = KS_U^*G^*$. The matrix

\[ S_U = \text{diag} (\sin \theta_1, \ldots, \sin \theta_k) \in \mathbb{R}^{k \times k} \]

contains the sine of the canonical angles between the subspaces spanned by the columns of $\hat{U}$ and $U_1$ [5, Equation (13)]. It is readily seen that

\[ GS_U^2 G^* = U_1^*(I - \mathcal{P}_U^*)U_1. \]

2. **Shrinking space.** In Algorithm 1, we had defined $Y = (AA^*)^q A \Omega$. It follows that

\[ U^*Y = \begin{bmatrix} \Sigma_1^{2q+1} \\ \Sigma_2^{2q+1} \end{bmatrix} (V^* \Omega) = \begin{bmatrix} \Sigma_1^{2q+1} \Omega_1 \\ \Sigma_2^{2q+1} \Omega_2 \end{bmatrix}, \]

where $\Omega_1 \equiv V_1^* \Omega$ and $\Omega_2 \equiv V_2^* \Omega$. Next, by Assumption 1, $\Omega_1$ has full row rank and therefore it has a right multiplicative inverse. Define

\[ Z \equiv U^*Y \Omega_1^\dagger \Sigma_1^{-2q-1} = \begin{bmatrix} I \\ F \end{bmatrix} \quad F \equiv \Sigma_2^{2q+1} \Omega_2 \Omega_1^\dagger \Sigma_1^{-2q-1}. \]

Recall that $Y = QR$ is the thin-QR factorization. Let $Q_1 R_1$ be the thin-QR factorization of $R \Omega_2 \Omega_1^\dagger$; here $Q_1 \in \mathbb{C}^{\ell \times k}, R_1 \in \mathbb{C}^{k \times k}$. From $Q_1 Q_1^* \succeq I$, the conjugation rule implies

\[ \mathcal{P}_Z = U^*Q_1 Q_1^* U \preceq U^* Q Q^* U = \mathcal{P}_U. \]

Since $\mathcal{R}(U^*Y) = \mathcal{R}(U^*Q) = \mathcal{R}(U^*\hat{U})$, they have the same projectors, so

\[ \mathcal{P}_Z \preceq \mathcal{P}_U \quad I - \mathcal{P}_U \preceq I - \mathcal{P}_Z. \]

Plug in $UU^* = I$ into (19), and use (20) to obtain

\[ U_1^*(I - \mathcal{P}_U)U_1 = U_1^*U(I - \mathcal{P}_U)U^*U_1 \preceq \begin{bmatrix} I & 0 \end{bmatrix} (I - \mathcal{P}_Z) \begin{bmatrix} I \\ 0 \end{bmatrix}. \]

3. **Simplifying $\mathcal{P}_Z$.** Since $\mathcal{P}_Z = ZZ^\dagger$, and $Z = \begin{bmatrix} I \\ F \end{bmatrix}$, we have

\[ \mathcal{P}_Z = \begin{bmatrix} I \\ F \end{bmatrix} (I + F^*F)^{-1} \begin{bmatrix} I \\ F \end{bmatrix}, \]

from which, it can be readily seen that

\[ \begin{bmatrix} I & 0 \end{bmatrix} (I - \mathcal{P}_Z) \begin{bmatrix} I \\ 0 \end{bmatrix} = I - (I + F^*F)^{-1} \]

\[ = (I + F^*F)^{-1/2} F^* F (I + F^*F)^{-1/2} \equiv H. \]

To summarize the story so far, $GS_U^2 G^* \preceq H$. 

12
4. Applying singular value inequalities. A straightforward SVD argument shows that the $j$-th singular value of $H$ satisfies

$$\sigma_j(H) = \sigma_j^2(F)/(1 + \sigma_j^2(F)) \quad j = 1, \ldots, k.$$ 

The singular value inequalities (18) imply

$$\sigma_j(F) \leq \sigma_1(\Sigma_2^{2q+1} \Omega_2 \Omega_1^T) \sigma_j(\Sigma_1^{-2q-1}) \leq \left(\frac{\sigma_{k+1}}{\sigma_{k-j+1}}\right)^{2q+1} \left\| \Omega_2 \Omega_1^T \right\|_2^2.$$

Plugging this inequality into $\sigma_j(H)$

$$\sigma_j^2(H) \leq \frac{\sigma_{k-j+1}^{4q+2} \left\| \Omega_2 \Omega_1^T \right\|_2^2}{1 + \sigma_{k-j+1}^{4q+2} \left\| \Omega_2 \Omega_1^T \right\|_2^2} \quad j = 1, \ldots, k.$$

Since $GS_2^\theta G^* \preceq H$, Weyl’s theorem implies $\sin \theta_{k-j+1} \leq \sigma_j^2(H)$. Take square roots on both sides and rename $j \leftarrow k - j + 1$ to get the desired result.

**Bounds for** $\sin \nu_j$ Let $GS_2^\theta G^*$ be the eigenvalue decomposition of $V_1^*(I - P_\theta) V_1$. Note that the diagonals of $S_V$ are the sine of the canonical angles $\angle(V_1, \tilde{V})$. Since $\tilde{V}$ is obtained from the thin SVD of $A^* Q$, $R(A^* Q) = R(\tilde{V})$ and $P_\tilde{V} = P_{A^* Q}$, since an orthogonal projection matrix is uniquely determined by the range. Next, consider $\tilde{Z}$ defined as

$$\tilde{Z} \equiv \Sigma^* U^* Y \Omega_1^T \Sigma_1^{-2q-2} = \begin{bmatrix} I \\ \tilde{F} \end{bmatrix} \quad \tilde{F} \equiv \Sigma_2^{2q+2} \Omega_2 \Omega_1^T \Sigma_1^{-2q-2},$$

from $(AV)^*Q = \Sigma^* U^* Q$, it can be verified that

$$R(\tilde{Z}) \subset R(\Sigma^* U^* Y) = R(\Sigma^* U^* Q) = R((AV)^*Q).$$

Using an argument similar to (20), we obtain

$$V_1^* V(I - P_\tilde{V}) V^* V_1 \preceq V_1^* V(I - P_{\tilde{Z}}) V^* V_1 = \begin{bmatrix} I & 0 \end{bmatrix} (I - P_{\tilde{Z}}) \begin{bmatrix} I \\ 0 \end{bmatrix}.$$ 

The right hand side simplifies to $I - (I + \tilde{F}^* \tilde{F})^{-1}$. The rest of the proof is similar to that of the proof for $\sin \theta_j$. \hfill \Box

**Proof (Theorem 2).** With the notation of Theorem 1, we follow steps 1-3 to obtain

$$GS_2^\theta G^* \preceq H \preceq F^* F.$$ 

Since the square root preserves partial ordering, implies

$$GS_2^\theta G^* \preceq (F^* F)^{1/2} \preceq \left\| \Omega_2 \Omega_1^T \right\|_2^{\frac{\left\| \Omega_2 \Omega_1^T \right\|_2^{\frac{1}{2}}}{\sigma_k}}.$$ 

Therefore, it follows that $\left\| \sin \angle(U_1, \tilde{U}) \right\| \leq \left\| \Omega_2 \Omega_1^T \right\|_2^{\frac{1}{2}} \frac{\left\| \Omega_2 \Omega_1^T \right\|_2}{\sigma_k}. \hfill \Box$
5.2. Proofs of Section 3.2 Theorems.

Proof (Theorem 3). Let $X = (I - \mathcal{P}_{\mathcal{U}_i})\mathcal{P}_{\mathcal{V}_i}$ and $Y = (I - \mathcal{P}_{\mathcal{V}_i})\mathcal{P}_{\mathcal{V}_i}$. The singular values of $X$ and $Y$ are $\{\sin \theta_j^i\}_{j=1}^k$ and $\{\sin \nu_j^i\}_{j=1}^k$, respectively, in decreasing order. Also, let $B \equiv \hat{A} - \lfloor \hat{A} \rfloor$. First, we observe that

$$E_{12} = (I - \mathcal{P}_{\mathcal{U}_i})(A - \hat{A})\mathcal{P}_{\mathcal{V}_i}$$

$$= (I - \mathcal{P}_{\mathcal{U}_i})[[A]]_k - (I - \mathcal{P}_{\mathcal{U}_i})\hat{A}\mathcal{P}_{\mathcal{V}_i}$$

$$= (I - \mathcal{P}_{\mathcal{U}_i})\mathcal{P}_{\mathcal{V}_i}[[A]]_k - (\hat{A} - \lfloor \hat{A} \rfloor)\mathcal{P}_{\mathcal{V}_i}$$

$$= X[[A]]_k - B(I - \mathcal{P}_{\mathcal{V}_i})\mathcal{P}_{\mathcal{V}_i} = X[[A]]_k - BY.$$

A similar calculation shows that $E_{21} = X^*B - [[A]]_kY^*$. From the first relation, since $\text{rank}(A) \geq k$, we have

$$X[[A]]_k[[A]]_k^T = (E_{12} + BY)[[A]]_k^T.$$

But $[[A]]_k[[A]]_k^T = \mathcal{P}_{\mathcal{U}_i}$ and $X\mathcal{P}_{\mathcal{U}_i} = X$. Applying (18), we have

$$\sigma_j(X) \leq (\|E_{12}\|_2 + \|B\|_2\|Y\|_2)/\sigma_{k-j+1}(A) \quad j = 1, \ldots, k.$$

A similar argument gives

$$\sigma_j(Y) \leq (\|E_{21}\|_2 + \|B\|_2\|X\|_2)/\sigma_{k-j+1}(A) \quad j = 1, \ldots, k.$$

Combining these relations

$$\max\{\sigma_j(X), \sigma_j(Y)\} \leq \frac{\max\{\|E_{21}\|_2, \|E_{12}\|_2\}}{\sigma_{k-j+1}(A)} + \frac{\|B\|_2}{\sigma_{k-j+1}(A)} \max\{\|X\|_2, \|Y\|_2\}.$$

Recognize that $\|B\|_2 = \sigma_{k+1}(\hat{A})$. Applying (5) in the spectral norm simplifies the expression since

$$\frac{1}{\sigma_{k-j+1}(A)} \left( 1 + \frac{\sigma_{k+1}(\hat{A})}{\sigma_k(A)} \right) = \frac{\sigma_k(A)}{\sigma_{k-j+1}(A)(\sigma_k(A) - \sigma_{k+1}(\hat{A}))}.$$

Therefore,

$$\max\{\sigma_j(X), \sigma_j(Y)\} \leq \frac{\sigma_k(A)}{\sigma_{k-j+1}(A)} \frac{\max\{\|E_{21}\|_2, \|E_{12}\|_2\}}{\text{gap}}.$$

Now $\sigma_j(X) = \sin \theta_j^i - j$ and $\sigma_j(Y) = \sin \nu_j^i - j + 1$. Rename $j \leftarrow k - j + 1$ to finish. \hfill \square

Proof (Theorem 4). We tackle each case independently.

**Unitarily invariant norms** Our proof involves simplifying each term in (5), and (6) and has several steps.

1. *Simplifying the gap.* Recall $\text{gap} = \sigma_k(A) - \sigma_{k+1}(\hat{A})$ and $\hat{A} = QQ^*A$. It can be shown that $\sigma_j(QQ^*A) = \sigma_j(Q^*A)$ for $j = 1, \ldots, \ell$. From the first part of Theorem 9

$$\text{gap} = \sigma_k(A) - \sigma_{k+1}(\hat{A}) \geq \sigma_k(A) - \sigma_{k+1}(A).$$
2. **Simplifying** $\|E_{12}\|$. First observe that $APV = [[A]]_k$. So 
\[ E_{12} = (I - P_{U_1})(I - QQ^*)APV = (I - P_{U_1})(I - QQ^*)[[A]]_k. \]
Then applying (12) along with submultiplicativity gives
\[ \|E_{12}\| \leq \|(I - QQ^*)[[A]]_k\| \leq \gamma_k^{2q} \|\Sigma_2 \Omega_1^j\|_2 \leq \gamma_k^{2q} \|\Sigma_2\| \|\Omega_2 \Omega_1^j\|_2. \]

3. **Simplifying** $\|E_{21}\|$. We can simplify $E_{21}$ in (6). First, $E_{21} = P_{U_1}(I - QQ^*)APV$, and since $\|P_{U_1}(I - QQ^*)\|_2 = \|\sin\angle(U_1, \tilde{U})\|_2$,
\[ \|E_{21}\| \leq \|\sin\angle(U_1, \tilde{U})\|_2 \|(I - QQ^*)A\|, \]
because of strong sub-multiplicativity. Applying Theorem 1 and (12)
\[ \|E_{21}\| \leq \frac{\gamma_k^{2q+1} \|\Omega_2 \Omega_1^j\|_2}{\sqrt{1 + \gamma_k^{4q+2} \|\Omega_2 \Omega_1^j\|_2^2}} \left(1 + \gamma_k^{2q} \|\Omega_2 \Omega_1^j\|_2\right) \|\Sigma_2\|. \]
Let $\beta = \gamma_k^{2q} \|\Omega_2 \Omega_1^j\|_2$. Then for $\beta \geq 0$, since $\gamma_k < 1$
\[ \frac{\gamma_k(1 + \beta)}{\sqrt{1 + \gamma_k^2 \beta^2}} \leq \frac{1 + \gamma_k \beta}{\sqrt{1 + \gamma_k^2 \beta^2}} \leq \sqrt{2}. \]
Therefore, $\|E_{21}\| \leq \sqrt{2} \gamma_k^{2q} \|\Sigma_2\| \|\Omega_2 \Omega_1^j\|_2$.

4. **Putting everything together.** Plugging in the intermediate quantities into (5), we have
\[ \max \left\{ \|\sin\angle(U_1, \tilde{U}_1)\|_2, \|\sin\angle(V_1, \tilde{V}_1)\|_2 \right\} \leq \sqrt{2} \gamma_k^{2q} \|\Omega_2 \Omega_1^j\|_2 \frac{\|\Sigma_2\|}{\sigma_k - \sigma_{k+1}}. \]
Dividing the numerator and denominator by $\sigma_k$ proves the stated result for unitarily invariant norms.

**Spectral/Frobenius norms** Let $\|\cdot\|_\xi$ denote the spectral and Frobenius norms. The first two steps are identical to the case for unitarily invariant norms. For the third step, using (12)
\[ \|E_{21}\|_\xi \leq \frac{\gamma_k^{2q+1} \|\Omega_2 \Omega_1^j\|_2}{\sqrt{1 + \gamma_k^{4q+2} \|\Omega_2 \Omega_1^j\|_2^2}} \|\Sigma_2\|_\xi \sqrt{1 + \gamma_k^{4q} \|\Omega_2 \Omega_1^j\|_2^2}. \]
With $\beta$ defined as before, since $\gamma_k < 1$, $\sqrt{\gamma_k^2 + \gamma_k^{4q} \beta^2} / \sqrt{1 + \gamma_k^2 \beta^2} \leq 1$. Therefore, $\|E_{21}\|_\xi \leq \gamma_k^{2q} \|\Sigma_2\|_\xi \|\Omega_2 \Omega_1^j\|_2$. The rest of the proof is the same.

**Canonical angles** Use Theorem 3. The term on the right hand side is nothing but $\max \left\{ \|\sin\angle(U_1, \tilde{U}_1)\|_2, \|\sin\angle(V_1, \tilde{V}_1)\|_2 \right\}$. The rest of the proof involves some simple manipulations.

**Proof (Theorem 5).** We first address (7). Following the steps of Theorem 1 we have
\[ \sin^2 \angle(u_j, \tilde{U}) = u_j^* U(I - P_{U^*})U^* u_j \leq [e_j^* \ 0] (I - P_Z) \begin{bmatrix} e_j \\ 0 \end{bmatrix}, \]
where \( e_j \) is the \( j \)-th column of the \( k \times k \) identity matrix. Therefore, we have \( \sin^2 \angle(u_j, \hat{U}) \leq e_j^*He_j \), where \( H \) was defined in (21). The inequality \( H \preceq F^*F \) implies
\[
\sin^2 \angle(u_j, \hat{U}) \leq \sigma_j^{-4q-2}e_j^*(\Omega_2\Omega_1^\dagger)\Sigma_2^{2q+2}(\Omega_2\Omega_1^\dagger)e_j \\
\quad \leq \sigma_j^{-4q-2}\left\| \Sigma_2^{2q+1}(\Omega_2\Omega_1^\dagger)e_j \right\|_2^2 \\
\quad \leq \gamma_j^{4q+2}\left\| \Omega_2\Omega_1^\dagger \right\|_2^2.
\]
Taking square-roots on both sides gives the desired results. The strategy for bounding the canonical angles \( \sin \angle(v_j, \hat{V}) \) is very similar and will be omitted.

We now address (8), which is a straightforward application of [13, Theorem 2.5]. Let \( P_u = QQ^* \) and \( P_V = I \). Then, in our notation, this result takes the form
\[
\max \{ \sin \angle(u_j, \hat{U}), \sin \angle(v_j, \hat{V}) \} \leq \sqrt{1 + \frac{\hat{\gamma}^2}{\hat{\beta}^2}} \max \{ \sin \angle(u_j, \hat{U}), \sin \angle(v_j, I) \},
\]
where \( \hat{\gamma}' = \max \{ 0, \| (I - QQ^*)A \|_2 \} \) and \( \hat{\beta} \) is as defined in the statement of the theorem. Theorem 8 for the spectral norm implies \( \hat{\gamma}' \leq \hat{\gamma} \), whereas Theorem 5 implies
\[
\max \{ \sin \angle(u_j, \hat{U}), \sin \angle(v_j, I) \} \leq \gamma_j^{2q}\left\| \Omega_2\Omega_1^\dagger \right\|_2.
\]
Plug in the intermediate steps to obtain the desired bound.

**Proof (Theorem 6).** With the constant \( C_\epsilon \) defined in [11, Theorem 5.8] says
\[
\mathbb{E} \left\| \Omega_2\Omega_1^\dagger \right\|_2 \leq C_\epsilon.
\]
Let \( \alpha > 0 \) be a constant. The map \( x \mapsto x/\sqrt{1 + \alpha x^2} \) is convex. Therefore, by Jensen’s inequality the results in expectation follow.

For the concentration inequalities, [11, Theorem 5.8] showed \( \left\| \Omega_2\Omega_1^\dagger \right\|_2 \leq C_d \) with a probability at least \( 1 - \delta \). Plug into Theorem 1 to obtain the desired bounds.

### 5.3. Proofs of Section 4 Theorems

It is useful to recall some properties of the unitarily invariant norms. Every unitarily invariant norm \( \| \cdot \| \) on \( \mathbb{C}^n \) is associated with a symmetric gauge function on \( \mathbb{R}^n \). The \( \| \cdot \| \) satisfies \( \| M \| = \| (M^*M)^{1/2} \| \), since both matrices have the same nonzero singular values. The following inequality for unitarily invariant norms, also known as strong sub-multiplicativity, will be useful [4, (IV.40)]
\[
\| ABC \| \leq \| A \|_2 \| C \|_2 \| B \|.
\]
We will also need

**Lemma 10.** Let \( A, B, D \in \mathbb{C}^{n \times n} \) such that with \( A, B \) hermitian and \( 0 \leq A \leq B \), then
\[
\left\| (D^*AD)^{1/2} \right\| \leq \left\| (D^*BD)^{1/2} \right\|.
\]

**Proof.** Combining the properties of the partial ordering, the eigenvalues of the scaled matrices satisfy \( \lambda_j(D^*AD)^{1/2} \leq \lambda_j(D^*BD)^{1/2} \) for all \( j = 1, \ldots, n \). Since the matrices are positive semidefinite, the eigenvalues are the singular values and
\((D^*AD)^{1/2}\|_k \leq (D^*BD)^{1/2}\|_k\) for every Ky-Fan-k norm \(k = 1, \ldots, n\). By the Fan dominance theorem [4, Theorem 4.2.2], the advertised inequality is true for every unitarily invariant norm.

We are finally ready to prove our main result for error analysis in unitarily invariant norms.

**Proof (Theorem 7).** Using the unitary invariance of the norms

\[
\|(I - \mathcal{P}_Q)A\| = \|(I - \mathcal{P}_{U^*Q})\Sigma\| = \left\|\left(\Sigma^T(I - \mathcal{P}_{U^*Q})\Sigma\right)^{1/2}\right\|.
\]

We use (20) combined with Lemma 10 to obtain

\[
\left\|\left(\Sigma^T(I - \mathcal{P}_{U^*Q})\Sigma\right)^{1/2}\right\| \leq \left\|\left(\Sigma^T(I - \mathcal{P}_{Z})\Sigma\right)^{1/2}\right\|.
\]

With \(M_1 \equiv I - (I + F^*F)^{-1}\) and \(M_2 \equiv I - F(I + F^*F)^{-1}F^*\), then \(\Sigma^T(I - \mathcal{P}_Z)\Sigma\) simplifies as

\[
\Sigma^*(I - \mathcal{P}_Z)\Sigma = \begin{bmatrix} \Sigma^T M_1 \Sigma_1 & \Sigma^T M_2 \Sigma_2 \end{bmatrix}.
\]

The square root function is concave on \([0, \infty)\) and \(\Sigma^T(I - \mathcal{P}_Z)\Sigma\) is positive semidefinite. Therefore, an extension to Rotfel’d’s theorem says [16, Theorem 2.1]

\[
\left\|\left(\Sigma^T(I - \mathcal{P}_Z)\Sigma\right)^{1/2}\right\| \leq \left\|\Sigma_1^T M_1 \Sigma_1\right\|^{1/2} + \left\|\Sigma_2^T M_2 \Sigma_2\right\|^{1/2}.
\]

Use the inequalities \(M_1 \preceq F^*F\) and \(M_2 \preceq I\), along with Lemma 10 gives

\[
\|(I - \mathcal{P}_Q)A\| \leq \|F\Sigma_1\| + \|\Sigma_2\|.
\]

Use \(F\Sigma_1 = \Sigma_2^{(2q+1)} \Omega_2 \Omega_1 \Sigma_1^{-2q}\) and the submultiplicativity to obtain the advertised bounds.

The proof for (12) is similar and is omitted. The main observation is that \([A]_k\) has only \(k\) nonzero singular values.

For (13), we follow the strategy in [7, Section 3.3]. Write \(A = [[A]]_k + A_{k,\perp}\) where \([[A]]_k\) is the best rank-\(k\) approximation to \(A\). Noting that \(Q[[Q^*A]]_k = \mathcal{P}_{U_1}A\), the triangle inequality gives

\[
\left\|\left(I - \mathcal{P}_{\hat{U}_1}\right)A\right\| \leq \left\|\left(I - \mathcal{P}_{\hat{U}_1}\right)[[A]]_k\right\| + \left\|\left(I - \mathcal{P}_{\hat{U}_1}\right)A_{k,\perp}\right\|.
\]

Since \([[A]]_k = \mathcal{P}_{U_1}[[A]]_k\), applying strong sub-multiplicativity

\[
\left\|\left(I - \mathcal{P}_{\hat{U}_1}\right)A\right\| \leq \left\|\left(I - \mathcal{P}_{\hat{U}_1}\right)\mathcal{P}_{U_1}\right\| \left\|[A]_k\right\|_2 + \|A_{k,\perp}\|.
\]

We recognize that \(\left\|\left(I - \mathcal{P}_{\hat{U}_1}\right)\mathcal{P}_{U_1}\right\| = \left\|\sin \angle(U_1, \hat{U}_1)\right\|\), apply Theorem 4 to complete the proof.

**Proof (Theorem 8).** The proof is similar to Theorem 7. Consider the term of interest \(\|(I - QQ^*)A\|_p^2\), which can be simplified to

\[
\|(I - QQ^*)A\|_p^2 = \|A^*(I - QQ^*)A\|_{p/2} = \left\|\Sigma^T(I - \mathcal{P}_{U^*Q})\Sigma\right\|_{p/2}.
\]
The first equality holds only for \( p \geq 2 \), whereas the last equality follows because of the unitary invariance. As in Theorem 1, we have

\[
Z = U^* Y \Omega^1 \Sigma^{-1} (2q+1) \quad F = \Sigma^2 (2q+1) \Omega^1 \Sigma^{-1} (2q+1).
\]

The use of (20) and Lemma 10 ensures

\[
\|\Sigma^T (I - \mathcal{P}_{U^* Q}) \Sigma\|_{p/2} \leq \|\Sigma^T (I - \mathcal{P}_Z) \Sigma\|_{p/2}.
\]

We apply [16, Theorem 2.1] to (23) with \( f(t) = t \) to obtain

\[
\|\Sigma^T (I - \mathcal{P}_Z) \Sigma\|_{p/2} \leq \|\Sigma^T \mathcal{M}_1 \Sigma_1\|_{p/2} + \|\Sigma^T \mathcal{M}_2 \Sigma_2\|_{p/2} = \|F \Sigma_1\|^2_p + \|\Sigma_2\|^2_p.
\]

We have used \( \mathcal{M}_1 \equiv I - (I + F^* F)^{-1} \leq F^* F \) and \( \mathcal{M}_2 \equiv I - F(I + F^* F)^{-1} F^* \leq I \).

The rest of the proof is similar to that of Theorem 7.

**Proof (Theorem 9).** From the inequality \( I \succeq QQ^* \), the conjugation rule gives

\[
A^* A \succeq A^* QQ^* A.
\]

Then Weyl’s theorem implies \( \lambda_i(A^* A) \geq \lambda_i(A^* QQ^* A) \) for \( i = 1, \ldots, k \). Relating the eigenvalues to the singular values proves the first inequality.

For the second inequality consider again \( A^* QQ^* A \). With the aid of (20)

\[
A^* QQ^* A = V \Sigma^T \mathcal{P}_{U^* Q} \Sigma V^* \succeq V \Sigma^T \mathcal{P}_Z \Sigma V^*.
\]

Therefore, \( \lambda_i(A^* QQ^* A) \geq \lambda_i(V \Sigma^T \mathcal{P}_Z \Sigma V^*) \) for \( i = 1, \ldots, k \).

Since \( V \Sigma^T \mathcal{P}_Z \Sigma V^* \) and \( \Sigma^T \mathcal{P}_Z \Sigma \) are similar, they share the same eigenvalues. It can be readily shown that

\[
\Sigma^T \mathcal{P}_Z \Sigma = \begin{bmatrix} \Sigma_1^T (I + F^* F)^{-1} \Sigma_1 & * \\ \ast & \ast \end{bmatrix}.
\]

Because of the partial ordering (24), [15, Corollary 7.7.4 (c)] implies

\[
\lambda_i(A^* QQ^* A) \geq \lambda_i(V \Sigma^T \mathcal{P}_Z \Sigma V^*) \geq \lambda_i(\Sigma_1^T (I + F^* F)^{-1} \Sigma_1).
\]

The second inequality follows from Cauchy interlacing theorem [19, Section 10-1]. Since \( \Sigma_1 \) is invertible,

\[
\Sigma_1^T (I + F^* F)^{-1} \Sigma_1 = (I + \Sigma_1^{-T} F^* F \Sigma_1^{-1})^{-1}.
\]

Next, consider the singular values of \( (I + \Sigma_1^{-T} F^* F \Sigma_1^{-1})^{-1} \) which, arranged in descending order, are equal to \( 1/(1 + \sigma_{k-j+1}^2 (F \Sigma_1^{-1})) \). Applying multiplicative singular value inequalities (18) as in Theorem 1 gives

\[
\sigma_j(F \Sigma_1^{-1}) \leq \gamma_{k-j+1} \sigma_{k-j+1}^{-1} (A) \|\Omega^1 \Sigma_1^{-1}\|_2^2,
\]

and therefore, it follows that

\[
\sigma_j(Q^* A) \geq \sigma_j(\Sigma_1^T (I + F^* F)^{-1} \Sigma_1) = \frac{1}{1 + \sigma_{k-j+1}^2 (F \Sigma_1^{-1})} \geq \frac{\sigma_j^2(A)}{1 + \gamma_{k-j} \|\Omega^1\|_2^2}.
\]

This proves the second inequality.
6. Numerical Results.

6.1. Test matrices. To demonstrate the performance of the bounds, we use the following test matrices

1. Controlled gap The first set of test matrices $A \in \mathbb{R}^{3000 \times 300}$ are constructed using the formula

$$A = \sum_{j=1}^{r} \frac{\text{gap}}{j} x_j y_j^\top + \sum_{j=r+1}^{300} \frac{1}{j} x_j y_j^\top,$$

where $x_j \in \mathbb{R}^{3000}$ and $y_j \in \mathbb{R}^{300}$ are sparse random vectors with non-negative entries generated using the MATLAB commands `sprand(3000,1,0.025)` and `sprand(300,1,0.025)` respectively. While the formula above is not an SVD, since the vectors are not orthogonal. Nonetheless, the singular values decay like $1/j$ and the gap between the singular values between 15 and 16 is controlled by the parameter gap. We consider three cases:

(a) Small gap (GapSmall) gap = 1,
(b) Medium gap (GapMedium) gap = 2,
(c) Large gap (GapLarge) gap = 10.

2. Low-rank plus noise The matrices are of the form

$$A = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} + \sqrt{\frac{\gamma_n}{2n^2}} (G + G^\top),$$

where $G \in \mathbb{R}^{n \times n}$ is a random Gaussian matrix. We consider three cases:

(a) Small noise (NoiseSmall) $\gamma_n = 10^{-2}$,
(b) Medium noise (NoiseMedium) $\gamma_n = 10^{-1}$,
(c) Large noise (NoiseLarge) $\gamma_n = 1.0$. 

Fig. 1. Matrices from the (left) ‘Controlled Gap’ example, (middle) ‘Low-rank plus noise’ example, (Right) ‘Low-rank plus decay’ example.
(c) Large noise (NoiseLarge) $\gamma_n = 1$.

3. **Low-rank plus decay** The matrices take the form

$$A = U\operatorname{diag}(1, 1, \ldots, 1, 2^{-d}, 3^{-d}, \ldots, (n-r+1)^{-d})V^*.$$ 

The unitary matrices $U, V$ are obtained by drawing a random Gaussian matrix, and taking its QR factorization. We distinguish between the following cases

(a) Slow decay (DecaySlow): $d = 0.5$,
(b) Medium decay (DecayMedium): $d = 1.0$,
(c) Fast decay (DecayFast): $d = 2.0$.

The first example is adapted from [21], whereas the second and third examples are drawn from [23]. In all the examples, the random matrices were fixed by setting the random seed and we set $r = 15$. The singular values of all the test matrices are plotted in Figure 1.

6.2. **Canonical angles.** For the first numerical example, we use the 9 test matrices in Section 6.1. For each matrix, we chose an oversampling parameter $\rho = 20$ and the target rank $k$ was chosen to be 25. The starting guess $\Omega$ was taken to be a random Gaussian matrix. We plot the canonical angles $\sin \angle(U_1, \hat{U})$ in solid lines, the corresponding bounds from Theorem 1 are also plotted in dashed lines. The results are displayed in Figure 2. We make the following observations

- The influence of the subspace iterations on the canonical angles is clear: the angles become smaller, as the number of iterations $q$ increases. This implies that the subspace is becoming more accurate.
- If there is a large singular value gap in the spectrum, this means that all the canonical angles below that index are captured accurately. This is prominently seen in Figure 2(e), in which there is a large gap between singular values 15 and 16. Similar observations can be made in the other figures.
- Similarly, as the decay in the singular values becomes more rapid, the corresponding canonical angles become smaller. The canonical angles for DecayFast is much smaller than DecaySlow.
- In all the figures the bounds are qualitatively informative, but in some figures, the bounds are also quantitatively accurate.
- Similar results were observed for $\sin \angle(V_1, \hat{V})$ and, therefore, omitted. Likewise, the bounds for Theorems 5 and 2 yield similar looking plots and, therefore, we did include them.

Our next experiment is to test the performance of the extraction step. We now compute $\sin \theta_j'$ and $\sin \nu_j'$ for the test matrices described above. We plot the quantities $\max\{\sin \theta_j', \sin \nu_j'\}$ for $j = 1, \ldots, k$ in solid lines. The corresponding bounds from Theorem 4 are plotted in dashed lines. Here the target rank was chosen to be $k = 15$, to exploit the singular value gap here. We make the following observations

- The extraction step did not significantly affect the canonical angles. Conclusions similar to Theorem 1 can, therefore, be drawn. The subspaces are more accurate as the number of iterations increase, and if there is a large singular value gap at index $j$, then the canonical angles with index $j' < j$ are captured accurately.
- Compared to Theorem 1, the bounds in Theorem 4 are only qualitatively good. One reason is that the error bounds in Theorem 1 are strictly less than
Fig. 2. Accuracy of the canonical angles $\sin \theta_j$ for $j = 1, \ldots, k$. The test matrices were described in § 6.1. The target rank $k = 25$ and an oversampling parameter of 20 was chosen for all the experiments. The solid lines correspond to the computed errors, the dashed lines correspond to bounds obtained using Theorem 1.

1, but this is not true for the bounds in Theorem 4. Furthermore, the bound in Theorem 4 has the factor $1/(1 - \gamma_k)$ in the denominator, which can be quite large. Better bounds may be possible, but we could not immediately see how to derive them.

• We also compared the accuracy of the individual singular vectors (not shown here). The results and the conclusions are similar.

6.3. Singular Values. We now consider the accuracy of the singular values. We use the same test matrices and the remaining parameters are kept fixed. The computed singular values are plotted against the upper and lower bounds. We make the following observations.

• For the large singular values, both the upper and lower bounds are very accurate for all the examples that we tested. The singular values are accurately captured for the NoiseMed and the DecayMedium examples, examples for which there is a large gap or the singular values decay rapidly. The bounds
were the worst for GapSmall.

- As the number of iterations increase, the singular values are increasingly accurately estimated and are practically indistinguishable from the upper bounds (the exact singular values). However, the lower bounds are not great approximations for the singular values close to the target rank. However, the bounds improve as the number of iterations increase.

- The bounds for the singular values captured the behaviour of the singular values far better than the bounds for the canonical angles.

7. Acknowledgements. The author would like to thank Ilse Ipsen and Andreas Stathopoulos for helpful conversations. He would also like to acknowledge Ivy Huang for her help with the figures.

REFERENCES
Fig. 4. Accuracy of the singular values. The test matrices were described in § 6.1. The target rank \( k = 25 \) and an oversampling parameter of 20 was chosen for all the experiments. The solid lines black and blue lines correspond to the upper and lower bounds respectively, the dashed red lines correspond to bounds obtained using Theorem 9.

[1] H. Avron, C. Boutsidis, S. Toledo, and A. Zouzias. Efficient dimensionality reduction for canonical correlation analysis. In International Conference on Machine Learning, pages 347–355, 2013.
[2] O. Balabanov and A. Nouy. Randomized linear algebra for model reduction. Part I: Galerkin methods and error estimation. arXiv preprint arXiv:1803.02602, 2018.
[3] F. L. Bauer. Das verfahren der treppeniteration und verwandte verfahren zur lösung algebraischer eigenwertprobleme. Zeitschrift für angewandte Mathematik und Physik ZAMP, 8(3):214–235, May 1957.
[4] R. Bhattacharyya. Matrix analysis, volume 169 of Graduate Texts in Mathematics. Springer-Verlag, New York, 1997.
[5] A. Björck and G. H. Golub. Numerical methods for computing angles between linear subspaces. Mathematics of computation, 27(123):579–594, 1973.
[6] C. Boutsidis, P. Kambadur, and A. Gittens. Spectral clustering via the power method-provably. In International Conference on Machine Learning, pages 40–48, 2015.
[7] P. Drineas and I. C. F. Ipsen. Low-rank matrix approximations do not need a singular value gap. arXiv preprint arXiv:1801.00670, 2018.
[8] P. Drineas, I. C. F. Ipsen, E. M. Kontopoulou, and M. Magdon-Ismail. Structural convergence
results for approximation of dominant subspaces from block Krylov spaces. *SIAM Journal on Matrix Analysis and Applications*, 39(2):567–586, 2018.

[9] N. B. Erichson, S. L. Brunton, and J. N. Kutz. Randomized dynamic mode decomposition. *arXiv preprint arXiv:1702.02912*, 2017.

[10] G. H. Golub and C. F. Van Loan. *Matrix Computations*. The Johns Hopkins University Press, Baltimore, fourth edition, 2013.

[11] M. Gu. Subspace iteration randomization and singular value problems. *SIAM Journal on Scientific Computing*, 37(3):A1139–A1173, 2015.

[12] N. Halko, P. G. Martinsson, and J. A. Tropp. Finding Structure with Randomness: Probabilistic Algorithms for Constructing Approximate Matrix Decompositions. *SIAM Rev.*, 53(2):217–288, 2011.

[13] M. E. Hochstenbach. Harmonic and refined extraction methods for the singular value problem, with applications in least squares problems. *BIT Numerical Mathematics*, 44(4):721–754, 2004.

[14] J. T. Holodnak, I. C. F Ipsen, and T. Wentworth. Conditioning of leverage scores and computation by QR decomposition. *SIAM Journal on Matrix Analysis and Applications*, 36(3):1143–1163, 2015.

[15] R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge University Press, Cambridge, second edition, 2013.

[16] E. Y. Lee. Extension of Rotfel’d’s theorem. *Linear Algebra and its Applications*, 435(4):735–741, 2011.

[17] M. W. Mahoney. Randomized algorithms for matrices and data. *Foundations and Trends® in Machine Learning*, 3(2):123–224, 2011.

[18] Y. Nakatsukasa. Accuracy of singular vectors obtained by projection-based SVD methods. *BIT Numerical Mathematics*, pages 1–16, 2017.

[19] B. N. Parlett. *The Symmetric Eigenvalue Problem*. Prentice Hall, Englewood Cliffs, 1980.

[20] Y. Saad. *Numerical methods for large eigenvalue problems*, volume 66 of *Classics in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2011. Revised edition of the 1992 original [1177405].

[21] D. C. Sorensen and M. Embree. A DEIM induced CUR factorization. *SIAM Journal on Scientific Computing*, 38(3):A1454–A1482, 2016.

[22] P. T. K. Tang and E. Polizzi. FEAST as a subspace iteration eigensolver accelerated by approximate spectral projection. *SIAM Journal on Matrix Analysis and Applications*, 35(2):354–390, 2014.

[23] J. A Tropp, A. Yurtsever, M. Udell, and V. Cevher. Practical sketching algorithms for low-rank matrix approximation. *SIAM Journal on Matrix Analysis and Applications*, 38(4):1454–1485, 2017.

[24] P. Å. Wedin. On angles between subspaces of a finite dimensional inner product space. In Bo Kågström and Axel Ruhe, editors, *Matrix Pencils*, pages 263–285, Berlin, Heidelberg, 1983. Springer Berlin Heidelberg.

[25] K. Ye and L. H. Lim. Schubert varieties and distances between subspaces of different dimensions. *SIAM Journal on Matrix Analysis and Applications*, 37(3):1176–1197, 2016.

[26] J. Zhang, A. K. Saibaba, M. E. Kilmer, and S. Aeron. A randomized tensor singular value decomposition based on the t-product. *To appear, Numerical Linear Algebra with applications. arXiv preprint arXiv:1609.07086*, 2018.