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

ARVIND K. SAIBABA[†]

Abstract. This paper analyzes 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

AMS subject classifications. 65F15, 65F30

DOI. 10.1137/18M1179432

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. These applications include principal component analysis, facial recognition, spectral clustering, model reduction techniques such as proper orthogonal decomposition and the discrete empirical interpolation method, and approximation algorithms for partial differential and integral equations. The celebrated Eckart–Young theorem [10] 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. For example, see the survey papers [12, 17]. Randomized methods have gained in popularity since they are easy to implement, computationally efficient, and numerically robust. Although randomized algorithms tend to have the same asymptotic cost compared to classical methods, they have several advantages that make them suitable for large-scale computing. Specifically, for datasets that are too large to fit in memory, randomized algorithms are able to exploit parallel computing efficiently and are efficient in the number of times they access the data. Randomized algorithms also have excellent numerical robustness and are very reliable in practical applications.

We focus on a specific randomized algorithm known as *randomized subspace iteration*. The main idea of this method is to use random sampling to identify a subspace that approximately captures the range of the matrix. A low-rank approximation to the matrix is then obtained by projecting the matrix onto this subspace.

[†]Department of Mathematics, North Carolina State University, Raleigh, NC 27695-8205 (asaibab@ncsu.edu).

A post-processing step is then performed to compress the low-rank representation to achieve a desired target rank, and a conversion step is performed to obtain an equivalent representation in the desired format (typically, a truncated SVD representation)—both these steps are deterministic.

Many advances have been made in the analysis of randomized algorithms for low-rank approximations. The analysis typically has two stages: a *structural, or deterministic, stage*, in which minimal assumptions are made about the distribution of the random matrix, and a *probabilistic stage*, in which the distribution of the random matrix is taken into account to derive bounds for expected and tail bounds of 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 of the 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 [2, 9], leverage score computation [14], spectral clustering [6], FEAST eigensolvers [23], and 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 subsection 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 subsection 3.2 quantify the accuracy in the extraction process.
3. Subsection 3.2 also presents 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 the accuracy to be expected. Furthermore, the truncation step to extract the k-dimensional subspaces does not significantly lower the accuracy of the subspaces.

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–Wielandt 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.

Generalization of sin theta theorem. The sin theta theorem [25] 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. This result may be of independent interest beyond the study of randomized algorithms.

2. Background and preliminaries.

2.1. Notation. Denote the target rank by k and let $1 \leq k \leq \text{rank}(A)$. Let the matrix $A \in \mathbb{C}^{m \times n}$ have the SVD

$$A = \begin{bmatrix} U_k & U_\perp \end{bmatrix} \begin{bmatrix} \Sigma_k & \\ & \Sigma_\perp \end{bmatrix} \begin{bmatrix} V_k^* \\ V_\perp^* \end{bmatrix}.$$

Here, $\Sigma_k \in \mathbb{C}^{k \times k}$ and $\Sigma_\perp \in \mathbb{C}^{(m-k) \times (n-k)}$; the columns of U_k and $U_\perp$ are the corresponding left singular vectors, and the columns of V_k and $V_\perp$ are the corresponding right singular vectors. We also denote $A_k = U_k \Sigma_k V_k^*$ as the best rank-k approximation to the matrix A, in any unitarily invariant norm (for a definition, see below). We also define $A_\perp = U_\perp \Sigma_\perp V_\perp^*$ and observe that

$$A = A_k + A_\perp.$$

Singular values and ratios. Let $\|\cdot\|_2$ denote the spectral norm, so that $\|\Sigma_\perp\|_2 = \sigma_{k+1}$ and $\left\|\Sigma_k^{-1}\right\|_2 = \frac{1}{\sigma_k}$. 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

$$(1) \qquad \gamma_j = \frac{\sigma_{k+1}}{\sigma_j}, \qquad j = 1, \ldots, k.$$

Since the singular values are monotonically decreasing, the singular value ratios are monotonically increasing, i.e., $\gamma_1 \leq \cdots \leq \gamma_k \leq 1$.

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, i.e., a norm that satisfies $\|\|QAZ\|\| = \|\|A\|\|$ for unitary matrices Q, Z. An example of the unitarily invariant norms is the Schatten-p class of norms, defined as the vector ℓ_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 with column space $\mathcal{R}(Z)$; $Z^\dagger$ is a left multiplicative inverse and where $\dagger$ represents the Moore–Penrose inverse. We define the (orthogonal) projection matrix $\mathcal{P}_Z = ZZ^\dagger$. An orthogonal projection matrix is uniquely defined by its range, and $\mathcal{R}(\mathcal{P}_Z) = \mathcal{R}(Z)$. For a matrix Q with orthonormal columns, the formula simplifies and $\mathcal{P}_Q = QQ^*$.

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 $\dim \mathcal{M} = \ell$, $\dim \mathcal{N} = k$, 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

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

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

$$u_j^* u, \quad v_j^* v = 0, \qquad j = 1, \ldots, i-1.$$

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 $\mathcal{P}_\mathcal{M} - \mathcal{P}_\mathcal{N}$.

We denote $\angle(\mathcal{M}, \mathcal{N})$ to be the canonical angles between subspaces $\mathcal{M}$ and $\mathcal{N}$. Let M and N be matrices with orthonormal columns, which form bases for subspaces $\mathcal{M}$ and $\mathcal{N}$, respectively. Then, the singular values of singular values of $(I - MM^*)N$ can be used to compute $\sin \angle(\mathcal{M}, \mathcal{N})$ and the singular values of M^*N can be used to compute $\cos \angle(\mathcal{M}, \mathcal{N})$ [5, section 3]. For ease of notation, in the rest of this paper, we write $\angle(M, N)$ instead of $\angle(\mathcal{M}, \mathcal{N})$.

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." A thin-QR factorization of Y is performed to obtain Q whose columns form an orthonormal basis for the range of Y. The main idea is that, under suitable conditions, the range of Q is a good approximation for the range of A. We obtain a low-rank approximation to A by the projection $\widehat{A} = QQ^*A$. The rest of the algorithm involves converting this low-rank approximation into the SVD format.

Algorithm 1. Idealized version of subspace iteration for approximate SVD.

Require: Matrix A, starting guess $\Omega \in \mathbb{C}^{n \times (k+\rho)}$, an integer $q \geq 0$.
1: Compute $Y = (AA^*)^q A\Omega$
2: Compute thin-QR factorization of Y, so that $Y = QR$.
3: Compute $B = Q^*A$ and its SVD $B = U_B \widehat{\Sigma} \widehat{V}^*$.
4: Compute $\widehat{U} = QU_B$.
5: **return** Matrices $\widehat{U}, \widehat{\Sigma}, \widehat{V}$ that define $\widehat{A} \equiv \widehat{U}\widehat{\Sigma}\widehat{V}^*$.

The algorithm to compute an approximate SVD, given starting guess $\Omega \in \mathbb{C}^{n \times (k+\rho)}$, is summarized in Algorithm 1. We say that 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 matrix-vector

products (matvecs) involving A; for more details regarding the implementation, the reader is referred to [12, 20]. In Algorithm 1, the output

$$\widehat{A} \equiv QQ^*A = \widehat{U}\widehat{\Sigma}\widehat{V}^*$$

may have a rank larger than (or equal to) k. If a rank-k approximation to A is desired, then it can be obtained by discarding the ρ smallest singular values of $\widehat{A}$. We denote this low-rank representation by

$$\widehat{A}_k = \widehat{U}_k \widehat{\Sigma}_k \widehat{V}_k^*.$$

This is summarized in Algorithm 2.

Algorithm 2. Truncated SVD of $\widehat{A} = QQ^*A$.

Require: Matrix $A \in \mathbb{C}^{m \times n}$ and $Q \in \mathbb{C}^{m \times (k+\rho)}$. Target rank $1 \le k\mathsf{rank}(A)$.
1: Form matrix $B = Q^*A$.
2: Compute the truncated SVD representation $B_k = \widehat{U}_{B,k}\widehat{\Sigma}_k\widehat{V}_k^*$.
3: Form $\widehat{U}_k = Q\widehat{U}_{B,k}$
4: **return** Matrices $\widehat{U}_k, \widehat{\Sigma}_k, \widehat{V}_k$ such that $\widehat{A}_k = \widehat{U}_k\widehat{\Sigma}_k\widehat{V}_k^*$.

Before we state the assumptions needed for our analysis, we introduce the following notation. The matrix $V^*\Omega$ captures the influence of the starting guess on the right singular matrix V. Partition this matrix as

$$(2) \qquad V^*\Omega = \begin{bmatrix} V_k^*\Omega \\ V_\perp^*\Omega \end{bmatrix} = \begin{bmatrix} \Omega_1 \\ \Omega_2 \end{bmatrix},$$

where $\Omega_1 = V_k^*\Omega \in \mathbb{C}^{k \times (k+\rho)}$ and $\Omega_2 = V_\perp^*\Omega \in \mathbb{C}^{(n-k) \times (k+\rho)}$. As mentioned earlier, we assume that the target rank k satisfies $1 \le k \le \mathsf{rank}(A)$. Additionally, the following assumptions will be required for our analysis.

Assumption 1. Let $\Omega_1 \in \mathbb{C}^{k \times (k+\rho)}$ be defined as above. We assume that

$$(3) \qquad \mathsf{rank}(\Omega_1) = k.$$

The singular value gap at index k is inversely proportional to the singular value ratio

$$(4) \qquad \gamma_k = \|\Sigma_\perp\|_2 \left\|\Sigma_k^{-1}\right\|_2 = \frac{\sigma_{k+1}}{\sigma_k} < 1.$$

The first assumption guarantees that the starting guess Ω has a significant influence over the right singular vectors, whereas the second assumption ensures that the k-dimensional subspace $\mathcal{R}(U_k)$ is well-defined. In practice, it is highly desirable that $\gamma_k \ll 1$, which ensures that there is a large singular value gap.

3. Accuracy of singular vectors. We want to understand how well $\mathcal{R}(\widehat{U})$ approximates $\mathcal{R}(U_k)$, measured in terms of the canonical angles between the subspaces. To this end, abbreviate the subspace angles between $\widehat{U} \in \mathbb{C}^{m \times \ell}$ and $U_k \in \mathbb{C}^{m \times k}$ as $\theta_1, \ldots, \theta_k$. Similarly, denote the angles between $\widehat{V} \in \mathbb{C}^{n \times \ell}$ and $V_k \in \mathbb{C}^{n \times k}$ by $\nu_1, \ldots, \nu_k$. We are also interested in obtaining bounds for the canonical angles $\angle(U_k, \widehat{U}_k)$ and $\angle(V_k, \widehat{V}_k)$. To distinguish these angles from $\angle(U_k, \widehat{U})$ and $\angle(V_k, \widehat{V})$, we call them θ_j' and ν_j' for $j = 1, \ldots, k$.

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

THEOREM 1. *Let $\widehat{U}$ and $\widehat{V}$ be obtained from Algorithm 1. With Assumption 1, the canonical angles θ_j and ν_j satisfy*

$$\sin\theta_j \leq \frac{\gamma_j^{2q+1} \left\|\Omega_2\Omega_1^\dagger\right\|_2}{\sqrt{1 + \gamma_j^{4q+2} \left\|\Omega_2\Omega_1^\dagger\right\|_2^2}} \qquad \sin\nu_j \leq \frac{\gamma_j^{2q+2} \left\|\Omega_2\Omega_1^\dagger\right\|_2}{\sqrt{1 + \gamma_j^{4q+4} \left\|\Omega_2\Omega_1^\dagger\right\|_2^2}}$$

for $j = 1, \ldots, k$.

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 equal to zero; that is, the randomized subspace iteration identifies the subspace exactly. On the other hand, when γ_k is very close to 1, the subspaces may not be well-defined and may be difficult to identify. In practice, it is highly desirable that $\gamma_k \ll 1$, so that the angles are captured accurately.

Second, the bounds for the canonical angles show explicit dependence on the singular value ratios γ_j. In particular, the canonical angles θ_j and ν_j converge to zero quadratically but at different rates depending on the singular value ratios. Specifically, the smaller the singular value ratio, smaller the canonical angles.

Third, the term $\|\Omega_2\Omega_1^\dagger\|_2$ can be written in terms of the right singular vector matrix V and the starting guess Ω as

$$\left\|\Omega_2\Omega_1^\dagger\right\|_2 = \left\|(V_\perp^*\Omega)(V_k^*\Omega)^\dagger\right\|_2.$$

When the columns of Ω are linearly independent, this quantity is nothing but the tangent of the largest canonical angle between $\mathcal{R}(V_k)$ 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, Ω contains the singular vectors in V_k. A discussion of the meaning and interpretation of this term is provided in [8, section 2.5]. In particular, when Ω is a Gaussian random matrix, $\|\Omega_2\Omega_1^\dagger\|_2$ is roughly on the order of $\sqrt{(n-k)k}$.

Fourth, the influence of $\|\Omega_2\Omega_1^\dagger\|_2$ is subdued by the singular value ratios γ_j^{2q+1}. With sufficiently large number of iterations q, the canonical angles are smaller than a user-defined tolerance. Rigorous bounds for the requisite number of iterations are provided in subsection 3.4.

Last, the bounds for the canonical angles θ_j are smaller than ν_j because the latter contains an additional power of γ_j. The reason for this higher accuracy is as follows: the columns of $\widehat{V}$ are the right singular vectors of Q^*A. Therefore, the multiplication step with Q amounts to an additional step of subspace iteration and gives the extra factor.

Remark 1. 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. With the same assumptions and notation as in Theorem 1, the relationship between the tangent and sine implies

$$\tan\theta_j \leq \gamma_j^{2q+1} \left\|\Omega_2\Omega_1^\dagger\right\|_2, \qquad \tan\nu_j \leq \gamma_j^{2q+2} \left\|\Omega_2\Omega_1^\dagger\right\|_2$$

for $j = 1, \ldots, k$. 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 the approximate singular vectors $\widehat{U}$ and $\widehat{V}$ for a matrix A be computed according to Algorithm* 1. *Under Assumption* 1, *for every unitarily invariant norm,*

(5)
$$\left\|\left\|\sin \angle(U_k, \widehat{U})\right\|\right\| \leq \gamma_k^{2q} \frac{\|\|\Sigma_\perp\|\|}{\sigma_k} \left\|\Omega_2 \Omega_1^\dagger\right\|_2,$$
$$\left\|\left\|\sin \angle(V_k, \widehat{V})\right\|\right\| \leq \gamma_k^{2q+1} \frac{\|\|\Sigma_\perp\|\|}{\sigma_k} \left\|\Omega_2 \Omega_1^\dagger\right\|_2.$$

The interpretation of this theorem is similar to that of Theorem 1. The connection between the two theorems follows from the identity $\sin \theta_k = \|\sin \angle(U_k, \widehat{U})\|_2$. If we specialize the result in Theorem 2 to the spectral norm, then it is clear that this result is weaker than the bound in Theorem 1.

3.2. Extraction of k-dimensional subspaces. In the previous subsection, the columns of $\widehat{U}$ and $\widehat{V}$ spanned $\ell = k + \rho$ dimensional subspaces. Many applications, however, require the extraction of k dimensional singular subspaces from the low-rank approximation $\widehat{A} \equiv QQ^*A$. One way to extract the appropriate subspaces is to first compute the optimal rank-k truncation of $\widehat{A}$, denoted by $\widehat{A}_k$. The singular vectors of $\widehat{A}_k$, denoted by $\widehat{U}_k$ and $\widehat{V}_k$, are then used instead of $\widehat{U}$ and $\widehat{V}$. See Algorithm 2 for details regarding implementation. The bounds derived in the previous subsection are not directly applicable since [26, Corollary 10] says

$$\theta_j \leq \theta'_j, \qquad \nu_j \leq \nu'_j, \qquad j = 1, \ldots, k.$$

To understand how much additional error is incurred during this extraction process, we present several results. The important conclusion of all these results is that the accuracy of the extracted subspaces of dimension k is comparable to the accuracy of the $k + \rho$ dimensional subspace provided the singular values are sufficiently well separated.

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 [25]. Let $A, \widehat{A}$ be two matrices of conformal dimensions. Assuming that

(6)
$$\zeta \equiv \sigma_k(A) - \sigma_{k+1}(\widehat{A}) > 0,$$

we have

(7)
$$\max\left\{\left\|\left\|\sin \angle(U_k, \widehat{U}_k)\right\|\right\|, \left\|\left\|\sin \angle(V_k, \widehat{V}_k)\right\|\right\|\right\} \leq \frac{\max\{\|\|E_{12}\|\|, \|\|E_{21}\|\|\}}{\zeta},$$

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

(8)
$$E_{12} = (I - \mathcal{P}_{\widehat{U}_k})(A - \widehat{A})\mathcal{P}_{V_k},$$
$$E_{21} = \mathcal{P}_{U_k}(A - \widehat{A})(I - \mathcal{P}_{\widehat{V}_k}).$$

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 $\mathsf{rank}\,(A) \geq k$ and let $\widehat{A}$ be the perturbed matrix with same dimensions. Suppose the singular value gap satisfies* (6). *Let $\widehat{A}_k = \widehat{U}_k \widehat{\Sigma}_k \widehat{V}_k^*$ be the truncated SVD of $\widehat{A}$. Then*

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

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

Our main result provides the following bounds for canonical angles between the exact and the approximate singular subspaces, when both the subspaces have the same dimension. The proof involves simplifying every term in (8).

THEOREM 4. *Let $\widehat{U}$ and $\widehat{V}$ be obtained from Algorithm* 1, *and matrices $\widehat{U}_k$ and $\widehat{V}_k$ from Algorithm* 2. *Under Assumption* 1, *the following hold:*
- *For every unitarily invariant norm*

$$\max\left\{\left\|\!\left\|\sin\angle(U_k, \widehat{U}_k)\right\|\!\right\|, \left\|\!\left\|\sin\angle(V_k, \widehat{V}_k)\right\|\!\right\|\right\} \leq \phi\frac{\gamma_k^{2q}}{1 - \gamma_k}\frac{\|\!\|\Sigma_\perp\|\!\|}{\sigma_k}\left\|\Omega_2\Omega_1^\dagger\right\|_2.$$

The factor ϕ takes different values depending on the specific norm used. For an arbitrary unitarily invariant norm, we have $\phi = \sqrt{2}$, whereas for the spectral and Frobenius norms, we have $\phi = 1$.
- *The canonical angles θ_j' and ν_j' satisfy*

$$\max\{\sin\theta_j', \sin\nu_j'\} \leq \gamma_j\frac{\gamma_k^{2q}}{1 - \gamma_k}\left\|\Omega_2\Omega_1^\dagger\right\|_2, \qquad j = 1, \ldots, k.$$

The interpretation of this theorem is as follows: (1) as the number of iterations q increases, the largest canonical angle converges to 0 quadratically, and (2) a larger singular value gap means that the subspace is computed more accurately. Comparing this result with Theorem 2, we see that the upper bound in Theorem 4 has additional factors which depend on the specific norm used. For an arbitrary unitarily invariant norm, there is an additional factor $\max\{1, \sqrt{2}\gamma_k\}/(1 - \gamma_k)$. For the spectral and Frobenius norms, the additional factor is $1/(1 - \gamma_k)$. Both factors are greater than 1, suggesting that the truncation process can introduce additional error. The additional factor is also independent of the number of iterations q, suggesting that it is a one-time price to be paid for the extraction process. The bound is devastating when $\gamma_k \approx 1$, but this also means that 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 $\widehat{U}$ and $\widehat{V}$ be computed according to Algorithm* 1. *With Assumption* 1, *we have the following inequalities:*

$$(9) \qquad \sin\angle(u_j, \widehat{U}) \leq \gamma_j^{2q+1}\left\|\Omega_2\Omega_1^\dagger\right\|_2, \qquad \sin\angle(v_j, \widehat{V}) \leq \gamma_j^{2q+2}\left\|\Omega_2\Omega_1^\dagger\right\|_2$$

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

$$(10) \qquad \max \left\{ \sin \angle(u_j, \hat{u}_j), \sin \angle(v_j, \hat{v}_j) \right\} \leq \sqrt{1 + 2\frac{\tilde{\gamma}^2}{\tilde{\delta}^2}} \, \gamma_j^{2q+1} \left\| \Omega_2 \Omega_1^\dagger \right\|_2.$$

Here, $\tilde{\gamma}^2 \equiv \|\Sigma_\perp\|_2^2 + \|\Sigma_\perp \Omega_2 \Omega_1^\dagger\|_2^2$ *and* $\tilde{\delta} \equiv \min\{\min_{\tilde{\sigma}_i \neq \tilde{\sigma}_j}\{|\sigma_j - \tilde{\sigma}_i|, \sigma_j\}\}$.

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 story similar 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 nondecreasing.

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 example, we refer to [19, Chapter 14]. Randomized subspace iteration has attracted a lot of attention in the last two decades, with a special emphasis on quantifying the influence of the starting guess Ω. In particular, recent research has focused on the choice of the distribution and the effect of the oversampling parameter ρ. The effect of randomized subspace iteration on the accuracy of singular vectors was studied in the context of spectral clustering in [6]. However, the authors made the rather strong assumption that $\Omega \in \mathbb{R}^{n \times k}$, which amounts to setting the oversampling parameter $\rho = 0$. This is a strong requirement since Assumption 1 now requires Ω_1 to be invertible. The authors were able to show (in our notation)

$$\left\| \sin \angle(U_k, \widehat{U}_k) \right\|_2 \leq \frac{\gamma_k^{2q+1} \left\| \Omega_2 \Omega_1^{-1} \right\|_2}{\sqrt{1 + \gamma_k^{4q+2} \left\| \Omega_2 \Omega_1^{-1} \right\|_2^2}}.$$

Notice that this bound coincides with Theorem 1 (for $\sin \theta_k$) when $\rho = 0$. Our results provide bounds for the right singular vectors as well as all the canonical angles.

Let us return to this assumption that $\mathsf{rank}\,(\Omega_1)$. When Ω is a standard Gaussian matrix, [21, Theorem 3.3] says

$$\|\Omega_1^{-1}\|_2 \leq \frac{2.35\sqrt{k}}{\delta}$$

with probability at least $1 - \delta$. For a small probability of failure $0 < \delta < 1$, this bound can be devastating. By contrast, if we let $\Omega_1 \in \mathbb{C}^{k \times (k+\rho)}$ with $\rho \geq 2$, and still suppose that Ω is a Gaussian random matrix. Then, with probability at least $1 - \delta$, [12, Proposition 10.4] says

$$\left\| \Omega_1^\dagger \right\|_2 \leq e \frac{\sqrt{k+\rho}}{\rho} \left(\frac{1}{\delta} \right)^{1/(\rho+1)}.$$

It is clear that when the random matrix is Gaussian, oversampling has an impact on the accuracy of the randomized subspace iteration. Specifically, the larger the oversampling, the more accurate the subspace.

Oversampling plays a bigger role for random matrices that have different distributions than Gaussian. When Ω is generated from the subsampled randomized Hadamard transform, or Rademacher distribution, a more aggressive form of oversampling $\ell \sim k \log k$ is necessary to ensure that $\mathsf{rank}\,(\Omega_1) = k$. Therefore, by allowing for oversampling, our bounds are applicable to starting guesses that are not restricted to 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 $\angle(U_k, \widehat{U})$ to $\angle(U_k, \widehat{U}_k)$. Our bounds for the canonical angles appear to be tighter than the result implied by [18, Corollary 1]. This may be because the analysis was applicable to arbitrary subspace projections, whereas ours is specialized to randomized subspace iteration; we do not go into a detailed comparison here. 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 $\mathsf{rank}\,(\Omega_1) = k$. In particular, Ω need not even be random and may be deterministic. However, more can be said about the random matrix Ω is drawn from a specific distribution.

In many applications, the matrix $\Omega \in \mathbb{R}^{n \times (k+\rho)}$ is taken to be the standard Gaussian random matrix. That is, the entries of Ω are independent and identically distributed $\mathcal{N}(0,1)$ random variables. Here we derive a few probabilistic results that provide insight into the accuracy of the subspaces. Let $\rho \geq 2$ and define the constant

$$(11) \qquad C_e = \sqrt{\frac{k}{\rho - 1}} + \frac{e\sqrt{(k+\rho)(n-k)}}{\rho}$$

and for $0 < \delta < 1$ define the constant

$$(12) \qquad C_d = \frac{e\sqrt{k+\rho}}{\rho+1} \left(\frac{2}{\delta}\right)^{1/(\rho+1)} \left(\sqrt{n-k} + \sqrt{k+\rho} + \sqrt{2\log\frac{2}{\delta}}\right).$$

THEOREM 6 (probabilistic bounds). *Let $\Omega \in \mathbb{R}^{n \times (k+\rho)}$ be a standard Gaussian random matrix with $\rho \geq 2$. Assume that the singular value ratio $\gamma_k < 1$. Let $\widehat{U}$ and $\widehat{V}$ be obtained from Algorithm 1. For $j = 1, \ldots, k$, the expected value of the canonical angles satisfies*

$$\mathbb{E}\left[\sin\theta_j\right] \leq \frac{\gamma_j^{2q+1}C_e}{\sqrt{1 + \gamma_j^{4q+2}C_e^2}}, \qquad \mathbb{E}\left[\sin\nu_j\right] \leq \frac{\gamma_j^{2q+2}C_e}{\sqrt{1 + \gamma_j^{4q+4}C_e^2}}.$$

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

$$\sin\theta_j \leq \frac{\gamma_{j+1}^{2q+1}C_d}{\sqrt{1 + \gamma_j^{4q+2}C_d^2}}, \qquad \sin\nu_j \leq \frac{\gamma_j^{2q+2}C_d}{\sqrt{1 + \gamma_j^{4q+4}C_d^2}}.$$

The main message of the theorem can be seen from the following bound on the number of subspace iterations q. Specifically, suppose $0 < \epsilon < 1$, and the number of subspace iterations q we take satisfies

$$q \geq \frac{1}{2}\left(\frac{\log \epsilon/C_e}{\log \gamma_k} - 1\right);$$

then $\mathbb{E} \sin \theta_j \leq \mathcal{O}(\epsilon^2)$ for $j = 1, \ldots, k$;

Several extensions of these results are possible. First, following the proof technique of Theorem 6, we can extend the probabilistic analysis to Theorems 2 and 5 as well. Second, following the strategy in [12], the probabilistic results can be extended to other distributions. However, we will not pursue these extensions here.

4. Low-rank approximation and singular values. In this section, we provide several structural bounds for the accuracy of the low-rank approximation and the accuracy of the 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 [11, 12, 27]. As 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 $\widehat{A} \in \mathbb{C}^{m \times n}$ be computed using Algorithm 1. Under Assumption 1, the following inequalities hold in every unitarily invariant norm:*

$$(13) \qquad \||(I - QQ^*)A\|| \leq \||\Sigma_\perp\|| + \gamma_k^{2q}\left\||\Sigma_\perp \Omega_2 \Omega_1^\dagger\right\||,$$

$$(14) \qquad \||(I - QQ^*)A_k\|| \leq \gamma_k^{2q}\left\||\Sigma_\perp \Omega_2 \Omega_1^\dagger\right\||.$$

Let $B = Q^*A$, and let B_k be its best rank-k approximation. If A is approximated using QB_k, then the error in the low-rank approximation is

$$(15) \qquad \||A - QB_k\|| \leq \left(1 + \frac{\sigma_1}{\sigma_k}\frac{\phi\gamma_k^{2q}}{1 - \gamma_k}\left\|\Omega_2\Omega_1^\dagger\right\|_2\right)\||\Sigma_\perp\||.$$

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 [27].

THEOREM 8. *Let $\widehat{A}$ be computed using Algorithm 1. Under Assumption 1, we have*

$$(16) \qquad \||(I - QQ^*)A\||_p^2 \leq \||\Sigma_\perp\||_p^2 + \gamma_k^{4q}\left\||\Sigma_\perp\Omega_2\Omega_1^\dagger\right\||_p^2.$$

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 $\widehat{A}$? 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 (20).

THEOREM 9. *Let $\widehat{A} = \widehat{U}\widehat{\Sigma}\widehat{V}^*$ be computed using Algorithm* 1. *Under Assumption* 1, *the approximate singular values $\sigma_j(\widehat{A})$ satisfy for $j = 1, \ldots, k$*

$$\sigma_j(A) \geq \sigma_j(\widehat{A}) \geq \frac{\sigma_j(A)}{\sqrt{1 + \gamma_j^{4q+2}\left\|\Omega_2\Omega_1^\dagger\right\|_2^2}}.$$

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.

Rather than quantify the accuracy of the individual singular values, the next results are of the Hoffman–Wielandt type and account for all the singular values together. Define the two matrices of conformal sizes

$$\Sigma = \begin{bmatrix} \Sigma_k & \\ & \Sigma_\perp \end{bmatrix}, \qquad \Sigma' = \begin{bmatrix} \widehat{\Sigma} & \\ & 0 \end{bmatrix}.$$

Under Assumption 1, the error in the singular values satisfies

$$(17) \qquad \|\|\Sigma - \Sigma'\|\| \leq \|\|\Sigma_\perp\|\| + \gamma_k^{2q}\|\|\Sigma_\perp\Omega_2\Omega_1^\dagger\|\|.$$

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

$$(18) \qquad \|\|\Sigma - \Sigma'\|\|_p \leq \sqrt{\|\|\Sigma_\perp\|\|_p^2 + \gamma_k^{4q}\|\|\Sigma_\perp\Omega_2\Omega_1^\dagger\|\|_p^2}.$$

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 semidefinite 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,

$$SMS^* \preceq SNS^* \qquad \forall\, S \in \mathbb{C}^{m \times n}.$$

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 semidefinite, then $M^{1/2} \preceq N^{1/2}$ [4, Proposition V.1.8] and $(I + N)^{-1} \preceq (I + M)^{-1}$.

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],

$$(19) \qquad \sigma_{i+j-1}(A + B) \leq \sigma_i(A) + \sigma_j(B),$$

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

$$(20) \qquad \sigma_{i+j-1}(AB^*) \leq \sigma_i(A)\sigma_j(B).$$

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 $i = 1, \ldots, \min\{m, n\}$.

Unitarily invariant norms. 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 submultiplicativity, will be useful [4, Equation IV.40]:

$$\|\|ABC\|\| \leq \|A\|_2 \, \|C\|_2 \, \|\|B\|\|.$$

We will need the following lemma.

LEMMA 10. *Let $A, B, D \in \mathbb{C}^{n \times n}$ such that A, B Hermitian and $0 \preceq A \preceq B$; then*

$$\left\|\left\|(D^*AD)^{1/2}\right\|\right\| \leq \left\|\left\|(D^*BD)^{1/2}\right\|\right\|.$$

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 IV.2.2], the advertised inequality is true for every unitarily invariant norm. $\square$

5.1. Proofs of subsection 3.1 theorems.

Proof of 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 eigenvalue decomposition.* We compute the thin SVD of $(I - \mathcal{P}_{\widehat{U}})U_k = KS_UG^*$. The matrix

$$S_U = \mathsf{diag}\left(\sin \theta_k, \ldots, \sin \theta_1\right) \in \mathbb{R}^{k \times k}$$

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

(21) $$GS_U^2G^* = U_k^*(I - \mathcal{P}_{\widehat{U}})U_k.$$

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

$$U^*Y = \begin{bmatrix} \Sigma_k^{2q+1} & \\ & (\Sigma_\perp \Sigma_\perp^\top)^q \Sigma_\perp \end{bmatrix} (V^*\Omega) = \begin{bmatrix} \Sigma_k^{2q+1}\Omega_1 \\ (\Sigma_\perp \Sigma_\perp^\top)^q \Sigma_\perp \Omega_2 \end{bmatrix},$$

where from (2), $\Omega_1 = V_k^*\Omega$ and $\Omega_2 = V_\perp^*\Omega$. Next, by Assumption 1, Ω_1 has full row rank and therefore it has a right multiplicative inverse. Define

$$Z \equiv U^*Y\Omega_1^\dagger \Sigma_k^{-(2q+1)} = \begin{bmatrix} I \\ F \end{bmatrix}, \qquad F \equiv (\Sigma_\perp \Sigma_\perp^\top)^q \Sigma_\perp \Omega_2 \Omega_1^\dagger \Sigma_k^{-(2q+1)}.$$

Recall that $Y = QR$ is the thin-QR factorization of Y. Let Q_1R_1 be the thin-QR factorization of $R\Omega_1^\dagger \Sigma_k^{-(2q+1)}$; here, $Q_1 \in \mathbb{C}^{(k+\rho)\times k}, R_1 \in \mathbb{C}^{k \times k}$.

From $Q_1Q_1^* \preceq I$, the conjugation rule implies

$$\mathcal{P}_Z = U^*QQ_1Q_1^*Q^*U \preceq U^*QQ^*U = \mathcal{P}_{U^*Q}.$$

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

$$(22) \qquad \mathcal{P}_Z \preceq \mathcal{P}_{U^*\widehat{U}}, \qquad I - \mathcal{P}_{U^*\widehat{U}} \preceq I - \mathcal{P}_Z.$$

Plug $UU^* = I$ into (21), and use (22) to obtain

$$U_k^*(I - \mathcal{P}_{\widehat{U}})U_k = U_k^*U(I - \mathcal{P}_{U^*\widehat{U}})U^*U_k \preceq \begin{bmatrix} I & 0 \end{bmatrix}(I - \mathcal{P}_Z)\begin{bmatrix} I \\ 0 \end{bmatrix}.$$

3. *Simplifying $\mathcal{P}_Z$.* Since $\mathcal{P}_Z = ZZ^\dagger$, 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

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

$$(23) \qquad\qquad\qquad\qquad = F^*F(I + F^*F)^{-1} \equiv H.$$

Note that H is positive semidefinite. To summarize the story so far, $GS_U^2G^* \preceq H$.

4. *Applying singular value inequalities.* A straightforward SVD argument shows that the jth 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 (20) imply

$$\sigma_j(F) \leq \sigma_1(\Sigma_\perp\Sigma_\perp^\top)^q\Sigma_\perp\Omega_2\Omega_1^\dagger)\sigma_j(\Sigma_k^{-2q-1}) \leq \left(\frac{\sigma_{k+1}}{\sigma_{k-j+1}}\right)^{2q+1}\left\|\Omega_2\Omega_1^\dagger\right\|_2.$$

Plugging this inequality into $\sigma_j(H)$

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

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

Bounds for $\sin\nu_j$. Let $GS_V^2G^*$ be the eigenvalue decomposition of $V_k^*(I - \mathcal{P}_{\widehat{V}})V_k$. Note that the diagonals of S_V are the sine of the canonical angles $\angle(V_k, \widehat{V})$. Since $\widehat{V}$ is obtained from the thin SVD of A^*Q, $\mathcal{R}(A^*Q) = \mathcal{R}(\widehat{V})$ and $\mathcal{P}_{\widehat{V}} = \mathcal{P}_{A^*Q}$, since an orthogonal projection matrix is uniquely determined by the range. Next, consider $\widehat{Z}$ defined as

$$(24) \qquad \widehat{Z} \equiv \Sigma^\top U^*Y\Omega_1^\dagger\Sigma_k^{-2q-2} = \begin{bmatrix} I \\ \widehat{F} \end{bmatrix}, \qquad \widehat{F} \equiv (\Sigma_\perp^\top\Sigma_\perp)^{q+1}\Omega_2\Omega_1^\dagger\Sigma_k^{-2q-2};$$

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

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

Using an argument similar to (22), we obtain

$$V_k^* V (I - \mathcal{P}_{\widehat{V}}) V^* V_k \preceq V_k^* V (I - \mathcal{P}_{\widehat{Z}}) V^* V_k = \begin{bmatrix} I & 0 \end{bmatrix} (I - \mathcal{P}_{\widehat{Z}}) \begin{bmatrix} I \\ 0 \end{bmatrix}.$$

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

Proof of Theorem 2. With the notation of Theorem 1, we follow steps 1–3 of the proof to obtain

$$G S_U^2 G^* \preceq H \preceq F^* F.$$

Since the square root preserves partial ordering, this implies $G S_U G^* \preceq (F^* F)^{1/2}$. Note that $(F^* F)^{1/2}$ and F have the same nonzero singular values. Therefore,

$$\left\| \left| \sin \angle (U_k, \widehat{U}) \right| \right\| \le \left\| \left| (F^* F)^{1/2} \right| \right\| = \|F\|.$$

By using strong submultiplicativity of the unitarily invariant norm, we have

$$\left\| \left| \sin \angle (U_k, \widehat{U}) \right| \right\| \le \gamma_k^{2q} \left\| \Omega_2 \Omega_1^\dagger \right\|_2 \frac{\| \Sigma_\perp \|}{\sigma_k}. \qquad \square$$

5.2. Proofs of subsection 3.2 theorems.

Proof of Theorem 3. Let $X = (I - \mathcal{P}_{\widehat{U}_k}) \mathcal{P}_{U_k}$ and $Y = (I - \mathcal{P}_{\widehat{V}_k}) \mathcal{P}_{V_k}$. In decreasing order, the singular values of X and Y are $\{\sin \theta'_j\}_{j=1}^k$ and $\{\sin \nu'_j\}_{j=1}^k$, respectively. Let $B \equiv \widehat{A} - \widehat{A}_k$. First, we observe that

$$\begin{aligned} E_{12} &= (I - \mathcal{P}_{\widehat{U}_k})(A - \widehat{A}) \mathcal{P}_{V_k} \\ &= (I - \mathcal{P}_{\widehat{U}_k}) A_k - (I - \mathcal{P}_{\widehat{U}_k}) \widehat{A} \mathcal{P}_{V_k} \\ &= (I - \mathcal{P}_{\widehat{U}_k}) \mathcal{P}_{U_k} A_k - (\widehat{A} - \widehat{A}_k) \mathcal{P}_{V_k} \\ &= X A_k - B (I - \mathcal{P}_{\widehat{V}_k}) \mathcal{P}_{V_k} = X A_k - B Y. \end{aligned}$$

A similar calculation shows that $E_{21} = X^* B - A_k Y^*$. From the first relation, since $\mathsf{rank}(A) \ge k$, we have

$$X A_k A_k^\dagger = (E_{12} + B Y) A_k^\dagger.$$

But $A_k A_k^\dagger = \mathcal{P}_{U_k}$ and $X \mathcal{P}_{U_k} = X$. Applying (20), we have

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

A similar argument gives

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

Combining these relations

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

Recognize that $\|B\|_2 = \sigma_{k+1}(\widehat{A})$. Applying (7) in the spectral norm simplifies the expression since

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

Therefore,

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

Now $\sigma_j(X) = \sin\theta'_{k-j+1}$ and $\sigma_j(Y) = \sin\nu'_{k-j+1}$. Rename $j \leftarrow k - j + 1$ to finish. $\square$

Proof of Theorem 4. We tackle each case independently.

Unitarily invariant norms. Our proof involves simplifying each term in (7), and (8) and has several steps.

1. *Simplifying the gap.* Recall $\zeta = \sigma_k(A) - \sigma_{k+1}(\widehat{A})$ and $\widehat{A} = QQ^*A$. From the first part of Theorem 9

$$\zeta = \sigma_k(A) - \sigma_{k+1}(\widehat{A}) \geq \sigma_k(A) - \sigma_{k+1}(A).$$

2. *Simplifying $\|E_{12}\|$.* First observe that $A\mathcal{P}_{V_k} = A_k$. So

$$E_{12} = (I - \mathcal{P}_{\widehat{U}_k})(I - QQ^*)A\mathcal{P}_{V_k} = (I - \mathcal{P}_{\widehat{U}_k})(I - QQ^*)A_k.$$

Then applying (14) along with submultiplicativity gives

$$\||E_{12}\|| \leq \||(I - QQ^*)A_k\|| \leq \gamma_k^{2q}\||\Sigma_\perp\Omega_2\Omega_1^\dagger\|| \leq \gamma_k^{2q}\||\Sigma_\perp\||\left\|\Omega_2\Omega_1^\dagger\right\|_2.$$

3. *Simplifying $\|E_{21}\|$.* First, $E_{21} = \mathcal{P}_{U_k}(I-QQ^*)A\mathcal{P}_{\widehat{V}_k}$, and since $\|\mathcal{P}_{U_k}(I-QQ^*)\|_2 = \|\sin\angle(U_k, \widehat{U})\|_2$,

$$\||E_{21}\|| \leq \left\|\sin\angle(U_k, \widehat{U})\right\|_2 \||(I - QQ^*)A\||,$$

because of strong submultiplicativity. Applying Theorem 1 and (14)

$$\||E_{21}\|| \leq \frac{\gamma_k^{2q+1}\left\|\Omega_2\Omega_1^\dagger\right\|_2}{\sqrt{1 + \gamma_k^{4q+2}\|\Omega_2\Omega_1^\dagger\|_2^2}}\left(1 + \gamma_k^{2q}\left\|\Omega_2\Omega_1^\dagger\right\|_2\right)\||\Sigma_\perp\||.$$

Let $\beta = \gamma_k^{2q}\|\Omega_2\Omega_1^\dagger\|_2$. 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, $\||E_{21}\|| \leq \sqrt{2}\gamma_k^{2q}\||\Sigma_\perp\||\|\Omega_2\Omega_1^\dagger\|_2$.

4. *Putting everything together.* Plugging in the intermediate quantities into (7), we have

$$\max\left\{\left\|\|\sin\angle(U_k, \widehat{U}_k)\right\|\|, \left\|\|\sin\angle(V_k, \widehat{V}_k)\right\|\|\right\} \leq \sqrt{2}\gamma_k^{2q}\left\|\Omega_2\Omega_1^\dagger\right\|_2 \frac{\||\Sigma_\perp\||}{\sigma_k - \sigma_{k+1}}.$$

Dividing the numerator and denominator by σ_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 proof for unitarily invariant norms. For the third step, using (14)

$$\|E_{21}\|_\xi \leq \frac{\gamma_k^{2q+1}\|\Omega_2\Omega_1^\dagger\|_2}{\sqrt{1 + \gamma_k^{4q+2}\|\Omega_2\Omega_1^\dagger\|_2^2}}\|\Sigma_\perp\|_\xi\sqrt{1 + \gamma_k^{4q}\left\|\Omega_2\Omega_1^\dagger\right\|_2^2}.$$

With β defined as before, since $\gamma_k < 1$, $\sqrt{\gamma_k^2 + \gamma_k^2 \beta^2}/\sqrt{1 + \gamma_k^2 \beta^2} \leq 1$. Therefore,

$$\|E_{21}\|_\xi \leq \gamma_k^{2q} \|\Sigma_\perp\|_\xi \left\|\Omega_2 \Omega_1^\dagger\right\|_2.$$

The rest of the proof is the same.

Canonical angles. The proof combines Theorem 3 with the above analysis for the spectral norm. The right-hand side contains the term

$$\max\left\{\left\|\sin\angle(U_k, \widehat{U}_k)\right\|_2, \left\|\sin\angle(V_k, \widehat{V}_k)\right\|_2\right\}.$$

The rest of the proof involves some simple manipulations. □

Proof of Theorem 5. We first address (9). Following the steps of the proof of Theorem 1, we have

$$\sin^2\angle(u_j, \widehat{U}) = u_j^* U(I - \mathcal{P}_{U^*Q})U^* u_j \preceq \begin{bmatrix} e_j^\top & 0 \end{bmatrix} (I - \mathcal{P}_Z) \begin{bmatrix} e_j \\ 0 \end{bmatrix},$$

where e_j is the jth column of the $k \times k$ identity matrix. Therefore, we have $\sin^2\angle(u_j, \widehat{U}) \leq e_j^\top H e_j$, where H was defined in (23). The inequality $H \preceq F^* F$ implies

$$\sin^2\angle(u_j, \widehat{U}) \leq \sigma_j^{-4q-2} \left\|(\Sigma_\perp \Sigma_\perp^\top)^q \Sigma_\perp (\Omega_2 \Omega_1^\dagger) e_j\right\|_2^2$$
$$\leq \gamma_j^{4q+2} \left\|\Omega_2 \Omega_1^\dagger\right\|_2^2.$$

Taking square roots on both sides gives the desired results. The strategy for bounding the canonical angles $\sin\angle(v_j, \widehat{V})$ is very similar and will be omitted.

We now address (10), which is a straightforward application of [13, Theorem 2.5]. Let $\mathcal{P}_\mathcal{U} = QQ^*$ and $\mathcal{P}_\mathcal{V} = I$. Then, in our notation, this result takes the form

$$\max\{\sin\angle(u_j, \widehat{u}_j), \sin\angle(v_j, \widehat{v}_j)\} \leq \sqrt{1 + 2\frac{\tilde{\gamma}'^2}{\tilde{\delta}^2}} \max\left\{\sin\angle(u_j, \widehat{U}), \sin\angle(v_j, I)\right\},$$

where $\tilde{\gamma}' = \max\{0, \|(I - QQ^*)A\|_2\}$ and $\tilde{\delta}$ is as defined in the statement of the theorem. Theorem 8 for the spectral norm implies $\tilde{\gamma}' \leq \tilde{\gamma}$, whereas Theorem 5 implies

$$\max\left\{\sin\angle(u_j, \widehat{U}), \sin\angle(v_j, I)\right\} \leq \gamma_j^{2q} \left\|\Omega_2 \Omega_1^\dagger\right\|_2.$$

Plug in the intermediate steps to obtain the desired bound. □

Proof of Theorem 6. In Theorem 1, bounds for $\|\Omega_2 \Omega_1^\dagger\|_2$ are available in the literature. From the proof of [12, Theorem 10.6] we find the inequality

$$\mathbb{E}\left\|\Omega_2 \Omega_1^\dagger\right\|_2 \leq C_e,$$

where the constant C_e was defined in (11). 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 that $\|\Omega_2 \Omega_1^\dagger\|_2 \leq C_d$ with a probability at least $1 - \delta$. Here, C_d was defined in (12). Plug into Theorem 1 to obtain the desired bounds. □

5.3. Proofs of section 4 theorems.

Proof of Theorem 7. *Proof of* (13). Using the unitary invariance of the norms

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

We use (22) combined with Lemma 10 to obtain

$$\left\||(\Sigma^\top(I - \mathcal{P}_{U^*Q})\Sigma)^{1/2}\right\|| \leq \left\||(\Sigma^\top(I - \mathcal{P}_Z)\Sigma)^{1/2}\right\||.$$

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

$$(25) \qquad \Sigma^\top(I - \mathcal{P}_Z)\Sigma = \begin{bmatrix} \Sigma_k M_1 \Sigma_k & * \\ * & \Sigma_\perp^\top M_2 \Sigma_\perp \end{bmatrix}.$$

The square root function is concave on $[0, \infty)$ and $\Sigma^\top(I - \mathcal{P}_Z)\Sigma$ is positive semidefinite. Therefore, an extension to Rotfel'd's theorem says [16, Theorem 2.1]

$$\left\||(\Sigma^\top(I - \mathcal{P}_Z)\Sigma)^{1/2}\right\|| \leq \left\||(\Sigma_k^\top M_1 \Sigma_k)^{1/2}\right\|| + \left\||(\Sigma_\perp^\top M_2 \Sigma_\perp)^{1/2}\right\||.$$

Using the inequalities $M_1 \preceq F^*F$ and $M_2 \preceq I$, along with Lemma 10, gives

$$(26) \qquad \begin{aligned} \||(I - \mathcal{P}_Q)A\|| &\leq \left\||(\Sigma_k F^*F \Sigma_k)^{1/2}\right\|| + \left\||(\Sigma_\perp \Sigma_\perp)^{1/2}\right\|| \\ &\leq \||F\Sigma_k\|| + \||\Sigma_\perp\||. \end{aligned}$$

Use $F\Sigma_k = (\Sigma_\perp \Sigma_\perp)^q \Sigma_\perp \Omega_2 \Omega_1^\dagger \Sigma_k^{-2q}$ and the submultiplicativity to obtain the advertised bounds.

Proof of (14). The proof for (14) is similar and is omitted. The main observation is that A_k has only k nonzero singular values.

Proof of (15). We follow the strategy in [7, section 3.3]. Recall that B_k is the best rank-k approximation to $B = Q^*A$. With the notation in Algorithm 2, note that

$$QB_k = Q\widehat{U}_{B,k}\widehat{U}_{B,k}^*B = \widehat{U}_k\widehat{U}_k^*A = \mathcal{P}_{\widehat{U}_k}A,$$

and the triangle inequality gives

$$\left\||(I - \mathcal{P}_{\widehat{U}_k})A\right\|| \leq \left\||(I - \mathcal{P}_{\widehat{U}_k})A_k\right\|| + \left\||(I - \mathcal{P}_{\widehat{U}_k})A_\perp\right\||.$$

Since $A_k = \mathcal{P}_{U_k}A_k$, applying strong submultiplicativity

$$\left\||(I - \mathcal{P}_{\widehat{U}_k})A\right\|| \leq \left\||(I - \mathcal{P}_{\widehat{U}_k})\mathcal{P}_{U_k}\right\|| \|A_k\|_2 + \||A_\perp\||.$$

We recognize that $\||(I - \mathcal{P}_{\widehat{U}_k})\mathcal{P}_{U_k}\|| = \||\sin\angle(U_k, \widehat{U}_k)\||$, apply Theorem 4 to complete the proof. □

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

$$\||(I - QQ^*)A\||_p^2 = \||A^*(I - QQ^*)A\||_{p/2} = \left\||\Sigma^\top(I - \mathcal{P}_{U^*Q})\Sigma\right\||_{p/2}.$$

The first equality holds only for $p \geq 2$, whereas the last equality follows because of the unitary invariance. As in the proof of Theorem 1, we have

$$Z = U^* Y \Omega_1^\dagger \Sigma_k^{-(2q+1)}, \qquad F = (\Sigma_\perp \Sigma_\perp^\top)^q \Sigma_\perp \Omega_2 \Omega_1^\dagger \Sigma_k^{-(2q+1)}.$$

The use of (22) and Lemma 10 ensures

$$\left\| \Sigma^\top (I - \mathcal{P}_{U^*Q}) \Sigma \right\|_{p/2} \leq \left\| \Sigma^\top (I - \mathcal{P}_Z) \Sigma \right\|_{p/2}.$$

We apply [16, Theorem 2.1] to (25) with $f(t) = t$ to obtain

$$\begin{aligned}
\left\| \Sigma^\top (I - \mathcal{P}_Z) \Sigma \right\|_{p/2} &\leq \left\| \Sigma_k M_1 \Sigma_k \right\|_{p/2} + \left\| \Sigma_\perp^\top M_2 \Sigma_\perp \right\|_{p/2} \\
&\leq \left\| \Sigma_k F^* F \Sigma_k \right\|_{p/2} + \left\| \Sigma_\perp^\top \Sigma_\perp \right\|_{p/2} \\
&= \left\| F \Sigma_k \right\|_p^2 + \left\| \Sigma_\perp \right\|_p^2.
\end{aligned}$$

We have used $M_1 \preceq F^* F$ and $M_2 \preceq I$. The rest of the proof is similar to that of Theorem 7. □

Proof of Theorem 9. The proof makes heavy use of the partial ordering which was reviewed at the start of section 5. From the inequality $I \succeq QQ^*$, the conjugation rule gives

$$A^* A \succeq A^* QQ^* A.$$

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

For the second inequality consider again $A^* QQ^* A$. With the aid of (22)

$$(27) \qquad A^* QQ^* A = V \Sigma^\top \mathcal{P}_{U^*Q} \Sigma V^* \succeq V \Sigma^\top \mathcal{P}_Z \Sigma V^*.$$

Therefore, $\lambda_j(A^* QQ^* A) \geq \lambda_j(V \Sigma^\top \mathcal{P}_Z \Sigma V^*)$ for $j = 1, \ldots, k$. Since $V \Sigma^\top \mathcal{P}_Z \Sigma V^*$ and $\Sigma^\top \mathcal{P}_Z \Sigma$ are similar, they share the same eigenvalues. It can be readily shown that

$$\Sigma^\top \mathcal{P}_Z \Sigma = \begin{bmatrix} \Sigma_k (I + F^* F)^{-1} \Sigma_k & * \\ * & * \end{bmatrix}.$$

For $j = 1, \ldots, k$, the eigenvalues of $A^* Q^* QA$ satisfy

$$(28) \qquad \lambda_j(A^* QQ^* A) \geq \lambda_j(V \Sigma^\top \mathcal{P}_Z \Sigma V^*) \geq \lambda_j(\Sigma_k (I + F^* F)^{-1} \Sigma_k).$$

The second inequality follows from the Cauchy interlacing theorem [19, section 10-1]. Applying the properties of partial ordering, we obtain

$$F^* F \preceq \sigma_{k+1}^{4q+2} \left\| \Omega_2 \Omega_1^\dagger \right\|_2^2 \Sigma_k^{-(4q+2)} = \left\| \Omega_2 \Omega_1^\dagger \right\|_2^2 \Gamma_k^{4q+2},$$

where $\Gamma_k = \mathsf{diag}\,(\gamma_1, \ldots, \gamma_k)$ is a diagonal matrix with the singular value gaps. Furthermore,

$$\Sigma_k(I + F^*F)^{-1}\Sigma_k \succeq \Sigma_k \left(I + \left\|\Omega_2\Omega_1^\dagger\right\|_2^2 \Gamma_k^{4q+2}\right)^{-1}\Sigma_k.$$

The diagonal matrix on the right-hand side has its singular values on the diagonals; this fact, combined with (28), gives for $j = 1, \ldots, k$

$$\sigma_j^2(Q^*A) = \lambda_j(A^*QQ^*A) \geq \lambda_j(\Sigma_k(I + F^*F)^{-1}\Sigma_k) \geq \frac{\sigma_j^2(A)}{1 + \left\|\Omega_2\Omega_1^\dagger\right\|_2^2 \gamma_j^{4q+2}}.$$

Taking square roots, we obtain the desired result. $\qquad\square$

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

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

where $x_j \in \mathbb{R}^{3000}$ and $y_j \in \mathbb{R}^{300}$ are sparse random vectors with nonnegative entries generated using the MATLAB commands `sprand(3000,1,0.025)` and `sprand(300,1,0.025)`, respectively. The formula above is not an SVD, since the vectors do not form an orthonormal set. 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

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

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), $\gamma_n = 1$.
3. Low-rank plus decay. The matrices take the form

$$A = U\mathsf{diag}\,(\underbrace{1, 1, \ldots, 1}_{r}, 2^{-d}, 3^{-d} \ldots, (n - r + 1)^{-d})V^*.$$
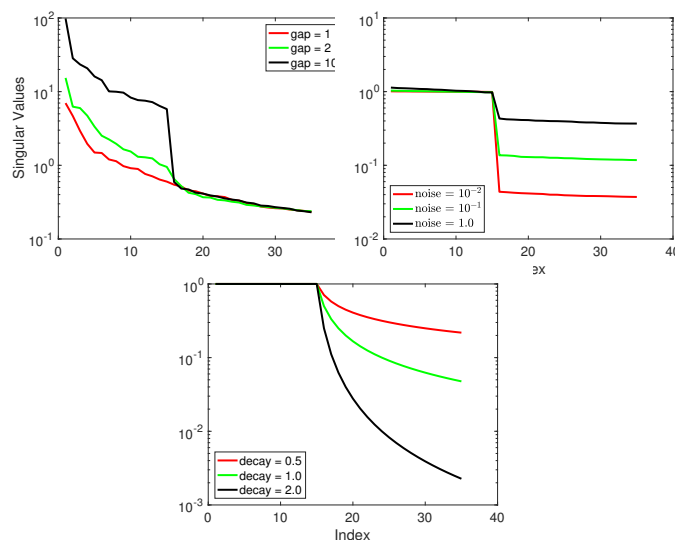
FIG. 1. *Singular value of the matrices from the (left) "controlled gap" example, (right) "low-rank plus noise" example, (below) "low-rank plus decay" example.*

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 [22], whereas the second and third examples are drawn from [24]. In all the examples, the random matrices were fixed by setting the random seed and we set the parameter $r = 15$. 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 nine test matrices in subsection 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 Ω was taken to be a random Gaussian matrix.

6.2.1. No extraction. We plot the canonical angles $\sin \angle(U_k, \widehat{U})$ 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 general 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(c), in which there is a large gap between singular values 15 and 16. Similar observations can be made in the other figures.
- As the decay rate of the singular values increases, the corresponding canonical angles become smaller.
- In most figures the bounds are qualitatively informative, but in some figures, the bounds are also quantitatively accurate (e.g., GapLarge).
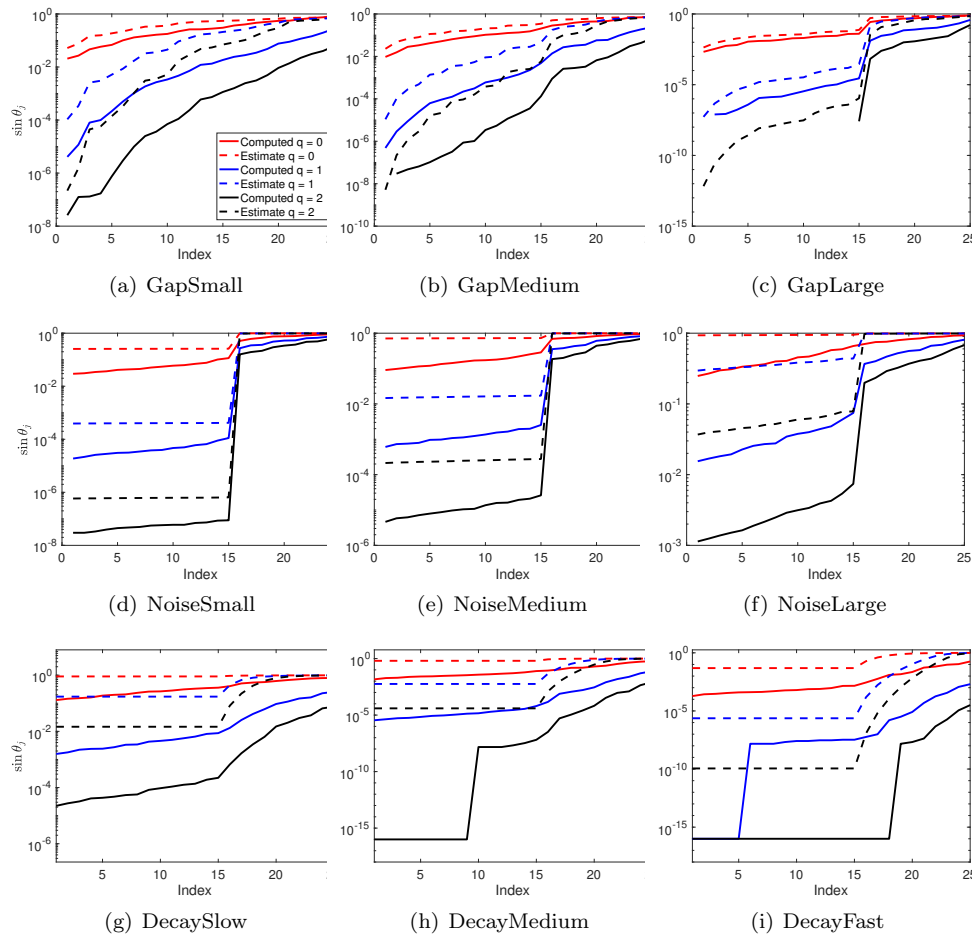- Similar results were observed for $\sin \angle(V_k, \widehat{V})$ and, therefore, are omitted.

FIG. 2. *Plots of* $\sin \theta_j$ *for* $j = 1, \ldots, k$. *The test matrices were described in subsection* 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 values, the dashed lines correspond to bounds obtained using Theorem* 1. *The parameter* q *corresponds to the number of subspace iterations.*

We now make observations specific to the test examples:

1. Gap examples. The computed canonical angles decrease as the gap increases, and with more iterations. The test matrices (GapMedium and GapLarge) have both a decay in the singular values and a prominent singular value gap between indices 15 and 16. These matrices satisfy the assumptions of our analysis, and therefore the bounds can be expected to be good. We see that as the size of the gap increases, the bounds become more accurate in accordance with Theorem 1. GapSmall has decay in the singular values but no special singular value gap. Even in this case, the bounds are qualitatively good.

2. Noise examples. NoiseSmall is close to a low-rank matrix and there is a large singular value gap at index 15. For this example, the bounds are qualitatively good. As the level of noise increases, the gap decreases and therefore, the computed angles increase, as predicted by Theorem 1. The bounds are unin-
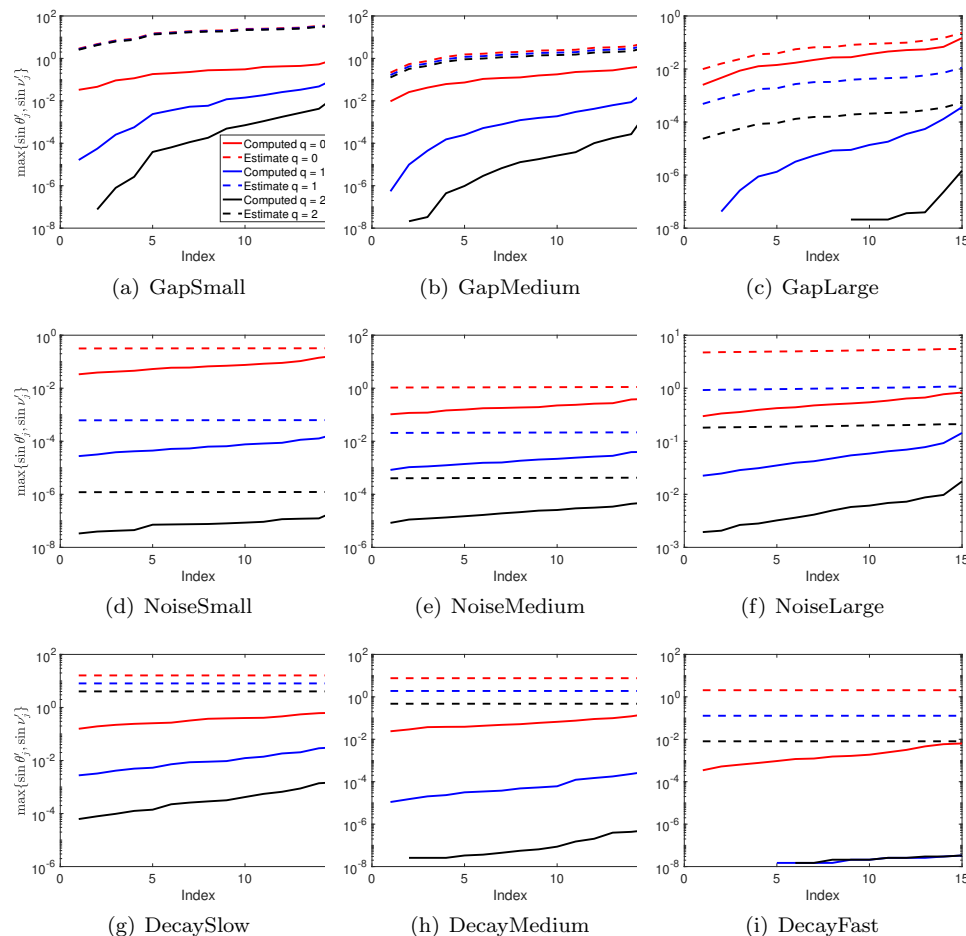
FIG. 3. *Plots of* $\max\{\sin\theta_j', \sin\nu_j'\}$ *for* $j = 1, \ldots, k$. *The test matrices were described in subsection* 6.1. *The target rank* $k = 15$ *and an oversampling parameter of* 20 *was chosen for all the experiments. The solid lines correspond to the computed values, the dashed lines correspond to bounds obtained using Theorem* 4. *The parameter* q *corresponds to the number of subspace iterations.*

    formative for $q = 0$ but qualitatively good for $q = 1$ and 2. Compared to the Gap examples, the bounds are not as sharp since there is very little decay in the singular values.

3. Decay examples. In these examples, the singular values decay beyond index 15 but there is no prominent gap. As the rate of decay increases, in general, the canonical angles decrease. It is also seen that the bounds are qualitatively accurate (except for $q = 0$).

    6.2.2. Extraction step. Our next experiment tests the effect of the extraction step on the accuracy of the canonical angles. We now compute $\sin\theta_j'$ and $\sin\nu_j'$ for the test matrices described in subsection 6.1. In Figure 3, 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 in the matrices. We make the following general observations:

- The extraction step did not significantly affect the canonical angles and the accuracy is comparable to Figure 2. The subspaces are more accurate as the number of iterations increases, and if there is a large singular value gap at index j, then the canonical angles with index $j' < j$ are captured accurately.
- Although the canonical angles are small, compared to Theorem 1, the bounds in Theorem 4 are not as accurate. One reason is that the upper bounds in Theorem 1 are at most 1, but the bounds in Theorem 4 are allowed to be greater than 1. Furthermore, the bound in Theorem 4 has the factor $1/(1 - \gamma_k)$ in the denominator, which can be quite large when there is a small singular value gap. It may be possible to derive better bounds, 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.

We now make observations specific to the test examples:

1. Gap examples. The behavior of the computed canonical angles is very similar to that without the extraction step. In general, the angles decrease as the gap increases. When the parameter gap is small, the singular value ratio γ_k is large, and $(1 - \gamma_k)^{-1}$ is small. This explains why the bounds are bad for GapSmall and GapMedium and show little improvement with more subspace iterations. Only for the GapLarge example with $q = 0$ are the bounds qualitatively good.
2. Noise examples. The computed canonical angles decrease as the noise decreases. In all three examples, the bounds are qualitatively good. The bounds are better for NoiseSmall and NoiseMedium because the singular value gap between indices 15 and 16 is bigger than that for NoiseLarge.
3. Decay examples. The computed canonical angles become smaller as the decay of the singular values increases. In these examples, there is no prominent gap, so the bounds don't capture the behavior well. However, the computed angles are small, and the subspace is accurate.

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. In Figure 4, the computed singular values are plotted against the upper and lower bounds. We make the following general observations:

- For the large singular values, both the upper and lower bounds are qualitatively good for all the examples that we tested.
- As the number of iterations increases, the singular values are computed more accurately and are close to the upper bounds (the exact singular values). However, for indices close to the target rank, the lower bounds are not tight. The bounds get tighter as the number of iterations q increases.
- The bounds for the singular values quantitatively better than the bounds for the canonical angles.

We now make observations specific to the test examples:

1. GapSmall. In these examples, the large singular values are captured accurately. As the number of iterations increases, both the lower bound and the approximate singular values approach the true singular values (upper bound). For GapMedium and GapLarge, the bounds were much more accurate.
2. NoiseMedium. There is a qualitatively different behavior before and after indices 15 and 16. The upper and lower bounds are tight before index 15, but only the upper bound is tight after index 16. The lower bound significantly underpredicts the singular values.
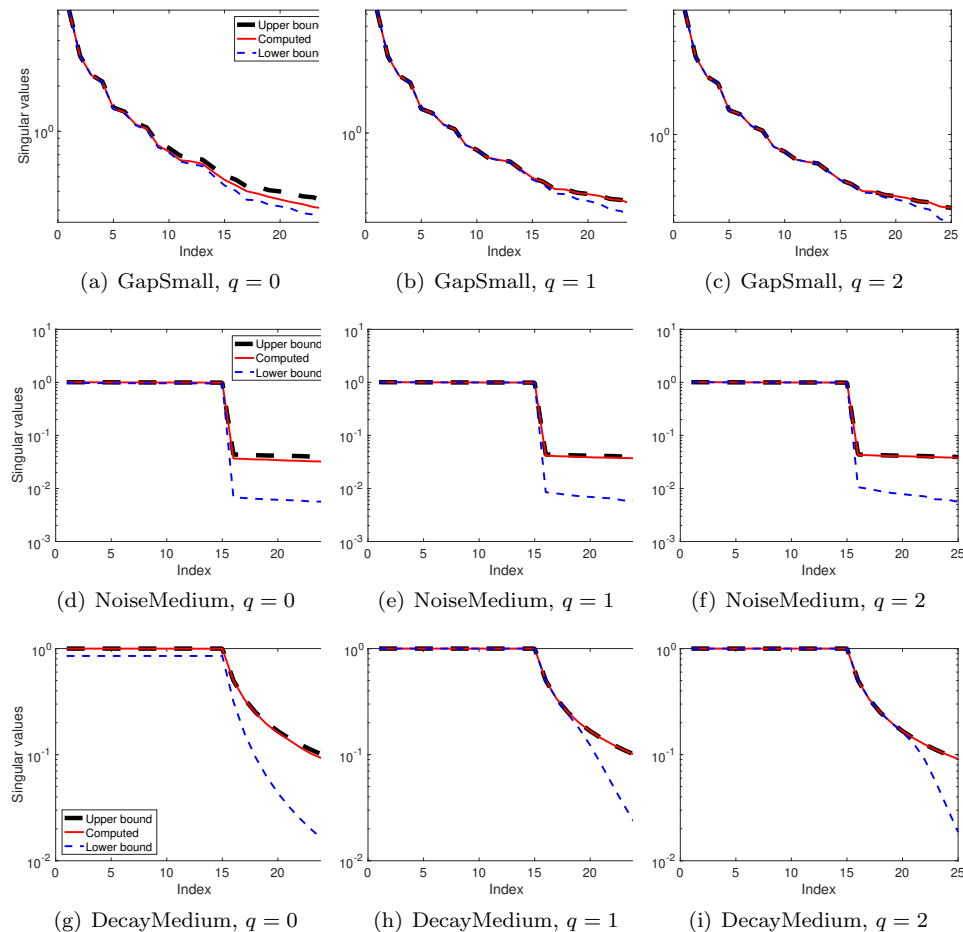
FIG. 4. *Plots of the singular values. The test matrices were described in subsection* 6.1*. The target rank k = 25 and an oversampling parameter of ρ = 20 were chosen for all the experiments. The solid lines black and blue lines correspond to the upper and lower bounds, respectively, and the dashed red lines correspond to bounds obtained using Theorem* 9*. The parameter q corresponds to the number of subspace iterations.*

3. DecayMedium. Similar to the previous example, the lower bounds are good before index 15 and improve with the number of iterations after index 15.

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

REFERENCES

[1] H. AVRON, C. BOUTSIDIS, S. TOLEDO, AND A. ZOUZIAS, *Efficient dimensionality reduction for canonical correlation analysis*, in Proceedings of the International Conference on Machine Learning, 2013, pp. 347–355.

[2] O. BALABANOV AND A. NOUY, *Randomized Linear Algebra for Model Reduction. Part* I*: Galerkin Methods and Error Estimation*, preprint, arXiv:1803.02602, 2018.

[3] F. L. BAUER, *Das verfahren der treppeniteration und verwandte verfahren zur lösung algebraischer eigenwertprobleme*, Z. Angew. Math. Phys. ZAMP, 8 (1957), pp. 214–235, https://doi.org/10.1007/BF01600502.

[4] R. Bhatia, *Matrix Analysis*, Grad. Texts in Math. 169, Springer-Verlag, New York, 1997, https://doi.org/10.1007/978-1-4612-0653-8.

[5] A. Björck and G. H. Golub, *Numerical methods for computing angles between linear subspaces*, Math. Comp., 27 (1973), pp. 579–594.

[6] C. Boutsidis, P. Kambadur, and A. Gittens, *Spectral clustering via the power method-provably*, in Proceedings of the International Conference on Machine Learning, 2015, pp. 40–48.

[7] P. Drineas and I. C. F. Ipsen, *Low-Rank Matrix Approximations Do Not Need a SIngular Value Gap*, 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 J. Matrix Anal. Appl., 39 (2018), pp. 567–586.

[9] N. B. Erichson, S. L. Brunton, and J. N. Kutz, *Randomized Dynamic Mode Decomposition*, preprint, arXiv:1702.02912, 2017.

[10] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 4th ed., Johns Hopkins University Press, Baltimore, 2013.

[11] M. Gu, *Subspace iteration randomization and singular value problems*, SIAM J. Sci. Comput., 37 (2015), pp. A1139–A1173.

[12] N. Halko, P. G. Martinsson, and J. A. Tropp, *Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions*, SIAM Rev., 53 (2011), pp. 217–288.

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

[14] J. T. Holodnak, I. C. F. Ipsen, and T. Wentworth, *Conditioning of leverage scores and computation by QR decomposition*, SIAM J. Matrix Anal. Appl., 36 (2015), pp. 1143–1163.

[15] R. A. Horn and C. R. Johnson, *Matrix Analysis*, 2nd ed., Cambridge University Press, Cambridge, UK, 2013.

[16] E. Y. Lee, *Extension of Rotfel'd theorem*, Linear Algebra Appl., 435 (2011), pp. 735–741.

[17] M. W. Mahoney, *Randomized algorithms for matrices and data*, Found. Trends Mach. Learn., 3 (2011), pp. 123–224.

[18] Y. Nakatsukasa, *Accuracy of singular vectors obtained by projection-based SVD methods*, BIT, 57 (2017), pp. 1137–1152.

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

[20] Y. Saad, *Numerical Methods for Large Eigenvalue Problems*, Classics in Appl. Math. 66, SIAM, Philadelphia, 2011, https://doi.org/10.1137/1.9781611970739.ch1.

[21] A. Sankar, D. A. Spielman, and S.-H. Teng, *Smoothed analysis of the condition numbers and growth factors of matrices*, SIAM J. Matrix Anal. Appl., 28 (2006), pp. 446–476.

[22] D. C. Sorensen and M. Embree, *A DEIM induced CUR factorization*, SIAM J. Sci. Comput., 38 (2016), pp. A1454–A1482.

[23] P. T. K. Tang and E. Polizzi, *FEAST as a subspace iteration eigensolver accelerated by approximate spectral projection*, SIAM J. Matrix Anal. Appl., 35 (2014), pp. 354–390.

[24] J. A. Tropp, A. Yurtsever, M. Udell, and V. Cevher, *Practical sketching algorithms for low-rank matrix approximation*, SIAM J. Matrix Anal. Appl., 38 (2017), pp. 1454–1485.

[25] P. Å. Wedin, *On angles between subspaces of a finite dimensional inner product space*, in Matrix Pencils, B. Kågström and A. Ruhe, eds., Springer, Berlin, 1983, pp. 263–285.

[26] K. Ye and L. H. Lim, *Schubert varieties and distances between subspaces of different dimensions*, SIAM J. Matrix Anal. Appl., 37 (2016), pp. 1176–1197.

[27] J. Zhang, A. K. Saibaba, M. E. Kilmer, and S. Aeron, *A randomized tensor singular value decomposition based on the t-product*, Numer. Linear Algebra Appl., 25 (2018), e2179.