CrossMark

BIT

Accuracy of singular vectors obtained by projection-based SVD methods

Yuji Nakatsukasa[1] (ID)

Abstract The standard approach to computing an approximate SVD of a large-scale matrix is to project it onto lower-dimensional trial subspaces from both sides, compute the SVD of the small projected matrix, and project it back to the original space. This results in a low-rank approximate SVD to the original matrix, and we can then obtain approximate left and right singular subspaces by extracting subsets from the approximate SVD. In this work we assess the quality of the extraction process in terms of the accuracy of the approximate singular subspaces, measured by the angle between the exact and extracted subspaces (relative to the angle between the exact and trial subspaces). The main message is that the extracted approximate subspaces are optimal usually to within a modest constant.

1 Introduction

The SVD plays a central role in a vast variety of applications, primarily due to its best low-rank approximation property [11, Thm. 7.4.9.1], which is valid in any unitarily

✉ Yuji Nakatsukasa
Yuji.Nakatsukasa@maths.ox.ac.uk

1   Mathematical Institute, University of Oxford, Oxford OX2 6GG, UK

invariant norm. Frequently in practice, the matrix is too large to compute the full SVD, and computing a partial and approximate SVD $A \approx U_1 \Sigma_1 V_1^*$ is of interest, where U_1, V_1 are tall-skinny matrices.

The standard approach to computing an approximate SVD of a large-scale matrix $A \in \mathbb{C}^{m \times n}$ is to project A onto lower-dimensional trial subspaces spanned by $\widehat{U} \in \mathbb{C}^{m \times \widehat{m}}$, $\widehat{V} \in \mathbb{C}^{n \times \widehat{n}}$ having orthonormal columns (how to choose $\widehat{U}, \widehat{V}$ is outside the scope; we refer e.g. to [1,7]), compute the SVD of the small $\widehat{m} \times \widehat{n}$ matrix $\widehat{A} := \widehat{U}^* A \widehat{V} = \widetilde{U} \widetilde{\Sigma} \widetilde{V}^*$ and obtain an approximate tall-skinny SVD as $A \approx (\widehat{U} \widetilde{U}) \widehat{\Sigma} (\widehat{V} \widetilde{V})^*$, represented explicitly as a low-rank (rank $\min(\widehat{m}, \widehat{n})$) matrix. Some of the columns of $\widehat{U} \widetilde{U}$ and $\widehat{V} \widetilde{V}$ then approximate the exact left and right singular vectors of A. This process is an analogue of the Rayleigh–Ritz process for the symmetric eigenvalue problem, sometimes called the Petrov–Galerkin method [2], since the residual $R = A - (\widehat{U} \widetilde{U}) \widehat{\Sigma} (\widehat{V} \widetilde{V})^*$ is orthogonal to $\widehat{U}, \widehat{V}$, that is, $\widehat{U}^* R = 0$, $R \widehat{V} = 0$.

For the Rayleigh–Ritz process for the symmetric eigenvalue problem, existing results are available for bounding the accuracy of the computed eigenvectors, most notably through results by Saad [20, Thm. 4.6], Knyazev [12, Thm. 4.3] and Stewart [21, Thm. 2]. Roughly speaking, these results show that the Rayleigh–Ritz process extracts an approximate set of eigenvectors that are optimal (for a fixed subspace) up to a certain constant factor, which depends on the residual norm and a gap between the exact and approximate eigenvalues (Ritz values).

This work derives analogous results for the SVD by establishing bounds for the accuracy of the computed left and right singular vectors, measured by the angle between the computed and desired subspaces. In essence, the message is the same as in the Rayleigh–Ritz process for eigenvalue problems: the projection algorithm obtains approximate sets of left and right singular vectors that are optimal to within a constant factor, again depending on the residual norm and a gap between exact and approximate singular values. The vector case ($m_1 = n_1 = 1$ in the notation below) was dealt with by Hochstenbach [10, Thm. 2.4]; the results here improve the bound slightly, and generalize to subspaces $m_1, n_1 > 1$. A preliminary version appeared in the author's PhD dissertation [17, Ch. 10].

Let us clarify the problem formulation. Let A be an $m \times n$ matrix and $\widehat{U} \in \mathbb{C}^{m \times \widehat{m}}$, $\widehat{V} \in \mathbb{C}^{n \times \widehat{n}}$ (throughout, matrices U, V and the variants $\widehat{U}, \widehat{V}, \widetilde{U}, \widetilde{V}$ have orthonormal columns) represent trial subspaces that are hoped to respectively approximately contain the subspace spanned by a set of desired $k \leq \min(\widehat{m}, \widehat{n})$ exact singular vectors U_1 and V_1, corresponding to some of the singular values, usually but not necessarily the largest ones. That is,

$$A = [U_1 \ U_1^\perp] \begin{bmatrix} \Sigma_1 \\ & \Sigma_{(\perp)} \end{bmatrix} [V_1 \ V_1^\perp]^* \tag{1}$$

is an exact full SVD, with $\Sigma_1 \in \mathbb{R}^{k \times k}$ and the singular values are arranged in an arbitrary (not necessarily decreasing) order. Let $[\widehat{U} \ \widehat{U}_3] \in \mathbb{C}^{m \times m}$ and $[\widehat{V} \ \widehat{V}_3] \in \mathbb{C}^{n \times n}$ be square unitary matrices and (we are using the subscript 3 for consistency with what follows)

$$(\widetilde{A} =) [\widehat{U} \ \widehat{U}_3]^* A [\widehat{V} \ \widehat{V}_3] = \begin{bmatrix} \widehat{\Sigma} & R \\ S & A_3 \end{bmatrix}. \tag{2}$$

In practice we usually do not have access to $\widehat{U}_3$ and $\widehat{V}_3$. Accordingly we do not know R or S, but their norms can be computed via $\|\widetilde{S}\| = \|A\widehat{V} - \widetilde{U}\widehat{\Sigma}\|$ and $\|R\| = \|\widehat{U}^*A - \widehat{\Sigma}\widehat{V}^*\|$, which hold for any unitarily invariant norm. Similarly, we will not need to know A_3, although the assumptions made in the results will have implications on its singular values.

If $\|R\| = \|S\| = 0$, then $\widehat{U}$ and $\widehat{V}$ correspond to exact left and right singular subspaces. When R, S are both small we can expect $\widehat{U}, \widehat{V}$ to be good approximations to some singular vectors. Bounds that assess the quality of the projection subspaces $\widehat{U}$ and $\widehat{V}$ are derived in a classical result by Wedin [24].

The focus of this paper is the quality of the projection process as a means to extract good singular vectors from the projection subspaces. Specifically, in practice, it often happens that the whole projection space $\widehat{U}$ is not very close to an exact singular subspace, but a subspace of $\widehat{U}$ is close to an exact singular subspace of lower dimension (this is the case e.g. in the Jacobi–Davidson context [9], in which $\widehat{U}, \widehat{V}$ contain approximate and search subspaces). In view of this, the hope is that if $\widehat{U}$ contains rich information in an exact singular subspace U_1 (which is usually of lower dimension k than $\widehat{m}$), then the projection method computes a good approximant $\widehat{U}_1$ to U_1. Quantifying this hope is the objective of this paper.

From the algorithmic viewpoint, we compute the projected matrix $\widehat{A} = \widehat{U}^*A\widehat{V} \in \mathbb{C}^{\widehat{m}\times\widehat{n}}$ and its (full) SVD

$$\widehat{A} = \widehat{U}^*A\widehat{V} = [\widetilde{U}_1\ \widetilde{U}_2]\begin{bmatrix}\widehat{\Sigma}_1 & \\ & \widehat{\Sigma}_2\end{bmatrix}[\widetilde{V}_1\ \widetilde{V}_2]^*,$$

where $\widehat{\Sigma}_1$ is $m_1 \times n_1$ and $\widehat{\Sigma}_2$ is $(\widehat{m} - m_1) \times (\widehat{n} - n_1)$; hence $\widetilde{U}_1 \in \mathbb{C}^{\widehat{m}\times m_1}, \widetilde{U}_2 \in \mathbb{C}^{\widehat{m}\times(\widehat{m}-m_1)}, \widetilde{V}_1 \in \mathbb{C}^{\widehat{n}\times n_1}, \widetilde{V}_2 \in \mathbb{C}^{\widehat{n}\times(\widehat{n}-n_1)}$. We do not impose any ordering in the singular values $\widehat{\Sigma}_1, \widehat{\Sigma}_2$. Then, defining $\widehat{U}_i = \widehat{U}\widetilde{U}_i$ and $\widehat{V}_i = \widehat{V}\widetilde{V}_i$ for $i = \{1, 2\}$, we have

$$\widetilde{A} := [\widehat{U}_1\ \widehat{U}_2\ \widehat{U}_3]^* A [\widehat{V}_1\ \widehat{V}_2\ \widehat{V}_3] = \begin{bmatrix}\widehat{\Sigma}_1 & 0 & R_1 \\ 0 & \widehat{\Sigma}_2 & R_2 \\ S_1 & S_2 & A_3\end{bmatrix}. \tag{3}$$

The goal here is to assess the quality of the extracted subspaces by examining the angles $\angle(U_1, \widehat{U}_1)$ and $\angle(V_1, \widehat{V}_1)$, the angles between extracted and exact singular subspaces, as compared with $\angle(U_1, \widehat{U})$ and $\angle(V_1, \widehat{V})$, the angles between the trial and exact subspace. Since $\mathrm{span}(\widehat{U}_1) \subseteq \mathrm{span}(\widehat{U})$, we trivially have $\angle(U_1, \widehat{U}) \leq \angle(U_1, \widehat{U}_1)$ and $\angle(V_1, \widehat{V}) \leq \angle(V_1, \widehat{V}_1)$. The quality of the extraction is measured by how far from best possible the extracted subspace $\widehat{U}_1, \widehat{V}_1$ are: the extraction is considered near-optimal if $\angle(U_1, \widehat{U}_1) \leq c\angle(U_1, \widehat{U})$ and $\angle(V_1, \widehat{V}_1) \leq c\angle(V_1, \widehat{V})$ hold for a constant c not much larger than 1.

Let us recall the definition of angles between subspaces. The angles θ_i between two subspaces spanned by $X \in \mathbb{C}^{m\times n_X}, Y \in \mathbb{C}^{m\times n_Y}$ $(n_X \leq n_Y(\leq m))$ with orthonormal columns are defined by $\theta_i = \mathrm{acos}(\sigma_i(X^*Y))$; they are known as the canonical angles or principal angles [5, Thm. 6.4.3]. These are connected to the CS decomposition for unitary matrices [5, Thm. 2.5.3], [18], in which

the matrix $\begin{bmatrix} Y & Y^\perp \end{bmatrix}^* X$ satisfies $\mathrm{diag}(Q_1, Q_2)(\begin{bmatrix} Y & Y^\perp \end{bmatrix}^* X)W = \begin{bmatrix} C \\ S \\ 0 \end{bmatrix}$, where

$Q_1 \in \mathbb{C}^{n_X \times n_X}$, $Q_2 \in \mathbb{C}^{(n-n_X) \times (n-n_X)}$, $W \in \mathbb{C}^{n_X \times n_X}$ are all square unitary and $C = \mathrm{diag}(\cos\theta_1, \ldots, \cos\theta_{n_X})$, $S = \mathrm{diag}(\sin\theta_1, \ldots, \sin\theta_{n_X})$. We write $\angle(X, Y) = \mathrm{diag}(\theta_1, \ldots, \theta_{n_X})$, and $\sin\angle(X, Y) = S$.

Applying a unitary transformation in (3) and (1), we see that $\angle(\widehat{U}_1, U_1) = \angle\left(\widetilde{U}_1, \begin{bmatrix} I_k \\ 0 \end{bmatrix}\right)$ and $\angle(\widehat{V}_1, V_1) = \angle\left(\widetilde{V}_1, \begin{bmatrix} I_k \\ 0 \end{bmatrix}\right)$, where $\widetilde{U}_1, \widetilde{V}_1$ are matrices of exact singular vectors of $\widetilde{A}$ in (3). We note that it is natural to take $k = m_1 = n_1$, although this is not necessary; k is an integer that one can choose from $1 \leq k \leq \min(m_1, n_1)$; the upper bound is required for the extracted subspace $\widetilde{U}_1, \widetilde{V}_1$ to contain the exact subspaces.

Here is the plan of this paper. First in Sect. 2 we review the results for the Rayleigh–Ritz process in symmetric eigenvalue problems. We then derive analogous results for the SVD in Sect. 3.

Notation. $\widehat{U}, \widehat{V}$ always respectively denote the left and right trial projection subspaces. $\widehat{U}_1$ and $\widehat{V}_1$ are the approximate subspaces obtained via the projection method and they satisfy $\mathrm{span}(\widehat{U}_1) \subseteq \mathrm{span}(\widehat{U})$, $\mathrm{span}(\widehat{V}_1) \subseteq \mathrm{span}(\widehat{V})$. U_1, V_1 are certain exact singular subspaces of A that $\widehat{U}_1, \widehat{V}_1$ approximate. I_n is the $n \times n$ identity matrix. $X^\perp$ denotes the orthogonal complement of X. $\| \cdot \|_2$ denotes the spectral norm, equal to the largest singular value. $\| \cdot \|_F$ is the Frobenius norm $\|A\|_F = \sqrt{\sum_{i,j} A_{ij}^2}$. Identities and inequalities involving $\| \cdot \|$ without subscripts indicate they hold for any unitarily invariant norm, and those involving $\| \cdot \|_{2,F}$ hold for both the spectral norm and the Frobenius norm, but not necessarily for every unitarily invariant norm.

$\sigma_i(A)$ denotes the ith largest singular value of A. For $A \in \mathbb{C}^{m \times n}$, by a full SVD we mean the decomposition $A = U\Sigma V^*$ where U, V are square unitary, hence Σ is $m \times n$. We define $\sigma_{\min}(A) = \sigma_{\min(m,n)}(A)$, and

$$\sigma_{\widetilde{\min}}(A) := \begin{cases} \sigma_n(A) & \text{if } m = n, \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

In words, $\sigma_{\widetilde{\min}}(A)$ is equal to the smallest singular value when $m = n$, but 0 otherwise. $\lambda(A)$ denotes the set of eigenvalues of A. The key identity we use pertaining to $\sigma_{\widetilde{\min}}(A)$ is $\|AB\| \geq \sigma_{\widetilde{\min}}(A)\|B\|$.

2 The Rayleigh–Ritz process for Hermitian eigenproblems and theorems of Saad, Knyazev and Stewart

The standard way of computing a subset of the eigenvalues and eigenvectors of a large-sparse Hermitian (or real symmetric) matrix $A \in \mathbb{C}^{n \times n}$ is to form a low-dimensional subspace $\mathrm{span}(\widehat{X})$ with $\widehat{X} \in \mathbb{C}^{n \times \widehat{n}}$ having orthonormal columns, which approximately contains the desired eigenvectors, and then extract approximate eigenpairs (called the *Ritz pairs*) from it by means of the *Rayleigh–Ritz process* [19, Ch 11]. This process computes the eigendecomposition of the small Hermitian matrix $\widehat{X}^* A \widehat{X} = \widetilde{X}\widehat{\Lambda}\widetilde{X}^*$,

from which the Ritz values are taken as the diagonals of $\widehat{\Lambda}$ and the Ritz vectors are the columns of $\widehat{X}\widetilde{X} =: [\widehat{X}_1 \ \widehat{X}_2]$. The Ritz pairs $(\widehat{\lambda}, \widehat{x})$ thus obtained satisfy $\widehat{x} \in \text{span}(\widehat{X})$, and $A\widehat{x} - \widehat{\lambda}\widehat{x} \perp \widehat{X}$. Observe the similarity between the Rayleigh–Ritz process and the Petrov–Galerkin projection method for the SVD.

A natural question is to ask how accurate the Ritz pairs are as approximation to the exact eigenpairs. The starting point is an analogue of (3): the partitioning of the projected A as

$$(\widetilde{A} =)[\widehat{X}_1 \ \widehat{X}_2 \ \widehat{X}_3]^* A [\widehat{X}_1 \ \widehat{X}_2 \ \widehat{X}_3] = \begin{bmatrix} \widehat{\Lambda}_1 & 0 & R_1^* \\ 0 & \widehat{\Lambda}_2 & R_2^* \\ R_1 & R_2 & A_3 \end{bmatrix}, \tag{5}$$

where $[\widehat{X}_1 \ \widehat{X}_2 \ \widehat{X}_3]$ is an orthogonal matrix, with $\widehat{X}_1 \in \mathbb{C}^{n \times n_1}$, $\widehat{X}_2 \in \mathbb{C}^{n \times (\widehat{n} - n_1)}$, and $\widehat{X}_3 \in \mathbb{C}^{n \times (n - \widehat{n})}$. The computable quantities are $\widehat{\Lambda}_1$, $\widehat{\Lambda}_2$ and the norms of each column of R_1, R_2. In the context of the Rayleigh–Ritz process, $\text{span}(\widehat{X}) = \text{span}([\widehat{X}_1 \ \widehat{X}_2])$ is the projection subspace, $\lambda(\widehat{\Lambda}_1)$ and $\lambda(\widehat{\Lambda}_2)$ are the Ritz values, and the goal here is to examine the accuracy of $\widehat{X}_1$ as an approximate eigenspace $X_1 \in \mathbb{C}^{n \times k}$ of A (or more precisely, invariant subspace) such that $AX_1 = X_1 \Lambda_1$.

We note that bounding the eigenvalue accuracy can be done using standard eigenvalue perturbation theory: for example by Weyl's theorem, the Ritz values Λ_i match some of those of A to within $\|R_i\|_2$ for $i = 1, 2$. For individual Ritz values, the corresponding residual norm $\|A\widehat{x}_i - \widehat{\lambda}_i\widehat{x}_i\|_2$ is a bound for the distance to the closest exact eigenvalue. Moreover, by using results in [13,15] one can often get tighter bounds for the eigenvalue accuracy, which scale quadratically with the norm of the residual.

2.1 Bounds for approximate eigenvector

Now we turn to eigenvectors and discuss the accuracy of the Ritz vectors. In the single-vector $n_1 = k = 1$ case, we bound the angle between approximate and exact eigenvectors $\widehat{x}$ and x. Saad [20, Thm. 4.6] proves the following theorem, applicable to (5) in the case where $\widehat{\Lambda}_1 = \widehat{\lambda}$ is a scalar and $\widehat{x} = \widehat{X}_1$ is a vector ($\widehat{X}$ is still a subspace of dimension > 1).

Theorem 1 *Let A be a Hermitian matrix and let (λ, x) be any of its eigenpairs. Let $(\widehat{\lambda}, \widehat{x})$ be a Ritz pair obtained from the subspace $\text{span}(\widehat{X})$, such that $\widehat{\lambda}$ is the closest Ritz value to λ. Suppose $\delta > 0$, where δ is the distance between λ and the set of Ritz values other than $\widehat{\lambda}$. Then*

$$\sin \angle(x, \widehat{x}) \leq \sin \angle(x, \widehat{X})\sqrt{1 + \frac{\|r\|_2^2}{\delta^2}},$$

where $r = A\widehat{x} - \widehat{\lambda}\widehat{x}$.

Recalling that $\sin \angle(x, \widehat{x}) \geq \sin \angle(x, \widehat{X})$ holds trivially because $x \in \text{span}(\widehat{X})$, we see that the above theorem shows that the Ritz vector is optimal up to the constant $\sqrt{1 + \frac{\|r\|_2^2}{\delta^2}}$.

Remark 1 This does not necessarily mean, however, that performing the Rayleigh–Ritz process is always a reliable way to extract approximate eigenpairs from a subspace; care is needed especially when computing interior eigenpairs, in which δ can be very small or zero. One remedy is to use the harmonic Rayleigh–Ritz process. For more on this issue, see for example [16] and [22, Sect. 5.1.4].

2.2 Bounds for approximate eigenspace

Knyazev [12, Thm. 4.3] derives an extension of Saad's result (Theorem 1) to approximate eigenspaces, presenting bounds in the spectral norm applicable in the context of self-adjoint linear operators in a Hilbert space. Stewart [21,22] proves an analogous result applicable to non-Hermitian matrices.

Here we state a result specializing to Hermitian matrices. We give a proof since we use the same line of argument later for the SVD; The bound below slightly improves the classical ones in two ways: First, the term $\|R_2\|$ in the bounds (6), (7) are $\|[R_1\ R_2]\|$ in [12]. Second, (6) holds for any unitarily invariant norm. Essentially the same bound as (6) appears in [14, Thm. 4.1] (under relaxed conditions on the spectrum; see "Appendix").

Theorem 2 *Let A be a Hermitian matrix. Let $\widetilde{A}$ be as in* (5)*; $(\widehat{\Lambda}_1, \widehat{X}_1)$ is a set of Ritz pairs with $\widehat{\Lambda}_1 \in \mathbb{R}^{n_1 \times n_1}$. Let (Λ_1, X_1) with $\Lambda_1 \in \mathbb{R}^{k \times k}$, $k \leq n_1$ be a set of exact eigenpairs whose eigenvalues lie in the interval $[\lambda_0 - d, \lambda_0 + d]$ for some λ_0 and $d > 0$. Suppose that $\delta = \min |\lambda(\widehat{\Lambda}_2) - \lambda_0| - d > 0$, where $\widehat{\Lambda}_2 \in \mathbb{R}^{(\widehat{n}-n_1) \times (\widehat{n}-n_1)}$ is as in* (5)*. Then for any unitarily invariant norm*

$$\left\| \sin \angle (X_1, \widehat{X}_1) \right\| \leq \left\| \sin \angle (X_1, \widehat{X}) \right\| \left(1 + \frac{\|R_2\|_2}{\delta} \right), \tag{6}$$

and for the spectral and Frobenius norms,

$$\left\| \sin \angle (X_1, \widehat{X}_1) \right\|_{2,F} \leq \left\| \sin \angle (X_1, \widehat{X}) \right\|_{2,F} \sqrt{1 + \frac{\|R_2\|_2^2}{\delta^2}}. \tag{7}$$

Note that when applicable, (7) is slightly tighter than (6).

Proof Let $\widetilde{X} = \begin{bmatrix} \widetilde{X}_1 \\ \widetilde{X}_2 \\ \widetilde{X}_3 \end{bmatrix}$ be a set of eigenvectors of $\widetilde{A}$ in (5) with orthonormal columns. A key step is to note from the CS decomposition that

$$\| \sin \angle (X_1, \widehat{X}) \| = \|\sigma_i((\widehat{X}^\perp)^* X_1)\| = \left\| \widetilde{X}_3 \right\| \tag{8}$$

and

$$\| \sin \angle (X_1, \widehat{X}_1) \| = \|\sigma_i((\widehat{X}_1^\perp)^* X_1)\| = \left\| \begin{bmatrix} \widetilde{X}_2 \\ \widetilde{X}_3 \end{bmatrix} \right\|. \tag{9}$$

The second block of $\widetilde{A}\widetilde{X} = \widetilde{X}\Lambda_1$ gives

$$\widehat{\Lambda}_2\widetilde{X}_2 + R_2^*\widetilde{X}_3 = \widetilde{X}_2\Lambda_1,$$

which is equivalent to

$$\widehat{\Lambda}_2\widetilde{X}_2 - \widetilde{X}_2\Lambda_1 = -R_2^*\widetilde{X}_3. \tag{10}$$

We now use the following fact [23, p. 251]: Let $F \in \mathbb{C}^{m \times m}$ and $G \in \mathbb{C}^{n \times n}$ be such that $1/\|F^{-1}\| - \|G\| = \delta > 0$. Then for any $W \in \mathbb{C}^{m \times n}$, we have $\|W\| \leq \frac{\|FW - WG\|}{\delta}$. We note that many results in eigenvector perturbation theory can be derived using this fact [23].

We use this in (10) as follows: Rewrite the equation as

$$(\widehat{\Lambda}_2 - \lambda_0 I_{\widehat{n}-n_1})\widetilde{X}_2 - \widetilde{X}_2(\Lambda_1 - \lambda_0 I_{n_1}) = -R_2^*\widetilde{X}_3,$$

and take $(\widehat{\Lambda}_2 - \lambda_0 I_{\widehat{n}-n_1}) := F$, $\widetilde{X}_2 := W$, and $\Lambda_1 - \lambda_0 I_{n_1} := G$. Then by the assumptions on the spectrum we have $1/\|F^{-1}\| \geq \min|\lambda(\widehat{\Lambda}_2) - \lambda_0| = \delta + d$ and $\|G\| \leq d$, so $1/\|F^{-1}\| - \|G\| \geq \delta > 0$. Therefore, we obtain

$$\|\widetilde{X}_2\| \leq \frac{\|R_2^*\widetilde{X}_3\|}{\delta}.$$

Hence using the fact [8, p. 327] that $\|XY\| \leq \|X\|_2\|Y\|$ for any unitarily invariant norm we obtain

$$\|\widetilde{X}_2\| \leq \frac{\|R_2\|_2}{\delta}\|\widetilde{X}_3\|. \tag{11}$$

Therefore, recalling (8) and (9) we have

$$\|\sin\angle(X_1, \widehat{X}_1)\| = \left\|\begin{bmatrix}\widetilde{X}_2 \\ \widetilde{X}_3\end{bmatrix}\right\| \leq \|\widetilde{X}_2\| + \|\widetilde{X}_3\|$$

$$\leq (1 + \frac{\|R_2\|_2}{\delta})\|\widetilde{X}_3\| = (1 + \frac{\|R_2\|_2}{\delta})\|\sin\angle(X_1, \widehat{X})\|, \tag{12}$$

which is (6). In the spectral or Frobenius norm we can use the inequality $\|\begin{bmatrix} A \\ B \end{bmatrix}\|_{2,F} \leq \sqrt{\|A\|_{2,F}^2 + \|B\|_{2,F}^2}$; note that this does not necessarily hold for other unitarily invariant norms, e.g. the nuclear (or trace) norm $\|A\|_* = \sum_i \sigma_i(A)$.

We thus obtain the slightly tighter bound

$$\|\sin\angle(X_1, \widehat{X}_1)\|_{2,F} \leq \sqrt{\|\widetilde{X}_2\|_{2,F}^2 + \|\widetilde{X}_3\|_{2,F}^2} = \sqrt{1 + \frac{\|R_2\|_2^2}{\delta^2}}\|\widetilde{X}_3\|_{2,F}$$

$$= \sqrt{1 + \frac{\|R_2\|_2^2}{\delta^2}}\|\sin(X_1, \widehat{X})\|_{2,F}, \tag{13}$$

which is (7). $\qquad\square$

As a historical sidenote, Theorem 2 (or more specifically the setup in (5)) is related to Question 10.2 in the classical paper by Davis and Kahan [4], where they ask for subspace angle bounds between three subspaces. Theorem 2 is related but not a direct answer to their question, since it provides bounds not on the direct subspace angles (as in Davis–Kahan's $\sin\theta$, $\tan\theta$ theorems) but the relative quality of the extracted subspace $\|\sin\angle(X_1, \widehat{X}_1)\|/\|\sin\angle(X_1, \widehat{X})\|$, where $\dim\widehat{X} > \dim X_1$. Note also that the Rayleigh–Ritz process imposes a structure in the residual (the zero $(1, 2)$ and $(2, 1)$-blocks) as in (5).

It is possible to relax the assumptions on the spectrum; in particular, we can allow the eigenvalues of Λ_1 and $\widehat{\Lambda}_2$ to interlace. To proceed straight to the main subject SVD, we defer the discussion to the "Appendix".

It is perhaps worth emphasizing that the "gap" δ is the gap between some of the Ritz values and some of the exact eigenvalues. In particular, the gap does not involve the eigenvalues of A_3, in contrast to the gap that arises in the context of quadratic eigenvalue perturbation bounds [13,15]. The same holds for the results for the SVD as we describe next.

3 Accuracy bounds for approximate singular subspaces

We now turn to our main subject SVD. Essentially, the goal is to derive the SVD analogues of Theorem 2. Regarding the accuracy of the singular values, just as in the symmetric eigenproblem, the errors can be bounded using standard perturbation theory, most importantly Weyl's bound; bounds that scale quadratically with the residual are also known [13]. Our focus here is the accuracy of the singular vectors.

Two approaches are commonly employed for extending results for the symmetric eigenproblem to the SVD of $A \in \mathbb{C}^{m \times n}$ (the discussion here assumes $m \geq n$):

1. Work with the Gram matrices A^*A and AA^*, which respectively have V and $[U\ U^\perp]$ as the eigenvector matrices.
2. Work with the Jordan–Wieldant matrix $\begin{bmatrix} 0 & A^* \\ A & 0 \end{bmatrix}$, whose eigenvalues are $\pm\sigma_i(A)$, along with $|m - n|$ copies of 0. The eigenvectors are (assuming $m \geq n$) $\begin{bmatrix} U & U & U^\perp \\ V & -V & 0 \end{bmatrix}$, where U, V are the matrices of left and right singular vectors of A, and $U^\perp \in \mathbb{C}^{m \times (m - \widehat{n})}$ is the orthogonal complement of U.

We avoid the Gram matrices because A^*A have eigenvalues $\sigma_i^2(A)$, through which the gap will be defined, and because of the squaring, the gap becomes smaller especially in the small singular values, resulting in unnecessarily large bounds.

Working with the Jordan–Wieldant matrix is an effective approach, and would give a bound on $\angle(\begin{bmatrix} \widehat{U} \\ \widehat{V} \end{bmatrix}, \begin{bmatrix} U \\ V \end{bmatrix})$ instead of $\angle(\widehat{U}, U)$ and $\angle(\widehat{V}, V)$. By using the technique employed in [23, p. 261] it is possible to deduce bounds on $\angle(U_1, \widehat{U}_1)$ and $\angle(V_1, \widehat{V}_1)$. Hochstenbach [10, Thm. 2.4] takes this approach to derive a bound for singular vectors (the case $m_1 = n_1 = k = 1$ in our setting). A slight issue with this approach is the presence of the $|m - n|$ extra eigenvalues at 0, which can result in the gap quantity

(δ below) being unnecessarily small; indeed in [10, Thm. 2.4] the gap is defined differently depending on whether $m = n$ or not. Our result below shows that this distinction is unnecessary (see remark at the end of Sect. 3.1). In addition, we remove a factor 2 in the bound [10], and more importantly, extend the analysis to the subspace case $m_1, n_1, k > 1$.

For the above reasons, in what follows we do not take the two approaches, and instead work directly with the matrix A and its projection.

3.1 Main result

We shall prove the following, which is the main result of this paper.

Theorem 3 *Let $A \in \mathbb{C}^{m \times n}$. Let $\widehat{U} = [\widehat{U}_1 \, \widehat{U}_2] \in \mathbb{C}^{m \times \widehat{m}}$ and $\widehat{V} = [\widehat{V}_1 \, \widehat{V}_2] \in \mathbb{C}^{n \times \widehat{n}}$ have orthonormal columns with $\widehat{U}_1 \in \mathbb{C}^{m \times m_1}$ and $\widehat{V}_1 \in \mathbb{C}^{n \times n_1}$, and let $\tilde{A}, \widehat{\Sigma}_i, R_i, S_i, A_3$ be as defined in (3), and let (1) be an exact SVD with a singular triplet (Σ_1, U_1, V_1), with $\Sigma_1 \in \mathbb{R}^{k \times k}$. Define (recall the definition of $\sigma_{\widetilde{\min}}$ in (4))*

$$\delta = \max(\sigma_{\min}(\Sigma_1) - \|\widehat{\Sigma}_2\|_2, \sigma_{\widetilde{\min}}(\widehat{\Sigma}_2) - \|\Sigma_1\|_2). \tag{14}$$

If $\delta > 0$, then

$$\max(\| \sin \angle(U_1, \widehat{U}_1)\|, \| \sin \angle(V_1, \widehat{V}_1)\|)$$
$$\leq \left(1 + \frac{\max(\|R_2\|_2, \|S_2\|_2)}{\delta}\right) \max(\| \sin \angle(U_1, \widehat{U})\|, \| \sin \angle(V_1, \widehat{V})\|) \tag{15}$$

and

$$\| \sin\angle(U_1, \widehat{U}_1)\| + \| \sin \angle(V_1, \widehat{V}_1)\|$$
$$\leq \left(1 + \frac{\max(\|R_2\|_2, \|S_2\|_2)}{\delta}\right) \left(\| \sin \angle(U_1, \widehat{U})\| + \| \sin \angle(V_1, \widehat{V})\|\right) \tag{16}$$

in any unitarily invariant norm. Moreover, in the spectral and Frobenius norms

$$\max(\| \sin \angle(U_1, \widehat{U}_1)\|_{2,F}, \| \sin \angle(V_1, \widehat{V}_1)\|_{2,F})$$
$$\leq \sqrt{1 + \frac{\max(\|R_2\|_2, \|S_2\|_2)^2}{\delta^2}} \max(\| \sin \angle(U_1, \widehat{U})\|_{2,F}, \| \sin \angle(V_1, \widehat{V})\|_{2,F}).$$
$$\tag{17}$$

Proof Let $(\Sigma_1, \widetilde{U}_1, \widetilde{V}_1)$ be exact singular triplets of $\tilde{A}$ and let $\widetilde{V}_1 = \begin{bmatrix} \widetilde{V}_{11} \\ \widetilde{V}_{21} \\ \widetilde{V}_{31} \end{bmatrix}, \widetilde{U}_1 = \begin{bmatrix} \widetilde{U}_{11} \\ \widetilde{U}_{21} \\ \widetilde{U}_{31} \end{bmatrix}$, so that

$$\begin{bmatrix} \widehat{\Sigma}_1 & 0 & R_1 \\ 0 & \widehat{\Sigma}_2 & R_2 \\ S_1 & S_2 & A_3 \end{bmatrix} \begin{bmatrix} \widetilde{V}_{11} \\ \widetilde{V}_{21} \\ \widetilde{V}_{31} \end{bmatrix} = \begin{bmatrix} \widetilde{U}_{11} \\ \widetilde{U}_{21} \\ \widetilde{U}_{31} \end{bmatrix} \Sigma_1, \tag{18}$$

and

$$\begin{bmatrix} \widetilde{U}_{11}^* & \widetilde{U}_{21}^* & \widetilde{U}_{31}^* \end{bmatrix} \begin{bmatrix} \widehat{\Sigma}_1 & 0 & R_1 \\ 0 & \widehat{\Sigma}_2 & R_2 \\ S_1 & S_2 & A_3 \end{bmatrix} = \Sigma_1 \begin{bmatrix} \widetilde{V}_{11}^* & \widetilde{V}_{21}^* & \widetilde{V}_{31}^* \end{bmatrix}. \tag{19}$$

As in (8), (9) we have

$$\| \sin \angle(U_1, \widehat{U}_1) \| = \left\| \begin{bmatrix} \widetilde{U}_{21} \\ \widetilde{U}_{31} \end{bmatrix} \right\|, \quad \| \sin \angle(V_1, \widehat{V}_1) \| = \left\| \begin{bmatrix} \widetilde{V}_{21} \\ \widetilde{V}_{31} \end{bmatrix} \right\|, \tag{20}$$

and recalling that $\widehat{U} = [\widehat{U}_1 \ \widehat{U}_2]$ and $\widehat{V} = [\widehat{V}_1 \ \widehat{V}_2]$,

$$\| \sin \angle(U_1, \widehat{U}) \| = \|\widetilde{U}_{31}\|, \quad \| \sin \angle(V_1, \widehat{V}) \| = \|\widetilde{V}_{31}\|. \tag{21}$$

To establish the theorem we shall bound $\|\widetilde{U}_{21}\|$ with respect to $\|\widetilde{U}_{31}\|$, and similarly bound $\|\widetilde{V}_{21}\|$ with respect to $\|\widetilde{V}_{31}\|$.

From the second block of (18) we obtain

$$\widehat{\Sigma}_2 \widetilde{V}_{21} + R_2 \widetilde{V}_{31} = \widetilde{U}_{21} \Sigma_1, \tag{22}$$

and from the second block of (19) we get

$$\widetilde{U}_{21}^* \widehat{\Sigma}_2 + \widetilde{U}_{31}^* S_2 = \Sigma_1 \widetilde{V}_{21}^*. \tag{23}$$

Now suppose that $\delta = \sigma_{\min}(\Sigma_1) - \|\widehat{\Sigma}_2\|_2 > 0$. Then, taking norms and using the triangular inequality and the fact $\|XY\| \le \|X\|_2 \|Y\|$ in (22) and (23), we obtain

$$\begin{aligned} \|\widetilde{U}_{21}\| \sigma_{\min}(\Sigma_1) - \|\widetilde{V}_{21}\| \|\widehat{\Sigma}_2\|_2 &\le \|R_2 \widetilde{V}_{31}\|, \\ \|\widetilde{V}_{21}\| \sigma_{\min}(\Sigma_1) - \|\widetilde{U}_{21}\| \|\widehat{\Sigma}_2\|_2 &\le \|\widetilde{U}_{31}^* S_2\|. \end{aligned} \tag{24}$$

By adding the first inequality times $\sigma_{\min}(\Sigma_1)$ and the second inequality times $\|\widehat{\Sigma}_2\|$, we eliminate the $\|\widetilde{V}_{21}\|$ term, and recalling the assumption $\sigma_{\min}(\Sigma_1) > \|\widehat{\Sigma}_2\|_2$ we obtain

$$\|\widetilde{U}_{21}\| \le \frac{\sigma_{\min}(\Sigma_1) \|R_2 \widetilde{V}_{31}\| + \|\widehat{\Sigma}_2\|_2 \|\widetilde{U}_{31}^* S_2\|}{(\sigma_{\min}(\Sigma_1))^2 - \|\widehat{\Sigma}_2\|_2^2}. \tag{25}$$

We can similarly obtain

$$\|\widetilde{V}_{21}\| \le \frac{\sigma_{\min}(\Sigma_1) \|\widetilde{U}_{31}^* S_2\| + \|\widehat{\Sigma}_2\|_2 \|R_2 \widetilde{V}_{31}\|}{(\sigma_{\min}(\Sigma_1))^2 - \|\widehat{\Sigma}_2\|_2^2}. \tag{26}$$

Again using the fact $\|XY\| \le \|X\|_2 \|Y\|$ we have

$$
\begin{aligned}
\max(\|\widetilde{U}_{21}\|, \|\widetilde{V}_{21}\|) &\le \frac{\max(\|\widetilde{U}_{31}^* S_2\|, \|R_2 \widetilde{V}_{31}\|)}{\sigma_{\min}(\Sigma_1) - \|\widehat{\Sigma}_2\|_2} \\
&\le \frac{\max(\|\widetilde{U}_{31}^*\|, \|\widetilde{V}_{31}\|) \max(\|R_2\|_2, \|S_2\|_2)}{\sigma_{\min}(\Sigma_1) - \|\widehat{\Sigma}_2\|_2} \\
&= \frac{\max(\|R_2\|_2, \|S_2\|_2)}{\delta} \max(\|\widetilde{U}_{31}^*\|, \|\widetilde{V}_{31}\|). \qquad (27)
\end{aligned}
$$

Now if $\delta = \sigma_{\widetilde{\min}}(\widehat{\Sigma}_2) - \|\Sigma_1\|_2 > 0$, we proceed similarly, replacing (24) with $\|\widetilde{V}_{21}\| \sigma_{\widetilde{\min}}(\widehat{\Sigma}_2) - \|\widetilde{U}_{21}\| \|\Sigma_1\|_2 \le \|R_2 \widetilde{V}_{31}\|$ and $\|\widetilde{U}_{21}\| \sigma_{\widetilde{\min}}(\widehat{\Sigma}_2) - \|\widetilde{V}_{21}\| \|\Sigma_1\|_2 \le \|\widetilde{U}_{31}^* S_2\|$ (using the fact $\|XY\| \ge \|X\| \sigma_{\widetilde{\min}}(Y)$) to arrive at the same conclusion (27).

Recalling (20) and (21), we obtain

$$
\begin{aligned}
\max(\| \sin \angle(U_1, \widehat{U}_1)\|, \| \sin \angle(V_1, \widehat{V}_1)\|) &= \max \left(\left\| \begin{bmatrix} \widetilde{U}_{21} \\ \widetilde{U}_{31} \end{bmatrix} \right\|, \left\| \begin{bmatrix} \widetilde{V}_{21} \\ \widetilde{V}_{31} \end{bmatrix} \right\| \right) \\
&\le \max(\|\widetilde{U}_{21}\| + \|\widetilde{U}_{31}\|, \|\widetilde{V}_{21}\| + \|\widetilde{V}_{31}\|) \\
&\le \left(1 + \frac{\max(\|R_2\|_2, \|S_2\|_2)}{\delta} \right) \max(\|\widetilde{U}_{31}\|, \|\widetilde{V}_{31}\|) \\
&= \left(1 + \frac{\max(\|R_2\|_2, \|S_2\|_2)}{\delta} \right) \max(\| \sin \angle(U_1, \widehat{U})\|, \| \sin \angle(V_1, \widehat{V})\|),
\end{aligned}
$$

giving (15). In the spectral or Frobenius norm we use the inequality $\left\| \begin{bmatrix} A \\ B \end{bmatrix} \right\|_{2,F} \le \sqrt{\|A\|_{2,F}^2 + \|B\|_{2,F}^2}$ to obtain the slightly tighter bound

$$
\begin{aligned}
\max\{\| \sin \angle(U_1, \widehat{U}_1)\|_{2,F}, \| \sin \angle(V_1, \widehat{V}_1)\|_{2,F}\} &= \max \left(\left\| \begin{bmatrix} \widetilde{U}_{21} \\ \widetilde{U}_{31} \end{bmatrix} \right\|_{2,F}, \left\| \begin{bmatrix} \widetilde{V}_{21} \\ \widetilde{V}_{31} \end{bmatrix} \right\|_{2,F} \right) \\
&\le \max(\sqrt{\|\widetilde{U}_{21}\|_{2,F}^2 + \|\widetilde{U}_{31}\|_{2,F}^2}, \sqrt{\|\widetilde{V}_{21}\|_{2,F}^2 + \|\widetilde{V}_{31}\|_{2,F}^2}) \\
&\le \sqrt{1 + \left(\frac{\max(\|R_2\|_2, \|S_2\|_2)}{\delta} \right)^2} \max(\| \sin \angle(U_1, [\widehat{U}_1 \ \widehat{U}_2])\|_{2,F}, \| \sin \angle(V_1, [\widehat{V}_1 \ \widehat{V}_2])\|_{2,F}).
\end{aligned}
$$

To obtain (16), adding the two inequalities (25) and (26) we have

$$
\|\widetilde{U}_{21}\| + \|\widetilde{V}_{21}\| \le \frac{\|R_2 \widetilde{V}_{31}\| + \|\widetilde{U}_{31}^* S_2\|}{(\sigma_{\widetilde{\min}}(\Sigma_1))^2 - \|\widehat{\Sigma}_2\|_2^2} \le (\|\widetilde{U}_{31}\| + \|\widetilde{V}_{31}\|) \frac{\max(\|R_2\|_2, \|S_2\|_2)}{\delta},
$$
(28)

where as before, the last bound holds in both cases $\delta = \sigma_{\widetilde{\min}}(\Sigma_1) - \|\widehat{\Sigma}_2\|_2$ and $\delta = \sigma_{\widetilde{\min}}(\widehat{\Sigma}_2) - \|\Sigma_1\|_2$. Hence we have

$$\| \sin \angle(U_1, \widehat{U}_1)\| + \| \sin \angle(V_1, \widehat{V}_1)\| = \left\| \begin{bmatrix} \widetilde{U}_{21} \\ \widetilde{U}_{31} \end{bmatrix} \right\| + \left\| \begin{bmatrix} \widetilde{V}_{21} \\ \widetilde{V}_{31} \end{bmatrix} \right\|$$

$$\leq \| \widetilde{U}_{21}\| + \|\widetilde{U}_{31}\| + \|\widetilde{V}_{21}\| + \|\widetilde{V}_{31}\|$$

$$\leq \left(1 + \frac{\max(\|R_2\|_2, \|S_2\|_2)}{\delta}\right) (\|\widetilde{U}_{31}\| + \|\widetilde{V}_{31}\|)$$

$$= \left(1 + \frac{\max(\|R_2\|_2, \|S_2\|_2)}{\delta}\right) \left(\| \sin \angle(U_1, \widehat{U})\| + \| \sin \angle(V_1, \widehat{V})\|\right),$$

completing the proof. □

Theorem 3 shows that the computed singular vectors extracted by the projection method is optimal up to the factor $\sqrt{1 + \left(\frac{\max(\|R_2\|_2, \|S_2\|_2)}{\delta}\right)^2}$ in the spectral norm, and the factor $1 + \frac{\max(\|R_2\|_2, \|S_2\|_2)}{\delta}$ in any unitarily invariant norm. Note the similarity of these factors between Theorems 2 and 3.

The quantity δ in the theorem plays essentially the same role as in the results in Sect. 2, and is the SVD analogue of the spectral gap. In a typical situation where the largest singular values are sought and $m_1 = n_1 = k$, we have $\widehat{\Sigma}_1 \approx \Sigma_1$ with Σ_1 containing the k largest singular values, and all the singular values of $\widehat{\Sigma}_2$ are smaller than $\sigma_k(A) = \sigma_{\widetilde{\min}}(\Sigma_1)$, and thus $\delta = \sigma_{\widetilde{\min}}(\Sigma_1) - \|\widehat{\Sigma}_2\|_2$. Similarly, when the smallest singular values are sought, $\widehat{\Sigma}_1 \approx \Sigma_1$ with Σ_1 containing the k smallest singular values, $\widehat{\Sigma}_2$ is square and all its singular values are larger than $\sigma_{\min(m,n)-k+1}(A) = \|\Sigma_1\|_2$, and thus $\delta = \sigma_{\widetilde{\min}}(\widehat{\Sigma}_2) - \|\Sigma_1\|_2$. It is perhaps worth noting that when $m \neq n$, the gap quantity given in [10, Thm. 2.4] becomes (in our notation) $\delta = \min(\sigma_{\widetilde{\min}}(\widehat{\Sigma}_2) - \|\Sigma_1\|_2, \|\Sigma_1\|_2)$, which is unnecessarily small when $\sigma_{\widetilde{\min}}(\widehat{\Sigma}_2) \gg \|\Sigma_1\|_2$.

3.2 When one-sided projection is used

Some of the state-of-the-art algorithms for an approximate SVD, such as [6,7], rely on one-sided projection methods, instead of two-sided projection as described so far. In this case one would approximate $A \approx (\widehat{U}\widehat{U})\widehat{\Sigma}\widetilde{V}^*$, where $\widetilde{U}\widehat{\Sigma}\widetilde{V}^* = \widehat{U}^*A$ is the SVD of the $\widehat{m} \times n$ matrix, obtained by projecting A only from the left side by $\widehat{U}$ (or from the right by $\widehat{V}$; we discuss left projection for definiteness).

Although one-sided projection is merely a special case of two-sided projection (as it corresponds to taking $\widehat{V} = I_n$), given its practical importance and the simplicities that accrue, we restate the above results when specialized to one-sided projections. In this case we do not have $\widehat{V}_3$, and we start with the equation

$$(\widetilde{A} =) [\widehat{U}_1 \ \widehat{U}_2 \ \widehat{U}_3]^* A[\widehat{V}_1 \ \widehat{V}_2] = \begin{bmatrix} \widehat{\Sigma}_1 & 0 \\ 0 & \widehat{\Sigma}_2 \\ S_1 & S_2 \end{bmatrix}. \tag{29}$$

Corollary 1 *In the notation of Theorem 3, suppose that $span(\widehat{V}) = \mathbb{R}^n$. Then*

$$\| \sin \angle(U_1, \widehat{U}_1)\| + \| \sin \angle(V_1, \widehat{V}_1)\| \leq \left(1 + \frac{\|S_2\|_2}{\delta}\right) \| \sin \angle(U_1, \widehat{U})\| \qquad (30)$$

in any unitarily invariant norm. Furthermore, in the spectral and Frobenius norms,

$$\max(\| \sin \angle(U_1, \widehat{U}_1)\|_{2,F}, \| \sin \angle(V_1, \widehat{V}_1)\|_{2,F}) \leq \sqrt{1 + \frac{\|S_2\|_{2,F}^2}{\delta^2}} \| \sin \angle(U_1, \widehat{U})\|_{2,F}.$$

$$(31)$$

Proof The bounds are obtained essentially by taking $R_i = 0$ in Theorem 3, and noting that $\angle(V_1, \widehat{V}) = 0$ because $\widehat{V}$ spans the entire space $\mathbb{R}^n$. □

Note that the inequality corresponding to (15) becomes strictly weaker than (30), thus omitted.

3.3 Subspace corresponding to smallest singular values

This work is expected to be relevant most in the context of finding an approximate and truncated SVD, in which one approximates the largest singular values and its associated subspaces. Nonetheless, the results obtained here are applicable to any set of singular vectors: for example, Theorem 3 can be used when $\widehat{U}_1, \widehat{V}_1$ approximate the singular subspaces corresponding to certain interior singular values. The same holds for the smallest singular values.

However, practical difficulties arise when one is not looking for the largest singular subspaces. The first difficulty is related to Remark 1 before Sect. 2.2; when computing interior singular values, "spurious" approximate singular values can be very close to an exact sought singular value, while having corresponding singular vectors that are nowhere near the correct ones; this phenomenon would manifest itself as δ being small. As mentioned at the beginning of Sect. 3, using the Jordan–Wieldant matrix introduces extra zero eigenvalues, resulting in unnecessarily small gaps for the smallest singular values—worsening the issue with spurious singular values. We avoid this by working directly with A. To remedy an (unavoidable) small gap, an approach analogous to harmonic Rayleigh–Ritz [10] might be helpful; the analysis of such alternatives would also be of interest.

Another difficulty is that even if the bounds in Theorem 3 are sufficiently small, this does not guarantee that $\widehat{U}_1, \widehat{V}_1$ capture the whole left and right null space of A: specifically, U_1, V_1 in the theorems above may merely be a subset of the null spaces, of the same size as $\widehat{U}_1, \widehat{V}_1$. Unfortunately there appears to be no easy way to check that $\widehat{U}_1, \widehat{V}_1$ contains the entire desired subspace (contrast this with when $\widehat{U}_1, \widehat{V}_1$ approximate the largest singular subspaces, in which case $A - \widehat{U}_1 \widehat{\Sigma}_1 \widehat{V}_1$ being small indicates $\widehat{U}_1, \widehat{V}_1$ contain the desired subspaces).

3.3.1 Finding null space

One situation that Theorem 3 does not cover is when the null space is desired of a (fat) rectangular matrix. For definiteness, suppose $A \in \mathbb{C}^{m \times n}$ with $m < n$ and one wants to compute the $n - m$ null vectors such that $AV = 0$. Then in the notation of Theorem 3, we would like to take Σ_1 to be the "$0 \times (n - m)$" matrix of zeros, which the theorem does not account for.

In this case we can modify the argument as follows.

Proposition 1 *Let $A \in \mathbb{C}^{m \times n}$ with $m < n$, and let $\widetilde{A}, \widehat{\Sigma}_i, \widehat{U}_i, \widehat{V}_i, R_i, S_i, A_3$ be as defined in (3). Let $V_1 \in \mathbb{C}^{n \times k}$ with $k \leq n - m$, $k \leq m_1, n_1$ be a null space of A such that $AV_1 = 0$. Then*

$$\| \sin \angle(V_1, \widehat{V}_1)\| \leq \left(1 + \frac{\|R_2\|_2)}{\sigma_{\min}(\widehat{\Sigma}_2)} \right) \| \sin \angle(V_1, \widehat{V})\|, \tag{32}$$

and

$$\| \sin \angle(V_1, \widehat{V}_1)\|_{2,F} \leq \sqrt{1 + \left(\frac{\|R_2\|_2}{\sigma_{\min}(\widehat{\Sigma}_2)} \right)^2} \| \sin \angle(V_1, \widehat{V})\|_{2,F}. \tag{33}$$

Proof We have

$$\begin{bmatrix} \widehat{\Sigma}_1 & 0 & R_1 \\ 0 & \widehat{\Sigma}_2 & R_2 \\ S_1 & S_2 & A_3 \end{bmatrix} \begin{bmatrix} \widetilde{V}_{11} \\ \widetilde{V}_{21} \\ \widetilde{V}_{31} \end{bmatrix} = 0, \tag{34}$$

where $\begin{bmatrix} \widetilde{V}_{11} \\ \widetilde{V}_{21} \\ \widetilde{V}_{31} \end{bmatrix} = [\widehat{V}_1 \ \widehat{V}_2 \ \widehat{V}^\perp]^* V_1$. The goal as before is to bound $\frac{\| \sin \angle(V_1, \widehat{V}_1)\|}{\| \sin \angle(V_1, \widehat{V})\|} = \left\| \begin{bmatrix} \widetilde{V}_{21} \\ \widetilde{V}_{31} \end{bmatrix} \right\| / \|\widetilde{V}_{31}\|$.

The second block of (34) gives $\widehat{\Sigma}_2 \widetilde{V}_{21} + R_2 \widetilde{V}_{31} = 0$, from which we obtain (we require that $\widehat{\Sigma}_2$ is square)

$$\|\widetilde{V}_{21}\| \leq \frac{\|R_2\|}{\sigma_{\min}(\widehat{\Sigma}_2)} \|\widetilde{V}_{31}\|.$$

Thus we obtain (32) and (33) using the same argument as in Theorem 3 as required. □

Appendix: When eigenvalues and Ritz values are interlacing

We revisit the eigenspace analysis in Sect. 2.2. As mentioned after Theorem 2, in the Frobenius norm, one can relax the assumption on the eigenvalues $\lambda(\Lambda_1)$ and Ritz values $\lambda(\widehat{\Lambda}_2)$, to allow them to interlace.

Proposition 2 *Let A be a Hermitian matrix, and let $\widetilde{A}$ be as in (5); $(\widehat{\Lambda}_1, \widehat{X}_1)$ is a set of Ritz pairs with $\widehat{\Lambda}_1 \in \mathbb{R}^{n_1 \times n_1}$. Let (Λ_1, X_1) with $\Lambda_1 \in \mathbb{R}^{k \times k}$, $k \leq n_1$ be a set of exact eigenpairs such that $\delta = \min |\lambda(\widehat{\Lambda}_2) - \lambda(\Lambda_1)| > 0$, where $\widehat{\Lambda}_2$ is as in (5). Then*

$$\left\| \sin \angle(X_1, \widehat{X}_1) \right\| \leq \left\| \sin \angle(X_1, \widehat{X}) \right\| \left(1 + \frac{\pi}{2} \frac{\|R_2\|_2}{\delta} \right), \tag{35}$$

$$\left\| \sin \angle(X_1, \widehat{X}_1) \right\|_2 \leq \left\| \sin \angle(X_1, \widehat{X}) \right\|_2 \sqrt{1 + \frac{\pi^2}{4} \frac{\|R_2\|^2}{\delta^2}}, \tag{36}$$

$$\left\| \sin \angle(X_1, \widehat{X}_1) \right\|_F \leq \left\| \sin \angle(X_1, \widehat{X}) \right\|_F \sqrt{1 + \frac{\|R_2\|_2^2}{\delta^2}}. \tag{37}$$

Proof The key too is the bound ([3], see also [14, Lem. 2.1]) for the Sylvester equation $c\|AX - XB\| \geq \|X\| \min |\lambda(A) - \lambda(B)|$, where $c \leq \frac{\pi}{2}$, and $c = 1$ in the Frobenius norm. Using this in (10) we obtain

$$\|R_2^* \widetilde{X}_3\| = \|\widehat{\Lambda}_2 \widetilde{X}_2 - \widetilde{X}_2 \Lambda_1\| \geq \frac{1}{c} \|\widetilde{X}_2\| \min |\lambda(\widehat{\Lambda}_2) - \lambda(\Lambda_1)| = \frac{\delta}{c} \|\widetilde{X}_2\|.$$

Hence, using $\|XY\| \leq \|X\|_2 \|Y\|$ [8, p. 327] as before, we obtain $\|\widetilde{X}_2\| \leq c \frac{\|R_2\|_2}{\delta} \|\widetilde{X}_3\|$. Therefore, recalling (8) and (9) we obtain

$$\|\sin \angle(X_1, \widehat{X}_1)\| = \left\| \begin{bmatrix} \widetilde{X}_2 \\ \widetilde{X}_3 \end{bmatrix} \right\| = \|\widetilde{X}_2\| + \|\widetilde{X}_3\| \leq \left(1 + c\left(\frac{\|R_2\|_2}{\delta} \right) \right) \|\widetilde{X}_3\|$$

$$\leq \left(1 + c\left(\frac{\|R_2\|_2}{\delta} \right) \right) \|\sin \angle(X_1, \widehat{X})\|,$$

which gives (35) by taking $c = \pi/2$. Moreover,

$$\|\sin \angle(X_1, \widehat{X}_1)\|_{2,F}^2 = \left\| \begin{bmatrix} \widetilde{X}_2 \\ \widetilde{X}_3 \end{bmatrix} \right\|_{2,F}^2 \leq \|\widetilde{X}_2\|_{2,F}^2 + \|\widetilde{X}_3\|_{2,F}^2$$

$$\leq \left(1 + c\left(\frac{\|R_2\|_2}{\delta} \right)^2 \right) \|\widetilde{X}_3\|_{2,F}$$

$$= \left(1 + c\left(\frac{\|R_2\|_2}{\delta} \right)^2 \right) \|\sin \angle(X_1, \widehat{X})\|_{2,F}^2.$$

Taking $c = \frac{\pi}{2}$ for the spectral norm gives (36), and $c = 1$ for the Frobenius norm we obtain (37).

$\square$

References

1. Bai, Z., Demmel, J., Dongarra, J., Ruhe, A., van der Vorst, H.: Templates for the Solution of Algebraic Eigenvalue Problems: A Practical Guide. SIAM, Philadelphia (2000)
2. Beattie, C.: Galerkin eigenvector approximations. Math. Comp. **69**, 1409–1434 (2000)
3. Bhatia, R., Davis, C., McIntosh, A.: Perturbation of spectral subspaces and solution of linear operator. Linear Algebra Appl. **52**, 45–67 (1983)
4. Davis, C., Kahan, W.M.: The rotation of eigenvectors by a perturbation. III. SIAM J. Numer. Anal. **7**(1), 1–46 (1970)
5. Golub, G.H., Van Loan, C.F.: Matrix Computations, 4th edn. The Johns Hopkins University Press, Baltimore (2012)
6. Gu, M.: Subspace Iteration Randomization and Singular Value Problems. arXiv preprint arXiv:1408.2208 (2014)
7. Halko, N., Martinsson, P.-G., Tropp, J.A.: Finding structure with randomness: probabilistic algorithms for constructing approximate matrix decompositions. SIAM Rev. **53**(2), 217–288 (2011)
8. Higham, N.J.: Functions of Matrices: Theory and Computation. SIAM, Philadelphia (2008)
9. Hochstenbach, M.E.: A Jacobi–Davidson type SVD method. SIAM J. Sci. Comp. **23**(2), 606–628 (2001)
10. Hochstenbach, M.E.: Harmonic and refined extraction methods for the singular value problem, with applications in least squares problems. BIT **44**(4), 721–754 (2004)
11. Horn, R.A., Johnson, C.R.: Matrix Analysis, 2nd edn. Cambridge University Press, Cambridge (2012)
12. Knyazev, A.V.: New estimates for Ritz vectors. Math. Comp. **66**(219), 985–995 (1997)
13. Li, C.-K., Li, R.-C.: A note on eigenvalues of perturbed Hermitian matrices. Linear Algebra Appl. **395**, 183–190 (2005)
14. Li, R.-C., Zhang, L.-H.: Convergence of the block Lanczos method for eigenvalue clusters. Numer. Math. **131**(1), 83–113 (2015)
15. Mathias, R.: Quadratic residual bounds for the Hermitian eigenvalue problem. SIAM J. Matrix Anal. Appl. **19**(2), 541–550 (1998)
16. Morgan, R.B.: Computing interior eigenvalues of large matrices. Linear Algebra Appl. **154–156**, 289–309 (1991)
17. Nakatsukasa, Y.: Algorithms and Perturbation Theory for Matrix Eigenvalue Problems and the Singular Value Decomposition. PhD thesis, University of California, Davis (2011)
18. Paige, C.C., Saunders, M.A.: Towards a generalized singular value decomposition. SIAM J. Numer. Anal. **18**(3), 398–405 (1981)
19. Parlett, B.N.: The Symmetric Eigenvalue Problem. SIAM, Philadelphia (1998)
20. Saad, Y.: Numerical Methods for Large Eigenvalue Problems, 2nd edn. SIAM, Philadelphia (2011)
21. Stewart, G.W.: A generalization of Saad's theorem on Rayleigh–Ritz approximations. Linear Algebra Appl. **327**, 115–119 (1999)
22. Stewart, G.W.: Matrix Algorithms Volume II: Eigensystems. SIAM, Philadelphia (2001)
23. Stewart, G.W., Sun, J.-G.: Matrix Perturbation Theory (Computer Science and Scientific Computing). Academic Press, London (1990)
24. Wedin, P.: Perturbation bounds in connection with singular value decomposition. BIT **12**, 99–111 (1972)