MATHEMATICAL ENGINEERING
TECHNICAL REPORTS

Accuracy of Singular Vectors Obtained by Projection-based SVD Methods

Yuji NAKATSUKASA

METR 2016–06 April 2016

DEPARTMENT OF MATHEMATICAL INFORMATICS
GRADUATE SCHOOL OF INFORMATION SCIENCE AND TECHNOLOGY
THE UNIVERSITY OF TOKYO
BUNKYO-KU, TOKYO 113-8656, JAPAN

WWW page: http://www.keisu.t.u-tokyo.ac.jp/research/techrep/index.html

Accuracy of singular vectors obtained by projection-based SVD methods

Yuji Nakatsukasa

Abstract

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 smaller-dimensional subspaces $\widehat{U} \in \mathbb{C}^{m\times m_1}, \widehat{V} \in \mathbb{C}^{n\times n_1}$ (allowing $m_1 = m$ or $n_1 = n$) to obtain a small $m_1 \times n_1$ matrix $\widetilde{A} := \widehat{U}^* A\widehat{V}$, compute the SVD $\widetilde{A} = \widetilde{U}\widehat{\Sigma}\widetilde{V}^*$, and approximate $A \approx (\widehat{U}\widetilde{U})\widehat{\Sigma}(\widehat{V}\widetilde{V})^*$. The columns in $\widehat{U}\widetilde{U}$ and $\widehat{V}\widetilde{V}$ are approximate left and right singular vectors. Similar projection methods are commonly employed for approximately computing a subset of singular vectors, not necessarily corresponding to the largest singular values. In this work we assess the validity of such projection methods from the viewpoint of singular vectors. Specifically, we give error bounds for a set of approximate singular vectors obtained by projection, i.e., a subset of columns of $\widehat{U}\widetilde{U}$ and $\widehat{V}\widetilde{V}$. The projection methods are an analogue of the Rayleigh-Ritz process for the symmetric eigenvalue problem, for which existing results are available for the accuracy of a set of computed eigenvectors. This work derives their SVD analogues, and show that as in Rayleigh-Ritz, the projection method for the SVD gives approximate singular subspaces that are optimal usually to within a certain constant factor.

1 Introduction

The SVD plays a central role in a vast variety of applications, primarily due to its best low-rank approximation property. Often the matrix is too large to compute the full SVD, and computing a partial and approximate SVD $A \approx U_1\Sigma_1V_1^*$ 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 smaller-dimensional subspaces $\widehat{U} \in \mathbb{C}^{m\times m_1}, \widehat{V} \in \mathbb{C}^{n\times n_1}$, compute the SVD of the small $m_1 \times n_1$ matrix $\widetilde{A} := \widehat{U}^* A\widehat{V} = \widetilde{U}\widehat{\Sigma}\widetilde{V}^*$ and obtain an apprxoimate tall-skinny SVD as $A \approx (\widehat{U}\widetilde{U})\widehat{\Sigma}(\widehat{V}\widetilde{V})^*$. This process is an analogue of the Rayleigh-Ritz process for the symmetric eigenvalue problem, sometimes called the Petrov-Galerkin method [1]. 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 [14, Thm. 4.6], Knyazev [7, Thm. 4.3] and Stewart [15, Thm. 2]. In a nutshell, these results show that the Rayleigh-Ritz process extracts an approximate set of eigenvectors that are optimal (given the subspace) up to a certain constant factor, which depends on the residual norm and a gap between exact eigenvalues and Ritz values (i.e., approximate eigenvalues).

This work derives analogous results for the SVD by establishing bounds for the computed left and right singular vectors. In essense, the message is the same as in the Rayleigh-Ritz process for the symmetric eigenvalue problem: 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. Somewhat surprisingly (given the importance of SVD and the prominence of projection-based algorithms for computing an approximate SVD), such extension appears to be unavailable in the literature. A preliminary version has appeared in the author's PhD dissertation [11, Ch. 10].

Let us clarify the problem formulation. Let A be an $m \times n$ matrix and $\widehat{U} \in \mathbb{C}^{m\times m_1}, \widehat{V} \in \mathbb{C}^{n\times n_1}$ be subspaces that are hoped to approximately contain a set of $k \leq \min(m_1, n_1)$ (exact) singular vectors U_1 and V_1, corresponding to some of the singular values, usually but not necessarily the largest ones. 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 instead of 2 for consistency with what follows)

$$(\widetilde{A} =)\ [\widehat{U}_1\ \widehat{U}_3]^* A [\widehat{V}_1\ \widehat{V}_3] = \begin{bmatrix} \widehat{\Sigma}_1 & R \\ S & A_3 \end{bmatrix}. \tag{1.1}$$

In practice we 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 $\|\widetilde{S}\| = \|A\widehat{V}_1 - \widetilde{U}_1\widehat{\Sigma}_1\|$ and $\|R\| = \|\widehat{U}_1^* A - \widehat{\Sigma}_1\widehat{V}_1^*\|$, which hold for any unitarily invariant norm. 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 $\widehat{U}, \widehat{V}$ correspond to exact singular subspaces. When R, S are both small we can expect $\widehat{U}_i, \widehat{V}_i$ to be good approximations to some singular vectors.

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 smaller dimension (this is the case e.g. in the Jacobi-Davidson context [6], 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 U_1 (which is usually of lower dimension k than m_1), then the projection method computes a good approximant $\widehat{U}_1$ to U_1, where $\widehat{U}_1$ and U_1 are both k-dimensional.

From the algorithmic viewpoint, we compute the projection $\widetilde{A} = \widehat{U}^* A \widehat{V}$ and its SVD

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

and defining $\widehat{U}_i = \widehat{U}\widetilde{U}_i$ and $\widehat{V}_i = \widehat{V}\widetilde{V}_i$ for $i = \{1, 2\}$, we have

$$(\widetilde{A} =)\ [\widehat{U}_1\ \widehat{U}_2\ \widehat{U}_3]^* A [\widehat{V}_1\ \widehat{V}_2\ \widehat{V}_3] = \begin{bmatrix} \widehat{\Sigma}_1 & 0 & R_1 \\ 0 & \widehat{\Sigma}_2 & R_2 \\ S_1 & S_2 & A_3 \end{bmatrix}. \tag{1.2}$$

The goal here is to bound the angles $\angle(U_1, \widehat{U}_1)$ and $\angle(V_1, \widehat{V}_1)$, the accuracy of the computed subspaces as approximants to exact singular subspaces, as compared with $\angle(U_1, \widehat{U})$ and $\angle(V_1, \widehat{V})$ respectively. Since $\mathrm{span}(\widehat{U}_1) \subseteq \mathrm{span}(\widehat{U})$ we have $\angle(U_1, \widehat{U}) \leq \angle(U_1, \widehat{U}_1)$ (similarly for the V's), the idea is that the projection method is regarded near-optimal if $\angle(U_1, \widehat{U}_1) \leq c\angle(U_1, \widehat{U})$ holds with a modest constant c.

We recall the definition of angles between subspaces. The angles θ_i between two subspaces spanned by $X \in \mathbb{C}^{m \times n_X}, Y \in \mathbb{C}^{m \times n_Y}$ $(n_X \leq n_Y)$ with orthonormal columns are defined by $\theta_i = \mathrm{acos}(\sigma_i(X^*Y))$; they are known as the canonical angles or principal angles [2, Thm. 2.5.3]. These are connected to the CS decomposition for orthogonal matrices [2, Thm. 2.5.3], [12], in which $C = \mathrm{diag}(\cos\theta_1, \ldots, \cos\theta_{n_X}), S = \mathrm{diag}(\sin\theta_1, \ldots, \sin\theta_{n_X})$, and we write $\angle(X, Y) = \mathrm{diag}(\theta_1, \ldots, \theta_{n_X})$.

By a unitary transformation we see that $\angle(\widehat{U}_1, U_1) = \angle(\widetilde{U}_1, \left[\begin{smallmatrix} I_k \\ 0 \end{smallmatrix}\right])$ and $\angle(\widehat{V}_1, V_1) = \angle(\widetilde{V}_1, \left[\begin{smallmatrix} I_k \\ 0 \end{smallmatrix}\right])$, where $\widetilde{U}_1, \widetilde{V}_1$ are matrices of exact singular vectors of $\widetilde{A}$ in (1.2). We note that k is an integer that one can choose in $1 \leq k \leq \min(m_1, n_1)$, and the bounds we derive are applicable to any choice.

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

Notation. $\widehat{U}, \widehat{V}$ always denote the left and right 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. A matrix norm $\|\cdot\|$ without subscripts denotes any unitariliy invariant norm, and $\|\cdot\|_2$ is 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 $\|\cdot\|_{2,F}$ indicates they hold for both the spectral norm and the Frobenius norm (not a mixture). $\sigma_i(A)$ denotes the ith largest singular value of A. $\lambda(A)$ denotes the set of eigenvalues of A.

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

The standard way of computing a part of the eigenvalues and eigenvectors of a large-sparse Hermitian (or real symmetric) matrix A is to form a low-dimensional subspace represented by $\mathrm{span}(\widehat{X})$ with $\widehat{X}$ 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*. This process computes the eigendecomposition of the small Hermitian matrix $\widehat{X}^* A \widehat{X} = Y \widehat{\Lambda} Y^*$, from which the Ritz values are taken as the diagonals of $\widehat{\Lambda}$ and the Ritz vectors as the columns of $\widehat{X}Y$. The Ritz pairs $(\widehat{\lambda}, \widehat{x})$ thus obtained satisfy $\widehat{x} \in \mathrm{span}(\widehat{X})$, and $A\widehat{x} - \widehat{\lambda}\widehat{x} \perp \widehat{X}$.

A natural question is to ask how accurate the Ritz pairs are as an approximation to the exact eigenpairs. The starting point is an analogue of (1.2): the partitioning of the projected

A as

$$(\widetilde{A} =)[\widehat{X}_1\ \widehat{X}_2\ \widehat{X}_3]^* A[\widehat{X}_1\ \widehat{X}_2\ \widehat{X}_3] = \begin{bmatrix} \widehat{\Lambda}_1 & 0 & R_1^* \\ 0 & \widehat{\Lambda}_2 & R_2^* \\ R_1 & R_2 & A_3 \end{bmatrix}, \tag{2.1}$$

where $[\widehat{X}_1\ \widehat{X}_2\ \widehat{X}_3]$ is an orthogonal matrix. 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, $\widehat{X} = [\widehat{X}_1\ \widehat{X}_2]$ 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 (or more precisely, invariant subspace) of A.

We note that bounding the eigenvalue accuracy can be done using standard eigenvalue perturbation theory: by Weyl's theorem, the Ritz values Λ_i match some of those of A to $\|R_i\|_2$ 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 [8, 9, 13] one can often get tighter bounds for the eigenvalue accuracy, which scale quadratically with the norm of the residual.

Now we turn to eigenvectors and discuss the accuracy of the Ritz vectors. In the 1-dimensional $k = 1$ case, we bound the angle between approximate and exact eigenvectors $\widehat{x}$ and x. Saad [14, Thm. 4.6] proves the following theorem, applicable to (2.1) 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 2.1 *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 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*

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

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

Recalling that $\sin\angle(x, \widehat{x}) \le \sin\angle(x, \widehat{X})$ holds trivially because $x \in \operatorname{span}(\widehat{X})$, we see that the above theorem claims that the Ritz vector is optimal up to the constant $\sqrt{1 + \frac{\|r\|_2^2}{\delta^2}}$. This does not 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 [10] and [16, § 5.1.4].

2.1 Knyazev and Stewart's bounds

Knyazev [7, Thm. 4.3] derived an extension of Saad's result (Theorem 2.1) to approximate eigenspaces, presenting bounds in the spectral norm applicable in the context of linear operators in a Hilbert space. Stewart [15, 16] proves an analogous result applicable to non-Hermitian matrices.

Here we state and prove a result specializing to Hermitian matrices, slightly improving existing results in two ways: First, the term $\|R_2\|$ in the bounds (2.2), (2.3) are $\|[R_1\ R_2]\|$ in the literature. Second, the bound (2.2) for any unitarily invariant norm appears to be new.

Theorem 2.2 *Let A be a Hermitian matrix. Let $(\widehat{\Lambda}_1, \widehat{X}_1)$ be a set of Ritz pairs as in* (2.1) *with $\widehat{\Lambda}_1 \in \mathbb{R}^{k\times k}$. Let (Λ_1, X_1) with $\Lambda_1 \in \mathbb{R}^{k\times k}$ be a set of exact eigenpairs of 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 $\widehat{\Lambda}_2$ is as in* (2.1) *Then for any unitarily invariant norm*

$$\left\|\sin\angle(X_1, \widehat{X}_1)\right\| \leq \left\|\sin\angle(X_1, \widehat{X})\right\| \left(1 + \frac{\|R_2\|_2}{\delta}\right), \tag{2.2}$$

and for the spectral and Frobenius norms,

$$\left\|\sin\angle(X_1, \widehat{X}_1)\right\|_{2,F} \leq \left\|\sin\angle(X_1, \widehat{X})\right\|_{2,F} \sqrt{1 + \frac{\|R_2\|_2^2}{\delta^2}}. \tag{2.3}$$

Note that when applicable, (2.3) is slightly tighter than (2.2).

PROOF. Let $\widetilde{X} = \begin{bmatrix} \widetilde{X}_1 \\ \widetilde{X}_2 \\ \widetilde{X}_3 \end{bmatrix}$ be a set of eigenvectors $\widetilde{A}$ with orthonormal columns. Note by the CS decomposition that we have

$$\|\sin\angle(X_1, \widehat{X})\| = \|\sigma_i((X_1^\perp)^* \widehat{X})\| = \left\|\left[\widetilde{X}_3\right]\right\| \tag{2.4}$$

and

$$\|\sin\angle(X_1, \widehat{X}_1)\| = \|\sigma_i((X_1^\perp)^* \widehat{X}_1)\| = \left\|\begin{bmatrix} \widetilde{X}_2 \\ \widetilde{X}_3 \end{bmatrix}\right\|. \tag{2.5}$$

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

$$\widehat{\Lambda}_2 \widetilde{X}_2 - \widetilde{X}_2 \Lambda_1 = -R_2^* \widetilde{X}_3. \tag{2.6}$$

We now use the following fact [17, p. 251]: Let A and B be square matrices such that $1/\|A^{-1}\| - \|B\| = \delta > 0$. Define $C = AX - XB$. Then we must have $\|X\| \leq \frac{\|C\|}{\delta}$ (many results in eigenvector perturbation theory use this fact [17]).

Using this fact in (2.6), we have

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

Hence using the fact [5, p. 327] that $\|XY\| \leq \|X\|_2 \|Y\|$ for any unitarily invariant norm we have

$$\|\widetilde{X}_2\| \leq \frac{\|R_2\|_2}{\delta} \|\widetilde{X}_3\|. \tag{2.7}$$

Therefore, recalling (2.4) and (2.5) we have

$$\begin{aligned}\|\sin\angle(X_1,\widehat{X}_1)\| &= \left\|\begin{bmatrix}\widetilde{X}_2\\ \widetilde{X}_3\end{bmatrix}\right\| \le \|\widetilde{X}_2\|+\|\widetilde{X}_3\| \\ &\le (1+\frac{\|R_2\|_2}{\delta})\|\widetilde{X}_3\| = (1+\frac{\|R_2\|_2}{\delta})\|\sin\angle(X_1,\widehat{X})\|, \qquad (2.8)\end{aligned}$$

which is (2.2). In the spectral or Frobenius norm we can use the inequality $\|\left[\begin{smallmatrix}A\\B\end{smallmatrix}\right]\|_{2,F}\le\sqrt{\|A\|_{2,F}^2+\|B\|_{2,F}^2}$; 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

$$\begin{aligned}\|\sin\angle(X_1,\widehat{X}_1)\|_{2,F} &\le \sqrt{\|\widetilde{X}_2\|_{2,F}^2+\|\widetilde{X}_3\|_{2,F}^2} = \sqrt{1+\frac{\|R_2\|_2^2}{\delta^2}}\|\widetilde{X}_3\|_{2,F} \\ &= \sqrt{1+\frac{\|R_2\|_2^2}{\delta^2}}\|\sin(X_1,\widehat{X})\|_{2,F}, \qquad (2.9)\end{aligned}$$

which is (2.3). □

It is perhaps worth emphasizing that the "gap" δ is the gap between some of the Ritz values and 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 [8, 9]. The same holds for the results for the SVD as we describe next.

3 Accuracy bounds for approximate singular subspaces

We now turn to the main subject SVD and derive accuracy bounds for approximate singular vectors or singular subspaces (just as in the symmetric eigenproblem, accuracy of singular values can be bounded using standard perturbation theory; most importantly Weyl's bound; bounds that scale quadratically with the residual are also known [8]). We derive the SVD analogues of Theorem 2.2 by employing the technique used in the proof given above.

3.1 Main result

Let $\widehat{U}_i,\widehat{V}_i,\widehat{\Sigma}_i,R_i,S_i,A_3$ be as defined in (1.2), and let $\left(\Sigma_1,\widetilde{U}_1,\widetilde{V}_1\right)$ be exact singular triplets of $\tilde{A}$ and let $\widetilde{V}_1=\begin{bmatrix}\widetilde{V}_{11}\\ \widetilde{V}_{21}\\ \widetilde{V}_{31}\end{bmatrix}, \widetilde{U}_1=\begin{bmatrix}\widetilde{U}_{11}\\ \widetilde{U}_{21}\\ \widetilde{U}_{31}\end{bmatrix}$. We shall prove the following, which is the main result of this paper.

Theorem 3.1 *Let $A\in\mathbb{C}^{m\times n}$. Let $\widehat{U}=[\widehat{U}_1\ \widehat{U}_2]\in\mathbb{C}^{m\times m_1}$ and $\widehat{V}=[\widehat{V}_1\ \widehat{V}_2]\in\mathbb{C}^{n\times n_1}$ have orthonormal columns with $\widehat{U}_1\in\mathbb{C}^{m\times k}$ and $\widehat{V}_1\in\mathbb{C}^{n\times k}$, and let $\widetilde{A},\widehat{\Sigma}_i,R_i,S_i,A_3$ be as defined* *in* (1.2). *Let*

$$A = [U_1\ U_2\ U_3] \begin{bmatrix} \Sigma_1 & & \\ & \Sigma_2 & \\ & & \Sigma_3 \end{bmatrix} [V_1\ V_2\ V_3]^* \tag{3.1}$$

be an (exact) SVD, with singular values arranged in an arbitrary (not necessarily decreasing) order. Define

$$\delta = \max(\sigma_{\min}(\Sigma_1) - \|\widehat{\Sigma}_2\|_2, \sigma_{\min}(\widehat{\Sigma}_2) - \|\Sigma_1\|_2). \tag{3.2}$$

If $\delta > 0$, *then*

$$\begin{aligned} &\max(\|\sin\angle(U_1,\widehat{U}_1)\|, \|\sin\angle(V_1,\widehat{V}_1)\|) \\ &\quad \le \left(1 + \frac{\max(\|R_2\|_2, \|S_2\|_2)}{\delta}\right) \max(\|\sin\angle(U_1,\widehat{U})\|, \|\sin\angle(V_1,\widehat{V})\|) \end{aligned} \tag{3.3}$$

and

$$\begin{aligned} &\|\sin\angle(U_1,\widehat{U}_1)\| + \|\sin\angle(V_1,\widehat{V}_1)\| \\ &\quad \le \big(1 + \frac{\max(\|R_2\|_2, \|S_2\|_2)}{\delta}\big)\big(\|\sin\angle(U_1,\widehat{U})\| + \|\sin\angle(V_1,\widehat{V})\|\big) \end{aligned} \tag{3.4}$$

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

$$\begin{aligned} &\max(\|\sin\angle(U_1,\widehat{U}_1)\|_{2,F}, \|\sin\angle(V_1,\widehat{V}_1)\|_{2,F}) \\ &\le \sqrt{1 + \frac{\max(\|R_2\|_2, \|S_2\|_2)^2}{\delta^2}} \max(\|\sin\angle(U_1,\widehat{U})\|_{2,F}, \|\sin\angle(V_1,\widehat{V})\|_{2,F}), \end{aligned} \tag{3.5}$$

PROOF. As in (2.4), (2.5) we have

$$\|\sin\angle(U_1,\widehat{U}_1)\| = \left\| \begin{bmatrix} \widetilde{U}_{21} \\ \widetilde{U}_{31} \end{bmatrix} \right\|, \quad \|\sin\angle(V_1,\widehat{V}_1)\| = \left\| \begin{bmatrix} \widetilde{V}_{21} \\ \widetilde{V}_{31} \end{bmatrix} \right\|, \tag{3.6}$$

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

$$\|\sin\angle(U_1,\widehat{U})\| = \|\widetilde{U}_{31}\|, \quad \|\sin\angle(V_1,\widehat{V})\| = \|\widetilde{V}_{31}\|. \tag{3.7}$$

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

From the second block of $\begin{bmatrix} \widehat{\Sigma}_1 & 0 & R_1 \\ 0 & \widehat{\Sigma}_2 & R_2 \\ S_1 & S_2 & A_3 \end{bmatrix} \begin{bmatrix} \widetilde{V}_{11} \\ \widetilde{V}_{21} \\ \widetilde{V}_{31} \end{bmatrix} = \begin{bmatrix} \widetilde{U}_{11} \\ \widetilde{U}_{21} \\ \widetilde{U}_{31} \end{bmatrix} \Sigma_1$ we obtain

$$\widehat{\Sigma}_2 \widetilde{V}_{21} + R_2 \widetilde{V}_{31} = \widetilde{U}_{21} \Sigma_1, \tag{3.8}$$

and also from the second block of $\begin{bmatrix} \widetilde{U}_{11}^* & \widetilde{U}_{21}^* & \widetilde{U}_{31}^* \end{bmatrix} \begin{bmatrix} \widehat{\Sigma}_1 & 0 & R_1 \\ 0 & \widehat{\Sigma}_2 & R_2 \\ S_1 & S_2 & A_3 \end{bmatrix} = \Sigma_1 \begin{bmatrix} \widetilde{V}_{11}^* & \widetilde{V}_{21}^* & \widetilde{V}_{31}^* \end{bmatrix}$ we get

$$\widetilde{U}_{21}^* \widehat{\Sigma}_2 + \widetilde{U}_{31}^* S_2 = \Sigma_1 \widetilde{V}_{21}^*. \tag{3.9}$$

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 (3.8) and (3.9) we obtain

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

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 assuming that $\sigma_{\min}(\Sigma_1) > \|\widehat{\Sigma}_2\|_2$ we get

$$\|\widetilde{U}_{21}\| \le \frac{\sigma_{\min}(\Sigma_1) \|R_2 \widetilde{V}_{31}\| + \|\widehat{\Sigma}_2\|_2 \|\widetilde{U}_{31}^* S_2\|}{(\sigma_{\min}(\Sigma_1))^2 - \|\widehat{\Sigma}_2\|_2^2}. \tag{3.10}$$

We can similarly obtain

$$\|\widetilde{V}_{21}\| \le \frac{\sigma_{\min}(\Sigma_1) \|\widetilde{U}_{31}^* S_2\| + \|\widehat{\Sigma}_2\|_2 \|R_2 \widetilde{V}_{31}\|}{(\sigma_{\min}(\Sigma_1))^2 - \|\widehat{\Sigma}_2\|_2^2}. \tag{3.11}$$

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

$$\begin{aligned} \max(\|\widetilde{U}_{21}\|, \|\widetilde{V}_{21}\|) &\le \frac{\max(\|\widetilde{U}_{31}^* S_2\|, \|R_2 \widetilde{V}_{31}\|)}{\sigma_{\min}(\Sigma_1) - \|\widehat{\Sigma}_2\|_2} \\ &\le \frac{\max(\|\widetilde{U}_{31}^*\|, \|\widetilde{V}_{31}\|) \max(\|R_2\|_2, \|S_2\|_2)}{\sigma_{\min}(\Sigma_1) - \|\widehat{\Sigma}_2\|_2} \\ &= \frac{\max(\|R_2\|_2, \|S_2\|_2)}{\delta} \max(\|\widetilde{U}_{31}^*\|, \|\widetilde{V}_{31}\|). \end{aligned} \tag{3.12}$$

Now if $\delta = \sigma_{\min}(\widehat{\Sigma}_2) - \|\Sigma_1\|_2 > 0$, we repeat the above analysis to arrive at the same conclusion (3.12).

Recalling (3.6) and (3.7), we obtain

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

giving (3.3). In the spectral or Frobenius norm we use the inequality $\| \left[\begin{smallmatrix} A \\ B \end{smallmatrix}\right] \|_2 \le \sqrt{\|A\|_2^2 + \|B\|_2^2}$ to obtain the slightly tighter bound

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

To obtain (3.4), adding the two inequalities (3.10) and (3.11) we have

$$\|\widetilde{U}_{21}\| + \|\widetilde{V}_{21}\| \le \frac{\|R_2\widetilde{V}_{31}\| + \|\widetilde{U}_{31}^* S_2\|}{(\sigma_{\min}(\Sigma_1))^2 - \|\widehat{\Sigma}_2\|_2^2} \le (\|\widetilde{U}_{31}\| + \|\widetilde{V}_{31}\|)\frac{\max(\|R_2\|_2, \|S_2\|_2)}{\delta}, \tag{3.13}$$

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

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

completing the proof. □

Theorem 3.1 shows that the computed singular vectors obtained by the projection method is optimal up to a factor $\sqrt{1 + \left(\frac{\max(\|R_2\|_2, \|S_2\|_2)}{\delta}\right)^2}$ in the spectral norm, and a factor $1 + \frac{\max(\|R_2\|_2, \|S_2\|_2)}{\delta}$ in any unitarily invariant norm. Note the similarity of these factor between Theorems 2.2 and 3.1.

3.2 When one-sided projection is used

Some of the state-of-the-art algorithms for an approximate SVD, such as [3, 4], rely on one-sided projection methods, instead of two-sided projection as described so far. In this case one would approximate $A \approx (\widehat{U}\widetilde{U})\widehat{\Sigma}\widetilde{V}^*$, where $\widetilde{U}\widehat{\Sigma}\widetilde{V}^* = \widehat{U}^* A$ is the SVD of the $m_1 \times n$ matrix, obtained by projecting A only from the left side (compare with the discussion in the introduction) by $\widehat{U}$.

Although one-sided projection is merely a special case of two-sided projection (as it corresponds to taking $\widehat{V} = I_n$), because of 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

$$(\widetilde{A} =)\ [\widehat{U}_1\ \widehat{U}_2\ \widehat{U}_3]^* A[\widehat{V}_1\ \widehat{V}_2] = \begin{bmatrix} \widehat{\Sigma}_1 & 0 \\ 0 & \widehat{\Sigma}_2 \\ S_1 & S_2 \end{bmatrix}. \tag{3.14}$$

Theorem 3.2 *In the notation of Theorem 3.1, suppose that span$(\widehat{V}) = \mathbb{R}^n$. Then*

$$\|\sin\angle(U_1, \widehat{U}_1)\| + \|\sin\angle(V_1, \widehat{V}_1)\|) \le (1 + \frac{\|S_2\|_2}{\delta})\|\sin\angle(U_1, \widehat{U})\| \tag{3.15}$$

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

$$\max(\|\sin\angle(U_1, \widehat{U}_1)\|_{2,F}, \|\sin\angle(V_1, \widehat{V}_1)\|_{2,F}) \le \sqrt{1 + \frac{\|S_2\|_{2,F}^2}{\delta^2}}\|\sin\angle(U_1, \widehat{U})\|_{2,F}. \tag{3.16}$$

PROOF. The bounds are obtained essentially by taking $R_i = 0$ in Theorem 3.1, and noting that $\angle(V_1, [\widehat{V}_1\ \widehat{V}_2]) = 0$ because $\widehat{V}$ spans the entire space $\mathbb{R}^n$. □

Note that the inequality corresponding to (3.3) becomes strictly weaker than (3.15), thus omitted.

3.3 Note on approximating null spaces

We expect that this work is most relevant in the context of finding a 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 3.1 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, and in particular, in the context of computing the null space.

However, practical difficulties arise when one is not finding the largest singular subspaces. The first difficulty is related to the remark before Section 2.1; 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. An approach analogous to harmonic Rayleigh-Ritz might be helpful here; the analysis of such alternatives would also be of interest.

Another difficulty (focusing on null spaces) is that even if the bounds in Theorem 3.1 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 are merely a subset of the null spaces, of the same size as $\widehat{U}_1, \widehat{V}_1$. Unfortunately there is no easy way to check that $\widehat{U}_1, \widehat{V}_1$ contains the entire desired null space (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). The blame is not on the projection algorithm but the input subspace $\widehat{U}, \widehat{V}$ that we started with.

A An alternative proof of Wedin's result

A classical result on perturbation of singular vectors is due to Wedin [18], which is an analogue of Davis-Kahan's $\sin\theta$ theorems for Hermitian eigenproblems. Wedin calls it the generalized $\sin\theta$ theorem in [18, § 3]. Here we show the line of argument that we used in this paper can be used to prove the result when the perturbation is off-diagonal. The proof below may look more accessible (an alternative derivation is via the Jordan-Wieldant matrix combined with the Davis-Kahan theory [17, § V.4.1]). We also prove slightly more in equation (A.2).

Theorem A.1 *Let $A, E \in \mathbb{C}^{m\times n}$ and $[U_1\ U_2] \in \mathbb{C}^{m\times m}$, $[V_1\ V_2] \in \mathbb{C}^{n\times n}$ be orthogonal matrices such that*

$$[U_1\ U_2]^*(A+E)[V_1\ V_2] = \begin{bmatrix}\Sigma_1 & \\ & \Sigma_2\end{bmatrix} + \begin{bmatrix}0 & R\\ S & 0\end{bmatrix},$$

where Σ_1, Σ_2 need not be diagonal. Let diag$(\widetilde{\Sigma}_1, \widetilde{\Sigma}_2)$ be a matrix of singular values of $A+E$ (singular values ordered arbitrarily), and define $\delta = \max(\sigma_{\min}(\Sigma_1) - \|\widetilde{\Sigma}_2\|_2, \sigma_{\min}(\widetilde{\Sigma}_2) - \|\Sigma_1\|_2)$. If $\delta > 0$, then the exact left and right singular subspaces $\widetilde{U}_1$ and $\widetilde{V}_1$ of $A+E$ corresponding to $\widetilde{\Sigma}_1$ satisfy, in any unitarily invariant norm,

$$\max(\|\sin\angle(V_1, \widetilde{V}_1)\|, \|\sin\angle(U_1, \widetilde{U}_1))\|) \le \frac{\max(\|R\|, \|S\|)}{\delta} \tag{A.1}$$

and

$$\|\sin\angle(V_1, \widetilde{V}_1)\| + \|\sin\angle(U_1, \widetilde{U}_1))\| \le \frac{\|R\| + \|S\|}{\delta}. \tag{A.2}$$

PROOF. Let $\widetilde{A} := [U_1\ U_2]^*(A+E)[V_1\ V_2]$ and $\widetilde{A} = \widetilde{U}\widetilde{\Sigma}\widetilde{V}^*$ be its SVD. Writing

$$\widetilde{U} = [\begin{bmatrix}\widetilde{U}_{11} & \widetilde{U}_{12}\\ \widetilde{U}_{21} & \widetilde{U}_{22}\end{bmatrix}], \quad \widetilde{V} = \begin{bmatrix}\widetilde{V}_{11} & \widetilde{V}_{12}\\ \widetilde{V}_{21} & \widetilde{V}_{22}\end{bmatrix},$$

as in (2.4), (2.5) we have

$$\|\sin\angle(U_1, \widetilde{U}_1)\| = \|\widetilde{U}_{21}\|, \qquad \|\sin\angle(V_1, \widetilde{V}_1)\| = \|\widetilde{V}_{21}\|. \tag{A.3}$$

By the second block row of $\begin{bmatrix}\Sigma_1 & R\\ S & \Sigma_2\end{bmatrix}\begin{bmatrix}\widetilde{V}_{11}\\ \widetilde{V}_{21}\end{bmatrix} = \begin{bmatrix}\widetilde{U}_{11}\\ \widetilde{U}_{21}\end{bmatrix}\widetilde{\Sigma}_1$ we get

$$S\widetilde{V}_{11} + \Sigma_2\widetilde{V}_{21} = \widetilde{U}_{21}\widetilde{\Sigma}_1,$$

so

$$\widetilde{U}_{21}\widetilde{\Sigma}_1 - \Sigma_2\widetilde{V}_{21} = S\widetilde{V}_{11},$$

If $\sigma_{\min}(\widetilde{\Sigma}_1) > \|\Sigma_2\|_2$, then taking norms we get

$$\sigma_{\min}(\widetilde{\Sigma}_1)\|\widetilde{U}_{21}\| - \|\Sigma_2\|_2\|\widetilde{V}_{21}\| \le \|S\|. \tag{A.4}$$

Similarly, the second block column of $\begin{bmatrix}\widetilde{U}_{11}^* & \widetilde{U}_{21}^*\end{bmatrix}\begin{bmatrix}\Sigma_1 & R \\ S & \Sigma_2\end{bmatrix} = \widetilde{\Sigma}_1\begin{bmatrix}\widetilde{V}_{11} & \widetilde{V}_{21}\end{bmatrix}$ gives $\widetilde{U}_{11}^*R + \widetilde{U}_{21}^*\Sigma_2 = \widetilde{\Sigma}_1\widetilde{V}_{21}$, so

$$\widetilde{\Sigma}_1\widetilde{V}_{21} - \widetilde{U}_{21}^*\Sigma_2 = \widetilde{U}_{11}^*R.$$

Taking norms we get

$$\sigma_{\min}(\widetilde{\Sigma}_1)\|\widetilde{V}_{21}\| - \|\Sigma_2\|_2\|\widetilde{U}_{21}\| \le \|R\|. \tag{A.5}$$

Eliminating the $\|\widetilde{V}_{21}\|$ term from (A.4) and (A.5), we get

$$((\sigma_{\min}(\widetilde{\Sigma}_1))^2 - \|\Sigma_2\|_2^2)\|\widetilde{U}_{21}\| \le \sigma_{\min}(\widetilde{\Sigma}_1)\|R\| + \|\Sigma_2\|_2\|S\|,$$

Hence, assuming $\sigma_{\min}(\widetilde{\Sigma}_1)) > \|\Sigma_2\|_2$ we obtain

$$\begin{aligned}\|\widetilde{U}_{21}\| &\le \frac{\sigma_{\min}(\widetilde{\Sigma}_1)\|R\| + \|\Sigma_2\|_2\|S\|}{(\sigma_{\min}(\widetilde{\Sigma}_1))^2 - \|\Sigma_2\|_2^2} &\text{(A.6)}\\ &\le \frac{\max(\|R\|,\|S\|)\sigma_{\min}(\widetilde{\Sigma}_1) + \|\Sigma_2\|_2}{(\sigma_{\min}(\widetilde{\Sigma}_1))^2 - \|\Sigma_2\|_2^2} = \frac{\max(\|R\|,\|S\|)}{(\sigma_{\min}(\widetilde{\Sigma}_1)) - \|\Sigma_2\|_2}\\ &= \frac{\max(\|R\|,\|S\|)}{\delta}. &\text{(A.7)}\end{aligned}$$

A similar argument, again assuming $\sigma_{\min}(\widetilde{\Sigma}_1)) > \|\Sigma_2\|_2$, yields a bound for $\|\widetilde{V}_{21}\|$:

$$\begin{aligned}\|\widetilde{V}_{21}\| &\le \frac{\|\Sigma_2\|_2\|R\| + \sigma_{\min}(\widetilde{\Sigma}_1)\|S\|}{(\sigma_{\min}(\widetilde{\Sigma}_1))^2 - \|\Sigma_2\|_2^2} &\text{(A.8)}\\ &\le \frac{\max(\|R\|,\|S\|)}{\delta}. &\text{(A.9)}\end{aligned}$$

As in the proof of Theorem 3.1, the bounds (A.7) and (A.9) hold even when $\sigma_{\min}(\Sigma_2) > \|\widetilde{\Sigma}_1\|_2$. Hence we conclude that

$$\max(\|\widetilde{V}_{21}\|,\|\widetilde{U}_{21}\|) \le \frac{\max(\|R\|,\|S\|)}{\delta}.$$

In view of (A.3) this establishes (A.1).

The remaining task is to prove (A.2). We add (A.6) and (A.8) to obtain (if $\sigma_{\min}(\widetilde{\Sigma}_1) > \|\Sigma_2\|_2$)

$$\|\widetilde{U}_{21}\| + \|\widetilde{V}_{21}\| \le \frac{\|R\| + \|S\|}{\sigma_{\min}(\widetilde{\Sigma}_1) - \|\Sigma_2\|_2} = \frac{\|R\| + \|S\|}{\delta},$$

and again by repeating the argument we see that the inequality $\|\widetilde{U}_{21}\| + \|\widetilde{V}_{21}\| \le \frac{\|R\|+\|S\|}{\delta}$ holds whether $\delta = \sigma_{\min}(\widetilde{\Sigma}_1) - \|\Sigma_2\|_2$ or $\delta = \sigma_{\min}(\Sigma_2) - \|\widetilde{\Sigma}_1\|_2$.

References

[1] C. Beattie. Galerkin eigenvector approximations. *Math. Comp.*, 69:1409–1434, 2000.

[2] G. H. Golub and C. F. Van Loan. *Matrix Computations.* The Johns Hopkins University Press, 4th edition, 2012.

[3] M. Gu. Subspace iteration randomization and singular value problems. *arXiv preprint arXiv:1408.2208*, 2014.

[4] N. Halko, P.-G. Martinsson, and J. A. Tropp. Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions. *SIAM Rev.*, 53(2):217–288, 2011.

[5] N. J. Higham. *Functions of Matrices: Theory and Computation.* SIAM, Philadelphia, PA, USA, 2008.

[6] M. E. Hochstenbach. A Jacobi–Davidson type SVD method. *SIAM J. Sci. Comp*, 23(2):606–628, 2001.

[7] A. V. Knyazev. New estimates for Ritz vectors. *Math. Comp.*, 66(219):985–995, 1997.

[8] C.-K. Li and R.-C. Li. A note on eigenvalues of perturbed Hermitian matrices. *Linear Algebra Appl.*, 395:183–190, 2005.

[9] R. Mathias. Quadratic residual bounds for the Hermitian eigenvalue problem. *SIAM J. Matrix Anal. Appl.*, 19(2):541–550, 1998.

[10] R. B. Morgan. Computing interior eigenvalues of large matrices. *Linear Algebra Appl.*, 154-156:289 – 309, 1991.

[11] Y. Nakatsukasa. *Algorithms and Perturbation Theory for Matrix Eigenvalue Problems and the Singular Value Decomposition.* PhD thesis, University of California, Davis, 2011.

[12] C. C. Paige and M. A. Saunders. Towards a generalized singular value decomposition. *SIAM J. Numer. Anal.*, 18(3):398–405, 1981.

[13] B. N. Parlett. *The Symmetric Eigenvalue Problem.* SIAM, Philadelphia, 1998.

[14] Y. Saad. *Numerical Methods for Large Eigenvalue Problems.* SIAM, Philadelphia, PA, USA, second edition, 2011.

[15] G. W. Stewart. A generalization of Saad's theorem on Rayleigh-Ritz approximations. *Linear Algebra Appl.*, 327:115–119, 1999.

[16] G. W. Stewart. *Matrix Algorithms Volume II: Eigensystems.* SIAM, Philadelphia, 2001.

[17] G. W. Stewart and J.-G. Sun. *Matrix Perturbation Theory (Computer Science and Scientific Computing)*. Academic Press, 1990.

[18] P. Wedin. Perturbation bounds in connection with singular value decomposition. *BIT*, 12:99–111, 1972.