Randomized Subspace Iteration Method for Eigenvalue Problems

Nikita Kapur

B.Sc. (Honours) Mathematics, University of Delhi, India, 2012

M.Sc. Mathematics, University of Delhi, India, 2014

B.Ed., Guru Gobind Singh Indraprastha University, India, 2015

Submitted to the graduate degree program in Department of Mathematics and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the degree of Master of Arts in Mathematics.

Committee members

Agnieszka Międlar, Chairperson

Emily E. Witt, Member

Hongguo Xu, Member

Arvind K. Saibaba, Member

Date defended: May 07, 2021

The Master Thesis Committee for Nikita Kapur certifies
that this is the approved version of the following master thesis :

Randomized Subspace Iteration Method for
Eigenvalue Problems

Agnieszka Międlar, Chairperson

Date approved: May 12, 2021

Abstract

This thesis presents the randomized subspace iteration method for eigenvalue problems. In our analysis, we have considered symmetric positive definite eigenvalue problem. We present deterministic and probabilistic bounds for three quantities. First, we present the deterministic and probabilistic bounds for the canonical angles between the exact and the approximate eigenvector subspaces. Second, we give deterministic and probabilistic bounds for the sine of angle between the eigenvectors of the exact eigenvector subspace and the approximated eigenvector subspace. Third, we also present deterministic and probabilistic bounds for the accuracy of eigenvalues using the randomized subspace iteration. The probabilistic bounds are provided when a Gaussian random matrix is used as the initial subspace. Finally, we illustrate our theoretical results numerically using several different test matrices.

Acknowledgements

I have received an immense support and guidance throughout the writing of this thesis. First, I would like to thank my advisor Dr. Agnieszka Międlar for her invaluable assistance in my research. Her impeccable methodology and optimism has made me an industrious researcher. Moreover, her insightful feedback has brought my work to a higher level. I would also like to acknowledge Dr. Arvind K. Saibaba from the North Carolina State University for his priceless support in my research. I am grateful to him for the opportunities he has given me to do research in this field. I would also like to thank my professor Dr. Emily Witt for her immense guidance in my studies. She has always encouraged me to try new things.

Last but not least, I would like to thank my parents and my brother for always believing in me. I would also like to thank my friends for their encouragement, especially my friend Xiaowen Li for helping me to learn MATLAB.

Contents

List of Figures

Chapter 1

Introduction and Motivation

Randomized algorithms are discussed extensively in the early survey article by P. Martinsson and J. A. Tropp [8]. Randomized (probabilistic) algorithms have been an integral part of scientific computing ever since the novel work of Ulam and von Neumann on Monte Carlo methods in 1940s. Since 1980s, the randomized methods have been used in Numerical Linear Algebra (NLA) [8, 9]. In late 1990s and early 2000s, the researchers in theoretical computer science showed that the storage costs are significantly lower for randomized embeddings in computations on streaming data than the classical algorithms. In the mid-2000s, numerical analysts have shown how randomized algorithms can be practically implemented for low-rank approximations and least square problems [7, 8]. Since then, randomized methods have been proven to be more efficient as compared to the classical methods in solving certain NLA problems [4].

In [3] and [14], randomized algorithms have been developed to give better approximations for low-rank decomposition when the singular values decay rapidly. As discussed earlier, randomized methods became more popular in last two decades to determine low-rank approximations of matrices. They have gained more popularity over the classical methods because they are easy to implement, numerically robust and computationally more efficient. Even though, they have same asymptotic cost as the classical methods, they are more suitable for large-scale computations. Moreover, they are very reliable in practical applications because of their numerical robustness.

In this thesis, we focus on a specific kind of randomized algorithm known as *randomized subspace iteration method*. Our analysis is a particular case of the analysis performed in [12], as we are considering symmetric positive definite matrices. The key idea is to identify a subspace that approximately captures the range of a matrix using random sampling. A low-rank approximation

is then obtained by projecting the matrix onto this approximated subspace. After computing the low-rank approximation, an additional post-processing step is applied to compress the low-rank representation to obtain a matrix with desired target rank. Then, a conversion step is performed in order to obtain an equivalent eigenvalue decomposition.

Chapter 2

Background and Notation

2.1 Preliminaries

Symmetric Positive Definite Eigenvalue Problem (SPDEVP): Given a symmetric positive definite matrix $A \in \mathbb{R}^{n\times n}$ we are interested in finding a scalar $\lambda \in \mathbb{R}$ (an eigenvalue) and a corresponding vector $\mathbf{u} \neq 0 \in \mathbb{R}^n$ (an eigenvector) of matrix A such that

$$A\mathbf{u} = \lambda\mathbf{u}. \tag{2.1}$$

A set of all eigenvalues of matrix A is called a *spectrum of A*. For a symmetric positive definite matrix A all its eigenvalues are real and positive, hence they can be arranged in a decreasing order

$$\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_k \geq \lambda_{k+1} \geq \cdots \geq \lambda_n > 0.$$

The corresponding eigenvectors are denoted as $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$.

Eigenvalue Decomposition: Given a symmetric positive definite matrix $A \in \mathbb{R}^{n\times n}$ and a positive integer $1 \leq k \leq \mathsf{rank}\,(A)$, we consider the following *eigenvalue decomposition* of matrix A

$$A = U\Lambda U^T = \begin{bmatrix} U_k & U_\perp \end{bmatrix} \begin{bmatrix} \Lambda_k & \\ & \Lambda_\perp \end{bmatrix} \begin{bmatrix} U_k^T \\ U_\perp^T \end{bmatrix}. \tag{2.2}$$

Here, $\Lambda_k \in \mathbb{R}^{k\times k}$ denotes a diagonal matrix containing the k largest eigenvalues of matrix A, i.e., $\Lambda_k = \mathsf{diag}\,(\lambda_1,\ldots,\lambda_k)$ and $\Lambda_\perp \in \mathbb{R}^{(n-k)\times(n-k)}$ a diagonal matrix containing the remaining $n-k$ eigenvalues of matrix A, i.e., $\Lambda_\perp = \mathsf{diag}\,(\lambda_{k+1},\lambda_{k+2},\ldots,\lambda_n)$. The columns of U_k and $U_\perp$ are the eigenvectors corresponding to eigenvalues in Λ_k and $\Lambda_\perp$, respectively. We also denote $A_k = U_k\Lambda_k U_k^T$ as the *best rank-k approximation of A*, for any unitarily invariant norm. Analogously, we define $A_\perp = U_\perp\Lambda_\perp U_\perp^T$ and notice that $A = A_k + A_\perp$.

Note that from [2], we have that *Spectral Theorem* gives a similar decomposition known as the *Singular Value Decomposition (SVD)*. The matrix A can be written as $A = U\Sigma V^T$, where $\Sigma \in \mathbb{R}^{n\times n}$ is the diagonal matrix with singular values as its diagonal entries and $U,V \in \mathbb{R}^{n\times n}$ are orthogonal matrices. The columns of U and the columns of V are called left and right singular vectors of A, respectively.

Eigenvalues and Eigenvalue Ratios: Let $\lambda_1,\lambda_2,\ldots,\lambda_n$ be the eigenvalues of A as defined before. Denote the spectral norm by $\|\cdot\|_2$. Then $\|\Lambda_\perp\|_2 = \lambda_{k+1}$ and $\|\Lambda_k^{-1}\|_2 = \frac{1}{\lambda_k}$. Further, we define the *eigenvalue ratios* as

$$\gamma_j = \frac{\lambda_{k+1}}{\lambda_j}, \quad j = 1,2,\ldots,k. \tag{2.3}$$

Since the eigenvalues of A are monotonically decreasing, the eigenvalue ratios are monotonically increasing, i.e., $\gamma_1 \le \gamma_2 \le \ldots \le \gamma_k \le 1$.

Projection Matrices: Suppose that the matrix Z has full column rank with column space range(Z). $Z^\dagger$ is a left multiplicative inverse with † denoting the Moore-Penrose inverse (or pseudo inverse). We will define the (orthogonal) projection matrix $\mathscr{P}_Z$, as $\mathscr{P}_Z = ZZ^\dagger$. An orthogonal matrix Q can be uniquely defined by its range, and range$(\mathscr{P}_Z)$ = range(Z). For any matrix Q with orthonormal columns, the above formula simplifies and we have $\mathscr{P}_Q = QQ^T$.

Canonical Angles: The canonical angles or principal angles measure the separation between subspaces. Let $\mathscr{X}$ and $\mathscr{Y}$ be two subspaces of $\mathbb{R}^n$ such that the $\dim(\mathscr{X}) = \ell$ and $\dim(\mathscr{Y}) = k$, and $\ell \ge k$. Then the canonical angles between the subspaces $\mathscr{X}$ and $\mathscr{Y}$ are defined as $0 \le \theta_i \le$

$\pi/2, \quad i = 1,2,\ldots,k$ such that

$$\cos\theta_i \quad = \max_{\mathbf{x}\in\mathscr{X},\mathbf{y}\in\mathscr{Y},\|\mathbf{x}\|_2=\|\mathbf{y}\|_2=1} \mathbf{y}^T\mathbf{x} \quad = \quad \mathbf{y}_i^T\mathbf{x}_i, \quad i = 1,2,\ldots,k,$$

where $\|\mathbf{x}_i\|_2 = \|\mathbf{y}_i\|_2 = 1$, and

$$\mathbf{x}_j^T\mathbf{x} = 0, \quad \mathbf{y}_j^T\mathbf{y} = 0 \quad j = 1,\ldots,i-1.$$

The canonical angles can be arranged in increasing order as

$$0 \le \theta_1 \le \theta_2 \le \cdots \le \theta_k \le \pi/2.$$

Moreover, it can easily be verified that $\sin\theta_i$ are singular values of $\mathscr{P}_{\mathscr{X}} - \mathscr{P}_{\mathscr{Y}}$.

Let $\angle(\mathscr{X},\mathscr{Y})$ denotes the canonical angles between the subspaces $\mathscr{X}$ and $\mathscr{Y}$. Let X and Y denote the matrices with orthonormal columns which form the bases for subspaces $\mathscr{X}$ and $\mathscr{Y}$, respectively. Then, the singular values of $(I - XX^T)Y$ can be used to find $\sin\angle(\mathscr{X},\mathscr{Y})$, and the singular values of X^TY can be used to find $\cos\angle(\mathscr{X},\mathscr{Y})$.

Rayleigh-Ritz Procedure: Here we give a generalized Rayleigh-Ritz procedure as described in [10]. Let $A \in \mathbb{R}^{n\times n}$ be a symmetric positive definite matrix and $\mathscr{V}$ be an m dimensional subspace of $\mathbb{R}^n$. We consider the eigenvalue problem (2.1). An orthogonal projection approach allows us to determine an approximate eigenpair $(\widehat{\lambda},\widehat{\mathbf{u}})$, with $\widehat{\lambda} \in \mathbb{R}$ and $\widehat{\mathbf{u}} \neq 0$ in $\mathscr{V}$ satisfying the following *Galerkin condition* given by

$$A\widehat{\mathbf{u}} - \widehat{\lambda}\widehat{\mathbf{u}} \perp \mathscr{V}.$$

This is equivalent to say that

$$\langle A\widehat{\mathbf{u}} - \widehat{\lambda}\widehat{\mathbf{u}}, \mathbf{v}\rangle = 0, \quad \text{for all } \mathbf{v} \in \mathscr{V} \tag{2.4}$$

Given an orthonormal basis $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m\}$ of $\mathcal{V}$ and denote $V = [\mathbf{v}_1\ \mathbf{v}_2\ \ldots\ \mathbf{v}_m]$, we can express an approximate eigenvector $\widehat{\mathbf{u}} \in \mathcal{V}$ as a linear combination of the basis $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m\}$, i.e.,

$$\widehat{\mathbf{u}} = V\mathbf{w}.$$

Therefore, equation (2.4) yields

$$\langle AV\mathbf{w} - \widehat{\lambda} V\mathbf{w}, \mathbf{v}_j \rangle = 0, \qquad \text{for all } \ j = 1, 2, \ldots, m.$$

Hence, $\mathbf{w}$ and $\widehat{\lambda}$ must satisfy

$$B_m\mathbf{w} = \widehat{\lambda}\mathbf{w},$$

where

$$B_m = V^T AV.$$

Each eigenvalue $\widehat{\lambda}_i, i = 1, 2, \ldots, m$ of B_m is called a *Ritz value*, and vector $V\mathbf{w}_i$ is called a *Ritz vector*, where $\mathbf{w}_i$ is the eigenvector of matrix B_m associated with the eigenvalue $\widehat{\lambda}_i$.

2.2 Subspace Iteration Method for Eigenvalue Problems

The *subspace iteration* (or simultaneous inverse iteration) method is an iterative method used for solving eigenvalue problems first introduced by Bauer under the name of *Treppeniteration* (staircase iteration) [10]. This method can be viewed as a generalization of the power method to account for multiple vectors at a time. In order to get better eigenpair approximations, the Rayleigh-Ritz projection step is used. The algorithm with the Rayleigh-Ritz projection step is as follows:

Algorithm 1 : Idealized Subspace Iteration with Rayleigh-Ritz Projection Step

Ensure: Let $A \in \mathbb{R}^{n\times n}$ be a symmetric positive definite matrix $X \in \mathbb{R}^{n\times m}$ matrix with $X^T X = I_m$

for $i = 1, 2, \ldots$ Do
    Compute $Y = AX$
    Compute $Y = QR$ {QR Factorization of matrix Y}
    Form $B = Q^T AQ$
    Compute $B =: W\widehat{\Lambda}W^T$ {Eigenvalue Decomposition of matrix B}
    $X = QW$
    Check Convergence
End for
$\widehat{U} = X$
return $\widehat{U}$ and $\widehat{\Lambda}$ that satisfy $\widehat{A} \equiv \widehat{U}\widehat{\Lambda}\widehat{U}^T$

Remark 1. *The diagonal entries $\widehat{\lambda}_1, \widehat{\lambda}_2, \ldots, \widehat{\lambda}_m$ of matrix $\widehat{\Lambda}$ (eigenvalues of matrix B) are the Ritz values of A and the columns $\widehat{\mathbf{u}}_i$ of matrix $\widehat{U}$ are the corresponding Ritz vectors.*

2.3 Random Matrices

In this section we give the common distributions and random matrices used in the random matrix theory as considered in [8, 17].

Distributions: Let us first discuss commonly used distributions.

- *Uniform Distribution:* generally, UNIF denotes the uniform distribution over a finite set. A scalar Rademacher random variable has a uniform distribution over a finite set $\{-1, +1\}$ [16]. A Rademacher random vector has independent and identically distributed (iid) entries, each distributed as a scalar Rademacher random variable. Sometimes, uniform distributions are considered over Borel subset of $\mathbb{F}^n$, equipped with Lebesgue measure [8].
- *Normal Distribution:* Normal distribution on $\mathbb{F}^n$ is denoted by Norm$(\mu, \mathbf{C})$ or $\mathcal{N}(\mu, \mathbf{C})$, where $\mathbb{F}$ is a field, μ is the expectation value and $\mathbf{C}$ is the positive semi-definite covariance matrix. A standard normal random variable or a random vector has expectation value *zero* ($\mu = 0$) and the identity covariance matrix ($\mathbf{C} = I$). We interchangeably use the term *Gaussian* to refer to normal distributions.

In practice, standard normal matrices are the most common random matrices used in randomized algorithms.

Standard Normal Matrix or Gaussian Matrix:

- *Real Standard Normal Matrix:* A matrix $\Omega \in \mathbb{R}^{m\times n}$ has the real standard normal distribution if its entries form an independent family of standard normal variables (i.e. Gaussian with mean zero and variance one).
- *Complex Standard Normal Matrix:* A matrix $\Omega \in \mathbb{C}^{m\times n}$ has complex standard normal distribution if it can be written as $\Omega = \Omega_1 + \imath\Omega_2$, where Ω_1 and Ω_2 are independent real standard normal matrices.

For the purpose of this work, matrix $\Omega \in \mathbb{R}^{n\times(k+p)}$ will be a *real Gaussian random matrix*, where k will be the *target rank* and p will be *oversampling parameter*. The entries of Ω will be independent and identically distributed $\mathcal{N}(0,1)$, i.e., with mean zero and variance one.

Chapter 3

Randomized Subspace Iteration Method

We will now consider the randomized subspace iteration method for solving eigenvalue problem (2.1) with a symmetric positive definite matrix A (SPDEVP). We start by taking a symmetric positive definite matrix $A \in \mathbb{R}^{n\times n}$ and $\Omega \in \mathbb{R}^{n\times(k+p)}$ is a starting guess, with parameter k denoting the *number of eigenvalues of interest* (desired size of the eigenspace), and $p > 0$ is the *oversampling parameter*. Moreover, let $q \geq 1$ be the maximum number of randomized subspace iteration steps performed to obtain matrix Y, also known as the *sketch* of matrix A. The algorithm proceeds by performing a thin-QR factorization of matrix Y in order to obtain a matrix Q whose columns form an orthonormal basis for the range of Y. We assume that the range of matrix Q gives a good approximation for the range of matrix A, i.e., given $\varepsilon > 0$

$$\|A - QQ^T A\|_2 \leq \varepsilon.$$

Moreover, by the triangle inequality we have that

$$\begin{aligned} \|A - QQ^T AQQ^T\|_2 &= \|A - QQ^T A + QQ^T A - QQ^T AQQ^T\|_2 \\ &\leq \|A - QQ^T A\|_2 + \|QQ^T(A - QQ^T A)\|_2 \\ &= 2\|(I - QQ^T)A\|_2 \leq 2\varepsilon. \end{aligned}$$

This observation tells us how to compute an eigenvalue decomposition of matrix A. With orthonormal basis Q at hand, we obtain a low-rank approximation $\widehat{A}$ of matrix A by the projection, i.e., $A \approx \widehat{A} = Q(Q^T AQ)Q^T$. We then compute the eigenvalue decomposition $W\widehat{\Lambda}W^T$ of a smaller

matrix Q^TAQ, which gives us the orthonormal matrix $\widehat{U} = QW$ satisfying $\widehat{A} = \widehat{U}\widehat{\Lambda}\widehat{U}^T$. The algorithm showing the randomized subspace iteration method for solving a symmetric positive definite eigenvalue problem is given in Algorithm 2.

Algorithm 2 : Randomized Subspace Iteration for SPDEP.

Ensure: Let $A \in \mathbb{R}^{n\times n}$ be a symmetric positive definite matrix and $\Omega \in \mathbb{R}^{n\times(k+p)}$ be a starting guess. Here k is the target rank (number of eigenvalues to be computed), p is the oversampling parameter and $q \geq 1$ is number of subspace iteration steps.

1: Set $X = \Omega$
2: for i = 1, 2,... q
3: Compute $Y = AX$
4: Compute $Y = QR$ {thin-QR factorization of matrix Y}
5: Form the small matrix $B = Q^TAQ$
6: Compute $B =: W\widehat{\Lambda}W^T$ {the eigenvalue decomposition of matrix B}
7: Form the orthonormal matrix $X = QW$
8: Check convergence
9: End for
10: $\widehat{U} := X$
11: **return** $\widehat{U}$ and $\widehat{\Lambda}$ that satisfy $A \equiv \widehat{A} = \widehat{U}\widehat{\Lambda}\widehat{U}^T$

The return matrix $\widehat{U}$ is the approximated eigenspace of dimension $n \times (k+p)$ and the matrix $\widehat{\Lambda}$ is a diagonal matrix of dimension $(k+p)\times(k+p)$ with approximated eigenvalues down the diagonal.

We now introduce the following notation which is needed in our analysis. The matrix $U^T\Omega$ captures the influence of initially chosen random Gaussian matrix Ω on the eigenvector matrix U. Partition this matrix as

$$U^T\Omega = \begin{bmatrix} U_k^T\Omega \\ U_\perp^T\Omega \end{bmatrix} =: \begin{bmatrix} \Omega_1 \\ \Omega_2 \end{bmatrix}, \tag{3.1}$$

where $\Omega_1 = U_k^\top\Omega \in \mathbb{R}^{k\times(k+p)}$ and $\Omega_2 = U_\perp^\top\Omega \in \mathbb{R}^{(n-k)\times(k+p)}$. We assume that the *target rank* k satisfies $1 \leq k \leq \mathsf{rank}(A)$. Additionally, we have the following assumptions:

Assumption 1: Let $\Omega_1 = U_k^\top\Omega \in \mathbb{R}^{k\times(k+p)}$ be defined as above. We assume that $\mathsf{rank}(\Omega_1) = k$.

Assumption 2: The *eigenvalue gap* at index k is inversely proportional to the *eigenvalue ratio*, i.e.,

$$\gamma_k = \|\Lambda_\perp\|_2 \|\Lambda_k^{-1}\|_2 = \frac{\lambda_{k+1}}{\lambda_k} < 1.$$

The first assumption ensures that the starting guess Ω has a significant influence over the eigenvectors, whereas the second assumption guarantees that the k-dimensional subspace range$(\mathrm{U_k})$ is well defined. In practice, it is highly desirable that $\gamma_k \ll 1$, which guarantees a large eigenvalue gap between eigenvalue λ_k and λ_{k+1}.

3.1 Deterministic Bounds

In this section we will discuss the accuracy of computed eigenpairs of problem (2.1) in the case of symmetric positive definite matrix A. The results presented in this section show how well range$(\widehat{\mathrm{U}})$ (an approximate eigenspace) approximates range$(\mathrm{U_k})$ (an exact eigenspace) corresponding to the k largest eigenvalues of A. This is measured in terms of the *canonical angles* between the exact and the approximate eigenspaces. Let $\theta_1, \theta_2, \ldots, \theta_k$ denote the subspace angles between $\widehat{U} \in \mathbb{R}^{n\times(k+p)}$ and $U_k \in \mathbb{R}^{n\times k}$. The following result is a direct consequence of [12, Theorem 1] and provides the bound for the *canonical angles* $\angle(U_k, \widehat{U})$ in the case of symmetric positive definite eigenvalue problem.

Theorem 3.1.1. *Let $\widehat{U}$ be obtained from the Algorithm 2. With Assumption 1 the canonical angles $\theta_j, j = 1,2,\ldots,k$, satisfy the following inequality*

$$\sin\theta_j \leq \frac{\gamma_j^q \|\Omega_2\Omega_1^\dagger\|_2}{\sqrt{1+\gamma_j^{2q}\|\Omega_2\Omega_1^\dagger\|_2^2}}.$$

Proof. The proof is similar to the one of [12, Theorem 1]. □

This result has several important consequences. First, if the matrix A has *exact rank* k, then all the *canonical angles* are uniformly zero. This implies that the randomized subspace iteration

has exactly identified the eigenvalues. On the other hand, if the *eigenvalue gap* γ_k is very close to 1, then the subspace may not be well defined, and it may be difficult to identify them using the randomized subspace iteration. In practice, it is highly advisable to take $\gamma_k \ll 1$, so that the angles can be captured accurately. Second, the bound for *canonical angles* show the explicit dependence on the *eigenvalue ratios* γ_j, in particular, the *canonical angles* θ_j converge to zero with the convergence rates depending on *eigenvalue ratios*. This is equivalent to say that the smaller the *eigenvalue ratio*, the smaller the *canonical angles*. Third, term $\|\Omega_2\Omega_1^\dagger\|_2$ can be written in terms of the eigenvector matrix U and the starting guess Ω as

$$\|\Omega_2\Omega_1^\dagger\|_2 = \|(U_\perp^T\Omega)(U_k^T\Omega)^\dagger\|_2.$$

It can be observed that when the columns of Ω are linearly independent, this quantity is the tangent of the largest *canonical angle* between the spaces range($\mathrm{U_k}$) and range(Ω). This is a common term in randomized linear algebra, and can be interpreted as the measure of the subspace overlap between the starting guess Ω and the matrix U_k containing the eigenvectors corresponding to the k largest eigenvalues of matrix A. In the ideal scenario, the eigenvectors in U_k are contained in Ω. This is extensively discussed in [1, Section 2.5]. In particular, when Ω is a Gaussian random matrix, $\|\Omega_2\Omega_1^\dagger\|_2$ is approximately of order $\sqrt{(n-k)k}$. Fourth, the influence of $\|\Omega_2\Omega_1^\dagger\|_2$ is restrained by the *eigenvalue ratios* γ_j^q. When the number of subspace iteration steps q is significantly large, then the *canonical angles* are smaller than the user defined tolerance.

Remark 2. *The above theorem gives the bound for the sines of the canonical angles; this bound can also be used to give the upper bound for tangents of the canonical angles and the lower bound for the cosines of the canonical angles. With the same notation and assumptions as in Theorem 3.1.1, the relationship between the sine and tangent implies that*

$$\tan\theta_j \leq \gamma_j^q\|\Omega_2\Omega_1^\dagger\|_2, \quad \textit{for} \quad j = 1,2\ldots,k.$$

The next theorem is a special case of a more general result from our ongoing work [5]. Let

$\lambda_1, \lambda_2, \ldots, \lambda_k$ be the eigenvalues of A as defined before, and let $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k$ be the corresponding eigenvectors. The following result quantifies the sine of the angle between the eigenvectors $\mathbf{u}_j$ and the subspace range$(\widehat{\mathrm{U}})$ measured in terms of $\|(I - \mathscr{P}_{\widehat{U}})\mathbf{u}_j\|_2$.

Theorem 3.1.2. *Let $\widehat{U}$ be obtained from Algorithm 2. With Assumption 1 we have that for $j = 1, 2, \ldots, k$*

$$\|(I - \mathscr{P}_{\widehat{U}})\mathbf{u}_j\|_2 \leq \frac{\gamma_j^q \|\Omega_2 \Omega_1^\dagger\|_2}{\sqrt{1 + \gamma_j^{2q} \|\Omega_2 \Omega_1^\dagger\|_2^2}}.$$

Proof. Consider the eigenvalue deconposition of matrix $A = U\Lambda U^T$ in (2.2) and let us denote $\Lambda_1 := (\Lambda_k)^q$ and $\Lambda_2 := (\Lambda_\perp)^q$. Hence, using (3.1) yields

$$\begin{aligned} Y = QR &= A^q\Omega = U\Lambda^q U^T\Omega \\ &= \begin{bmatrix} U_k & U_\perp \end{bmatrix} \begin{bmatrix} \Lambda_k^q & \\ & \Lambda_\perp^q \end{bmatrix} \begin{bmatrix} U_k^T \\ U_\perp^T \end{bmatrix} \Omega = U \begin{bmatrix} \Lambda_1\Omega_1 \\ \Lambda_2\Omega_2 \end{bmatrix}. \end{aligned}$$

Since by assumption $\mathsf{rank}\,(\Omega_1) = k$, $\Omega_1\Omega_1^\dagger = I_k$. Therefore,

$$Y\Omega_1^\dagger\Lambda_1^{-1} = U \begin{bmatrix} I \\ F \end{bmatrix} \quad \text{with} \quad F = \Lambda_2\Omega_2\Omega_1^\dagger\Lambda_1^{-1}.$$

Now define the matrix Z such that

$$Z = Y\Omega_1^\dagger\Lambda_1^{-1}(I_k + F^TF)^{-1/2} = U \begin{bmatrix} I_k \\ F \end{bmatrix} (I_k + F^TF)^{-1/2}. \tag{3.2}$$

Since $I - \mathscr{P}_{\widehat{U}}$ is an orthogonal projector, $0 \preceq I - \mathscr{P}_{\widehat{U}} \preceq I$. The conjugation rule [3, Proposition 8.1] yields

$$(I - \mathscr{P}_Z)(I - \mathscr{P}_{\widehat{U}})(I - \mathscr{P}_Z) \preceq I - \mathscr{P}_Z. \tag{3.3}$$

Since range(Z) $\subset$ range($\widehat{\mathrm{U}}$), following [3, Proposition 8.5] we obtain

$$(I-\mathscr{P}_{\mathrm{Z}})(I-\mathscr{P}_{\widehat{U}})=(I-\mathscr{P}_{\widehat{U}}),$$

and

$$\begin{aligned}(I-\mathscr{P}_{\mathrm{Z}})(I-\mathscr{P}_{\widehat{U}})(I-\mathscr{P}_{\mathrm{Z}}) &= (I-\mathscr{P}_{\widehat{U}})(I-\mathscr{P}_{\mathrm{Z}})\\ &= \big((I-\mathscr{P}_{\mathrm{Z}})(I-\mathscr{P}_{\widehat{U}})\big)^T\\ &= (I-\mathscr{P}_{\widehat{U}})^T\\ &= (I-\mathscr{P}_{\widehat{U}}).\end{aligned}$$

Combining the last equation with (3.3), we have

$$I-\mathscr{P}_{\widehat{U}} \preceq I-\mathscr{P}_{\mathrm{Z}}.$$

This gives $\|(I-\mathscr{P}_{\widehat{U}})\mathbf{u}_j\|_2 \leq \|(I-\mathscr{P}_{\mathrm{Z}})\mathbf{u}_j\|_2$. Considering the square of this upper bound and using the fact that $\mathscr{P}_Z = ZZ^{\dagger}$ we have,

$$\begin{aligned}\|(I-\mathscr{P}_{\mathrm{Z}})\mathbf{u}_j\|_2^2 &\leq \mathbf{u}_j^T(I-\mathscr{P}_{\mathrm{Z}})\mathbf{u}_j = \mathbf{e}_j^T U^T(I-\mathscr{P}_{\mathrm{Z}})U\mathbf{e}_j\\ &= \mathbf{e}_j^T(I-\mathscr{P}_{\mathrm{Z}})\mathbf{e}_j = \mathbf{e}_j^T(I_k-(I_k+F^TF)^{-1})\mathbf{e}_j. \qquad (3.4)\end{aligned}$$

By the proof of [12, Theorem 9], we have

$$F^TF \preceq \|\Omega_2\Omega_1^{\dagger}\|_2^2\,\mathsf{diag}\,(\gamma_1^{2q},\gamma_2^{2q},\ldots,\gamma_k^{2q}),$$

where $\mathsf{diag}\,(\gamma_1^{2q},\gamma_2^{2q},\ldots,\gamma_k^{2q})$ is a diagonal matrix with the *eigenvalue ratios*. Here $M \preceq N$ means $N-M$ is positive semi-definite. Furthermore,

$$\mathbf{e}_j^T(I_k+F^TF)^{-1}\mathbf{e}_j \succeq \mathbf{e}_j^T(I_k+\|\Omega_2\Omega_1^{\dagger}\|_2^2\,\mathsf{diag}\,(\gamma_1^{2q},\gamma_2^{2q},\ldots,\gamma_k^{2q}))^{-1}\mathbf{e}_j.$$

Then we have

$$\mathbf{e}_j^T (I_k + F^T F)^{-1} \mathbf{e}_j \geq \frac{1}{1+\gamma_j^{2q} \|\Omega_2 \Omega_1^\dagger\|_2^2},$$

which yields

$$\mathbf{e}_j^T (I_k - (I_k + F^T F)^{-1}) \mathbf{e}_j \leq 1 - \Big(\frac{1}{1+\gamma_j^{2q} \|\Omega_2 \Omega_1^\dagger\|_2^2}\Big) = \frac{\gamma_j^{2q} \|\Omega_2 \Omega_1^\dagger\|_2^2}{1+\gamma_j^{2q} \|\Omega_2 \Omega_1^\dagger\|_2^2}.$$

Combining this inequality with equation (3.4), and taking the square root on both the sides gives

$$\|(I - \mathscr{P}_{\widehat{U}}) \mathbf{u}_j\|_2 \leq \frac{\gamma_j^q \|\Omega_2 \Omega_1^\dagger\|_2}{\sqrt{1+\gamma_j^{2q} \|\Omega_2 \Omega_1^\dagger\|_2^2}},$$

which completes the proof. □

The inequality, $\gamma_1 \leq \gamma_2 \leq \cdots \leq \gamma_k$ suggests that the approximations to the eigenvectors in the subspace range$(\widehat{\mathrm{U}})$ are converging at different rates. Observe that as already discussed in preliminaries, the eigenvalues of the matrix A are arranged in decreasing order. The next result gives the accuracy of the eigenvalues of A computed using the randomized subspace iteration method presented in Algorithm 2.

Theorem 3.1.3. *Assume $\ell = k$, i.e. $p = 0$ and $\Omega_1 \in \mathbb{R}^{k\times k}$ is invertible. Let $\lambda_1, \lambda_2, \ldots, \lambda_k$ be the exact eigenvalues of A arranged in decreasing order, and $\widehat{\lambda}_1, \widehat{\lambda}_2, \ldots, \widehat{\lambda}_k$ be the eigenvalues of $Q^T AQ$ (Ritz values) arranged in decreasing order. Then, for $j = 1, \ldots, k$, we have*

$$\max_{1\leq j\leq k} |\lambda_j - \widehat{\lambda}_j| \leq (\lambda_{max} - \lambda_{min}) \frac{\gamma_k^q \|\Omega_2 \Omega_1^{-1}\|_2}{\sqrt{1+\gamma_k^{2q} \|\Omega_2 \Omega_1^{-1}\|_2^2}},$$

where λ_{max} and λ_{min} are the largest and the smallest exact eigenvalues of A, respectively.

Proof. By applying [6, Theorem 5], we have that

$$\max_{1\leq j\leq k} |\lambda_j - \widehat{\lambda}_j| \leq (\lambda_{max} - \lambda_{min}) \|\sin\Theta\|_2, \tag{3.5}$$

where Θ denote the diagonal matrix with diagonal entries as the *canonical angles* between range$(\widehat{\mathrm{U}})$ (or range(Q)) and range$(\mathrm{U_k})$. Since $k = \ell$, and Ω_1 is a square matrix with $\mathsf{rank}\,(\Omega_1) = k$, we have that Ω_1 is invertible, i.e., $\Omega_1^{\dagger} = \Omega_1^{-1}$. Equation (3.5) and Theorem 3.1.1 yields

$$\max_{1\le j\le k} |\lambda_j - \widehat{\lambda}_j| \le (\lambda_{max} - \lambda_{min}) \frac{\gamma_k^q \|\Omega_2 \Omega_1^{-1}\|_2}{\sqrt{1+\gamma_k^{2q}\|\Omega_2\Omega_1^{-1}\|_2^2}},$$

which completes the proof. □

3.2 Probabilistic Bounds

The results presented in Section 3.1 do not take into account any information about the distribution of the random matrix $\Omega \in \mathbb{R}^{n\times\ell}$, where $\ell = k+p$. Although, as discussed in Section 2.3 , the entries of matrix Ω can be drawn from different distributions, but the lack of underlying results from the random matrix theory restricts our analysis to a class of standard Gaussian random matrices. In this section we derive a few probabilistic results that will help us better understand the accuracy of the subspaces.

Before we introduce the main results of this section, we first need the following lemma [13, Lemma 2].

Lemma 3.2.1. *Let $\Omega \in \mathbb{R}^{n\times\ell}$ be a standard Gaussian matrix, such that $\mathsf{rank}\,(\Omega_1) = k$. Further let us assume that the oversampling parameter $p \ge 2$ and $\ell = k+p \le n$. Then with probability at least $1-\delta$, we have*

$$\|\Omega_2\Omega_1^{\dagger}\|_2 \le C_g,$$

with

$$C_g \equiv \frac{e\sqrt{\ell}}{p}\left(\frac{2/\delta}{\sqrt{2\pi(p+1)}}\right)^{\frac{1}{(p+1)}}\left(\sqrt{n-k}+\sqrt{\ell}+\sqrt{2\log\frac{2}{\delta}}\right).$$

Theorem 3.2.2 provides probabilistic bounds for the sines of the *canonical angles* between the subspaces range$(\mathrm{U_k})$ and range$(\widehat{\mathrm{U}})$, and for the sines of angle between the eigenvectors $\mathbf{u}_j$ and the

subspace range$(\widehat{\mathrm{U}})$.

Theorem 3.2.2. *Let $\Omega \in \mathbb{R}^{n\times\ell}$ be a standard Gaussian matrix such that* $\mathsf{rank}\,(\Omega_1) = k$. *Moreover, let the oversampling parameter $p \geq 2$ and $\ell = k+p \leq n$. Then, due to Lemma 3.2.1, each of the following results hold with the probability at least $1-\delta$*

(i)

$$\sin\theta_j \leq \frac{\gamma_j^q C_g}{\sqrt{1+\gamma_j^{2q}C_g^2}}, \quad \textit{for} \quad j = 1,\ldots,k,$$

and

(ii)

$$\|(I-\mathscr{P}_{\hat{U}})\mathbf{u}_j\|_2 \leq \frac{\gamma_j^q C_g}{\sqrt{1+\gamma_j^{2q}C_g^2}}, \quad \textit{for} \quad j = 1,\ldots,k,$$

with $\quad C_g \equiv \frac{e\sqrt{\ell}}{p}\left(\frac{2/\delta}{\sqrt{2\pi(p+1)}}\right)^{\frac{1}{(p+1)}}\left(\sqrt{n-k}+\sqrt{\ell}+\sqrt{2\log\frac{2}{\delta}}\right)$.

Proof. (i) Let us first recall the main result of Theorem 3.1.1, i.e.,

$$\sin\theta_j \leq \frac{\gamma_j^q\|\Omega_2\Omega_1^\dagger\|_2}{\sqrt{1+\gamma_j^{2q}\|\Omega_2\|\Omega_1^\dagger\|_2^2}}. \tag{3.6}$$

Lemma 3.2.1 and the fact that $\gamma_j^q > 0$ yield

$$\gamma_j^q\|\Omega_2\Omega_1^\dagger\|_2 \leq \gamma_j^q C_g.$$

Let us define a function, $f(t) := \frac{t}{\sqrt{1+t^2}}$ for $t \in (0,\infty)$, which obviously is an increasing function on the interval $(0,\infty)$. Since $\gamma_j^q\|\Omega_2\Omega_1^\dagger\|_2 \leq \gamma_j^q C_g$, we get

$$f(\gamma_j^q\|\Omega_2\Omega_1^\dagger\|_2) \leq f(\gamma_j^q C_g),$$

and consequently

$$\frac{\gamma_j^q\|\Omega_2\Omega_1^\dagger\|_2}{\sqrt{1+\gamma_j^{2q}\|\Omega_2\Omega_1^\dagger\|_2^2}} \leq \frac{\gamma_j^q C_g}{\sqrt{1+\gamma_j^{2q}C_g^2}}. \tag{3.7}$$

Combining equations (3.6) and (3.7) completes the first part of the proof.

(ii) Following the first part of the proof, we get

$$\frac{\gamma_j^q \|\Omega_2\Omega_1^\dagger\|_2}{\sqrt{1+\gamma_j^{2q}\|\Omega_2\Omega_1^\dagger\|_2^2}} \leq \frac{\gamma_j^q C_g}{\sqrt{1+\gamma_j^{2q}C_g^2}}. \tag{3.8}$$

Equation (3.8) and the result of Theorem 3.1.2 together yield the desired result

$$\|(I-\mathscr{P}_{\hat{U}})\mathbf{u}_j\|_2 \leq \frac{\gamma_j^{2q}\|\Omega_2\Omega_1^\dagger\|_2}{\sqrt{1+\gamma_j^{2q}\|\Omega_2\Omega_1^\dagger\|_2^2}}.$$

□

Before we present our next and final result, we need the following Lemma.

Lemma 3.2.3. *Let $\Omega \in \mathbb{R}^{n\times\ell}$ be a standard Gaussian matrix such that rank$(\Omega_1) = k$, with $\ell = k$ and Ω_1 is invertible. Then with probability $1-\delta$*

$$\|\Omega_2\Omega_1^{-1}\|_2 \leq N_g,$$

with $N_g = \frac{2\sqrt{k}}{\delta}\left(\sqrt{n-k}+\sqrt{k}+\sqrt{2\log\frac{2}{\delta}}\right)$.

Proof. First, by the submultiplicativity of the 2-norm

$$\|\Omega_2\Omega_1^{-1}\|_2 \leq \|\Omega_2\|_2\|\Omega_1^{-1}\|_2. \tag{3.9}$$

From [13, Lemma2], we have

$$\|\Omega_2\|_2 \leq \left(\sqrt{n-k}+\sqrt{k}+\sqrt{2\log\frac{2}{\delta}}\right),$$

with the probability of failure at most $\delta/2$. Since Ω_1 is a standard Gaussian random matrix and

$\mathsf{rank}\,(\Omega_1) = k$, by [15, Theorem 3.4] we have that the probability

$$P\big(\|{\Omega_1}^{-1}\|_2 \geq x\big) \leq \frac{\sqrt{k}}{x}. \tag{3.10}$$

Choosing $x = \frac{2\sqrt{k}}{\delta}$ in (3.10), we get the following

$$P(\|\Omega_1^{-1}\|_2 \geq \frac{2\sqrt{k}}{\delta}) \leq \frac{\delta}{2}.$$

Hence, with the probability of failure at most $\frac{\delta}{2}$ we have that

$$\|\Omega_1^{-1}\|_2 \leq \frac{2\sqrt{k}}{\delta}. \tag{3.11}$$

Combining (3.9) and (3.11), and using the union bound, we get

$$P\Big(\|\Omega_2\|_2\|\Omega_1^{-1}\|_2 \geq N_g\Big) \leq \delta/2 + \delta/2 = \delta,$$

with $N_g = \frac{2\sqrt{k}}{\delta}\Big(\sqrt{n-k} + \sqrt{k} + \sqrt{2\log\frac{2}{\delta}}\Big)$. Consequently, with the probability at least $1-\delta$, we have the desired bound

$$\|\Omega_2\Omega_1^{-1}\|_2 \leq \|\Omega_2\|_2\|\Omega_1^{-1}\|_2 \leq N_g,$$

where $N_g = \frac{2\sqrt{k}}{\delta}\Big(\sqrt{n-k} + \sqrt{k} + \sqrt{2\log\frac{2}{\delta}}\Big)$. □

Our final result gives the probabilistic bounds for the accuracy of eigenvalues of matrix A computed using the randomized subspace iteration method presented in Algorithm 2.

Theorem 3.2.4. *Assume $k = \ell$ and $\Omega_1 \in \mathbb{R}^{k\times k}$ is invertible. Let $\lambda_1, \lambda_2, \ldots, \lambda_k$ be the eigenvalues of matrix A, and $\widehat{\lambda}_1, \widehat{\lambda}_2, \ldots, \widehat{\lambda}_k$ be the eigenvales of $Q^T AQ$ (Ritz values). Then, by Lemma 3.2.3, for $j = 1, \ldots, k$ the following result holds with the probability at least $1-\delta$*

$$\max_{1\leq j\leq k} |\lambda_j - \widehat{\lambda}_j| \leq (\lambda_{\max} - \lambda_{\min})\frac{\gamma_k^q N_g}{\sqrt{1+\gamma_k^{2q}{N_g}^2}},$$

with

$$N_g = \frac{2\sqrt{k}}{\delta}\Big(\sqrt{n-k}+\sqrt{k}+\sqrt{2\log\frac{2}{\delta}}\Big).$$

Proof. From Lemma 3.2.3, we have that

$$\|\Omega_2\Omega_1^{-1}\|_2 \leq N_g.$$

Following the proof of Theorem 3.2.2 part (i), with an increasing function $f(t) := \frac{t}{\sqrt{1+t^2}}$ on $(0,\infty)$ and $\gamma_j > 0$, we have that

$$\frac{\gamma_k^q\|\Omega_2\Omega_1^{-1}\|_2}{\sqrt{1+\gamma_k^{2q}\|\Omega_2\Omega_1^{-1}\|_2^2}} \leq \frac{\gamma_k^q N_g}{\sqrt{1+\gamma_k^{2q}N_g^2}}. \tag{3.12}$$

Then, from (3.12) and Theorem 3.1.3, we obtain the final result

$$\max_{1\leq j\leq k}|\lambda_j-\widehat{\lambda}_j| \leq (\lambda_{\max}-\lambda_{\min})\frac{\gamma_k^q N_g}{\sqrt{1+\gamma_k^{2q}{N_g}^2}},$$

with

$$N_g = \frac{2\sqrt{k}}{\delta}(\sqrt{n-k}+\sqrt{k}+\sqrt{2\log\frac{2}{\delta}}).$$

□

Chapter 4

Numerical Experiments

In this chapter we will numerically illustrate the angles between the subspaces range($\mathrm{U_k}$) and range($\widehat{\mathrm{U}}_\mathrm{k}$) and the error bounds presented in Theorem 3.1.1. We will follow numerical experiments performed in [12] and adapt the MATLAB implementation of randomized singular value decomposition (`randsvd`) provided in [11] to use randomized subspace iteration for solving symmetric positive definite eigenvalue problem. All presented numerical experiments were performed using MATLAB 9.4.0.813654 (R2018a).

4.1 Example 1: Low-rank Plus Noise

For our first numerical example, we will illustrate the the angles between the subspaces range($\mathrm{U_k}$) and range($\widehat{\mathrm{U}}_\mathrm{k}$) and the error bounds introduced in Theorem 3.1.1 using the *low-rank plus noise* test matrix introduced in [12]. Given a Gaussian random matrix $G \in \mathbb{R}^{n\times n}$, we construct a positive definite matrix $A \in \mathbb{R}^{n\times n}$

$$A = \frac{1}{2}(\Delta + \Delta^T), \tag{4.1}$$

with $\Delta \in \mathbb{R}^{n\times n}$ given by

$$\Delta = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} + \sqrt{\frac{noise \cdot r}{2n^2}}(G + G^T).$$

Here, *noise* denotes the size of the noise introduced into the matrix, and r denotes the location of the *eigenvalue gap*. To verify our theoretical results we will present the computed values of *canonical angles* $\sin\theta_j$ for $j = 1,\ldots,k$ as well as the upper bounds provided in Theorem 3.1.1. We will consider three values for parameter *noise*, i.e.,

Small Noise with $noise = 10^{-2}$,

Medium Noise with $noise = 10^{-1}$,

Large Noise with $noise = 1$.

We further consider matrix of size $n = 300$, the *oversampling parameter* $p = 20$, and the *target rank* $k = 25$. For our experiment, we consider Ω to be a Gaussian random matrix of dimension $n \times (k+p)$ generated using MATLAB command `randn(n,k+p)`. The location of the *eigenvalue gap* is chosen as $r = 15$, i.e., there will be a gap between the eigenvalue λ_{15} and λ_{16}.

We have performed $q = 1, 2, 3$ steps of randomized subspace iteration method on these matrices with varied noises in order to check the validity of our proposed angle bounds.

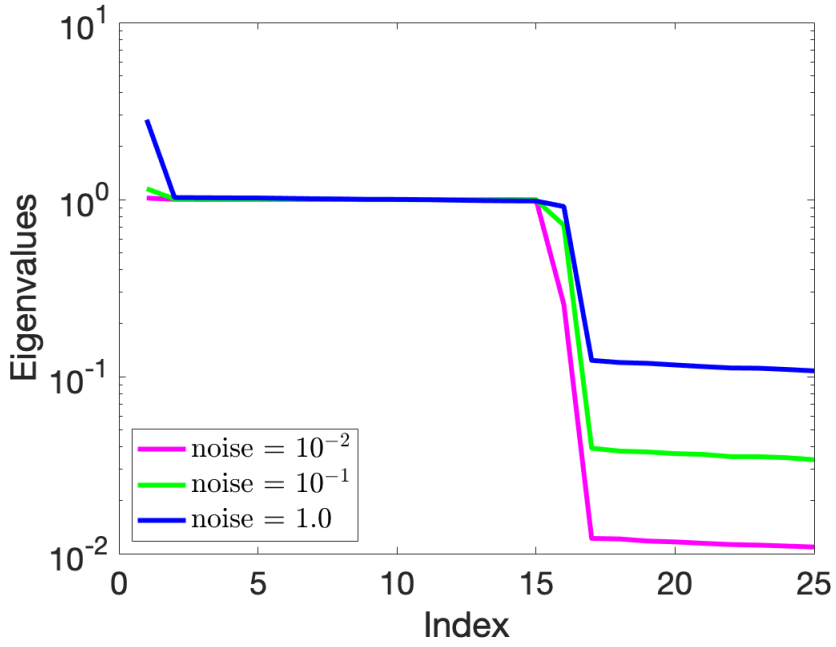
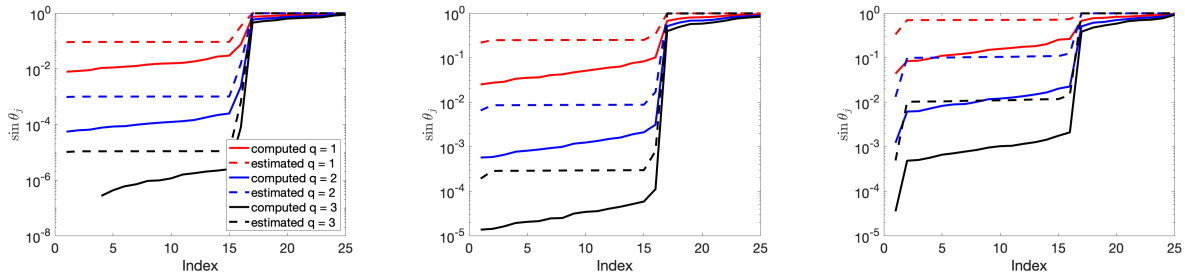


Figure 4.1: $k = 25$ largest eigenvalues of the matrix A defined in (4.1).

Figure 4.1 depicts the first $k = 25$ eigenvalues of the *low-rank plus noise* matrices defined in (4.1). It can be observed that for every case there is a sharp decay in the eigenvalues after the index 15, where the *eigenvalue gap* is located.

Figure 4.2 illustrates the computed *canonical angles* and the angle bounds presented in Theorem 3.1.1 for different noise sizes. The *target rank* is $k = 25$ and the *oversampling parameter*



(a) Small Noise $noise = 10^{-2}$ (b) Medium Noise $noise = 10^{-1}$ (c) Large Noise $noise = 1$

Figure 4.2: Plots of $\sin\theta_j$ for $j = 1,\ldots,k$ for the test matrices *Low-rank plus noise*. (4.1).

is $p = 20$. The *solid lines* correspond to the *computed values*, and the *dashed lines* correspond to the bounds obtained using Theorem 3.1.1. The parameter q corresponds to the number of steps in randomized subspace iteration method. It can be observed that the sines of the computed *canonical angles* are smaller than the angle bounds for all the three cases. The Small Noise is quite close to a low rank matrix and there is a large *eigenvalue gap* at index 15. For this example, the bounds are qualitatively good. As the level of noise increases the *eigenvalue gap* decreases, therefore the computed angles increase as predicted by Theorem 3.1.1. The bounds are not informative for the $q = 1$ but are qualitatively good for $q = 2$ and $q = 3$. In this experiment, we have illustrated the fact that the sines of computed *canonical angles* between the exact k-dimensinal eigenspace U_k and the approximated eigenspace $\widehat{U}_k$ are indeed bounded from above by the right-hand side $\frac{\gamma_j^q \|\Omega_2\Omega_1^\dagger\|_2}{\sqrt{1+\gamma_j^{2q}\|\Omega_2\Omega_1^\dagger\|_2^2}}$ of Theorem 3.1.1. These observations confirm our theoretical results.

4.2 Example 2: Controlled Gap

For our second numerical example, we will illustrate the angles between the subspaces range$(\mathrm{U_k})$ and range$(\widehat{\mathrm{U}}_\mathrm{k})$ and the error bounds established in Theorem 3.1.1 using the *controlled gap* test matrix introduced in [12]. We construct a positive definite matrix $A \in \mathbb{R}^{n\times n}$

$$A = \frac{1}{2}(\Delta + \Delta^T), \tag{4.2}$$

with a sparse matrix $\Delta \in \mathbb{R}^{300\times 300}$ given by

$$\Delta = \sum_{j=1}^{r} \frac{gap}{j} x_j x_j^T + \sum_{j=r+1}^{300} \frac{1}{j} x_j x_j^T,$$

where $x_j \in \mathbb{R}^{300}$ is a sparse random vectors with non-negative entries generated using the MATLAB command `sprand(300,1,0.25)`. The eigenvalues decay like $\frac{1}{j}$ and the gap between the eigenvalues λ_{15} and λ_{16} is controlled by the parameter *gap*.

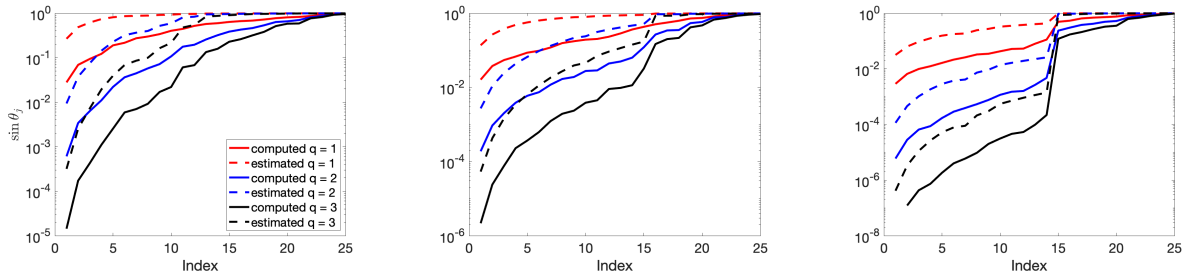
To verify our theoretical results we will present the sines of the computed *canonical angles* $\sin\theta_j$ for $j = 1,\ldots,k$ along with the upper bounds given in Theorem 3.1.1. We will consider three values for parameter *gap*, i.e.,

Small Gap with $gap = 1$,

Medium Gap with $gap = 2$,

Large Gap with $gap = 10$.

Here again we consider the *oversampling parameter* $p = 20$ and the *target rank* $k = 25$. For this example, we use the random matrix Ω as described in Example 1. The location of the *eigenvalue gap* is chosen as $r = 15$, i.e., there will be a gap between the eigenvalue λ_{15} and λ_{16}. We have performed $q = 1,2,3$ steps of the randomized subspace iteration method on these matrices with different gaps in order to verify our theoretical bounds.



(a) Small Gap $gap = 1$ (b) Medium Gap $gap = 2$ (c) Large Gap $gap = 10$

Figure 4.4: Plots of $\sin\theta_j$ for $j = 1,\ldots,k$ for the test matrices *controlled gap* defined in (4.2).

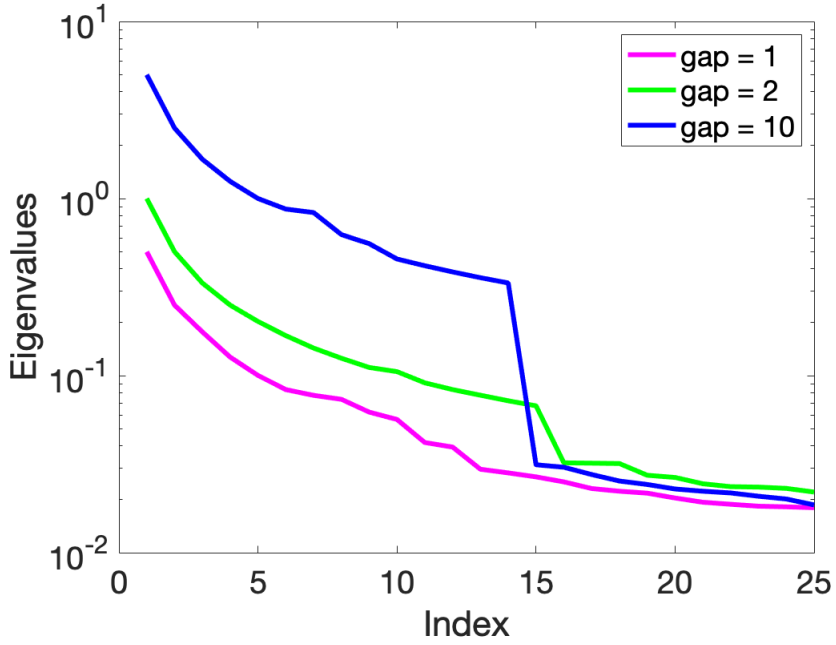


Figure 4.3: $k = 25$ largest eigenvalues of the matrix A defined in (4.2).

Figure 4.3 shows the k largest eigenvalues of controlled gap matrices defined as in (4.2) for different values of parameter *gap*. It is easy to see that there is a slow decay in the eigenvalues for Small Gap and Medium Gap test matrices, but the eigenvalues decrease rapidly for the Large Gap test matrix near the index 15. Figure 4.4 illustrates the computed *canonical angle* and the angle bounds given in Theorem 3.1.1 for different gaps.The *target rank* is $k = 25$ and the *oversampling parameter* is $p = 20$. The parameter q corresponds to the number of steps in randomized subspace iteration method.The *solid lines* correspond to the *computed values*, and the *dashed lines* correspond to the bounds obtained using Theorem 3.1.1. It can be observed that the the angle bounds

are in accordance with the Theorem 3.1.1.The angle bounds are closer to the computed *canonical angles* in all the three cases for every iteration. As we increase the gap and number of iterations, the computed *canonical angles* decrease.The text matrices Medium Gap and Large Gap show a decay in eigenvalues as well as show the prominent *eigenvalue gap* between the eigenvalues λ_{15} and λ_{16}. These matrices satisfy the assumptions of our analysis, and hence forth the bounds are expected to be nice. Small Gap has shown a decay in the eigenvalues but there is no prominent *eigenvalue gap*. Even in this case the bounds are somewhat qualitatively good except for $q = 1$. Also note that the bounds in this experiment are sharper than the bounds given in the first experiment for Low-rank noise. Through this experiment, we have illustrated the fact that the computed *canonical angles* between the exact k-dimensional eigenspace U_k and the approximated eigenspace $\widehat{U}_k$ are indeed bounded from above by the angle bound $\frac{\gamma_j^q \|\Omega_2 \Omega_1^\dagger\|_2}{\sqrt{1+\gamma_j^{2q}\|\Omega_2 \Omega_1^\dagger\|_2^2}}$ of Theorem 3.1.1. These observations confirm our theoretical results.

4.3 Observations From Numerical Experiments

We plot the *canonical angles* $\sin\angle(U_k, \widehat{U})$ in solid lines; the corresponding bounds from Theorem 3.1.1 are also plotted in dashed lines. The results are depicted in Figure 4.2 and Figure 4.4 . We make the general observations here:

- The influence of the number of steps in the randomized subspace iteration method on the *canonical angles* is clear from the plots for both the experiments. The angles become smaller as the number of iterations q increases, implying that the resulting subspace is becoming closer to the exact eigenspace.
- A large *eigenvalue gap* in the spectrum, implies that the *canonical angles* below the index locating the gap are captured accurately. This is prominently observed in the Figure 4.4a, in which there is a large gap between eigenvalues λ_{15} and λ_{16}. We can make similar observations in other figures.
- The bounds depicted for most experiments are qualitatively informative, but in some experi-

ments the bounds are quantitatively accurate as well, e.g., Large Gap matrix..

- As the decay rate of eigenvalues increases, it can be observed that the corresponding *canonical angles* are becoming smaller.

Chapter 5

Conclusions and Future Research

In this work we have presented the deterministic and probabilistic bounds for using the randomized subspace iteration for symmetric positive definite eigenvalue problems. We have proposed deterministic and bounds for the canonical angles between the exact and the approximate eigenvector subspaces, the sine of angle between the eigenvectors of the exact eigenvector subspace and the approximated eigenvector subspace, and for the accuracy of eigenvalues using the randomized subspace iteration. We have also shown the validity of our theoretical angle bounds by implementing two numerical examples. We were able to successfully show the similar illustrations as given in [12] but with symmetric positive definite matrices and using randomized subspace iteration for the SPDEVP. This guarantees that our results are in consistence with the results presented in [12] for obtaining the singular value decomposition (SVD) using the randomized subspace iteration method. We can further our analysis by presenting similar results for the generalized eigenvalue problems. We are currently working in that direction developing and analyzing the randomized FEAST algorithm for the generalized eigenvalue problems.

Bibliography

[1] Petros Drineas, Ilse C. F. Ipsen, Eugenia-Maria Kontopoulou, and Malik Magdon-Ismail. Structural convergence results for approximation of dominant subspaces from block Krylov spaces. *SIAM J. Matrix Anal. Appl.*, 39(2):567–586, 2018.

[2] Gene H. Golub and Charles F. van Loan. *Matrix Computations*. JHU Press, fourth edition, 2013.

[3] Nathan Halko, Per-Gunnar Martinsson, and Joel A. Tropp. Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions. *SIAM Review*, 53(2):217–288, 2011.

[4] Ravindran Kannan and Santosh Vempala. Randomized algorithms in numerical linear algebra. *Acta Numerica*, 26:95–135, 05 2017.

[5] Nikita Kapur, Agnieszka Międlar, Arvind K. Saibaba, and Eric de Sturler. Randomized FEAST Algorithm, 2021. In preparation.

[6] Andrew V. Knyazev and Merico E. Argentati. On proximity of Rayleigh quotients for different vectors and Ritz values generated by different trial subspaces. *Linear Algebra and its Applications*, 415(1):82–95, 2006.

[7] Michael W Mahoney. Randomized algorithms for matrices and data. *arXiv:1104.5557*, 2011.

[8] Per-Gunnar Martinson and Joel A. Tropp. Randomized Numerical Linear Algebra: Foundations & Algorithms. *arXiv:2002.01387*, 2020.

[9] Rajeev Motwani and Prabhakar Raghavan. *Randomized Algorithms*. Cambridge University Press, Cambridge; NY, 1995.

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

[11] Arvind. K. Saibaba. `https://github.com/arvindks/randsvs`.

[12] Arvind K. Saibaba. Randomized subspace iteration: Analysis of canonical angles and unitarily invariant norms. *SIAM J. Matrix Anal. Appl.*, 40(1):23–48, 2019.

[13] Arvind K. Saibaba, Joseph Hart, and Bart van Bloemen Waanders. Randomized algorithms for generalized singular value decomposition with application to sensitivity analysis. *arXiv preprint arXiv:2002.02812*, 2020.

[14] Arvind K. Saibaba, Jonghyun Lee, and Peter K. Kitanidis. Randomized algorithms for generalized Hermitian eigenvalue problems with application to computing Karhunen-Loève expansion. *Numer. Linear Algebra Appl.*, 23(2):314–339, 2016.

[15] Arvind Sankar, Daniel A. Spielman, and Shang-Hua Teng. Smoothed analysis of the condition numbers and growth factors of matrices. *SIAM J. Matrix Anal. Appl.*, 28(2):446–476, 2006.

[16] Joel A. Tropp. An Introduction to Matrix Concentration Inequalities. *arxiv:1501.01571*, 2015. To appear in Foundations and Trends in Machine Learning.

[17] Joel A. Tropp, Alp Yurtsever, Madeleine Udell, and Volkan Cevher. Practical sketching algorithms for low-rank approximation. *SIAM J. Matrix Anal. Appl.*, 38(4):1454–1485, 2017.