MATHEMATICAL ENGINEERING TECHNICAL REPORTS

Global Convergence of the Restarted Lanczos Method and Jacobi-Davidson Method for Symmetric Eigenvalue Problems

Kensuke AISHIMA

METR 2013–27 October 2013

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

Global Convergence of the Restarted Lanczos Method and Jacobi-Davidson Method for Symmetric Eigenvalue Problems

Kensuke AISHIMA

Department of Mathematical Informatics
Graduate School of Information Science and Technology
The University of Tokyo
Kensuke_Aishima@mist.i.u-tokyo.ac.jp

October, 2013

Abstract

The Lanczos method is well known to compute the extremal eigenvalues of symmetric matrices. For efficiency and robustness a restart strategy is employed in practice, but this makes the convergence analysis less straightforward. We prove global convergence of the restarted Lanczos method in exact arithmetic by using certain convergence properties of the Rayleigh-Ritz procedure due to Crouzeix, Philippe and Sadkane. For the restarted Lanczos, Sorensen's previous analysis establishes global convergence to the largest eigenvalues under the technical assumption that the absolute values of the off-diagonal elements of the Lanczos tridiagonal matrix are larger than a positive constant throughout the iterations. In this paper, we prove global convergence without any such assumption. The only assumption is that the initial vector is not orthogonal to the exact eigenvectors. Our analysis covers a dynamic restarting procedure where the restarting points are dynamically determined. The convergence theorem is extended to restarted Lanczos for computing both the largest and smallest eigenvalues. Moreover, we derive certain global convergence theorems of the block Lanczos method and the Jacobi-Davidson method: for both algorithms, the Ritz values converge to exact eigenvalues, although not necessarily to the extremal ones.

1 Introduction

Suppose one wants to compute one or more extremal eigenvalues and their corresponding eigenvectors of a symmetric matrix A. There exist a number of efficient iterative methods for the task. Among them, the Lanczos method [17] is a classical and powerful technique [3, 9, 11, 13, 22]. The Lanczos method performs the so-called Rayleigh-Ritz procedure on a Krylov subspace. In order to reduce the computational and memory costs, a restart strategy [25, 28, 29] should be employed from the practical point of view. In fact, the restarted Lanczos method is currently implemented in MATLAB's built-in function `eigs`.

In this paper, we prove global convergence of the restarted Lanczos method. Key tools for our proof are certain convergence properties of the Rayleigh-Ritz procedure due to Crouzeix, Philippe and Sadkane [5, §2]. As well as Lanczos, there are a number of efficient algorithms using the Rayleigh-Ritz procedure, including Davidson's method [5, 8, 18], the Jacobi-Davidson method [24] proposed by Sleijpen and van der Vorst, and LOBPCG [15] proposed by Knyazev. The block Lanczos method [6, 7, 10], which is mathematically equivalent to the band Lanczos [23], is also effective for computing multiple eigenvalues. Among them, we prove certain global convergence properties of the restarted block Lanczos method and Jacobi-Davidson method.

Existing studies of the global convergence of the restarted Lanczos method can be summarized as follows. In 1951, Karush derived global convergence for the restarted strategy to compute one largest eigenvalue [14], and Knyazev and Skorokhodov gave its convergence proof based on certain properties of the steepest descent method [16]. To the best of the author's knowledge, the most general result about the global convergence is Sorensen's theorem of the implicitly restarted Lanczos method for computing the largest more than one eigenvalues [25]. However, in Sorensen's theorem there is a technical assumption that the absolute values of the off-diagonal elements of the Lanczos tridiagonal matrix are larger than a positive constant throughout the iterations.

In this paper, we prove global convergence without any such assumption. The only assumption is that the initial vector is not orthogonal to the exact eigenvectors. In addition, the convergence theorem is extended to restarted Lanczos for computing both the largest and smallest eigenvalues. As for the restarted block Lanczos method [2, 7], some convergence property of a restarted strategy to compute the largest eigenvalues has been already proven by Crouzeix, Philippe and Sadkane [5, Corollary 2.2]. More specifically, the Ritz values converge to exact eigenvalues, although not necessarily to the largest ones. We extend this result to the restart strategy in [6] for computing both the largest and smallest eigenvalues.

Regarding the Jacobi-Davidson method, the local asymptotic convergence rate has been well studied [19, 20, 21, 24]. However, despite many efforts over a decade, the theory for global convergence remains an open problem. On the other hand, in the Davidson method with suitable starting vectors, the largest Ritz values always converge to eigenvalues, although not necessarily to the largest ones. This is proven by Crouzeix, Philippe and Sadkane [5]. In [5], they have proven global convergence of the restarted Rayleigh-Ritz procedure under some general assumptions [5, Theorem 2.1]. As a corollary, they establish the convergence of the Davidson method. Actually, by carefully reading [5, Theorem 2.1], we can prove a global convergence property of the Jacobi-Davidson with one restarting vector. This is a proof by contradiction. In this paper, we slightly modify the convergence theorem [5, Theorem 2.1] in order to provide a direct convergence proof for multiple restarting vectors. See Section 4 for the details. Specifically, we prove that the largest Ritz value of the restarted Jacobi-Davidson converges to an exact eigenvalue, although not necessarily to the largest one.

It is worth noting that, there are a number of improved versions of the Lanczos and Jacobi-Davidson combined with other effective restarted strategies [1, 3, 4, 26, 27, 28, 29]. However, in this paper, we investigate global convergence properties of the typical and basic versions. We also note that our results cover a dynamic restarting procedure [5], where the restarting points are dynamically determined. Moreover, our results can be extended to the generalized symmetric eigenvalue problems $Ax = \lambda Bx$, where A is a symmetric matrix and B is a symmetric positive definite matrix, because the Lanczos and the Jacobi-Davidson for $Ax = \lambda Bx$ with the B-orthogonal basis are mathematically equivalent to those for $L^{-1}AL^{-\mathrm{T}}$ with the standard orthogonal basis, where L is the Cholesky factor of B [3, 22].

This paper is organized as follows. Section 2 is devoted to a description of the Lanczos method. In Section 3, Sorensen's convergence analysis for the restarted Lanczos is briefly summarized. We present certain global convergence properties of the restarted Rayleigh-Ritz procedure in Section 4. We use them to prove global convergence of the restarted Lanczos in Section 5. This result includes Sorensen's convergence theorem. In addition, global convergence properties of the restarted block Lanczos and Jacobi-Davidson methods are shown in Section 6 and Section 7, respectively.

Notation and assumptions. Throughout this paper, A is an $N \times N$ symmetric matrix whose eigenvalues are $\lambda_1 \geq \cdots \geq \lambda_N$ with corresponding normalized eigenvectors $x_1, \ldots, x_N$. However, multiple eigenvalues need some care. See Reark 1 for details. $[x_1, \ldots, x_k]$ denotes the matrix whose ith column is x_i. I is the identity matrix, and O is the zero matrix. Let $k(= k_1 + k_2)$ be the number of desired eigenvalues, where k_1 denotes the largest ones and k_2 denotes the smallest ones.

2 Lanczos method

This section is devoted to a description of the Lanczos method. The Lanczos method starts with a properly chosen starting vector v and builds up an orthonormal basis V_m of the Krylov subspace,

$$\mathcal{K}^m(A, v) = \mathrm{span}\{v, Av, A^2v, \ldots, A^{m-1}v\} = \mathrm{span}\{V_m\}. \tag{1}$$

V_m is given by the Gram-Schmidt orthogonalization for $v, Av, A^2v, \ldots, A^{m-1}v$. The Lanczos method is summarized as follows.

1. Compute the orthonormal basis $\{v_i\}_{i=1,\ldots,m}$ of $\mathcal{K}^m(A, v)$. Let $V_m = [v_1, v_2, \ldots, v_m]$.
2. Compute $T_m = V_m^{\mathrm{T}} A V_m$.
3. Compute the eigenvalues of T_m and select the k desired ones $\theta_i, i = 1, 2, \ldots, k$, where $k \leq m$ (for instance the largest ones).
4. Compute the eigenvectors $y_i, i = 1, \ldots, k$, of T_m associated with $\theta_i, i = 1, \ldots, k$, and the corresponding approximate eigenvectors of A, $u_i = V y_i, i = 1, \ldots, k$.

If the residuals $\|Au_i - \theta_i u_i\|$ for $i = 1, \ldots, k$ are sufficiently small, approximate eigenpairs θ_i and u_i for $i = 1, \ldots, k$ are obtained. In the Lanczos method, $\theta_i, i = 1, 2, \ldots, k$ are referred to as the Ritz values and the vectors $u_i, i = 1, \ldots, k$ are the associated Ritz vectors, as the Lanczos method is the Rayleigh-Ritz procedure on the Krylov subspace [3, 22]. See Section 4 for the Rayleigh-Ritz procedure. Importantly, T_m is a tridiagonal matrix due to a special feature of the Krylov subspace. The Lanczos method is an effective method but, from the practical point of view, the restart strategy is needed in order to reduce the computational costs [3, 25]. In other words, if the residuals $\|Au_i - \theta_i u_i\|$ for $i = 1, \ldots, k$ are not sufficiently small, the Ritz pairs should be computed again by another Krylov subspace that is generated by some refined starting vector.

There exist a number of restarting strategies. Among them, we focus on the following restarted Lanczos method for computing the k largest eigenvalues [25]. Let $m_\ell \geq k + 1$ be the maximum iteration numbers of the inner loop for $\ell = 0, 1, \ldots$.

Algorithm 1 The restarted Lanczos method

Initialization pick a unit vector $v_1^{(0)}$

1: compute the orthonormal basis $V_{m_0}^{(0)} = [v_1^{(0)}, \ldots, v_{m_0}^{(0)}]$ of $\mathcal{K}^{m_0}(A, v_1^{(0)})$ by the Lanczos process

2: **for** $\ell := 0, 1, \ldots,$ **do**

3: compute $T_{m_\ell}^{(\ell)} = {V_{m_\ell}^{(\ell)}}^{\mathrm{T}} A V_{m_\ell}^{(\ell)}$

4: compute eigenvalues of $T_{m_\ell}^{(\ell)}$: $\theta_1^{(\ell)} \geq \cdots \geq \theta_k^{(\ell)} \geq \mu_1^{(\ell)} \cdots \geq \mu_{m_\ell - k}^{(\ell)}$

5: compute the QR decomposition: $\widehat{Q}^{(\ell)} \widehat{R}^{(\ell)} = (T_{m_\ell}^{(\ell)} - \mu_1^{(\ell)} I) \cdots (T_{m_\ell}^{(\ell)} - \mu_{m_\ell - k}^{(\ell)} I)$

6: compute $V_k^{(\ell+1)} = [v_1^{(\ell)}, \ldots, v_{m_\ell}^{(\ell)}] \widehat{Q}_k^{(\ell)}$, where $\widehat{Q}_k^{(\ell)}$ denotes the matrix of the first k columns of $\widehat{Q}^{(\ell)}$ (Then $V_k^{(\ell+1)} = [u_1^{(\ell)}, \ldots, u_k^{(\ell)}] Q^{(\ell)}$, where $u_1^{(\ell)}, \ldots, u_k^{(\ell)}$ are the Ritz vectors corresponding to $\theta_1^{(\ell)} \geq \cdots \geq \theta_k^{(\ell)}$ and $Q^{(\ell)}$ is some orthogonal matrix)

7: compute $v_{k+1}^{(\ell+1)}, \ldots, v_{m_{\ell+1}}^{(\ell+1)}$ by the Lanczos process to obtain a new orthonormal basis $V_{m_{\ell+1}}^{(\ell+1)} = [v_1^{(\ell+1)}, \ldots, v_{m_{\ell+1}}^{(\ell+1)}]$

8: **end for**

Remark 1. *Multiple eigenvalues need some care: the Lanczos method is unable to compute the multiplicity correctly [3]. Hence, let $\lambda_1 > \cdots > \lambda_n$ denote all the distinct eigenvalues of $A \in \mathbb{R}^{N \times N}$ in this section. Let x_i $(i = 1, \ldots, n)$ denote normalized vectors in the eigenspaces corresponding λ_i for $i = 1, \ldots, n$. Also in Sections 3 and 5, we use the same notation.* ■

It is empirically known that

$$T_k^{(\ell)} \approx \mathrm{diag}(\lambda_1, \ldots, \lambda_k), \tag{2}$$

$$V_k^{(\ell)} \approx X_k = [x_1, \ldots, x_k] \tag{3}$$

for sufficiently large ℓ. Therefore, the k largest eigenvalues $\lambda_1, \ldots, \lambda_k$ and the corresponding eigenvectors $x_1, \ldots, x_k$ are obtained by Algorithm 1.

Algorithm 1 is mathematically equivalent to Sorensen's implicitly restarted Lanczos method [25] with the so-called exact shifts. Furthermore, it is also mathematically equivalent to the thick restarted Lanczos method [28] that is a more recently proposed efficient algorithm. The aim of this paper is a theoretical convergence analysis, and hence we discuss the convergence behavior of Algorithm 1 in the following sections.

3 Convergence analysis of the restarted Lanczos method by Sorensen

In this section, we summarize the convergence analysis of the restarted Lanczos method by Sorensen [25]. Firstly, the following simple property can be stated.

Lemma 1 ([25])**.** *Suppose m_l is a fixed m for all ℓ. Then, each $\{\theta_i^{(\ell)} : \ell = 1, 2, \ldots\}$ is an increasing convergent sequence for each $i = 1, 2, \ldots, k$.* ■

Lemma 1 is easily extended to the restarted Rayleigh-Ritz procedure. See Lemma 2 in Section 4 for its proof. Based on Lemma 1, Sorensen proved the next theorem that states global convergence under a certain assumption.

Theorem 1 ([25])**.** *Let m_l be a fixed m for all ℓ. Suppose that the initial starting vector $v_1^{(0)}$ satisfies $x_j^{\mathrm{T}} v_1^{(0)} \neq 0$ for $j = 1, 2, \ldots, k$, where x_j is the eigenvector of A corresponding to the eigenvalue λ_j with the eigenvalues of A listed in descending order. Let $\beta_i^{(\ell)}$ be the ith subdiagonal element of $T_k^{(\ell)}$ and assume that $\exists \epsilon > 0$ s.t. $|\beta_i^{(\ell)}| > \epsilon$ for all i, ℓ. Then the sequences $\theta_j^{(\ell)} \to \theta_j = \lambda_j$ as $\ell \to \infty$.* ■

The assumption that subdiagonal elements of $T_k^{(\ell)}$ are not smaller than ϵ is not restrictive from the practical point of view because the so-called deflation is applied when a subdiagonal element becomes smaller than machine epsilon ϵ in practice.

However, as mentioned before, the more specific convergence properties (2) and (3) are empirically known. The aim of this paper is to prove the convergence properties (2) and (3) without the assumptions on the subdiagonal elements of $T_k^{(\ell)}$. Indeed, $\beta_i^{(\ell)} \to 0$.

4 Convergence property of the restarted Rayleigh-Ritz procedure

In this section we investigate the convergence behavior of the restarted Rayleigh-Ritz procedure. The Rayleigh-Ritz procedure is described as follows.

1. Compute an orthonormal basis $\{v_i\}_{i=1,\ldots,m}$. Let $V = [v_1, v_2, \ldots, v_m]$.
2. Compute $B = V^{\mathrm{T}} AV$.
3. Compute the eigenvalues of B and select the k desired ones $\theta_i, i = 1, 2, \ldots, k$, where $k \leq m$ (for instance the largest ones).
4. Compute the eigenvectors $y_i, i = 1, \ldots, k$, of B associated with $\theta_i, i = 1, \ldots, k$, and the corresponding approximate eigenvectors of A, $u_i = V y_i, i = 1, \ldots, k$.

It is easily seen that the Lanczos method is the Rayleigh-Ritz procedure on the Krylov subspace. As in the Lanczos method, the restart strategy is incorporated into the Rayleigh-Ritz procedure. Although Algorithm 1 computes the k largest eigenvalues, the aim of this paper is not only to prove the convergence of Algorithm 1 but also to discuss the convergence behavior of the other algorithms for computing other eigenvalues. Hence, we focus on the following restart strategy to compute both the k_1 largest eigenvalues and the k_2 smallest eigenvalues.

[The restarted Rayleigh-Ritz procedure]

1: compute an orthonormal basis $V_{m_0}^{(0)} = [v_1^{(0)}, \ldots, v_{m_0}^{(0)}]$
2: **for** $\ell := 0, 1, \ldots,$ **do**
3: compute $B^{(\ell)} = V_{m_\ell}^{(\ell)\mathrm{T}} A V_{m_\ell}^{(\ell)}$
4: compute eigenvalues of $B^{(\ell)}$: $\theta_1^{(\ell)} \geq \cdots \geq \theta_{k_1}^{(\ell)} \geq \mu_1^{(\ell)} \cdots \geq \mu_{m_\ell - k}^{(\ell)} \geq \vartheta_1^{(\ell)} \geq \cdots \geq \vartheta_{k_2}^{(\ell)}$ $(k = k_1 + k_2)$
5: compute the Ritz vectors $u_1^{(\ell)}, \ldots, u_k^{(\ell)}$ associated with $\theta_1^{(\ell)} \geq \cdots \geq \theta_k^{(\ell)} \geq \vartheta_1^{(\ell)} \geq \cdots \geq \vartheta_{k_2}^{(\ell)}$
6: compute $[v_1^{(\ell+1)}, \ldots, v_k^{(\ell+1)}] = [u_1^{(\ell)}, \ldots, u_k^{(\ell)}]Q^{(\ell)}$, where $Q^{(\ell)}$ is some orthogonal matrix
7: compute $v_{k+1}^{(\ell+1)}, \ldots, v_{m_{\ell+1}}^{(\ell+1)}$ to obtain a new orthonormal basis $V_{m_{\ell+1}}^{(\ell+1)} = [v_1^{(\ell+1)}, \ldots, v_{m_{\ell+1}}^{(\ell+1)}]$
8: **end for**

For the convergence analysis, write the matrix $B^{(\ell)}$ as

$$B^{(\ell)} = \begin{pmatrix} B_{11}^{(\ell)} & B_{12}^{(\ell)} \\ B_{12}^{(\ell)\mathrm{T}} & B_{22}^{(\ell)} \end{pmatrix},$$

where $B_{11}^{(\ell)}$ is $k \times k$, $B_{12}^{(\ell)}$ is $k \times (m_\ell - k)$, and $B_{22}^{(\ell)}$ is $(m_\ell - k) \times (m_\ell - k)$. Note that Lemma 1 is easily extended to the Rayleigh-Ritz procedure as follows.

Lemma 2. *In the restarted Rayleigh-Ritz procedure, $\theta_1^{(\ell)} \geq \cdots \geq \theta_{k_1}^{(\ell)} \geq \vartheta_1^{(\ell)} \geq \cdots \geq \vartheta_{k_2}^{(\ell)}$ are the eigenvalues of $B_{11}^{(\ell+1)}$. Each $\{\theta_i^{(\ell)} : \ell = 1, 2, \ldots\}$ is a nondecreasing convergent sequence for each $i = 1, 2, \ldots, k_1$, and each $\{\vartheta_i^{(\ell)} : \ell = 1, 2, \ldots\}$ is a nonincreasing convergent sequence for each $i = 1, 2, \ldots, k_2$. In other words, $\theta_1^{(\ell)} \geq \cdots \geq \theta_{k_1}^{(\ell)} \geq \vartheta_1^{(\ell)} \geq \cdots \geq \vartheta_{k_2}^{(\ell)}$ are corresponding to the eigenvalues of $B_{11}^{(\ell)}$ as $\ell \to \infty$.*

Proof. From line 6 of the restarted Rayleigh-Ritz procedure, it is easy to see that the eigenvalues of $B_{11}^{(\ell+1)} = V_k^{(\ell+1)\mathrm{T}} A V_k^{(\ell+1)}$ are $\theta_1^{(\ell)} \geq \cdots \geq \theta_k^{(\ell)} \geq \vartheta_1^{(\ell)} \geq \cdots \geq \vartheta_{k_2}^{(\ell)}$. Noting the well-known Courant-Fischer theorem [22] that characterizes the eigenvalues of a symmetric matrix, we see that the eigenvalues of $B^{(\ell+1)}$ satisfy $\theta_i^{(\ell+1)} \geq \theta_i^{(\ell)}$ for $i = 1, 2, \ldots, k_1$ and $\vartheta_i^{(\ell+1)} \leq \vartheta_i^{(\ell)}$ for $i = 1, 2, \ldots, k_2$. Since the Ritz values are confined to an interval, namely, $[\lambda_n, \lambda_1]$, $\{\theta_i^{(\ell)} : \ell = 1, 2, \ldots\}$ is a nondecreasing convergent sequence and $\{\vartheta_i^{(\ell)} : \ell = 1, 2, \ldots\}$ is a nonincreasing convergent sequence. $\square$

Another crucial convergence property is $\lim_{\ell \to \infty} B_{12}^{(\ell)} = O$ as the next lemma.

Lemma 3. *In the restarted Rayleigh-Ritz procedure, write $B^{(\ell)}$ as*

$$B^{(\ell)} = \begin{pmatrix} B_{11}^{(\ell)} & B_{12}^{(\ell)} \\ B_{12}^{(\ell)\mathrm{T}} & B_{22}^{(\ell)} \end{pmatrix},$$

where $B_{11}^{(\ell)}$ is a $k \times k$ symmetric matrix, $B_{12}^{(\ell)}$ is a $k \times (m_\ell - k)$ matrix, and $B_{22}^{(\ell)}$ is an $(m_\ell - k) \times (m_\ell - k)$ symmetric matrix. Then

$$\lim_{\ell \to \infty} B_{12}^{(\ell)} = O$$

holds.

Proof. Let $\widehat{B}_{11}^{(\ell)}$ denote the leading principal $(k+1) \times (k+1)$ submatrix of $B^{(\ell)}$ and $\widehat{\mu}_1^{(n)} \geq \cdots \geq \widehat{\mu}_{k+1}^{(n)}$ denote the eigenvalues of $\widehat{B}_{11}^{(\ell)}$. As the proof of Lemma 2, the well-known Courant-Fischer theorem [22] ensures that

$$\theta_i^{(\ell-1)} \leq \widehat{\mu}_i^{(\ell)} \leq \theta_i^{(\ell)} \quad (i = 1, \ldots, k_1) \tag{4}$$

$$\vartheta_i^{(\ell-1)} \geq \widehat{\mu}_{k_1+1+i}^{(\ell)} \geq \theta_i^{(\ell)} \quad (i = 1, \ldots, k_2), \tag{5}$$

where $\theta_1^{(\ell-1)} \geq \cdots \geq \theta_{k_1}^{(\ell-1)} \geq \vartheta_1^{(\ell-1)} \geq \cdots \geq \vartheta_{k_2}^{(\ell-1)}$ are the eigenvalues of $B_{11}^{(\ell)}$. Here we define $c^{(\ell)}$ as the first column of $B_{12}^{(\ell)}$ and $d^{(\ell)}$ as the upper left element of $B_{22}^{(\ell)}$. In other words, $\widehat{B}_{11}^{(\ell)}$ is divided into

$$\widehat{B}_{11}^{(\ell)} = \left(\begin{array}{c|c} B_{11}^{(\ell)} & c^{(\ell)} \\ \hline c^{(\ell)\mathrm{T}} & d^{(\ell)} \end{array}\right).$$

Firstly, we prove $\lim_{\ell \to \infty} \|c^{(\ell)}\| = 0$ as follows. The square of the Frobenius norm of $\widehat{B}_{11}^{(\ell)}$ is

$$\sum_{i=1}^{k+1} \widehat{\mu}_i^{(\ell)2} = \sum_{i=1}^{k_1} {\theta_i^{(\ell-1)}}^2 + \sum_{i=1}^{k_2} {\vartheta_i^{(\ell-1)}}^2 + {d^{(\ell)}}^2 + 2\|c^{(\ell)}\|^2.$$

It then follows that

$$\begin{aligned} 2\|c^{(\ell)}\|^2 \quad = \quad & \sum_{i=1}^{k_1} (\widehat{\mu}_i^{(\ell)} - \theta_i^{(\ell-1)})(\widehat{\mu}_i^{(\ell)} + \theta_i^{(\ell-1)}) \\ & + \sum_{i=1}^{k_2} (\widehat{\mu}_{k_1+1+i}^{(\ell)} - \vartheta_i^{(\ell-1)})(\widehat{\mu}_i^{(\ell)} + \vartheta_i^{(\ell-1)}) \\ & + (\widehat{\mu}_{k_1+1}^{(\ell)} - d^{(\ell)})(\widehat{\mu}_{k_1+1}^{(\ell)} + d^{(\ell)}). \end{aligned} \tag{6}$$

Noting that the trace of $\widehat{B}_{11}^{(\ell)}$ is

$$\sum_{i=1}^{k_1} \widehat{\mu}_i^{(\ell)} + \widehat{\mu}_{k_1+1}^{(\ell)} + \sum_{i=k_1+2}^{k+1} \widehat{\mu}_i^{(\ell)} = \sum_{i=1}^{k_1} \theta_i^{(\ell-1)} + \sum_{i=1}^{k_2} \vartheta_i^{(\ell-1)} + d^{(\ell)},$$

we see

$$\widehat{\mu}_{k_1+1}^{(\ell)} - d^{(\ell)} = -\sum_{i=1}^{k_1}(\widehat{\mu}_i^{(\ell)} - \theta_i^{(\ell-1)}) - \sum_{i=1}^{k_2}(\widehat{\mu}_{k_1+1+i}^{(\ell)} - \vartheta_i^{(\ell-1)})$$

in the right-hand side of (6). Hence, we have

$$\begin{aligned}
2\|c^{(\ell)}\|^2 &= \sum_{i=1}^{k_1}(\widehat{\mu}_i^{(\ell)} - \theta_i^{(\ell-1)})(\widehat{\mu}_i^{(\ell)} + \theta_i^{(\ell-1)} - \widehat{\mu}_{k_1+1}^{(\ell)} - d^{(\ell)}) \\
&\quad + \sum_{i=1}^{k_2}(\widehat{\mu}_{k_1+1+i}^{(\ell)} - \vartheta_i^{(\ell-1)})(\widehat{\mu}_{k_1+1+i}^{(\ell)} + \vartheta_i^{(\ell-1)} - \widehat{\mu}_{k_1+1}^{(\ell)} - d^{(\ell)}) \\
&\leq 4\|A\|(\sum_{i=1}^{k_1}|\widehat{\mu}_i^{(\ell)} - \theta_i^{(\ell-1)}| + \sum_{i=1}^{k_2}|\widehat{\mu}_{k_1+1+i}^{(\ell)} - \vartheta_i^{(\ell-1)}|), \\
&\leq 4\|A\|(\sum_{i=1}^{k_1}|\theta_i^{(\ell)} - \theta_i^{(\ell-1)}| + \sum_{i=1}^{k_2}|\vartheta_i^{(\ell)} - \vartheta_i^{(\ell-1)}|),
\end{aligned}$$

where the first inequality is due to $|v^{\mathrm{T}}Av| \leq \|A\|$ for any unit vector v, the second inequality is due to (4) and (5). Therefore, by Lemma 2, we obtain $\lim_{\ell\to\infty}\|c^{(\ell)}\| = 0$, where $c^{(\ell)}$ is the first column of $B_{12}^{(\ell)}$. By permuting the column of $B_{12}^{(\ell)}$, we can see $\lim_{\ell\to\infty} B_{12}^{(\ell)} = O$. □

Finally, we review a closely related convergence analysis by Crouzeix, Philippe and Sadkane [5]. In fact, for the restart strategy for computing the k largest eigenvalues, namely $k = k_1$, we can see Lemmas 2 and 3 by [5, §2]. In this section, we have proven Lemmas 2 and 3 for the general $k_1,\ k_2$ along with [5, §2]. In addition, we should note [5, Theorem 2.1], which shows that under some assumptions the k largest Ritz values always converge to eigenvalues of A as follows.

Theorem 2 ([5])**.** *In the restarted Rayleigh-Ritz procedure for the k largest eigenvalues, each $\{\theta_i^{(\ell)} : \ell = 1, 2, \ldots\}$ is a nondecreasing convergent sequence for each $i = 1, 2, \ldots, k$. Moreover, let a set of matrices $\{C_{i,j}^{(\ell)}\}$ satisfy the following assumption: there exist $K_1,\ K_2 > 0$ such that $K_1\|v\| \leq v^{\mathrm{T}}C_{i,j}^{(\ell)}v \leq K_2\|v\|$ for any vector $v \in span\{V_{jk}^{(\ell)}\}^{\perp}$, where $V_{jk}^{(\ell)} = [v_1^{(\ell)}, v_2^{(\ell)}, \ldots, v_{jk}^{(\ell)}]$ with $j \geq 1$. Suppose that the vector $(I - V_{jk}^{(\ell)}{V_{jk}^{(\ell)}}^{\mathrm{T}})C_{i,j}^{(\ell)}(A - \widehat{\theta}_{i,j}^{(\ell)}I)\widehat{u}_{i,j}^{(\ell)}$ belongs to span$\{V_{(j+1)k}^{(\ell)}\}$, where $\widehat{\theta}_{1,j}^{(\ell)} \geq \cdots \geq \widehat{\theta}_{k,j}^{(\ell)}$ are the k largest Ritz values of ${V_{jk}^{(\ell)}}^{\mathrm{T}}AV_{jk}^{(\ell)}$, and $\widehat{u}_{1,j}^{(\ell)}, \ldots, \widehat{u}_{k,j}^{(\ell)}$ are the corresponding Ritz vectors. Then, as $\ell \to \infty$, $\widehat{\theta}_{i,j}^{(\ell)}$ $(i = 1, \ldots, k)$ converge to eigenvalues of A and $\widehat{u}_{i,j}^{(\ell)}$ $(i = 1, \ldots, k)$ converge to the corresponding eigenvectors.* ■

It is easy to see that the block Lanczos method with k starting vectors $v_1^{(0)}, \ldots, v_k^{(0)}$ corresponds to the situation where $C_{i,j}^{(\ell)}$ is the identity matrix, and hence the k largest Ritz values converge to some eigenvalues of A [5, Corollary 2.2]. In contrast, we would like to prove that the Ritz values converge to the k largest eigenvalues for the restarted Lanczos method with one starting vector $v_1^{(0)}$ in the next section. This proof is not readily accessible from Theorem 2. Moreover, in order to extend the proof to the restart strategy to compute both the largest and smallest eigenvalues, we use Lemmas 2 and 3 in the following sections.

Regarding the Jacobi-Davidson method with one restarting vector $v_1^{(\ell)}$ ($\ell = 0, 1, \ldots$), the situation corresponds to $C_{i,j}^{(\ell)} = (A - \widehat{\theta}_{i,j}^{(\ell)} I)^{-2}$, $i = 1$ (see Section 7, [12, §2] and [24]). Therefore, we can prove a global convergence property based on Theorem 2. This is a proof by contradiction. Noting that the Ritz value $\theta_{1,j}^{(\ell)}$ is convergent, we assume $\theta_{1,j}^{(\infty)}$ is not an eigenvalue of A. Let $d_{\max}$ be $\max_{1 \leq i \leq N} |\theta_{1,j}^{(\infty)} - \lambda_i|$ and $d_{\min}$ be $\min_{1 \leq i \leq N} |\theta_{1,j}^{(\infty)} - \lambda_i|$. Then we have ${d_{\max}}^{-2} \leq v^{\mathrm{T}}(A - \widehat{\theta}_{1,j}^{(\infty)} I)^{-2} v \leq {d_{\min}}^{-2}$ for any unit vector v. In other words, for all sufficiently large ℓ, $C_{1,j}^{(\ell)} = (A - \widehat{\theta}_{1,j}^{(\ell)} I)^{-2}$ satisfy the conditions in Theorem 2. Hence, the Ritz value $\theta_{1,j}^{(\ell)}$ should converge to an exact eigenvalue of A. Thus Theorem 2 ensures this global convergence property of the restarted Jacobi-Davidson. In contrast, we provide a direct convergence proof for multiple restarting vectors $v_1^{(\ell)}, \ldots, v_{\widehat{m}}^{(\ell)}$ ($\ell = 0, 1, \ldots$) in Section 7. To this end, we use Lemmas 2 and 3 in the following sections.

5 Global convergence of the restarted Lanczos method

In this section, we prove global convergence of the restarted Lanczos method. More specifically, (2), (3) are guaranteed for all sufficiently large ℓ in the following theorem.

Theorem 3. *In Algorithm 1, suppose the initial starting vector satisfies* $x_j^{\mathrm{T}} v_1^{(0)} \neq 0$ $(j = 1, \ldots, k)$. *Then we have*

$$\lim_{\ell \to \infty} T_k^{(\ell)} = \mathrm{diag}(\lambda_1, \ldots, \lambda_k), \tag{7}$$

$$\lim_{\ell \to \infty} V_k^{(\ell)} = X_k. \tag{8}$$

Proof. In view of Lemma 3, the $(k+1, k)$ element of $T_{m_\ell}^{(\ell)}$ converges to 0. Therefore, $\|Au_j^{(\ell)} - \theta_j^{(\ell)} u_j^{(\ell)}\| \to 0$ $(j = 1, \ldots, k)$ as $\ell \to \infty$ [9, Theorem 7.2], where $\theta_j^{(\ell)}$ $(j = 1, \ldots, k)$ are the Ritz values and $u_j^{(\ell)}$ $(j = 1, \ldots, k)$ are the corresponding Ritz vectors. From Lemma 2, the Ritz values $\theta_j^{(\ell)}$ $(j = 1, \ldots, k)$ converge to some eigenvalues, and hence the Ritz vectors $u_j^{(\ell)}$ $(j = 1, \ldots, k)$ also converge to the corresponding eigenvectors.

Next, we prove $\theta_j^{(\infty)} = \lambda_j$, $u_j^{(\infty)} = x_j$ $(j = 1, \ldots, k)$. Importantly, the starting vector $v_1^{(\ell)}$ for all $\ell = 0, 1, \ldots$ satisfy

$$w_1^{(\ell+1)} = (A - \mu_1^{(\ell)} I) \cdots (A - \mu_{m_\ell - k}^{(\ell)} I) v_1^{(\ell)}, \tag{9}$$

$$v_1^{(\ell+1)} = w_1^{(\ell+1)} / \|w_1^{(\ell+1)}\|, \tag{10}$$

where $\mu_1^{(\ell)}, \ldots, \mu_{m_\ell - k}^{(\ell)}$ are unwanted eigenvalues of $T_{m_\ell}^{(\ell)}$ in line 4 of Algorithm 1. In other words, the starting vector $v_1^{(\ell+1)}$ is given by multiplying a filter polynomial to $v_1^{(\ell)}$. See [25] for details. The starting vector $v_1^{(0)}$ is expressed by the eigenvectors as

$$v_1^{(0)} = c_1 x_1 + c_2 x_2 + \cdots + c_n x_n, \tag{11}$$

and hence

$$\widehat{w}_1^{(\ell)} = \sum_{j=1}^{n} c_j \prod_{\widehat{\ell}=0}^{\ell-1} \prod_{i=1}^{m_\ell - k} (\lambda_j - \mu_i^{(\widehat{\ell})}) x_j \tag{12}$$

$$v_1^{(\ell)} = \widehat{w}_1^{(\ell)} / \|\widehat{w}_1^{(\ell)}\| \tag{13}$$

by the equations (9), (10). Here we define

$$c_j^{(\ell)} = c_j \prod_{\widehat{\ell}=0}^{\ell-1} \prod_{i=1}^{m_\ell - k} (\lambda_j - \mu_i^{(\widehat{\ell})}) \quad (j = 1, \ldots, n). \tag{14}$$

We have already proven that $\theta_j^{(\infty)}$ $(j = 1, \ldots, k)$ are eigenvalues of A and $u_j^{(\infty)}$ $(j = 1, \ldots, k)$ are the corresponding eigenvectors. Suppose that $\theta_k^{(\infty)} = \lambda_{\widehat{k}},\ u_k^{(\infty)} = x_{\widehat{k}}$ $(\widehat{k} \geq k)$. We show $\widehat{k} = k$ as follows. Since $v_1^{(\infty)} \in \operatorname{span}\{u_1^{(\infty)}, \ldots, u_k^{(\infty)}\}$ and $u_i^{(\infty)} \neq x_j$ $(i = 1, \ldots, k,\ j = \widehat{k} + 1, \ldots, n)$, we see

$$\lim_{\ell \to \infty} c_j^{(\ell)} / c_{\widehat{k}}^{(\ell)} = 0 \quad (j = \widehat{k} + 1, \ldots, n). \tag{15}$$

Furthermore, with the aid of line 4 of Algorithm 1,

$$\lambda_{\widehat{k}} > \theta_k^{(\ell)} > \mu_j^{(\ell)} \quad (j = 1, \ldots, m_\ell - k) \tag{16}$$

for all ℓ. Therefore, the expansion coefficients $c_1^{(\ell)}, \ldots, c_{\widehat{k}}^{(\ell)}$ satisfy

$$\lim_{\ell \to \infty} c_{j+1}^{(\ell)} / c_j^{(\ell)} = 0 \quad (j = 1, \ldots, \widehat{k} - 1) \tag{17}$$

from (14), (16). Thus we see

$$\lim_{\ell \to \infty} c_i^{(\ell)} / c_j^{(\ell)} = 0 \quad (i = k + 1, \ldots, n,\ j = 1, \ldots, k). \tag{18}$$

It then follows that the Ritz vectors $u_j^{(\ell)}$ $(j = 1, \ldots, k)$ converge to the eigenvectors x_j $(j = 1, \ldots, k)$ corresponding to $c_j^{(\ell)}$ $(j = 1, \ldots, k)$, namely $\widehat{k} = k$. Therefore, we obtain

$$\lim_{\ell \to \infty} \theta_j^{(\ell)} = \lambda_j,\ \lim_{\ell \to \infty} u_j^{(\ell)} = x_j \quad (j = 1, \ldots, k). \tag{19}$$

The final task is to show (7) and (8). From (17) with $\widehat{k} = k$, we have

$$\lim_{\ell \to \infty} c_{j+1}^{(\ell)} / c_j^{(\ell)} = 0 \quad (j = 1, \ldots, k - 1). \tag{20}$$

This implies $v_j^{(\infty)} = x_j$ $(j = 1, \ldots, k)$. Therefore, we obtain (7), (8). $\square$

Note that the proof above is similar to Sorensen's proof of Theorem 1. See [25] for details. However, in Theorem 3, the assumption about the off-diagonal elements of $T_k^{(\ell)}$ is not needed. In other words, Theorem 3 is regarded as an extension of Theorem 1.

Recall Section 4, which is devoted to showing the convergence properties of the Rayleigh-Ritz procedure. Lemmas 2 and 3 are applied to the shift strategy to compute both the k_1 largest eigenvalues and the k_2 smallest eigenvalues. Hence, let us consider the convergence behavior of the following restarted Lanczos method.

Algorithm 2 The restarted Lanczos method for the k_1 largest eigenvalues and the k_2 smallest eigenvalues

---

Initialization pick a unit vector $v_1^{(0)}$

1: compute the orthonormal basis $V_{m_0}^{(0)} = [v_1^{(0)}, \dots, v_{m_0}^{(0)}]$ of $\mathcal{K}^{m_0}(A, v_1^{(0)})$ by the Lanczos process

2: **for** $\ell := 0, 1, \dots,$ **do**

3: compute $T_{m_\ell}^{(\ell)} = {V_{m_\ell}^{(\ell)}}^{\mathrm{T}} A V_{m_\ell}^{(\ell)}$

4: compute eigenvalues of $T_{m_\ell}^{(\ell)}$: $\theta_1^{(\ell)} \geq \dots \geq \theta_{k_1}^{(\ell)} \geq \mu_1^{(\ell)} \cdots \geq \mu_{m_\ell - k}^{(\ell)} \geq \vartheta_1^{(\ell)} \geq \dots \geq \vartheta_{k_2}^{(\ell)}$ $(k = k_1 + k_2)$

5: compute the QR decomposition: $\widehat{Q}^{(\ell)} \widehat{R}^{(\ell)} = (T_{m_\ell}^{(\ell)} - \mu_1^{(\ell)} I) \cdots (T_{m_\ell}^{(\ell)} - \mu_{m_\ell - k}^{(\ell)} I)$

6: compute $V_k^{(\ell+1)} = [v_1^{(\ell)}, \dots, v_{m_\ell}^{(\ell)}] \widehat{Q}_k^{(\ell)}$, where $\widehat{Q}_k^{(\ell)}$ denotes the matrix of the first k columns of $\widehat{Q}^{(\ell)}$ (Then $V_k^{(\ell+1)} = [u_1^{(\ell)}, \dots, u_k^{(\ell)}] Q^{(\ell)}$, where $u_1^{(\ell)}, \dots, u_k^{(\ell)}$ are the Ritz vectors corresponding to $\theta_1^{(\ell)} \geq \dots \geq \theta_{k_1}^{(\ell)} \geq \vartheta_1^{(\ell)} \geq \dots \geq \vartheta_{k_2}^{(\ell)}$ and $Q^{(\ell)}$ is some orthogonal matrix)

7: compute $v_{k+1}^{(\ell+1)}, \dots, v_{m_{\ell+1}}^{(\ell+1)}$ by the Lanczos process to obtain a new orthonormal basis $V_{m_{\ell+1}}^{(\ell+1)} = [v_1^{(\ell+1)}, \dots, v_{m_{\ell+1}}^{(\ell+1)}]$

8: **end for**

Similarly to Theorem 3, we prove global convergence of Algorithm 2, which states the convergence of the Ritz values and the corresponding Ritz vectors. The next theorem is regarded as another extension of Theorem 1.

Theorem 4. *In Algorithm 2, suppose the initial starting vector satisfies* $x_j^{\mathrm{T}} v_1^{(0)} \neq 0$ $(j = 1, \dots, k_1)$ *and* $x_{n-j+1}^{\mathrm{T}} v_1^{(0)} \neq 0$ $(j = 1, \dots, k_2)$*. Then we have* $\lim_{\ell \to \infty} \theta_j^{(\ell)} = \lambda_j$, $\lim_{\ell \to \infty} \vartheta_j^{(\ell)} = \lambda_{n-j+1}$*, the Ritz vectors* $u_1^{(\ell)}, \dots, u_k^{(\ell)}$ *converge to the corresponding eigenvectors.*

Proof. Similarly to the proof of Theorem 3, it is easy to see that the Ritz values $\theta_j^{(\ell)}$ $(j = 1, \dots, k_1)$ and $\vartheta_j^{(\ell)}$ $(j = 1, \dots, k_2)$ converge to eigenvalues of A from Lemmas 2 and 3. We prove $\theta_j^{(\infty)} = \lambda_j$ $(j = 1, \dots, k_1)$ and $\vartheta_j^{(\infty)} = \lambda_{n-j+1}$ $(j = 1, \dots, k_2)$ as follows.

The following discussion is almost the same as the proof of Theorem 3. Firstly, we have

$$w_1^{(\ell+1)} = (A - \mu_1^{(\ell)} I) \cdots (A - \mu_{m_\ell - k}^{(\ell)} I) v_1^{(\ell)}, \tag{21}$$

$$v_1^{(\ell+1)} = w_1^{(\ell+1)} / \| w_1^{(\ell+1)} \|, \tag{22}$$

from [25]. The initial starting vector $v_1^{(0)}$ is expressed by the eigenvectors as

$$v_1^{(0)} = c_1 x_1 + c_2 x_2 + \cdots + c_n x_n, \tag{23}$$

and hence

$$\widehat{w}_1^{(\ell)} = \sum_{j=1}^{n} c_j \prod_{\widehat{\ell}=0}^{\ell-1} \prod_{i=1}^{m_\ell - k} (\lambda_j - \mu_i^{(\widehat{\ell})}) x_j \tag{24}$$

$$v_1^{(\ell)} = \widehat{w}_1^{(\ell)} / \|\widehat{w}_1^{(\ell)}\| \tag{25}$$

by the equations (21), (22). Here we define

$$c_j^{(\ell)} = c_j \prod_{\widehat{\ell}=0}^{\ell-1} \prod_{i=1}^{m_\ell - k} (\lambda_j - \mu_i^{(\widehat{\ell})}) \quad (j = 1, \ldots, n). \tag{26}$$

Suppose that $\theta_{k_1}^{(\infty)} = \lambda_{\widehat{k}_1}$, $u_{k_1}^{(\infty)} = x_{\widehat{k}_1}$ $(\widehat{k}_1 \geq k_1)$ and $\vartheta_1^{(\infty)} = \lambda_{n-\widehat{k}_2+1}$, $u_{\widehat{k}_1+1}^{(\infty)} = x_{n-\widehat{k}_2+1}$ $(\widehat{k}_2 \geq k_2)$. Since $v_1^{(\infty)} \in \operatorname{span}\{u_1^{(\infty)}, \ldots, u_k^{(\infty)}\}$ and $u_i^{(\infty)} \neq x_j$ $(i = 1, \ldots, k,\ j = \widehat{k}_1 + 1, \ldots, n - \widehat{k}_2)$, we see

$$\lim_{\ell\to\infty} c_j^{(\ell)} / c_{\widehat{k}_1}^{(\ell)} = 0,\ \lim_{\ell\to\infty} c_j^{(\ell)} / c_{n-\widehat{k}_2+1}^{(\ell)} = 0 \quad (j = \widehat{k}_1 + 1, \ldots, n - \widehat{k}_2). \tag{27}$$

Furthermore, with the aid of line 4 of Algorithm 2,

$$\lambda_{\widehat{k}_1} > \theta_{k_1}^{(\ell)} > \mu_j^{(\ell)} > \vartheta_1^{(\ell)} > \lambda_{n-\widehat{k}_2+1} \quad (j = 1, \ldots, m_\ell - k) \tag{28}$$

for all ℓ. Therefore, the expansion coefficients satisfy

$$\lim_{\ell\to\infty} c_{j+1}^{(\ell)} / c_j^{(\ell)} = 0 \quad (j = 1, \ldots, \widehat{k}_1 - 1) \tag{29}$$

$$\lim_{\ell\to\infty} c_j^{(\ell)} / c_{j+1}^{(\ell)} = 0 \quad (j = n - \widehat{k}_2 + 1, \ldots, n) \tag{30}$$

from (26), (28). Thus we see

$$\lim_{\ell\to\infty} c_i^{(\ell)} / c_j^{(\ell)} = 0 \quad (i = k_1 + 1, \ldots, n - k_2,\ j = 1, \ldots, k_1, n - k_2 + 1, \ldots, n).$$

It then follows that the Ritz vectors $u_j^{(\ell)}$ $(j = 1, \ldots, k)$ converge to the eigenvectors x_j $(j = 1, \ldots, k_1, n - k_2 + 1, \ldots, n)$ corresponding to $c_j^{(\ell)}$ $(j = 1, \ldots, k_1, n - k_2 + 1, \ldots, n)$, namely $\widehat{k}_1 = k_1$, $\widehat{k}_2 = k_2$. Therefore, the theorem is established. $\square$

6 Global convergence of the restarted block Lanczos method

From the practical point of view, there are situations where the use of a block of k starting vectors, instead of a single starting vector, is preferable. One such case is that of matrices with multiple or closely clustered eigenvalues. To obtain basis vectors for the eigenspace corresponding to such a cluster of k eigenvalues, block Krylov subspaces induced by A and a block of k starting vectors need to be used.

For an $N \times k$ matrix $V = [v_1, \ldots, v_k]$ with orthonormal columns, the block Krylov subspace is defined as

$$\mathcal{K}^m(A, V) = \mathrm{span}\{v_1, \ldots, v_k, Av_1, \ldots, Av_k, \ldots, A^{m-1}v_1, \ldots, A^{m-1}v_k\}. \tag{31}$$

The block Lanczos method is the Rayleigh-Ritz procedure based on the block Krylov subspace, which is mathematically equivalent to the band Lanczos method. Similarly to the Krylov subspace, the orthonormal basis $\{v_i\}_{i=1,\ldots,km}$ of $\mathcal{K}^m(A, V)$ is computed. Let $V_{km} = [v_1, \ldots, v_{km}]$. Then $T_{km} = {V_{km}}^{\mathrm{T}} A V_{km}$ is a $km \times km$ symmetric block tridiagonal matrix of the form

$$T_{km} = \begin{pmatrix} T_{1,1} & T_{1,2} & & \\ {T_{1,2}}^{\mathrm{T}} & T_{2,2} & \ddots & \\ & \ddots & \ddots & T_{m-1,m} \\ & & {T_{m-1,m}}^{\mathrm{T}} & T_{m,m} \end{pmatrix}, \tag{32}$$

with $k \times k$ symmetric diagonal blocks $T_{1,1}, \ldots, T_{m,m}$.

There exist a number of the restart strategy depending on the desired eigenvalues. The most typical restart block Lanczos method for the k_1 largest eigenvalues and the k_2 smallest eigenvalues is described as follows [6].

Algorithm 3 The restarted block Lanczos method for the k_1 largest eigenvalues and the k_2 smallest eigenvalues [6]

Initialization pick an $N \times k$ matrix $V_1^{(0)}$ with orthonormal columns

1: compute the orthonormal basis $V_{km_0}^{(0)}$ of $\mathcal{K}^{m_0}(A, V_1^{(0)})$
2: **for** $\ell := 0, 1, \ldots,$ **do**
3: compute $T_{km_\ell}^{(\ell)} = {V_{km_\ell}^{(\ell)}}^{\mathrm{T}} A V_{km_\ell}^{(\ell)}$
4: compute the k_1 largest eigenvalues and the k_2 smallest eigenvalues of $T_{km}^{(\ell)}$: $\theta_1^{(\ell)} \geq \cdots \geq \theta_k^{(\ell)}$ $(k = k_1 + k_2)$
5: compute the Ritz vectors $u_1^{(\ell)}, \ldots, u_k^{(\ell)}$ associated with $\theta_1^{(\ell)} \geq \cdots \geq \theta_k^{(\ell)}$
6: $V_1^{(\ell+1)} := [u_1^{(\ell)}, \ldots, u_k^{(\ell)}]$
7: compute the orthonormal basis $V_{km_{\ell+1}}^{(\ell+1)}$ of $\mathcal{K}^{m_{\ell+1}}(A, V_1^{(\ell+1)})$
8: **end for**

Here we show a global convergence theorem of Algorithm 3.

Theorem 5. *In Algorithm 3, the Ritz values $\theta_1^{(\ell)},\dots,\theta_k^{(\ell)}$ converge to eigenvalues of A and the Ritz vectors $u_1^{(\ell)},\dots,u_k^{(\ell)}$ converge to the corresponding eigenvectors.*

Proof. From (32), $T_{km_\ell}^{(\ell)}$ is a $km_\ell \times km_\ell$ symmetric block tridiagonal matrix of the form

$$T_{km_\ell}^{(\ell)} = \begin{pmatrix} T_{1,1}^{(\ell)} & T_{1,2}^{(\ell)} & & \\ {T_{1,2}^{(\ell)}}^{\mathrm{T}} & T_{2,2}^{(\ell)} & \ddots & \\ & \ddots & \ddots & T_{m_\ell-1,m_\ell}^{(\ell)} \\ & & {T_{m_\ell-1,m_\ell}^{(\ell)}}^{\mathrm{T}} & T_{m_\ell,m_\ell}^{(\ell)} \end{pmatrix},$$

with $k\times k$ symmetric diagonal blocks $T_{1,1}^{(\ell)},\dots,T_{m_\ell,m_\ell}^{(\ell)}$. From Lemma 3, we have $\lim_{\ell\to\infty} T_{1,2}^{(\ell)} = O$. By noting that the Ritz values $\theta_1^{(\ell)},\dots,\theta_k^{(\ell)}$ are the eigenvalues of $T_{1,1}^{(\ell+1)}$ from Lemma 2, we obtain the theorem. □

Although Theorem 5 states a global convergence, it is not shown that $\theta_1^{(\ell)},\dots,\theta_k^{(\ell)}$ converge to the k_1 largest eigenvalues and the k_2 smallest eigenvalues of A. This property is differ from Theorem 4.

7 Global convergence of the restarted Jacobi-Davidson method

In this section, we show global convergence of the restarted Jacobi-Davidson method proposed by Sleijpen and van der Vorst [3, 24].

Firstly, we briefly summarize the Jacobi-Davidson method. See [24] for details. The Jacobi-Davidson method is also based on the Rayleigh-Ritz procedure. For the orthonormal basis $V_j = [v_1,\dots,v_j]$, the new basis v_{j+1} is obtained as follows.

The strategy for computing the largest eigenvalue is shown here. Let $\widehat{\theta}_{1,j}$ denote the largest Ritz value, namely, the largest eigenvalue of $V_j^{\mathrm{T}}AV_j$, and u_j denote the corresponding Ritz vector. Moreover, let $r_j = Au_j - \widehat{\theta}_{1,j}u_j$. Then we find t_{j+1} that satisfies the equations

$$(I - u_ju_j^{\mathrm{T}})(A - \widehat{\theta}_{1,j}I)(I - u_ju_j^{\mathrm{T}})t_{j+1} = -r_j,\ u_j^{\mathrm{T}}t_{j+1} = 0. \tag{33}$$

The new basis vector v_{j+1} is obtained by orthogonalizing t_{j+1} to V_j. In other words, v_{j+1} is given by

$$\begin{aligned} w_{j+1} &= t_{j+1} - V_j(V_j^{\mathrm{T}}t_{j+1}) \\ v_{j+1} &= w_{j+1}/\|w_{j+1}\|. \end{aligned}$$

In this section, we focus on the fact that the process above is regarded as one iteration of the Rayleigh quotient iterations method as is shown below (see also [12, §2] and [24]). If (33) is solved exactly, we have

$$(A - \widehat{\theta}_{1,j}I)t_{j+1} = -r_j + cu_j \tag{34}$$

for a constant c such that $u_j^{\mathrm{T}} t_{j+1} = 0$. It follows that

$$t_{j+1} = -u_j + c(A - \widehat{\theta}_{1,j} I)^{-1} u_j \tag{35}$$

in view of $r_j = Au_j - \widehat{\theta}_{1,j} u_j$. From the orthogonality $u_j^{\mathrm{T}} t_{j+1} = 0$, we see $c = (u_j^{\mathrm{T}}(A - \widehat{\theta}_{1,j} I)^{-1} u_j)^{-1}$. Since the Ritz vector u_j satisfies $u_j \in \mathrm{span}\{v_1, \ldots, v_j\}$, v_{j+1} is given by orthogonalizing $(A - \widehat{\theta}_{1,j} I)^{-1} u_j$ to V_j. Note that $(A - \widehat{\theta}_{1,j} I)^{-1} u_j$ is given by the Rayleigh quotient iteration for u_j. Thus we see that v_{j+1} satisfies

$$\widehat{w}_{j+1} = (A - \widehat{\theta}_{1,j} I)^{-1} u_j - V_j(V_j^{\mathrm{T}}(A - \widehat{\theta}_{1,j} I)^{-1} u_j) \tag{36}$$

$$v_{j+1} = \widehat{w}_{j+1} / \|\widehat{w}_{j+1}\|. \tag{37}$$

As the discussion before, the restart strategy is incorporated to the Jacobi-Davidson method. The algorithm reads as follows.

Algorithm 4 The restarted Jacobi-Davidson method for the largest eigenvalue [24]

Initialization pick an $N \times \widehat{m}$ matrix $V_{\widehat{m}}^{(0)} := [v_1^{(0)}, \ldots, v_{\widehat{m}}^{(0)}]$ with orthonormal columns

for $\ell := 0, 1, \ldots,$ **do**
  for $j := \widehat{m}, \ldots, m_\ell - 1$ **do**
    compute the largest eigenvalue $\widehat{\theta}_{1,j}^{(\ell)}$ of ${V_j^{(0)}}^{\mathrm{T}} A V_j^{(0)}$
    compute the corresponding Ritz vector $u_j^{(\ell)}$
    compute the residual vector $r_j^{(\ell)} := Au_j^{(\ell)} - \widehat{\theta}_{1,j}^{(\ell)} u_j^{(\ell)}$
    compute a new vector $t_{j+1}^{(\ell)}$ such that $(I - u_j^{(\ell)} {u_j^{(\ell)}}^{\mathrm{T}})(A - \widehat{\theta}_{1,j}^{(\ell)} I)(I - u_j^{(\ell)} {u_j^{(\ell)}}^{\mathrm{T}}) t_{j+1}^{(\ell)} = -r_j^{(\ell)}$, ${u_j^{(\ell)}}^{\mathrm{T}} t_{j+1}^{(\ell)} = 0$
    compute $w_{j+1}^{(\ell)} := t_{j+1}^{(\ell)} - V_j^{(\ell)}({V_j^{(\ell)}}^{\mathrm{T}} t_{j+1}^{(\ell)})$
    compute $v_{j+1}^{(\ell)} := w_{j+1}^{(\ell)} / \|w_{j+1}^{(\ell)}\|$
    $V_{j+1}^{(\ell)} := [v_1^{(\ell)}, \ldots, v_{j+1}^{(\ell)}]$
  end for
  compute $B_{m_\ell}^{(\ell)} := {V_{m_\ell}^{(\ell)}}^{\mathrm{T}} A V_{m_\ell}^{(\ell)}$
  compute the Ritz vectors $u_1^{(\ell)}, \ldots, u_{\widehat{m}}^{(\ell)}$ corresponding to the $\widehat{m}$ largest eigenvalues of $B_{m_\ell}^{(\ell)}$
  $V_{\widehat{m}}^{(\ell+1)} := [u_1^{(\ell)}, \ldots, u_{\widehat{m}}^{(\ell)}]$
end for

It is empirically known that, in Algorithm 4, the largest Ritz value $\theta_1^{(\ell)}$ for ${V_{m_\ell}^{(\ell)}}^{\mathrm{T}} A V_{m_\ell}^{(\ell)}$ usually converges to the largest eigenvalue. In this section, we prove that the Ritz value converges to eigenvalue, although not necessarily to the largest one.

Theorem 6. *In Algorithm 4, if breakdown does not occur, the largest Ritz value $\theta_1^{(\ell)}$ converges to an eigenvalue of A. When the Ritz value $\theta_1^{(\ell)}$ converges to the largest eigenvalue λ_1, the Ritz vector $u_1^{(\ell)}$ converges to the corresponding eigenvector x_1.*

Proof. Similarly to Lemma 3, write $B_{m_\ell}^{(\ell)}$ as

$$B_{m_\ell}^{(\ell)} = \begin{pmatrix} B_{m_\ell,11}^{(\ell)} & B_{m_\ell,12}^{(\ell)} \\ {B_{m_\ell,12}^{(\ell)}}^{\mathrm{T}} & B_{m_\ell,22}^{(\ell)} \end{pmatrix},$$

where $B_{m_\ell,11}^{(\ell)}$ is $\widehat{m} \times \widehat{m}$, $B_{m_\ell,12}^{(\ell)}$ is $\widehat{m} \times (m_\ell - \widehat{m})$, and $B_{m_\ell,22}^{(\ell)}$ is $(m_\ell - \widehat{m}) \times (m_\ell - \widehat{m})$. Then $B_{m_\ell,11}^{(\ell)} = \mathrm{diag}(\theta_1^{(\ell)}, \dots, \theta_{\widehat{m}}^{(\ell)})$ in view of line 13 of Algorithm 4. Since $\lim_{\ell\to\infty} B_{m_\ell,12}^{(\ell)} = O$ from Lemma 3, $\lim_{\ell\to\infty} {v_1^{(\ell)}}^{\mathrm{T}} A v_{\widehat{m}+1}^{(\ell)} = 0$ holds, where ${v_1^{(\ell)}}^{\mathrm{T}} A v_{\widehat{m}+1}^{(\ell)}$ is the $(1, \widehat{m}+1)$ element of $B_{m_\ell,12}^{(\ell)}$. Similarly to (36) and (37), let

$$\widehat{w}_{\widehat{m}+1}^{(\ell)} = (A - \widehat{\theta}_{1,\widehat{m}}^{(\ell)} I)^{-1} v_1^{(\ell)} - V_{\widehat{m}}^{(\ell)} ({V_{\widehat{m}}^{(\ell)}}^{\mathrm{T}} (A - \widehat{\theta}_{1,\widehat{m}}^{(\ell)} I)^{-1} v_1^{(\ell)}). \tag{38}$$

Then we have

$$v_{\widehat{m}+1}^{(\ell)} = \widehat{w}_{\widehat{m}+1}^{(\ell)} / \|\widehat{w}_{\widehat{m}+1}^{(\ell)}\| \tag{39}$$

by (36), (37). Thus we see

$$\lim_{\ell\to\infty} {v_1^{(\ell)}}^{\mathrm{T}} A \widehat{w}_{\widehat{m}+1}^{(\ell)} / \|\widehat{w}_{\widehat{m}+1}^{(\ell)}\| = 0. \tag{40}$$

Firstly, we prove ${v_1^{(\ell)}}^{\mathrm{T}} A \widehat{w}_{\widehat{m}+1}^{(\ell)} = 1$ as follows. For the equation (38), we see

$${v_1^{(\ell)}}^{\mathrm{T}} A V_{\widehat{m}}^{(\ell)} ({V_{\widehat{m}}^{(\ell)}}^{\mathrm{T}} (A - \widehat{\theta}_{1,\widehat{m}}^{(\ell)} I)^{-1} v_1^{(\ell+1)}) = \widehat{\theta}_{1,\widehat{m}}^{(\ell)} {v_1^{(\ell)}}^{\mathrm{T}} (A - \widehat{\theta}_{1,\widehat{m}}^{(\ell)} I)^{-1} v_1^{(\ell)}$$

in view of $V_{\widehat{m}}^{(\ell)} = [u_1^{(\ell-1)}, \dots, u_{\widehat{m}}^{(\ell-1)}]$, where $u_1^{(\ell-1)}, \dots, u_{\widehat{m}}^{(\ell-1)}$ are the Ritz vectors, which satisfy ${u_i^{(\ell-1)}}^{\mathrm{T}} A u_j^{(\ell-1)} = 0$ for $i \neq j$. It follows that

$$\begin{aligned} {v_1^{(\ell)}}^{\mathrm{T}} A \widehat{w}_{\widehat{m}+1}^{(\ell)} &= {v_1^{(\ell)}}^{\mathrm{T}} A (A - \widehat{\theta}_{1,\widehat{m}}^{(\ell)} I)^{-1} v_1^{(\ell)} - \widehat{\theta}_{1,\widehat{m}}^{(\ell)} {v_1^{(\ell)}}^{\mathrm{T}} (A - \widehat{\theta}_{1,\widehat{m}}^{(\ell)} I)^{-1} v_1^{(\ell)} \\ &= {v_1^{(\ell)}}^{\mathrm{T}} (A - \widehat{\theta}_{1,\widehat{m}}^{(\ell)} I)(A - \widehat{\theta}_{1,\widehat{m}}^{(\ell)} I)^{-1} v_1^{(\ell)} \\ &= 1. \end{aligned} \tag{41}$$

Thus we have

$$\lim_{\ell\to\infty} {v_1^{(\ell)}}^{\mathrm{T}} A v_{\widehat{m}+1}^{(\ell)} = \lim_{\ell\to\infty} {v_1^{(\ell)}}^{\mathrm{T}} A \widehat{w}_{\widehat{m}+1}^{(\ell)} / \|\widehat{w}_{\widehat{m}+1}^{(\ell)}\| = \lim_{\ell\to\infty} 1 / \|\widehat{w}_{\widehat{m}+1}^{(\ell)}\| = 0. \tag{42}$$

Noting

$$\|\widehat{w}_{\widehat{m}+1}^{(\ell)}\| \leq \|(A - \widehat{\theta}_{1,\widehat{m}}^{(\ell)} I)^{-1} v_1^{(\ell)}\|$$

from (38), we have

$$\lim_{\ell\to\infty} 1 / \|(A - \widehat{\theta}_{1,\widehat{m}}^{(\ell)} I)^{-1} v_1^{(\ell)}\| = 0,$$

which means that $A - \widehat{\theta}_{1,\widehat{m}}^{(\infty)} I$ is a singular matrix. In other words, $\theta_1^{(\infty)} = \widehat{\theta}_{1,\widehat{m}}^{(\infty)}$ is an eigenvalue of A. When $\theta_1^{(\ell)}$ converges to the largest eigenvalue, the Ritz vector $u_1^{(\ell)}$ converges to the corresponding eigenvector, which is the property of the Rayleigh-Ritz procedure. □

Finally, we would like to mention that a similar result is established for the Jacobi-Davidson method with Harmonic Ritz values in [24] for computing the eigenvalue with the smallest absolute value. This algorithm is mathematically equivalent to Algorithm 4 for A^{-1}. If A is nonsingular, then the largest eigenvalue of $B_{m_\ell}^{(\ell)}$ converges to an eigenvalue of A^{-1} by Theorem 6. When the convergent value is the largest eigenvalue of A^{-1}, the corresponding Harmonic Ritz vector converges to the corresponding eigenvector, which is the property of the Rayleigh-Ritz procedure.

Acknowledgments

The author is grateful to Masaaki Sugihara, Yusaku Yamamoto, and Yuji Nakatsukasa for their valuable comments and suggestions. The author is supported by JSPS Grant-in-Aid for Research Activity Start-up.

References

[1] J. Baglama, D. Calvetti, and L. Reichel: Iterative methods for the computation of a few eigenvalues of a large symmetric matrix, *BIT Numer. Math.*, vol. 36 (1996), pp. 400–421.

[2] J. Baglama, D. Calvetti, and L. Reichel: IRBL: an implicitly restarted block Lanczos method for largescale Hermitian eigenproblems, *SIAM J. Sci. Comput.*, vol. 24 (2003), pp. 1650–1677.

[3] Z. Bai, J. Demmel, J. Dongarra, A. Ruhe, and H. van der Vorst: *Templates for the Solution of Algebraic Eigenvalue Problems: A Practical Guide*, SIAM, Philadelphia, 2000.

[4] D. Calvetti, L. Reichel, and D. C. Sorensen: An implicitly restarted Lanczos method for large symmetric eigenvalue problems, *Electron. Trans. Numer. Anal.*, vol. 2 (1994), 1–21.

[5] M. Crouzeix, B. Philippe, and M. Sadkane: The Davidson method, *SIAM J. Sci. Comput.*, vol. 15 (1994), pp. 62–76.

[6] J. K. Cullum: The simultaneous computation of a few of the algebraically largest and smallest eigenvalues of a large, symmetric, sparse matrix, *BIT*, vol. 18 (1978), pp. 265–275.

[7] J. K. Cullum and W. E. Donath: A block Lanczos algorithm for computing the q algebraically largest eigenvalues and a corresponding eigenspace for large, sparse symmetric matrices, *In Proceedings of the 1994 IEEE Conference on Decision and Control*, pages 505–509. IEEE Press, Piscataway, NJ, 1974.

[8] E. R. Davidson: The iterative calculation of a few of the lowest eigenvalues and corresponding eigenvectors of large real-symmetric matrices, *J. Comput. Phys.*, vol. 17 (1975), pp. 87–94.

[9] J. W. Demmel: *Applied Numerical Linear Algebra*, SIAM, Philadelphia, 1997.

[10] G. H. Golub and R. Underwood: The block Lanczos method for computing eigenvalues, *In J. Rice, editor, Mathematical Software* III, pages 364–377. Academic Press, New York, 1977.

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

[12] M. E. Hochstenbach and G. L. G. Sleijpen: Two-sided and alternating Jacobi-Davidson, *Linear Algebra Appl.*, 358 (2003) 145–172.

[13] R. A. Horn and C. R. Johnson: *Matrix Analysis*, 2nd edition, Cambridge University, 2012.

[14] W. Karush: An iterative method for finding characteristic vectors of a symmetric matrix, *Pacific J. Math.*, vol. 1 (1951), pp. 233–248.

[15] A. V. Knyazev: Toward the optimal preconditioned eigensolver: locally optimal block preconditioned conjugate gradient method, *SIAM J. Sci. Comput.*, vol. 23 (2001), pp. 517–541.

[16] A. V. Knyazev and A. L. Skorokhodov: On exact estimates of the convergence rate of the steepest ascent method in the symmetric eigenvalue problem, *Linear Algebra Appl.*, vol. 154-156 (1991), pp. 245–257.

[17] C. Lanczos: An iteration method for the solution of the eigenvalue problem of linear differential and integral operators, *J. Res. Nat. Bur. Standards*, vol. 45 (1950), pp. 255–282.

[18] R. B. Morgan and D. S. Scott: Generalizations of Davidson's method for computing eigenvalues of sparse symmetric matrices, *SIAM J. Sci. Statist. Comput.*, vol. 7 (1986), pp. 817–825.

[19] Y. Notay: Is Jacobi-Davidson faster than Davidson? *SIAM J. Matrix Anal. Appl.*, vol. 26 (2005), pp. 522–543.

[20] E. Ovtchinnikov: Convergence estimates for the generalized Davidson method for symmetric eigenvalue problems I: the preconditioning aspect, *SIAM J. Numer. Anal.*, vol. 41 (2003), pp. 258–271.

[21] E. Ovtchinnikov: Convergence estimates for the generalized Davidson method for symmetric eigenvalue problems II: the preconditioning aspect, *SIAM J. Numer. Anal.*, vol. 41 (2003), pp. 272–286.

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

[23] A. Ruhe: Implementation aspects of band Lanczos algorithms for computation of eigenvalues of large sparse symmetric matrices, *Math. Comp.*, vol. 33 (1979), pp. 680–687.

[24] G. L. G. Sleijpen and H. A. van der Vorst: A Jacobi-Davidson iteration method for linear eigenvalue problems, *SIAM J. Matrix Anal. Appl.*, vol. 17 (1996), pp. 401–425.

[25] D. C. Sorensen: Implicit application of polynomial filters in a k-step Arnoldi method, *SIAM J. Matrix Anal. Appl.*, vol. 13 (1992), pp. 357–385.

[26] A. Stathopoulos: Nearly optimal preconditioned methods for Hermitian eigenproblems under limited memory. Part I: seeking one eigenvalue, *SIAM J. Sci. Comput.*, vol. 29 (2007), pp. 481–514.

[27] A. Stathopoulos: Nearly optimal preconditioned methods for Hermitian eigenproblems under limited memory. Part II: seeking many eigenvalues, *SIAM J. Sci. Comput.*, vol. 29 (2007), pp. 2162–2188.

[28] K. Wu and H. Simon: Thick-restarted Lanczos method for large symmetric eigenvalue problems, *SIAM J. Matrix Anal. Appl.*, vol. 22 (2000), pp. 602–616.

[29] I. Yamazaki, Z. Bai, H. Simon, L. Wang, and K. Wu: Adaptive projection subspace dimension for the thick-restart Lanczos method, *ACM Trans. Math. Softw.*, vol. 47 (2010), Article No. 27.