MATHEMATICAL ENGINEERING TECHNICAL REPORTS

On Convergence of Iterative Projection Methods for Symmetric Eigenvalue Problems

Kensuke AISHIMA

METR 2015–25 August 2015

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

On Convergence of Iterative Projection Methods 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

August 2015

Abstract

We prove global convergence of particular iterative projection methods using the so-called shift-and-invert technique for solving symmetric generalized eigenvalue problems. In particular, we aim to provide a variant of the convergence theorem obtained by Crouzeix, Philippe, and Sadkane for the generalized Davidson method. Our result covers the Jacobi-Davidson and the rational Krylov methods with restarting and preconditioning that are important techniques for modern eigensolvers. More specifically, we prove that the Ritz pairs converge to exact eigenpairs, even though they are not necessarily the target eigenpairs. We would like to emphasize that our proof is not a routine consideration of Crouzeix, Philippe, and Sadkane. To complete the proof, we discover a key lemma, which leads to a very simple convergence proof, resulting in a new theorem similar to that of Crouzeix, Philippe, and Sadkane.

1 Introduction

In this study, we focus on convergence theory for iterative projection methods for finding smallest eigenvalues of the generalized eigenvalue problem

$$Ax = \lambda Bx, \quad A, B \in \mathbb{R}^{n\times n}, \tag{1}$$

where A is symmetric, and B is symmetric positive definite. Solving such eigenvalue problems is important in many scientific and engineering applications. For example, (1) arises from discretization of a self-adjoint operator for elliptic partial differential equations. Some discretization methods are associated with projection to a finite-dimensional subspace.

For large matrix eigenvalue problems, projection methods are also effective, as in the case of the discretization of infinite dimensional operators. Nearly all the effective projection methods for generalized symmetric eigenvalue problems are based on the Rayleigh-Ritz procedure for a subspace of the Euclidean space $\mathbb{R}^n$ [4, 12, 13, 34, 36]. The best-known method is the Lanczos method using a Krylov subspace [20]. In recent times, the rational Krylov method has also attracted much attention [7, 14, 21, 22, 23, 35]. See [3, 5, 6, 46, 48] for recent developments regarding Krylov subspace methods. Furthermore, the Davidson [10] and Jacobi-Davidson methods [39] are familiar. Moreover, a generalized Davidson method exists [25, 31, 32], which can be viewed as a general framework that includes the Lanczos, Davidson, and Jacobi-Davidson methods. Furthermore, the steepest descent and conjugate gradient methods for minimizing the Rayleigh-quotient

$$\rho(x) := \frac{x^{\mathrm{T}}Ax}{x^{\mathrm{T}}Bx} \tag{2}$$

to obtain the eigenvalue often use the Rayleigh-Ritz procedure [17, 19]. Moreover, some contour integral methods [15, 45] can be viewed as belonging to the class of the Rayleigh-Ritz procedure.

This study discusses theoretical global convergence properties for such methods. In 1994, Crouzeix, Philippe, and Sadkane derived a global convergence theorem [8, Theorem 2.1] for the restarted generalized Davidson method that covers the restarted block Lanczos method, whereas Sorensen proved global convergence of the restarted Lanczos in 1992 [40, Theorem 5.9]. Furthermore, in 2015, Sorensen's result was extended to general situations [1, Theorems 3 and 4] to cover modern sophisticated restart strategies [46, 48].

The steepest descent (PSD) and locally optimal conjugate gradient methods (LOPCG) with preconditioning for the Rayleigh-quotient are also effective with convergence rates that can be derived for an appropriate initial guess and preconditioning [17, 18, 27, 28]. In recent times, it was shown in [44] that global and asymptotic convergence of the basic conjugate gradient (CG) method applies to nonlinear Hermitian eigenvalue problems. From another standpoint, PSD and LOPCG can be viewed as a sort of restarted generalized Davidson method. In 2003, Ovtchinnikov presented convergence estimates from this viewpoint in [32]. Although such an approach has derived sharper convergence estimates than that of the generalized Davidson method with suitable initial guesses for more than a decade [30, 33], they do not show global convergence properties of the Jacobi-Davidson method for any initial guess, which cannot be covered by Crouzeix, Philippe, and Sadkane [8, Theorem 2.1]. Global convergence of the restarted Jacobi-Davidson method was proved in [1, Theorem 6] for the first time in 2015. In particular, [1, Theorem 6] shows that the smallest Ritz value converges to an exact eigenvalue, although not necessarily to the smallest one, in the same manner as [8, Theorem 2.1]. This global convergence property is also similar to the Rayleigh-quotient iteration [34, Theorem 4.9.1].

In connection with recent developments of the (inexact) Rayleigh-quotient iteration [2, 11, 16, 29, 37, 43, 47], the (inexact) rational Krylov method using the so-called shift-and-invert technique or the Cayley transform has recently been thoroughly investigated [7, 14, 22, 23]. The shift-and-invert or Cayley transform techniques can be viewed as a sort of preconditioning, which has been also thoroughly investigated to accelerate general iterative projection methods [24, 41, 42].

With such a background of studies, we would like to construct a general framework concerning global convergence for iterative projection methods. In particular, we aim to extend the convergence proof for the Jacobi-Davidson [1, Theorem 6] to more general methods in the same manner as that of Crouzeix, Philippe, and Sadkane [8, Theorem 2.1]. Our result covers the rational Krylov method with a restart strategy and preconditioning. We would like to emphasize that our proof is not a routine consideration of [8, Theorem 2.1]. We need a key lemma (Lemma 1) to complete the convergence proof. This key lemma leads to a very simple convergence proof, though the convergence rate cannot be derived from it.

This paper is organized as follows. Section 2 is devoted to descriptions of the Rayleigh-Ritz procedure with restart strategy in an abstract form and the convergence proof by Crouzeix, Philippe, and Sadkane [8] for the generalized Davidson method for solving standard symmetric eigenvalue problems. This theorem is extended to generalized symmetric eigenvalue problems, and our goal is clarified in Section 3. In Section 4, we derive a new convergence theorem including the Jacobi-Davidson and rational Krylov methods with restarting and preconditioning in an abstract form.

Notation. Throughout this study, $A \in \mathbb{R}^{n\times n}$ is symmetric, $B \in \mathbb{R}^{n\times n}$ is symmetric positive definite, and the generalized eigenvalues for (A, B) are $\lambda_1 \leq \cdots \leq \lambda_n$. Furthermore, $X_i := [x_1, \dots, x_i]$ the matrix whose jth column is the corresponding eigenvector x_j to λ_j for any $j = 1, \dots, i$, normalized as $X_i^{\mathrm{T}} B X_i = I$, where I is the identity matrix. For any $V \in \mathbb{R}^{n\times i}$, let $\mathrm{span}\{V\}$ be the subspace spanned by the columns of V. Moreover, k is the number of desired smallest eigenvalues. For any vector $v \in \mathbb{R}^n$, let $\|v\|$ be $\sqrt{v^{\mathrm{T}} v}$ and $\|v\|_B$ be $\sqrt{v^{\mathrm{T}} B v}$.

2 Rayleigh-Ritz procedure and convergence theory [8]

We describe the Rayleigh-Ritz procedure with restart strategy in an abstract mathematical form as follows.

Algorithm 1 A framework of iterative projection methods with restarting for computing the k smallest eigenvalues of $Ax = \lambda Bx$.

Input: $A, B \in \mathbb{R}^{n\times n}$ and $V^{(0)} = [v_1^{(0)}, \dots, v_{m_0}^{(0)}] \in \mathbb{R}^{n\times m_0}$

for $\ell := 0, 1, \dots,$ **do**
  compute $A^{(\ell)} = {V^{(\ell)}}^{\mathrm{T}} A V^{(\ell)}$, $B^{(\ell)} = {V^{(\ell)}}^{\mathrm{T}} B V^{(\ell)}$
  compute the k smallest eigenvalues for $(A^{(\ell)}, B^{(\ell)})$: $\lambda_1^{(\ell)} \le \cdots \le \lambda_k^{(\ell)}$
  compute the corresponding Ritz vectors $x_1^{(\ell)}, \dots, x_k^{(\ell)}$
  compute $V^{(\ell+1)} := [v_1^{(\ell+1)}, \dots, v_{m_{\ell+1}}^{(\ell+1)}]$, where $\mathrm{span}\{V^{(\ell+1)}\} \ni x_i^{(\ell)}$ for $i = 1, \dots, k$
end for

For example, the restarted Lanczos method corresponds to the situation where $V^{(\ell)}$ for any $\ell \in \mathbb{N}$ is a Krylov subspace. The relationship of Algorithm 1 and modern solvers to generalized symmetric eigenvalue problems is discussed in the next section. In general, most iterative projection methods are related to Algorithm 1, and are of one of two types, i.e., the residual based and the Rayleigh-quotient iteration methods as shown in Table 1 in the next section. In this section, we first discuss global convergence of the above algorithm for standard eigenvalue problems, for which $B = I$.

2.1 Convergence proof by Crouzeix, Philippe, and Sadkane

Here we focus on standard symmetric eigenvalue problems. Crouzeix, Philippe, and Sadkane [8] proposed the following algorithm, an instance of Algorithm 1.

Algorithm 2 The generalized Davidson method for $Ax = \lambda x$ in [8].

Input: $A \in \mathbb{R}^{n\times n}$ and $V^{(0)} = [v_1^{(0)}, \dots, v_k^{(0)}] \in \mathbb{R}^{n\times m_0}$ with ${V^{(0)}}^{\mathrm{T}} V^{(0)} = I$

1: **for** $\ell := 0, 1, \dots,$ **do**
2:   compute $A^{(\ell)} = {V^{(\ell)}}^{\mathrm{T}} A V^{(\ell)}$
3:   compute the k smallest eigenvalues of $A^{(\ell)}$: $\lambda_1^{(\ell)} \le \cdots \le \lambda_k^{(\ell)}$
4:   compute the corresponding Ritz vectors $x_1^{(\ell)}, \dots, x_k^{(\ell)}$
5:   compute the residuals $r_i^{(\ell)} := A x_i^{(\ell)} - \lambda_i^{(\ell)} x_i^{(\ell)}$ $(1 \le i \le k)$
6:   compute the new directions $t_i^{(\ell)} := C_i^{(\ell)} r_i^{(\ell)}$ $(1 \le i \le k)$
7:   **if** $\dim(\mathrm{span}\{V^{(\ell)}\}) \le m - k$ **then**
8:     $V^{(\ell+1)} := \mathrm{GS}(V^{(\ell)}, t_1^{(\ell)}, \dots, t_k^{(\ell)})$
9:   **else**
10:     $V^{(\ell+1)} := \mathrm{GS}(x_1^{(\ell)}, \dots, x_k^{(\ell)}, t_1^{(\ell)}, \dots, t_k^{(\ell)})$
11:   **end if**
12: **end for**

In lines 8 and 10, GS is the Gram-Schmidt orthogonalization. The block

Lanczos method corresponds to $C_i^{(\ell)} = I$ $(i = 1, \dots, k,\ \ell \in \mathbb{N})$, which satisfy the convergence conditions. In general, $C_i^{(\ell)}$ $(i = 1, \dots, k,\ \ell \in \mathbb{N})$ in line 6 can be regarded as the preconditioning required to obtain the new directions $t_i^{(\ell)} := C_i^{(\ell)} r_i^{(\ell)}$ $(i = 1, \dots, k,\ \ell \in \mathbb{N})$. To prove the convergence of Algorithm 2, Crouzeix, Philippe, and Sadkane derived a convergence theorem for Algorithm 1 in an abstract form, as follows [8].

Theorem 1 ([8])**.** *Suppose that Algorithm 1 is applied to the symmetric eigenvalue problem $Ax = \lambda x$. Then, for $i = 1, \dots, k$, the sequences $\{\lambda_i^{(\ell)}\}_{\ell \in \mathbb{N}}$ are nonincreasing and convergent. Let ${V^{(\ell)}}^{\mathrm{T}} V^{(\ell)} = I$ for all $\ell \in \mathbb{N}$. Moreover, let a set of matrices $\{C_i^{(\ell)}\}$ satisfy the following assumption: for any $i = 1, \dots, k$, there exist $K_1,\ K_2 > 0$ such that for any $\ell \in \mathbb{N}$ and for any vector $v \in span\{V^{(\ell)}\}^{\perp}$: $K_1 \|v\|^2 \leq v^{\mathrm{T}} C_i^{(\ell)} v \leq K_2 \|v\|^2$. Furthermore, we assume that, for any $i = 1, \dots, k,\ \ell \in \mathbb{N}$, the vector $(I - V^{(\ell)} {V^{(\ell)}}^{\mathrm{T}}) C_i^{(\ell)} (A - \lambda_i^{(\ell)} I) x_i^{(\ell)}$ belongs to $span\{V^{(\ell+1)}\}$. Then, for any $i = 1, \dots k$, $\lim_{\ell \to \infty} \lambda_i^{(\ell)}$ is an eigenvalue of A, and the accumulation points of $\{x_i^{(\ell)}\}_{\ell \in \mathbb{N}}$ are the corresponding eigenvectors.*

In [8], it is shown that $C_i^{(\ell)}$ $(i = 1, \dots, k,\ \ell \in \mathbb{N})$ are always positive definite diagonal matrices with suitable $V^{(0)}$ for the block Davidson method (Algorithm 2).

Our aim is to prove global convergence of other methods, such as the rational Krylov and Jacobi-Davidson methods, which incorporate restart strategies. For this purpose, we let

$$X_i^{(\ell)} := [x_1^{(\ell)}, \dots, x_i^{(\ell)}] \tag{3}$$

for the Ritz vectors, and we rewrite Theorem 1 as follows. Although the proof is nearly the same as the proof in [8], we present the proof to explain our contribution later.

Theorem 2. *Suppose that Algorithm 1 is applied to the symmetric eigenvalue problem $Ax = \lambda x$. Then, for $i = 1, \dots, k$, the sequences $\{\lambda_i^{(\ell)}\}_{\ell \in \mathbb{N}}$ are nonincreasing and convergent. For each $i = 1, \dots, k$, define $U_{m(i)}^{(\ell)} \in \mathbb{R}^{n \times m(i)}$ normalized to ${U_{m(i)}^{(\ell)}}^{\mathrm{T}} U_{m(i)}^{(\ell)} = I$ for $m(i) \geq i$, satisfying $span\{X_i^{(\ell)}\} \subseteq span\{U_{m(i)}^{(\ell)}\} \subseteq span\{V^{(\ell+1)}\}$ for all $\ell \in \mathbb{N}$. Let $u_i^{(\ell)}$ be the Ritz vector corresponding to the i-th smallest Ritz value $\theta_i^{(\ell)}$ for the subspace $span\{U_{m(i)}^{(\ell)}\}$. Moreover, let a set of matrices $\{C_i^{(\ell)}\}$ satisfy the following assumption: for any $i = 1, \dots, k$, there exist $K_1,\ K_2 > 0$ such that for any $\ell \in \mathbb{N}$, $s_i^{(\ell)} := (A - \theta_i^{(\ell)} I) u_i^{(\ell)}$ and $w_i^{(\ell)} := (I - U_{m(i)}^{(\ell)} {U_{m(i)}^{(\ell)}}^{\mathrm{T}}) C_i^{(\ell)} s_i^{(\ell)}$ satisfy*

$$K_1 \|s_i^{(\ell)}\|^2 \leq {s_i^{(\ell)}}^{\mathrm{T}} C_i^{(\ell)} s_i^{(\ell)},\ \|w_i^{(\ell)}\|^2 \leq K_2. \tag{4}$$

Furthermore, we assume that, for any $i = 1, \dots, k$, $\ell \in \mathbb{N}$, the vector $w_i^{(\ell)}$ belongs to span$\{V^{(\ell+1)}\}$. Then, for any $i = 1, \dots k$, $\lim_{\ell\to\infty} \lambda_i^{(\ell)}$ is an eigenvalue of A, and the accumulation points of $\{x_i^{(\ell)}\}_{\ell\in\mathbb{N}}$ are the corresponding eigenvectors.

Proof. It is easy to observe that for $i = 1, \dots, k$, the sequences $\{\lambda_i^{(\ell)}\}_{\ell\in\mathbb{N}}$ are nonincreasing and convergent from the Cauchy interlace theorem.

In what follows, we prove that, for any $i = 1, \dots k$, $\lim_{\ell\to\infty} \lambda_i^{(\ell)}$ is an eigenvalue of A, and the accumulation points of $\{x_i^{(\ell)}\}_{\ell\in\mathbb{N}}$ are the corresponding eigenvectors. Using $w_i^{(\ell)} := (I - U_{m(i)}^{(\ell)} U_{m(i)}^{(\ell)\,\mathrm{T}}) C_i^{(\ell)} s_i^{(\ell)}$, we see

$$u_i^{(\ell)\mathrm{T}} A w_i^{(\ell)} = s_i^{(\ell)\mathrm{T}} C_i^{(\ell)} s_i^{(\ell)}. \tag{5}$$

Let $\Pi_i^{(\ell)} := [u_i^{(\ell)}, w_i^{(\ell)}/\|w_i^{(\ell)}\|]$. Note that $\Pi_i^{(\ell)\mathrm{T}} \Pi_i^{(\ell)} = I$ and let

$$\Pi_i^{(\ell)\mathrm{T}} A \Pi_i^{(\ell)} = \begin{pmatrix} \theta_i^{(\ell)} & \alpha_i^{(\ell)} \\ \alpha_i^{(\ell)} & \beta_i^{(\ell)} \end{pmatrix}. \tag{6}$$

Then, we see

$$\alpha_i^{(\ell)} = \frac{u_i^{(\ell)\mathrm{T}} A w_i^{(\ell)}}{\|w_i^{(\ell)}\|} = \frac{s_i^{(\ell)\mathrm{T}} C_i^{(\ell)} s_i^{(\ell)}}{\|w_i^{(\ell)}\|} \tag{7}$$

from (5). To prove $\lim_{\ell\to\infty} s_i^{(\ell)} = 0$, where $s_i^{(\ell)} = (A - \theta_i^{(\ell)} I) u_i^{(\ell)}$, we investigate the behavior of $\alpha_i^{(\ell)}$ as follows.

Let $\widehat{\theta}_i^{(\ell)}$ denote the smallest eigenvalue of $\Pi_i^{(\ell)\mathrm{T}} A \Pi_i^{(\ell)}$. Noting the characteristic equation

$$\det(\Pi_i^{(\ell)\mathrm{T}} A \Pi_i^{(\ell)} - \widehat{\theta}_i^{(\ell)} I) = (\theta_i^{(\ell)} - \widehat{\theta}_i^{(\ell)})(\beta_i^{(\ell)} - \widehat{\theta}_i^{(\ell)}) - |\alpha_i^{(\ell)}|^2 = 0, \tag{8}$$

we have

$$|\alpha_i^{(\ell)}|^2 \le 2\|A\|(\theta_i^{(\ell)} - \widehat{\theta}_i^{(\ell)}) \tag{9}$$

because of $|\beta_i^{(\ell)}| \le \|A\|$ and $|\widehat{\theta}_i^{(\ell)}| \le \|A\|$. Using $\lambda_i^{(\ell+1)} \le \widehat{\theta}_i^{(\ell)} \le \theta_i^{(\ell)} \le \lambda_i^{(\ell)}$ and $\lim_{\ell\to\infty}(\lambda_i^{(\ell)} - \lambda_i^{(\ell+1)}) = 0$, we have

$$\lim_{\ell\to\infty} |\alpha_i^{(\ell)}| = 0. \tag{10}$$

From the relation (7), we have

$$|s_i^{(\ell)\mathrm{T}} C_i^{(\ell)} s_i^{(\ell)}| = \|w_i^{(\ell)}\| |\alpha_i^{(\ell)}|. \tag{11}$$

From (4) and (10), it follows that $\lim_{\ell\to\infty} s_i^{(\ell)} = 0$, where $s_i^{(\ell)} = (A - \theta_i^{(\ell)} I)u_i^{(\ell)}$. Hence, $\lim_{\ell\to\infty} \theta_i^{(\ell)}$ is an eigenvalue of A and the accumulation points of $\{u_i^{(\ell)}\}_{\ell\in\mathbb{N}}$ are the corresponding eigenvectors. Recall that $u_i^{(\ell)} \in \text{span}\{V^{(\ell+1)}\}$ for $i = 1, \dots, k$. From the properties of the Rayleigh-Ritz procedure, for any $i = 1, \dots k$, $\lim_{\ell\to\infty} \lambda_i^{(\ell)}$ is an eigenvalue of A and the accumulation points of $\{x_i^{(\ell)}\}_{\ell\in\mathbb{N}}$ are the corresponding eigenvectors. □

In this proof, the relation (7) is crucial. If we establish (7), the proof is naturally derived through easy calculations. In this study, to derive the new convergence theorem, we adapt such a relation to the proof for other eigensolvers. Furthermore, to complete our proof, we establish another crucial lemma (Lemma 1) in Section 4.

3 Toward generalized eigenvalue problems

In this section, we examine an easy extension of Theorem 2 to generalized eigenvalue problems as preparation for deriving our main result later.

Theorem 3. *In Algorithm 1, for $i = 1, \dots, k$, the sequences $\{\lambda_i^{(\ell)}\}_{\ell\in\mathbb{N}}$ are nonincreasing and convergent. For each $i = 1, \dots, k$, define $U_{m(i)}^{(\ell)} \in \mathbb{R}^{n\times m(i)}$ normalized to ${U_{m(i)}^{(\ell)}}^{\mathrm{T}} B U_{m(i)}^{(\ell)} = I$ for $m(i) \geq i$, satisfying span$\{X_i^{(\ell)}\} \subseteq$ span$\{U_{m(i)}^{(\ell)}\} \subseteq$ span$\{V^{(\ell+1)}\}$ for all $\ell \in \mathbb{N}$. Let $u_i^{(\ell)}$ be the Ritz vector corresponding to the i-th smallest Ritz value $\theta_i^{(\ell)}$ for the subspace span$\{U_{m(i)}^{(\ell)}\}$. Moreover, let a set of matrices $\{C_i^{(\ell)}\}$ satisfy the following assumption: for any $i = 1, \dots, k$, there exist $K_1,\ K_2 > 0$ such that for any $\ell \in \mathbb{N}$, $s_i^{(\ell)} := (A - \theta_i^{(\ell)} B)u_i^{(\ell)}$ and $w_i^{(\ell)} := (I - U_{m(i)}^{(\ell)} {U_{m(i)}^{(\ell)}}^{\mathrm{T}} B) C_i^{(\ell)} s_i^{(\ell)}$ satisfy*

$$K_1 \|s_i^{(\ell)}\|^2 \leq s_i^{(\ell)\mathrm{T}} C_i^{(\ell)} s_i^{(\ell)},\ \ \|w_i^{(\ell)}\|_B^2 \leq K_2. \tag{12}$$

Furthermore, we assume that, for any $i = 1, \dots, k,\ \ell \in \mathbb{N}$, the vector $w_i^{(\ell)}$ belongs to span$\{V^{(\ell+1)}\}$. Then, for any $i = 1, \dots k$, $\lim_{\ell\to\infty} \lambda_i^{(\ell)}$ is an eigenvalue for (A, B), and the accumulation points of $\{x_i^{(\ell)}\}_{\ell\in\mathbb{N}}$ are the corresponding eigenvectors.

Proof. It is easy to observe that, for $i = 1, \dots, k$, the sequences $\{\lambda_i^{(\ell)}\}_{\ell\in\mathbb{N}}$ are nonincreasing and convergent, as in the proof of Theorem 2.

In what follows, we prove that, for any $i = 1, \dots k$, $\lim_{\ell\to\infty} \lambda_i^{(\ell)}$ is an eigenvalue and the accumulation points of $\{x_i^{(\ell)}\}_{\ell\in\mathbb{N}}$ are the corresponding eigenvectors. Using $w_i^{(\ell)} := (I - U_{m(i)}^{(\ell)} {U_{m(i)}^{(\ell)}}^{\mathrm{T}} B) C_i^{(\ell)} s_i^{(\ell)}$, we see

$$u_i^{(\ell)\mathrm{T}} A w_i^{(\ell)} = s_i^{(\ell)\mathrm{T}} C_i^{(\ell)} s_i^{(\ell)}.$$

Let $\Pi_i^{(\ell)} := [u_i^{(\ell)}, w_i^{(\ell)}/\|w_i^{(\ell)}\|_B]$. Note $\Pi_i^{(\ell)\mathrm{T}}\Pi_i^{(\ell)} = I$ and let

$$\Pi_i^{(\ell)\mathrm{T}} A \Pi_i^{(\ell)} = \begin{pmatrix} \theta_i^{(\ell)} & \alpha_i^{(\ell)} \\ \alpha_i^{(\ell)} & \beta_i^{(\ell)} \end{pmatrix}.$$

in the same way as (6). It then follows that

$$\alpha_i^{(\ell)} = \frac{u_i^{(\ell)\mathrm{T}} A w_i^{(\ell)}}{\|w^{(\ell)}\|_B} = \frac{s_i^{(\ell)\mathrm{T}} C_i^{(\ell)} s_i^{(\ell)}}{\|w^{(\ell)}\|_B}. \tag{13}$$

Let $\widehat{\theta}_i^{(\ell)}$ be the smallest eigenvalue of $\Pi_i^{(\ell)\mathrm{T}} A \Pi_i^{(\ell)}$. We observe

$$\det(\Pi_i^{(\ell)\mathrm{T}} A \Pi_i^{(\ell)} - \widehat{\theta}_i^{(\ell)} I) = (\theta_i^{(\ell)} - \widehat{\theta}_i^{(\ell)})(\beta_i^{(\ell)} - \widehat{\theta}_i^{(\ell)}) - |\alpha_i^{(\ell)}|^2 = 0.$$

Hence, similar to (9), we have

$$|\alpha_i^{(\ell)}|^2 \leq 2\|B^{-1}A\|(\theta_i^{(\ell)} - \widehat{\theta}_i^{(\ell)})$$

because of $|\beta_i^{(\ell)}| \leq \|B^{-1}A\|$ and $|\widehat{\theta}_i^{(\ell)}| \leq \|B^{-1}A\|$. Using $\lambda_i^{(\ell+1)} \leq \widehat{\theta}_i^{(\ell)} \leq \theta_i^{(\ell)} \leq \lambda_i^{(\ell)}$ and $\lim_{\ell\to\infty}(\lambda_i^{(\ell)} - \lambda_i^{(\ell+1)}) = 0$, we have $\alpha_i^{(\infty)} = 0$. Therefore, we have $\lim_{\ell\to\infty} s_i^{(\ell)} = 0$ in the same manner as in the proof of Theorem 2. This completes the proof. □

This theorem establishes global convergence in the case of $C_i^{(\ell)} = I$ ($i = 1, \ldots, k,\ \ell \in \mathbb{N}$), though the convergence speed might be very slow. The block Lanczos method corresponds to the situation $C_i^{(\ell)} = B^{-1}$ ($i = 1, \ldots, k,\ \ell \in \mathbb{N}$), which is an ideal case. If solving the linear system requires high computational cost for the large matrix B, then the linear system is often solved approximately, which corresponds, in a sense, to the situation $C_i^{(\ell)} \approx B^{-1}$ ($i = 1, \ldots, k,\ \ell \in \mathbb{N}$). For example, if we determine $C_i^{(\ell)}$ ($i = 1, \ldots, k,\ \ell \in \mathbb{N}$) as the inverse matrices of the diagonal part of B in the same way as in the Davidson method, then the convergence is also guaranteed.

Here we classify the popular projection methods based on the new direction expanding subspace in Theorem 3. In general, there are two classes, i.e., the residual-based and Rayleigh-quotient iteration methods. The former aims to add the residual $(A - \theta_i^{(\ell)} B)u_i^{(\ell)}$ (improved by B^{-1}) to the new subspace, which is also interpreted as one power iteration to obtain a new direction. Hence, this covers the Lanczos method (e.g., see [8]). The latter aims to filter the target eigenpairs directly using the Rayleigh-quotient as $(A - \theta_i^{(\ell)} B)^{-1} B u_i^{(\ell)}$. From this viewpoint, we classify the popular projection methods as in Table 1 (see [4] for details).

Table 1: Classification of the popular eigensolvers from the new direction expanding subspace.

$B^{-1}(A-\theta_i^{(\ell)}B)u_i^{(\ell)}$ (with a slight modification)	$(A-\theta_i^{(\ell)}B)^{-1}Bu_i^{(\ell)}$ (with a slight modification)
Lanczos [20, 40]	Jacobi-Davidson [39]
PSD (PINVIT) [19, 26]	Rational Krylov [35]
LOPCG [17]	

More precisely, the new directions of the PSD and LOPCG are $T(A-\theta_i^{(\ell)}B)u_i^{(\ell)}$, where T is often assumed to be a positive definite matrix filtering the target eigenvector in the same manner as in the right part of Table 1. In other words, $(A-\theta_i^{(\ell)}B)u_i^{(\ell)}$ is the steepest descent direction for the Rayleigh-quotient $u_i^{(\ell)\mathrm{T}}Au_i^{(\ell)}/u_i^{(\ell)\mathrm{T}}Bu_i^{(\ell)}$, then T can be viewed as the preconditioning matrix. Since the generalized Davidson is a general framework, it exists in both the classes. Furthermore, there are block versions (e.g., block Lanczos [9], LOBPCG [17], block Jacobi-Davidson [38], and so forth). Although Theorem 3 covers the left part of Table 1, it does not cover the right part where the new directions are obtained using $(A-\theta_i^{(\ell)}B)^{-1}Bu_i^{(\ell)}$ $(i = 1,\dots,k,\ \ell \in \mathbb{N})$, such as the Jacobi-Davidson [39] (see also [1, §7]), and rational Krylov [23, 35] (see also [4, §8.5]). If we apply Theorem 3 to such situations, we must let $C_i^{(\ell)} := (A-\theta_i^{(\ell)}B)^{-1}B(A-\theta_i^{(\ell)}B)^{-1}$ $(i = 1,\dots,k,\ \ell \in \mathbb{N})$. For simplicity, we assume $k = 1$ for the Jacobi-Davidson method. If we assume $\theta_1^{(\infty)}$ is not equal to an eigenvalue, then the conditions in (12) are satisfied. Hence, $\theta_1^{(\ell)}$ is convergent to an eigenvalue. This is a contradiction. Therefore, $\theta_1^{(\infty)} = \lambda_1^{(\infty)}$ would be expected to be an eigenvalue. However, in such a case, $\|w_1^{(\ell)}\|_B^2$ is not bounded in view of (13), $\alpha_1^{(\infty)} = 0$, and $C_1^{(\ell)} := (A-\theta_1^{(\ell)}B)^{-1}B(A-\theta_1^{(\ell)}B)^{-1}$. In other words, $\|w_1^{(\ell)}\|_B^2 \le K_2$ in (12) does not hold in Theorem 3 for the condition of $\lim_{\ell\to\infty} s_1^{(\ell)} = 0$, and hence the convergence of $x_1^{(\ell)}$ cannot be theoretically guaranteed.

With such a background, we derive another convergence theorem that covers the Ritz vectors in the case of $C_i^{(\ell)} := (A-\theta_i^{(\ell)}B)^{-1}B(A-\theta_i^{(\ell)}B)^{-1}$ in the next section. Note that the following analysis also focuses on the situation in which $(A-\theta_i^{(\ell)}B)^{-1}Bu_i^{(\ell)}$ for $i = 1,\dots,k,\ \ell \in \mathbb{N}$ are solved approximately from the practical point of view.

4 New convergence theorem

In this section, we derive another new convergence theorem for particular iterative projection methods, such as the Jacobi-Davidson, and rational Krylov methods and some methods with preconditioning related to $(A - \theta_i^{(\ell)} B)^{-1}$.

To this end, we provide the following crucial lemma.

Lemma 1. *Suppose that A is symmetric, and B is symmetric positive definite. Let $\{U^{(\ell)}\}_{\ell \in \mathbb{N}}$ and $\{V^{(\ell)}\}_{\ell \in \mathbb{N}}$ be the sequences of matrices satisfying span$\{U^{(\ell)}\} \subseteq$ span$\{V^{(\ell+1)}\}$ for all $\ell \in \mathbb{N}$. For span$\{U^{(\ell)}\}$, let $\theta_i^{(\ell)}$ denote the i-th smallest Ritz value and $u_i^{(\ell)}$ the corresponding Ritz vector. For span$\{V^{(\ell)}\}$, let $\lambda_i^{(\ell)}$ and $x_i^{(\ell)}$ be the Ritzpair in a similar manner. Furthermore, let $\{y^{(\ell)}\}_{\ell \in \mathbb{N}}$ be a sequence satisfying $\|y^{(\ell)}\|_B \leq K$ for all $\ell \in \mathbb{N}$ and positive constant $K > 0$. Assume that $\{\theta_i^{(\ell)}\}_{\ell \in \mathbb{N}}$ and $\{\lambda_i^{(\ell)}\}_{\ell \in \mathbb{N}}$ are convergent, that $\theta_i^{(\infty)} = \lambda_i^{(\infty)}$, and that $\{y_i^{(\ell)}\}_{\ell \in \mathbb{N}}$ satisfies $(A - \theta_i^{(\ell)} B)^{-1} y^{(\ell)} \in$ span$\{V^{(\ell+1)}\}$ and $\lim_{\ell \to \infty} 1/\|(A - \theta_i^{(\ell)} B)^{-1} y^{(\ell)}\|_B = 0$. Then, $\lambda_i^{(\infty)}$ is an eigenvalue for (A, B) and the accumulation points of $\{x_i^{(\ell)}\}_{\ell \in \mathbb{N}}$ are the corresponding eigenvectors.*

Proof. Since $\theta_i^{(\ell)}$ is convergent and $\lim_{\ell \to \infty} 1/\|(A-\theta_i^{(\ell)} B)^{-1} y^{(\ell)}\|_B = 0$, $(A - \theta_i^{(\infty)} B)$ is singular. In other words, $\lambda_i^{(\infty)} (= \theta_i^{(\infty)})$ is an eigenvalue. Therefore, the accumulation points of $\{(A - \theta_i^{(\ell)} B)^{-1} y^{(\ell)} / \|(A - \theta_i^{(\ell)} B)^{-1} y^{(\ell)}\|_B\}_{\ell \in \mathbb{N}}$ are the eigenvectors corresponding to the eigenvalue $\lambda_i^{(\infty)} (= \theta_i^{(\infty)})$ for (A, B). Moreover, noting $(A - \theta_i^{(\ell)} B)^{-1} y^{(\ell)} \in$ span$\{V^{(\ell+1)}\}$, we see that the accumulation points of $\{x_i^{(\ell)}\}_{\ell \in \mathbb{N}}$ are also the corresponding eigenvectors. □

In this proof, the convergence of $\theta_i^{(\ell)}$ to an eigenvalue is obvious because $(A - \theta_i^{(\infty)} B)$ must be singular. In addition, $u_i^{(\ell)}$ corresponding to $\theta_i^{(\ell)}$ is also regarded as the Ritz vector for span$\{V^{(\ell+1)}\}$ as $\ell \to \infty$ in view of span$\{U^{(\ell)}\} \subseteq$ span$\{V^{(\ell+1)}\}$ and $\theta_i^{(\infty)} = \lambda_i^{(\infty)}$. Hence, most researchers might try to prove that $\{u_i^{(\ell)}\}_{\ell \in \mathbb{N}}$ is also convergent to the eigenvector, but this cannot be proved directly. To overcome this limitation, we track the behavior of $\{(A - \theta_i^{(\ell)} B)^{-1} y^{(\ell)} / \|(A - \theta_i^{(\ell)} B)^{-1} y^{(\ell)}\|_B\}_{\ell \in \mathbb{N}}$, resulting in the finding of the corresponding eigenvector in span$\{V^{(\ell)}\}$. This finding is crucial, even though proving its convergence is not so difficult. Since $\{u_i^{(\ell)}\}_{\ell \in \mathbb{N}}$ corresponds to $\{\theta_i^{(\ell)}\}_{\ell \in \mathbb{N}}$ converging to an eigenvalue, it is difficult to find out $\{(A - \theta_i^{(\ell)} B)^{-1} y^{(\ell)}\}_{\ell \in \mathbb{N}}$ using $\{y^{(\ell)}\}_{\ell \in \mathbb{N}}$, which has nothing to do with $\{\text{span}\{V^{(\ell)}\}\}_{\ell \in \mathbb{N}}$. In fact, the latest paper [1] cannot find such a sequence, and hence the convergence of the Ritz vector in the Jacobi-Davidson method cannot be proved in the general situation [1, Theorem 6]. In this study, combining this finding with Theorem 3 using appropriate definitions, we obtain the following main result covering the eigenvectors.

Theorem 4. *In Algorithm 1, for $i = 1, \ldots, k$, the sequences $\{\lambda_i^{(\ell)}\}_{\ell \in \mathbb{N}}$ are nonincreasing and convergent. For each $i = 1, \ldots, k$, define $U_{m(i)}^{(\ell)} \in \mathbb{R}^{n \times m(i)}$*

normalized to $U_{m(i)}^{(\ell)}{}^{\mathrm{T}} B U_{m(i)}^{(\ell)} = I$ *for* $m(i) \geq i$*, satisfying span*$\{X_i^{(\ell)}\} \subseteq$ *span*$\{U_{m(i)}^{(\ell)}\} \subseteq$ *span*$\{V^{(\ell+1)}\}$ *for all* $\ell \in \mathbb{N}$*. Let* $u_i^{(\ell)}$ *be the Ritz vector corresponding to the* i*-th smallest Ritz value* $\theta_i^{(\ell)}$ *for the subspace span*$\{U_{m(i)}^{(\ell)}\}$*. Moreover, let a set of matrices* $\{C_i^{(\ell)}\}$ *satisfy the following assumption: for any* $i = 1, \ldots, k$*, there exist* $K_1, K_2 > 0$ *such that for any* $\ell \in \mathbb{N}$*,* $u_i^{(\ell)}$ *satisfies*

$$K_1 \leq u_i^{(\ell)\mathrm{T}} C_i^{(\ell)} u_i^{(\ell)}, \ \|C_i^{(\ell)} u_i^{(\ell)}\|_B^2 \leq K_2. \tag{14}$$

Furthermore, we assume that, for any $i = 1, \ldots, k$*,* $\ell \in \mathbb{N}$*, the vector* $(A - \theta_i^{(\ell)} B)^{-1} C_i^{(\ell)} u_i^{(\ell)}$ *belongs to span*$\{V^{(\ell+1)}\}$*. Then, for any* $i = 1, \ldots k$*,* $\lim_{\ell\to\infty} \lambda_i^{(\ell)}$ *is an eigenvalue for* (A, B)*, and the accumulation points of* $\{x_i^{(\ell)}\}_{\ell\in\mathbb{N}}$ *are the corresponding eigenvectors.*

Proof. Since the convergence $\lim_{\ell\to\infty} \lambda_i^{(\ell)}$ for any $i = 1, \ldots k$ is obvious, we prove that, for any $i = 1, \ldots k$, $\lim_{\ell\to\infty} \lambda_i^{(\ell)}$ is an eigenvalue of A, and the accumulation points of $\{x_i^{(\ell)}\}_{\ell\in\mathbb{N}}$ are the corresponding eigenvectors. Letting

$$w_i^{(\ell)} := (I - U_{m(i)}^{(\ell)} U_{m(i)}^{(\ell)}{}^{\mathrm{T}} B)(A - \theta_i^{(\ell)} B)^{-1} C_i^{(\ell)} u_i^{(\ell)}, \tag{15}$$

we see

$$u_i^{(\ell)\mathrm{T}} A w_i^{(\ell)} = u_i^{(\ell)\mathrm{T}} C_i^{(\ell)} u_i^{(\ell)}. \tag{16}$$

Define $\alpha_i^{(\ell)}$ in the same manner as in Theorem 3. Then, we have

$$\alpha_i^{(\ell)} = \frac{u_i^{(\ell)\mathrm{T}} A w_i^{(\ell)}}{\|w_i^{(\ell)}\|_B} = \frac{u_i^{(\ell)\mathrm{T}} C_i^{(\ell)} u_i^{(\ell)}}{\|w_i^{(\ell)}\|_B} \tag{17}$$

and $\alpha_i^{(\infty)} = 0$ in the same manner as in the proof of Theorem 3. Noting the definition (15), and $K_1 \leq u_i^{(\ell)\mathrm{T}} C_i^{(\ell)} u_i^{(\ell)}$ in (14), we have

$$\lim_{\ell\to\infty} \frac{K_1}{\|(I - U_{m(i)}^{(\ell)} U_{m(i)}^{(\ell)}{}^{\mathrm{T}} B)(A - \theta_i^{(\ell)} B)^{-1} C_i^{(\ell)} u_i^{(\ell)}\|_B} = 0, \tag{18}$$

where $K_1 > 0$. Then, it follows that $\lim_{\ell\to\infty} 1/\|(A - \theta_i^{(\ell)} B)^{-1} C_i^{(\ell)} u_i^{(\ell)}\|_B = 0$. Hence, letting $y^{(\ell)} := C_i^{(\ell)} u_i^{(\ell)}$ in Lemma 1, we obtain the theorem. □

In this proof, the relation (17) found in [8] and Lemma 1 are crucial. Note that, as a consequence of Lemma 1, the theorem is organized similar to Theorem 3.

Our new theorem covers any solver in the right part of Table 1. The rational Krylov method corresponds to the situation $C_i^{(\ell)} := B$ $(i = 1, \ldots, k, \ \ell \in$

$\mathbb{N}$) [23, 35] (see also [4, §8.5]). The Jacobi-Davidson method also corresponds to the situation $k = 1$ and $C_1^{(\ell)} := B$ for all $\ell \in \mathbb{N}$, in which the so-called correction equations are solved exactly [39] (see also [1, §7]). A block variant of Jacobi-Davidson [38, Algorithm 9.3] is also covered because it corresponds to the situation $C_i^{(\ell)} := B$ $(i = 1, \dots, k,\ \ell \in \mathbb{N})$ in the same way as the rational Krylov method.

4.1 Inexact inner iteration and preconditioning

From the practical point of view, $(A - \theta_i^{(\ell)} B)^{-1} C_i^{(\ell)} u_i^{(\ell)}$ for $i = 1, \dots, k$ are often computed using sparse LU factorization (e.g., see [23]). Otherwise, $(A - \theta_i^{(\ell)} B)^{-1} C_i^{(\ell)} u_i^{(\ell)}$ for $i = 1, \dots, k$ are solved approximately using some linear system solvers such as the generalized minimal residual method (GMRES) together with preconditions filtering the target eigenvectors. Theorem 4 covers such a situation as follows. Let the new directions be $t_i^{(\ell)} := P_i^{(\ell)} B u_i^{(\ell)}$ $(i = 1, \dots, k,\ \ell \in \mathbb{N})$, where $P_i^{(\ell)}$ can be regarded as the product of the preconditioning matrix and an approximation of $(A - \theta_i^{(\ell)} B)^{-1}$. Also note that $t_i^{(\ell)}$ can be obtained without explicitly constructing $P_i^{(\ell)}$. Theorem 4 states that, for any $i = 1, \dots, k$, if $C_i^{(\ell)} u_i^{(\ell)} := (A - \theta_i^{(\ell)} B) t_i^{(\ell)}$ satisfies the convergence conditions in (14) for some positive constants $K_1, K_2 > 0$ (i.e., $K_1 \leq u_i^{(\ell)\mathrm{T}} (A - \theta_i^{(\ell)} B) t_i^{(\ell)}$ and $\|(A - \theta_i^{(\ell)} B) t_i^{(\ell)}\|_B \leq K_2$), then the eigenpairs can be obtained. As an example, we choose the new direction $t_i^{(\ell)}$ using the GMRES iterations for solving $B u_i^{(\ell)} = (A - \theta_i^{(\ell)} B) t_i^{(\ell)}$ approximately, for which the number of iterations is determined as $u_i^{(\ell)\mathrm{T}} (A - \theta_i^{(\ell)} B) t_i^{(\ell)}$ and $\|(A - \theta_i^{(\ell)} B) t_i^{(\ell)}\|_B$ are less than K_1 and K_2, respectively. Then, the convergence is guaranteed from Theorem 4. Note that since $u_i^{(\ell)\mathrm{T}} (A - \theta_i^{(\ell)} B) t_i^{(\ell)}$ and $\|(A - \theta_i^{(\ell)} B) t_i^{(\ell)}\|_B$ can be computed during the GMRES iteration, we can stop the GMRES iteration appropriately in order to guarantee the global convergence.

4.2 Extension to extremal eigenvalues

In modern eigensolvers, it is important to use both the smallest and largest Ritz values and the corresponding Ritz vectors for the restart strategies [9, 46, 48]. Hence, we prove global convergence for such strategies, as demonstrated in [1, Lemma 3].

To this end, we describe the Rayleigh-Ritz procedure with restart strategy for computing both the p smallest and the q largest ones.

Algorithm 3 A framework of iterative projection methods with restarting for computing extremal eigenvalues of $Ax = \lambda Bx$.

Input: $A, B \in \mathbb{R}^{n\times n}$ and $V^{(0)} = [v_1^{(0)}, \dots, v_{m_0}^{(0)}] \in \mathbb{R}^{n\times m_0}$

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

  compute $A^{(\ell)} = {V^{(\ell)}}^{\mathrm{T}} A V^{(\ell)}, B^{(\ell)} = {V^{(\ell)}}^{\mathrm{T}} B V^{(\ell)}$

  compute the p smallest eigenvalues for $(A^{(\ell)}, B^{(\ell)})$: $\lambda_1^{(\ell)} \le \dots \le \lambda_p^{(\ell)}$

  compute the q largest eigenvalues for $(A^{(\ell)}, B^{(\ell)})$: $\lambda_{m_\ell-q+1}^{(\ell)} \le \dots \le \lambda_{m_\ell}^{(\ell)}$

  compute the corresponding Ritz vectors $x_1^{(\ell)}, \dots, x_p^{(\ell)}, x_{m_\ell-q+1}^{(\ell)}, \dots, x_{m_\ell}^{(\ell)}$

  compute $V^{(\ell+1)} := [v_1^{(\ell+1)}, \dots, v_{m_{\ell+1}}^{(\ell+1)}]$, where $\mathrm{span}\{V^{(\ell+1)}\} \ni x_i^{(\ell)}$ for $i = 1, \dots, p, m_\ell - q + 1, \dots, m_\ell$

end for

For simplicity, let

$$X^{(\ell)} := [x_1^{(\ell)}, \dots, x_p^{(\ell)}, x_{m_\ell-q+1}^{(\ell)}, \dots, x_{m_\ell}^{(\ell)}]$$

in the following discussion. Similar to Theorem 4, we obtain the following convergence theorem.

Theorem 5. *In Algorithm 3, for $i = 1, \dots, p$, the sequences $\{\lambda_i^{(\ell)}\}_{\ell\in\mathbb{N}}$ are nonincreasing and convergent. and for $i = m_\ell - q + 1, \dots, m_\ell$, the sequences $\{\lambda_i^{(\ell)}\}_{\ell\in\mathbb{N}}$ are nondecreasing and convergent. For each $i = 1, \dots, p, m_\ell - q + 1, \dots, m_\ell$, define $U_{m(i)}^{(\ell)} \in \mathbb{R}^{n\times m(i)}$ normalized to ${U_{m(i)}^{(\ell)}}^{\mathrm{T}} B U_{m(i)}^{(\ell)} = I$ for $m(i) \ge i$, satisfying span$\{X^{(\ell)}\} \subseteq$ span$\{U_{m(i)}^{(\ell)}\} \subseteq$ span$\{V^{(\ell+1)}\}$ for all $\ell \in \mathbb{N}$. Let $u_i^{(\ell)}$ be the Ritz vector corresponding to the i-th smallest Ritz value $\theta_i^{(\ell)}$ for the subspace span$\{U_{m(i)}^{(\ell)}\}$. Moreover, let a set of matrices $\{C_i^{(\ell)}\}$ satisfy the following assumption: for any $i = 1, \dots, p, m_\ell - q + 1, \dots, m_\ell$, there exist $K_1, K_2 > 0$ such that for any $\ell \in \mathbb{N}$, $u_i^{(\ell)}$ satisfies*

$$K_1 \le u_i^{(\ell)\mathrm{T}} C_i^{(\ell)} u_i^{(\ell)}, \ \|C_i^{(\ell)} u_i^{(\ell)}\|_B^2 \le K_2. \tag{19}$$

Furthermore, we assume that, for any $i = 1, \dots, p, m_\ell - q + 1, \dots, m_\ell$, $\ell \in \mathbb{N}$, the vector $(A - \theta_i^{(\ell)} B)^{-1} C_i^{(\ell)} u_i^{(\ell)}$ belongs to span$\{V^{(\ell+1)}\}$. Then, for any $i = 1, \dots, p, m_\ell - q + 1, \dots, m_\ell$, $\lim_{\ell\to\infty} \lambda_i^{(\ell)}$ is an eigenvalue for (A, B), and the accumulation points of $\{x_i^{(\ell)}\}_{\ell\in\mathbb{N}}$ are the corresponding eigenvectors.

Proof. Since the convergence $\lim_{\ell\to\infty} \lambda_i^{(\ell)}$ for any $i = 1, \dots, p, m_\ell - q + 1, \dots, m_\ell$ is obvious, we prove that, for any $i = 1, \dots, p, m_\ell - q + 1, \dots, m_\ell$, $\lim_{\ell\to\infty} \lambda_i^{(\ell)}$ is an eigenvalue of A, and the accumulation points of $\{x_i^{(\ell)}\}_{\ell\in\mathbb{N}}$

are the corresponding eigenvectors. The following proof is almost the same as that of Theorem 4. Letting $w_i^{(\ell)} := (I - U_{m(i)}^{(\ell)} U_{m(i)}^{(\ell)\mathrm{T}} B)(A - \theta_i^{(\ell)} B)^{-1} C_i^{(\ell)} u_i^{(\ell)}$, we see $u_i^{(\ell)\mathrm{T}} A w_i^{(\ell)} = u_i^{(\ell)\mathrm{T}} C_i^{(\ell)} u_i^{(\ell)}$. Define $\alpha_i^{(\ell)}$ in the same manner as in Theorem 4. Then, we have

$$\alpha_i^{(\ell)} = \frac{u_i^{(\ell)\mathrm{T}} A w_i^{(\ell)}}{\|w_i^{(\ell)}\|_B} = \frac{u_i^{(\ell)\mathrm{T}} C_i^{(\ell)} u_i^{(\ell)}}{\|w_i^{(\ell)}\|_B}$$

and $\alpha_i^{(\infty)} = 0$, resulting in $\lim_{\ell\to\infty} 1/\|(A - \theta_i^{(\ell)} B)^{-1} C_i^{(\ell)} u_i^{(\ell)}\|_B = 0$ in the same manner as in the proof of Theorem 4. Noting that Lemma 1 is also established for the largest Ritz values, we obtain the theorem. □

Acknowledgement
This work was supported by JSPS Grant-in-Aid for Young Scientists (Grant Number 25790096).

References

[1] K. Aishima, Global convergence of the restarted Lanczos and Jacobi–Davidson methods for symmetric eigenvalue problems, Numer. Math., DOI: 10.1007/s00211-015-0699-4.

[2] K. Aishima, A note on the Rayleigh quotient iteration for symmetric eigenvalue problems, Jpn. J. Ind. Appl. Math., 31 (2014) 575–581.

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

[4] 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.

[5] C. Beattie, M. Embree, and J. Rossi, Convergence of restarted Krylov subspaces to invariant subspaces, SIAM J. Matrix Anal. Appl., 25 (2004) 1074–1109.

[6] C. Beattie, M. Embree, D. C. Sorensen, Convergence of polynomial restart Krylov methods for eigenvalue computations, SIAM Rev., 47 (2005) 492–515.

[7] B. Beckermann, S. Güttel, R. Vandebril, On the convergence of rational Ritz values, SIAM J. Matrix Anal. Appl., 31 (2009/10), no. 4, 1740–1774.

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

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

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

[11] M. A. Freitag and A. Spence, Convergence of inexact inverse iteration with application to preconditioned iterative solvers, BIT, 47 (2007) 27–44.

[12] G.H. Golub and C.F. Van Loan, Eigenvalue computation in the 20th century, J. Comput. Appl. Math., 123 (2000) 35–65.

[13] G.H. Golub and C.F. Van Loan, Matrix Computations, 4th edition., Johns Hopkins University, 2013

[14] K. Henrik, A. Olsson, and A. Ruhe, Rational Krylov for eigenvalue computation and model order reduction, BIT, 46 (2006) 99–111.

[15] A. Imakura, L. Du, and T. Sakurai A block Arnoldi-type contour integral spectral projection method for solving generalized eigenvalue problems, Appl. Math. Letters, 32 (2014) 22–27.

[16] Z. Jia, On convergence of the inexact Rayleigh quotient iteration with MINRES, J. Comput. Appl. Math., 236 (2012) 4276–4295.

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

[18] A.V. Knyazev and K. Neymeyr, Gradient flow approach to geometric convergence analysis of preconditioned eigensolvers, SIAM J. Matrix Anal. Appl., 31 (2009) 621–628.

[19] 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., 154-156 (1991) 245–257.

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

[21] R. B. Lehoucq and K. Meerbergen, Using generalized Cayley transformations within an inexact rational Krylov sequence method, SIAM J. Matrix Anal. Appl., 20 (1998) 131–148.

[22] T. Mach, M. S. Pranić, and R. Vandebril, Computing approximate (block) rational Krylov subspaces without explicit inversion with extensions to symmetric matrices, Electron. Trans. Numer. Anal., 43 (2014) 100–124.

[23] K. Meerbergen, Changing poles in the rational Lanczos method for the Hermitian eigenvalue problem, Numer. Linear Algebra Appl., 8 (2001) 33–52.

[24] R.B. Morgan, Preconditioning eigenvalues and some comparison of solvers, J. Comput. Appl. Math., 123 (2010) 101–115.

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

[26] K. Neymeyr A geometric theory for preconditioned inverse iteration applied to a subspace, Math. Comp., 71 (2002) 197–216.

[27] K. Neymeyr A geometric convergence theory for the preconditioned steepest descent iteration, SIAM J. Numer. Anal., 50 (2012) 3188–3207.

[28] K. Neymeyr, E. Ovtchinnikov, and M. Zhou, Convergence analysis of gradient iterations for the symmetric eigenvalue problem, SIAM J. Matrix Anal. Appl., 32 (2011) 443–456.

[29] Y. Notay, Convergence analysis of inexact Rayleigh quotient iteration, SIAM J. Matrix Anal. Appl., 24 (2003) 627–644.

[30] Y. Notay, Is Jacobi–Davidson faster than Davidson?, SIAM J. Matrix Anal. Appl., 26 (2005) 522–543.

[31] S. Oliveira, On the convergence rate of a preconditioned subspace eigensolver, Computing, 63 (1999) 219–231.

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

[33] E. Ovtchinnikov, Jacobi correction equation, line search, and conjugate gradients in Hermitian eigenvalue computation I: Computing an extreme eigenvalue, SIAM J. Numer. Anal., 46 (2008) 2567–2592.

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

[35] A. Ruhe, Rational Krylov: A practical algorithm for large sparse nonsymmetric matrix pencils SIAM J. Sci Comput., 19 (1998) 1535–1551.

[36] Y. Saad, Numerical Methods for Large Eigenvalue Problems: Revised Edition, SIAM, 2011.

[37] V. Simoncini, L. Eldén, Inexact Rayleigh quotient-type methods for eigenvalue computations, BIT, 42 (2002) 159–182.

[38] G.L.G. Sleijpen, L. Booten, R. Fokkema, and A. van der Vorst, Jacobi-Davidson type methods for generalized eigenproblems and polynomial eigenproblems, BIT, 36 (1996) 595–633.

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

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

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

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

[43] D. B. Szyld and F. Xue, Efficient preconditioned inner solves for inexact Rayleigh quotient iteration and their connections to the single-vector Jacobi-Davidson method, SIAM J. Matrix Anal. Appl., 32 (2011) 993–1018.

[44] D. B. Szyld and F. Xue, Preconditioned eigensolvers for large-scale nonlinear Hermitian eigenproblems with variational characterizations. I. Conjugate gradient methods, Research Report 14-08-26, Department of Mathematics, Temple University, August 2014. Revised April 2015.

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

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

[47] F. Xue and H. C. Elman, Convergence analysis of iterative solvers in inexact Rayleigh quotient iteration, SIAM J. Matrix Anal. Appl., 31 (2009) 877–899.

[48] 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. 37 (2010) (Article No. 27).