Research Article

A Study on the Convergence Analysis of the Inexact Simplified Jacobi–Davidson Method

Jutao Zhao and Pengfei Guo

1Department of Mathematics, Changzhi University, Changzhi 046011, China
2School of Mathematics and Statistics, Hainan Normal University, Haikou 571158, China

Correspondence should be addressed to Pengfei Guo; 050116@hainnu.edu.cn

Received 2 October 2021; Accepted 20 November 2021; Published 7 December 2021

Abstract

The Jacobi–Davidson iteration method is very efficient in solving Hermitian eigenvalue problems. If the correction equation involved in the Jacobi–Davidson iteration is solved accurately, the simplified Jacobi–Davidson iteration is equivalent to the Rayleigh quotient iteration which achieves cubic convergence rate locally. When the involved linear system is solved by an iteration method, these two methods are also equivalent. In this paper, we present the convergence analysis of the simplified Jacobi–Davidson method and present the estimate of iteration numbers of the inner correction equation. Furthermore, based on the convergence factor, we can see how the accuracy of the inner iteration controls the outer iteration.

1. Introduction

Let $A$ be a sparse and Hermitian matrix. Then, we are supposed to compute the smallest eigenvalue $\lambda$ of $A$ and the associated eigenvector $x$ of $\lambda$ with large, i.e.,

$$Ax = \lambda x, \quad \text{with} \ |x| = 1.$$  \hspace{1cm} (1)

Here and in the following, $\| \cdot \|$ indicates the induced Euclidean norm for either a vector or a matrix. In the literature, there have been many methods developed involving gradient-type methods and subspace methods. The Lanczos, Arnoldi, Davidson, and Jacobi–Davidson methods are classical and effective methods for solving eigenproblem (1), and there have been many convergence results developed for these methods. It should be mentioned that, for the Lanczos and Arnoldi methods, the involved projection subspace should be restricted to the Krylov subspace. For more details about these methods as well as their variants, refer to in [1–7].

The main framework for the subspace methods is generating a sequence of enlarging subspaces $V_k$, which contain more and more information for the desired eigenvalue or eigenvector of matrix $A$. The central problem for this method, which can be accomplished by the Rayleigh–Ritz procedure, is to extract the approximation to the desired eigenvalue or eigenvector from the projection subspace. For more details of the Rayleigh–Ritz procedure, refer to in [4].

As we know, for Hermitian matrices, the exact Rayleigh quotient iteration (RQI) [4] converges cubically. However, an ill-conditioned linear system of equations should be solved exactly which is expensive in each step of the iteration as the approximation is close to the target eigenvalue. The idea replacing the exact solution by a cheaper approximate solution results in an inexact Rayleigh quotient iteration (IRQI) [8, 9]; however, this replacement may destroy the local convergence property of the RQI.

The Jacobi–Davidson (JD) iteration method [6] proposed by Sleijpen and van der Vorst can overcome these difficulties. A so-called correction equation

$$(I - uu^*) (A - \theta I)(I - uu^*)^t = -r,$$  \hspace{1cm} (2)

where $u$ is the current approximation to the desired eigenvector with norm unity, $\theta = u^* Au$ is the Rayleigh quotient, and $r = Au - \theta u$ is the current residual, is solved to expand the projection subspace which contains more and more information of the desired eigenvector $x$ theoretically.
As we know, if (2) is solved exactly, the JD method converges as fast as the RQI method. However, what we are interested in is the iterative solver with an ideal preconditioner for it. Although the shifted matrix $A - \theta I$ becomes ill-conditioned as the approximation $\theta$ is near to the desired eigenvalue, correction equation (2) remains well conditioned, thanks to the projection onto the orthogonal complement of $u$. As Notay in [10] pointed out, with proper implementations, the potential indefiniteness of the coefficient matrix in system (2) cannot spoil the method.

At present, we have not seen the analysis of the connection between the solution of the correction equation and the convergence of the approximate vector $u$ towards the wanted eigenvector in this paper. In this paper, we study the convergence of the simplified JD iteration. Through the convergence factor, we try to analyze how the inner linear system controls the outer iteration. Moreover, we can also see that, under some assumptions, the JD method attains quadratic convergence rate locally when the involved correction equation is solved by a Krylov subspace method and attains cubic convergence rate locally when the involved correction equation is solved by a Krylov subspace method. In the last part of this paper, we give the convergence analysis in terms of the residual norms, from which we can see that these results are asymptotically identical to those derived in the former section; we also gave the analysis on the iteration number of the inner linear system.

We should mention that some analyses for the JD method or, more generally, the Newton method have been established so far; see, e.g., [7, 11–13] and the references therein. In particular, Bai and Miao in [11] presented the convergence of the JD iteration method, and they proved that the JD iteration method attains quadratic convergence locally when the involved correction equation is solved by a Krylov subspace method and attains cubic convergence rate when the correction equation is solved to a prescribed precision proportional to the norm of the current residual vector. In this paper, we will use a completely different technique to demonstrate the convergence of the JD method from another point of view. In addition, we have further studied the connection between the iteration steps of the inner correction equation with the convergence of the outer iteration.

This paper is organized as follows. In Section 2, we give some preliminaries of the JD method. In Section 3, we give some known results and then built several new results concerning the convergence property of the simplified JD method whose involved correction equation is inexactly solved by Krylov solvers. In Section 4, we give the estimate on the iteration number of the inner linear system. Finally, we give some concluding remarks.

2. Preliminaries

Let $\{x_i\}_{i=1}^n$ be the eigenvectors of the Hermitian matrix $A$ associated with the eigenvalues $\{\lambda_i\}_{i=1}^n$ with the ascending order $\lambda_1 < \lambda_2 \leq \ldots \leq \lambda_n$. In the following discussion, we want to compute the eigenvector $x_1$ associated with the simple smallest eigenvalue $\lambda_1$. Denote by $\phi$ the angle between the current approximation $u$ and the wanted eigenvector $x_1$.

We first present the algorithmic description of the simplified JD method in Algorithm 1.

As we know, if the correction equation in Algorithm 1 is solved accurately, the simplified JD method is equivalent to the RQI method. In fact, according to (2), we have $(A - \theta I)u = (u^*(A - \theta I)u)u = (u^*(A - \theta I)u)u$. We can see that the new approximation $u = u + t\|u + l\|$ is the Rayleigh quotient iteration vector. In fact, Theorem 4.2 in [14] tells us that the inexact simplified JD method and the IRQI method can also be equivalent if the inner linear system is solved by Krylov subspace methods. Here, the ‘inexact’ method means that the inner linear system is solved by an iteration method.

Based on the property of the correction equation, the inner linear system in Algorithm 1 is solved approximately by Krylov subspace methods, and we will use the obtained vector to update the current eigenvector approximation.

Suppose that we have obtained an approximate eigenvector $u$ which is close to the wanted eigenvector $x$ in Algorithm 1; we decompose it in the following way:

$$u = a_1 x_1 + a_2 x_2 + \cdots + a_n x_n = a_1 x_1 + \sqrt{1 - a_1^2} \sum_{j=2}^n a_j x_j \sqrt{1 - a_1^2}$$

$$= x_1 \cos \phi + u \sin \phi,$$

(3)

where $u \perp x_1$ with $\|u\| = 1$ and $\phi$ is the angle between vectors $u$ and $x_1$. Also, in this paper, the approximate vector $u$ is close to the wanted eigenvector $x_1$ which means $\theta \sin^2 \phi = |\lambda_1 - \theta| \ll |\lambda_2 - \theta|$.

In the following, we give a lemma which reveals the relation between two Krylov subspaces.

Lemma 2.1. Let $u$ be a normalized vector, $\theta = u^* Au$, and $r = Au - \theta u$. Denote by $\Pi = I - uu^*$ and $A(\theta, u) = \Pi (A - \theta I) \Pi$. Then, for $k \geq 2$, the two Krylov subspaces $V^R_k$ and $V^R_k$ have the following relation:

$$V^R_k \subseteq V^R_k,$$

(4)

where

$$V^R_k = \text{span}\{r, A(\theta, u)r, \ldots, A(\theta, u)^{k-1} r\},$$

(5)

and

$$V^R_k = \text{span}\{u, (A - \theta I)u, \ldots, (A - \theta I)^{k-1} u\}.$$  (6)

Proof. We prove this lemma by induction over $k$. Obviously, for $k = 2$, $V^R_2 = \text{span}\{r\} \subseteq V^R_2 = \text{span}\{u, (A - \theta I)u\}$. Assume that, for all $i < k - 1$, we have $V^R_i \subseteq V^R_i$. Next, we will prove that this relation satisfies for $i = k$. Denote $v = A(\theta, u)^{k-1} r$; based on the fact $u \perp r $ and the induction hypothesis, we have $v \in V^R_{k-1} \subseteq V^R_k$. Then, we have $A(\theta, u)^{k-2} r = A(\theta, u) A(\theta, u)^{k-3} r = A(\theta, u)^{k-3} r = A(\theta, u)^{k-3} r$. We have verified the relation $V^R_{k-1} \subseteq V^R_k$. $\square$
3. Convergence Analysis of the JD Iteration

In the following discussion, we solve the correction equation approximately by Krylov subspace methods, such as CG or MINRES with the initial vector being zero, to obtain an approximate solution to update the new eigenvector approximation \( u \). That is to say, solution \( t \) satisfies the following equation:

\[
(1 - uu^*) (A - \theta I)(1 - uu^*) t = -r + \eta_m,
\]

where \( \eta_m \) is the residual at step \( m \).

In order to analyze conveniently, (7) can also be represented by other equivalent forms such as (30).

As we know, the JD method is one of the “inner-outer” type iterations. In the method of this type, it is essential to know how the inner iteration controls the outer iteration. In other words, we want to know how accuracy or how many steps should be solved for the inner iteration equation to ensure the convergence of the outer iteration. As Sleijpen and van der Vorst pointed out in [6], we cannot answer the question at present. All these problems may be based on the convergence analysis of the algorithm. In the following, we first give some convergence results.

**Lemma 3.1** (see [7]). If Algorithm 1 is applied to seek the smallest simple eigenvalue of the Hermitian matrix \( A \) and we assume that the approximate solution \( t \) satisfies the relation in (30), then for \( \bar{u} = u + t/\|u + t\| \), we asymptotically have

\[
\tan \phi \leq \left( \frac{\|\eta_m\|^2 - |q_{m+1}(\lambda_1 - \theta)\cos \phi|^2}{|q_{m+1}(\lambda_1 - \theta) + r|} \right)^{1/2} + |r| \sin \phi \frac{|\lambda_1 - u^* Aw|}{|\lambda_2 - \theta| \cos \phi} \sin^2 \phi,
\]

where \( \phi \) is the angle between the new approximation \( \bar{u} \) and \( x_1 \) and \( r = u^* (A - \theta I)t_m \).

**Proof.** Suppose that \( u \) is the current approximation to the wanted eigenvector \( x_1 \), and we decompose it in the way of (3). If we solve the correction equation \( (I - uu^*) (A - \theta I)(1 - uu^*) t = -r \) inexacty with zero starting vector by the preconditioned Krylov subspace method, then the approximate solution \( t_m \) of the linear system in step \( m \) is in the Krylov subspace \( \mathcal{V}_m = \text{span} \{ r, A(\theta, u)r, A(\theta, u)^2r, \ldots, A(\theta, u)^{m-1} r \} \), where \( A(\theta, u) \) and \( \mathcal{V}_m \) are defined as Lemma 2.1. Based on Lemma 2.1, we have \( t_m \in \mathcal{V}_m \subseteq \mathcal{V}_{m+1} \) for \( t_m \perp u \). Then, the solution \( t_m = \bar{q}_m (A - \theta I)u \), and we get the new approximation

\[
\bar{u} = u + t_m = u + \bar{q}_m (A - \theta I)u
\]

where \( \bar{q}_m(\lambda) = \bar{q}_m(\lambda) + 1 \).

The residual of the correction equation at step \( m \) is

\[
|\tan \phi| \leq \xi \left| \frac{\lambda_n - \lambda}{\lambda_2 - \lambda} \right| \sin \phi + O(\sin^2 \phi).
\]

**Lemma 3.2** (see [13]). If Algorithm 1 is applied to seek the smallest simple eigenvalue of the Hermitian matrix \( A \) and we assume that the approximate solution \( t \) satisfies the relation in (30), then for \( \bar{u} = u + t/\|u + t\| \), we asymptotically have

\[
|\tan \phi| \leq \xi \left| (A - \theta I) \right| f \|\cos \phi| \sin \phi + O(\sin^2 \phi).
\]

The two convergence results above are analyzed based on the angle between vectors \( u \) and \( x_1 \). Proposition 1 in [15] gives us another result for the convergence analysis based on the metric of \( \|u - x_1\| \). However, all these results cannot answer the question asked above, and we could not answer how the convergence order varies as the inner correction equation is being solved.

**Theorem 3.1.** Let \( (\lambda_1, x_1) \) be a simple eigenpair of the Hermitian matrix \( A \) in Algorithm 1, \( (\lambda_1, x_1) \) be the eigenpair of the shifted matrix \( A - \theta I \) with the ascending order \( |\lambda_1 - \theta| < |\lambda_2 - \theta| \leq \ldots \leq |\lambda_n - \theta| \), \( t_m \) be the approximate solution obtained by using the unpreconditioned Krylov subspace method with zero starting vector, and \( \eta_m = (I - uu^*) (A - \theta I)(I - uu^*) t_m + r = q_{m+1}(A - \theta I) u \) be the associated residual. Then, the following estimate holds:

\[
\eta_m = (I - uu^*) (A - \theta I)(I - uu^*) t_m + r = (A - \theta I) t_m - \tau u + r = (A - \theta I)(u + t_m) - \tau u = (A - \theta I) q_m (A - \theta I) u - \tau u = q_{m+1} (A - \theta I) u,
\]

with \( \tau = u^* (A - \theta I) t_m \) and \( q_{m+1}(\lambda) = \lambda q_m(\lambda) - \tau \).

According to the decomposition in (3), we have

\[
\bar{u} = q_m (A - \theta I) u = q_m (A - \theta I) x_1 \cos \phi + q_m (A - \theta I) w \sin \phi.
\]

Thus, we obtain

\[
\tan \phi = \frac{\| q_m (A - \theta I) w \|}{| q_m (\lambda_1 - \theta) |} \tan \phi.
\]

Note that \( q_{m+1}(\lambda) = \lambda q_m(\lambda) - \tau \); that is, \( q_m(\lambda) = q_{m+1}(\lambda) + \tau/\lambda \); then,
\[ \tan \phi = \frac{\| (A - \theta I)^{-1} (q_{m+1} (A - \theta I) + \tau I) w \|}{|q_{m+1} (1 - \theta) + \tau|} |\lambda_1 - \theta| \tan \phi. \]  

(15)

Based on (3) and (12), we get
\[ \| \eta_m \|^2 = |q_{m+1} (1 - \theta) \cos \phi|^2 + \| q_{m+1} (A - \theta I) w \|^2 \sin^2 \phi. \]  

(16)

Thus, we have
\[ \| q_{m+1} (A - \theta I) w \| = \left( \frac{\| \eta_m \|^2 - |q_{m+1} (1 - \theta) \cos \phi|^2}{\sin \phi} \right)^{1/2}. \]  

(17)

\[ \tan \phi \leq \left( \frac{\| \eta_m \|^2 - |q_{m+1} (1 - \theta) \cos \phi|^2}{|q_{m+1} (1 - \theta) + \tau|} \right)^{1/2} + |\tau| \sin \phi \left( \frac{|\lambda_1 - \theta A u|}{|\lambda_2 - \theta| \cos \phi} \right) \sin^2 \phi. \]  

(20)

In addition, it obviously holds that
\[ \| (A - \theta I)^{-1} (q_{m+1} (A - \theta I) + \tau I) w \| \leq \frac{\| q_{m+1} (A - \theta I) w \| + |\tau|}{|\lambda_2 - \theta|}. \]  

(18)

\[ \lambda_1 - \theta = \lambda_1 - u^* A u \]
\[ = \lambda_1 - (x_1 \cos \phi + w \sin \phi)^* A (x_1 \cos \phi + w \sin \phi) \]
\[ = \lambda_1 - (\lambda_1 \cos \phi + w^* A w \sin^2 \phi) \]
\[ = (\lambda_1 - w^* A w) \sin^2 \phi. \]

Thereby, based on (15), (17), (18), and (19), we have the estimate

Proof. \( \forall x \in \text{span}\{u\} \) with \( \| x \| = 1 \), we decompose \( u \) and \( x \) as \( u = \sum_{i=1}^{n} \alpha_i x_i \) and \( x = \sum_{i=1}^{n} \beta_i x_i \), where \( \sum_{i=1}^{n} |\alpha_i|^2 = \sum_{i=1}^{n} |\beta_i|^2 = 1 \). Then, we have

\[ x^* (A - \theta I)^2 x = \left( \sum_{i=1}^{n} \beta_i x_i \right)^* (A - \theta I)^2 \left( \sum_{i=1}^{n} \beta_i x_i \right) \]
\[ = \left( \sum_{i=1}^{n} \beta_i x_i \right)^* \left( \sum_{i=1}^{n} \beta_i (\lambda_1 - \theta)^2 x_i \right) \]
\[ = \sum_{i=1}^{n} |\beta_i|^2 (\lambda_1 - \theta)^2 \]
\[ \geq (\lambda_2 - \theta)^2 \sum_{i=2}^{n} |\beta_i|^2 + |\beta_1|^2 (\lambda_1 - \theta)^2 \]
\[ = (\lambda_2 - \theta)^2 - |\beta_1|^2 (\lambda_2 - \lambda_1)^2 \]
\[ = (\lambda_2 - \theta)^2 - |\beta_1|^2 (\lambda_2 + \lambda_1 - 2\theta) (\lambda_2 - \lambda_1). \]

(23)

It follows from \( x^* u = 0 \) that \( \sum_{i=1}^{n} \alpha_i \beta_i = 0 \); equivalently, we have \( \alpha_i \beta_i = -\sum_{i=2}^{n} \alpha_i \beta_i \). Moreover, we have

\[ |\beta_1|^2 = \frac{\left( \sum_{i=1}^{n} \alpha_i \beta_i \right)^2}{|\alpha_1|^2} \leq \frac{\left( \sum_{i=2}^{n} |\alpha_i|^2 \right) \left( \sum_{i=2}^{n} |\beta_i|^2 \right)}{|\alpha_1|^2} \leq \frac{1 - |\alpha_1|^2}{|\alpha_1|^2}. \]

(24)

The first assumption in (21) indicates that \( \lambda_2 + \lambda_1 - 2\theta < \lambda_2 - \lambda_1 \) and \( \lambda_2 - \theta > \lambda_2 - \lambda_1/2 \) hold; then, we have the estimate

\[ |\beta_1|^2 \leq \frac{1 - |\alpha_1|^2}{|\alpha_1|^2}. \]

(23)
Here, we use the fact $\alpha_i = \cos \phi$. Combining the above estimate with the second assumption in (21), we frequently obtain the estimate in (22).

Thus, combining (26) and the correct equation in (7), we have

$$\frac{\| -r + \eta_m \|}{\|t_m\|} \geq \sigma_{\text{min}}^i (A - \theta I).$$

(27)

It further indicates that

$$\|t_m\| \leq \frac{\| -r + \eta_m \|}{\sigma_{\text{min}}^i (A - \theta I)} \leq \frac{\| r \| + \| \eta_m \|}{\sigma_{\text{min}}^i (A - \theta I)}$$

(28)

and, at last, we obtain

$$|r| = |u^* (A - \theta I) t_m| \leq \| r \| \| t_m \| \leq \frac{\| r \| + \| \eta_m \|}{\sigma_{\text{min}}^i (A - \theta I)} \| r \|.$$

(29)

To see the behavior of the convergence for the JD method clearly, the stopping criterion we adopt is that the norm of the current residual is reduced by a factor $\xi$ from that of the initial residual. That is, $t$ satisfies the following equation:

$$(I - uu^*) (A - \theta I) (I - uu^*) t = -r + \xi \| r \| d,$$

(30)

where $d$ is the residual direction and $\xi$ is the stopping factor.

In the following, we give an estimate of $|r|$ with $r = u^* (A - \theta I) t_m$.

Theorem 3.2. Given $\xi_1, \xi_2 < 1$, if $\| \eta_m \| \leq \xi_1 \| r \|$, then the JD method converges linearly as follows:

$$\frac{\tan \phi}{C_1} \leq \frac{\| r \|}{\| r \| \| \eta_m \| \| r \|} \leq \frac{\| r \| + \| \eta_m \|}{\sigma_{\text{min}}^i (A - \theta I)} \| r \|.$$

(31)

under the assumption

$$\min_{\| r \| \leq \xi_2 \| r \|} \lambda_j - \theta q_{m+1} (\lambda_j - \theta) > \lambda_1 - \theta (\lambda_n + 2 \lambda_1) (\lambda_n - \lambda_1) \frac{1 + \xi_2}{\lambda_1 - \theta} \| r \| \| \eta_m \| \| r \|,$$

(32)

If $\| \eta_m \| \leq \xi_2 \| r \|$, then the JD method converges quadratically as follows:

$$\tan \phi \leq \frac{\alpha^2 (\xi_2 + (1 + \alpha \xi_1 \sin \phi) \| \eta_m \| \sin \phi)}{C_2} \frac{\| r \| + \| \eta_m \|}{\sigma_{\text{min}}^i (A - \theta I)} \| r \|.$$

(33)

under the assumption

$$\frac{\eta_m \perp \mathcal{Y}^I}{\eta_m \perp \mathcal{Y}^I} = \text{span}\{r, A(\theta, u)r, \ldots, A(\theta, u)^{m-1} r\}.$$

(37)

Specially, we have $\eta_m \perp r$; combining with the factorization in (3), we have

$$r^* \eta_m = u^* (A - \theta I) q_{m+1} (\lambda_1 - \theta) u
\begin{align*}
&= (x_1 \cos \phi + w \sin \phi)^* (A - \theta I) q_{m+1} \\
&\quad \cdot (A - \theta I) (x_1 \cos \phi + w \sin \phi) \\
&= (\lambda_1 - \theta) q_{m+1} (\lambda_1 - \theta) \cos^2 \phi + w^* \\
&\quad \cdot (A - \theta I) q_{m+1} (A - \theta I) w \sin^2 \phi \\
&= 0.
\end{align*}$$

(38)

Proof. According to the properties of the Krylov subspace method, e.g., the conjugate gradient method, we have

$$x^* (A - \theta I)^2 x \geq \frac{(\lambda_2 - \lambda_1)^2}{2} |\tan \phi|^2 (\lambda_2 - \lambda_1)^2.$$

(25)
Then, we further have
\[
|q_{m+1}(\lambda_1 - \theta)| = \frac{|w^*(A - \theta I)q_{m+1}(A - \theta)w|}{|\lambda_1 - \theta|\cos^2 \phi} \sin^2 \phi
\]
which implies that
\[
|r| \leq \alpha \sin \phi. \tag{41}
\]
If \(\|q_m\| \leq \xi_m\|r\|\), according to the estimate of \(|r|\), we get
\[
|r| \leq \frac{1 + \xi}{\sigma_{\min^+}} \|r\| \leq \alpha^2 \frac{1 + \xi}{\sigma_{\min^+}} \sin^2 \phi. \tag{42}
\]
Thereby, we obtain
\[
\|r\|^2 = (\lambda_1 - \omega^*Aw)^2 \sin^4 \phi \cos^2 \phi + \|(A - \theta I)w\|^2 \sin^2 \phi,
\]
\[
|q_{m+1}(\lambda_1 - \theta)| \geq \frac{\min_{\lambda \in \mathbb{R}}\left|\frac{\lambda_1 - \theta}{\lambda_1 - \lambda}\right|}{|\lambda_1 - \theta|\cos^2 \phi} \sin^2 \phi - \frac{\alpha^2}{\sigma_{\min^+}} \frac{1 + \xi}{\sigma_{\min^+}} \sin^2 \phi
\]
which implies that
\[
|r| \leq \alpha \sin \phi. \tag{41}
\]
If \(\|q_m\| \leq \xi_m\|r\|\), according to the estimate of \(|r|\), we get
\[
|r| \leq \frac{1 + \xi}{\sigma_{\min^+}} \|r\| \leq \alpha^2 \frac{1 + \xi}{\sigma_{\min^+}} \sin^2 \phi. \tag{42}
\]
Thereby, we obtain
\[
\|r\|^2 = (\lambda_1 - \omega^*Aw)^2 \sin^4 \phi \cos^2 \phi + \|(A - \theta I)w\|^2 \sin^2 \phi,
\]
\[
|q_{m+1}(\lambda_1 - \theta)| \geq \frac{\min_{\lambda \in \mathbb{R}}\left|\frac{\lambda_1 - \theta}{\lambda_1 - \lambda}\right|}{|\lambda_1 - \theta|\cos^2 \phi} \sin^2 \phi - \frac{\alpha^2}{\sigma_{\min^+}} \frac{1 + \xi}{\sigma_{\min^+}} \sin^2 \phi
\]
which implies that
\[
|r| \leq \alpha \sin \phi. \tag{41}
\]
If \(\|q_m\| \leq \xi_m\|r\|\), according to the estimate of \(|r|\), we get
\[
|r| \leq \frac{1 + \xi}{\sigma_{\min^+}} \|r\| \leq \alpha^2 \frac{1 + \xi}{\sigma_{\min^+}} \sin^2 \phi. \tag{42}
\]
Thereby, we obtain
\[
\|r\|^2 = (\lambda_1 - \omega^*Aw)^2 \sin^4 \phi \cos^2 \phi + \|(A - \theta I)w\|^2 \sin^2 \phi,
\]
\[
|q_{m+1}(\lambda_1 - \theta)| \geq \frac{\min_{\lambda \in \mathbb{R}}\left|\frac{\lambda_1 - \theta}{\lambda_1 - \lambda}\right|}{|\lambda_1 - \theta|\cos^2 \phi} \sin^2 \phi - \frac{\alpha^2}{\sigma_{\min^+}} \frac{1 + \xi}{\sigma_{\min^+}} \sin^2 \phi
\]
which implies that
\[
|r| \leq \alpha \sin \phi. \tag{41}
\]
If \(\|q_m\| \leq \xi_m\|r\|\), according to the estimate of \(|r|\), we get
\[
|r| \leq \frac{1 + \xi}{\sigma_{\min^+}} \|r\| \leq \alpha^2 \frac{1 + \xi}{\sigma_{\min^+}} \sin^2 \phi. \tag{42}
\]
Thereby, we obtain
\[
\|r\|^2 = (\lambda_1 - \omega^*Aw)^2 \sin^4 \phi \cos^2 \phi + \|(A - \theta I)w\|^2 \sin^2 \phi,
\]
\[
|q_{m+1}(\lambda_1 - \theta)| \geq \frac{\min_{\lambda \in \mathbb{R}}\left|\frac{\lambda_1 - \theta}{\lambda_1 - \lambda}\right|}{|\lambda_1 - \theta|\cos^2 \phi} \sin^2 \phi - \frac{\alpha^2}{\sigma_{\min^+}} \frac{1 + \xi}{\sigma_{\min^+}} \sin^2 \phi
\]
Thus, combining with the estimate in (10), we have the following estimate:

\[
\tan \phi \leq \frac{\xi_1 \|r\| + |r| \sin \phi}{\frac{|\lambda_1 - w^* A w|}{C_1 \sin^2 \phi} \left| \frac{|\lambda_1 - w^* A w|}{C_1 \sin^2 \phi} \right| \sin^2 \phi \leq \frac{\xi_1 |r| + |r|}{\frac{|\lambda_1 - w^* A w|}{C_1 \sin^2 \phi} \left| \frac{|\lambda_1 - w^* A w|}{C_1 \sin^2 \phi} \right| \sin^2 \phi}. \quad (45)
\]

Similar to the above proof, if \(\|q_m\| \leq \xi_2 \|r\|^2\), according to the estimate of \(|r|\), we get

\[
|r| \leq \frac{1 + \xi_2 \|r\|^2}{\sigma_{\min}^2} \leq \frac{1}{\sigma_{\min}^2} \frac{1 + \alpha \xi_2 \sin \phi}{\sin^2 \phi}. \quad (46)
\]

Note that

\[
1 + \alpha \xi_2 \sin \phi = 1 + \xi_2 \left( |\lambda_1 - \theta| |\lambda_1 - w^* A w| + \|A - \theta I\|\|\| w \|^2 \right) \sin \phi
\]

\[
\leq 1 + \xi_2 \left( |\lambda_1 - \theta| (\lambda_n - \lambda_1) + (\lambda_n - \theta)^2 \right) \sin \phi \quad (47)
\]

\[
\leq 1 + \xi_2 \left( (\theta - \lambda_1) (\lambda_n - \lambda_1) + (\lambda_n - \theta) (\lambda_n - \lambda_1) \right) \sin \phi
\]

\[
= 1 + \xi_2 (\lambda_n - \lambda_1) \sin \phi,
\]

and combining with the assumption in (33), we obtain

\[
|q_{m+1}(\lambda_1 - \theta)| - |r| \geq \frac{\min \left( \frac{|(\lambda_j - \theta) q_{m+1}(\lambda_j - \theta)|}{|\lambda_j - \theta| \cos^2 \phi} \sin^2 \phi - \alpha^2 \frac{1 + \alpha \xi_2 \sin \phi}{\sigma_{\min}^2} \sin^2 \phi \right)}{\frac{|\lambda_1 - \theta| \cos^2 \phi}{\sin^2 \phi} - \alpha^2 \frac{1 + \alpha \xi_2 \sin \phi}{\sigma_{\min}^2} \sin^2 \phi}
\]

\[
= \left( \frac{\min \left( |(\lambda_1 - \theta) q_{m+1}(\lambda_j - \theta)| / |\lambda_1 - \theta| \cos^2 \phi \right) - \alpha^2 \frac{1 + \alpha \xi_2 \sin \phi}{\sigma_{\min}^2} \sin^2 \phi}{\sin^2 \phi} \frac{1 + \xi_2 (\lambda_n - \lambda_1) \sin \phi}{\sigma_{\min}^2} \sin^2 \phi \right)
\]

\[
\geq \left( \frac{\lambda_n - \lambda_1}{\lambda_2 - \lambda_1 + 2(\theta - \lambda_1)} \right) \frac{1 + \xi_2 (\lambda_n - \lambda_1) \sin \phi}{\sigma_{\min}^2} \sin^2 \phi.
\quad (48)
\]

Utilizing the estimate in (10) again, we have the following estimate:

\[
\tan \phi \leq \frac{\xi_2 \|r\|^2 + |r| \sin \phi}{\frac{|\lambda_1 - w^* A w|}{C_2 \sin^2 \phi} \left| \frac{|\lambda_1 - w^* A w|}{C_2 \sin^2 \phi} \right| \sin^2 \phi \leq \frac{\alpha^2 (\xi_2 + 1 + \alpha \xi_2 \sin \phi / \sigma_{\min}^2 \sin \phi) \left| \frac{|\lambda_1 - w^* A w|}{C_2 \sin^2 \phi} \right| \sin^2 \phi}{\left| \frac{|\lambda_1 - w^* A w|}{C_2 \sin^2 \phi} \right| \sin^2 \phi}. \quad (49)
\]

Note that the assumptions in (32) and (34) will easily be satisfied if \(\min_{2 \leq j \leq n} |\lambda_j - \theta| q_{m+1}(\lambda_j - \theta)|\) is not very small because the right terms of the two assumptions have the factor \(|\lambda_1 - \theta| = O(\sin^2 \phi)\), which would be small if the current approximate eigenvector is near to the desired one.

From Theorem 3.1, we cannot fully understand the convergence of the JD method because it just gives us a preliminary convergence analysis and includes some unknown factors to be explored. Thus, we further explored these unknown factors in Theorem 3.2 and established the
Algorithm 1: The simplified Jacobi-Davidson iteration.

Given a normalized vector \( \mathbf{u} \) arbitrary and a stopping factor \( \epsilon \)

For \( m = 1, 2, \ldots, \) do

1. Compute \( \vartheta = \mathbf{u}^* \mathbf{A} \mathbf{u} \)
2. Compute \( \tau = \mathbf{A} \mathbf{u} - \vartheta \mathbf{u} \)
3. Test the convergence; if \( ||\tau|| \leq \epsilon \), stop
4. Solve the correction equation \( (I - \mathbf{u} \mathbf{u}^*) (A - \theta I) (I - \mathbf{u} \mathbf{u}^*) \mathbf{t} = -\mathbf{r} \) for \( \mathbf{t} \perp \mathbf{u} \)
5. Let \( \mathbf{u}' = (\mathbf{u} + \mathbf{t}) / ||\mathbf{u} + \mathbf{t}|| \)

End for

The convergence of the JD method, from which we can see clearly how the inner correction equation controls the convergence of the outer iteration.

In addition, from Theorem 3.2, we can see that the JD method converges linearly if the accuracy of the correction equation is roughly \( \Theta (||\tau||) \) and converges quadratically with the accuracy of the correction equation being \( \Theta (||\tau||^2) \). Moreover, observing the convergence factor, we can see that the method can gain cubic convergence rate ideally.

4. Estimate for the Iteration Number

In this section, we first give the bounds of the residual norms of the JD iteration method. Through this bound, we may analyze the relation between the outer iteration and the inner iteration; more clearly speaking, we can see how the inner iteration controls the convergence property of the outer iteration.

Theorem 4.1. Let \((\theta, \mathbf{u})\) be the approximate eigenpair obtained by Algorithm (1) with \( ||\mathbf{u}|| = 1 \), \( \vartheta = \mathbf{u}^* \mathbf{A} \mathbf{u} \), and the residual \( \mathbf{r} = \mathbf{A} \mathbf{u} - \vartheta \mathbf{u} \). If we solve the JD correction equation by the Krylov subspace method with the zero initial vector, then we get the following estimate:

\[
\|\mathbf{r}\| \leq \frac{1}{\|\mathbf{r}\|} \left( \frac{1 + \xi_m}{\sigma_{\min}^+} \right) \|\mathbf{r}\| + \xi_m \|\mathbf{r}\|, \quad (50)
\]

where \( \sigma_{\min} = \sigma_{\min}^+ (A - \theta I) \) and \( \mathbf{u}' = q_m (A - \theta I) \mathbf{u} \) is the new approximation defined as (11).

Proof. Using the minimal residual property of Ritz values (Fact 1.9 in [4]), we have

\[
\|\mathbf{r}\| = \left\| \mathbf{A} \frac{\mathbf{u}'}{\|\mathbf{u}'\|} - \vartheta \frac{\mathbf{u}'}{\|\mathbf{u}'\|} \right\| \leq \frac{1}{\|\mathbf{u}'\|} \left\| (A - \theta I) \mathbf{u}' \right\|. \quad (51)
\]

Based on the relation in (12) of the residual equation, we have

\[
(A - \theta I) q_m (A - \theta I) u = \tau u + \eta_m \text{ with } \tau = \mathbf{u}^* (A - \theta I) t_m. \quad (52)
\]

By making use of the estimate of \( ||\mathbf{r}|| \), we get

\[
\|\mathbf{r}\| \leq \frac{1}{\|\mathbf{r}\|} \left( \|\tau\| + \|\eta_m\| \right) \leq \frac{1}{\|\mathbf{r}\|} \left( \frac{\|\tau\| + \|\eta_m\|}{\sigma_{\min}^+} \|\mathbf{r}\| + \|\eta_m\| \right) = \frac{1}{\|\mathbf{r}\|} \left( \frac{1 + \xi_m}{\sigma_{\min}^+} + \xi_m \right) \|\mathbf{r}\|. \quad (53)
\]

From the above theorem, we can choose a considerate stopping factor \( \xi \) to obtain a higher convergence rate as follows.

Corollary 4.1. Under the assumption of Theorem 4.1, if the stopping factor \( \xi \) satisfies \( \xi \leq c \|\mathbf{r}\| \) (equivalent to \( \eta_m \leq c \|\mathbf{r}\|^2 \)) with \( c \leq 1 \), then the following relation holds:

\[
\|\mathbf{r}\| \leq \frac{1}{\|\mathbf{r}\|} \left( \frac{1 + \xi_m}{\sigma_{\min}^+} + c \right) \|\mathbf{r}\|^2. \quad (54)
\]

Next, we give a rough estimate for \( \|\mathbf{u}'\| \), based on \( \mathbf{u}' = q_m (A - \theta I) \mathbf{u} \) and \( q_{m+1} (\lambda) = \lambda q_m (\lambda) - \tau \), we have

\[
\begin{align*}
\mathbf{u}' (A - \theta I)^{-1} (q_{m+1} (A - \theta I) + \tau) & = (\lambda_1 - \theta)^{-1} (q_{m+1} (\lambda_1 - \theta) + \tau) x_1 \cos \phi + (A - \theta I)^{-1} (q_{m+1} (A - \theta I) + \tau) w \sin \phi, \\
\|\mathbf{u}'\| & \geq \frac{|q_{m+1} (\lambda_1 - \theta) + \tau|}{|\lambda_1 - \theta|} \cos \phi. \quad (55)
\end{align*}
\]

which further indicates that
We can see that, under some assumptions, e.g., $|q_{m+1}|(\lambda_1 - \theta) > 0$ and some assumptions on $|\theta| > 0$ and $\sin^2 \phi$, the asymptotical convergence properties of Theorem 4.1 and Corollary 4.1 are identical to those of Theorem 3.2.

Theorem 4.1 provides us with an asymptotical description of the convergence property from the point of view of the residual norm. We can see that, at the starting iterate, i.e., $\|x\|$ is not very small relative to the stopping precision, the stopping factor $\xi$ need not be very small; we can obtain a fast convergence rate; that is to say, it only needs a few iterate numbers for the inner iteration. As the approximate eigenvector $u$ is near to the desired eigenvector $x_*$, we need a relative high accuracy for the inner iteration to gain an ideal convergence rate.

In theory, we can give an estimate whether the JD method converges or not through the factor $q = 1/\|r\| ((|\xi_m|\|r\|)/\sigma_{\min} + \xi_m)$. From the above factor, we can see that it is $\|r\|$ and the inner residual factor $\xi_m$ which determine whether the outer residual norm decreases or not. So, we can use $q < 1$ as the inner iteration stopping criterion. Perhaps, we can also estimate the convergence property through judging the norm of the new approximation $\|u\|$.

In the following discussion, let us give an estimate on the iteration number roughly by using the conjugate gradient method to solve the JD correction equation.

First, we illustrate that it is reasonable to solve the correction equation by using the conjugate gradient method as the approximation $u$ is close to the desired vector $x_*$.

**Lemma 4.1.** For any $v \in \text{span}(u)^\perp$, the following inequality satisfies

\[
(f(u,v),v) \geq (\lambda_1 + \lambda_2 - \mu - \lambda(u))\|v\|^2,
\]

where $f(u,v) = (I - uu^*)(A - \mu I)(I - uu^*)$.

From the above lemma, we can see that if $\theta < (\lambda_1 + \lambda_2)/2$, which is satisfied under the assumption $|\lambda_1 - \theta| < |\lambda_2 - \theta|$, $f(\theta,u)$ is positive definite in the subspace span$(u)^\perp$.

Next, we give a simple convergence property of the conjugate gradient method for the linear equation $Ax = b$ with $A$ being positive definite.

**Lemma 4.2.** Solving the symmetric positive definite linear system of equations $Ax = b$ by applying the conjugate gradient algorithm, the following estimate holds for the residual norm:

\[
\|\eta_m\| \leq 2k^2 \left(\frac{\sqrt{k - 1}}{\sqrt{k + 1}}\right)^m \|\eta_0\|,
\]

where $x_m$ is the approximate solution obtained at the $m$th step of the conjugate gradient algorithm, $\eta_m = Ax_m - b$ is the corresponding residual, and $k = \|A\|\|A^{-1}\|$ is the condition number of $A$.

**Proof.** According to Theorem 6.6 in [5], we have

\[
\|x_m - x_*\| \leq 2\left(\frac{\sqrt{k - 1}}{\sqrt{k + 1}}\right)^m \|x_0 - x_*\| A, (59)
\]

where $x_*$ is the exact solution and $\|x\|_A$ is the $A$-norm defined as $\|x\|_A = \sqrt{x^T A x}$.

Based on the minimax theorem, we know that

\[
\lambda_{\min}(A)\|x\|^2 \leq \|x\|_A^2 = x^T A x \leq \lambda_{\max}(A)\|x\|^2,
\]

and by straightforward computations, we frequently obtain

\[
\|x_m - x_*\| \leq 2\sqrt{k} \left(\frac{\sqrt{k - 1}}{\sqrt{k + 1}}\right)^m \|x_0 - x_*\|.
\]

Thus, the residual norm has the following estimate:

\[
\|\eta_m\| = \|Ax_m - b\| \\
\leq \|A\|\|x_m - x_*\| \\
\leq 2\|A\|\|A\|^{-1} \|x\|_A \|x_0 - x_*\| \\
\leq 2\|A\|\|A\|^{-1} \|x\|_A \|x_0 - x_*\| \\
\leq 2\|A\|\|A\|^{-1} \|x\|_A \|x_0 - x_*\| \\
= 2k^2 \left(\frac{\sqrt{k - 1}}{\sqrt{k + 1}}\right)^m \|\eta_0\|,
\]

which completed the proof of this theorem.

Next, we will combine Theorem 4.1 with Lemma 4.2 to give an estimate on the iteration number of the correction equation roughly.

In fact, the coefficient operator $(I - uu^*)(A - \theta I)(I - uu^*)$ of the correction equation can be seen as the operator $A - \theta I$ restricted to the subspace span$(u)^\perp$, which is denoted by $(A - \theta I)^\perp$.

Assume that, at the current outer iterate, all approximations $\{\bar{u}_i\}$ obtained by the conjugate gradient method satisfy $\mu_1 \leq \|\bar{u}_i\| \leq \mu_2$ for $i = 1, 2, \ldots, m$, where $m$ is the maximum iteration number of the correction equation.

**Theorem 4.2.** In the JD method, solving the correction equation by applying the conjugate gradient method with the zero initial vector, the iteration number $m$ of the correction equation should satisfy

\[
m \geq \frac{\ln 2k_s^{\sqrt{2}} + \ln |r| + \sigma_{\min}^1 / \mu_s \sigma_{\min}^1 - \|r\|}{\ln \sqrt{k_s^{-1} + 1/\sqrt{k_s^{-1} - 1}}},
\]

where $k_s$ is the condition number of the matrix $(A - \theta I)^\perp$ to ensure the convergence of the algorithm.

**Proof.** By assumption, the convergence factor $q$ in Theorem 4.1 satisfies
\[
q = \frac{1}{\|a\|} \left( \frac{(1 + \xi_m)\|r\|}{\sigma_{\text{min}}^+ + \xi_m} \right) \leq \frac{1}{\mu_1} \left( \frac{(1 + \xi_m)\|r\|}{\sigma_{\text{min}}^+ + \xi_m} \right). 
\]

(64)

Thus, if \( 1/\mu_1 \left( (1 + \xi_m)\|r\|/\sigma_{\text{min}}^+ + \xi_m \right) < 1 \), that is,

\[
\xi_m \leq \frac{\mu_1 \sigma_{\text{min}}^+ - \|r\|}{\|r\| + \sigma_{\text{min}}^+},
\]

(65)

the JD method converges. It indicates that the residual norm of the correction equation \( \|\text{res}^m\| \) satisfies

\[
\|\text{res}^m\| \leq \frac{\mu_1 \sigma_{\text{min}}^+ - \|r\|}{\|r\| + \sigma_{\text{min}}^+}. 
\]

(66)

Based on Lemma 4.2, we know that if \( 2k^{3/2}_n \left( \sqrt{k_n - 1} / \sqrt{k_n + 1} \right)^m \|r\| \leq \mu_1 \sigma_{\text{min}}^+ - \|r\|/\|r\| + \sigma_{\text{min}}^+ \|r\| \), the JD method converges. By straightforward computations, we frequently get

\[
m \geq \frac{\ln 2k^{3/2}_n + \ln \|r\| + \sigma_{\text{min}}^+ / \mu_1 \sigma_{\text{min}}^+ - \|r\|}{\ln \sqrt{k_n - 1} / \sqrt{k_n + 1}}. \quad \Box
\]

5. Concluding Remarks

We have proved that the inexact simplified Jacobi–Davidson iteration method for Hermitian eigenvalue problems can attain cubic convergence rate locally, and it is asymptotically convergent as fast as the Rayleigh quotient iteration. Thus, both exact and inexact simplified Jacobi–Davidson methods are competitive with the exact and inexact Rayleigh quotient iterations. Moreover, we give an estimate of iteration numbers of the inner correction equation. Based on these theoretical results, we can see clearly how the accuracy of the inner correction equation controls the outer iteration.

Data Availability

No data were used to support this study.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

The authors are grateful to Dr. Cunqiang Miao for his helpful discussions. This work was supported by the National Science Foundation of China (nos. 1206030 and 11661031) and Key Project of Teaching Reform in Shanxi Province of China (no. J2015111).

References


