Journal of Computational Mathematics
Vol.36, No.5, 2018, 739–760.

http://www.global-sci.org/jcm
doi:10.4208/jcm.1703-m2016-0580

CONVERGENCE ANALYSIS OF A LOCALLY ACCELERATED PRECONDITIONED STEEPEST DESCENT METHOD FOR HERMITIAN-DEFINITE GENERALIZED EIGENVALUE PROBLEMS*

Yunfeng Cai
LMAM & School of Mathematical Sciences, Peking University, Beijing, 100871, China
Email: yfcai@math.pku.edu.cn
Zhaojun Bai
Department of Computer Science and Department of Mathematics, University of California, Davis CA 95616, USA
Email : bai@cs.ucdavis.edu
John E. Pask
Physics Division, Lawrence Livermore National Laboratory, Livermore, CA 94550, USA
Email: pask1@llnl.gov
N. Sukumar
Department of Civil and Environmental Engineering, University of California, Davis, CA 95616, USA
Email: nsukumar@ucdavis.edu

Abstract

By extending the classical analysis techniques due to Samokish, Faddeev and Faddeeva, and Longsine and McCormick among others, we prove the convergence of the preconditioned steepest descent with implicit deflation (PSD-id) method for solving Hermitian-definite generalized eigenvalue problems. Furthermore, we derive a nonasymptotic estimate of the rate of convergence of the PSD-id method. We show that with a proper choice of the shift, the indefinite shift-and-invert preconditioner is a locally accelerated preconditioner, and is asymptotically optimal that leads to superlinear convergence. Numerical examples are presented to verify the theoretical results on the convergence behavior of the PSD-id method for solving ill-conditioned Hermitian-definite generalized eigenvalue problems arising from electronic structure calculations. While rigorous and full-scale convergence proofs of preconditioned block steepest descent methods in practical use still largely eludes us, we believe the theoretical results presented in this paper shed light on an improved understanding of the convergence behavior of these block methods.

Mathematics subject classification: 65F08, 65F15, 65Z05, 15A12.
Key words: Eigenvalue problem, Steepest descent method, Preconditioning, Superlinear convergence.

1. Introduction

We consider the Hermitian-definite generalized eigenvalue problem

$$Hu = \lambda Su, \tag{1.1}$$

where H and S are n-by-n Hermitian matrices and S is positive-definite. The scalar λ and nonzero vector u satisfying (1.1) are called *eigenvalue* and *eigenvector*, respectively. The pair

* Received March 21, 2016 / Revised version received November 28, 2016 / Accepted March 27, 2017 / Published online June 22, 2018 /

(λ, u) is called an *eigenpair*. All eigenvalues of (1.1) are known to be real. Our task is to compute few smallest eigenvalues and the corresponding eigenvectors. We are particularly interested in solving the eigenvalue problem (1.1), where the matrices H and S are large and sparse, and there is *no obvious* gap between the eigenvalues of interest and the rest. Furthermore, S is nearly singular and H and S share a near-nullspace. It is called an ill-conditioned generalized eigenvalue problem in [5], a term we will adopt in this paper. The ill-conditioned generalized eigenvalue problem is considered to be an extremely challenging problem.[1)]

Beside examples such as those cited in [5], the ill-conditioned eigenvalue problem (1.1) arises from the discretization of enriched Galerkin methods. The partition-of-unity finite element (PUFE) method [14], which falls within the class of enriched Galerkin methods, is a promising approach in quantum-mechanical materials calculations, see [3] and references therein. In the PUFE method, physics-based basis functions are added to the classical finite element (polynomial basis) approximation, which affords the method improved accuracy at reduced costs versus existing techniques. However, due to near linear-dependence between the polynomial and enriched basis functions, the system matrices that stem from such methods are ill-conditioned, and share a large common near-nullspace. Furthermore, there is in general no clear gap between the eigenvalues that will be sought and the rest. Another example of the ill-conditioned eigenvalue problem (1.1) arises from modeling protein dynamics using normal-mode analysis [2, 10, 11, 17].

In this paper, we focus on a preconditioned steepest descent with implicit deflation method, PSD-id method in short, to solve the eigenvalue problems (1.1). The basic idea of the PSD-id method is simple. Denote all the eigenpairs of (1.1) by (λ_1, u_1), (λ_2, u_2), $\ldots$, (λ_n, u_n), and the eigenvalue and eigenvector matrices by $\Lambda = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$ and $U = [u_1\ u_2\ \cdots\ u_n]$, respectively. Assume that the eigenvalues $\{\lambda_i\}$ are in an ascending order $\lambda_1 \le \lambda_2 \le \cdots \le \lambda_n$. The following variational principles are well-known, see [27, p.99] for example:

$$\lambda_i = \min_{U_{i-1}^H S z = 0} \rho(z) \quad \text{and} \quad u_i = \operatorname*{argmin}_{U_{i-1}^H S z = 0} \rho(z), \tag{1.2}$$

where $U_{i-1} = [u_1\ u_2\ \cdots\ u_{i-1}]$ and $\rho(z)$ is the Rayleigh quotient

$$\rho(z) = \frac{z^H H z}{z^H S z}. \tag{1.3}$$

On assuming that U_{i-1} is known, one can find the ith eigenpair by minimizing the Rayleigh quotient $\rho(z)$ with z being S-orthogonal against U_{i-1} under the algorithmic framework of the preconditioned steepest descent minimization.

The idea of computing the algebraically largest eigenvalue and its corresponding eigenvector of (1.1) (with $B = I$) using the steepest descent (SD) method dates back to early 1950s [7] and [4, Chap.7]. In [13], block steepest descent (BSD) methods are proposed to compute several eigenpairs simultaneously. The preconditioned steepest descent (PSD) method was introduced around late 1950s [24, 25]. The block PSD (BPSD) methods have appeared in the literature, see [1, 16] and references therein. Like the PSD method, the PSD-id method studied in this paper computes one eigenpair at a time. To compute the ith eigenpair, the search subspace of PSD-id is *implicitly* orthogonalized against the previously computed $i-1$ eigenvectors. The preconditioner at each iteration of PSD-id is flexible (i.e., could change at every iteration) and can be indefinite, instead of being fixed and positive definite as in [1, 16, 25].

[1)] W. Kahan, Refining the general symmetric definite eigenproblem, poster presentation at Householder Symposium XVIII 2011, available http://www.cs.berkeley.edu/∼wkahan/HHXVIII.pdf

Over the past six decades, there has been significant work on the convergence analysis of the SD, PSD and BPSD methods. The convergence of the SD method to compute a single eigenpair is presented in [4, Chap.7]. For the BSD method, the convergence of the first eigenpair is presented in [13] and "ordered convergence" for multiple eigenpairs is declared. The (nonasymptotic) rate of convergence of the PSD method is first studied in [25], which later is proven to be sharp [20]. A comprehensive review of the convergence estimates of the PSD method, is presented in [1]. The theoretical proofs of the convergence of the BPSD method have still largely eluded us, we refer the readers to [1,19] and two recent papers [15,16]. In this paper, we present two main results (Theorems 3.1 and 3.2) on the convergence and nonasymptotic rate of convergence of the PSD-id method. These results extend the classical ones due to Faddeev and Faddeeva [4, sec.74] and Samokish [25] for the SD and PSD methods. Specifically, we have included the implicit deflation and preconditioning in the classical convergence analysis of the SD method presented in [4, sec.74], and does not require the positive definiteness of the preconditioners as in [25]. For practical applications, we show that with a proper choice of the shift, the well-known indefinite shift-and-invert preconditioner is a flexible and locally accelerated preconditioner, and is asymptotically optimal which leads to superlinear convergence of the PSD-id method. Numerical examples show the superlinear convergence of the PSD-id method with locally accelerated preconditioners for solving ill-conditioned generalized eigenvalue problems (1.1) arising from full self-consistent electronic structure calculations.

We would like to note that the main objective of this paper is to provide a rigorous convergence analysis of the PSD-id method with flexible and locally accelerated preconditioners rather than to advocate the usage of the PSD-id method in practice. The BPSD methods [1,16] and a recent proposed locally accelerated BPSD (LABPSD) presented in our previous work [3] have demonstrated their efficiency for finding several eigenpairs simultaneously. While a rigorous and full-scale convergence proof of the BPSD methods still largely eludes us, we believe the analysis of the PSD-id method presented in this paper can shed light on an improved understanding of the convergence behavior of the BPSD methods such as the LAPBSD method [3] for solving the ill-conditioned generalized eigenvalue problem (1.1) arising from the PUFE simulation of electronic structure calculations.

The rest of this paper is organized as follows. In section 2, we present the PSD-id method and discuss its basic properties. In section 3, we provide a convergence proof and a nonasymptotic estimate of the convergence rate of the PSD-id method. An asymptotically optimal preconditioner is discussed in section 4. Numerical examples to illustrate the theoretical results are presented in section 5. We close with some final remarks in section 6.

In the spirit of reproducible research, MATLAB scripts of an implementation of the PSD-id method, and the data that used to generate numerical results presented in this paper can be obtained from the URL http://dsec.pku.edu.cn/∼yfcai/psdid.html.

2. Algorithm

Assuming that U_{i-1} is already known, by (1.2), one can find the ith eigenpair by minimizing the Rayleigh quotient $\rho(z)$ with z being S-orthogonal against U_{i-1}. Specifically, let us denote by $(\lambda_{i;j}, u_{i;j})$ the jth approximation of (λ_i, u_i) and assume that

$$U_{i-1}^H S u_{i;j} = 0, \quad \|u_{i;j}\|_S = 1 \quad \text{and} \quad \lambda_{i;j} = \rho(u_{i;j}). \tag{2.1}$$

To compute the $(j+1)$st approximate eigenpair $(\lambda_{i;j+1}, u_{i;j+1})$, by the steepest descent approach, the steepest decreasing direction of $\rho(z)$ is opposite to the gradient of $\rho(z)$ at $z = u_{i;j}$:

$$\nabla\rho(u_{i;j}) = 2(H - \lambda_{i;j}S)u_{i;j} = 2r_{i;j}.$$

Furthermore, to accelerate the convergence, we use the following preconditioned search vector

$$p_{i;j} = -K_{i;j}r_{i;j}, \tag{2.2}$$

where $K_{i;j}$ is a preconditioner and satisfies $K_{i;j}^H = K_{i;j}$. By a Rayleigh-Ritz projection based implementation, the $(j+1)$st approximate eigenpair $(\lambda_{i;j+1}, u_{i;j+1})$ computed by the preconditioned steepest descent method is given by

$$(\lambda_{i;j+1}, u_{i;j+1}) = (\gamma_i, Z_j w_i), \tag{2.3}$$

where (γ_i, w_i) is the ith eigenpair of the projected matrix pair $(H_R, S_R) = (Z_j^H H Z_j, Z_j^H S Z_j)$, $\|w_i\|_{S_R} = 1$, and $Z_j = [U_{i-1}\ u_{i;j}\ p_{i;j}]$ is the basis matrix of the projection subspace. Here we assume that Z_j is of full column rank.

Algorithm 2.1 is a summary of the aforementioned procedure. Since the first eigenvectors U_{i-1} are implicitly deflated in the Rayleigh-Ritz procedure, we call Algorithm 2.1 *a preconditioned steepest descent with implicit deflation*, PSD-id in short. We note that the preconditioner $K_{i;j}$ is flexible. It can be changed at each iteration. If the preconditioner is fixed as a uniform positive definite matrix, i.e., $K_{i;j} = K > 0$, then Algorithm 2.1 is the SIRQIT-G2 algorithm in [13] with $K = I$ and initial vectors $X^{(0)} = [U_{i-1}\, u_{i;0}]$, and is the BPSD method [9] with initial vectors $[U_{i-1}\, u_{i;0}]$.

Algorithm 2.1 PSD-id

Input: U_{i-1} and initial vector $u_{i;0}$

Output: Approximate eigenpair $(\lambda_{i;j}, u_{i:j})$ of (λ_i, u_i)

1: $\lambda_{i;0} = \rho(u_{i;0})$
2: **for** $j = 0, 1, \ldots,$ until convergence **do**
3: compute $r_{i;j} = Hu_{i;j} - \lambda_{i;j}Su_{i;j}$
4: precondition $p_{i;j} = -K_{i;j}r_{i;j}$
5: compute the ith eigenpair (γ_i, w_i) of $(H_R, S_R) = (Z_j^H H Z_j, Z_j^H S Z_j)$, $Z_j = [U_{i-1}\ u_{i;j}\ p_{i;j}]$, $\|w_i\|_{S_R} = 1$
6: update $\lambda_{i;j+1} = \gamma_i$ and $u_{i;j+1} = Z_j w_i$
7: **end for**

If Algorithm 2.1 does not breakdown, i.e., the matrices Z_j on line 5 are of full column rank for all j, then a sequence of approximate eigenpairs $\{(\lambda_{i;j}, u_{i;j})\}_j$ is produced. The following proposition gives basic properties of the sequence. In particular, if the initial vector $u_{i;0}$ does not satisfy the assumption (2.1), the first approximate vector $u_{i;1}$ computed by Algorithm 2.1 will suffice.

Proposition 2.1. *If Z_j is of full column rank, then*

(a) $U_{i-1}^H S u_{i;j+1} = 0$.

(b) $\|u_{i;j+1}\|_S = 1$.

(c) $\lambda_{i;j+1} \geq \lambda_i$.

(d) $\lambda_{i;j+1} \leq \lambda_{i;j}$.

Proof. Results (a) and (b) are verified by straightforward calculations. The result (c) follows from the inequality

$$\lambda_{i;j+1} = \rho(u_{i;j+1}) \geq \min_{U_{i-1}^H S z = 0} \rho(z) = \lambda_i.$$

Finally, the result (d) follows from the facts that

$$\lambda_{i;j+1} = \lambda_i(H_R, S_R) = \min_{U_{i-1}^H S Z_j w = 0} \rho(Z_j w) \leq \rho(Z_j w)|_{w=e_i} = \rho(u_{i;j}) = \lambda_{i;j},$$

where e_i is the ith column vector of identity matrix of order $i+1$. □

The following proposition shows that with a proper choice of the preconditioner $K_{i;j}$, the basis matrix $Z_j = [U_{i-1}\ u_{i;j}\ p_{i;j}]$ is of full column rank, which implies that Algorithm 2.1 does not breakdown.

Proposition 2.2. *If* $r_{i;j} \neq 0$ *and* $K_{i;j}$ *is chosen such that*

$$K_{i;j}^c := (U_{i-1}^c)^H S K_{i;j} S U_{i-1}^c > 0, \tag{2.4}$$

then the basis matrix Z_j *is of full column rank. Here* U_{i-1}^c *is the complementary eigenvector matrix of* U_{i-1}, $U_{i-1}^c = [u_i\ \cdots\ u_n]$.

Proof. We prove that Z_j is of full column rank by showing that

$$\det(H_R - \lambda_{i;j} S_R) = \det(Z_j^H (H - \lambda_{i;j} S) Z_j) \neq 0.$$

First, it can be verified that the projected matrix pair (H_R, S_R) can be factorized as follows:

$$(H_R, S_R) = L^{-1} \left(\begin{bmatrix} \Lambda_{i-1} & 0 \\ 0 & H_\perp \end{bmatrix}, \begin{bmatrix} I_{i-1} & 0 \\ 0 & S_\perp \end{bmatrix} \right) L^{-H}, \tag{2.5}$$

where

$$L = \begin{bmatrix} I_{i-1} & 0 & 0 \\ 0 & 1 & 0 \\ -p_{i;j}^H S U_{i-1} & 0 & 1 \end{bmatrix}, \quad H_\perp = Z_\perp^H H Z_\perp, \quad S_\perp = Z_\perp^H S Z_\perp,$$

$Z_\perp = [u_{i;j}\ p_\perp]$ and $p_\perp = U_{i-1}^c (U_{i-1}^c)^H S p_{i;j}$. Consequently, we have

$$\det(H_R - \lambda_{i;j} S_R) = \det(\Lambda_{i-1} - \lambda_{i;j} I) \det(H_\perp - \lambda_{i;j} S_\perp). \tag{2.6}$$

By Proposition 2.1(c), we have $\lambda_{i;j} \geq \lambda_i$. Since $r_{i;j} \neq 0$, $\lambda_{i;j} > \lambda_i$. Hence, we conclude that

$$\det(\Lambda_{i-1} - \lambda_{i;j} I) \neq 0. \tag{2.7}$$

Next we show that $\det(H_\perp - \lambda_{i;j} S_\perp) \neq 0$. We first note that since $U_{i-1}^H S u_{i;j} = 0$, there exists a vector a such that $u_{i;j} = U_{i-1}^c a$. Then it follows that

$$r_{i;j} = (H - \lambda_{i;j} S) U_{i-1}^c a = S U_{i-1}^c (\Lambda_{i-1}^c - \lambda_{i;j} I) a = S U_{i-1}^c (U_{i-1}^c)^H r_{i;j}, \tag{2.8}$$

where $\Lambda_{i-1}^c = \mathrm{diag}(\lambda_i, \ldots, \lambda_n)$. Note that $(U_{i-1}^c)^H r_{i;j} \neq 0$ since $r_{i;j} \neq 0$. Furthermore, using (2.8) and (2.4), we have

$$\begin{aligned}
\det(H_\perp - \lambda_{i;j} S_\perp) &= \det \begin{bmatrix} 0 & r_{i;j}^H p_\perp \\ p_\perp^H r_{i;j} & p_\perp^H (H - \lambda_{i;j} S) p_\perp \end{bmatrix} \\
&= -|p_\perp^H r_{i;j}|^2 \\
&= -|r_{i;j}^H U_{i-1}^c (U_{i-1}^c)^H S K_{i;j} S U_{i-1}^c (U_{i-1}^c)^H r_{i;j}|^2 \\
&= -|r_{i;j}^H U_{i-1}^c K_{i;j}^c (U_{i-1}^c)^H r_{i;j}| < 0. \qquad (2.9)
\end{aligned}$$

By (2.6), (2.7) and (2.9), we conclude that $H_R - \lambda_{i;j} S_R$ is nonsingular, which implies that Z_j is of full column rank. □

Definition 2.1. *A preconditioner $K_{i;j}$ satisfying the condition* (2.4) *is called an* effectively positive definite preconditioner.

We note that an effectively positive definite preconditioner $K_{i;j}$ with $i > 1$ is not necessarily to be symmetric positive definite. For example, for any $\lambda_1 < \sigma < \lambda_i$ and σ is not an eigenvalue of (H, S), $K_{i;j} = (H - \sigma S)^{-1}$ is effectively positive definite, although $K_{i;j}$ is indefinite.

If the preconditioner $K_{i;j}$ is chosen such that the search vector $p_{i;j} = -K_{i;j} r_{i;j}$ satisfies

$$U^H S(u_{i;j} + p_{i;j}) = \xi = (\xi_1, \xi_2, \ldots, \xi_n)^H \quad \text{with } \xi_i \neq 0 \text{ and } \xi_j = 0 \text{ for } j > i, \qquad (2.10)$$

then

$$\begin{aligned}
\lambda_{i;j+1} &= \min_{U_{i-1}^H S Z_j w = 0} \rho(Z_j w) \\
&= \min_w \rho(U_{i-1}^c (U_{i-1}^c)^H S Z_j w) \\
&= \min_v \rho(U_{i-1}^c (U_{i-1}^c)^H S [u_{i;j}\ p_{i;j}] v) \\
&\leq \rho(U_{i-1}^c (U_{i-1}^c)^H S [u_{i;j}\ p_{i;j}] v)|_{v=[1\ 1]^\top} \\
&= \rho(U_{i-1}^c (U_{i-1}^c)^H S (u_{i;j} + p_{i;j})) \\
&= \rho(U_{i-1}^c (U_{i-1}^c)^H S U \xi) \\
&= \rho(\xi_i u_i) = \lambda_i. \qquad (2.11)
\end{aligned}$$

Therefore, combining the inequality (2.11) and Proposition 2.1(c), we have $\lambda_{i;j+1} = \lambda_i$. In this case, we refer to $p_{i;j}$ satisfying the equation (2.10) as an *ideal search direction*. The notion of an ideal search direction not only helps assessing the quality of a preconditioned search direction, but also tells the desired property for the solution of the preconditioning equation $p_{i;j} = -K_{i;j} r_{i;j}$.

3. Convergence Analysis

In this section, we prove the convergence of the PSD-id method and derive a nonasymptotic estimate of the convergence rate. For brevity, we assume that for the desired ith eigenvalue λ_i, it satisfies $\lambda_{i-1} < \lambda_i < \lambda_{i+1}$. Otherwise by replacing λ_{i+1} with the smallest eigenvalue of (H, S) which is larger than λ_i, all results in this section still hold, and the proofs are similar.

3.1. Convergence results

Assume that the preconditioner $K_{i;j}$ is effectively positive definite. By the definition of $\lambda_{i;j+1}$ in (2.3), the identity (2.6), we know that $\lambda_{i;j+1}$ is the smaller root of the quadratic polynomial $\det(H_\perp - \mu S_\perp) = 0$ of μ. Using (2.9), we have that $\lambda_{i;j+1}$ is strictly less than $\lambda_{i;j}$,

$$\lambda_{i;j+1} < \lambda_{i;j}. \tag{3.1}$$

Furthermore, by Proposition 2.1(c) and (3.1), the approximate eigenvalue sequence $\{\lambda_{i;j}\}_j$ is monotonically decreasing and is bounded below by λ_i, i.e.,

$$\lambda_{i;0} > \lambda_{i;1} > \cdots > \lambda_{i;j} > \lambda_{i;j+1} > \cdots \geq \lambda_i. \tag{3.2}$$

Therefore, the sequence $\{\lambda_{i;j}\}_j$ must converge. Does it converge to the ith eigenvalue λ_i of (H, S)? How about the corresponding $\{u_{i;j}\}_j$? We will answer these questions in this subsection. First, we give the following lemma to quantify the difference between two consecutive approximates $\lambda_{i;j}$ and $\lambda_{i;j+1}$ of λ_i.

Lemma 3.1. *If $r_{i;j} \neq 0$ and the preconditioner $K_{i;j}$ is effectively positive definite, then*

$$\lambda_{i;j} - \lambda_{i;j+1} \geq \sqrt{g^2 + \phi^2} - g, \tag{3.3}$$

where $g = (\lambda_n - \lambda_i)/2$ and $\phi = \|r_{i;j}\|_{S^{-1}}/\kappa(K^c_{i;j})$, $\kappa(K^c_{i;j})$ is the condition number of $K^c_{i;j}$ defined in (2.4).

Proof. Define $H_\perp$, $S_\perp$, $Z_\perp$ and $p_\perp$ as in (2.5). By (2.5) and Proposition 2.2, we know that $Z_\perp$ is of full column rank. Let $\hat{p} = (I - u_{i;j}u^H_{i;j}S)p_\perp$, then $[u_{i;j}\ \hat{p}] = Z_\perp \begin{bmatrix} 1 & -u^H_{i;j}p_\perp \\ 0 & 1 \end{bmatrix}$. Therefore, $[u_{i;j}\ \hat{p}]$ is also of full column rank, and hence it holds that $\hat{p}^H S\hat{p} > 0$. By straightforward calculations, we have

$$\begin{aligned}
\det(H_\perp - \mu S_\perp) &= \det(Z_\perp^H(H - \mu S)Z_\perp) = \det([u_{i;j}\ \hat{p}]^H(H - \mu S)[u_{i;j}\ \hat{p}]) \\
&= \det\left(\begin{bmatrix} \lambda_{i;j} & u^H_{i;j}H\hat{p} \\ \hat{p}^H H u_{i;j} & \hat{p}^H H\hat{p} \end{bmatrix} - \mu \begin{bmatrix} 1 & 0 \\ 0 & \hat{p}^H S\hat{p} \end{bmatrix}\right) \\
&= \hat{p}^H S\hat{p}(\lambda_{i;j} - \mu)(\rho(\hat{p}) - \mu) - |u^H_{i;j}H\hat{p}|^2 \\
&= \hat{p}^H S\hat{p}\left[(\lambda_{i;j} - \mu)^2 + (\rho(\hat{p}) - \lambda_{i;j})(\lambda_{i;j} - \mu) - \frac{|u^H_{i;j}H\hat{p}|^2}{\hat{p}^H S\hat{p}}\right].
\end{aligned} \tag{3.4}$$

By the definition of $\lambda_{i;j+1}$ in (2.3), the identity (2.6), we know that $\lambda_{i;j+1}$ is the smaller root of the quadratic polynomial (3.4) of μ. In addition, by (3.2), we know that $\lambda_{i;j} - \lambda_{i;j+1}$ is positive. Therefore $\lambda_{i;j} - \lambda_{i;j+1}$ is the positive root of the following quadratic equation in t:

$$t^2 + (\rho(\hat{p}) - \lambda_{i;j})t - \frac{|u^H_{i;j}H\hat{p}|^2}{\hat{p}^H S\hat{p}} = 0.$$

Then it follows that

$$\lambda_{i;j} - \lambda_{i;j+1} = -\frac{\rho(\hat{p}) - \lambda_{i;j}}{2} + \sqrt{\left(\frac{\rho(\hat{p}) - \lambda_{i;j}}{2}\right)^2 + \frac{|u^H_{i;j}H\hat{p}|^2}{\hat{p}^H S\hat{p}}}. \tag{3.5}$$

In what follows, we give the estimates of the quantities $|\rho(\hat{p}) - \lambda_{i;j}|$, $|u_{i;j}^H H\hat{p}|^2$ and $\hat{p}^H S\hat{p}$, respectively.

For the quantity $|\rho(\hat{p}) - \lambda_{i;j}|$, using the fact that for any nonzero z satisfying $U_{i-1}^H Sz = 0$, it holds $\lambda_i \le \rho(z) \le \lambda_n$, then using $U_{i-1}^H S\hat{p} = 0$ and $U_{i-1}^H S u_{i;j} = 0$, we have

$$0 \le |\rho(\hat{p}) - \lambda_{i;j}| \le \lambda_n - \lambda_i = 2g. \tag{3.6}$$

For the quantity $|u_{i;j}^H H\hat{p}|^2$, we have

$$\begin{aligned}
|u_{i;j}^H H\hat{p}| &= |u_{i;j}^H H(I - u_{i;j}u_{i;j}^H S)U_{i-1}^c (U_{i-1}^c)^H S K_{i;j} r_{i;j}| && \text{(3.7a)}\\
&= |\big[u_{i;j}^H H - \lambda_{i;j} u_{i;j}^H S\big] U_{i-1}^c \big[(U_{i-1}^c)^H S K_{i;j} S U_{i-1}^c\big] (U_{i-1}^c)^H r_{i;j}| && \text{(3.7b)}\\
&= |r_{i;j}^H U_{i-1}^c K_{i;j}^c (U_{i-1}^c)^H r_{i;j}| && \text{(3.7c)}\\
&\ge \lambda_{\min}(K_{i;j}^c) \|(U_{i-1}^c)^H r_{i;j}\|^2 && \text{(3.7d)}\\
&= \lambda_{\min}(K_{i;j}^c) \|r_{i;j}\|_{S^{-1}}^2, && \text{(3.7e)}
\end{aligned}$$

where (3.7a) uses the definition of $\hat{p}$ and (2.2), (3.7b) uses the fact that $r_{i;j} = SU_{i-1}^c (U_{i-1}^c)^H r_{i;j}$, (3.7c) and (3.7d) use the definition of $K_{i;j}^c$ in (2.4) and the assumption that $K_{i;j}^c$ is symmetric positive definite, and (3.7e) is based on the following calculations:

$$\begin{aligned}
\|(U_{i-1}^c)^H r_{i;j}\|^2 &= r_{i;j}^H U_{i-1}^c (U_{i-1}^c)^H r_{i;j}\\
&= r_{i;j}^H U_{i-1}^c (U_{i-1}^c)^H r_{i;j} + r_{i;j}^H U_{i-1} U_{i-1}^H r_{i;j} && (U_{i-1}^H r_{i;j} = 0)\\
&= r_{i;j}^H U U^H r_{i;j} = r_{i;j}^H S^{-1} r_{i;j} && (UU^H = S^{-1})\\
&= \|r_{i;j}\|_{S^{-1}}^2.
\end{aligned}$$

For the quantity $\hat{p}^H S\hat{p}$, we have

$$\begin{aligned}
&\hat{p}^H S\hat{p}\\
=&(r_{i;j})^H K_{i;j} S U_{i-1}^c \big[(U_{i-1}^c)^H (I - S u_{i;j}u_{i;j}^H) S (I - u_{i;j}u_{i;j}^H S) U_{i-1}^c\big] (U_{i-1}^c)^H S K_{i;j} r_{i;j}\\
\le&\|(U_{i-1}^c)^H (I - S u_{i;j}u_{i;j}^H) S (I - u_{i;j}u_{i;j}^H S) U_{i-1}^c\| \|(U_{i-1}^c)^H S K_{i;j} r_{i;j}\|^2\\
\le&\|(U_{i-1}^c)^H S K_{i;j} r_{i;j}\|^2 = \|K_{i;j}^c (U_i^c)^H r_{i;j}\|^2\\
\le&\lambda_{\max}(K_{i;j}^c)^2 \|(U_{i-1}^c)^H r_{i;j}\|^2 = \lambda_{\max}(K_{i;j}^c)^2 \|r_{i;j}\|_{S^{-1}}^2,
\end{aligned} \tag{3.8}$$

where the second inequality uses the fact that

$$\begin{aligned}
&\|(U_{i-1}^c)^H (I - S u_{i;j}u_{i;j}^H) S (I - u_{i;j}u_{i;j}^H S) U_{i-1}^c\|\\
=\ &\|(U_{i-1}^c)^H S^{\frac{1}{2}} (I - S^{\frac{1}{2}} u_{i;j}u_{i;j}^H S^{\frac{1}{2}}) (I - S^{\frac{1}{2}} u_{i;j}u_{i;j}^H S^{\frac{1}{2}}) S^{\frac{1}{2}} U_{i-1}^c\|\\
\le\ &\|(I - S^{\frac{1}{2}} u_{i;j}u_{i;j}^H S^{\frac{1}{2}})\|^2 \|S^{\frac{1}{2}} U_{i-1}^c\|^2 \le 1.
\end{aligned}$$

Finally, we arrive at

$$\begin{aligned}
\lambda_{i;j} - \lambda_{i;j+1} &\ge -\frac{|\rho(\hat{p}) - \lambda_{i;j}|}{2} + \sqrt{\left(\frac{\rho(\hat{p}) - \lambda_{i;j}}{2}\right)^2 + \frac{|u_{i;j}^H H\hat{p}|^2}{\hat{p}^H S\hat{p}}} && \text{(3.9a)}\\
&\ge -\frac{\lambda_n - \lambda_i}{2} + \sqrt{\left(\frac{\lambda_n - \lambda_i}{2}\right)^2 + \frac{|u_{i;j}^H H\hat{p}|^2}{\hat{p}^H S\hat{p}}} && \text{(3.9b)}\\
&\ge -\frac{\lambda_n - \lambda_i}{2} + \sqrt{\left(\frac{\lambda_n - \lambda_i}{2}\right)^2 + \frac{\|r_{i;j}\|_{S^{-1}}^2}{\kappa^2(K_{i;j}^c)}} && \text{(3.9c)}\\
&= -g + \sqrt{g^2 + \phi^2},
\end{aligned}$$

where (3.9a) uses (3.5), (3.9b) uses (3.6) and (3.9c) uses (3.7e) and (3.8). This completes the proof. □

We note that in [4, Chap.7], for the steepest descent method to compute the largest eigenvalue λ_n of a Hermitian matrix, it shows that

$$\lambda_{n;j+1} - \lambda_{n;j} \geq \frac{\|r_{n;j}\|^2}{\lambda_n - \lambda_1}.$$

Then it is established that $\lambda_{n;j}$ converges to λ_n, and $u_{n;j}$ converges to u_n directionally. Lemma 3.1 and the following theorem are generalizations that are not limited to the largest eigenpair, and include the usage of flexible preconditioners.

Theorem 3.1. *If the initial estimate eigenvalue $\lambda_{i;0}$ satisfies $\lambda_i < \lambda_{i;0} < \lambda_{i+1}$, and the flexible preconditioners $K_{i;j}$ are effectively positive definite for all j and $\sup_j \kappa(K_{i;j}^c)\|r_{i;j}\|_{S^{-1}} = q < \infty$, then the sequence $\{(\lambda_{i;j}, u_{i;j})\}_j$ generated by the PSD-id method converges to the desired pair (λ_i, u_i), i.e.,*

(a) $\lim_{j\to\infty} \lambda_{i;j} = \lambda_i$.

(b) $\lim_{j\to\infty} \|r_{i;j}\|_{S^{-1}} = 0$, *namely $u_{i;j}$ converges to u_i directionally.*

Proof. To prove (a), we first notice that $\{\lambda_{i;j}\}_j$ is a monotonic decreasing sequence, and is bounded by λ_i from below. So there exists a real number $\tilde{\lambda}_i$ such that $\lambda_{i;j} \to \tilde{\lambda}_i$ as $j \to \infty$. Now we show by contradiction that $\tilde{\lambda}_i = \lambda_i$. For any $u_{i;j}$ ($\|u_{i;j}\|_S = 1$), we have

$$\begin{aligned}\|r_{i;j}\|_{S^{-1}} &= \|(H - \lambda_{i;j}S)u_{i;j}\|_{S^{-1}} \\ &\geq \|(H - \tilde{\lambda}_i S)u_{i;j}\|_{S^{-1}} - (\lambda_{i;j} - \tilde{\lambda}_i)\|Su_{i;j}\|_{S^{-1}} \\ &\geq \min_k |\lambda_k - \tilde{\lambda}_i| - (\lambda_{i;j} - \tilde{\lambda}_i).\end{aligned}$$

As $\lim_{j\to\infty} \lambda_{i;j} = \tilde{\lambda}_i$, there exists a j_0 such that for any $j \geq j_0$,

$$\|r_{i;j}\|_{S^{-1}} > \frac{1}{2}\min_k |\lambda_k - \tilde{\lambda}_i|. \tag{3.10}$$

By the assumption $\sup_j \kappa(K_{i;j}^c)\|r_{i;j}\|_{S^{-1}} = q < \infty$, we know that for any j,

$$\kappa(K_{i;j}^c) \leq q/\|r_{i;j}\|_{S^{-1}}. \tag{3.11}$$

By defining $d(r, \kappa) := -g + \sqrt{g^2 + (r/\kappa)^2}$, and using (3.10) and (3.11), by Lemma 3.1 we know that for any $j \geq j_0$, it holds

$$\begin{aligned}\lambda_{i;j} - \lambda_{i;j+1} &\geq d(\|r_{i;j}\|_{S^{-1}}, \kappa(K_{i;j}^c)) \\ &\geq d(\|r_{i;j}\|_{S^{-1}}, q/\|r_{i;j}\|_{S^{-1}}) > d(\min_k |\lambda_k - \tilde{\lambda}_i|^2/4, q),\end{aligned}$$

which in the limit becomes

$$0 > d(\min_k |\lambda_k - \tilde{\lambda}_i|^2/4, q).$$

This is a contradiction to the fact that $d(\min_k |\lambda_k - \tilde{\lambda}_i|^2/4, q)$ is a positive constant.

To prove (b), using $\lim_{j\to\infty} \lambda_{i;j} = \lambda_i$, (3.11) and Lemma 3.1, we have

$$\lim_{j\to\infty} d(\|r_{i;j}\|_{S^{-1}}^2, q) \leq \lim_{j\to\infty} d(\|r_{i;j}\|_{S^{-1}}, \kappa(K_{i;j}^c)) \leq \lim_{j\to\infty} (\lambda_{i;j} - \lambda_{i;j+1}) = 0,$$

Consequently, $\lim_{j\to\infty} d(\|r_{i;j}\|^2_{S^{-1}}, q) = 0$, which leads to $\lim_{j\to\infty} \|r_{i;j}\|_{S^{-1}} = 0$ since $1 \le q < \infty$. □

We note that in Theorem 3.1, if without assuming $\lambda_{i;0} < \lambda_{i+1}$, then by similar argument, we can conclude that $\{\lambda_{i;j}\}_j$ monotonically converges to an eigenvalue λ_k, where $k \in \{i, i+1, \ldots, n\}$, and meanwhile $\{u_{i;j}\}_j$ directionally converges to the corresponding eigenvector u_k. In the case when $k > i$, λ_k is a saddle point of the restricted Rayleigh quotient $\rho(z)|_{U_{i-1}^H S z=0}$, and hence unstable. In practice, due to rounding-off error, $\{\lambda_{i;j}\}_j$ seldom converges to λ_k with $k > i$.

3.2. Rate of convergence

Theorem 3.1 concludes the convergence of the sequence $\{\lambda_{i;j}\}_j$, in what follows we derive a nonasymptotic estimate of the convergence rate of $\{\lambda_{i;j}\}_j$ based on the work of Samokish in 1958 [25]. We begin by recalling the following equalities for the projection matrix $P_{i-1} = I - U_{i-1}U_{i-1}^H S = U_{i-1}^c (U_{i-1}^c)^H S$:

$$P_{i-1}u_{i;j} = u_{i;j}, \tag{3.12a}$$
$$P_{i-1}^2 = P_{i-1}, \tag{3.12b}$$
$$P_{i-1}^H (H - \lambda_i S) = (H - \lambda_i S)P_{i-1}, \tag{3.12c}$$
$$P_{i-1}^H S = S P_{i-1}, \tag{3.12d}$$

First, we have the following lemma.

Lemma 3.2. *Define*

$$M = P_{i-1}^H (H - \lambda_i S) P_{i-1} \tag{3.13}$$

and assume that $K_{i;j}$ is effectively positive definite.

(a) M is positive semi-definite and $M = GG^H$, where $G = SU_i^c(\Lambda_i^c - \lambda_i I)^{\frac{1}{2}}$ is of full column rank.

(b) All eigenvalues of $G^H K_{i;j} G$ are positive.

(c) The eigenvalues of $K_{i;j}M$ are given by

$$\lambda(K_{i;j}M) = \{0_{[i]}\} \cup \lambda\big(G^H K_{i;j} G\big), \tag{3.14}$$

where $0_{[i]}$ stands for the multiplicity i of the number 0.

Proof. (a) By the definitions of M and P_{i-1}, it easy to see that

$$\begin{aligned} M &= SU_{i-1}^c (U_{i-1}^c)^H (H - \lambda_i S) U_{i-1}^c (U_{i-1}^c)^H S \\ &= SU_i^c (\Lambda_i^c - \lambda_i I)(U_i^c)^H S = GG^H \ge 0, \end{aligned}$$

where $G = SU_i^c(\Lambda_i^c - \lambda_i I)^{\frac{1}{2}}$.

(b) Direct calculation leads to

$$G^H K_{i;j} G = (\Lambda_i^c - \lambda_i I)^{\frac{1}{2}} \widetilde{K}_{22} (\Lambda_i^c - \lambda_i I)^{\frac{1}{2}},$$

where $\widetilde{K}_{22}$ is the trailing $(n-i)$-by-$(n-i)$ principal submatrix of $K_{i;j}^c$ by deleting its first $i-1$ rows and first $i-1$ columns. Since $K_{i;j}$ is effectively positive definite, we know that $K_{i;j}^c > 0$ and hence $\widetilde{K}_{22} > 0$. Thus all eigenvalues of $G^H K_{i;j} G$ are positive.

(c) It follows that

$$\lambda(K_{i;j}M) = \lambda(K_{i;j}GG^H) = \{0_{[i]}\} \cup \lambda(G^H K_{i;j} G),$$

where we use the well-known identity $\lambda(AB) = \{0_{[m-n]}\} \cup \lambda(BA)$ for $A \in \mathbb{C}^{m\times n}$ and $B \in \mathbb{C}^{n\times m}$ and $m \geq n$. □

We now give a nonasymptotic estimate of the convergence rate of PSD-id (Algorithm 2.1).

Theorem 3.2. *Let $\epsilon_{i;j} = \lambda_{i;j} - \lambda_i$ and $\lambda_{i;j}$ be* localized, *namely*

$$\tau(\sqrt{\theta_{i;j}\epsilon_{i;j}} + \delta_{i;j}\epsilon_{i;j}) < 1, \tag{3.15}$$

then

$$\epsilon_{i;j+1} \leq \left[\frac{\Delta + \tau\sqrt{\theta_{i;j}\epsilon_{i;j}}}{1 - \tau(\sqrt{\theta_{i;j}\epsilon_{i;j}} + \delta_{i;j}\epsilon_{i;j})}\right]^2 \epsilon_{i;j}, \tag{3.16}$$

where $\theta_{i;j} = \|S^{\frac{1}{2}} K_{i;j} M K_{i;j} S^{\frac{1}{2}}\|$, $\delta_{i;j} = \|S^{\frac{1}{2}} K_{i;j} S^{\frac{1}{2}}\|$, $\Delta = (\Gamma - \gamma)/(\Gamma + \gamma)$, Γ and γ are the largest and smallest positive eigenvalues of $K_{i;j}M$, respectively, and $\tau = 2/(\Gamma + \gamma)$.

Proof. Recall $Z_\perp = [\, u_{i;j} \;\; P_{i-1}p_{i;j} \,]$ defined in (2.5). It is easy to see that by using $Z_\perp$, the $(j+1)$st approximate eigenpair $(\lambda_{i;j+1}, u_{i;j})$ can be written as

$$\lambda_{i;j+1} = \min_v \rho(Z_\perp v).$$

Considering a choice of the vector v for the line search, we have

$$\lambda_{i;j+1} = \min_v \rho(Z_\perp v) \leq \rho(Z_\perp v)|_{v=[1\ \tau]^\top} = \rho(z),$$

where $z = Z_\perp [1\ \tau]^\top = u_{i;j} + \tau P_{i-1}p_{i;j}$. Consequently, we have

$$\epsilon_{i;j+1} = \lambda_{i;j+1} - \lambda_i \leq \rho(z) - \lambda_i = \frac{z^H(H - \lambda_i S)z}{z^H S z}. \tag{3.17}$$

In the following, we provide estimates for the numerator and denominator of the upper bound (3.17).

For the numerator of the upper bound in (3.17), it follows that

$$\begin{aligned}
&z^H(H-\lambda_i S)z \\
=&(u_{i;j} + \tau P_{i-1}p_{i;j})^H P_{i-1}^H (H - \lambda_i S) P_{i-1}(u_{i;j} + \tau P_{i-1}p_{i;j}) && \text{(3.18a)}\\
=&\|u_{i;j} + \tau P_{i-1}p_{i;j}\|_M^2 \\
=&\|u_{i;j} - \tau P_{i-1}K_{i;j}(H - \lambda_{i;j}S)u_{i;j}\|_M^2 \\
=&\|u_{i;j} - \tau P_{i-1}K_{i;j}[(H - \lambda_i S) - \epsilon_{i;j}S]u_{i;j}\|_M^2 \\
=&\|[I - \tau P_{i-1}K_{i;j}(H - \lambda_i S)]u_{i;j} + \tau\epsilon_{i;j}P_{i-1}K_{i;j}Su_{i;j}\|_M^2 \\
=&\|[I - \tau P_{i-1}K_{i;j}P_{i-1}^H(H - \lambda_i S)P_{i-1}]u_{i;j} + \tau\epsilon_{i;j}P_{i-1}K_{i;j}Su_{i;j}\|_M^2 && \text{(3.18b)}\\
\leq&\big(\|[I - \tau P_{i-1}K_{i;j}M]u_{i;j}\|_M + \tau\epsilon_{i;j}\|P_{i-1}K_{i;j}Su_{i;j}\|_M\big)^2, && \text{(3.18c)}
\end{aligned}$$

where the equality (3.18a) uses the identities (3.12a) and (3.12b), (3.18b) uses (3.12a) and (3.12c). The inequality (3.18c) uses the triangular inequality of the vector semi-norm induced by the semi-positive definite matrix M. For the first term in (3.18c), using $M = GG^H$ and $G^H P_{i-1} = G^H$, where G is defined in Lemma 3.2, we have

$$\begin{aligned}&\|[I - \tau P_{i-1} K_{i;j} M] u_{i;j}\|_M = \|G^H [I - \tau P_{i-1} K_{i;j} G G^H] u_{i;j}\| \\ =&\|(I - \tau G^H K_{i;j} G)(G^H u_{i;j})\| \leq \|(I - \tau G^H K_{i;j} G)\| \|u_{i;j}\|_M. \end{aligned} \tag{3.19}$$

Note that by Lemma 3.2, it yields that

$$\begin{aligned}&\|(I - \tau G^H K_{i;j} G)\| = \max_k |1 - \tau \lambda_k (G^H K_{i;j} G)| \\ =&\max\{|1 - \tau\gamma|, |1 - \tau\Gamma|\} = \frac{\Gamma - \gamma}{\Gamma + \gamma} = \Delta. \end{aligned} \tag{3.20}$$

Consequently, we can rewrite (3.19) as

$$\|[I - \tau P_{i-1} K_{i;j} M] u_{i;j}\|_M \leq \Delta \, \|u_{i;j}\|_M = \Delta \sqrt{\epsilon_{i;j}}. \tag{3.21}$$

For the second term in (3.18c):

$$\begin{aligned}\|P_{i-1} K_{i;j} S u_{i;j}\|_M^2 &= |u_{i;j}^H S K_{i;j} P_{i-1}^H M P_{i-1} K_{i;j} S u_{i;j}| \\ &\leq \|S^{\frac{1}{2}} K_{i;j} M K_{i;j} S^{\frac{1}{2}}\| \|S^{\frac{1}{2}} u_{i;j}\|^2 \qquad (3.22a) \\ &= \|S^{\frac{1}{2}} K_{i;j} M K_{i;j} S^{\frac{1}{2}}\| \qquad (3.22b) \\ &= \theta_{i;j}, \end{aligned}$$

where (3.22a) uses (3.12b), (3.22b) uses the fact $\|u_{i;j}\|_S = 1$. Combining (3.21) and (3.22), an estimate of the numerator of the upper bound in (3.17) is given by

$$z^H (H - \lambda_i S) z \leq (\Delta + \tau \sqrt{\theta_{i;j} \epsilon_{i;j}})^2 \epsilon_{i;j}. \tag{3.23}$$

For the denominator of the upper bound (3.17), we first note that

$$\begin{aligned}z^H S z &= \|u_{i;j} + \tau P_{i-1} p_{i;j}\|_S^2 \\ &\geq (\|u_{i;j}\|_S - \tau \|P_{i-1} p_{i;j}\|_S)^2 \\ &= (1 - \tau \|P_{i-1} p_{i;j}\|_S)^2. \end{aligned} \tag{3.24}$$

By calculations, we have the following upper bound for $\|P_{i-1} p\|_S$:

$$\begin{aligned}&\|P_{i-1} p_{i;j}\|_S \\ =&\|P_{i-1} K_{i;j} (H - \lambda_{i;j} S) u_{i;j}\|_S \\ =&\|P_{i-1} K_{i;j} (H - \lambda_i S) u_{i;j} - \epsilon_{i;j} P_{i-1} K_{i;j} S u_{i;j}\|_S \\ \leq&\|P_{i-1} K_{i;j} (H - \lambda_i S) u_{i;j}\|_S + \epsilon_{i;j} \|P_{i-1} K_{i;j} S u_{i;j}\|_S \\ =&\|P_{i-1} K_{i;j} M u_{i;j}\|_S + \epsilon_{i;j} \|P_{i-1} K_{i;j} S u_{i;j}\|_S \qquad (3.25a) \\ \leq&\|S^{\frac{1}{2}} P_{i-1} S^{-\frac{1}{2}}\| \|S^{\frac{1}{2}} K_{i;j} M^{\frac{1}{2}}\| \|M^{\frac{1}{2}} u_{i;j}\| + \epsilon_{i;j} \|S^{\frac{1}{2}} P_{i-1} S^{-\frac{1}{2}}\| \|S^{\frac{1}{2}} K_{i;j} S^{\frac{1}{2}}\| \|S^{\frac{1}{2}} u_{i;j}\| \\ \leq&\sqrt{\theta_{i;j} \epsilon_{i;j}} + \delta_{i;j} \epsilon_{i;j}, \qquad (3.25b) \end{aligned}$$

where the equality (3.25a) uses (3.12a) and (3.12c), and the inequality (3.25b) uses the fact that $\|S^{\frac{1}{2}} P_{i-1} S^{-\frac{1}{2}}\| \leq 1$.

By (3.24) and (3.25b), if

$$\tau(\sqrt{\theta_{i;j}\epsilon_{i;j}} + \delta_{i;j}\epsilon_{i;j}) < 1,$$

then the denominator of the upper bound (3.17) satisfies

$$z^H S z \geq (1 - \tau(\sqrt{\theta_{i;j}\epsilon_{i;j}} + \delta_{i;j}\epsilon_{i;j}))^2. \tag{3.26}$$

By combining (3.17), (3.23) and (3.26), we derive the desired estimate (3.16). This concludes the proof. □

We note that the localization assumption (3.15) of Theorem 3.2 essentially requires that the approximate eigenvalue $\lambda_{i;j}$ is sufficiently close to λ_i. Theorem 3.2 indicates that if $\lambda_{i;j}$ is localized, then as $\Delta + \tau\sqrt{\theta_{i;j}\epsilon_{i;j}} \to 0$ as $j \to \infty$, the PSD-id algorithm converges *superlinearly*. In this case, we may call the preconditioner $Ki; j$ *asymptotically optimal*. We will consider one such preconditioner in next section and show how to check the localization condition (3.15) in practice in section 5.

To end this section we note that for the smallest eigenvalue λ_1, if the preconditioner $K_{i;j}$ is chosen to be fixed and positive definite, i.e., $K_{i;j} = K > 0$, one can verify that $\theta_{i;j} \leq \Gamma\delta_{i;j}$. Theorem 3.2 becomes the classical Samokish's theorem [20, 25], which remains asymptotically most accurate estimate of the convergence rate of the PSD method and is proven to be *sharp* [20]. The proof of Theorem 3.2 relies on the triangular inequality (3.18), which is inspired by the proof of Samokish's theorem presented in [20]. However, the treatment of each term in (3.17) needs to be handled diligently in order to accommodates the effects of the projection matrix P_{i-1} and the flexible preconditioner $K_{i;j}$.

4. An Asymptotically Optimal Preconditioner

In this section, we consider the shift-and-invert preconditioner:

$$\widehat{K}_{i;j} = \big(H - \beta_{i;j} S\big)^{-1}, \tag{4.1}$$

where $\beta_{i;j}$ is the shift. The following theorem shows that with a proper choice of $\beta_{i;j}$, $\widehat{K}_{i;j}$ is asymptotically optimal and consequently, the PSD-id method converges superlinearly.

Theorem 4.1. *Consider the shift*

$$\beta_{i;j} = \lambda_{i;j} - c\|r_{i;j}\|_{S^{-1}}, \tag{4.2}$$

where the constant $c = \inf_{k\geq 1} \sqrt{(\lambda_{i;k} - \lambda_i)(\lambda_{i+1} - \lambda_{i;k})}/\|r_{i;k}\|_{S^{-1}}$. [1] *If*

$$c > 3\sqrt{\Delta_{i;j}}, \tag{4.3a}$$

$$0 < \Delta_{i;j} < \min\left\{\frac{\Delta_i^2}{4}, \frac{1}{10}\right\}, \tag{4.3b}$$

where $\Delta_i = (\lambda_i - \lambda_{i-1})/(\lambda_{i+1} - \lambda_i)$ *and* $\Delta_{i;j} = (\lambda_{i;j} - \lambda_i)/(\lambda_{i+1} - \lambda_{i;j})$. *Then*

[1] In PSD-id, the approximate eigenvector $u_{i;k}$ corresponding with $\lambda_{i;k}$ can be given by $u_{i;k} = U_{i-1}^c a$ for some vector a. Let $a = [a_i \ \dots \ a_n]^\top$ with $|a_i|^2 = 1 - \epsilon^2$, $\sum_{j=i}^n |a_j|^2 = 1$, where ϵ^2 is sufficiently small such that $\lambda_{i;k} < \lambda_{i+1}$. We can show that $\|r_{i;k}\|_{S^{-1}} = O(\epsilon)$ and $|\lambda_{i;k} - \lambda_i| = O(\epsilon^2)$. Therefore, the constant c defined here is positive.

(a) $\beta_{i;j} < \lambda_i$ *and* $\widehat{K}_{i;j}$ *is effectively positive definite.*

(b) $\beta_{i;j} \to \lambda_i$ *as* $j \to \infty$.

(c) The condition (3.15) *of Theorem* 3.2 *is satisfied, namely,* $\lambda_{i;j}$ *is* localized.

(d) $\Delta + \tau\sqrt{\theta_{i;j}\epsilon_{i;j}} \to 0$ *as* $j \to \infty$, *where* Δ *is defined in Theorem* 3.2.

By (c) and (d), the preconditioner $\widehat{K}_{i;j}$ *is asymptotically optimal.*

Proof. (a) By the condition (4.3b), the relation $0 < \Delta_{i;j} < 0.1$ implies that λ_i is the closest eigenvalue to $\lambda_{i;j}$. Let $u_{i;j} = U_{i-1}^c a$ for some a, then $(\lambda_{i;j}, a)$ is an approximated eigenpair of Λ_{i-1}^c. Using the Kato-Temple inequality [8], we have

$$(\lambda_{i;j} - \lambda_i)(\lambda_{i+1} - \lambda_{i;j}) \le \|(\Lambda_{i-1}^c - \lambda_{i;j} I)a\|_2 = \|r_{i;j}\|_{S^{-1}}^2. \tag{4.4}$$

Therefore, the result $\beta_{i;j} < \lambda_i$ is verified as follows:

$$\begin{aligned}\beta_{i;j} - \lambda_i &= \lambda_{i;j} - c\|r_{i;j}\|_{S^{-1}} - \lambda_i \\ &\le \lambda_{i;j} - \lambda_i - c\sqrt{(\lambda_{i;j} - \lambda_i)(\lambda_{i+1} - \lambda_{i;j})} \\ &= (\lambda_{i+1} - \lambda_{i;j})(\Delta_{i;j} - c\sqrt{\Delta_{i;j}}) < 0,\end{aligned}$$

where for the last inequality we used the condition (4.3a).

The preconditioner $\widehat{K}_{i;j}$ is effectively positive definite since

$$\widehat{K}_{i;j}^c = (U_{i-1}^c)^H S \widehat{K}_{i;j} S U_{i-1}^c = \operatorname{diag}\Big(\frac{1}{\lambda_i - \beta_{i;j}}, \frac{1}{\lambda_{i+1} - \beta_{i;j}}, \ldots, \frac{1}{\lambda_n - \beta_{i;j}}\Big) \tag{4.5}$$

and $\beta_{i;j} < \lambda_i$.

(b) By (4.5), we have $\kappa(\widehat{K}_{i;j}^c) = \frac{\lambda_n - \beta_{i;j}}{\lambda_i - \beta_{i;j}}$. Using the definition of c, we have

$$c\|r_{i;j}\|_{S^{-1}} \le \sqrt{(\lambda_{i;j} - \lambda_i)(\lambda_{i+1} - \lambda_{i;j})}. \tag{4.6}$$

Then by calculations, we get

$$\kappa(\widehat{K}_{i;j}^c)\|r_{i;j}\|_{S^{-1}} = \frac{(\lambda_n - \lambda_{i;j} + c\|r_{i;j}\|_{S^{-1}})\|r_{i;j}\|_{S^{-1}}}{c\|r_{i;j}\|_{S^{-1}} - (\lambda_{i;j} - \lambda_i)} \tag{4.7a}$$

$$\le \frac{\big(\lambda_n - \lambda_i + \sqrt{(\lambda_{i;j} - \lambda_i)(\lambda_{i+1} - \lambda_{i;j})}\big)\|r_{i;j}\|_{S^{-1}}}{c\|r_{i;j}\|_{S^{-1}} - (\lambda_{i;j} - \lambda_i)} \tag{4.7b}$$

$$\le \frac{\big(\lambda_n - \lambda_i + \sqrt{(\lambda_{i;j} - \lambda_i)(\lambda_{i+1} - \lambda_{i;j})}\big)\|r_{i;j}\|_{S^{-1}}}{c\sqrt{(\lambda_{i;j} - \lambda_i)(\lambda_{i+1} - \lambda_{i;j})} - (\lambda_{i;j} - \lambda_i)} \tag{4.7c}$$

$$\le \frac{\lambda_n + \lambda_{i+1} - 2\lambda_i}{c} \frac{\sqrt{(\lambda_{i;j} - \lambda_i)(\lambda_{i+1} - \lambda_{i;j})}}{c\sqrt{(\lambda_{i;j} - \lambda_i)(\lambda_{i+1} - \lambda_{i;j})} - (\lambda_{i;j} - \lambda_i)} \tag{4.7d}$$

$$= \frac{\lambda_n + \lambda_{i+1} - 2\lambda_i}{c(c - \sqrt{\Delta_{i;j}})} \le \frac{\lambda_n + \lambda_{i+1} - 2\lambda_i}{c(c - \sup_j \sqrt{\Delta_{i;j}})} := q, \tag{4.7e}$$

where (4.7a) uses (4.2), (4.7b) uses $\lambda_{i;j} \ge \lambda_i$ and (4.6), (4.7c) uses (4.4) and (4.7d) uses $\lambda_i \le \lambda_{i;j} \le \lambda_{i+1}$ and (4.6).

Using (4.3a), we know that q defined in (4.7e) is a positive constant. Consequently, we can apply Theorem 3.1 to get the result (b).

(c) With the choice of $\beta_{i;j}$ in (4.2), for $\theta_{i;j}$, we have

$$\theta_{i;j} = \|S^{\frac{1}{2}}\widehat{K}_{i;j}M\widehat{K}_{i;j}S^{\frac{1}{2}}\| = \max_{i+1\le k\le n}\frac{\lambda_k-\lambda_i}{(\lambda_k-\beta_{i;j})^2} = \frac{\lambda_{i+1}-\lambda_i}{(\lambda_{i+1}-\beta_{i;j})^2}, \tag{4.8}$$

where for the last equality, we only need to show that $f'(x) < 0$ for $x \ge \lambda_{i+1}$, where $f(x) = \frac{x-\lambda_i}{(x-\beta_{i;j})^2}$. By calculations, we have

$$f'(x) = \frac{2\lambda_i - x - \beta_{i;j}}{(x-\beta_{i;j})^3} < 0$$

since $x - \beta_{i;j} > 0$ and

$$\begin{aligned}&2\lambda_i - x - \beta_{i;j} \le 2\lambda_i - \lambda_{i+1} - \lambda_{i;j} + c\|r_{i;j}\|_{S^{-1}}\\ <&\lambda_{i;j} - \lambda_{i+1} + \sqrt{(\lambda_{i;j}-\lambda_i)(\lambda_{i+1}-\lambda_{i;j})}\\ =&(\lambda_{i+1}-\lambda_{i;j})(-1+\sqrt{\Delta_{i;j}}) < 0.\end{aligned}$$

For $\delta_{i;j}$, we have

$$\delta_{i;j} = \|S^{\frac{1}{2}}\widehat{K}_{i;j}S^{\frac{1}{2}}\| = \frac{1}{\min_{1\le k\le n}|\lambda_k-\beta_{i;j}|} = \frac{1}{\lambda_i-\beta_{i;j}}, \tag{4.9}$$

where for the last equality, we only need to show that $\beta_{i;j} - \lambda_{i-1} > \lambda_i - \beta_{i;j}$, which is equivalent to

$$2c\|r_{i;j}\|_{S^{-1}} < 2\lambda_{ij} - \lambda_i - \lambda_{i-1}.$$

Notice that the right hand side of the above inequality is no less than $\lambda_i - \lambda_{i-1}$, thus, we only need to show

$$2c\|r_{i;j}\|_{S^{-1}} < \lambda_i - \lambda_{i-1}.$$

By calculations, we have

$$\frac{\lambda_i-\lambda_{i-1}}{2c\|r_{i;j}\|_{S^{-1}}} \ge \frac{\lambda_i-\lambda_{i-1}}{2\sqrt{(\lambda_{i;j}-\lambda_i)(\lambda_{i+1}-\lambda_{i;j})}} > \frac{\Delta_i}{2\sqrt{\Delta_{i;j}}} \ge 1.$$

In addition, using Lemma 3.2(c) and (4.5), it is easy to see that

$$\Gamma = \frac{\lambda_n-\lambda_i}{\lambda_n-\beta_{i;j}} \quad \text{and} \quad \gamma = \frac{\lambda_{i+1}-\lambda_i}{\lambda_{i+1}-\beta_{i;j}}. \tag{4.10}$$

Then it follows that

$$\begin{aligned}&\tau = 2/(\Gamma+\gamma) \le 1/\gamma,\\ &\tau\sqrt{\theta_{i;j}\epsilon_{i;j}} \le \frac{1}{\gamma}\sqrt{\theta_{i;j}\epsilon_{i;j}} = \frac{\lambda_{i+1}-\beta_{i;j}}{\lambda_{i+1}-\lambda_i}\sqrt{\frac{\lambda_{i+1}-\lambda_i}{(\lambda_{i+1}-\beta_{i;j})^2}(\lambda_{i;j}-\lambda_i)} = \sqrt{\Delta_{i;j}},\\ &\frac{1}{\gamma} = \frac{\lambda_{i+1}-\lambda_{i;j}+c\|r_{i;j}\|_{S^{-1}}}{\lambda_{i+1}-\lambda_i} < \frac{\lambda_{i+1}-\lambda_{i;j}+\sqrt{(\lambda_{i;j}-\lambda_i)(\lambda_{i+1}-\lambda_{i;j})}}{\lambda_{i+1}-\lambda_{i;j}} = 1+\sqrt{\Delta_{i;j}},\\ &\delta_{i;j}\epsilon_{i;j} = \frac{\lambda_{i;j}-\lambda_i}{\lambda_i-\beta_{i;j}} = \frac{\lambda_{i;j}-\lambda_i}{\lambda_i-\lambda_{i;j}+c\|r_{i;j}\|_{S^{-1}}} < \frac{\lambda_{i;j}-\lambda_i}{\lambda_i-\lambda_{i;j}+3\sqrt{\Delta_{i;j}}\|r_{i;j}\|_{S^{-1}}}\\ &\qquad\le \frac{\lambda_{i;j}-\lambda_i}{\lambda_i-\lambda_{i;j}+3\sqrt{\Delta_{i;j}}\sqrt{(\lambda_{i;j}-\lambda_i)(\lambda_{i+1}-\lambda_{i;j})}} = \frac{1}{-1+3} = \frac{1}{2}.\end{aligned}$$

Therefore,

$$\tau(\sqrt{\theta_{i;j}\epsilon_{i;j}} + \delta_{i;j}\epsilon_{i;j}) \leq \sqrt{\Delta_{i;j}} + \frac{1+\sqrt{\Delta_{i;j}}}{2} < 1.$$

(d) By the expressions (4.10) of Γ and γ, we have

$$\begin{aligned}\Delta = \frac{\Gamma-\gamma}{\Gamma+\gamma} &= \frac{(\lambda_n-\lambda_i)(\lambda_{i+1}-\beta_{i;j}) - (\lambda_{i+1}-\lambda_i)(\lambda_n-\beta_{i;j})}{(\lambda_n-\lambda_i)(\lambda_{i+1}-\beta_{i;j}) + (\lambda_{i+1}-\lambda_i)(\lambda_n-\beta_{i;j})} \\ &= \frac{(\lambda_n-\lambda_{i+1})(\lambda_i-\beta_{i;j})}{(\lambda_n-\lambda_i)(\lambda_{i+1}-\beta_{i;j}) + (\lambda_{i+1}-\lambda_i)(\lambda_n-\beta_{i;j})} \\ &< \frac{(\lambda_n-\lambda_{i+1})(\lambda_i-\beta_{i;j})}{2(\lambda_n-\lambda_i)(\lambda_{i+1}-\lambda_i)} < \frac{\lambda_i-\beta_{i;j}}{2(\lambda_{i+1}-\lambda_i)}.\end{aligned}$$

Combining the above estimates of Δ, $\theta_{i;j}\epsilon_{i;j}$ and τ, we have

$$\Delta + \tau\sqrt{\theta_{i;j}\epsilon_{i;j}} < \frac{\lambda_i-\beta_{i;j}}{2(\lambda_{i+1}-\lambda_i)} + \sqrt{\Delta_{i;j}}. \tag{4.11}$$

By Theorem 3.1(a) and the result (b) of this theorem, the upper bound of (4.11) converges to zero as $j \to \infty$. □

Four remarks are in order.

Remark 4.1. By the definition of the shift $\beta_{i;j}$ in (4.2) and the inequality (4.6) on the upper bound of $c\|r_{i;j}\|_{S^{-1}}$, the shift

$$\beta_{i;j} = \lambda_{i;j} + \mathcal{O}((\lambda_{i;j}-\lambda_i)^{\frac{1}{2}}). \tag{4.12}$$

Since the preconditioning equation is solved approximately, the shift $\beta_{i;j}$ is chosen to be $\lambda_{i;j}$ in practice. We call the preconditioner

$$\widetilde{K}_{i;j} = (H-\lambda_{i;j}S)^{-1}. \tag{4.13}$$

a *locally accelerated* preconditioner.

Remark 4.2. With the locally accelerated preconditioner $\widetilde{K}_{i;j}$, the corresponding search vector $\widetilde{p}_{i;j} = -\widetilde{K}_{i;j}r_{i;j}$. A direct calculation gives rise to

$$U^H S(u_{i;j}+\widetilde{p}_{i;j}) = \left[0 \quad \dots \quad 0 \quad \tfrac{\lambda_{i;j}-\beta_{i;j}}{\lambda_i-\beta_{i;j}}a_i \quad \dots \quad \tfrac{\lambda_{i;j}-\beta_{i;j}}{\lambda_n-\beta_{i;j}}a_n\right]^\top,$$

where we use the fact $U^HSu_{i;j} = a = [0,\dots,0,a_i,\dots,a_n]^\top$. By the facts $\lambda_{i;j}-\lambda_i = \mathcal{O}(\epsilon^2)$ and $\|r_{i;j}\|_{S^{-1}} = \mathcal{O}(\epsilon)$ as shown in Theorem 4.1(b), we have $\lambda_i-\beta_{i;j} = \lambda_i-\lambda_{i;j}+c\|r_{i;j}\|_2 = \mathcal{O}(\epsilon)$, and $\lambda_{i;j}-\beta_{i;j} = c\|r_{i;j}\|_{S^{-1}} = \mathcal{O}(\epsilon)$. Hence, we can conclude that $U^HS(u_{i;j}+\widetilde{p}_{i;j})$ converges to $e_i^\top$ directionally as $j\to\infty$. In the notion of an ideal search vector introduced at the end of section 2, the search vector $\widetilde{p}_{i;j}$ is an asymptotically ideal search vector.

Remark 4.3. Before $\lambda_{i;j}$ is localized, we can use a fixed preconditioner $K_{i;j} = K$ for all j. An obvious choice is to set $K_{i;j} \equiv K_\sigma = (H-\sigma S)^{-1}$ for some $\sigma < \lambda_1$. K_σ is symmetric positive definite and can be regarded as a global preconditioner for the initial few iterations. By the convergence of PSD-id (Theorem 3.1), it is guaranteed that the sequence $\{\lambda_{i;j}\}_j$ is strictly monotonically decreasing, albeit the convergence may be slow before the locally accelerated preconditioner $\widetilde{K}_{i;j}$ is applied, see the numerical illustration in section 5.

Remark 4.4. As we discussed in section 1, we are particularly interested in solving ill-conditioned Hermitian-definite generalized eigenvalue problem (1.1) where H and S sharing a common near-nullspace $\mathcal{V}$, whose dimension can be large. If we set the preconditioner $K_{i;j} \equiv I$, then $K_{i;j}M = M = SU_i^c(\Lambda_i^c - \lambda_i I)(U_i^c)^H S$, which has a near-nullspace $\mathcal{V}$, and a nullspace $\mathrm{span}(U_i)$. As $\dim(\mathcal{V}) > \dim(\mathrm{span}(U_i))$, $K_{i;j}M$ has very small positive eigenvalues. Therefore, $\Gamma/\gamma \gg 1$, and $\Delta \approx 1$. By Theorem 3.2, we know that the PSD-id method would converge linearly. By a similar arguments, we can declare that for any well-conditioned preconditioner $K_{i;j}$, the PSD-id method would also converge linearly. Therefore, in order to achieve a fast convergence, one has to apply an ill-conditioned preconditioner such as the locally accelerated preconditioner $\widetilde{K}_{i;j}$.

5. Numerical Examples

In this section, we use a MATLAB implementation for the PSD-id method (Algorithm 2.1) with locally accelerated preconditioners $\widetilde{K}_{i;j}$ defined in (4.13) to verify the convergence and the rate of convergence of the method discussed in Sections 3.1 and 3.2. To illustrate the efficiency of the method, we focus on two ill-conditioned generalized eigenvalue problems (1.1) arising from the PUFE approach to solve differential eigenvalue equations arising in quantum mechanics. MATLAB scripts of the implementation of the PSD-id method and the data that used to generate numerical results presented in this section can be obtained from the URL http://dsec.pku.edu.cn/∼yfcai/psdid.html.

To apply $\widetilde{K}_{i;j}$, we need to test the localization conditions (4.3a) and (4.3b) of the jth approximate eigenvalue $\lambda_{i;j}$. For the condition (4.3a), note that c is a constant and $\Delta_{i;j}$ in limit is zero. Therefore, when the residual $r_{i;j}$ is sufficiently small, $\lambda_{i;j}$ is close enough to λ_i, then the condition (4.3a) will be satisfied. Therefore, the test of the condition (4.3a) can be replaced by the following residual test:

$$\mathrm{Res}[\lambda_{i;j}, u_{i;j}] = \frac{\|Hu_{i;j} - \lambda_{i;j}Su_{i;j}\|}{\|Hu_{i;j}\| + |\lambda_{i;j}|\|Su_{i;j}\|} \le \tau_1, \tag{5.1}$$

where τ_1 is some prescribed threshold, say $\tau_1 = 0.1$.

For the condition (4.3b), we need the estimates of eigenvalues λ_i and λ_{i+1} to approximate the quantities $\Delta_i = (\lambda_i - \lambda_{i-1})/(\lambda_{i+1} - \lambda_i)$ and $\Delta_{i;j} = (\lambda_{i;j} - \lambda_i)/(\lambda_{i+1} - \lambda_{i;j})$. For Δ_i, it is natural to take the jth approximates $\lambda_{i;j}$ and $\lambda_{i+1;j}$ of λ_i and λ_{i+1} respectively and yields the following estimate of Δ_i

$$\Delta_i \approx \widehat{\Delta}_i = \frac{\lambda_{i;j} - \lambda_{i-1}}{\lambda_{i+1;j} - \lambda_{i;j}}.$$

For $\Delta_{i;j}$, if we simply use $\lambda_{i;j}$ to estimate λ_i, then it leads to $\Delta_{i;j} = 0$. This violates the condition (4.3b). A better estimate of λ_i is to use the linear extrapolation $\widehat{\lambda}_i = 2\lambda_{i;j} - \lambda_{i;j-1}$ of $\lambda_{i;j-1}$ and $\lambda_{i;j}$ for $j > 1$. Note that when $j = 1$, all approximated eigenvalues are assumed to be not localized. Then it yields the following estimate of $\Delta_{i;j}$:

$$\Delta_{i;j} \approx \widehat{\Delta}_{i;j} = \frac{\lambda_{i;j} - \widehat{\lambda}_i}{\lambda_{i+1;j} - \lambda_{i;j}} = \frac{\lambda_{i;j-1} - \lambda_{i;j}}{\lambda_{i+1;j} - \lambda_{i;j}}.$$

In order to estimate λ_{i+1}, the Rayleigh-Ritz projection subspace in PSD-id is spanned by $Z = [U_{i-1}\ u_{i;j}\ \ldots\ u_{i+\ell;j}\ p_{i;j}]$ with extra ℓ columns for some $\ell > 1$, say $\ell = 4$. For $k = i, i+1, \ldots, i+\ell$,

$(\lambda_{k;j+1}, u_{k;j+1})$ is updated as (γ_k, Zw_k), where (γ_k, w_k) is the kth eigenpair of $(Z^H HZ, Z^H SZ)$. Consequently, $\lambda_{i+1;j}$ can be used to approximate λ_{i+1}.

By the estimates $\widehat{\Delta}_i$ and $\widehat{\Delta}_{i;j}$, the localization condition (4.3b) of the jth approximate eigenvalue $\lambda_{i;j}$ of λ_i can be verified by the following condition

$$\widehat{\Delta}_{i;j} < \min\left\{\frac{1}{4}\widehat{\Delta}_i^2, 0.1\right\} \equiv \tau_2. \tag{5.2}$$

Note that for computing the smallest eigenvalue λ_1, we let the initial approximate $\lambda_{0,j} = \sigma$ for some $\sigma < \lambda_1$. Here σ is a user given parameter or a lower bound of λ_1, say $\lambda_{1;j} - \|r_{1;j}\|_{S^{-1}} \approx \lambda_{1;j} - \|r_{1;j}\|$.

We use the preconditioned MINRES [21] to compute the preconditioned search vector

$$p_{i;j} = -\widetilde{K}_{i;j} r_{i;j} = -(H - \lambda_{i;j} S)^{-1} r_{i;j}. \tag{5.3}$$

In practice, the vector $p_{i;j}$ is just needed to be computed approximately such that

$$\|(H - \lambda_{i;j})p_{i;j} + r_{i;j}\| \le \eta_{i;j}\|r_{i;j}\|, \tag{5.4}$$

where $\eta_{i;j} < 1$ is a parameter. In our numerical experiments, the preconditioner of the MINRES is S^{-1}, $\eta_{i;j} = \mathrm{Res}[\lambda_{i;j}, u_{i;j}]$, and the maximum number of MINRES iterations is set to be 200.

All numerical experiments are performed on a quad-core Intel® Xeon® Processor E5-2643 running at 3.30GHz with 31.3GB RAM, machine epsilon $\varepsilon \approx 2.2 \times 10^{-16}$.

Example 5.1. Consider the following Schrödinger equation for a one-dimensional harmonic oscillator studied in quantum mechanics [6, 12]:

$$-\frac{1}{2}\psi''(x) + \frac{1}{2}x^2\psi(x) = E\psi(x), \quad -L \le x \le L \tag{5.5a}$$

$$\psi(-L) = \psi(L) = 0, \tag{5.5b}$$

where E is the energy, $\psi(x)$ is the wavefunction. If $L = \infty$, the eigenvalues of the equation (5.5a) are $\lambda_i = i - 0.5$ and the corresponding eigenfunctions are $\psi_i(x) = H_i(x)e^{-0.5x^2}$, where $H_i(x)$ is the ith order Hermite polynomial [18, Chap. 18].

For numerical experiments, we set $L = 10$ since $\psi_i(x)$ is numerically zero for $|x| > 10$. We discretize the equation (5.5a) by linear finite element (FE), cubic FE and partition of unit FE (PUFE) [14], respectively. In all three cases, the eigenfunction $\psi(x)$ is approximated by

$$\psi^h(x) = \sum_i c_i\phi_i(x) + \sum_\alpha \sum_j c_{j\alpha}\phi_j^{PU}(x)\tilde{\psi}_\alpha(x) \equiv \sum_{k=1}^{n} u_k \Phi_k(x), \tag{5.6}$$

where $\phi_i(x)$ are the FE basis functions, ϕ_j^{PU} are the FE basis function to form enriched basis functions, $\tilde{\psi}_\alpha(x)$ are enrichment functions, and c_i and $c_{j\alpha}$ are coefficients. The enrichment term vanishes in the linear and cubic FE cases. In our numerical experiments, the interval $[-10, 10]$ is divided uniformly, and for PUFE, $\phi_i(x)$, $\phi_j^{PU}(x)$ are chosen to be cubic and linear, respectively, and $\tilde{\psi}_\alpha(x) = e^{-0.4x^2}$ for $x \in [-5, 5]$ and zero elsewhere.

Converting (5.5a) into its weak form, and using Φ_i as the test functions, we obtain an algebraic generalized eigenvalue problem (1.1), where $u = [u_1\, u_2\, \ldots\, u_n]$ and (i, j) elements h_{ij} and s_{ij} of H and S are given by

$$h_{ij} = \int_{-10}^{10} (\Phi'(x)\Phi_j'(x) + \frac{1}{2}x^2\Phi_i(x)\Phi_j(x))dx \quad \text{and} \quad s_{ij} = \int_{-10}^{10} \Phi_i(x)\Phi_j(x)dx,$$

respectively. The left plot of Figure 5.1 shows the errors of the sums of the four smallest eigenvalues of (H, S) with respect to the number of FEs of three different finite element discretizations. The matrix sizes of linear FE are $7, 15, 31, 63, 127, 255$ and 511. The matrix sizes for the cubic FE are $23, 47, 95$ and 191. The matrix sizes for the PUFE are $28, 56$ 112. We can see that to achieve the same accuracy, the matrix sizes of the PUFE are much smaller. However, the condition numbers of PUFE matrices H, S are large; $(\kappa_2(H), \kappa_2(S)) = (3.0 \times 10^6, 5.0 \times 10^6), (6.5 \times 10^8, 2.7 \times 10^9), (8.8 \times 10^{10}, 8.1 \times 10^{11})$, respectively.

For demonstrating the convergence behavior of PSD-id, let us compute $m = 4$ smallest eigenvalues of the PUFE matrices H and S of order $n = 112$, which corresponds to the mesh size $h = 2L/32$. The matrices H and S are ill-conditioned, $(\kappa_2(H), \kappa_2(S)) = (8.8 \times 10^{10}, 8.1 \times 10^{11})$. Furthermore, H and S share a common near-nullspace, namely there exists a subspace span(V) of dimension 17 such that $\|HV\| = \|SV\| = \mathcal{O}(10^{-5})$. To compute 4 smallest eigenpairs, we run the PSD-id algorithm for $i = 1, 2, 3, 4$ with $\ell = 4$. The accuracy threshold of computed eigenvalues is $\tau_{\text{eig}} = 10^{-9}$. $\tau_1 = 0.1$ is used for the residual test (5.1).

The right plot of Figure 5.1 shows the convergence history in the relative residuals Res$[\lambda_{i;j}, u_{i;j}]$ of the PSD-id method for computing four smallest eigenvalues. The localization (i.e., the conditions (5.1) and (5.2) are satisfied) of the j approximate eigenpair $(\lambda_{i;j}, u_{i;j})$ for computing the ith eigenvalue λ_i is marked by "+" sign. The locally accelerated preconditioner $\widetilde{K}_{i;j} = (H - \lambda_{i;j} S)^{-1}$ is used once $\lambda_{i;j}$ is localized. As Theorem 4.1 predicts, the locally accelerated preconditioner $\widetilde{K}_{i;j}$ is asymptotically optimal and leads to superlinear convergence of the PSD-id algorithm.

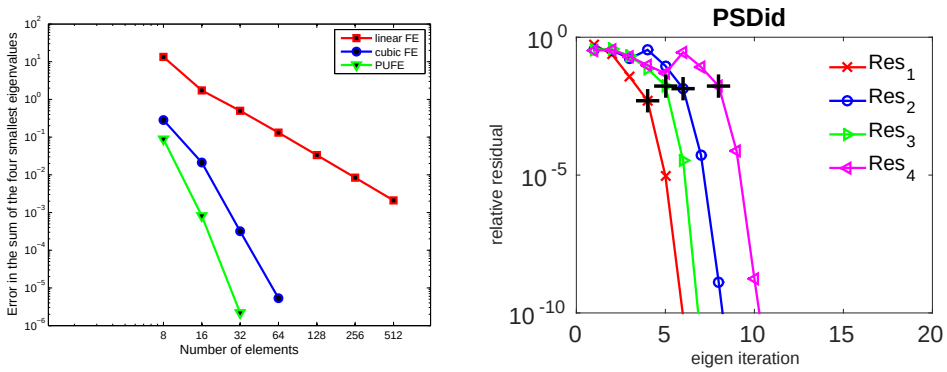

Fig. 5.1. Left: error of the sum of four smallest eigenvalues of (H, S) with respect to the number of FEs of three different FE discretizatoins in Example 5.1. Right: convergence of the PSD-id method for computing four smallest eigenvalues.

Example 5.2. The Hermitian-definite generalized eigenvalue problem (1.1) is a computational kernel in quantum mechanical methods employing a nonorthogonal basis for *ab initio* three-dimensional electronic structure calculations, see [3] and references therein. In this example, we select a sequence of eigenproblems produced by the PUFE method for a self-consistent pseudopotential density functional calculation for metallic, triclinic CeAl [22, 23, 26]. The Brillouin zone is sampled at two $\mathbf{k}$-points: $\mathbf{k} = (0.00, 0.00, 0.00)$ and $\mathbf{k} = (0.12, -0.24, 0.37)$. The PUFE approximation for the wavefunction is of the form given in the equation (5.6) and we apply a standard Galerkin procedure to set up the discrete system matrices. The unit cell is a triclinic

box, with atoms displaced from ideal positions. The primitive lattice vectors and the position of the atomic centers are

$$\mathbf{a}_1 = a(1.00 \quad 0.02 \quad -0.04), \quad \mathbf{a}_2 = a(0.01 \quad 0.98 \quad 0.03), \quad \mathbf{a}_3 = a(0.03 \quad -0.06 \quad 1.09)$$

and

$$\tau_{\mathrm{Ce}} = a(0.01 \quad 0.02 \quad 0.03), \quad \tau_{\mathrm{Al}} = a(0.51 \quad 0.47 \quad 0.55),$$

with lattice parameter $a = 5.75$ bohr. Since Ce has a full complement of s, p, d, and f states in valence, it requires 17 enrichment functions to span the occupied space. The near-dependencies between the finite element basis functions and the enriched basis functions lead to an ill-conditioned generalized eigenvalue problem (1.1).

In this numerical example, the matrix size of H and S is $n = 7 \times 8^3 + 1752 = 5336$. Both H and S are ill conditioned and their condition numbers are $(\kappa_2(H), \kappa_2(S)) = (1.1641 \times 10^{10}, 2.5731 \times 10^{11})$. Furthermore, H and S share a common near-nullspace span(V) of dimension 1000 such that $\|HV\| = \|SV\| = O(10^{-4})$, where V is orthonormal. This is an extremely ill-conditioned eigenvalue problem. Figure 5.2 shows the convergence history of the PSD-id method for computing four smallest eigenvalues. As in Figure 5.1, the localization of the j approximate eigenpair $(\lambda_{i;j}, u_{i;j})$ is marked by "+" sign. Once $\lambda_{i;j}$ is localized, the locally accelerated preconditioner $\widetilde{K}_{i;j} = (H - \lambda_{i;j}S)^{-1}$ is used. Again, as Theorem 4.1 predicts, $\widetilde{K}_{i;j}$ leads to superlinear convergence of the PSD-id algorithm.

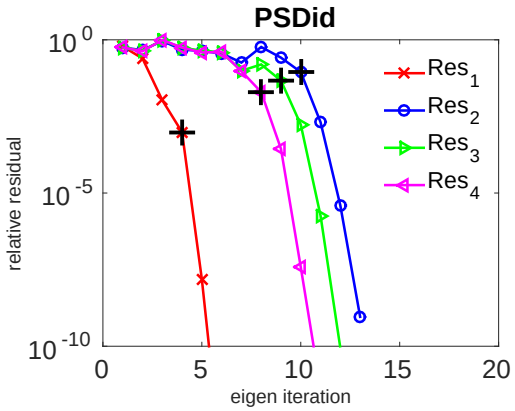

Fig. 5.2. Convergence of the PSD-id method for computing four smallest eigenvalues of the CeAl matrix pair described in Example 5.2.

6. Conclusion

We proved the convergence of the PSD-id method, and derived a nonasymptotic estimate of the rate of convergence of the method. We show that with the proper choice of the shift, the indefinite shift-and-invert preconditioner is a locally accelerated preconditioner and leads to superlinear convergence. Two numerical examples are presented to verify the theoretical results on the convergence behavior of the PSD-id method for solving ill-conditioned Hermitian-definite generalized eigenvalue problems.

Acknowledgments. The authors express their gratitude to the referees for their valuable comments and suggestions to improve the presentation of the paper. Cai was supported in part by NSFC grants 11301013, 11671023 and 11421101. Bai was supported in part by NSF grants DMS-1522697 and CCF-1527091.

References

[1] J.H. Bramble, J.E. Pasciak, and A.V. Knyazev. A subspace preconditioning algorithm for eigenvector/eigenvalue computation. *Adv. Comput. Math.*, **6**:1 (1996), 159–189.

[2] B.R. Brooks, D. Janežič, and M. Karplus. Harmonic analysis of large systems. I. methodology. *J. Comput. Chem.*, **16**:12 (1995), 1522–1542.

[3] Y. Cai, Z. Bai, J.E. Pask, and N. Sukumar. Hybrid preconditioning for iterative diagonalization of ill-conditioned generalized eigenvalue problems in electronic structure calculations. *J. Comput. Phys.*, **255** (2013), 16 –30.

[4] D.K. Faddeev and V.N. Faddeeva. *Computational Methods of Linear Algebra*. W. H. Freeman and Company, San Francisco and London, 1963.

[5] G. Fix and R. Heiberger. An algorithm for the ill-conditioned generalized eigenvalue problem. *SIAM J. Numer. Anal.*, **9**:1 (1972), 78–88.

[6] D.J. Griffiths. *Introduction to Quantum Mechanics (2nd Edition)*. Pearson Prentice Hall, 2004.

[7] M.R. Hestenes and W. Karush. A method of gradients for the calculation of the characteristic roots and vectors of a real symmetric matrix. *J. Res. Nat. Bur. Stand.*, **47**:1 (1951), 45–61.

[8] T. Kato. Upper and lower bounds of eigenvalues. *Phys. Rev.*, **77**:3 (1950), 413.

[9] A.V. Knyazev and K. Neymeyr. Efficient solution of symmetric eigenvalue problems using multigrid preconditioners in the locally optimal block conjugate gradient method. *Electron. Trans. Numer. Anal.*, **15** (2003), 38–55.

[10] M. Levitt. Private communication, January 2015.

[11] M. Levitt, C. Sander, and P.S. Stern. Protein normal-mode dynamics: trypsin inhibitor, crambin, ribonuclease and lysozyme. *J. Mol. Biol.*, **181**:3 (1985), 423–447.

[12] R.L. Liboff. *Introductory Quantum Mechanics (4th Edition)*. Addison-Wesley, 2003.

[13] D.E. Longsine and S.F. McCormick. Simultaneous rayleigh-quotient minimization methods for Ax= λBx. *Linear Algebra Appl.*, **34** (1980), 195–234.

[14] J.M. Melenk and I. Babuška. The partition of unity finite element method: basic theory and applications. *Comput. Methods Appl. Mech. Engrg.*, **139**:1 (1996), 289–314.

[15] K. Neymeyr and M. Zhou. The block preconditioned steepest descent iteration for elliptic operator eigenvalue problems. *Electron. Trans. Numer. Anal.*, **41** (2014), 93–108.

[16] K. Neymeyr and M. Zhou. Iterative minimization of the rayleigh quotient by block steepest descent iterations. *Numer. Linear Algebra Appl.*, **21**:5 (2014), 604–617.

[17] T. Nishikawa and N. Gō. Normal modes of vibration in bovine pancreatic trypsin inhibitor and its mechanical property. *Proteins: Struct., Funct., Bioinf.*, **2**:4 (1987), 308–329.

[18] F.W.J. Olver, D.W. Lozier, R.F. Boisvert and C.W. Clark. *NIST Handbook of Mathematical Functions*. Cambridge University Press, 2010.

[19] E. Ovtchinnikov. Cluster robustness of preconditioned gradient subspace iteration eigensolvers. *Linear Algebra Appl.*, **415**:1 (2006), 140–166.

[20] E.E. Ovtchinnikov. Sharp convergence estimates for the preconditioned steepest descent method for hermitian eigenvalue problems. *SIAM J. Numer. Anal.*, **43**:6 (2006), 2668–2689.

[21] C.C. Paige and M.A. Saunders. Solution of sparse indefinite systems of linear equations. *SIAM J. Numer. Anal.*, **12**:4 (1975), 617–629.

[22] J.E. Pask and N. Sukumar, Partition of unity finite element method for quantum mechanical materials calculations, Extreme Mech. Lett., Vol.11 (2017), 8–17.

[23] J.E. Pask, N. Sukumar, and S.E. Mousavi. Linear scaling solution of the all-electron Coulomb problem in solids. *Int. J. Multiscale Comput. Eng.*, **10**:1 (2012), 83–99.

[24] W.V. Petryshyn. On the eigenvalue problem $Tu - \lambda Su = 0$ with unbounded and nonsymmetric operators T and S. *Philos. Trans. R. Soc. Math. Phys. Sci.*, **262**:1130 (1986), 413–458.

[25] B.A. Samokish. The steepest descent method for an eigenvalue problem with semi-bounded operators. *Izv. Vyssh. Uchebn. Zaved. Mat.*, **5** (1958), 105–114.

[26] N. Sukumar and J.E. Pask. Classical and enriched finite element formulations for Bloch-periodic boundary conditions. *Int. J. Numer. Meth. Eng.*, **77**:8 (2009), 1121–1138.

[27] J.H. Wilkinson. *The Algebraic Eigenvalue Problem.* Clarendon Press, Oxford, 1965.