Backward Stability of Explicit External Deflation for the Symmetric Eigenvalue Problem

Chao-Ping Lin\textsuperscript{*}  Ding Lu\textsuperscript{†}  Zhaojun Bai\textsuperscript{‡}

May 5, 2021

Abstract

A thorough backward stability analysis of Hotelling’s deflation, an explicit external deflation procedure through low-rank updates for computing many eigenpairs of a symmetric matrix, is presented. Computable upper bounds of the loss of the orthogonality of the computed eigenvectors and the symmetric backward error norm of the computed eigenpairs are derived. Sufficient conditions for the backward stability of the explicit external deflation procedure are revealed. Based on these theoretical results, the strategy for achieving numerical backward stability by dynamically selecting the shifts is proposed. Numerical results are presented to corroborate the theoretical analysis and to demonstrate the stability of the procedure for computing many eigenpairs of large symmetric matrices arising from applications.

1 Introduction

Let $A$ be an $n \times n$ real symmetric matrix with ordered eigenvalues $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$, and the corresponding eigenvectors $v_1, v_2, \ldots, v_n$. Let $\mathcal{I} = [\lambda_{\text{low}}, \lambda_{\text{upper}}]$ be an interval containing the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_{n_e}$ at the lower end of the spectrum of $A$. We are interested in computing the partial eigenvalue decomposition

$$AV_{n_e} = V_{n_e} \Lambda_{n_e},$$

where $\Lambda_{n_e} = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_{n_e})$, $V_{n_e} = [v_1, v_2, \ldots, v_{n_e}]$ and $V_{n_e}^T V_{n_e} = I_{n_e}$. We are particularly interested in the case where the interval is of the width $|\mathcal{I}| = \lambda_{\text{upper}} - \lambda_{\text{low}} \leq \frac{1}{2} \|A\|_2$, and may contain a large number of eigenvalues of $A$.

A majority of subspace projection methods for computing the partial eigenvalue decomposition (1.1), such as the Lanczos methods \cite{13, 1}, typically converge first to the eigenvalues on the periphery of the spectrum. A projection subspace of a large dimension is necessary to compute the eigenvalues located deep inside the spectrum. Maintaining the orthogonality of a large projection subspace is computationally expensive. To reduce the dimensions of projection subspaces and accelerate the convergence, a number of techniques such as the restarting \cite{17, 8, 20} and the filtering \cite{11} have been developed.

This paper revisits a classical deflation technique known as Hotelling’s deflation \cite{6, 19, 13}. Hotelling’s deflation computes the partial eigenvalue decomposition (1.1) through a sequence of low-rank updates. In the simplest case, supposing that the lowest eigenpair $(\lambda_1, v_1)$ of $A$ is computed

\textsuperscript{*} Department of Mathematics, University of California, Davis, CA 95616, USA, cplin@ucdavis.edu
\textsuperscript{†} Department of Mathematics, University of Kentucky, Lexington, KY 40506, USA, Ding.Lu@uky.edu
\textsuperscript{‡} Department of Mathematics and Department of Computer Science, University of California, Davis, CA 95616, USA, zbai@ucdavis.edu
by an eigensolver, by choosing a real shift $\sigma_1$ and defining the rank-one update matrix
$$A_1 \equiv A + \sigma_1 v_1 v_1^T,$$
the eigenpair $(\lambda_1, v_1)$ of $A$ is displaced to an eigenpair $(\lambda_1 + \sigma_1, v_1)$ of $A_1$, while all the other eigenpairs of $A$ and $A_1$ are the same. If the shift $\sigma_1 > \lambda_{\text{upper}} - \lambda_1$, then the eigenpair $(\lambda_1, v_1)$ is shifted outside the interval $I$ and the eigenpair $(\lambda_2, v_2)$ of $A$ becomes the lowest eigenpair of $A_1$. Subsequently, we can use the eigensolver again to compute the lowest eigenpair $(\lambda_2, v_2)$ of $A_1$. The procedure is repeated until all the eigenvalues in the interval $I$ are found. Since the low-rank update of the original $A$ can be done outside of the eigensolver, we refer to this strategy as an explicit external deflation (EED) procedure, to distinguish from the deflation techniques used inside an eigensolver, such as TRLan [20] and ARPACK [9]. Recently the EED has been combined with Krylov subspace methods to progressively compute the eigenpairs toward the interior of the spectrum of symmetric eigenvalue problems [12, 21], and extended to structured eigenvalue problems [2, 3].

In the presence of finite precision arithmetic, the EED procedure is susceptible to numerical instability due to the accumulation of numerical errors from previous computations; see, e.g., [19, p. 585] and [15, p. 96]. In [13, Chap. 5.1], Parlett showed that the change to the smallest eigenvalue in magnitude by deflating out the largest computed eigenvalue would be at the same order of the round-off error incurred. In [14], for nonsymmetric matrices, Saad derived an upper bound on the backward error norm and showed that the stability is determined by the angle between the computed eigenvectors and the deflated subspace. Saad concluded that “if the angle between the computed eigenvector and the previous invariant subspace is small at every step, the process may quickly become unstable. On the other hand, if this is not the case then the process is quite safe, for small number of requested eigenvalues.”

However, none of the existing studies had examined the effect of the choice of shifts on the numerical stability. In this paper, we present a thorough backward stability analysis of the EED procedure. We first derive governing equations of the EED procedure in the presence of finite precision arithmetic, and then derive upper bounds of the loss of the orthogonality of the computed eigenvectors and the symmetric backward error norm of the computed eigenpairs. From these upper bounds, under mild assumptions, we conclude that the EED procedure is backward stable if the shifts $\sigma_j$ are dynamically chosen such that (i) the spectral gaps, defined as the separation between the computed eigenvalues and shifted ones, are maintained at the same order of the norm of the matrix $A$, and (ii) the shift-gap ratios, defined as the ratios between the largest shift in magnitude and the spectral gaps, are at the order of one. Our analysis implies that the re-orthogonalization of computed eigenvectors and the large angles between the computed eigenvectors and the previous invariant subspaces are unnecessary. Based on theoretical analysis, we provide a practical shift selection scheme to guarantee the backward stability. For a set of test matrices from applications, we demonstrate that the EED in combination with an established eigensolver can compute a large number of eigenvalues with backward stability.

The rest of the paper is organized as follows. In Section 2, we present the governing equations of the EED in finite precision arithmetic. In Section 3, we derive the upper bounds on the loss of orthogonality of computed eigenvectors and the symmetric backward error norm of computed eigenpairs. From these upper bounds, we derive conditions for the backward stability of the EED procedure. In Section 4, we discuss how to properly choose the shifts to satisfy the stability conditions and outline an algorithm for combining the EED procedure with an eigensolver. Numerical results to verify the sharpness of the upper bounds and to demonstrate the backward stability of the EED for symmetric eigenvalue problems from applications are presented in Section 5. Concluding remarks are given in section 6.
Following the convention of matrix computations, we use the upper case letters for matrices and the lower case letters for vectors. In particular, we use $I_n$ for the identity matrix of dimension $n$. If not specified, the dimensions of matrices and vectors conform to the dimensions used in the context. We use $.^T$ for transpose, and $\| \cdot \|_2$ and $\| \cdot \|_F$ for 2-norm and the Frobenius norm, respectively. The range of a matrix $A$ is denoted by $\mathcal{R}(A)$. We also use $\sigma_{\text{min}}(\cdot)$ to denote the minimal singular values of a matrix. Other notations will be explained as used.

2 EED in finite precision arithmetic

Suppose that we use an established eigensolver called EIGSOL, such as TRLan [20] and ARPACK [9], to successfully compute the lowest eigenpair $(\hat{\lambda}, \hat{v})$ of $A$ with

$$A\hat{v} = \hat{\lambda}\hat{v} + \eta,$$

where $\|\hat{v}\|_2 = 1$ and the residual vector $\eta$ satisfies

$$\|\eta\|_2 \leq \text{tol} \cdot \|A\|_2 \quad (2.1)$$

for a prescribed convergence tolerance $\text{tol}$. The EED procedure starts with computing the lowest eigenpair $(\hat{\lambda}_1, \hat{v}_1)$ of $A$ by EIGSOL satisfying

$$A\hat{v}_1 = \hat{\lambda}_1\hat{v}_1 + \eta_1,$$

where $\hat{\lambda}_1 \in \mathcal{I} = [\lambda_{\text{low}}, \lambda_{\text{upper}}]$, and the residual vector $\eta_1$ satisfies (2.1). At the first EED step, we choose a shift $\sigma_1$ and define

$$\hat{A}_1 \equiv A + \sigma_1 \hat{v}_1\hat{v}_1^T.$$

By choosing the shift $\sigma_1 > \lambda_{\text{upper}} - \hat{\lambda}_1$, the lowest eigenpair of $\hat{A}_1$ is an approximation of the second eigenpair $(\lambda_2, v_2)$ of $A$. Subsequently, we use EIGSOL to compute the lowest eigenpair $(\hat{\lambda}_2, \hat{v}_2)$ of $\hat{A}_1$ satisfying

$$\hat{A}_1\hat{v}_2 = \hat{\lambda}_2\hat{v}_2 + \eta_2,$$

where the residual vector $\eta_2$ satisfies (2.1). Meanwhile, expressing the computed eigenpair $(\hat{\lambda}_1, \hat{v}_1)$ in terms of $\hat{A}_1$, we have

$$\hat{A}_1\hat{v}_1 = (\hat{\lambda}_1 + \sigma_1)\hat{v}_1 + \eta_1.$$

Proceeding to the second EED step, we choose a shift $\sigma_2$ and define

$$\hat{A}_2 \equiv \hat{A}_1 + \sigma_2 \hat{v}_2\hat{v}_2^T = A + \hat{V}_2\Sigma_2\hat{V}_2^T,$$

where $\hat{V}_2 = [\hat{v}_1, \hat{v}_2]$ and $\Sigma_2 = \text{diag}(\sigma_1, \sigma_2)$. By choosing the shift $\sigma_2 > \lambda_{\text{upper}} - \hat{\lambda}_2$, the lowest eigenpair of $\hat{A}_2$ is an approximation of the third eigenpair $(\lambda_3, v_3)$ of $A$. Then we use EIGSOL again to compute the lowest eigenpair $(\hat{\lambda}_3, \hat{v}_3)$ of $\hat{A}_2$ satisfying

$$\hat{A}_2\hat{v}_3 = \hat{\lambda}_3\hat{v}_3 + \eta_3,$$

where the residual vector $\eta_3$ satisfies (2.1). Meanwhile, expressing the computed eigenpairs $(\hat{\lambda}_1, \hat{v}_1)$ and $(\hat{\lambda}_2, \hat{v}_2)$ in terms of $\hat{A}_2$, we have

$$\hat{A}_2\hat{V}_2 = \hat{V}_2(\hat{\Lambda}_2 + \Sigma_2) + \hat{V}_2\Sigma_2\Phi_2 + E_2,$$
where $\hat{\Lambda}_2 = \text{diag}(\hat{\lambda}_1, \hat{\lambda}_2)$, $E_2 = [\eta_1, \eta_2]$, and $\Phi_j \in \mathbb{R}^{2 \times 2}$ is the strictly lower triangular part of the matrix $\hat{V}_j^T \hat{V}_j - I_2$, i.e., $\Phi_2 + \Phi_2^T = \hat{V}_j^T \hat{V}_j - I_2$.

In general, at the $j$-th EED step, we choose a shift $\sigma_j$ and define

$$\hat{A}_j \equiv \hat{A}_{j-1} + \sigma_j \hat{v}_j \hat{v}_j^T = A + \hat{V}_j \Sigma_j \hat{V}_j^T,$$

where $\hat{V}_j \equiv [\hat{v}_1, \ldots, \hat{v}_j]$ and $\Sigma_j = \text{diag}(\sigma_1, \ldots, \sigma_j)$ with $A_0 \equiv A$. Then by choosing the shift $\sigma_j > \lambda_{\text{upper}} - \hat{\lambda}_j$, the lowest eigenpair of $\hat{A}_j$ is an approximation of the $(j + 1)$-th eigenpair $(\lambda_{j+1}, v_{j+1})$ of $A$. We use EIGSOL to compute the lowest eigenpair $(\hat{\lambda}_{j+1}, \hat{v}_{j+1})$ of $\hat{A}_j$ satisfying

$$\hat{A}_j \hat{v}_{j+1} = \hat{\lambda}_{j+1} \hat{v}_{j+1} + \eta_{j+1},$$

where $\|\hat{v}_{j+1}\|_2 = 1$ and the residual vector $\eta_{j+1}$ satisfies (2.1), i.e.,

$$\|\eta_{j+1}\|_2 \leq \text{tol} \cdot \|A\|_2.$$

Meanwhile, for the computed eigenpairs $(\hat{\lambda}_j, \hat{v}_j)$ in terms of $\hat{A}_j$, we have

$$\hat{A}_j \hat{v}_j = \hat{A}_{j-1} \hat{v}_j + \sigma_j \hat{v}_j = (\lambda_j + \sigma_j) \hat{v}_j + \eta_j,$$

and for the computed eigenpairs $(\hat{\lambda}_i, \hat{v}_i)$ with $1 \leq i \leq j - 1$ in terms of $\hat{A}_j$, we have

$$\hat{A}_j \hat{v}_i = \left( \hat{A}_{j-1} + \sigma_j \hat{v}_i \hat{v}_i^T \right) \hat{v}_i$$

$$= \left( \hat{A}_{i-1} + \sigma_i \hat{v}_i \hat{v}_i^T + \hat{V}_{i+1:j} \Sigma_{i+1:j} \hat{V}_{i+1:j}^T \right) \hat{v}_i$$

$$= \left( \hat{A}_{i-1} + \sigma_i \hat{v}_i \hat{v}_i^T \right) \hat{v}_i + \hat{V}_{i+1:j} \Sigma_{i+1:j} \hat{V}_{i+1:j}^T \hat{v}_i$$

$$= \hat{\lambda}_i \hat{v}_i + \eta_i + \sigma_i \hat{v}_i + \hat{V}_{i+1:j} \Sigma_{i+1:j} \hat{V}_{i+1:j}^T \hat{v}_i$$

$$= (\hat{\lambda}_i + \sigma_i) \hat{v}_i + \hat{V}_{i+1:j} \Sigma_{i+1:j} \hat{V}_{i+1:j}^T \hat{v}_i + \eta_i$$

$$= (\hat{\lambda}_i + \sigma_i) \hat{v}_i + \hat{V}_j \Sigma_j \begin{bmatrix} 0 \\ \hat{V}_{i+1:j}^T \hat{v}_i \end{bmatrix} + \eta_i,$$

where $\hat{V}_{i+1:j} \equiv [\hat{v}_{i+1}, \ldots, \hat{v}_j]$ and $\Sigma_{i+1:j} \equiv \text{diag}(\sigma_{i+1}, \ldots, \sigma_j)$.

Combining (2.5) and (2.6), we have

$$\hat{A}_j \hat{V}_j = \hat{V}_j (\hat{\Lambda}_j + \Sigma_j) + \hat{V}_j \Sigma_j \Phi_j + E_j,$$

where $\hat{\Lambda}_j = \text{diag}(\hat{\lambda}_1, \ldots, \hat{\lambda}_j)$, $E_j = [\eta_1, \ldots, \eta_j]$, $\Phi_j$ is the strictly lower triangular part of the matrix $\hat{V}_j^T \hat{V}_j - I_j$ and $\Phi_j + \Phi_j^T = \hat{V}_j^T \hat{V}_j - I_j$. Eqs. (2.3) and (2.7) are referred to as the governing equations of $j$ steps of (inexact) EED procedure.

For the backward stability analysis of the EED procedure, we introduce the following two quantities associated with the shifts $\sigma_1, \ldots, \sigma_j$ for a $j$-step EED:

- the spectral gap of $\hat{A}_j$, defined as the separation between the computed eigenvalues and the shifted ones:

$$\gamma_j \equiv \min_{\lambda, \theta \in I_{j+1}, \theta \in J_j} |\lambda - \theta| > 0,$$

where $I_{j+1} \equiv \{\hat{\lambda}_1, \ldots, \hat{\lambda}_j, \hat{\lambda}_{j+1}\}$, the set of computed eigenvalues, and $J_j \equiv \{\hat{\lambda}_1 + \sigma_1, \ldots, \hat{\lambda}_j + \sigma_j\}$, the set of computed eigenvalues after shifting (see Figure 1 for an illustration);
3 Backward stability analysis of EED

In this section, we first review the notion of backward stability for the computed eigenpairs. Then we derive upper bounds on the loss of orthogonality of computed eigenvectors and the backward error norm of computed eigenpairs by the EED procedure, and reveal conditions for the backward stability of the procedure.

3.1 Notion of backward stability

By the well-established notion of backward stability analysis in numerical linear algebra \[5, 18\], the following two quantities measure the accuracy of the approximate eigenpairs \((\hat{\Lambda}_{j+1}, \hat{V}_{j+1})\) of \(A\) computed by the EED procedure:

- the loss of orthogonality of the computed eigenvectors \(\hat{V}_{j+1}\),
  \[
  \omega_{j+1} \equiv \|\hat{V}_{j+1}^T \hat{V}_{j+1} - I_{j+1}\|_F, \tag{3.1}
  \]

- the symmetric backward error norm of the computed eigenpairs \((\hat{\Lambda}_{j+1}, \hat{V}_{j+1})\),
  \[
  \delta_{j+1} \equiv \min_{\Delta \in \mathcal{H}_{Q_{j+1}}} \|\Delta\|_F, \tag{3.2}
  \]

where \(\mathcal{H}_{Q_{j+1}}\) is the set of the symmetric backward errors for the orthonormal basis \(Q_{j+1}\) from the polar decomposition\(^1\) of \(\hat{V}_{j+1}\), namely,
\[
\mathcal{H}_{Q_{j+1}} \equiv \left\{ \Delta \mid (A + \Delta)Q_{j+1} = Q_{j+1}\hat{\Lambda}_{j+1}, \Delta = \Delta^T \in \mathbb{R}^{n \times n} \right\}. \tag{3.3}
\]

For a prescribed tolerance \(tol\) of the stopping criterion \((2.4)\) for an eigensolver EIGSOL, the EED procedure is considered to be backward stable if
\[
\omega_{j+1} = O(tol) \tag{3.4}
\]
and
\[
\delta_{j+1} = O(tol \cdot \|A\|_2), \tag{3.5}
\]
where the constants in the big-O notations are low-degree polynomials in the number \(j\) of the EED steps.

\(^1\)The polar decomposition of a matrix \(X\) is given by \(X = UP\) where \(U\) is orthonormal and \(P\) is symmetric positive semi-definite.
3.2 Loss of orthogonality

We first prove the following lemma to reveal the structure of the orthogonality between the computed eigenvectors.

**Lemma 3.1.** By the governing equations (2.3) and (2.7) of \( j \) steps of EED, if \( \tau_j \omega_j < \sqrt{2} \), then for \( i = 1, 2, \ldots, j \), the matrices \( \Gamma_i = \tilde{\Lambda}_i + \Sigma_i - \lambda_i + I_i \) and \( I_i + \Phi_i^T \Sigma_i \Gamma_i^{-1} \) are non-singular, and

\[
\hat{V}_i^T \hat{v}_{i+1} = \Gamma_i^{-1} (I_i + \Phi_i^T \Sigma_i \Gamma_i^{-1})^{-1} \left[ \hat{V}_i^T \eta_{i+1} - E_i^T \hat{v}_{i+1} \right].
\] (3.6)

Furthermore,

(i) \( \| \Gamma_i^{-1} \|_2 \leq \gamma_j^{-1} \),

(ii) \( \|(I_i + \Phi_i^T \Sigma_i \Gamma_i^{-1})^{-1}\|_2 \leq (1 - \tau_j \omega_j / \sqrt{2})^{-1} \),

where \( \gamma_j \) and \( \tau_j \) are the spectral gap and the shift-gap ratio defined in (2.8) and (2.9), respectively, and \( \omega_j \) is the loss of the orthogonality defined in (3.1).

**Proof.** By the governing equations (2.3) and (2.7) of \( j \) steps of EED, for \( 1 \leq i \leq j \), we have

\[
\hat{V}_i^T \hat{A}_i \hat{v}_{i+1} = \tilde{\Lambda}_{i+1} \hat{V}_i^T \hat{v}_{i+1} + \hat{V}_i^T \hat{\eta}_{i+1}
\]

and

\[
\hat{V}_i^T \hat{A}_i \hat{v}_{i+1} = (\tilde{\Lambda}_i + \Sigma_i) \hat{V}_i^T \hat{v}_{i+1} + \Phi_i^T \Sigma_i \hat{V}_i^T \hat{v}_{i+1} + E_i^T \hat{v}_{i+1}.
\]

Consequently,

\[
(\Gamma_i + \Phi_i^T \Sigma_i) \hat{V}_i^T \hat{v}_{i+1} = \hat{V}_i^T \hat{\eta}_{i+1} - E_i^T \hat{v}_{i+1},
\] (3.7)

where \( \Gamma_i = \tilde{\Lambda}_i + \Sigma_i - \lambda_i + I_i \) is a diagonal matrix.

By the definition (2.8) of the spectral gap \( \gamma_j \),

\[
\sigma_{\min}(\Gamma_i) = \min_{1 \leq k \leq i} |\lambda_k + \sigma - \lambda_{i+1}| \geq \gamma_j > 0.
\] (3.8)

Hence the matrix \( \Gamma_i \) is non-singular and the bound (i) holds.

Since \( \Phi_i \) is the strictly lower triangular part of the matrix \( \hat{V}_i^T \hat{V}_i - I_i \),

\[
\| \Phi_i^T \|_2 \leq \| \Phi_i^T \|_F = \frac{\omega_i}{\sqrt{2}} \leq \frac{\omega_j}{\sqrt{2}}.
\]

Consequently,

\[
\| \Phi_i^T \Sigma_i \Gamma_i^{-1} \|_2 \leq \| \Phi_i^T \|_2 \| \Gamma_i^{-1} \|_2 \| \Sigma_i \|_2 \leq \frac{\omega_j}{\sqrt{2}} \cdot \gamma_j^{-1} \cdot \| \Sigma_j \|_2 = \frac{\omega_j}{\sqrt{2}} \cdot \tau_j < 1,
\] (3.9)

where for the last inequality, we use the assumption \( \tau_j \omega_j < \sqrt{2} \). By (3.9), the matrix \( I_i + \Phi_i^T \Sigma_i \Gamma_i^{-1} \) is non-singular and the bound (ii) holds due to \( \|(I + X)^{-1}\|_2 \leq (1 - \|X\|_2)^{-1} \) for any \( X \) with \( \|X\|_2 < 1 \).

Since both matrices \( \Gamma_i \) and \( I_i + \Phi_i^T \Sigma_i \Gamma_i^{-1} \) are invertible, the identity (3.6) follows from (3.7). \( \square \)

Next we exploit the structure of the product \( \hat{V}_i^T \hat{v}_{i+1} \) to derive a computable upper bound on the loss of orthogonality \( \omega_{j+1} \) of computed eigenvectors \( \hat{V}_{j+1} \).
Theorem 3.1. By the governing equations \((2.3)\) and \((2.7)\) of \(j\) steps of EED, if \(\tau_j \omega_j < \sqrt{2}\), then the loss of orthogonality \(\omega_{j+1}\) of the computed eigenvectors \(\hat{V}_{j+1}\) defined in \((3.1)\) satisfies

\[
\omega_{j+1} \leq 2 \frac{c_j \gamma_j}{\gamma_j} \left( 1 + 2 \frac{c_j \gamma_j}{\gamma_j} \|E_{j+1}\|_F \right) \|E_{j+1}\|_F,
\]

where \(c_j = (1 - \tau_j \omega_j / \sqrt{2})^{-1}\), and \(\gamma_j\) and \(\tau_j\) are the spectral gap and the shift-gap ratio defined in \((2.8)\) and \((2.9)\), respectively.

Proof. By the definition \((3.1)\), we have

\[
\omega_{j+1}^2 = 2 \cdot \|\Phi_{j+1}^T\|_F^2 = 2 \cdot \sum_{i=1}^{j} \|\hat{V}_i^T \hat{v}_{i+1}\|_2^2.
\]

Recall Lemma 3.1 that, for any \(1 \leq i \leq j\),

\[
\|\hat{V}_i^T \hat{v}_{i+1}\|_2 \leq \frac{c_j}{\gamma_j} \cdot \|\hat{V}_i^T \hat{v}_{i+1} - E_i^T \hat{v}_{i+1}\|_2.
\]

Hence we can derive from \((3.11)\) that

\[
\omega_{j+1}^2 \leq \frac{2c_j^2}{\gamma_j^2} \cdot \sum_{i=1}^{j} \|\hat{V}_i^T \hat{v}_{i+1} - E_i^T \hat{v}_{i+1}\|_2 \leq \frac{2c_j^2}{\gamma_j^2} \cdot \|\hat{V}_{j+1}^T \hat{v}_{j+1} - E_{j+1}^T \hat{v}_{j+1}\|_F \leq \frac{4c_j^2}{\gamma_j^2} \cdot \|\hat{V}_{j+1}^T \|_2^2 \|E_{j+1}\|_F^2.
\]

Since

\[
\|\hat{V}_{j+1}^T \|_2 = \|\hat{V}_{j+1}^T \hat{V}_{j+1}\|_2 \leq \|I_{j+1}\|_2 + \|\hat{V}_{j+1}^T \hat{V}_{j+1} - I_{j+1}\|_2 \leq 1 + \omega_{j+1},
\]

we arrive at

\[
\omega_{j+1}^2 \leq \frac{4c_j^2}{\gamma_j^2} \cdot (1 + \omega_{j+1}) \cdot \|E_{j+1}\|_F^2.
\]

Letting \(t = \omega_{j+1} / \chi_{j+1}\), where \(\chi_{j+1} = 2c_j \|E_{j+1}\|_F / \gamma_j\), then the inequality \((3.13)\) is recast as

\[
t^2 - \chi_{j+1} t - 1 \leq 0.
\]

By the fact that the quadratic polynomial in \((3.14)\) is concave, we conclude that

\[
t \leq \frac{1}{2} \cdot \left( \chi_{j+1} + \sqrt{4 + \chi_{j+1}^2} \right) \leq \frac{1}{2} \cdot (\chi_{j+1} + 2 + \chi_{j+1}) \leq 1 + \chi_{j+1}.
\]

This proves the upper bound in \((3.10)\).

Remark 3.1. A straightforward way to derive an upper bound of \(\omega_{j+1}\) is to use the relation

\[
\omega_{j+1}^2 = \omega_j^2 + 2 \cdot \|\hat{V}_j^T \hat{v}_{j+1}\|_2^2,
\]
and the following bound from Lemma \ref{lem:3.1}
\[ \| \hat{V}_j^T \hat{e}_{j+1} \|_2^2 \leq \frac{c_j^2}{\tau_j} \left( \| \hat{V}_j \|_2 \| e_j \|_2 + \| E_j \|_F \right)^2 \approx \frac{c_j^2}{\gamma_j^2} (\| E_j+1 \|_F^2 + 2 \| e_{j+1} \|_2 \| E_j \|_F). \]

However, this would lead to a pessimistic upper bound of \( \omega_{j+1} \). In contrast, in Theorem \ref{thm:3.1} we exploit the structure of the orthogonality \( \hat{V}_j^T \hat{e}_{j+1} \) to obtain a tighter bound (3.10) of \( \omega_{j+1} \). Eq. (3.12) in the proof is the key step. In Example 5.1 in Section 5, we will demonstrate numerically that the bound (3.10) is indeed a tight upper bound on the loss of orthogonality \( \omega_{j+1} \).

### 3.3 Symmetric backward error norm

In this section, we derive a computable upper bound on the symmetric backward error norm \( \delta_{j+1} \) of computed eigenpairs \( (\hat{\Lambda}_{j+1}, \hat{V}_{j+1}) \) of \( A \) defined in (3.2). First, the following lemma gives an upper bound of the norm of the residual for \( (\hat{\Lambda}_{j+1}, \hat{V}_{j+1}) \):

\[ R_{j+1} = A \hat{V}_{j+1} - \hat{V}_{j+1} \hat{\Lambda}_{j+1}. \tag{3.15} \]

**Lemma 3.2.** By the governing equations (2.3) and (2.7) of \( j \) steps of EED, if \( \tau_j \omega_j \leq \sqrt{2} \), then for the computed eigenpairs \( (\hat{\Lambda}_{j+1}, \hat{V}_{j+1}) \) of \( A \), the Frobenius norm of the residual \( R_{j+1} \) defined in (3.15) satisfies

\[ \| R_{j+1} \|_F \leq \left( 1 + \sqrt{2} c_j \tau_j (1 + \omega_{j+1}) \right) \| E_{j+1} \|_F, \tag{3.16} \]

where \( c_j = (1 - \tau_j \omega_j / \sqrt{2})^{-1} \), and \( \gamma_j \) and \( \tau_j \) are the spectral gap and the shift-gap ratio defined in (2.8) and (2.9), respectively.

**Proof.** From the governing equation (2.7) of the EED procedure after \( j + 1 \) steps, we have

\[ \hat{A}_{j+1} \hat{V}_{j+1} = \hat{V}_{j+1} (\hat{\Lambda}_{j+1} + \mu_{j+1}) + \hat{V}_{j+1} \Sigma_{j+1} \Phi_{j+1} + E_{j+1}. \tag{3.17} \]

On the other hand, by the definition (2.2) of \( \hat{A}_{j+1} \), we have

\[ \hat{A}_{j+1} \hat{V}_{j+1} = A \hat{V}_{j+1} + \hat{V}_{j+1} \Sigma_{j+1} \hat{V}_j^T \hat{V}_{j+1} \]

\[ = A \hat{V}_{j+1} + \hat{V}_{j+1} \Sigma_{j+1} (\Phi_{j+1} + I_{j+1} + \Phi_{j+1}^T). \tag{3.18} \]

Combining (3.17) and (3.18), we obtain the residual

\[ R_{j+1} = A \hat{V}_{j+1} - \hat{V}_{j+1} \hat{\Lambda}_{j+1} = E_{j+1} - \hat{V}_{j+1} \Sigma_{j+1} \Phi_{j+1}^T. \tag{3.19} \]

Consequently, the norm of the residual \( R_{j+1} \) is bounded by

\[ \| R_{j+1} \|_F \leq \| E_{j+1} \|_F + \| \hat{V}_{j+1} \|_2 \| \Sigma_{j+1} \Phi_{j+1}^T \|_F. \tag{3.20} \]

Note that \( \Sigma_{j+1} = \text{diag}(\sigma_1, \ldots, \sigma_{j+1}) \) and \( \Phi_{j+1}^T \) is the strictly upper triangular part of the matrix \( \hat{V}_{j+1}^T \hat{V}_{j+1} - I_{j+1} \), and we have

\[ \| \hat{V}_{j+1} \|_2 \| \Sigma_{j+1} \Phi_{j+1}^T \|_F \leq \| \hat{V}_{j+1} \|_2 \| \Sigma_{j+1} \|_2 \| \Phi_{j+1}^T \|_F \]

\[ \leq \| \hat{V}_{j+1} \|_2 \| \Sigma_{j+1} \|_2 \frac{1}{\sqrt{2}} \omega_{j+1} \]

\[ \leq \frac{1}{\sqrt{2}} \| \Sigma_{j+1} \|_2 \sqrt{(1 + \omega_{j+1}) \omega_{j+1}^2}, \tag{3.21} \]
where, for the third inequality, we again use the fact that \( \| \hat{V}_{j+1} \|_2 \leq \sqrt{1 + \omega_{j+1}} \) by the definition of the loss of orthogonality \( \omega_{j+1} \). Left-multiplying (3.13) by \( 1 + \omega_{j+1} \), we know that

\[
(1 + \omega_{j+1})\omega_{j+1}^2 \leq \frac{4c_j^2}{\gamma_j^2} \cdot (1 + \omega_{j+1})^2 \cdot \| E_{j+1} \|_F^2.
\]

Plugging into (3.21), we obtain

\[
\| \hat{V}_{j+1} \|_2 \| \Sigma_{j+1} \Phi_{j+1}^T \|_F \leq \sqrt{2} \| \Sigma_j \|_2 c_j \gamma_j^{-1} \cdot (1 + \omega_{j+1}) \cdot \| E_{j+1} \|_F
\]

\[
= \sqrt{2} c_j \tau_j \cdot (1 + \omega_{j+1}) \cdot \| E_{j+1} \|_F.
\]

Combine with (3.20) and we arrive at the upper bound (3.16) of \( \| R_{j+1} \|_F \).

Now we recall the following theorem by Sun [18].

**Theorem 3.2** (Thm.3.1 in [18]). Let \((A, X)\) be approximate eigenpairs of a symmetric matrix \(A\). Let \(Q\) be the orthonormal basis from the polar decomposition of \(X\). Define the set

\[
\mathcal{H}_Q = \{ \Delta \mid (A + \Delta)Q = QA, \Delta = \Delta^T \in \mathbb{R}^{n \times n} \}. \tag{3.22}
\]

Then \(\mathcal{H}_Q\) is non-empty and there exists a unique \(\Delta_Q \in \mathcal{H}_Q\) such that

\[
\min_{\Delta \in \mathcal{H}_Q} \| \Delta \|_F = \| \Delta_Q \|_F \leq \sqrt{\| R \|_F^2 + \| P_X^\perp R \|_F^2} / \sigma_{\min}(X), \tag{3.23}
\]

where \(R = AX - XA\) is the residual and \(P_X^\perp\) is the orthogonal projection onto the orthogonal complement of \(R(X)\).

From Lemma 3.2 and Theorem 3.2, we have the following computable upper bound on the symmetric backward error norm \(\delta_{j+1}\) of the computed eigenpairs \((\hat{\Lambda}_{j+1}, \hat{V}_{j+1})\) of \(A\).

**Theorem 3.3.** By the governing equations (2.3) and (2.7) of \(j\) steps of EED, if \(\tau_j \omega_j < \sqrt{2}\) and \(\omega_{j+1} < 1\), then the symmetric backward error norm \(\delta_{j+1}\) of the computed eigenpairs \((\hat{\Lambda}_{j+1}, \hat{V}_{j+1})\) of \(A\) defined in (3.2) has the following upper bound

\[
\delta_{j+1} \leq \sqrt{2} \left( \frac{1 + c_j \tau_j (1 + \omega_{j+1})}{\sqrt{1 - \omega_{j+1}}} \right) \| E_{j+1} \|_F, \tag{3.24}
\]

where \(c_j = (1 - \tau_j \omega_j / \sqrt{2})^{-1}\), and \(\gamma_j\) and \(\tau_j\) are the spectral gap and the shift-gap ratio defined in (2.8) and (2.9), respectively.

**Proof.** For the computed eigenvectors \(\hat{V}_{j+1}\) of \(A\), let \(Q_{j+1}\) be the orthonormal basis from the polar decomposition of \(\hat{V}_{j+1}\), and the set \(\mathcal{H}_{Q_{j+1}}\) be defined in (3.3). It follows from the definition (3.2) and Theorem 3.2 that

\[
\delta_{j+1} = \min_{\Delta \in \mathcal{H}_{Q_{j+1}}} \| \Delta \|_F \leq \frac{1}{\sigma_{\min}(\hat{V}_{j+1})} \sqrt{\| R_{j+1} \|_F^2 + \| P^\perp R_{j+1} \|_F^2}, \tag{3.25}
\]

where \(R_{j+1}\) is the residual of \((\hat{\Lambda}_{j+1}, \hat{V}_{j+1})\) defined in (3.15), and \(P^\perp\) is the orthogonal projection onto the orthogonal complement of the subspace \(R(\hat{V}_{j+1})\), i.e.,

\[
P^\perp = I - \hat{V}_{j+1} (\hat{V}_{j+1}^T \hat{V}_{j+1})^{-1} \hat{V}_{j+1}^T.
\]
By the equation (3.19), \( P^\bot R_{j+1} = P^\bot E_{j+1} \). Hence we have
\[
\|P^\bot R_{j+1}\|_F = \|P^\bot E_{j+1}\|_F \leq \|E_{j+1}\|_F. \tag{3.26}
\]
On the other hand, by the definition (3.1) of \( \omega_{j+1} \) and the assumption \( \omega_{j+1} < 1 \), we have
\[
|\sigma^2_{\min}(\hat{V}_{j+1}) - 1| \leq \|\hat{V}_{j+1}^T \hat{V}_{j+1} - I_{j+1}\|_2 \leq \omega_{j+1} < 1,
\]
which implies the following lower bound of the singular value
\[
\sigma_{\min}(\hat{V}_{j+1}) \geq \sqrt{1 - \omega_{j+1}}. \tag{3.27}
\]
Plug (3.26) and (3.27) into (3.25) and recall the upper bound of \( \|R_{j+1}\|_F \) in Lemma 3.2, and then we obtain
\[
\delta_{j+1} \leq \sqrt{1 + \left(1 + \sqrt{2} c_j \tau_j (1 + \omega_{j+1})\right)^2} \cdot \frac{\|E_{j+1}\|_F}{\sqrt{1 - \omega_{j+1}}} \\
\leq \sqrt{2} \left(1 + c_j \tau_j (1 + \omega_{j+1})\right) \cdot \frac{\|E_{j+1}\|_F}{\sqrt{1 - \omega_{j+1}}},
\]
where the second inequality is due to \(1 + (1 + \sqrt{2}a)^2 \leq 2(1 + 2a + a^2) = 2(1 + a)^2\). This completes the proof.

\[\square\]

**Remark 3.2.** We note from the inequalities (3.25) and (3.27) that, when \( \omega_{j+1} \ll 1 \), the residual norm \( \|R_{j+1}\|_F \) can be used as an easy-to-compute estimate of \( \delta_{j+1} \). We will use this observation in numerical examples in Section 5.

### 3.4 Backward stability of EED

In Theorems 3.1 and 3.3, the upper bounds (3.10) and (3.24) for \( \delta_{j+1} \) and \( \omega_{j+1} \) involve the quantity \( \omega_j \) from the previous EED step. In this section, under a mild assumption, we derive explicit upper bounds for \( \omega_{j+1} \) and \( \delta_{j+1} \), and then reveal conditions for the backward stability of the EED procedure.

**Lemma 3.3.** Consider \( j \) steps of EED governed by Eqs. (2.3) and (2.7). Assume
\[
\tau_j \frac{\|A\|_2}{\gamma_j} \cdot 4\sqrt{j + 1} \cdot \text{tol} < 0.1. \tag{3.28}
\]
Then
(i) it holds that
\[
\tau_i \omega_i < 0.11 \quad \text{and} \quad c_i = (1 - \tau_i \omega_i/\sqrt{2})^{-1} < 2 \quad \text{for } i = 1, 2, \ldots, j; \tag{3.29}
\]
(ii) the loss of orthogonality \( \omega_{j+1} \) is bounded by
\[
\omega_{j+1} \leq \left(\frac{\|A\|_2}{\gamma_j} \cdot 5\sqrt{j + 1}\right) \cdot \text{tol}; \tag{3.30}
\]
(iii) the backward error norm \( \delta_{j+1} \) is bounded by
\[
\delta_{j+1} \leq \left(\tau_j \cdot 5\sqrt{j + 1}\right) \cdot \text{tol} \cdot \|A\|_2. \tag{3.31}
\]
Proof. First observe that by the definitions (2.8) and (2.9), \( \gamma_i \) is monotonically decreasing with the index \( i \) and \( \tau_i \geq 1 \) is monotonically increasing with \( i \). Therefore, the assumption (3.28) implies the inequalities

\[
\|A\|_2 \cdot \frac{4}{\gamma_i} \cdot \sqrt{i} \cdot tol < 0.1 \quad \text{and} \quad \tau_i \frac{\|A\|_2}{\gamma_{i-1}} \cdot \sqrt{i} \cdot tol < 0.1 \quad \text{for all} \quad i \leq j. \tag{3.32}
\]

Since the stopping criterion (2.4) of EIGSOL implies

\[
\|E_i\|_F = \|[\eta_1, \ldots, \eta_i]\|_F \leq \sqrt{i} \cdot tol \cdot \|A\|_2, \tag{3.33}
\]

inequalities (3.32) leads to

\[
\frac{4}{\gamma_i} \cdot \|E_{i+1}\|_F < 0.1 \quad \text{and} \quad \tau_i \frac{4}{\gamma_{i-1}} \cdot \|E_i\|_F < 0.1 \quad \text{for all} \quad i \leq j. \tag{3.34}
\]

(i) We prove the inequality (3.29) by induction. To begin with, recall that \( \|\tilde{v}_1\|_2 = 1 \), which implies \( \omega_1 = \|\tilde{v}_1^T \tilde{v}_1 - 1\|_F = 0, \tau_1 \omega_1 = 0 < 0.11 \), and \( c_1 = 1 < 2 \). Hence (3.29) holds for \( i = 1 \). Now, for \( 2 \leq i \leq j \), assume that \( \tau_{i-1} \omega_{i-1} < 0.11 \) and \( c_{i-1} < 2 \). Since \( \tau_{i-1} \omega_{i-1} < 0.11 \), we can apply Theorem 3.1 and derive from (3.10) that

\[
\tau_i \omega_i \leq \tau_i \frac{2c_{i-1}}{\gamma_{i-1}} \|E_i\|_F \cdot \left(1 + \frac{2c_{i-1}}{\gamma_{i-1}} \|E_i\|_F\right) < 0.1 \cdot (1 + 0.1) = 0.11, \tag{3.35}
\]

where the last inequality of (3.35) is by \( 2c_{i-1} < 4 \) and (3.34). This implies immediately

\[
c_i = (1 - \tau_i \omega_i / \sqrt{2})^{-1} \leq (1 - 0.11 / \sqrt{2})^{-1} < 2.
\]

Therefore, (3.29) follows by induction.

(ii) Since we have \( \tau_j \omega_j < 0.11 \) and \( c_j < 2 \) by (3.29), we can apply Theorem 3.1 and derive from (3.10) that

\[
\omega_{j+1} \leq \frac{2c_j}{\gamma_j} \cdot \|E_{j+1}\|_F \cdot \left(1 + \frac{2c_j}{\gamma_j} \|E_{j+1}\|_F\right) \leq \frac{4}{\gamma_j} \cdot \|E_{j+1}\|_F \cdot (1 + 0.1), \tag{3.36}
\]

where in the second inequality we used \( 2c_j < 4 \) and the first inequality in (3.34). Recall the error bound of \( \|E_{j+1}\|_F \) from (3.33) and we obtain (3.30).

(iii) We have \( \tau_j \omega_j < 0.11 \) and \( c_j < 2 \) by (3.29). It also follows from (3.36) and (3.34) that \( \omega_{j+1} < 0.11 \). Therefore, we can apply Theorem 3.3 and derive from (3.24) that

\[
\delta_{j+1} \leq \sqrt{2} \left(1 + \frac{c_j \tau_j (1 + \omega_{j+1})}{\sqrt{1 - \omega_{j+1}}}\right) \|E_{j+1}\|_F \leq \sqrt{2} \left(1 + 2 \tau_j (1 + 0.11)\right) \|E_{j+1}\|_F,
\]

where in second inequality we used \( 0 \leq \omega_{j+1} < 0.11 \). Since \( \tau_j \geq 1 \) by definition (2.9), we can relax the leading constant as \( \sqrt{2} (1.06 + 2.36 \cdot \tau_j) \leq \sqrt{2} (3.42 \cdot \tau_j) < 5 \tau_j \). Recall the error bound of \( \|E_{j+1}\|_F \) from (3.33) and we prove (3.31).

By the error bounds (3.30) and (3.31) in Lemma 3.3 we can see that the quantities \( \gamma_j^{-1} \|A\|_2 \) and \( \tau_j \) play important roles for the stability of the EED procedure. A sufficient condition to achieve the backward stability (3.4) and (3.5) is given by \( \gamma_j^{-1} \|A\|_2 = O(1) \) and \( \tau_j = O(1) \). In summary, we have the following theorem for the backward stability of the EED procedure.
Theorem 3.4. Under the assumptions of the residual norm \( \| \eta_i \|_2 \) of EIGSOL satisfying (2.1) and the inequality (3.28), the backward stability of the EED procedure, in the sense of (3.4) and (3.5), is guaranteed if the shifts \( \sigma_1, \ldots, \sigma_j \) are dynamically chosen such that
\[
\gamma_j^{-1} \| A \|_2 = O(1) \quad \text{and} \quad \tau_j = O(1).
\] (3.37)

We note that when the shifts \( \sigma_j \) are dynamically chosen such that the conditions (3.37) are satisfied, the assumption of the inequality (3.28) is indeed mild.

Remark 3.3. From the upper bound (3.30) of the loss of orthogonality \( \omega_{j+1} \), we see that if the spectral gap \( \gamma_j \) is too small, i.e., \( \gamma_j \ll \| A \|_2 \), then \( \omega_{j+1} \) could be amplified by a factor of \( \gamma_j^{-1} \| A \|_2 \). On the other hand, from the upper bound (3.31) of the symmetric backward error norm \( \delta_{j+1} \), we see that when \( \tau_j \) is too large, i.e., \( \tau_j \gg 1 \), \( \delta_{j+1} \) could be amplified by a factor of \( \tau_j \). We will demonstrate these observations numerically in Example 5.2 in Section 5.

4 EED in practice

In Section 3.4, we derived the required conditions (3.37) of the spectral gap \( \gamma_j \) and the shift-gap ratio \( \tau_j \) for the backward stability of the EED procedure. The quantities \( \gamma_j \) and \( \tau_j \) are controlled by the selection of the shifts \( \sigma_1, \ldots, \sigma_j \). In this section, we discuss a practical selection scheme of the shifts to satisfy the conditions (3.37), and then present an algorithm that combines an eigensolver EIGSOL with the EED procedure to compute eigenvalues in a prescribed interval \( I = [\lambda_{\text{low}}, \lambda_{\text{upper}}] \).

4.1 Choice of the shifts

We consider the following choice of the shift at the \( j \)-th EED step,
\[
\sigma_j = \mu - \hat{\lambda}_j,
\] (4.1)
where \( \mu \in \mathbb{R} \) is a parameter with \( \mu > \lambda_{\text{upper}} \). The shifting scheme (4.1) has been used in several previous works, although without elaboration on the choice of parameter \( \mu \); see [13, Chap.5.1] and [12, 21]. In the following, we will discuss how to choose the parameter \( \mu \) such that the conditions (3.37) can hold.

To begin with, the choice of the shifts in (4.1) implies that the spectral gap \( \gamma_j \) in (2.8) satisfies
\[
\gamma_j = \min_{\theta \in I_{j+1}, \lambda \in J_j} |\lambda - \theta| = \min_{1 \leq i \leq j+1} |\mu - \hat{\lambda}_i|,
\] (4.2)
where \( I_{j+1} = \{\hat{\lambda}_1, \ldots, \hat{\lambda}_j, \hat{\lambda}_{j+1}\} \) and \( J_j = \{\hat{\lambda}_1 + \sigma_1, \ldots, \hat{\lambda}_j + \sigma_j\} = \{\mu\} \). On the other hand, it also implies that the shift-gap ratio \( \tau_j \) in (2.9) satisfies
\[
\tau_j = \frac{1}{\gamma_j} \cdot \max_{1 \leq i \leq j} |\sigma_i| = \frac{\max_{1 \leq i \leq j+1} |\mu - \hat{\lambda}_i|}{\min_{1 \leq i \leq j+1} |\mu - \hat{\lambda}_i|}.
\] (4.3)
Now recall that \( \mu > \lambda_{\text{upper}} \) and the computed eigenvalues \( \hat{\lambda}_i \in [\lambda_{\text{low}}, \lambda_{\text{upper}}] \), for \( i = 1, 2, \ldots, j+1 \), so we have
\[
\mu - \lambda_{\text{upper}} \leq \min_{1 \leq i \leq j+1} |\mu - \hat{\lambda}_i| \leq \max_{1 \leq i \leq j+1} |\mu - \hat{\lambda}_i| \leq \mu - \lambda_{\text{low}}.
\]
Hence, (4.2) and (4.3) lead to
\[ \gamma_g \leq \gamma_j \leq \gamma_g \tau_g \quad \text{and} \quad \tau_j \leq \tau_g, \tag{4.4} \]
where
\[ \gamma_g \equiv \mu - \lambda_{\text{upper}} \quad \text{and} \quad \tau_g \equiv \frac{\mu - \lambda_{\text{low}}}{\mu - \lambda_{\text{upper}}}. \]
The problem is then turned to the choice of parameter \( \mu \) such that the quantities \( \gamma_g \) and \( \tau_g \) satisfy the conditions (3.37).

Let us consider a frequently encountered case in practice where the width of the interval \( I = [\lambda_{\text{low}}, \lambda_{\text{upper}}] \) satisfies
\[ \lambda_{\text{upper}} - \lambda_{\text{low}} \leq \frac{1}{2} \| A \|_2. \]
Then by setting
\[ \mu = \hat{\lambda}_1 + \| A \|_2, \]
we have
\[ \frac{1}{2} \| A \|_2 \leq \gamma_g = \left( 1 - \frac{\lambda_{\text{upper}} - \hat{\lambda}_1}{\| A \|_2} \right) \| A \|_2 \leq \| A \|_2 \]
and
\[ \tau_g = 1 + \frac{\lambda_{\text{upper}} - \lambda_{\text{low}}}{\gamma_g} \leq 2. \]
Consequently, by (4.4), the spectral gap \( \gamma_j \) and the shift-gap ratio \( \tau_j \) satisfy the desired conditions (3.37).

In summary, for the EED procedure in practice, we recommend the use of the following strategy for the choice of the shift \( \sigma_j \) at the \( j \)-th EED,
\[ \sigma_j = \mu - \hat{\lambda}_j \quad \text{with} \quad \mu = \hat{\lambda}_1 + \| A \|_2, \tag{4.5} \]
to compute the eigenvalues in an interval \( I = [\lambda_{\text{low}}, \lambda_{\text{upper}}] \), where \( \lambda_{\text{upper}} - \lambda_{\text{low}} \leq \frac{1}{2} \| A \|_2 \).

### 4.2 Algorithm

An eigensolver EIGSOL combined with the EED procedure to compute the eigenvalues in an interval \( I = [\lambda_{\text{low}}, \lambda_{\text{upper}}] \) at the lower end of the spectrum of \( A \) is summarized in Algorithm 1.

**Algorithm 1** An eigensolver EIGSOL combined with the EED procedure

**Input:** the interval \( I = [\lambda_{\text{low}}, \lambda_{\text{upper}}] \) at the lower end of the spectrum of \( A \); the relative tolerance \( tol \) in (2.1) for EIGSOL.

**Output:** the approximate eigenpairs \((\hat{\lambda}_i, \hat{v}_i)\) in the interval \( I \).

1. \( \hat{A}_0 = A \);
2. use EIGSOL to compute the lowest eigenpair \((\hat{\lambda}_1, \hat{v}_1)\) of \( \hat{A}_0 \) and an estimate \( \text{Anorm} \) of \( \| A \|_2 \);
3. \( \mu = \hat{\lambda}_1 + \text{Anorm} \);
4. **for** \( j = 1, 2, \ldots \) **do**
5. \( \sigma_j = \mu - \hat{\lambda}_j \);
6. \( \hat{A}_j = \hat{A}_{j-1} + \sigma_j \hat{v}_j \hat{v}_j^T = A + \hat{V}_j \Sigma_j \hat{V}_j^T \);
7. compute the lowest eigenpair \((\hat{\lambda}_{j+1}, \hat{v}_{j+1})\) of \( \hat{A}_j \) by EIGSOL;
8. check if all the eigenpairs in the interval \( I \) have been computed;
9. **end for**
10. return the approximate eigenpairs \((\hat{\lambda}_i, \hat{v}_i)\) in the interval \( I \);

A few remarks are in order:

- In practice, we never need to form the matrix \( \hat{A}_j \) at step 6 explicitly. We can assume that the only operation that is required by EIGSOL is the matrix-vector product \( y := \hat{A}_j x \).

- At step 7, the computation of the lowest eigenpair \((\hat{\lambda}_{j+1}, \hat{v}_{j+1})\) of \( \hat{A}_j \) can be accelerated by warm starting the EIGSOL with the lowest unconverged eigenvectors of \( \hat{A}_{j-1} \). This is possible for most iterative eigensolvers, such as TRLan [20] and ARPACK [9].

- At step 8, an ideal validation method is to use the inertias of the shifted matrix \( A - \lambda_{\text{upper}} I \). However, computation of the inertias is a prohibitive cost for large matrices. An empirical validation is to monitor the lowest eigenvalue \( \hat{\lambda}_{j+1} \) of \( \hat{A}_j \). All eigenpairs in the interval \( I \) are considered to be found when \( \hat{\lambda}_{j+1} \) is outside the interval \( I \).

5 Numerical experiments

In this section, we first use synthetic examples to verify the sharpness of the upper bounds (3.10) and (3.24) on the loss of orthogonality and the symmetric backward error norm of the EED procedure under the choice (4.5) of the shifts \( \sigma_j \). We present the cases where improper choices of the shifts \( \sigma_j \) may lead to numerical instability of the EED procedure. Then we demonstrate the numerical stability of the EED procedure for a set of large sparse symmetric matrices arising from applications.

We use TRLan as the eigen solver. TRLan is a C implementation of the thick-restart Lanczos method with adaptive sizes of the projection subspace [20, 22, 23]. The convergence criterion of an approximate eigenpair \((\hat{\lambda}_{j+1}, \hat{v}_{j+1})\) is the residual norm satisfying

\[
\| \eta_{j+1} \|_2 = \| \hat{A}_j \hat{v}_{j+1} - \hat{\lambda}_{j+1} \hat{v}_{j+1} \|_2 < \text{tol} \cdot \text{Anorm},
\]

where \( \text{tol} \) is a user-specified tolerance and \( \text{Anorm} \) is a 2-norm estimate of \( A \) computed by TRLan. The starting vector is a random vector.

**Example 5.1.** In this example, we demonstrate the sharpness of the upper bounds (3.10) and (3.24) on the loss of orthogonality and the symmetric backward error norm with the choice (4.5) of the shifts \( \sigma_j \).

We consider a diagonal matrix \( A \) with diagonal elements

\[
a_{kk} = \begin{cases} 
\frac{1}{2} d_k, & \text{if } 1 \leq k \leq n/2, \\
\frac{1}{2} (1 + d_k - n/2), & \text{if } n/2 < k \leq n,
\end{cases}
\]

where \( d_k = 10^{-5(1 - \frac{k-1}{n-1})} \) and the matrix size \( n = 500 \). The spectrum range of \( A \) is \((0, 1]\). The eigenvalues of \( A \) are clustered around 0 and 0.5. We are interested in computing the \( n_e = 65 \) eigenvalues in the interval \( I = [0, 10^{-4}] \). The computed 2-norm of \( A \) is \( \text{Anorm} = 1.00 \).

To closely observe the convergence, TRLan is slightly modified so that the convergence test is performed at each Lanczos iteration. The maximal dimension \( m \) of the projection subspace is set to 40.

Numerical results of TRLan with the EED procedure for computing all the \( n_e \) eigenvalues in the interval \( I \) are summarized in Table 1, where the 4th column is the loss of orthogonality \( \omega_{n_e} \), the 5th column is the upper bound (3.10) of \( \omega_{n_e} \), the 6th column is the residual norm \( \| R_{n_e} \|_F \) and
Table 1: Numerical stability of TRLan with EED for different tolerances \( tol \) (Example 5.1)

| \( tol \) | \( \mu \) | \( \gamma_g \) | \( \omega_{ne} \) | bound (3.10) | \( \| R_{ne} \|_F \) | bound (3.24) |
|-----------|---------|-------------|---------------|-------------|----------------|-------------|
| \( 10^{-6} \) | 1.00 | 1.00 | \( 2.37 \cdot 10^{-6} \) | \( 1.59 \cdot 10^{-5} \) | \( 7.87 \cdot 10^{-6} \) | \( 2.24 \cdot 10^{-5} \) |
| \( 10^{-8} \) | 1.00 | 1.00 | \( 1.78 \cdot 10^{-8} \) | \( 1.58 \cdot 10^{-7} \) | \( 7.95 \cdot 10^{-8} \) | \( 2.24 \cdot 10^{-7} \) |
| \( 10^{-10} \) | 1.00 | 1.00 | \( 1.82 \cdot 10^{-10} \) | \( 1.58 \cdot 10^{-9} \) | \( 7.94 \cdot 10^{-10} \) | \( 2.24 \cdot 10^{-9} \) |

Figure 2: The loss of orthogonality \( \omega_{j+1} \) and the upper bound (3.10) of \( \omega_{j+1} \) against the computed eigenvalues \( \hat{\lambda}_{j+1} \) for \( 2 \leq j+1 \leq n_e \), \( tol = 10^{-8} \) (Example 5.1).

From Table 1, we observe that with the choice (4.5) of the shifts \( \sigma_j \), \( \gamma_g^{-1} \text{Anorm} \approx 1 \) and \( \tau_g \approx 1 \). Therefore, the conditions (3.37) of the spectral gap \( \gamma_j \) and the shift-gap ratio \( \tau_j \) for the backward stability are satisfied. Consequently, the loss of orthogonality of the computed eigenvectors is \( \omega_{ne} = O(tol) \) and the symmetric backward error norm of the computed eigenpairs \( (\hat{A}_{ne}, \hat{V}_{ne}) \) is \( \delta_{ne} = O(tol \cdot \text{Anorm}) \). In addition, we observe that the upper bounds (3.10) and (3.24) of \( \omega_{ne} \) and \( \delta_{ne} \) are tight within an order of magnitude.

**Example 5.2.** In this example we illustrate that improperly chosen shifts \( \sigma_j \) may lead to instability of the EED procedure.

In Remark 3.3, we stated that if the shifts \( \sigma_j \) are chosen such that the spectral gaps \( \gamma_j \) are too small, i.e., \( \gamma_j \ll \text{Anorm} \), then the loss of orthogonality of the computed eigenvectors could be amplified by a factor of \( \gamma_j^{-1} \cdot \text{Anorm} \). Here is an example using the same diagonal matrix \( A \) in Example 5.1. The combination of TRLan and EED is used to compute the \( n_e = 65 \) eigenvalues in the interval \( \mathcal{I} = [0, 10^{-4}] \). Let us set the shifts \( \sigma_j = \mu - \hat{\lambda}_j \) with \( \mu = 2 \cdot 10^{-4} \), which is much smaller than the recommended value of \( \mu = \hat{\lambda}_1 + \| A \|_2 \approx 1.00 \). Numerical results are summarized in Table 2, where the tolerance \( tol = 10^{-8} \) for TRLan. We observe that \( \gamma_j = O(\gamma_g) \ll \text{Anorm} \), and the loss of orthogonality of the computed eigenvectors is indeed amplified by a factor of \( \gamma_j^{-1} \cdot \text{Anorm} \). We note that since \( \tau_j = O(1) \), the symmetric backward error norms \( \delta_{ne} = O(tol \cdot \text{Anorm}) \).

In Remark 3.3, we also stated that if the shifts \( \sigma_j \) are chosen such that the shift-gap ratios \( \tau_j \) are too large, then the symmetric backward error norm \( \delta_j \) could be amplified by a factor of \( \tau_j \). Here is an example. We flip the sign of the diagonal elements of \( A \) defined in Example 5.1, and set...
Table 2: Instability of TRLan with EED when the spectral gaps $\gamma_j = O(\gamma_g)$ are too small.

| $tol$  | $\mu$   | $\gamma_g$ | $\omega_{ne}$ | bound (3.10) | $\|R_{ne}\|_F$ | bound (3.24) |
|--------|---------|-------------|----------------|--------------|----------------|--------------|
| $10^{-6}$ | 2 \cdot 10^{-4} | 10^{-4} | 8.26 \cdot 10^{-3} | 1.79 \cdot 10^{-4} | 8.00 \cdot 10^{-6} | 3.29 \cdot 10^{-5} |
| $10^{-8}$ | 2 \cdot 10^{-4} | 10^{-4} | 8.28 \cdot 10^{-5} | 1.53 \cdot 10^{-3} | 7.96 \cdot 10^{-8} | 3.22 \cdot 10^{-7} |
| $10^{-10}$ | 2 \cdot 10^{-4} | 10^{-4} | 8.27 \cdot 10^{-7} | 1.52 \cdot 10^{-5} | 7.95 \cdot 10^{-10} | 3.22 \cdot 10^{-9} |

Figure 3: The loss of orthogonality $\omega_{j+1}$ (left) and the residual norm $\|R_{j+1}\|_F$ (right) against the computed eigenvalues $\hat{\lambda}_{j+1}$ for $2 \leq j + 1 \leq n_e$. The red lines are $tol$ (left) and $tol \cdot \text{Anorm}$ (right).

$n = 200$. We compute $n_e = 74$ eigenvalues in the interval $\mathcal{I} = [-1.0, -0.5001]$ using TRLan with EED procedure. The computed 2-norm of $A$ is $\text{Anorm} = 1.00$.

Instead of the choice (4.5) for the shifts $\sigma_j$, we set $\sigma_j = \mu - \hat{\lambda}_j$ with $\mu = -0.5$. The blue $\times$-lines in Figure 3 are the loss of orthogonality and the residual norms for the computed eigenpairs $(\hat{\lambda}_{j+1}, \hat{v}_{j+1})$ for $2 \leq j + 1 \leq n_e$. As we observe that for the first 6 computed eigenvalues in the subinterval $[-1.0, -0.75]$ of $\mathcal{I}$, since the spectral gap $\gamma_j \geq 0.25$ and the shift-gap ratio $\tau_j \leq 2$, the computed eigenpairs are backward stable with $\omega_0 = 2.48 \cdot 10^{-9} = O(tol)$ and $\|R_0\|_F = 9.05 \cdot 10^{-9} = O(tol \cdot \text{Anorm})$. However, for the computed eigenvalues in the subinterval $[-0.75, -0.5001]$ of $\mathcal{I}$, the computed eigenpairs are not backward stable due to the facts that the spectral gaps $\gamma_j$ become small, $\gamma_j \approx 1.03 \cdot 10^{-4}$, and the shift-gap ratios $\tau_j$ grows up to $\tau_j \approx 4.86 \cdot 10^3$. Consequently, the loss of orthogonality $\omega_{ne}$ and the residual norm $\|R_{ne}\|_F$ are increased by a factor of up to $10^3$, respectively. The stability are restored if the shifts are chosen according to the recommendation (4.5) as shown by the green $\pm$-lines in Figure 3.

Example 5.3. In this example, we demonstrate the numerical stability of TRLan with the EED procedure for a set of large sparse symmetric matrices from applications.

The statistics of the matrices are summarized in Table 3, where $n$ is the size of the matrix, $\text{nnz}$ is the number of nonzero entries of the matrix, $[\lambda_{\text{min}}, \lambda_{\text{max}}]$ is the spectrum range, and $n_e$ is the number of eigenvalues in the interval $\mathcal{I} = [\lambda_{\text{low}}, \lambda_{\text{upper}}]$. The quantities $n_e$ are calculated by computing the inertias of the shifted matrices $A - \lambda_{\text{upper}}I$. The Laplacian is the negative 2D Laplacian on a 200-by-200 grids with Dirichlet boundary condition [16]. The worms20 is the graph Laplacian worms20_10NN in machine learning datasets [4]. The Si0, Si34H36, Ge87H76 and Ge99H100 are Hamiltonian matrices from PARSEC collection [4].

We run TRLan with a maximal number $m$ of Lanczos vectors to compute the lowest eigenpairs of the matrix $A_j$. The convergence test is performed at each restart of TRLan. All the converged eigenvalues in the interval $\mathcal{I}$ are shifted by EED. Meanwhile, we also keep a maximal number $m_0$
of the lowest unconverged eigenvectors as the starting vectors of TRLan for the matrix $\tilde{A}_{j+1}$. All the eigenvalues in $I$ are assumed to be computed when the lowest converged eigenvalue is outside the interval $I$. This combination of TRLan is referred to as TRLED.

TRLED is compiled using the icc compiler (version 2021.1) with the optimization flag -O2, and linked to BLAS and LAPACK available in Intel Math Kernel Library (version 2021.1.1). The experiments are conducted on a MacBook with 1.6 GHz Intel Core i5 CPU and 8GB of RAM.

For numerical experiments, we set the maximal number of Lanczos vectors $m = 150$. When starting TRLED for $\tilde{A}_{j+1}$, the maximal number of the starting vectors is $n_0 = 75$. The convergence tolerance for the residual norm was set to $tol = 10^{-8}$ as a common practice for solving large scale eigenvalue problems with double precision; see for example [11].

Numerical results of TRLED are summarized in Table 4, where the 2nd column is the number $\tilde{n}_e$ of the computed eigenpairs $(\tilde{\lambda}_e, \tilde{\omega}_e)$ in the interval $I$, the 3rd column is the number $j_{\max}$ of steps of EED performed, the 4th column is the loss of orthogonality $\omega_{\tilde{n}_e}$, and the 5th column is the relative residual norm $\|R_{\tilde{n}_e}\|_F/Anorm$ of the computed eigenpairs $(\Lambda_{n_e}, \tilde{V}_{n_e})$. From the quantities $n_e$ in Table 3 and $\tilde{n}_e$ in Table 4, we see that for all test matrices the eigenvalues in the prescribed intervals $I$ are successfully computed with the desired backward stability.

The left plot of Figure 4 is a profile of the number of converged eigenvalues at each external deflation of a total of 74 EEDs for the matrix Ge99H100. The right plot of Figure 4 shows the relative residual norms of all 372 computed eigenpairs in the interval. We observe that a large number of converged eigenvalues are deflated and shifted away at some EED steps.

To examine whether the multiple explicit external deflations lead to a significant increase in execution time, in the 6th and 7th columns of Table 4 we record the CPU time of TRLED and TRLan for computing all eigenvalues in the same intervals. For TRLan, we set the maximal number of Lanczos vectors to $n_e + 150$. The restart scheme with restart=1 is used. TRLan is compiled and executed under the same setting as TRLED. We observe comparable execution time of TRLED and TRLan. In fact, TRLED is slightly faster. We think this might be due to the facts that TRLED

| matrix   | $n$   | nz    | $[\lambda_{\min}, \lambda_{\max}]$ | $[\lambda_{\text{low}}, \lambda_{\text{upper}}]$ | $n_e$ |
|----------|-------|-------|------------------------------------|-----------------------------------------------|------|
| Laplacian| 40,000| 199,200| $[0, 7.9995]$                      | $[0, 0.07]$                                   | 205  |
| worms20  | 20,055| 260,881| $[0, 6.0450]$                      | $[0, 0.05]$                                   | 289  |
| Si0      | 33,401| 1,317,655| $[-1.6745, 84.3139]$               | $[-1.7, 2.0]$                                 | 182  |
| Si34H36  | 97,569| 5,156,379| $[-1.1586, 42.9396]$               | $[-1.2, 0.4]$                                 | 310  |
| Ge87H76  | 112,985| 7,892,195| $[-1.214, 32.764]$                | $[-1.3, -0.0053]$                            | 318  |
| Ge99H100 | 112,985| 8,451,395| $[-1.226, 32.703]$                | $[-1.3, -0.0096]$                            | 372  |

| matrix   | $\tilde{n}_e$ | $j_{\max}$ | $\omega_{\tilde{n}_e}$ | $\|R_{\tilde{n}_e}\|_F/Anorm$ | CPU time (sec.) |
|----------|---------------|-------------|------------------------|-------------------------------|-----------------|
| Laplacian| 205           | 60          | $1.93 \cdot 10^{-8}$  | $6.33 \cdot 10^{-8}$          | 66.5            | 86.0 |
| worms20  | 289           | 86          | $2.63 \cdot 10^{-8}$  | $7.24 \cdot 10^{-8}$          | 57.3            | 74.8 |
| Si0      | 182           | 41          | $2.33 \cdot 10^{-8}$  | $4.71 \cdot 10^{-8}$          | 42.4            | 47.1 |
| Si34H36  | 310           | 72          | $3.41 \cdot 10^{-8}$  | $7.50 \cdot 10^{-8}$          | 309.9           | 310.4|
| Ge87H76  | 318           | 66          | $4.08 \cdot 10^{-8}$  | $8.50 \cdot 10^{-8}$          | 388.7           | 421.0|
| Ge99H100 | 372           | 74          | $3.65 \cdot 10^{-8}$  | $7.63 \cdot 10^{-8}$          | 501.1           | 533.4|
uses fewer Lanczos vectors, which compensates the cost of the matrix-vector products \( y := \hat{A}_j x \) and the warm start with those unconverged eigenvectors from the previous deflated matrix \( \hat{A}_{j-1} \). These are subjects of future study.

6 Concluding remarks

Based on the governing equations of the EED procedure in finite precision arithmetic, we derived the upper bounds on the loss of the orthogonality of the computed eigenvectors and the symmetric backward error norm of the computed eigenpairs in terms of two key quantities, namely the spectral gaps and the shift-gap ratios. Consequently, we revealed the required conditions of these two key quantities for the backward stability of the EED procedure. We present a practical strategy on the dynamically selection of the shifts such that the spectral gaps and the shift-gap ratios satisfy the required conditions.

There are a number of topics that can be explored further. One of them is the acceleration effect of the warm starting the EIGSOL with the unconverged eigenvectors of \( \hat{A}_{j-1} \) for computing the lowest eigenpairs of \( \hat{A}_j \). Another one is how to efficiently perform the matrix-vector products \( y := \hat{A}_j x \) in EIGSOL by exploiting the sparse plus low rank structure of the matrix \( \hat{A}_j \). Communication-avoid algorithms for sparse-plus-low-rank matrix-vector products can be found in [10, 7]. A preliminary numerical study for high performance computing is reported in [21].

References

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

[2] Z. Bai, R.-C. Li, and W. Lin. Linear response eigenvalue problem solved by extended locally optimal preconditioned conjugate gradient methods. Sci. China Math., 259:1443–1460, 2016.

[3] P.-Y. Chen, B. Zhang, and M. Al Hasan. Incremental eigenpair computation for graph laplacian matrices: theory and applications. Social Network Analysis and Mining, 8(1):4, 2018.
[4] T. Davis and Y. Hu. The University of Florida sparse matrix collection. *ACM Trans. Math. Software (TOMS)*, 38(1):1–25, 2011.

[5] D. Higham and N. Higham. Structured backward error and condition of generalized eigenvalue problems. *SIAM J. Mat. Anal. Appl.*, 20(2):493–512, 1998.

[6] H. Hotelling. Analysis of a complex of statistical variables into principal components. *J. of Educational Psychology*, 24(6):417, 1933.

[7] N. Knight, E. Carson, and J. Demmel. Exploiting data sparsity in parallel matrix powers computations. In *the proceedings of the international conference on parallel processing and applied mathematics (PPAM)*, 2014.

[8] R. Lehoucq and D. Sorensen. Deflation techniques for an implicitly restarted Arnoldi iteration. *SIAM J. Matrix Anal. Appl.*, 17(4):789–821, 1996.

[9] R. Lehoucq, D. Sorensen, and C. Yang. *ARPACK Users’ Guide: Solution of Large-Scale Eigenvalue Problems with Implicitly Restarted Arnoldi Methods*. SIAM, Philadelphia, PA, 1998.

[10] C. Leiserson, S. Rao, and S. Toledo. Efficient out-of-core algorithms for linear relaxation using blocking covers. *J. Comput. Syst. Sci. Int.*, 54:332–344, 1997.

[11] R. Li, Y. Xi, E. Vecharynskio, C. Yang, and Y. Saad. A thick-restart Lanczos algorithm with polynomial filtering for Hermitian eigenvalue problems. *SIAM J. Sci. Comput.*, 38:A2512–A2534, 2016.

[12] J. Money and Q. Ye. Algorithm 845: EIGIFP: a MATLAB program for solving large symmetric generalized eigenvalue problems. *ACM Trans. Math. Software (TOMS)*, 31(2):270–279, 2005.

[13] B. N. Parlett. *The Symmetric Eigenvalue Problem*. SIAM, Philadelphia, PA, 1998.

[14] Y. Saad. Numerical solution of large nonsymmetric eigenvalue problems. *Comput. Phys. Commun.*, 53(1-3):71–90, 1989.

[15] Y. Saad. *Numerical Methods for Large Eigenvalue Problems: revised edition*, volume 66. SIAM, Philadelphia, PA, 2011.

[16] B. Smith and A. Knyazev. *Laplacian in 1D, 2D or 3D*, 2015 (accessed March 24, 2020). [https://www.mathworks.com/matlabcentral/fileexchange/27279-laplacian-in-1d-2d-or-3d](https://www.mathworks.com/matlabcentral/fileexchange/27279-laplacian-in-1d-2d-or-3d).

[17] Danny C Sorensen. Implicit application of polynomial filters in a k-step Arnoldi method. *SIAM J. Matrix. Anal. Appl.*, 13(1):357–385, 1992.

[18] J.-G. Sun. A note on backward perturbations for the Hermitian eigenvalue problem. *BIT Numer. Math.*, 35(3):385–393, 1995.

[19] J. H. Wilkinson. *The algebraic eigenvalue problem*. Clarendon Press, Oxford, 1965.

[20] K. Wu and H. Simon. Thick-restart Lanczos method for large symmetric eigenvalue problems. *SIAM J. Matrix Anal. Appl.*, 22(2):602–616, 2000.
[21] I. Yamazaki, Z. Bai, D. Lu, and J. Dongarra. Matrix powers kernels for thick-restart Lanczos with explicit external deflation. In *2019 IEEE International Parallel Distributed Processing Symposium (IPDPS)*, pages 472–481. IEEE, 2019.

[22] I. Yamazaki, Z. Bai, H. Simon, L.-W. Wang, and K. Wu. Adaptive projection subspace dimension for the thick-restart Lanczos method. *ACM Trans. Math. Software (TOMS)*, 37(3):27, 2010.

[23] I. Yamazaki, K. Wu, and H. Simon. nu-TRLan User Guide. Technical report, LBNL-1288E, 2008. Software https://codeforge.lbl.gov/projects/trlan/.