SIAM J. MATRIX ANAL. APPL.
Vol. 34, No. 2, pp. 392–416

MINIMIZATION PRINCIPLES FOR THE LINEAR RESPONSE EIGENVALUE PROBLEM II: COMPUTATION*

ZHAOJUN BAI† AND REN-CANG LI‡

Abstract. In Part I of this paper we presented minimization principles and related theoretical results for the linear response eigenvalue problem. Here we develop best approximations for the few smallest eigenvalues with the positive sign via a structure-preserving subspace projection. Then we present four-dimensional subspace search conjugate gradient-like algorithms for simultaneously computing these eigenvalues and their associated eigenvectors. Finally, we present numerical examples to illustrate convergence behaviors of the proposed methods with and without preconditioning.

Key words. eigenvalue, eigenvector, minimization principle, random phase approximation, quantum linear response

AMS subject classifications. 65L15, 15A18, 81Q15

DOI. 10.1137/110838972

1. Introduction.

This is the second paper of ours in the sequel. Building upon the theoretical results in [2], here we will focus on the numerical aspect of the linear response (LR) eigenvalue problem:

$$Hz \equiv \begin{bmatrix} 0 & K \\ M & 0 \end{bmatrix} \begin{bmatrix} y \\ x \end{bmatrix} = \lambda \begin{bmatrix} y \\ x \end{bmatrix} \equiv \lambda z, \tag{1.1}$$

where K and M are $n \times n$ symmetric positive semidefinite matrices and one of them is definite. It is an equivalent problem obtained from the original LR (a.k.a. random phase approximation (RPA)) eigenvalue problem:

$$\begin{bmatrix} A & B \\ -B & -A \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = \lambda \begin{bmatrix} u \\ v \end{bmatrix} \tag{1.2}$$

by an orthogonal similarity transformation to give $K = A - B$ and $M = A + B$, where A and B are $n \times n$ real symmetric matrices such that the symmetric matrix $\left[\begin{smallmatrix} A & B \\ B & A \end{smallmatrix}\right]$ is positive definite [36, 45].

It is easy to see that the eigenvalue problem (1.1) is equivalent to any one of the following product eigenvalue problems:

$$KMy = \lambda^2 y \quad \text{and} \quad MKx = \lambda^2 x. \tag{1.3}$$

Their equivalences have led to solving (1.1) through attempting to solve one of the eigenvalue problems in (1.3). They also imply that the eigenvalues of H come in $\pm\lambda$ pairs. As in [2], we will denote the eigenvalues with the positive sign[1] of H by λ_i

*Received by the editors June 29, 2011; accepted for publication (in revised form) by M. Hochbruck January 18, 2013; published electronically April 25, 2013.
http://www.siam.org/journals/simax/34-2/83897.html

†Department of Computer Science and Department of Mathematics, University of California, Davis, CA 95616 (bai@cs.ucdavis.edu). This author was supported in part by NSF grants DMR-1035468 and DMS-1115817 and DOE grant DE-FC02-06ER25794.

‡Department of Mathematics, University of Texas at Arlington, Arlington, TX 76019 (rcli@uta.edu). This author was supported in part by NSF grants DMS-0810506 and DMS-1115834.

[1]Note our convention of assigning the positive sign to half of the eigenvalues 0 and the negative sign to the other half in [2].

$(1 \le i \le n)$ and

$$-\lambda_n \le \cdots \le -\lambda_2 \le -\lambda_1 \le \lambda_1 \le \lambda_2 \le \cdots \le \lambda_n.$$

An important minimization principle due to Thouless [44] is that when both K and M are symmetric positive definite, we have

$$\lambda_1 = \min_{u,v} \varrho(u, v), \tag{1.4}$$

where $\varrho(u, v)$ is defined by

$$\varrho(u, v) = \frac{\begin{bmatrix} u \\ v \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} A & B \\ B & A \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix}}{|u^{\mathrm{T}}u - v^{\mathrm{T}}v|}, \tag{1.5}$$

and the minimization is taken among all vectors u, v such that $u^{\mathrm{T}}u - v^{\mathrm{T}}v \neq 0$. For H, as shown in [48], this minimization principle translates into

$$\lambda_1 = \min_{x,y} \rho(x, y), \tag{1.6}$$

where

$$\rho(x, y) = \frac{x^{\mathrm{T}}Kx + y^{\mathrm{T}}My}{2|x^{\mathrm{T}}y|}, \tag{1.7}$$

where the minimization is taken among all x and y such that either $x^{\mathrm{T}}y \neq 0$ or $x^{\mathrm{T}}y = 0$ but $x^{\mathrm{T}}Kx + y^{\mathrm{T}}My > 0$. This removes those x and y that annihilate both the numerator and the denominator from the domain. In particular $x = y = 0$ is excluded. In [2], the minimization principle (1.6) is extended to include the case where K and M are symmetric positive semidefinite and one of them is definite.

Since the RPA and linear response theory was proposed by Bohm and Pines for studying the collective motion of many particles in the early 1950s [6], the development of numerical methods for solving the eigenvalue problem (1.2) and equivalently (1.1) has been an active research subject in computational (quantum) physics and chemistry for over four decades. In [10], it was suggested to solve the equivalent product eigenvalue problems (1.3) instead by converting it to the symmetric eigenvalue problem of $R^{\mathrm{T}}KR$ using the Cholesky decomposition of $M = R^{\mathrm{T}}R$. In [30, 41], Davidson's algorithm for the symmetric eigenvalue problem was extended to the large scale eigenvalue problem (1.2). In [16, 17, 47, 48, 49], Lanczos-like algorithms were proposed. Given the minimization principles (1.4) and (1.6), conjugate gradient (CG) methods become natural choices for finding the smallest positive eigenvalue and indeed they have; see, e.g., [9, 26, 28].

Meanwhile, the eigenvalue problems in the forms of (1.1), (1.3), and (1.2) have also attracted a great deal of attention in numerical analysis community over the past four decades. As early as in the 1960s, Wilkinson discussed the product eigenvalue problems (1.3) and proposed the method of transforming them to the standard symmetric eigenvalue problems by using the Cholesky decomposition [52, p. 35, p. 337]. Wilkinson's method is implemented as LAPACK's routine `xSYGVD` [1], where the product eigenvalue problems in (1.3) are classified as *the types* 2 *and* 3 of the generalized symmetric definite eigenvalue problems, respectively. Alternatively the structure-preserving GR algorithm, a generalization of the well-known QR algorithm,

can also be used for small to medium size problems [51, Chapter 8]. For large and sparse cases, the Lanczos algorithm, Krylov–Schur algorithm and Jacobi–Davidson algorithm all have been generalized to the product eigenvalue problems [19, 23, 50]. On the other hand, since the RPA eigenvalue problem (1.2) is a special case of the *Hamiltonian matrix eigenvalue problem*, an extension of the QR algorithm made for Hamiltonian matrix eigenvalue problems can be used to solve the problems of small to medium sizes [5, 8, 13, 51]. In particular, the work [13] treated a more general linear response eigenvalue problem. Algorithms for large scale Hamiltonian eigenvalue problems can be found in [4, 3] and the references therein. An RPA test case is given in [4] to illustrate the computational efficiency of a Hamiltonian Krylov–Schur-type algorithm.

A recent survey study [46] compared four numerical methods (namely Lanczos, Arnoldi, Davidson, and CG) and discussed the limitations of each of these methods for developing an efficient linear-scaling eigensolver for the RPA eigenvalue problem (1.2). In the study, severe limitations were experienced for the Lanczos-type methods due to the orthogonality constraints (also see [47]), for the CG type methods to compute several eigenpairs simultaneously, and for incorporating preconditioning techniques (see also [26]).

In [2], we obtained a trace (or subspace) version of (1.6):

$$\sum_{i=1}^{k} \lambda_i = \frac{1}{2} \inf_{U^{\mathrm{T}}V=I_k} \mathrm{trace}(U^{\mathrm{T}}KU + V^{\mathrm{T}}MV), \tag{1.8}$$

as well as Cauchy-like interlacing inequalities for a structure-preserving projection H_{SR} of H. Based on this newly developed theory, we are presented with an opportunity to develop efficient numerical methods for the LR eigenvalue problem (1.1) in much the same way as the conjugate gradient and Lanczos methods for solving the large scale symmetric eigenvalue problem. In this paper, we will show an important computational implication of the minimization principle (1.8) that is that it lends itself to seek approximations to a cluster of smallest eigenvalues λ_i $(1 \le i \le k)$ with the positive sign simultaneously through minimizing the objective function $\mathrm{trace}(U^{\mathrm{T}}KU + V^{\mathrm{T}}MV)$ subject to $U^{\mathrm{T}}V = I_k$ and to that $\mathrm{span}(U)$ and $\mathrm{span}(V)$ are restricted inside two suitably built subspaces $\mathcal{U}$ and $\mathcal{V}$, respectively:

$$\sum_{i=1}^{k} \lambda_i \approx \frac{1}{2} \inf_{\substack{U^{\mathrm{T}}V=I_k \\ \mathrm{span}(U)\subseteq\mathcal{U},\, \mathrm{span}(V)\subseteq\mathcal{V}}} \mathrm{trace}(U^{\mathrm{T}}KU + V^{\mathrm{T}}MV), \tag{1.9}$$

where $\mathrm{span}(U)$ denotes the subspace spanned by the column vectors of U.

The minimization problem in the right-hand side of (1.9) does not look easy to solve at first sight. But we introduce a structure-preserving projection matrix H_{SR} and show that the sum of its first k smallest eigenvalues with the positive sign is the infimum. In this sense, H_{SR} is the best projection matrix from the given subspaces $\mathcal{U}$ and $\mathcal{V}$, and solving its eigenvalue problem yields the best approximations to λ_i $(1 \le i \le k)$ and their associated eigenvectors. Moreover, H_{SR} has the same block structure as H. With these new developments, we will be able to construct efficient numerical algorithms that can compute several smallest eigenvalues with the positive sign of H simultaneously. Indeed we will present our versions of locally optimal conjugate gradient type algorithms, including blocked versions for computing these smallest eigenvalues simultaneously and preconditioned versions for speedy convergence. We

point out that these new algorithms are not straightforward applications of the standard steepest decent and nonlinear conjugate gradient algorithms, but improved ones to take advantage of the best projection matrix H_{SR} we have uncovered. All these are made possible by our new theory, parallel to some of the well-known and important results for the symmetric eigenvalue problem; see, for example, [33, 40].

The rest of this paper is organized as follows. Section 2 presents an algorithm to construct approximate eigenpairs for H, given a pair of approximate deflating subspaces $\{\mathcal{U}, \mathcal{V}\}$. It is derived from the result in [2] for the case when the subspaces do consist of a pair of deflating subspaces. Section 3 and the appendix discuss how to construct the best approximations to some of the eigenpairs of H, given a pair of approximate deflating subspaces $\{\mathcal{U}, \mathcal{V}\}$. The results in section 3 justifies the algorithm in section 2 from a different perspective. In section 4, we apply newly established minimization principles in [2] to derive CG type algorithms for computing a set of the smallest eigenvalues with the positive sign. In section 5, we present numerical results to illustrate the convergence behaviors of CG methods. Concluding remarks are in section 6.

Notation. We will follow the notation as specified at the end of section 1 in [2]. In particular, K, $M \in \mathbb{R}^{n\times n}$ are assumed, by default, to be symmetric positive semidefinite and one of them is definite, unless explicitly stated differently.

2. Approximate deflating subspaces. Recall that $\{\mathcal{U}, \mathcal{V}\}$ is a *pair of deflating subspaces* of $\{K, M\}$ if

$$K\mathcal{U} \subseteq \mathcal{V} \quad \text{and} \quad M\mathcal{V} \subseteq \mathcal{U}. \tag{2.1}$$

Each such pair will yield a subset of H's eigenvalues and corresponding eigenvectors associated with the eigenvalues in the subset [2, section 2.2]. But in practical computations, rarely pairs of exact deflating subspaces are known, only approximate ones. The question then arises: how to compute approximate eigenpairs of H given a pair of *approximate* deflating subspaces.

Let $\{\mathcal{U}, \mathcal{V}\}$ be a pair of *approximate* deflating subspaces with $\dim(\mathcal{U}) = \dim(\mathcal{V}) = \ell$ such that $W \overset{\text{def}}{=} U^{\text{T}}V$ is nonsingular. In [2, section 2], we defined a structure-preserving projection

$$H_{\text{SR}} = \begin{bmatrix} 0 & W_1^{-\text{T}}U^{\text{T}}KUW_1^{-1} \\ W_2^{-\text{T}}V^{\text{T}}MVW_2^{-1} & 0 \end{bmatrix} \tag{2.2}$$

of H onto the pair of the subspaces $\{\mathcal{U}, \mathcal{V}\}$, where $W_i \in \mathbb{R}^{\ell\times\ell}$ are from factorizing $W = W_1^{\text{T}}W_2$ and nonsingular. This H_{SR} in many ways, as will become clear later, plays the same role for H as the Rayleigh quotient matrix for the symmetric eigenvalue problem.

Theorem 2.6 in [2] shows how to construct the eigenpairs of H from those of H_{SR} when $\{\mathcal{U}, \mathcal{V}\}$ *is* a pair of deflating subspaces of $\{K, M\}$. The way of the construction there naturally leads us to propose the following algorithm.

ALGORITHM 2.1. *Given the basis matrices U and V of a pair of approximate deflating subspaces $\{\mathcal{U}, \mathcal{V}\}$ of $\{K, M\}$, this algorithm returns approximate eigenvalues and eigenvectors for H as follows.*

1. *Construct H_{SR} as in* (2.2) *(assume $U^{\text{T}}V$ is nonsingular);*
2. *Compute the eigenpairs $\{\hat{\lambda}, \left[\begin{smallmatrix}\hat{y}\\ \hat{x}\end{smallmatrix}\right]\}$ of H_{SR};*
3. *The computed eigenvalues $\hat{\lambda}$ approximate some eigenvalues of H, and the associated approximate eigenvectors are $\left[\begin{smallmatrix}VW_2^{-1}\hat{y}\\ UW_1^{-1}\hat{x}\end{smallmatrix}\right]$.*

In view of [2, Theorem 2.6], as far as the eigenvalue problem of H is concerned, in theory any one of H_{SR} associated with a given pair of *approximate* deflating subspaces is just as good as another. Numerically, however, we should pick basis matrices that are sufficiently well-conditioned, like with orthonormal columns.

Remark 2.1. Algorithm 2.1 requires that $W = U^{\text{T}}V$ be nonsingular. This necessarily entails $\dim(\mathcal{U}) = \dim(\mathcal{V})$. Conceivably, we may have a pair of approximate deflating subspaces $\{\mathcal{U}, \mathcal{V}\}$ for which either $\dim(\mathcal{U}) \neq \dim(\mathcal{V})$ or $W = U^{\text{T}}V$ is (numerically) singular. When that happens, H_{SR} cannot be defined as in (2.2). How then are we going to find approximate eigenpairs of H within the approximate deflating subspaces? This question will be answered in the appendix, where it is concluded that a smaller eigenvalue problem for a different structure-preserving projection matrix $\hat{H}_{\text{SR}}$ must be solved (see Theorem A.1).

3. Best approximations by a pair of subspaces. The two most important aspects in solving a large scale eigenvalue problem are

1. building subspaces to which the desired eigenvectors (or invariant subspaces) are close, and
2. seeking *best possible* approximations from the suitably built subspaces.

In this section, we shall address the second aspect for our current problem at hand, i.e., seeking *best possible* approximations to a few smallest eigenvalues with the positive sign of H and their associated eigenvectors from given pair of subspaces. We will prove that H_{SR} provides best approximations. We leave the first aspect to the later sections when we present our computational algorithms.

The concept of *best possible* comes with a quantitative measure as to what constitutes *best possible*. There may not be such a measure in general. In [33, section 11.4], Parlett uses three different ways to justify the use of the Rayleigh–Ritz procedure for the symmetric eigenvalue problem. For the eigenvalue problem here, each of the minimization principles we established in [2] provides a quantitative measure.

Recall the default assumption that K, $M \in \mathbb{R}^{n\times n}$ are symmetric positive semidefinite and one of them is definite. Let $\{\mathcal{U}, \mathcal{V}\}$ be a pair of approximate deflating subspaces of $\{K, M\}$ and $\dim(\mathcal{U}) = \ell_1$ and $\dim(\mathcal{V}) = \ell_2$. Motivated by the minimization principles in [2], we will seek

1. the best approximation to λ_1 in the sense of

$$\inf_{x\in\mathcal{U},\, y\in\mathcal{V}} \rho(x, y) \tag{3.1}$$

   and its associated approximate eigenvector;
2. the best approximations to λ_j $(1 \le j \le k)$ in the sense of

$$\frac{1}{2} \inf_{\substack{\text{span}(\hat{U})\subseteq\mathcal{U},\text{span}(\hat{V})\subseteq\mathcal{V}\\ \hat{U}^{\text{T}}\hat{V}=I_k}} \text{trace}(\hat{U}^{\text{T}}K\hat{U} + \hat{V}^{\text{T}}M\hat{V}) \tag{3.2}$$

   and their associated approximate eigenvectors.

To this end, we divide our investigation into two cases. Let $U \in \mathbb{R}^{n\times\ell_1}$, $V \in \mathbb{R}^{n\times\ell_2}$ be the basis matrices of $\mathcal{U}$ and $\mathcal{V}$, respectively, and set $W = U^{\text{T}}V$. The two cases are

1. $W = U^{\text{T}}V$ is nonsingular. (Necessarily, $\ell_1 = \ell_2$. Set $\ell = \ell_i$.)
2. $W = U^{\text{T}}V$ is singular or $\ell_1 \neq \ell_2$.

For the first case, i.e., $W = U^{\text{T}}V$ is nonsingular, we factorize $W = W_1^{\text{T}}W_2$, where $W_i \in \mathbb{R}^{\ell\times\ell}$ are nonsingular. How this factorization is done is not essential mathematically. But it is included to accommodate cases when such a factorization

may offer certain conveniences. In general, simply taking $W_1 = W^{\mathrm{T}}$ and $W_2 = I_\ell$ or $W_1 = I_\ell$ and $W_2 = W$ may just be good enough.

For the best approximation to λ_1 by (3.1), we note that any $x \in \mathcal{U}$ and $y \in \mathcal{V}$ can be written as $x = U\hat{u}$ and $y = V\hat{v}$ for some $\hat{u}$, $\hat{v} \in \mathbb{R}^\ell$ and vice versa. Therefore, we have

$$\begin{aligned} \rho(x,y) &= \frac{\hat{u}^{\mathrm{T}}U^{\mathrm{T}}KU\hat{u} + \hat{v}^{\mathrm{T}}V^{\mathrm{T}}MV\hat{v}}{2|\hat{u}^{\mathrm{T}}W\hat{v}|} \\ &= \frac{\hat{x}^{\mathrm{T}}W_1^{-\,\mathrm{T}}U^{\mathrm{T}}KUW_1^{-1}\hat{x} + \hat{y}^{\mathrm{T}}W_2^{-\,\mathrm{T}}V^{\mathrm{T}}MVW_2^{-1}\hat{y}}{2|\hat{x}^{\mathrm{T}}\hat{y}|}, \end{aligned} \tag{3.3}$$

where $\hat{x} = W_1\hat{u}$ and $\hat{y} = W_2\hat{v}$. By (3.3) and [2, Theorem 3.1], we immediately conclude that the quantity in (3.1) is the smallest eigenvalue with the positive sign of H_{SR} defined in (2.2).

Now turn to the best approximations to λ_j $(1 \le j \le k)$ by (3.2). Note that any $\hat{U}$ and $\hat{V}$ such that $\mathrm{span}(\hat{U}) \subseteq \mathcal{U}$, $\mathrm{span}(\hat{V}) \subseteq \mathcal{V}$, and $\hat{U}^{\mathrm{T}}\hat{V} = I_k$ can be written as

$$\hat{U} = UW_1^{-1}\hat{X}, \quad \hat{V} = VW_2^{-1}\hat{Y},$$

where $\hat{X}$, $\hat{Y} \in \mathbb{R}^{\ell\times k}$ and $\hat{X}^{\mathrm{T}}\hat{Y} = I_k$, and vice versa. Hence we have

$$\hat{U}^{\mathrm{T}}K\hat{U} + \hat{V}^{\mathrm{T}}M\hat{V} = \hat{X}^{\mathrm{T}}W_1^{-\,\mathrm{T}}U^{\mathrm{T}}KUW_1^{-1}\hat{X} + \hat{Y}^{\mathrm{T}}W_2^{-\,\mathrm{T}}V^{\mathrm{T}}MVW_2^{-1}\hat{Y}$$

and thus

$$\begin{aligned} &\inf_{\substack{\mathrm{span}(\hat{U})\subseteq\mathcal{U},\mathrm{span}(\hat{V})\subseteq\mathcal{V}\\ \hat{U}^{\mathrm{T}}\hat{V}=I_k}} \mathrm{trace}(\hat{U}^{\mathrm{T}}K\hat{U} + \hat{V}^{\mathrm{T}}M\hat{V}) \\ &\qquad = \inf_{\hat{X}^{\mathrm{T}}\hat{Y}=I_k} \mathrm{trace}(\hat{X}^{\mathrm{T}}W_1^{-\,\mathrm{T}}U^{\mathrm{T}}KUW_1^{-1}\hat{X} + \hat{Y}^{\mathrm{T}}W_2^{-\,\mathrm{T}}V^{\mathrm{T}}MVW_2^{-1}\hat{Y}). \end{aligned} \tag{3.4}$$

By [2, Theorem 3.2], we know that the right-hand side of (3.4) is the sum of the k smallest eigenvalues with the positive sign of H_{SR}.

In summary, the *best approximations to the first k eigenvalues with the positive sign of H within the pair of approximate deflating subspaces are the eigenvalues of* H_{SR}. Algorithmically, denote by μ_j $(j = 1, \ldots, \ell)$ the eigenvalues with the positive sign of H_{SR} in the ascending order and by $\hat{z}_j$ the associated eigenvectors, i.e., $0 \le \mu_1 \le \cdots \le \mu_\ell$, and

$$H_{\mathrm{SR}}\hat{z}_j = \mu_j\hat{z}_j, \quad \hat{z}_j = \begin{bmatrix}\hat{y}_j\\ \hat{x}_j\end{bmatrix}. \tag{3.5}$$

It can be verified that

$$\rho(UW_1^{-1}\hat{x}_j, VW_2^{-1}\hat{y}_j) = \mu_j \quad \text{for } j = 1, \ldots, \ell.$$

Naturally, according to Algorithm 2.1, we take $\lambda_j \approx \mu_j$ and the corresponding approximate eigenvectors of H as

$$\widetilde{z}_j \equiv \begin{bmatrix}\widetilde{y}_j\\ \widetilde{x}_j\end{bmatrix} = \begin{bmatrix}VW_2^{-1}\hat{y}_j\\ UW_1^{-1}\hat{x}_j\end{bmatrix} \quad \text{for } j = 1, \ldots, \ell. \tag{3.6}$$

In practice, not all of the approximate eigenpairs $(\mu_j, \widetilde{z}_j)$ are equally accurate to the same level. Usually the first few pairs are more accurate than the rest.

For the ease of reference, we summarize the findings for the first case of nonsingular $W = U^{\mathrm{T}}V$ into the following theorem.

THEOREM 3.1. *Suppose that one of* $K, M \in \mathbb{R}^{n\times n}$ *is definite. Let* $\{\mathcal{U}, \mathcal{V}\}$ *be a pair of approximate deflating subspaces of* $\{K, M\}$ *with* $\dim(\mathcal{U}) = \dim(\mathcal{V}) = \ell$*, and let* $U, V \in \mathbb{R}^{n\times\ell}$ *be the basis matrices of* $\mathcal{U}$ *and* $\mathcal{V}$*, respectively. If* $W \stackrel{\text{def}}{=} U^{\mathrm{T}}V$ *is nonsingular, then*

$$\sum_{j=1}^{k}\mu_j = \frac{1}{2}\inf_{\substack{\mathrm{span}(\hat{U})\subseteq\mathcal{U},\mathrm{span}(\hat{V})\subseteq\mathcal{V}\\ \hat{U}^{\mathrm{T}}\hat{V}=I_k}} \mathrm{trace}(\hat{U}^{\mathrm{T}}K\hat{U} + \hat{V}^{\mathrm{T}}M\hat{V}),$$

and the best approximations to λ_1 *in the sense of* (3.1) *or to* λ_j *(*$1 \le j \le k$*) in the sense of* (3.2) *are the eigenvalues* $\{\mu_j\}$ *of* H_{SR} *defined in* (2.2) *with the corresponding approximate eigenvectors given by* (3.6).

We recall that even though H_{SR} is not uniquely determined by the given subspaces $\mathcal{U}$ and $\mathcal{V}$, the approximate eigenpairs $(\mu_j, \widetilde{z}_j)$ are uniquely determined, as guaranteed by [2, Theorem 2.9].

The second case for W being singular or $\ell_1 \neq \ell_2$ is much involved, and we defer its treatment to the appendix. The conclusion is similar in that both optimization problems in (3.1) and (3.2) can still be solved through solving a smaller eigenvalue problem for a structure-preserving projection matrix $\hat{H}_{\mathrm{SR}}$ to be defined in the appendix. Similar results to these in Theorem 3.1 can be found in Theorem A.1.

Remark 3.1. The best approximation technique so far is based on the minimization principles in [2, Theorems 3.1 and 3.2]. Naturally one may wonder if a similar technique could be devised using the minimization principles in [2, Theorem 3.4] for the original LR eigenvalue problem (1.2). But that seems hard, if at all possible. The difficulty lies in that there appears no good way to define a proper structure-preserving projection matrix of $\left[\begin{smallmatrix} A & B \\ B & A\end{smallmatrix}\right]$ or of $\left[\begin{smallmatrix} A & B \\ -B & -A\end{smallmatrix}\right]$ onto the given subspaces.

4. 4D CG algorithms.

4.1. Partial gradients. The partial gradients of the Thouless functional $\rho(x, y)$ with respect to x and y will be needed later for minimization. To find the gradients, we perturb x and y to $x + p$ and $y + q$, respectively, where p and q are assumed to be small in magnitude. Assuming $x^{\mathrm{T}}y \neq 0$, up to the first order in p and q, we have

$$\begin{aligned}\rho(x+p, y+q) &= \frac{(x+p)^{\mathrm{T}}K(x+p) + (y+q)^{\mathrm{T}}M(y+q)}{2|(x+p)^{\mathrm{T}}(y+q)|}\\ &= \frac{x^{\mathrm{T}}Kx + 2p^{\mathrm{T}}Kx + y^{\mathrm{T}}My + 2q^{\mathrm{T}}My}{2\,|x^{\mathrm{T}}y|}\left(1 - \frac{p^{\mathrm{T}}y + q^{\mathrm{T}}x}{x^{\mathrm{T}}y}\right)\\ &= \rho(x,y) + \frac{1}{x^{\mathrm{T}}y}\,p^{\mathrm{T}}\left[Kx - \rho(x,y)\,y\right] + \frac{1}{x^{\mathrm{T}}y}\,q^{\mathrm{T}}\left[My - \rho(x,y)\,x\right].\end{aligned}$$

Therefore the partial gradients of $\rho(x, y)$ with respect to x and y are given by

$$\nabla_x\rho = \frac{1}{x^{\mathrm{T}}y}\left[Kx - \rho(x,y)\,y\right], \quad \nabla_y\rho = \frac{1}{x^{\mathrm{T}}y}\left[My - \rho(x,y)\,x\right]. \tag{4.1}$$

There is a close relation between these two partial gradients and the residual:

$$Hz - \rho(x,y)z \equiv \begin{bmatrix} 0 & K \\ M & 0\end{bmatrix}\begin{bmatrix} y \\ x\end{bmatrix} - \rho(x,y)\begin{bmatrix} y \\ x\end{bmatrix} = x^{\mathrm{T}}y\begin{bmatrix}\nabla_x\rho \\ \nabla_y\rho\end{bmatrix}. \tag{4.2}$$

Namely the block vector obtained by stacking $\nabla_x\rho$ over $\nabla_y\rho$ is parallel to the residual.

4.2. 4D search. The line search is a common approach in the process of optimizing a function value. For our case, we are interested in solving $\inf_{x,y} \rho(x,y)$ in order to compute λ_1 and its associated eigenvector of H. From the theoretical point of view, this task of minimizing $\rho(x,y)$ may end up with no optimal arguments because possibly no x and y attend the infimum, unless both K and M are definite. One may argue that in this case, λ_1 is already known, i.e., $+0$, when the infimum cannot be attained and it happens if one of K and M is singular. Naturally one can compute corresponding eigenvectors (by, e.g., the inverse iteration) and deflate out the eigenvalues 0. But in practice, that one of them is singular may not be known a priori, except that both are semidefinite is usually known from the problem setup. A likely scenario would be that one may still attempt to minimize $\rho(x,y)$ anyway. What would happen then? First numerically rarely a matrix is exactly singular. This means that the singular K or M is not actually singular (even might be slightly indefinite). With carefully written computer codes, one may safely regard the singular one barely definite. We find from our numerical tests that with a preconditioner approximately H^{-1}, computations by minimizing $\rho(x,y)$ can still yield meaningful numerical results: the computed λ_1 is very tiny, and one of x and y is negligible compared to the other. Therefore, despite the implied theoretical impasse by [2, Theorem 3.1] when one of K and M is singular, attempting to minimize $\rho(x,y)$, with a suitable preconditioner, is still a worthwhile thing to do in seeking λ_1 and its associated eigenvector of H.

Given a search direction $\left[\begin{smallmatrix} q \\ p \end{smallmatrix}\right]$ from the current position $\left[\begin{smallmatrix} y \\ x \end{smallmatrix}\right]$, the basic idea of the line search[2] [26, 28] is to look for the best possible scalar argument t to minimize ρ:

$$\min_t \rho(x + tp, y + tq) \tag{4.3}$$

on the line

$$\left\{ \begin{bmatrix} y \\ x \end{bmatrix} + t \begin{bmatrix} q \\ p \end{bmatrix} \,:\, t \in \mathbb{R} \right\}. \tag{4.4}$$

Recently, in [9], a dual-channel extension of the line search is introduced by solving the minimization problem

$$\min_{s,t} \rho(x + sp, y + tq), \tag{4.5}$$

where the search directions p and q are selected as the partial gradients $\nabla_x \rho$ and $\nabla_y \rho$ as in (4.1). The minimization problem (4.5) is then solved iteratively by freezing one of s and t and minimizing the functional ρ over the other in an alternative manner.

However, we did not pursue these ideas for reasons to be detailed in a moment. Instead, we shall look for four scalars α, β, s, and t for the minimization problem

$$\inf_{\alpha,\beta,s,t} \rho(\alpha x + sp, \beta y + tq) = \min_{u \in \mathrm{span}(U),\ v \in \mathrm{span}(V)} \rho(u,v), \tag{4.6}$$

where $U = [x,p]$ and $V = [y,q]$. This no longer performs a line search (4.4) or dual channel optimization (4.5), but a 4-*dimensional subspace search* (*4D search* for short)

[2]Since ρ is homogeneous of degree 0, i.e., $\rho(tx,ty) \equiv \rho(x,y)$ for any scalar t, minimizing ρ along the line (4.4) is in fact minimizing ρ in $\{\alpha\left[\begin{smallmatrix} y \\ x \end{smallmatrix}\right] + \alpha t\left[\begin{smallmatrix} q \\ p \end{smallmatrix}\right] \,:\, \alpha,\, t \in \mathbb{R}\}$ which, in general, form a plan in $\mathbb{R}^{2n}$ spanned by $\left[\begin{smallmatrix} y \\ x \end{smallmatrix}\right]$ and $\left[\begin{smallmatrix} q \\ p \end{smallmatrix}\right]$, excluding the line $\{t\left[\begin{smallmatrix} q \\ p \end{smallmatrix}\right] \,:\, t \in \mathbb{R}\}$. Therefore, the standard line search becomes a defacto plane search for ρ. To be consistent with the standard terminology in optimization, we still call it *the line search*.

within the 4-*dimensional subspace*:

$$\left\{ \begin{bmatrix} \beta y + tq \\ \alpha x + sp \end{bmatrix} \quad \text{for all scalars } \alpha,\ \beta,\ s, \text{ and } t \right\}. \tag{4.7}$$

The right-hand side of (4.6) can be solved by the methods given in section 3 if $U^{\mathrm{T}}V$ is nonsingular (the common case) or in the appendix otherwise (the rare case).

We prefer our 4D search for the following reasons:

1. While we have no formal proof, it seems that the optimization problems (4.3) and (4.5) cannot be recasted into a (much) smaller eigenvalue problem of a matrix having the same block structure as H.
2. The line search and dual channel optimization are not readily extensible to the subspace search, a crucial technique for our development in simultaneously computing few smallest eigenvalues with the positive sign and corresponding eigenvectors of H.
3. Both optimization problems (4.4) and (4.5) yield the best possible approximation in a subspace that is contained in the 4 dimensional subspace over which our 4D search minimizes. Starting with the same p and q, the solution by our 4D search is and can be much better at about the same cost.
4. Although we restricted our developments so far on real K and M, they are actually valid for Hermitian K and M after minor changes, i.e., replacing all transposes $(\cdot)^{\mathrm{T}}$ by complex conjugate transposes $(\cdot)^{\mathrm{H}}$. When K and M are Hermitian and some of their entries are complex, our 4D search is truly a 4D search over a 4 dimensional subspace in $\mathbb{C}^{2n}$ (the $2n$-dimensional Euclidean vector space over the complex field), whereas the standard line search does not minimize ρ over a straight line in $\mathbb{C}^{2n}$ because t is restricted to be real. Conceivably the standard line search solution could be even worse in the complex case.

4.3. 4D CG algorithms. The minimization principle in [44, 48] and the newly established one in [2, Theorem 3.2] make it tempting to apply memory-efficient nonlinear CG algorithms (see, e.g., [34, 29, 42]) to solve LR eigenvalue problems. Not surprisingly, such applications had been attempted [9, 26, 28]. Conceivably when only one eigenvalue and its associated eigenvector are requested, it matters little, if any, to apply CG to (1.4) for the Hamiltonian matrix in (1.2) or to (1.6) for H in (1.1), much like the CG method for symmetric eigenvalue problems [7, 39, 25]. But it is a very different story if more than one eigenpair are requested, in which case block algorithms are better options. It seems hard, if at all possible, to create a block CG algorithm for the Hamiltonian matrix eigenvalue problem (1.2) directly, even with our new minimization principle in [2, Theorem 3.3] for the same reason as we pointed out in Remark 3.1. On the other hand, the developments in section 3 and the appendix make it possible for us to design efficient block CG algorithms for H to compute its first few smallest eigenvalues λ_j with the positive sign and their corresponding eigenvectors simultaneously, based on the minimization principle in [2, Theorem 3.2] and the Cauchy-like interlacing inequalities in [2, Theorem 4.1]. This is the precise reason we prefer to work with H.

The *locally optimal CG algorithm* [34, 21] was born as a result of the following:

1. the observation that the next approximation from applying a (classical) nonlinear CG algorithm to an optimization problem lies in the subspace spanned by the most recent approximation, the most recent searching direction, and the gradient at the most recent approximation, and

2. that subspace is the same as the one spanned by the two most recent approximations and the gradient, and thus we could compute the next approximation as the optimal solution within that subspace.

It has been noted that the locally optimal CG algorithm is often better suited for solving large scale symmetric eigenvalue problems, especially with a proper preconditioner, than the (classical) nonlinear CG algorithms [22, 31]. It converges fast, has no parameters to tune, and is easy to implement. Inspired by this, we present in what follows our locally optimal 4D search CG algorithms (with or without preconditioners).

For many nonlinear optimization problems, even solving simple line searches poses challenges. But for the eigenvalue problem for H, thanks to Theorems 3.1 and A.1, the optimal approximate solution within a pair of subspaces of dimension higher than one is easily computed, very much like the case for the standard symmetric eigenvalue problem for which Knyazev [22] proposed the *locally optimal block preconditioned conjugate gradient* (LOBPCG) method.

Algorithm 4.1 summarizes four *locally optimal* 4*D CG algorithms* in one. We attach 4D to them because of their relation to the 4D search idea in subsection 4.2. Their creation follows the idea of "*local optimality*" in the locally optimal CG algorithm in that each step the optimal solution is searched within the subspace spanned by the two most recent approximations and the partial gradients. Each of the four algorithms below is realized through setting its integer parameter k and preconditioner Φ in Algorithm 4.1:

- *Locally optimal* 4*D CG algorithm* (LO4DCG):

$$k = 1 \quad \text{and} \quad \Phi = \begin{bmatrix} 0 & I_n \\ I_n & 0 \end{bmatrix}. \tag{4.8}$$

- *Locally optimal preconditioned* 4*D CG algorithm* (LOP4DCG): $k = 1$ and preconditioner Φ.
- *Locally optimal block* 4*D CG algorithm* (LOB4DCG): $k > 1$ and Φ as in (4.8).
- *Locally optimal block preconditioned* 4*D CG algorithm* (LOBP4DCG): $k > 1$ and preconditioner Φ.

The key iterative step in these locally optimal 4D CG algorithms is to seek the best possible approximations in the subspace spanned by the two most recent approximations and the (preconditioned) gradients at the most recent approximations, except for the first iterative step for which the search subspace is simply spanned by the initial approximations and the (preconditioned) gradients at the approximations. A straightforward application would be to search the next approximations within

$$\text{span} \bigcup_{1 \le j \le k} \left\{ \begin{bmatrix} y_j^{(i)} \\ x_j^{(i)} \end{bmatrix}, \begin{bmatrix} y_j^{(i-1)} \\ x_j^{(i-1)} \end{bmatrix}, \begin{bmatrix} q_j \\ p_j \end{bmatrix} \right\}, \tag{4.9}$$

where the superscripts $(i-1)$ and (i) indicate that they are for the $(i-1)$st and ith iterative steps, respectively, and

$$\begin{bmatrix} q_j \\ p_j \end{bmatrix} = \Phi \left. \begin{bmatrix} \nabla_x \rho \\ \nabla_y \rho \end{bmatrix} \right|_{(x,y)=(x_j^{(i)}, y_j^{(i)})} \tag{4.10}$$

and Φ is a preconditioner. For the first iterative step, the vectors in (4.9) with the superscript $(i-1)$ should be deleted from the list because they are not available yet.

To utilize the best approximation methods in section 3 and the appendix, we modify this approach by using the search space

$$\operatorname{span} \bigcup_{1\le j\le k} \left\{ \begin{bmatrix} y_j^{(i)} \\ 0 \end{bmatrix}, \begin{bmatrix} y_j^{(i-1)} \\ 0 \end{bmatrix}, \begin{bmatrix} q_j \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ x_j^{(i)} \end{bmatrix}, \begin{bmatrix} 0 \\ x_j^{(i-1)} \end{bmatrix}, \begin{bmatrix} 0 \\ p_j \end{bmatrix} \right\}.$$

Breaking each vector into two in such a way is a common technique today in developing structure-preserving algorithms (see, e.g., [20, 14, 24]). We are now ready to give our four locally optimal 4D CG algorithms collectively in one.

ALGORITHM 4.1.

0 Select initial approximations X_0 and Y_0 having k columns such that columns of $Z_0 = \left[\begin{smallmatrix} Y_0 \\ X_0 \end{smallmatrix}\right]$ are approximate eigenvectors of H associated with λ_j, $1 \le j \le k$.
1 for $i = 0, 1, \ldots$ until convergence:
2 $\quad \rho_j = \rho((X_i)_{(:,j)}, (Y_i)_{(:,j)})$, $1 \le j \le k$;
3 $\quad P_i = KX_i - Y_i \operatorname{diag}(\rho_1, \ldots, \rho_k)$, $Q_i = MY_i - X_i \operatorname{diag}(\rho_1, \ldots, \rho_k)$;
3.1 $\quad \left[\begin{smallmatrix} Q_i \\ P_i \end{smallmatrix}\right] \leftarrow \Phi \left[\begin{smallmatrix} P_i \\ Q_i \end{smallmatrix}\right]$ if the preconditioner Φ is given;
4.1 $\quad$ For $i = 0$: $U = [X_i, P_i]$, $V = [Y_i, Q_i]$;
4.2 $\quad$ For $i > 0$: $U = [X_i, X_{i-1}, P_i]$, $V = [Y_i, Y_{i-1}, Q_i]$;
4.3 $\quad$ Orthogonalize the columns of U and V;
4.4 $\quad W = U^{\mathrm{T}}V = W_1^{\mathrm{T}} W_2$;
5 $\quad$ Construct H_{SR} as in (2.2) (assume W is nonsingular);
6 $\quad$ Compute the k smallest eigenvalues with the positive sign of H_{SR}, and the associated eigenvectors as in (3.5);
7 $\quad X_{i+1} = UW_1^{-1}[\hat{x}_1, \ldots, \hat{x}_k]$, $Y_{i+1} = VW_2^{-1}[\hat{y}_1, \ldots, \hat{y}_k]$;
8 $\quad$ Normalize each column of $Z_{i+1} = \left[\begin{smallmatrix} Y_{i+1} \\ X_{i+1} \end{smallmatrix}\right]$.
9 end

A few comments are in order for Algorithm 4.1:

1. At line 2, evaluations of ρ_j are needed only for iterative step $i = 0$; for $i \ge 1$, they are the k smallest eigenvalues with the positive sign of H_{SR} in the previous iterative step.
2. For the convergence test, we can use the relative residual norm of the approximate eigenpair $(\rho(x_j^{(i)}, y_j^{(i)}), z_j^{(i)})$ of H, where $z_j^{(i)} = (Z_i)_{(:,j)}$ (see section 5).
3. U and V constructed at line 4.1 or line 4.2 may be ill-conditioned, especially when near convergence because then the gradients tend to the zero vector and some columns of X_i and/or Y_i are almost converged. To ensure that U and V are well-conditioned for better numerical stability, we may have to orthogonalize their columns via, e.g., the (classical/modified) Gram–Schmidt orthogonalization process. This is the reason we have line 4.3 there.

   An idea presented in [22] and [18] can be adapted to alleviate the ill-conditioning of the U and V in line 4.2. It computes different U and V without altering their column spaces which are all that matter for the algorithm. It goes as follows. Change the assignments to U and V at line 4.2 to $U = [X_i, \widetilde{X}_i, P_i]$, $V = [Y_i, \widetilde{Y}_i, Q_i]$. The needed $\widetilde{X}_i$ and $\widetilde{Y}_i$ are calculated in the previous iterative step at line 7: besides X_{i+1} and Y_{i+1}, also compute

$$\widetilde{X}_{i+1} = UW_1^{-1}\widehat{X}, \quad \widetilde{Y}_{i+1} = VW_2^{-1}\widehat{Y},$$

   where $\widehat{X}$ is $[\hat{x}_1, \ldots, \hat{x}_k]$ with its first k rows zeroed out and $\widehat{Y}$ is $[\hat{y}_1, \ldots, \hat{y}_k]$ with its first k rows also zeroed out. In order for this idea to work, we need to

make sure that the orthogonalization at line 4.3 does not alter the subspace spanned by the first k columns of U and the subspace spanned by the first k columns of V. This requirement is easy to satisfy, however. But we emphasize that with the newly proposed line 4.2, something like the orthogonalization at line 4.3 is still necessary for the same reason as argued in [18]. We tested this idea in our numerical examples in section 5 and found little difference from Algorithm 4.1 as is.

4. From line 5 to line 8, we leave out the case when $U^{\mathrm{T}}V$ is singular or the numbers of columns in U and V are different for simplicity. Actual implementation should include the case for which the optimal solution has been given in detail in the appendix. Specifically, instead of H_{SR} as in (2.2), we compute $\hat{H}_{\text{SR}}$ as in (A.5) and its $\min\{k, r\}$ smallest eigenvalues with the positive sign and the associated eigenvectors as in (A.7), and finally compute $Z_{i+1} = \left[\widetilde{z}_1, \ldots, \widetilde{z}_{\min\{k,r\}}\right]$ by (A.8)–(A.10), where r is the (numerical) rank of W.
   There are two factors that affect the (non)singularity of $U^{\mathrm{T}}V$: (1) the choices of U and V as the basis matrices of $\mathcal{U} = \text{span}\{X_i, X_{i-1}, P_i\}$ and $\mathcal{V} = \text{span}\{Y_i, Y_{i-1}, Q_i\}$, respectively, and (2) the angle $\angle(\mathcal{U}, \mathcal{V})$ between the two subspaces. The first factor can be alleviated by picking orthonormal bases as suggested at line 4.3, albeit an expensive step. See our third comment above. A similar issue came up in the LOBPCG [18, 22]. But the second factor is an intrinsic one. That the numbers of columns in U and V may be different is due to the fact that numerically $\dim(\mathcal{U})$ and $\dim(\mathcal{V})$ may be different, especially when some approximate eigenpairs have converged. Handling these cases requires the technique detailed in the appendix. Even with the technique, a complete and robust implementation must confront the important issue of deciding the numerical rank of $U^{\mathrm{T}}V$ and the numerical dimensions of $\mathcal{U}$ and $\mathcal{V}$ to balance convergence speed and numerical stability. This is one of the issues that warrant further study.
5. At line 6, LAPACK's routine `xSYSVD` can be used to solve the eigenvalue problem of H_{SR} because of its small size. In theory half of the eigenvalues of H_{SR} have the positive sign and the other half are opposite. But when the off-diagonal blocks of H_{SR} is semi- or barely definite, some of the tiny eigenvalues with the positive sign may be computed by `xSYSVD` negative or even complex with tiny magnitude. This can be easily detected and corrected by simple postprocessing.
6. At line 8, we can simply scale each column of Z_{i+1} to be a unit vector in some vector norm.
7. Sometimes it can be helpful to use a k that is somewhat bigger than the actual number of requested eigenpairs for accelerating convergence.
8. Φ as in (4.8) gives the plain 4D CG algorithm (i.e., without preconditioning). A *generic* preconditioner to compute the eigenvalues of H close to a prescribed point μ is $\Phi = (H - \mu I_{2n})^{-1}$. Then the vectors p_j and q_j defined by (4.10) can be computed through approximately solving a linear system with the coefficient matrix $H - \mu I_{2n}$. Note that the arrangement of the two blocks in the matrix applied to by Φ is not mistaken. In fact each column of the matrix is parallel to the corresponding residual vector as given by (4.2). The modified directions are parallel to the ones obtained from one step of the inverse power iteration on the residual. When μ is closer to the desired eigenvalues than any

others, the preconditioned directions should have "larger" components in the desired eigenvectors than the ones obtained without preconditioning. Since we are particularly interested in the smallest eigenvalues with the positive sign, $\mu = 0$ is often an obvious choice. Then

$$\Phi \begin{bmatrix} \nabla_x \rho \\ \nabla_y \rho \end{bmatrix} = \begin{bmatrix} 0 & M^{-1} \\ K^{-1} & 0 \end{bmatrix} \begin{bmatrix} \nabla_x \rho \\ \nabla_y \rho \end{bmatrix} = \begin{bmatrix} M^{-1} \nabla_y \rho \\ K^{-1} \nabla_x \rho \end{bmatrix} =: \begin{bmatrix} q \\ p \end{bmatrix}. \tag{4.11}$$

In this case, both p and q can be computed by using the linear conjugate gradient method [11, 15]. The search direction in the x-component depends only on $\nabla_x \rho$ while the search direction in the y-component depends only on $\nabla_y \rho$. This, in part, also justifies the correct block ordering in the vector applied to by Φ in (4.10).

5. Numerical examples. In this section, we present numerical results obtained within the MATLAB environment to illustrate the essential convergence behaviors of locally optimal 4D CG algorithms in section 4. The normalized residual norms and relative eigenvalue errors for the jth approximate eigenpair $(\lambda_j^{(i)}, z_j^{(i)})$ at the ith iterative step to the exact jth eigenpairs $(\lambda_j^{\text{exact}}, z_j)$ are defined by

$$\frac{\|H z_j^{(i)} - \lambda_j^{(i)} z_j^{(i)}\|_1}{(\|H\|_1 + \lambda_j^{(i)}) \|z_j^{(i)}\|_1} \quad \text{and} \quad \frac{|\lambda_j^{\text{exact}} - \lambda_j^{(i)}|}{|\lambda_j^{\text{exact}}|},$$

respectively, where λ_j^{exact} are computed by the QR algorithm (via MATLAB's function `eig`) and considered to be the "exact" eigenvalues, and $\|\cdot\|_1$ is the vector 1-norm or the matrix ℓ_1-operator norm, depending on its argument.

Example 5.1. In this example, we use a pair of matrices K and M from linear response analysis of silane (SiH4) compound generated by the turboTDDFT code in QUANTUM ESPRESSO (QE), an electronic structure calculation code that implements density functional theory (DFT) using plane-waves as the basis set and pseudopotentials [12]. Such small molecules are often used as benchmark tests to assess various simulation models, functionals, and methods (see, e.g., [27]). The order of the symmetric positive definite matrices K and M is 5660. Consequently, the dimension of the LR eigenvalue problem of H is 11320.

Our goal is to compute four smallest positive eigenvalues $0 < \lambda_1 < \lambda_2 < \lambda_3 < \lambda_4$ and corresponding eigenvectors z_1, z_2, z_3, z_4 of H. The preconditioner is chosen to be the generic one,

$$\Phi = H^{-1} = \begin{bmatrix} 0 & M^{-1} \\ K^{-1} & 0 \end{bmatrix}. \tag{5.1}$$

The preconditioned search vectors q_i and p_i are computed by using the linear CG method [11, 15] to solve the associated linear systems of equations (see (4.11)). Often very crude approximations of q_i and p_i are good enough. In this example, the linear CG iterations are set with the stopping tolerance 10^{-2} or maximal 20 iterations.

Figure 5.1 shows the normalized residual norms and the relative eigenvalue errors of LOB4DCG and LOBP4DCG (Algorithm 4.1 with $k = 4$). The initial approximate eigenvectors of z_j are chosen as $\left[e_j^{\mathrm{T}}, e_j^{\mathrm{T}}\right]^{\mathrm{T}}$ for $j = 1, \ldots, k$. We observe the significant improvement in the rate of convergence by using preconditioning. In terms of the cost for applying the preconditioner at line 3.1 of LOBP4DCG iteration, there are k

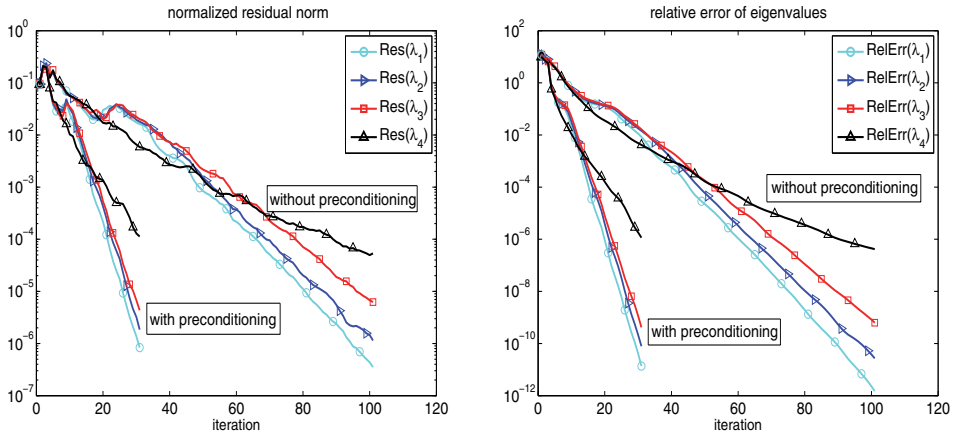

FIG. 5.1. *The convergence behaviors of the locally optimal block 4D CG algorithms with and without preconditioning for computing the four smallest positive eigenvalues (excitation states) of a synthesized SiH_4 compound: normalized residual norms (left) and relative eigenvalue errors (right).*

linear systems with coefficient matrix K and k ones with M for applying the generic preconditioner H^{-1}. We recorded that there were a total of 1048 linear CG iterations associated with the matrix K and 1010 associated with the matrix M over the 30 LOBP4DCG iterations to converge to the desired eigenpairs. Therefore, the total number of matrix-vector multiplications for applying the preconditioner H^{-1} is 2058, and the average number of inner linear CG iterations for each eigenpair is 8.575 per LOBP4DCG iteration.

We should note that the generic preconditioner H is not the natural preconditioner for this example. For the plane wave-based calculations, it is more natural to use a proper scaled diagonal-like preconditioner proposed in [43]. Excellent performance of such preconditioner in the the turboTDDFT code has been reported [37].

Example 5.2. We examine the convergence behavior of the the LOBP4DCG algorithm (Algorithm 4.1) for a case where K is symmetric positive semi-definite (i.e., K is singular) and M is definite. Specifically, K is a tridiagonal matrix such that $K_{(i,i+1)} = K_{(i+1,i)} = -1$ for $1 \le i \le n-1$, $K_{(i,i)} = 2$ for $2 \le i \le n-1$, $K_{(1,1)} = K_{(n,n)} = 1$ and $K_{(i,j)} = 0$ elsewhere. $M = \mathrm{diag}(1, 2, \ldots, n)$. We use the LOBP4DCG to compute the first four smallest eigenvalues with the positive sign (including $\lambda_1 = +0$) of H with $n = 2000$. The initial Z_0 are chosen as a random $2n \times k$ matrix. Figure 5.2 shows the normalized residual norms and the absolute error in $\lambda_1^{(i)}$ (since $\lambda_1 = +0$) and relative errors in $\lambda_j^{(i)}$ $(j = 2, 3, 4)$ by LOBP4DCG with the generic preconditioner $\Phi = H^{-1}$, which is again implemented by the linear CG method[3] to solve the associated linear systems with stopping tolerance 10^{-2} or maximal 50 iterations. Despite the singularity of K, as shown in Figure 5.2, we still observe decent convergence rate towards the desired eigenvalues. Note for this example, LOB4DCG, i.e., without the preconditioner $\Phi = H^{-1}$, does not converge. This example shows that when one of K and M is singular, the LR eigenvalue problem (1.1) becomes very hard, though not impossible, for LOBP4DCG to solve. Perhaps a better way would be deflated out the zero eigenvalues first before LOBP4DCG is

[3]Since M is diagonal, the linear systems with M are simple to solve. But we pretend not to know this and use CG anyway for testing purpose.

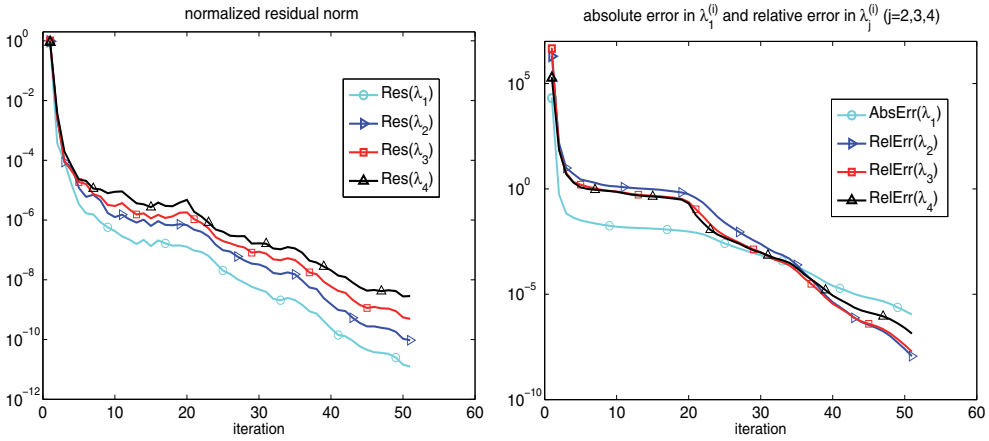

FIG. 5.2. *The convergence behaviors of the locally optimal block preconditioned 4D CG algorithms with preconditioning for computing the four smallest eigenvalues with the positive sign for the artificially constructed H: normalized residual norms (left) and absolute or relative eigenvalue errors (right).*

applied. We will investigate this issue elsewhere.

6. Concluding remarks. Based on the theoretical foundation laid out in [2] for the LR eigenvalue problem (1.1), we developed a 4D search technique and devised locally optimal CG methods that are capable of computing the first few smallest eigenvalues with the positive sign and corresponding eigenvectors simultaneously. Two numerical examples illustrate the effectiveness of the new algorithms, especially with suitable preconditioners.

Recently, in [38] we developed a 4D block steepest descent (4DBSD) algorithm and successfully solved the LR eigenvalue problems of dimension up to $2n = 5,650,410$ for a fully converged first-principle calculation of the excitation states of the benzene molecule. The 4DBSD algorithm is a variant of Algorithm 4.1, which is a block implementation of the 4D search technique, similar to the simultaneous Rayleigh quotient minimization method due to Longsine and McCormick [25].

We have applied the generic preconditioner $\Phi = (H - \mu I)^{-1}$ so far because of its generality for the case when no special structural property of K and/or M are known and can thus be exploited. When (1.1) is gotten from a plane-wave discretization, K and M are not explicitly constructed but rather exist in their mixed physical and frequency domain representations. Some natural "diagonal" preconditioners (similar to the one in [43] for the ground state computations) that reduce the high-wave-number components while leaving the low wave numbers untouched are emerging as very promising [37]. These issues will be a subject of future study.

In both [2] and this paper, we have focused on the case where the LR (RPA) eigenvalue problem has only real eigenvalues with eigenvalues 0 allowed. There are cases in which imaginary eigenvalues do occur. For example, the positive-definiteness condition of $A + B$ and/or $A - B$ is not met in [32, 35]. The development of efficient numerical methods for treating such large scale problems needs to be investigated, too.

Appendix. Best approximations: The singular/unequal dimension case. This appendix continues the investigation in section 3 to seek best approximate eigen-

pairs of H for given $\{\mathcal{U},\mathcal{V}\}$, a pair of approximate deflating subspaces of $\{K,M\}$ with $\dim(\mathcal{U})=\ell_1$ and $\dim(\mathcal{V})=\ell_2$. In section 3, we have treated the case in which $\ell_1=\ell_2$ and $W \stackrel{\text{def}}{=} U^{\mathrm{T}}V$ is nonsingular, where $U\in\mathbb{R}^{n\times\ell_1}$, $V\in\mathbb{R}^{n\times\ell_2}$ are the basis matrices of $\mathcal{U}$ and $\mathcal{V}$, respectively.

What follows is for the most general case: ℓ_1 and ℓ_2 may or may not be the same and W may or may not have full rank. Thus the case in section 3 is a special one whose treatment, however, is much simpler than that below for the general case.

Factorize

$$W = W_1^{\mathrm{T}} W_2, \quad W_i \in \mathbb{R}^{r\times\ell_i}, \quad r=\operatorname{rank}(W) \le \min_i \ell_i. \tag{A.1}$$

So both W_i have full row rank. Factorize[4]

$$W_i^{\mathrm{T}} = Q_i \begin{bmatrix} R_i \\ 0 \end{bmatrix} \quad \text{for } i=1,2, \tag{A.2}$$

where $R_i\in\mathbb{R}^{r\times r}$, $Q_i\in\mathbb{R}^{\ell_i\times\ell_i}$ $(i=1,2)$ are nonsingular.

Consider the best approximation to λ_1 by (3.1). We still have (3.3):

$$\rho(x,y) = \frac{\hat{u}^{\mathrm{T}}U^{\mathrm{T}}KU\hat{u} + \hat{v}^{\mathrm{T}}V^{\mathrm{T}}MV\hat{v}}{2|\hat{u}^{\mathrm{T}}W\hat{v}|}, \tag{3.3}$$

where $x=U\hat{u}$ and $y=V\hat{v}$ for some $\hat{u}\in\mathbb{R}^{\ell_1}$, $\hat{v}\in\mathbb{R}^{\ell_2}$. Note the correspondence between $x\in\mathcal{U}$ and $\hat{u}\in\mathbb{R}^{\ell_1}$ and that between $y\in\mathcal{V}$ and $\hat{v}\in\mathbb{R}^{\ell_2}$ are one-one. Let $\hat{x}=W_1\hat{u}\in\mathbb{R}^r$ and $\hat{y}=W_2\hat{v}\in\mathbb{R}^r$. Since $r\le\ell_i$, $\hat{u}$ may not be uniquely defined by $\hat{x}$; neither may $\hat{v}$ by $\hat{y}$. But using (A.2), we see that[5]

$$\hat{u} = Q_1^{-\mathrm{T}} \begin{bmatrix} R_1^{-\mathrm{T}}\hat{x} \\ u \end{bmatrix}, \quad \hat{v} = Q_2^{-\mathrm{T}} \begin{bmatrix} R_2^{-\mathrm{T}}\hat{y} \\ v \end{bmatrix},$$

where $u\in\mathbb{R}^{\ell_1-r}$, $v\in\mathbb{R}^{\ell_2-r}$ are arbitrary. Partition

$$Q_1^{-1}U^{\mathrm{T}}KUQ_1^{-\mathrm{T}} = \begin{matrix} r \\ \ell_1-r \end{matrix} \begin{bmatrix} \overset{r}{K_{11}} & \overset{\ell_1-r}{K_{12}} \\ K_{12}^{\mathrm{T}} & K_{22} \end{bmatrix}, \tag{A.3a}$$

$$Q_2^{-1}V^{\mathrm{T}}MVQ_2^{-\mathrm{T}} = \begin{matrix} r \\ \ell_2-r \end{matrix} \begin{bmatrix} \overset{r}{M_{11}} & \overset{\ell_2-r}{M_{12}} \\ M_{12}^{\mathrm{T}} & M_{22} \end{bmatrix}. \tag{A.3b}$$

We have

$$\hat{u}^{\mathrm{T}}U^{\mathrm{T}}KU\hat{u} = \begin{bmatrix} R_1^{-\mathrm{T}}\hat{x} \\ u \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} K_{11} & K_{12} \\ K_{12}^{\mathrm{T}} & K_{22} \end{bmatrix} \begin{bmatrix} R_1^{-\mathrm{T}}\hat{x} \\ u \end{bmatrix},$$

$$\hat{v}^{\mathrm{T}}V^{\mathrm{T}}MV\hat{v} = \begin{bmatrix} R_2^{-\mathrm{T}}\hat{y} \\ v \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} M_{11} & M_{12} \\ M_{12}^{\mathrm{T}} & M_{22} \end{bmatrix} \begin{bmatrix} R_2^{-\mathrm{T}}\hat{y} \\ v \end{bmatrix}.$$

[4]Computationally, this can be realized by the QR decompositions of W_i^{T}. For more generality in presentation, we do not assume that they have to be QR decompositions.

[5]A vector block in a partitioned vector is an empty block if it has dimension 0. The same can be said about a matrix block in a partitioned matrix if its row/column dimension is 0.

Given $\hat{x}$, $\hat{u}^{\mathrm{T}}U^{\mathrm{T}}KU\hat{u}$ is minimized at these u such that $K_{22}u = -K_{12}^{\mathrm{T}}R_1^{-\mathrm{T}}\hat{x}$. This equation always has a solution because that $Q_1^{-1}U^{\mathrm{T}}KUQ_1^{-\mathrm{T}}$ is positive semidefinite implies $\mathrm{span}(K_{12}^{\mathrm{T}}) \subseteq \mathrm{span}(K_{22})$, and its solution is not unique if K_{22} is singular. But the nonuniqueness does not matter as far as the minimal value of $\hat{u}^{\mathrm{T}}U^{\mathrm{T}}KU\hat{u}$ is concerned. The same can be said about $\hat{v}^{\mathrm{T}}V^{\mathrm{T}}MV\hat{v}$. In fact,

$$\min_u \hat{u}^{\mathrm{T}}U^{\mathrm{T}}KU\hat{u} = \hat{x}^{\mathrm{T}}R_1^{-1}\left(K_{11} - K_{12}K_{22}^{\dagger}K_{12}^{\mathrm{T}}\right)R_1^{-\mathrm{T}}\hat{x}, \tag{A.4a}$$

$$\min_v \hat{v}^{\mathrm{T}}V^{\mathrm{T}}MV\hat{v} = \hat{y}^{\mathrm{T}}R_2^{-1}\left(M_{11} - M_{12}M_{22}^{\dagger}M_{12}^{\mathrm{T}}\right)R_2^{-\mathrm{T}}\hat{y}, \tag{A.4b}$$

where $K_{22}^{\dagger}$ and $M_{22}^{\dagger}$ are the Moore–Penrose inverses of K_{22} and M_{22}, respectively. The minimums in (A.4) are attained at those u and v satisfying

$$K_{22}u = -K_{12}^{\mathrm{T}}R_1^{-\mathrm{T}}\hat{x}, \quad M_{22}v = -M_{12}^{\mathrm{T}}R_2^{-\mathrm{T}}\hat{y}.$$

Finally, the quantity in (3.1) is

$$\inf_{\hat{x},\hat{y}} \frac{\hat{x}^{\mathrm{T}}R_1^{-1}\left(K_{11} - K_{12}K_{22}^{\dagger}K_{12}^{\mathrm{T}}\right)R_1^{-\mathrm{T}}\hat{x} + \hat{y}^{\mathrm{T}}R_2^{-1}\left(M_{11} - M_{12}M_{22}^{\dagger}M_{12}^{\mathrm{T}}\right)R_2^{-\mathrm{T}}\hat{y}}{2|\hat{x}^{\mathrm{T}}\hat{y}|}$$

which, by [2, Theorem 3.1], is the smallest eigenvalue with the positive sign of $\hat{H}_{\mathrm{SR}}$:

$$\hat{H}_{\mathrm{SR}} = \begin{bmatrix} 0 & R_1^{-1}\left(K_{11} - K_{12}K_{22}^{\dagger}K_{12}^{\mathrm{T}}\right)R_1^{-\mathrm{T}} \\ R_2^{-1}\left(M_{11} - M_{12}M_{22}^{\dagger}M_{12}^{\mathrm{T}}\right)R_2^{-\mathrm{T}} & 0 \end{bmatrix} \in \mathbb{R}^{2r\times 2r}. \tag{A.5}$$

Now we turn to the best approximations to λ_j $(1 \le j \le k)$ by (3.2). Assume (A.1) and (A.2). Any $\hat{U}$, $\hat{V} \in \mathbb{R}^{n\times k}$ such that $\mathrm{span}(\hat{U}) \in \mathcal{U}$, $\mathrm{span}(\hat{V}) \in \mathcal{V}$, and $\hat{U}^{\mathrm{T}}\hat{V} = I_k$ can be written as

$$\hat{U} = UQ_1^{-\mathrm{T}}\widetilde{U}, \quad \hat{V} = VQ_2^{-\mathrm{T}}\widetilde{V},$$

where $\widetilde{U} \in \mathbb{R}^{\ell_1\times k}$, $\widetilde{V} \in \mathbb{R}^{\ell_2\times k}$ that make $\hat{U}^{\mathrm{T}}\hat{V} = I_k$, and vice versa. We note that necessarily

$$k = \mathrm{rank}(I_k) = \mathrm{rank}(\hat{U}^{\mathrm{T}}\hat{V}) \le \mathrm{rank}(W) = r.$$

We first look into what constraint is needed on $\widetilde{U}$ and $\widetilde{V}$ in order to enforce $\hat{U}^{\mathrm{T}}\hat{V} = I_k$. To this end, we partition

$$\widetilde{U} = \begin{matrix} r \\ \ell_1 - r \end{matrix} \begin{bmatrix} \widetilde{U}_1 \\ \widetilde{U}_2 \end{bmatrix}, \quad \widetilde{V} = \begin{matrix} r \\ \ell_2 - r \end{matrix} \begin{bmatrix} \widetilde{V}_1 \\ \widetilde{V}_2 \end{bmatrix}.$$

We have

$$\hat{U}^{\mathrm{T}}\hat{V} = \widetilde{U}^{\mathrm{T}}Q_1^{-1}W_1^{\mathrm{T}}W_2Q_2^{-\mathrm{T}}\widetilde{V} = \widetilde{U}^{\mathrm{T}}\begin{bmatrix} R_1 \\ 0 \end{bmatrix}\left(R_2^{\mathrm{T}}, 0\right)\widetilde{V} = \widetilde{U}_1^{\mathrm{T}}R_1R_2^{\mathrm{T}}\widetilde{V}_1.$$

Let $\hat{X} = R_1^{\mathrm{T}}\widetilde{U}_1$, $\hat{Y} = R_2^{\mathrm{T}}\widetilde{V}_1 \in \mathbb{R}^{r\times k}$. Then $\hat{U}^{\mathrm{T}}\hat{V} = I_k$ is equivalent to $\hat{X}^{\mathrm{T}}\hat{Y} = I_k$ which will be enforced henceforth, while $\widetilde{U}_2$ and $\widetilde{V}_2$ are arbitrary. Assume the partitioning in (A.3). We have

$$\hat{U}^{\mathrm{T}}K\hat{U} = \widetilde{U}^{\mathrm{T}}Q_1^{-1}U^{\mathrm{T}}KUQ_1^{-\mathrm{T}}\widetilde{U} = \begin{bmatrix} R_1^{-\mathrm{T}}\hat{X} \\ \widetilde{U}_2 \end{bmatrix}^{\mathrm{T}}\begin{bmatrix} K_{11} & K_{12} \\ K_{12}^{\mathrm{T}} & K_{22} \end{bmatrix}\begin{bmatrix} R_1^{-\mathrm{T}}\hat{X} \\ \widetilde{U}_2 \end{bmatrix},$$

$$\hat{V}^{\mathrm{T}}M\hat{V} = \widetilde{V}^{\mathrm{T}}Q_2^{-1}V^{\mathrm{T}}KVQ_2^{-\mathrm{T}}\widetilde{V} = \begin{bmatrix} R_2^{-\mathrm{T}}\hat{Y} \\ \widetilde{V}_2 \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} M_{11} & M_{12} \\ M_{12}^{\mathrm{T}} & M_{22} \end{bmatrix} \begin{bmatrix} R_2^{-\mathrm{T}}\hat{Y} \\ \widetilde{V}_2 \end{bmatrix}.$$

Given $\hat{X}$ and $\hat{Y}$, it can be verified that

$$\min_{\widetilde{U}_2} \operatorname{trace}(\hat{U}^{\mathrm{T}}K\hat{U}) = \operatorname{trace}(\hat{X}^{\mathrm{T}}R_1^{-1}[K_{11} - K_{12}K_{22}^{\dagger}K_{12}^{\mathrm{T}}]R_1^{-\mathrm{T}}\hat{X}), \tag{A.6a}$$

$$\min_{\widetilde{V}_2} \operatorname{trace}(\hat{V}^{\mathrm{T}}M\hat{V}) = \operatorname{trace}(\hat{Y}^{\mathrm{T}}R_2^{-1}[M_{11} - M_{12}M_{22}^{\dagger}M_{12}^{\mathrm{T}}]R_2^{-\mathrm{T}}\hat{Y}) \tag{A.6b}$$

with the minimums attained at those $\widetilde{U}_2$ and $\widetilde{V}_2$ satisfying

$$K_{22}\widetilde{U}_2 = -K_{12}^{\mathrm{T}}R_1^{-\mathrm{T}}\hat{X}, \quad M_{22}\widetilde{V}_2 = -M_{12}^{\mathrm{T}}R_2^{-\mathrm{T}}\hat{Y}.$$

Therefore the quantity in (3.2) is

$$\inf_{\hat{X}^{\mathrm{T}}\hat{Y}=I_k} \operatorname{trace}\left(\hat{X}^{\mathrm{T}}R_1^{-1}[K_{11} - K_{12}K_{22}^{\dagger}K_{12}^{\mathrm{T}}]R_1^{-\mathrm{T}}\hat{X} + \hat{Y}^{\mathrm{T}}R_2^{-1}[M_{11} - M_{12}M_{22}^{\dagger}M_{12}^{\mathrm{T}}]R_2^{-\mathrm{T}}\hat{Y}\right)$$

which, by [2, Theorem 3.2], is the sum of the k smallest eigenvalues with the positive sign of $\hat{H}_{\mathrm{SR}}$ defined by (A.5).

In summary, the *best approximations to some of the eigenvalues of H within the pair of approximate deflating subspaces are the eigenvalues of $\hat{H}_{\mathrm{SR}}$*. Denote by μ_j $(j = 1, \ldots, r)$ the eigenvalues with the positive sign of $\hat{H}_{\mathrm{SR}}$ in the ascending order and by $\hat{z}_j$ the associated eigenvectors:

$$\hat{H}_{\mathrm{SR}}\hat{z}_j = \mu_j\hat{z}_j, \quad \hat{z}_j = \begin{bmatrix} \hat{y}_j \\ \hat{x}_j \end{bmatrix}. \tag{A.7}$$

Following the derivations above, we conclude

$$\rho(\widetilde{x}_j, \widetilde{y}_j) = \mu_j \quad \text{for } j = 1, \ldots, r,$$

where

$$\widetilde{x}_j = UQ_1^{-\mathrm{T}} \begin{bmatrix} R_1^{-\mathrm{T}}\hat{x}_j \\ u_j \end{bmatrix}, \quad \widetilde{y}_j = VQ_2^{-\mathrm{T}} \begin{bmatrix} R_2^{-\mathrm{T}}\hat{y}_j \\ v_j \end{bmatrix} \tag{A.8}$$

for u_j and v_j satisfying

$$K_{22}u_j = -K_{12}^{\mathrm{T}}R_1^{-\mathrm{T}}\hat{x}_j, \quad M_{22}v_j = -M_{12}^{\mathrm{T}}R_2^{-\mathrm{T}}\hat{y}_j. \tag{A.9}$$

Naturally the approximate eigenvectors of H should be taken as

$$\widetilde{z}_j = \begin{bmatrix} \widetilde{y}_j \\ \widetilde{x}_j \end{bmatrix} \quad \text{for } j = 1, \ldots, r. \tag{A.10}$$

For the ease of reference, we summarize our findings into the following theorem.

THEOREM A.1. *Suppose that one of K, $M \in \mathbb{R}^{n\times n}$ is definite. Let $\{\mathcal{U}, \mathcal{V}\}$ be a pair of approximate deflating subspaces of $\{K, M\}$ with $\dim(\mathcal{U}) = \ell_1$ and $\dim(\mathcal{V}) = \ell_2$, and let $U \in \mathbb{R}^{n\times\ell_1}$, $V \in \mathbb{R}^{n\times\ell_2}$ be the basis matrices of $\mathcal{U}$ and $\mathcal{V}$, respectively. Let $\hat{H}_{\mathrm{SR}}$ be defined by* (A.5). *Then the best approximations to λ_1 in the sense of* (3.1) *or to λ_j $(1 \le j \le k)$ in the sense of* (3.2) *are the corresponding eigenvalues μ_j of $\hat{H}_{\mathrm{SR}}$, with the corresponding approximate eigenvectors given by* (A.8)–(A.10).

In [2, Theorem 2.9], we proved the approximate eigenpairs are unique for given $\{\mathcal{U},\mathcal{V}\}$ with nonsingular $U^{\mathrm{T}}V$, even though there are infinitely many different H_{SR} associated with the pair of subspaces. We are faced with the same question for $\hat{H}_{\text{SR}}$ in whose construction there are three nonunique choices:

$$\text{(A.11)}\qquad \begin{cases} 1.\ \text{Factorizations in (A.2) are not unique.} \\ 2.\ \text{Factorization } W = W_1^{\mathrm{T}}W_2 \text{ in (A.1) is not unique.} \\ 3.\ \text{Basis matrices } U \text{ and } V \text{ are not unique.} \end{cases}$$

The question would arise if different $\hat{H}_{\text{SR}}$ could produce different approximate eigenpairs. This is addressed by the following theorem.

THEOREM A.2. *Suppose that one of K, $M \in \mathbb{R}^{n\times n}$ is definite. Let $\mathcal{U}$ and $\mathcal{V}$ be two subspaces of $\mathbb{R}^n$ with basis matrices $U \in \mathbb{R}^{n\times\ell_1}$, $V \in \mathbb{R}^{n\times\ell_2}$, respectively. Define $\hat{H}_{\text{SR}}$ by* (A.5)*. Then*

1. *the approximate eigenvalues, i.e., the eigenvalues of $\hat{H}_{\text{SR}}$, are invariant with respect to any of the nonuniqueness listed in* (A.11) *for constructing $\hat{H}_{\text{SR}}$;*
2. *the approximate eigenvectors by* (A.8)–(A.10) *are invariant with respect to any of the nonuniqueness listed in* (A.11) *if and only if both K_{22} and M_{22} are definite.*[6]

Proof. To see the first conclusion, we notice that the infimum (3.2) only depends on $\{\mathcal{U},\mathcal{V}\}$ and is invariant with respect to any of the nonuniqueness in (A.11) for $1 \le k \le r$. Since the infimum is the sum of the first k smallest eigenvalues with the positive sign of $\hat{H}_{\text{SR}}$, let k go from 1 to r to conclude that the eigenvalues with the positive sign of $\hat{H}_{\text{SR}}$ are invariant with respect to any of the nonuniqueness in (A.11); so are all eigenvalues of $\hat{H}_{\text{SR}}$.

For the second conclusion, let us first select one choice for each of them in (A.11), namely basis matrices U and V, a factorization $W = W_1^{\mathrm{T}}W_2$ in (A.1), and two factorizations in (A.2). Let $H_0 \overset{\text{def}}{=} \hat{H}_{\text{SR}}$ with these selected choices, and suppose that both K_{22} and M_{22} are definite. We shall now prove that the approximate eigenvectors are invariant with respect to any variation to the selected ones. Along the way, we will also see the definiteness of K_{22} and M_{22} does not change with the variations either.

1. *Invariance with respect to different choices of factorizations in* (A.2). Any factorizations other than the given ones in (A.2) can be written as

$$\text{(A.12)}\qquad W_i^{\mathrm{T}} = Q_i \begin{bmatrix} S_{i1} & \\ & S_{i2} \end{bmatrix} \begin{bmatrix} S_{i1}^{-1}R_i \\ 0 \end{bmatrix},$$

for some nonsingular $S_{i1} \in \mathbb{R}^{r\times r}$, $S_{i2} \in \mathbb{R}^{(\ell_i-r)\times(\ell_i-r)}$. Denote by $H_1 \overset{\text{def}}{=} \hat{H}_{\text{SR}}$ with given U, V and (A.1), and (A.12). Perform substitutions

$$Q_i \begin{bmatrix} S_{i1} & \\ & S_{i2} \end{bmatrix} \leftarrow Q_i,\ S_{i1}^{-1}R_i \leftarrow R_i,\ S_{1i}^{-1}K_{ij}S_{1j}^{-\mathrm{T}} \leftarrow K_{ij},\ S_{2i}^{-1}M_{ij}S_{2j}^{-\mathrm{T}} \leftarrow M_{ij}$$

to see $H_0 = H_1$ and that the approximate eigenvectors for H by (A.8)–(A.10) do not change. Also the definiteness of K_{22} and M_{22} does not change with the variation in (A.12).

[6] Although K_{22} and M_{22} vary with the nonunique choices in (A.11), this condition is well-stated because the ranks of K_{22} and M_{22} remain unchanged with respect to the choices. This will be confirmed in the proof of this theorem as well.

2. *Invariance with respect to different choice of factorization* $W = W_1^{\mathrm{T}} W_2$. Any factorization other than the given one in (A.1), can be written as

$$W = W_1^{\mathrm{T}} S S^{-1} W_2 = (S^{\mathrm{T}} W_1)^{\mathrm{T}} (S^{-1} W_2) \tag{A.13}$$

for some nonsingular $S \in \mathbb{R}^{r\times r}$. Define $H_1 \overset{\text{def}}{=} \hat{H}_{\text{SR}}$ with given U, V, and (A.13) and

$$(S^{\mathrm{T}} W_1)^{\mathrm{T}} = Q_1 \begin{bmatrix} R_1 S \\ 0 \end{bmatrix}, \quad (S^{-1} W_2)^{\mathrm{T}} = Q_2 \begin{bmatrix} R_2 S^{-\mathrm{T}} \\ 0 \end{bmatrix}. \tag{A.14}$$

Since we just proved the invariance with respect to different choices of factorizations in (A.2), it suffices to prove that the approximate eigenvectors obtained through H_0 and H_1 are the same. Upon using substitutions $R_1 S \leftarrow R_1$ and $R_2 S^{-\mathrm{T}} \leftarrow R_2$, we find

$$H_1 = \left(S^{-1} \oplus S^{\mathrm{T}}\right) H_0 \left(S^{-1} \oplus S^{\mathrm{T}}\right)^{-1},$$

and thus the relationships between the eigenvectors for H_0 and H_1. It can then be verified that the approximate eigenvectors obtained through H_0 and H_1 via (A.8)–(A.10) are the same. Also, the definiteness of K_{22} and M_{22} does not change with the variation in (A.13).

3. *Invariance with respect to different choices of basis matrices.* Given basis matrices U and V of $\mathcal{U}$ and $\mathcal{V}$, respectively, any other basis matrices can be written as UR and VS for some nonsingular $R \in \mathbb{R}^{\ell_1\times\ell_1}$, $S \in \mathbb{R}^{\ell_2\times\ell_2}$. Define $H_1 \overset{\text{def}}{=} \hat{H}_{\text{SR}}$ with UR and VS, and

$$W = (UR)^{\mathrm{T}}(VS) = R^{\mathrm{T}} U^{\mathrm{T}} V S = (W_1 R)^{\mathrm{T}} (W_2 S), \tag{A.15}$$

and

$$(W_1 R)^{\mathrm{T}} = R^{\mathrm{T}} Q_1 \begin{bmatrix} R_1 \\ 0 \end{bmatrix}, \quad (W_2 S)^{\mathrm{T}} = S^{\mathrm{T}} Q_2 \begin{bmatrix} R_2 \\ 0 \end{bmatrix}. \tag{A.16}$$

By the two invariance properties we just proved, it suffices to prove that the approximate eigenvectors obtained through H_0 and H_1 are the same. Perform substitutions

$$UR \leftarrow U, \quad VS \leftarrow V, \quad R^{\mathrm{T}} Q_1 \leftarrow Q_1, \quad S^{\mathrm{T}} Q_2 \leftarrow Q_2$$

to see $H_0 = H_1$ and that the approximate eigenvectors for H by (A.8)–(A.10) do not change. Again the definiteness of K_{22} and M_{22} does not change with the variation from U to UR and from V to VS.

Finally, if K_{22} is singular, then u_j satisfying the first equation in (A.9) is not unique. In fact, if u_j is one, any $u_j + g$ is another, for any g in the kernel of K_{22}. So the defining equation in (A.8) for $\widetilde{x}_j$ gives

$$\widetilde{x}_j = U Q_1 \begin{bmatrix} R_1^{-\mathrm{T}} \hat{x}_j \\ -K_{22}^{\dagger} K_{12}^{\mathrm{T}} R_1^{-\mathrm{T}} \hat{x}_j + g \end{bmatrix}, \tag{A.17}$$

leading to different approximate eigenvectors as g varies within the kernel of K_{22}. The same thing happens if M_{22} is singular. ∎

The proof of Theorem A.2 exposes the cause for the approximate eigenvectors by (A.8)–(A.10) not to be uniquely determined, namely, one of the equations in (A.9) may have infinitely many solutions.[7] When that's the case, we can either always take

$$u_j = -K_{22}^{\dagger} K_{12}^{\mathrm{T}} R_1^{-\,\mathrm{T}} \hat{x}_j, \quad v_j = -M_{22}^{\dagger} M_{12}^{\mathrm{T}} R_1^{-\,\mathrm{T}} \hat{y}_j$$

or settle the nonuniqueness by

$$\min_{g,h} \left\{ \|K\widetilde{x}_j - \mu_j \widetilde{y}_j\|_2^2 + \|M\widetilde{y}_j - \mu_j \widetilde{x}_j\|_2^2 \right\} \tag{A.18}$$

over all g in the kernel of K_{22} and h in the kernel of M_{22}, upon noticing (A.17) and

$$\widetilde{y}_j = V Q_2^{-\,\mathrm{T}} \begin{bmatrix} R_2^{-\,\mathrm{T}} \hat{y}_j \\ -M_{22}^{\dagger} M_{12}^{\mathrm{T}} R_2^{-\,\mathrm{T}} \hat{y}_j + h \end{bmatrix}. \tag{A.19}$$

Finally it can be seen that (A.18) is a least squares problem in g and h.

The next theorem too says that there are Cauchy-like interlacing inequalities for $\widehat{H}_{\mathrm{SR}}$, similarly to [2, Theorem 4.1].

THEOREM A.3. *Assume the conditions of Theorem* A.1. *Then*

$$\lambda_i \le \mu_i \le \lambda_{i+2n-(\ell_1+\ell_2)} \quad \textit{for } 1 \le i \le r, \tag{A.20}$$

where we set $\lambda_{i+2n-(\ell_1+\ell_2)} = \infty$ *if* $i + 2n - (\ell_1 + \ell_2) > n$.

Proof. Suppose for the moment that both K and M are positive definite. Recall the equivalence between the eigenvalue problem (1.1) and the one for

$$\boldsymbol{A} - \lambda \boldsymbol{B} \equiv \begin{bmatrix} M & 0 \\ 0 & K \end{bmatrix} - \lambda \begin{bmatrix} 0 & I_n \\ I_n & 0 \end{bmatrix}. \tag{A.21}$$

$\boldsymbol{A}$ is symmetric positive definite if K and M are. Let

$$Z = \begin{bmatrix} V Q_2^{-\,\mathrm{T}} (R_2^{-\,\mathrm{T}} \oplus I_{\ell_2 - r}) & \\ & U Q_1^{-\,\mathrm{T}} (R_1^{-\,\mathrm{T}} \oplus I_{\ell_1 - r}) \end{bmatrix},$$

which has full column rank. It can be verified, upon using (A.3), that

$$Z^{\mathrm{T}} \boldsymbol{A} Z = \begin{bmatrix} \widehat{M} & 0 \\ 0 & \widehat{K} \end{bmatrix}, \quad Z^{\mathrm{T}} \boldsymbol{B} Z = \begin{bmatrix} 0 & \widehat{I}^{\mathrm{T}} \\ \widehat{I} & 0 \end{bmatrix},$$

where

$$\begin{aligned} \widehat{M} &= \begin{bmatrix} R_2^{-1} & \\ & I_{\ell_2 - r} \end{bmatrix} \begin{bmatrix} M_{11} & M_{12} \\ M_{12}^{\mathrm{T}} & M_{22} \end{bmatrix} \begin{bmatrix} R_2^{-\,\mathrm{T}} & \\ & I_{\ell_2 - r} \end{bmatrix}, \\ \widehat{K} &= \begin{bmatrix} R_1^{-1} & \\ & I_{\ell_1 - r} \end{bmatrix} \begin{bmatrix} K_{11} & K_{12} \\ K_{12}^{\mathrm{T}} & K_{22} \end{bmatrix} \begin{bmatrix} R_1^{-\,\mathrm{T}} & \\ & I_{\ell_1 - r} \end{bmatrix}, \\ \widehat{I} &= \begin{bmatrix} I_r & \\ & 0 \end{bmatrix} \in \mathbb{R}^{\ell_1 \times \ell_2}. \end{aligned}$$

$Z^{\mathrm{T}} \boldsymbol{A} Z$ is positive definite because $\boldsymbol{A}$ is. Note the eigenvalues of $\boldsymbol{B} - \lambda \boldsymbol{A}$ are $\pm \lambda_i^{-1}$ and

$$-\lambda_1^{-1} \le -\lambda_2^{-1} \le \cdots \le -\lambda_n^{-1} < \lambda_n^{-1} \le \cdots \le \lambda_2^{-1} \le \lambda_1^{-1}.$$

[7] By default, one of K and M is definite. Thus at most one of K_{22} and M_{22} is singular.

Denote the nonzero eigenvalues[8] of $Z^{\mathrm{T}}\boldsymbol{B}Z - \lambda Z^{\mathrm{T}}\boldsymbol{A}Z$ by $\pm\sigma_i$ ordered as

$$-\sigma_1 \le \cdots \le -\sigma_r < \sigma_r \le \cdots \le \sigma_1.$$

All the rest $\ell_1 + \ell_2 - 2r$ eigenvalues of $Z^{\mathrm{T}}\boldsymbol{B}Z - \lambda Z^{\mathrm{T}}\boldsymbol{A}Z$ are 0s. Now apply Cauchy's interlacing inequalities (extended for the generalized eigenvalue problem) to $\boldsymbol{B} - \lambda\boldsymbol{A}$ and $Z^{\mathrm{T}}\boldsymbol{B}Z - \lambda Z^{\mathrm{T}}\boldsymbol{A}Z$ to get for $1 \le i \le r$,

$$\lambda_i^{-1} \ge \sigma_i \ge \begin{cases} \lambda_{i+2n-(\ell_1+\ell_2)}^{-1} & \text{if } i + 2n - (\ell_1 + \ell_2) \le n, \\ 0 & \text{otherwise.} \end{cases}$$

Equivalently, $\lambda_i \le \sigma_i^{-1} \le \lambda_{i+2n-(\ell_1+\ell_2)}$ for $1 \le i \le r$. It remains to show that $\mu_i = \sigma_i^{-1}$ for $1 \le i \le r$. To this end, we let

$$Z_1 = \begin{bmatrix} I_r & 0 \\ -K_{22}^{-1}K_{12}^{\mathrm{T}}R_1^{-\,\mathrm{T}} & I_{\ell_1-r} \end{bmatrix}, \quad Z_2 = \begin{bmatrix} I_r & 0 \\ -M_{22}^{-1}M_{12}^{\mathrm{T}}R_2^{-\,\mathrm{T}} & I_{\ell_2-r} \end{bmatrix}.$$

It can be verified that $Z_1^{\mathrm{T}}\widehat{I}Z_2 = \widehat{I}$ and

$$Z_1^{\mathrm{T}}\widehat{K}Z_1 = \begin{bmatrix} R_1^{-1}\widehat{K}_{11}R_1^{-\,\mathrm{T}} & 0 \\ 0 & K_{22} \end{bmatrix}, \quad \widehat{K}_{11} = K_{11} - K_{12}K_{22}^{-1}K_{12}^{\mathrm{T}},$$
$$Z_2^{\mathrm{T}}\widehat{M}Z_2 = \begin{bmatrix} R_2^{-1}\widehat{M}_{11}R_2^{-\,\mathrm{T}} & 0 \\ 0 & M_{22} \end{bmatrix}, \quad \widehat{M}_{11} = M_{11} - M_{12}M_{22}^{-1}M_{12}^{\mathrm{T}}.$$

The eigenvalues of $Z^{\mathrm{T}}\boldsymbol{B}Z - \lambda Z^{\mathrm{T}}\boldsymbol{A}Z$ are the same as these of

$$(Z_1 \oplus Z_2)^{\mathrm{T}}\left(Z^{\mathrm{T}}\boldsymbol{B}Z - \lambda Z^{\mathrm{T}}\boldsymbol{A}Z\right)(Z_1 \oplus Z_2)$$

which is a 4×4 block matrix pencil and becomes, after switching its second and third row and its second and third column,

$$\begin{bmatrix} 0 & I_r \\ I_r & 0 \end{bmatrix} \oplus 0_{(\ell_1+\ell_2-2r)\times(\ell_1+\ell_2-2r)} - \lambda \begin{bmatrix} R_2^{-1}\widehat{M}_{11}R_2^{-\,\mathrm{T}} & 0 \\ 0 & R_1^{-1}\widehat{K}_{11}R_1^{-\,\mathrm{T}} \end{bmatrix} \oplus \begin{bmatrix} M_{22} & 0 \\ 0 & K_{22} \end{bmatrix}.$$

Thus the nonzero eigenvalues of $Z^{\mathrm{T}}\boldsymbol{B}Z - \lambda Z^{\mathrm{T}}\boldsymbol{A}Z$ are the same as these of

$$\begin{bmatrix} 0 & I_r \\ I_r & 0 \end{bmatrix} - \lambda \begin{bmatrix} R_2^{-1}\widehat{M}_{11}R_2^{-\,\mathrm{T}} & 0 \\ 0 & R_1^{-1}\widehat{K}_{11}R_1^{-\,\mathrm{T}} \end{bmatrix},$$

which in turn are the same as the reciprocals of the eigenvalues of $\widehat{H}_{\mathrm{SR}}$, i.e., $\sigma_i = \mu_i^{-1}$, as expected.

Consider now that K is singular. We seek to perturb K to a positive definite $K(\epsilon)$ for all $\epsilon > 0$ in such a way that the resulting $\widehat{H}_{\mathrm{SR}}(\epsilon) \to \widehat{H}_{\mathrm{SR}}$ as $\epsilon \to 0^+$ and then use the limiting argument. In (A.3a) since $\mathrm{span}(K_{12}^T) \subseteq \mathrm{span}(K_{22})$, there exists orthogonal $P \in \mathbb{R}^{(\ell_1-r)\times(\ell_1-r)}$ such that

$$P^T K_{12}^T = \begin{matrix} r_1 \end{matrix} \begin{bmatrix} \overset{r}{\widetilde{K}_{12}^T} \\ 0 \end{bmatrix}, \quad P^T K_{22} P = \begin{matrix} r_1 \end{matrix} \begin{bmatrix} \overset{r_1}{\widetilde{K}_{22}} & 0 \\ 0 & 0 \end{bmatrix} \tag{A.22}$$

[8] These nonzero eigenvalues are the same as those of $(Z^{\mathrm{T}}\boldsymbol{A}Z)^{-1/2}(Z^{\mathrm{T}}\boldsymbol{B}Z)(Z^{\mathrm{T}}\boldsymbol{A}Z)^{-1/2} = \left[\begin{smallmatrix} 0 & C \\ C^{\mathrm{T}} & 0 \end{smallmatrix}\right]$, where $C = \widehat{M}^{-1/2}\widehat{I}\widehat{K}^{-1/2}$. Thus σ_i are the nonzero singular values of C.

and $\widetilde{K}_{22}$ is positive definite and diagonal. Let $E = UQ_1^{-\mathrm{T}}(I_r \oplus P) \in \mathbb{R}^{n\times\ell_1}$. It can be seen that $\mathrm{rank}(E) = \ell_1$, and hence there exists $E_\perp \in \mathbb{R}^{n\times(n-\ell_1)}$ such that $[E, E_\perp]$ is invertible. Let $G = [E, E_\perp]^{-1}$ and set

$$K(\epsilon) = K + \epsilon G^T G$$

which is positive definite. Define accordingly $H(\epsilon)$ and denote its eigenvalues by $\pm\lambda_i(\epsilon)$, and define $\widehat{H}_{\text{SR}}(\epsilon)$ and denote its eigenvalues by $\pm\mu_i(\epsilon)$ as well. Using (A.3a) and $E^T G^T G E = I_{\ell_1}$, we have

$$E^{\mathrm{T}} K(\epsilon) E = \begin{bmatrix} K_{11} & \widetilde{K}_{12} & 0 \\ \widetilde{K}_{12}^T & \widetilde{K}_{22} & 0 \\ 0 & 0 & 0 \end{bmatrix} + \epsilon I_{\ell_1},$$

$$Q_1^{-1} U^{\mathrm{T}} K(\epsilon) U Q_1^{-\mathrm{T}} = \begin{bmatrix} K_{11} & K_{12} \\ K_{12}^{\mathrm{T}} & K_{22} \end{bmatrix} + \epsilon I_{\ell_1}.$$

Let $K_{11}(\epsilon) := K_{11} + \epsilon I$ and $K_{22}(\epsilon) := K_{22} + \epsilon I$. We have by (A.22)

$$\begin{aligned}[K_{22}(\epsilon)]^{-1} K_{12}^{\mathrm{T}} &= \left(P \begin{bmatrix} \widetilde{K}_{22} & 0 \\ 0 & 0 \end{bmatrix} P^T + \epsilon I \right)^{-1} P \begin{bmatrix} \widetilde{K}_{12}^T \\ 0 \end{bmatrix} \\ &= P \begin{bmatrix} (\widetilde{K}_{22} + \epsilon I)^{-1} \widetilde{K}_{12}^T \\ 0 \end{bmatrix} \to P \begin{bmatrix} \widetilde{K}_{22}^{-1} \widetilde{K}_{12}^T \\ 0 \end{bmatrix} = K_{22}^{\dagger} K_{12}^{\mathrm{T}}\end{aligned}$$

as $\epsilon \to 0^+$. Recall how $\widehat{H}_{\text{SR}}$ is constructed, we can see that the resulting $\widehat{H}_{\text{SR}}(\epsilon)$ as the result of the perturbation is the same as $\widehat{H}_{\text{SR}}$, except in its $(1,2)$ block which is

$$R_1^{-1} \left(K_{11}(\epsilon) - K_{12}[K_{22}(\epsilon)]^{-1} K_{12}^{\mathrm{T}} \right) R_1^{-\mathrm{T}} \to R_1^{-1} \left(K_{11} - K_{12} K_{22}^{\dagger} K_{12}^{\mathrm{T}} \right) R_1^{-\mathrm{T}}$$

as $\epsilon \to 0^+$. That is $\widehat{H}_{\text{SR}}(\epsilon) \to \widehat{H}_{\text{SR}}$ as $\epsilon \to 0^+$. Therefore, by the eigenvalue continuity with respect to a matrix's entries [40], we have $\mu_i(\epsilon) \to \mu_i$ as $\epsilon \to 0^+$. For the same reason, $\lambda_i(\epsilon) \to \lambda_i$, too, as $\epsilon \to 0^+$. Finally, since $K(\epsilon)$ and M are positive definite, by what we just proved, we have

$$\lambda_i(\epsilon) \le \mu_i(\epsilon) \le \lambda_{i+2n-(\ell_1+\ell_2)}(\epsilon) \quad \text{for } 1 \le i \le r. \tag{A.23}$$

Now letting $\epsilon \to 0^+$ in (A.23) leads to the desired inequalities in (A.20). ∎

Remark A.1. Noticeably, the treatment above is much more involved than the nonsingular case in section 3. Certainly an argument can be made to not use $\{\mathcal{U}, \mathcal{V}\}$ with a singular W at all, since by [2, Lemma 2.7], we know that W is nonsingular if $\{\mathcal{U}, \mathcal{V}\}$ is exact and if $\dim\mathcal{U} = \dim\mathcal{V}$. But, in practice, especially at the beginning of an iterative process, it is hard to guarantee this is so at all times. The treatment presented here, albeit involved, shows that the optimums in (3.1) and (3.2) can still be realized. An alternative and simpler treatment for the singular case at a tradeoff of achieving only suboptimal approximations to (3.1) and (3.2) is as follows. Suppose (A.1) and (A.2). We have

$$(R_1^{-1}, 0) Q_1^{-1} U^{\mathrm{T}} V Q_2^{-\mathrm{T}} \begin{bmatrix} R_2 \\ 0 \end{bmatrix} = I_r.$$

After substitutions

$$U \leftarrow U Q_1^{-\mathrm{T}} \begin{bmatrix} R_1 \\ 0 \end{bmatrix}, \quad V \leftarrow V Q_2^{-\mathrm{T}} \begin{bmatrix} R_2 \\ 0 \end{bmatrix},$$

two new subspaces $\mathcal{U}$ and $\mathcal{V}$ with dimension r are born with new basis matrices U and V satisfying $U^{\mathrm{T}}V = I_r$, returning to the nonsingular W case in section 3.

Acknowledgments. We thank Dr. D. Rocca for providing us with the the test matrices used in Example 5.1. We are grateful to the anonymous referees for their careful reading of the manuscript and valuable comments.

REFERENCES

[1] E. Anderson, Z. Bai, C. Bischof, S. Blackford, J. Demmel, J. Dongarra, J. Du Croz, A. Greenbaum, S. Hammarling, A. McKenney, and D. Sorensen, *LAPACK Users' Guide* 3rd ed., SIAM, Philadelphia, 1999.

[2] Z. Bai and R.-C. Li, *Minimization principles for the linear response eigenvalue problem,* I*: Theory*, SIAM J. Matrix Anal. Appl., 33 (2012), pp. 1075–1100.

[3] P. Benner and H. Fassbender, *An implicitly restarted symplectic Lanczos method for the Hamiltonian eigenvalue problem*, Linear Algebra Appl., 263 (1997), pp. 75–111.

[4] P. Benner, H. Fassbender, and M. Stoll, *A Hamiltonian Krylov-Schur-type method based on the symplectic Lanczos process*, Linear Algebra Appl., 435 (2011), pp. 578–600.

[5] P. Benner, V. Mehrmann, and H. Xu, *A numerically stable structure preserving method for computing the eigenvalues of real Hamiltonian or symplectic pencils*, Numer. Math., 78 (1998), pp. 329–357.

[6] D. Bohm and D. Pines, *A collective description of electron interactions:* III*. Coulomb interactions in a degenerate electron gas*, Phys. Rev. (2), 92 (1953), pp. 609–625.

[7] W. W. Bradbury and R. Fletcher, *New iterative methods for solution of the eigenproblem*, Numer. Math., 9 (1966), pp. 259–267.

[8] A. Bunse-Gerstner and V. Mehrmann, *A symplectic QR like algorithm for the solution of the real algebraic Riccati equation*, IEEE Trans. Automat. Control, 31 (1986), pp. 1104–1113.

[9] M. Challacombe, *Linear scaling solution of the time-dependent self-consisten-field equations*, available online at arXiv:1001.2586v2, 2010.

[10] B. E. Chi, *The eigenvalue problem for collective motion in the random phase approximation*, Nuclear Phys., A146 (1970), pp. 449–456.

[11] J. Demmel, *Applied Numerical Linear Algebra*, SIAM, Philadelphia, 1997.

[12] P. Giannozzi et al., *QUANTUM ESPRESSO: A modular and open-source software project for quantum simulations of materials*, J. Phys. Condensed Matter, 21 (2009), 395502, available online at http://www.quantum-espresso.org/.

[13] U. Flaschka, W.-W. Lin, and J.-L. Wu, *A KQZ algorithm for solving linear-response eigenvalue equations*, Linear Algebra Appl., 165 (1992), pp. 93–123.

[14] R. W. Freund, *SPRIM: Structure-preserving reduced-order interconnect macromodeling*, in Proceedings of the International Conference of Computer Aided Design, IEEE Computer Society, Los Alamitos, CA, 2004, pp. 80–87.

[15] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd ed., Johns Hopkins University Press, Baltimore, MD, 1996.

[16] M. Grüning, A. Marini, and X. Gonze, *Exciton-plasmon states in nanoscale materials: Breakdown of the Tamm-Dancoff approximation*, Nano Letters, 9 (2009), pp. 2820–2824.

[17] M. Grüning, A. Marini, and X. Gonze, *Implementation and Testing of Lanczos-based Algorithms for Random-Phase Approximation Eigenproblems*, Technical report, 2011, arXiv:1102.3909v1.

[18] U. Hetmaniuk and R. Lehoucq, *Basis selection in LOBPCG*, J. Comput. Phys., 218 (2006), pp. 324–332.

[19] M. E. Hochstenbach, *A Jacobi-Davidson type method for the product eigenvalue problem*, J. Comput. Appl. Math., 212 (2008), pp. 46–62.

[20] K. J. Kerns and A. T. Yang, *Preservation of passivity during RLC network reduction via split congruence transformations*, IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst., 17 (1998), pp. 582–591.

[21] A. V. Knyazev, *A preconditioned conjugate gradient method for eigenvalue problems and its implementation in a subspace*, Internat Ser. Numer. Math., 96 (1991), pp. 143–154.

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

[23] D. Kressner, *A periodic Krylov-Schur algorithm for large matrix products*, Numer. Math., 103 (2006), pp. 461–483.

[24] R.-C. Li and Z. Bai, *Structure-preserving model reductions using a Krylov subspace projection formulation*, Commun. Math. Sci., 3 (2005), pp. 179–199.
[25] D. E. Longsine and S. F. McCormick, *Simultaneous Rayleigh-quotient minimization methods for $Ax = \lambda Bx$*, Linear Algebra Appl., 34 (1980), pp. 195–234.
[26] M. J. Lucero, A. M. N. Niklasson, S. Tretiak, and M. Challacombe, *Molecular-orbital-free algorithm for excited states in time-dependent perturbation theory*, J. Chem. Phys., 129 (2008), 064114.
[27] M. A. Marques, A. Castro, and A. Rubio, *Assessment of exchange-correlation functionals for the calculation of dynamical properties of small clusters in time-dependent density functional theory*, J. Chem. Phys., 115 (2001), pp. 3006–3014.
[28] A. Muta, J.-I. Iwata, Y. Hashimoto, and K. Yabana, *Solving the RPA eigenvalue equation in real-space*, Progr. Theoret. Phys., 108 (2002), pp. 1065–1076.
[29] J. Nocedal and S. Wright, *Numerical Optimization*, 2nd ed., Springer, New York, 2006.
[30] J. Olsen, H. J. Aa. Jensen, and P. Jørgensen, *Solution of the large matrix equations which occur in response theory*, J. Comput. Phys., 74 (1988), pp. 265–282.
[31] E. E. Ovtchinnikov, *Jacobi correction equation, line search, and conjugate gradients in Hermitian eigenvalue computation* I*: Computing an extreme eigenvalue*, SIAM J. Numer. Anal., 46 (2008), pp. 2567–2592.
[32] P. Papakonstantinou, *Reduction of the RPA eigenvalue problem and a generalized Cholesky decomposition for real-symmetric matrices*, Europhys. Lett., EPL, 78 (2007), 12001.
[33] B. N. Parlett, *The Symmetric Eigenvalue Problem*, SIAM, Philadelphia, 1998.
[34] B. T. Polyak, *Introduction to Optimization*, Optimization Software, New York, 1987.
[35] P. Ring, Z.-Y. Ma, N. Van Giai, D. Vretenar, A. Wandelt, and L.-G. Gao, *The time-dependent relativistic mean-field theory and the random phase approximation*, Nuclear Phys. A, 249 (2001), pp. 249–268.
[36] P. Ring and P. Schuck, *The Nuclear Many-Body Problem*, Springer-Verlag, New York, 1980.
[37] D. Rocca, *Iterative Diagonalization of Non-Hermitian Eigenproblems in Time-Dependent Density Functional and Many-Body Perturbation Theory*, Presentation at Session B39, the APS Marching Meeting, Boston, 2012.
[38] D. Rocca, Z. Bai, R.-C. Li, and G. Galli, *A block variational procedure for the iterative diagonalization of non-Hermitian random-phase approximation matrices*, J. Chem. Phys., 136 (2012), 034111.
[39] A. Ruhe, *Iterative eigenvalue algorithms for large symmetric matrices*, in Numerische Behandlung von Eigenwertaufgaben (Tagung, Oberwolfach, 1972), ISNM 24, Birkhäuser, Basel, 1974, pp. 97–115.
[40] G. W. Stewart and J.-G. Sun, *Matrix Perturbation Theory*, Academic Press, Boston, 1990.
[41] R. E. Stratmann, G. E. Scuseria, and M. J. Frisch, *An efficient implementation of time-dependent density-functional theory for the calculation of excitation of large molecules*, J. Chem. Phys., 109 (1998), pp. 8218–8824.
[42] W. Sun and Y. Yuan, *Optimization Theory and Methods: Nonlinear Programming*, Springer, New York, 2010.
[43] M. P. Teter, M. C. Payne, and D. C. Allan, *Solution of Schrödinger's equation for large systems*, Phys. Rev. B, 40 (1989), pp. 12255–12263.
[44] D. J. Thouless, *Vibrational states of nuclei in the random phase approximation*, Nuclear Phys., 22 (1961), pp. 78–95.
[45] D. J. Thouless, *The Quantum Mechanics of Many-Body Systems*, Academic Press, New York, London, 1972.
[46] S. Tretiak, C. M. Isborn, A. M. N. Niklasson, and M. Challacombe, *Representation independent algorithms for molecular response calculations in time-dependent self-consistent field theories*, J. Chem. Phys., 130 (2009), 054111.
[47] S. Tretiak and S. Mukamel, *Density matrix analysis and simulation of electronic excitations in conjugated and aggregated molecules*, Chem. Rev., 102 (2002), pp. 3171–3212.
[48] E. V. Tsiper, *Variational procedure and generalized Lanczos recursion for small-amplitude classical oscillations*, JETP Lett., 70 (1999), pp. 751–755.
[49] E. V. Tsiper, *A classical mechanics technique for quantum linear response*, J. Phys. B: At. Mol. Opt. Phys., 34 (2001), pp. L401–L407.
[50] H. A. van der Vorst, *A generalized Lanczos scheme*, Math. Comp., 39 (1982), pp. 559–561.
[51] D. S. Watkins, *The Matrix Eigenvalue Problems: GR and Krylov Subspace Methods*, SIAM, Philadelphia, 2007.
[52] J. H. Wilkinson, *The Algebraic Eigenvalue Problem*, Oxford University Press, Oxford, 1965.