

Minimization Principles for the Linear Response Eigenvalue Problem I: Theory

Zhaojun Bai* Ren-Cang Li †

June 28, 2011

Abstract

We present two theoretical results for the linear response eigenvalue problem. The first result is a minimization principle for the sum of the smallest few positive eigenvalues. The second result is a couple of Cauchy-like interlacing inequalities. Although the linear response eigenvalue problem is a nonsymmetric eigenvalue problem, these results mirror the well-known Courant-Fischer trace minimization principle and Cauchy interlacing inequalities for the symmetric eigenvalue problem.

2000 Mathematics Subject Classification. Primary 65L15. Secondary 15A18, 81Q15

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

1 Introduction

In this paper, we consider the eigenvalue problem of the form:

$$Hz \equiv \begin{pmatrix} 0 & K \\ M & 0 \end{pmatrix} \begin{pmatrix} y \\ x \end{pmatrix} = \lambda \begin{pmatrix} y \\ x \end{pmatrix} \equiv \lambda z, \quad (1.1)$$

where K and M are $n \times n$ symmetric positive semi-definite matrices and one of them is definite. We refer to it as a *linear response (LR) eigenvalue problem* for the reason to be explained later. It can be seen that the eigenvalue problem (1.1) is equivalent to any one of the following product eigenvalue problems

$$KM y = \lambda^2 y, \quad (1.2a)$$

$$MK x = \lambda^2 x. \quad (1.2b)$$

Theoretically, if any one of them is solved, the solutions to the other two can be constructed from the solved one with little effort.

The LR eigenvalue problem (1.1) arises from computing excitation states (energies) of physical systems in the study of collective motion of many particle systems, ranging from silicon nanoparticles and nanoscale materials to analysis of interstellar clouds (see for example [5, 11, 15]). In computational quantum chemistry and physics, the excitation states are described by the *random phase approximation* (RPA), a linear response perturbation analysis in

*Department of Computer Science and Department of Mathematics, University of California, Davis, CA 95616. (bai@cs.ucdavis.edu)

†Department of Mathematics, University of Texas at Arlington, P.O. Box 19408, Arlington, TX 76019 (rcli@uta.edu.)

the time-dependent Hartree-Fock and time-dependent density function theories. There are a great deal of recent work and interests in developing efficient numerical algorithms and simulation techniques for excitation response calculations of molecules for materials design in energy science [7, 19, 20, 12].

The heart of (nonrelativistic) RPA calculation is to compute a few smallest positive eigenvalues and their corresponding eigenvectors of the following eigenvalue problem

$$\begin{pmatrix} A & B \\ -B & -A \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} = \lambda \begin{pmatrix} u \\ v \end{pmatrix}, \quad (1.3)$$

where A and B are $n \times n$ real symmetric matrices such that the symmetric matrix $\begin{pmatrix} A & B \\ B & A \end{pmatrix}$ is positive definite¹ [18, 23]. In physics and chemistry literature, it is this eigenvalue problem that is referred to as *the LR eigenvalue problem* (see, e.g., [14]), or *the RPA eigenvalue problem* (see e.g., [6, 22, 11]). The eigenvalue problem (1.3) is also a special case of the *Hamiltonian eigenvalue problem* (see, e.g., [2, 3, 27]) because the matrix in (1.3) is a Hamiltonian matrix.

Define the symmetric orthogonal matrix

$$J = \frac{1}{\sqrt{2}} \begin{pmatrix} I_n & I_n \\ I_n & -I_n \end{pmatrix}, \quad (1.4)$$

where I_n is the $n \times n$ identity matrix. It can be verified that $J^T J = J^2 = I_{2n}$ and

$$J^T \begin{pmatrix} A & B \\ -B & -A \end{pmatrix} J = \begin{pmatrix} 0 & A - B \\ A + B & 0 \end{pmatrix}, \quad (1.5)$$

which is H in (1.1) with

$$K = A - B, \quad M = A + B. \quad (1.6)$$

Hence the Hamiltonian matrix in (1.3) and the matrix H in (1.1) with (1.6) are similar through J , making it equivalent to solve the eigenvalue problem for one by the one for the other. In fact, both have the same eigenvalues with corresponding eigenvectors related by

$$\begin{pmatrix} y \\ x \end{pmatrix} = J^T \begin{pmatrix} u \\ v \end{pmatrix}, \quad \begin{pmatrix} u \\ v \end{pmatrix} = J \begin{pmatrix} y \\ x \end{pmatrix}. \quad (1.7)$$

Furthermore, the positive definiteness of the matrix $\begin{pmatrix} A & B \\ B & A \end{pmatrix}$ is equivalent to that both of K and M are positive definite since

$$J^T \begin{pmatrix} A & B \\ B & A \end{pmatrix} J = \begin{pmatrix} A + B & 0 \\ 0 & A - B \end{pmatrix}. \quad (1.8)$$

By the equivalence of the eigenvalue problems (1.3) and (1.1), in this paper, we also refer to the eigenvalue problem (1.1) as the *linear response eigenvalue problem*.

When both K and M are symmetric positive definite, as in the case for RPA [17, 24, 26], it can be shown that the Hamiltonian matrix in (1.3) and thus the matrix H in (1.1) have only nonzero real eigenvalues and their nonzero eigenvalues come in $\pm\lambda$ pairs (see section 2). In this case, Thouless [24] showed that the smallest positive eigenvalue λ_{\min} admits the following minimization principle:

$$\lambda_{\min} = \min_{u,v} \varrho(u, v), \quad (1.9)$$

¹In this article we will focus on very much this case, except that the eigenvalue 0 is allowed, i.e., (1.3) has only real eigenvalues.

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

$$\varrho(u, v) = \frac{\begin{pmatrix} u \\ v \end{pmatrix}^T \begin{pmatrix} A & B \\ B & A \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix}}{|u^T u - v^T v|}, \quad (1.10)$$

and the minimization is taken among all vectors u, v such that $u^T u - v^T v \neq 0$. By the similarity transformation (1.5) and using the relationships in (1.7), we have

$$\varrho(u, v) \equiv \rho(x, y) \stackrel{\text{def}}{=} \frac{x^T K x + y^T M y}{2|x^T y|}, \quad (1.11)$$

and thus equivalently [26]

$$\lambda_{\min} = \min_{x, y \in \mathbb{D}} \rho(x, y), \quad (1.12)$$

where the domain \mathbb{D} consists of all x and y such that either $x^T y \neq 0$ or $x^T y = 0$ but $x^T K x + y^T M y > 0$. This removes those x and y that annihilate both the numerator and the denominator from the domain. In particular $x = y = 0$ is not in the domain \mathbb{D} .

We will refer to both $\varrho(u, v)$ and $\rho(x, y)$ as the *Thouless functionals* but in different forms. Although $\varrho(u, v) \equiv \rho(x, y)$ under (1.7), in this paper we primarily work with $\rho(x, y)$ to develop extensions of (1.12) and efficient numerical methods.

The theoretical part of our contributions in this paper are three-folds:

1. We extended the minimization principle (1.12) to include the case when one of K and M is singular and thus $\lambda_{\min} = 0$ for which “min” needs to be replaced by “inf”.
2. We introduced and proved a subspace version of the minimization principle (1.9):

$$\sum_{i=1}^k \lambda_i = \frac{1}{2} \inf_{\substack{U, V \in \mathbb{R}^{n \times k} \\ U^T V = I_k}} \text{trace}(U^T K U + V^T M V), \quad (1.13)$$

where λ_i ($1 \leq i \leq k$) are the first k smallest positive eigenvalues² of H , and $U, V \in \mathbb{R}^{n \times k}$. Moreover, “inf” can be replaced by “min” if both K and M are definite.

Equation (1.13) suggests that

$$\frac{1}{2} \text{trace}(U^T K U + V^T M V) \quad \text{subject to } U, V \in \mathbb{R}^{n \times k} \text{ and } U^T V = I_k \quad (1.14)$$

is a *proper subspace version* of the Thouless functional in the form of $\rho(\cdot, \cdot)$. By exploiting the close relation through (1.7) between ρ and ϱ , we also obtained a subspace version of the minimization principle (1.9) in Theorem 3.3 for the original LR eigenvalue problem (1.3) and, at the same time, a *proper subspace version* of the Thouless functional in the form of $\varrho(\cdot, \cdot)$.

3. We proved that the i th positive eigenvalue of a structure-preserving projection matrix H_{SR} of H onto a pair of subspaces is no smaller than the corresponding λ_i of H . In many ways, H_{SR} plays the same role for the LR eigenvalue problem (1.1) as the Rayleigh quotient matrix for the symmetric eigenvalue problem [16].

² H has an even number of eigenvalue 0, if any. For convenience, the plus sign is artificially assigned to half of the 0s and the negative sign to the other half. Although $+0 = -0$ in value, we regard $+0$ as positive. Doing so allows us to say that H has n positive eigenvalues and n negative eigenvalues without causing any ambiguity.

These theoretical contributions mirror the well-known results for the symmetric eigenvalue problem, namely the minimization principle of the Rayleigh quotient, the trace minimization, and Cauchy interlacing inequalities, see for examples [10, 16, 21].

This is the first paper of ours in a sequel of two. Here we focus on treating the theoretical aspect of the LR eigenvalue problem. The numerical aspect will be the subject of study in the second paper [1]. The rest of this paper is organized as follows. In section 2, we review basic theoretical results about the eigenvalue problem (1.1) and then introduce the concept of a pair of deflating subspaces and its approximation properties. In section 3, we extend the minimization principle (1.12) by Thouless and Tsiper to include several eigenvalues and present our Cauchy-like interlacing inequalities and more. We will also discuss the metric about the best approximation from a pair of approximate deflating subspaces. For simplicity of exposition, most proofs are deferred to appendix A. Concluding remarks are in section 4.

Throughout this paper, $\mathbb{R}^{n \times m}$ is the set of all $n \times m$ real matrices, $\mathbb{R}^n = \mathbb{R}^{n \times 1}$, and $\mathbb{R} = \mathbb{R}^1$. I_n (or simply I if its dimension is clear from the context) is the $n \times n$ identity matrix, and e_j is its j th column. The superscript “ \cdot^T ” takes transpose only. We shall also adopt MATLAB-like convention to access the entries of vectors and matrices. $i : j$ is the set of integers from i to j inclusive. For a vector u and an matrix X , $u_{(j)}$ is u 's j th entry, $X_{(i,j)}$ is X 's (i, j) th entry; X 's submatrices $X_{(k:\ell, i:j)}$, $X_{(k:\ell, :)}$, and $X_{(:, i:j)}$ consist of intersections of row k to row ℓ and column i to column j , row k to row ℓ , and column i to column j , respectively. If X is nonsingular, $\kappa(X) \stackrel{\text{def}}{=} \|X\|_2 \|X^{-1}\|_2$ is its spectral condition number, where $\|\cdot\|_2$ denotes the ℓ_2 -norm of a vector or the spectral norm of a matrix. For matrices or scalars X_i , both $\text{diag}(X_1, \dots, X_k)$ and $X_1 \oplus \dots \oplus X_k$ denote the same matrix

$$\begin{pmatrix} X_1 & & \\ & \ddots & \\ & & X_k \end{pmatrix}.$$

The assignments in (1.1) will be assumed, namely H is always defined that way for given $K, M \in \mathbb{R}^{n \times n}$ which are assumed by default to be symmetric positive semi-definite and one of which is definite, unless explicitly stated differently. This assumption is essential to our main contributions in this paper and its following one [1], although a few results do not require this. We will point them out along the way.

2 Basic theory and pair of deflating subspaces

2.1 Basic theory

In this subsection, we discuss some basic theoretical results on the eigenvalue problem (1.1). Most results are likely known, but cannot be found in one place. They are collected here for the convenience of our later developments. As discussed in section 1, the eigenvalue problem (1.1) of H and the eigenvalue problems (1.2) of KM and MK are all equivalent in the sense that if any one of them is solved, the solutions to the other two can be constructed from the solved one with little effort. Detail is given in Theorem 2.1 below.

Theorem 2.1. *1. If λ is an eigenvalue of H and (1.1) holds for $z \neq 0$, then equations in (1.2) hold, and λ^2 is an eigenvalue of KM if $y \neq 0$ and λ^2 is an eigenvalue of MK if $x \neq 0$. In particular if $\lambda \neq 0$, then both $x \neq 0$ and $y \neq 0$ and thus λ^2 is an eigenvalue of both KM and MK .*

2. Equation (1.2a) with $y \neq 0$ implies (1.1) for the nonzero z defined by

$$z = \begin{cases} \begin{pmatrix} y \\ \lambda^{-1}My \end{pmatrix}, & \text{if } \lambda \neq 0, \\ \begin{pmatrix} y \\ 0 \end{pmatrix}, & \text{if } \lambda = 0 \text{ and } My = 0, \\ \begin{pmatrix} 0 \\ My \end{pmatrix}, & \text{if } \lambda = 0 \text{ and } My \neq 0. \end{cases}$$

3. Equation (1.2b) with $x \neq 0$ implies (1.1) for the nonzero z defined by

$$z = \begin{cases} \begin{pmatrix} \lambda^{-1}Kx \\ x \end{pmatrix}, & \text{if } \lambda \neq 0, \\ \begin{pmatrix} 0 \\ x \end{pmatrix}, & \text{if } \lambda = 0 \text{ and } Kx = 0, \\ \begin{pmatrix} Kx \\ 0 \end{pmatrix}, & \text{if } \lambda = 0 \text{ and } Kx \neq 0. \end{cases}$$

4. Each nonzero $\mu = \lambda^2$ as an eigenvalue of KM (and MK) leads to two distinct eigenvalues of H and two corresponding eigenvectors z .

5. The number of zero eigenvalues of H is twice as many as the number of zero eigenvalues of KM (or MK).

Proof. We note that (1.1) is equivalent to

$$Kx = \lambda y, \quad My = \lambda x. \quad (2.1)$$

1. Only the part for $\lambda \neq 0$ needs a proof. For that, it suffices to show if $\lambda \neq 0$, then both $x \neq 0$ and $y \neq 0$. Suppose to the contrary that $\lambda \neq 0$ and $y = 0$. Then $x \neq 0$ since either x or y or both must not be zero. $Kx = 0$ by (2.1), and thus $\lambda^2 x = 0 \Rightarrow \lambda = 0$ or $x = 0$ by (1.2b), a contradiction. Similarly $\lambda \neq 0$ and $x = 0$ cannot happen either.

2. and 3. can be verified straightforwardly.

4. For each $\mu = \lambda^2 \neq 0$, there are two distinct square roots λ 's which yield two nonzero z 's in 2. and 3. as corresponding eigenvectors.

5. It is well-known that KM and MK have the same eigenvalues. Since H and H^2 have the same number of zero eigenvalues and

$$H^2 = \begin{pmatrix} KM & 0 \\ 0 & MK \end{pmatrix},$$

the conclusion follows. □

REMARK 2.1. Our implicit assumption that K and M are real and symmetric positive semi-definite and one of them is definite is not used in the proof. Thus Theorem 2.1 is actually valid for all square matrices K and M . ◇

Suppose that K and M are symmetric positive semi-definite. Since $KM = K^{1/2}K^{1/2}M$ has the same eigenvalues as $K^{1/2}MK^{1/2}$ which is also symmetric positive semi-definite, all eigenvalues of KM are real and nonnegative. Denote these eigenvalues by λ_i^2 ($1 \leq i \leq n$) in ascending order, i.e.,

$$0 \leq \lambda_1^2 \leq \lambda_2^2 \leq \cdots \leq \lambda_n^2, \quad (2.2)$$

where all $\lambda_i \geq 0$ and thus $0 \leq \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$. The eigenvalues of MK are λ_i^2 ($1 \leq i \leq n$), too. Theorem 2.1 implies the eigenvalues of H are

$$\pm\lambda_i \quad \text{for } i = 1, 2, \dots, n. \quad (2.3)$$

An immediate consequence of this is that the eigenvalues of H come in $\pm\lambda$ pairs. In particular, it has an even number of zero eigenvalues. As we mentioned in the footnote 2, for convenience, we shall associate half of 0 eigenvalues with the positive sign and the other half with the negative sign. Although $+0 = -0$ in value, we regard $+0$ as positive and doing so legitimizes the use of the phrase “the first k smallest positive eigenvalues of H ” to refer to λ_i for $1 \leq i \leq k$ without ambiguity even when $\lambda_1 = +0$. Throughout this paper, we will use λ_i^2 ($1 \leq i \leq n$) in ascending order as in (2.2) to denote the eigenvalues of KM .

Set

$$\mathcal{J} = \begin{pmatrix} 0 & I_n \\ I_n & 0 \end{pmatrix} \quad (2.4)$$

which is symmetric but indefinite. The matrix \mathcal{J} induces an *indefinite inner product* on \mathbb{R}^{2n} :

$$\langle z_1, z_2 \rangle_{\mathcal{J}} \stackrel{\text{def}}{=} z_1^{\text{T}} \mathcal{J} z_2.$$

The following theorem tells us some orthogonality properties among the eigenvectors of H . It does not require that one of K and M are definite.

Theorem 2.2. *Suppose K and M are symmetric and positive semi-definite.*

1. Let (λ, z) be an eigenpair of H , i.e., $H z = \lambda z$ and $z = \begin{pmatrix} y \\ x \end{pmatrix} \neq 0$, where $x, y \in \mathbb{R}^n$. Then $\lambda \langle z, z \rangle_{\mathcal{J}} = 2\lambda x^{\text{T}} y > 0$ if $\lambda \neq 0$. In particular, this implies $\langle z, z \rangle_{\mathcal{J}} = 2x^{\text{T}} y \neq 0$ if $\lambda \neq 0$.
2. Let (α_i, z_i) ($i = 1, 2$) be two eigenpairs of H . Partition $z_i = \begin{pmatrix} y_i \\ x_i \end{pmatrix} \neq 0$, where $x_i, y_i \in \mathbb{R}^n$.

(a) If $\alpha_1 \neq \alpha_2$, then $\langle z_1, z_2 \rangle_{\mathcal{J}} = y_1^{\text{T}} x_2 + x_1^{\text{T}} y_2 = 0$.

(b) If $\alpha_1 \neq \pm\alpha_2 \neq 0$, then $y_1^{\text{T}} x_2 = x_1^{\text{T}} y_2 = 0$.

Proof. 1. $H z = \lambda z$ gives $Kx = \lambda y$ and $My = \lambda x$. We then have $x^{\text{T}} K x = \lambda x^{\text{T}} y$ and $y^{\text{T}} M y = \lambda y^{\text{T}} x$. So $\lambda x^{\text{T}} y \geq 0$. It suffices to show $x^{\text{T}} y \neq 0$ if $\lambda \neq 0$. Suppose to the contrary that $\lambda \neq 0$ but $x^{\text{T}} y = 0$. Then

$$x^{\text{T}} y = 0 \Rightarrow x^{\text{T}} K x = y^{\text{T}} M y = 0 \Rightarrow K x = M y = 0 \Rightarrow \lambda y = \lambda x = 0 \Rightarrow x = y = 0,$$

contradicting $z \neq 0$.

2. Since

$$\mathcal{J} H = \begin{pmatrix} M & 0 \\ 0 & K \end{pmatrix} = (\mathcal{J} H)^{\text{T}}$$

is symmetric and positive semi-definite, we have

$$\alpha_2 z_1^{\text{T}} \mathcal{J} z_2 = z_1^{\text{T}} \mathcal{J} (H z_2) = z_1^{\text{T}} (\mathcal{J} H) z_2 = z_1^{\text{T}} (\mathcal{J} H)^{\text{T}} z_2 = (H z_1)^{\text{T}} \mathcal{J} z_2 = \alpha_1 z_1^{\text{T}} \mathcal{J} z_2,$$

and therefore $(\alpha_2 - \alpha_1)z_1^T \mathcal{J} z_2 = 0$, which implies that $z_1^T \mathcal{J} z_2 = 0$ if $\alpha_1 \neq \alpha_2$. Suppose that $\alpha_1 \neq \pm\alpha_2$. Note that $-\alpha_2$ is an eigenvalue of H , too, with eigenvector $\begin{pmatrix} -y_2 \\ x_2 \end{pmatrix}$. By what we just proved, we have

$$y_1^T x_2 + x_1^T y_2 = 0, \quad y_1^T x_2 + x_1^T (-y_2) = 0$$

which yield $y_1^T x_2 = x_1^T y_2 = 0$. □

More can be said when one of K and M is definite. For the sake of presentation, we shall always either assume that M is definite or only provide proofs for definite M whenever one of K and M is required to be definite. Doing so loses no generality because the interchangeable roles played by K and M makes it rather straightforward to create a version for the case when K is definite by simply swapping K and M in each of their appearances. The following theorem is critical to our theoretical developments.

Theorem 2.3. *Suppose that M is definite. Then the following statements are true:*

1. *There exists a nonsingular $Y \in \mathbb{R}^{n \times n}$ such that*

$$K = Y\Lambda^2 Y^T, \quad M = XX^T, \tag{2.5}$$

where $\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$ and $X = Y^{-T}$.

2. *If K is also definite, then all $\lambda_i > 0$ and H is diagonalizable:*

$$H \begin{pmatrix} Y\Lambda & Y\Lambda \\ X & -X \end{pmatrix} = \begin{pmatrix} Y\Lambda & Y\Lambda \\ X & -X \end{pmatrix} \begin{pmatrix} \Lambda & \\ & -\Lambda \end{pmatrix}. \tag{2.6}$$

3. *H is not diagonalizable if and only if $\lambda_1 = 0$ which happens when and only when K is singular.*

4. *The i th column of $Z = \begin{pmatrix} Y\Lambda \\ X \end{pmatrix}$ are the eigenvector corresponding to λ_i and it is unique if*

(a) λ_i is a simple eigenvalue of H , or

(b) $i = 1$, $\lambda_1 = +0 < \lambda_2$. In this case, 0 is a double eigenvalue of H but there is only one eigenvector associated with it.

5. *If $0 = \lambda_1 = \dots = \lambda_\ell < \lambda_{\ell+1}$, then H 's Jordan canonical form is*

$$\underbrace{\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \oplus \dots \oplus \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}}_{\ell} \oplus \text{diag}(\lambda_{\ell+1}, -\lambda_{\ell+1}, \dots, \lambda_n, -\lambda_n). \tag{2.7}$$

Thus H has 0 as an eigenvalue of algebraic multiplicity 2ℓ with only ℓ linear independent eigenvectors which are the columns of $\begin{pmatrix} 0 \\ X_{(:,1:\ell)} \end{pmatrix}$.

6. *The eigen-decompositions of KM and MK are*

$$(KM)Y = Y\Lambda^2, \quad (MK)X = X\Lambda^2, \tag{2.8}$$

respectively.

7. *The eigen-decomposition for the matrix pencil $MKM - \lambda M$ is*

$$MKM = X\Lambda^2 X^T, \quad M = XX^T. \tag{2.9}$$

Proof. 1. Since M^{-1} is symmetric and positive definite, it has a Cholesky decomposition $M^{-1} = R^T R$, where $R \in \mathbb{R}^{n \times n}$ is nonsingular. That $R^{-T} K R^{-1}$ is symmetric and positive semi-definite implies that it has the eigen-decomposition $R^{-T} K R^{-1} = Q \Lambda^2 Q^T$, where $Q \in \mathbb{R}^{n \times n}$ is orthogonal. Now take $Y = R^T Q$ to give (2.5).

2. Since $X = Y^{-T}$, we have from (2.5)

$$KX = Y\Lambda^2, \quad MY = X. \quad (2.10)$$

Rewrite (2.10) as

$$KX = (Y\Lambda)\Lambda, \quad M(Y\Lambda) = X\Lambda, \quad (2.11)$$

which gives (2.6). It can be verified that $\begin{pmatrix} Y\Lambda & Y\Lambda \\ X & -X \end{pmatrix}$ is nonsingular if all $\lambda_i > 0$. So (2.6) implies that H is diagonalizable when all $\lambda_i > 0$.

For items 3. to 5. we deduce from (2.10)

$$H \begin{pmatrix} Y & \\ & X \end{pmatrix} = \begin{pmatrix} Y & \\ & X \end{pmatrix} \begin{pmatrix} & \Lambda^2 \\ I_n & \end{pmatrix}. \quad (2.12)$$

Since $\begin{pmatrix} Y & \\ & X \end{pmatrix}$ is always nonsingular, it suffices to investigate the same issues for $\begin{pmatrix} & \Lambda^2 \\ I_n & \end{pmatrix}$.

With the so-called perfect shuffle, we have

$$P^T \begin{pmatrix} & \Lambda^2 \\ I_n & \end{pmatrix} P = \begin{pmatrix} & \lambda_1^2 \\ 1 & \end{pmatrix} \oplus \begin{pmatrix} & \lambda_2^2 \\ 1 & \end{pmatrix} \oplus \cdots \oplus \begin{pmatrix} & \lambda_n^2 \\ 1 & \end{pmatrix},$$

where $P = (e_1, e_{n+1}, e_2, e_{n+2}, \dots, e_n, e_{2n})$. The eigenvalues of each 2-by-2 matrix

$$\begin{pmatrix} & \lambda_i^2 \\ 1 & \end{pmatrix}$$

are $\pm\lambda_i$, and such a matrix is diagonalizable if and only if $\lambda_i \neq 0$:

$$\begin{pmatrix} & \lambda_i^2 \\ 1 & \end{pmatrix} \begin{pmatrix} \lambda_i & \lambda_i \\ 1 & -1 \end{pmatrix} = \begin{pmatrix} \lambda_i & \lambda_i \\ 1 & -1 \end{pmatrix} \begin{pmatrix} \lambda_i & \\ & -\lambda_i \end{pmatrix}.$$

Therefore H is not diagonalizable if and only if $\lambda_1 = 0$.

The verifications of items 6. and 7. are rather straightforward. \square

We note that Jordan canonical form of the matrix H in the previous theorem is a special case of the canonical forms of doubly structured matrices in the work of Mehl, Mehrmann and Xu [13].

2.2 Pair of deflating subspaces

Let $\mathcal{U}, \mathcal{V} \subseteq \mathbb{R}^n$ be subspaces. We call $\{\mathcal{U}, \mathcal{V}\}$ a *pair of deflating subspaces* of $\{K, M\}$ if

$$K\mathcal{U} \subseteq \mathcal{V} \quad \text{and} \quad M\mathcal{V} \subseteq \mathcal{U}. \quad (2.13)$$

Let $U \in \mathbb{R}^{n \times k}$ and $V \in \mathbb{R}^{n \times \ell}$ be the basis matrices for the subspaces \mathcal{U} and \mathcal{V} , respectively, where $\dim(\mathcal{U}) = k$ and $\dim(\mathcal{V}) = \ell$. Then (2.13) implies that there exist $K_R \in \mathbb{R}^{\ell \times k}$ and $M_R \in \mathbb{R}^{k \times \ell}$ such that

$$KU = VK_R, \quad MV = UM_R. \quad (2.14)$$

Given U and V , both K_R and M_R are uniquely determined by respective equations in (2.14), but there are numerous way to express them. In fact for any left generalized inverses U^{-1} and V^{-1} of U and V , respectively, i.e., $U^{-1}U = I_k$ and $V^{-1}V = I_\ell$,

$$K_R = V^{-1}KU, \quad M_R = U^{-1}MV. \quad (2.15)$$

There are infinitely many left generalized inverses U^{-1} and V^{-1} . For example, two of them for U are

$$U^{-1} = (U^T U)^{-1} U^T$$

or

$$U^{-1} = (V^T U)^{-1} V^T \quad \text{if } (V^T U)^{-1} \text{ exists.} \quad (2.16)$$

But still K_R and M_R are unique. The second left generalized inverse (2.16) will become important later in preserving symmetry in K and M .

Define

$$H_R = \begin{pmatrix} 0 & K_R \\ M_R & 0 \end{pmatrix}. \quad (2.17)$$

Then H_R is the restriction of H onto $\mathcal{V} \oplus \mathcal{U}$ with respect to the basis matrix $V \oplus U$:

$$\begin{pmatrix} 0 & K \\ M & 0 \end{pmatrix} \begin{pmatrix} V \\ U \end{pmatrix} = \begin{pmatrix} V \\ U \end{pmatrix} \begin{pmatrix} 0 & K_R \\ M_R & 0 \end{pmatrix}. \quad (2.18)$$

This also says that $\mathcal{V} \oplus \mathcal{U}$ is an invariant subspace of H . On the other hand, every invariant subspace of H yields a pair of deflating subspaces of $\{K, M\}$ as well.

Theorem 2.4. *1. If $\{\mathcal{U}, \mathcal{V}\}$ is a pair of deflating subspaces of $\{K, M\}$, then $\mathcal{V} \oplus \mathcal{U}$ is an invariant subspace of H .*

2. Let \mathcal{Z} be invariant subspace of H and let $Z = \begin{pmatrix} V \\ U \end{pmatrix}$ be a basis matrix of \mathcal{Z} , where $V \in \mathbb{R}^{n \times \ell}$. Then $\{\text{span}(U), \text{span}(V)\}$ is a pair of deflating subspaces of $\{K, M\}$.

Proof. 1. That $\mathcal{V} \oplus \mathcal{U}$ is an invariant subspace of H is a consequence of (2.18).

2. There is a matrix D such that $HZ = ZD$ which leads to $KU = VD$ and $MV = UD$. Thus (2.13) holds for $\mathcal{U} = \text{span}(U)$ and $\mathcal{V} = \text{span}(V)$. \square

The following theorem says a subset of eigenvalues and eigenvectors of H can be recovered from those of H_R .

Theorem 2.5. *Let $\{\mathcal{U}, \mathcal{V}\}$ be a pair of deflating subspaces of $\{K, M\}$ and $U \in \mathbb{R}^{n \times k}$ and $V \in \mathbb{R}^{n \times \ell}$ be the basis matrices for the subspaces \mathcal{U} and \mathcal{V} , respectively. Define K_R , M_R , and H_R by equations in (2.14) and (2.17). Then*

$$H_R \hat{z} \equiv \begin{pmatrix} 0 & K_R \\ M_R & 0 \end{pmatrix} \begin{pmatrix} \hat{y} \\ \hat{x} \end{pmatrix} = \lambda \begin{pmatrix} \hat{y} \\ \hat{x} \end{pmatrix} \equiv \lambda \hat{z}$$

implies (1.1) with $x = U\hat{x}$ and $y = V\hat{y}$, where $\hat{z} = \begin{pmatrix} \hat{y} \\ \hat{x} \end{pmatrix}$ conformably partitioned.

Proof. $H_R \hat{z} = \lambda \hat{z}$ yields $K_R \hat{x} = \lambda \hat{y}$ and $M_R \hat{y} = \lambda \hat{x}$. Therefore $KU\hat{x} = VK_R \hat{x} = \lambda V\hat{y}$ and $MV\hat{y} = UM_R \hat{y} = \lambda U\hat{x}$, as was to be shown. \square

H_R in (2.17) inherits the block structure in H in (1.1): zero blocks remain zero blocks. But when K and M are symmetric, as in the RPA case, in general H_R may lose the symmetry property in its off-diagonal blocks K_R and M_R , not to mention positive semi-definiteness of K and M . We propose a modification to H_R to overcome this potential loss, when $W \stackrel{\text{def}}{=} U^T V$ is nonsingular. Factorize $W = W_1^T W_2$, where W_1 and W_2 are nonsingular, and define

$$H_{\text{SR}} = \begin{pmatrix} 0 & W_1^{-T} U^T K U W_1^{-1} \\ W_2^{-T} V^T M V W_2^{-1} & 0 \end{pmatrix}. \quad (2.19)$$

Note H_{SR} shares not only the block structure in H but also the symmetry and semi-definiteness in its off-diagonal blocks. In defining H_{SR} here, it is assumed that $U^T V$ is nonsingular. For this, we have the following lemma.

Lemma 2.1. *Suppose that one of K and M is definite. Let $\{\mathcal{U}, \mathcal{V}\}$ be a pair of deflating subspaces of $\{K, M\}$ with $\dim(\mathcal{U}) = \dim(\mathcal{V}) = k$, and let $U \in \mathbb{R}^{n \times k}$ and $V \in \mathbb{R}^{n \times k}$ be the basis matrices of the subspaces \mathcal{U} and \mathcal{V} , respectively. Then $U^T V$ is nonsingular.*

Proof. Equations in (2.14) hold for some K_R and M_R . Thus

$$U^T K U = U^T V K_R, \quad V^T M V = V^T U M_R.$$

Suppose that M is definite. Then $V^T M V$ is definite and thus nonsingular; so $V^T U$ is nonsingular from the second equation. \square

Theorem 2.6. *Let $\{\mathcal{U}, \mathcal{V}\}$ be a pair of deflating subspaces of $\{K, M\}$ and $U \in \mathbb{R}^{n \times k}$ and $V \in \mathbb{R}^{n \times \ell}$ be the basis matrices for the subspaces \mathcal{U} and \mathcal{V} , respectively. Suppose that $W \stackrel{\text{def}}{=} U^T V$ is nonsingular and is factorized as $W = W_1^T W_2$ with both W_1 and W_2 being nonsingular, and define H_{SR} by (2.19). Then H_{SR} is the restriction of H onto $\mathcal{V} \oplus \mathcal{U}$ with respect to the basis matrix $VW_2^{-1} \oplus UW_1^{-1}$:*

$$H \begin{pmatrix} VW_2^{-1} & \\ & UW_1^{-1} \end{pmatrix} = \begin{pmatrix} VW_2^{-1} & \\ & UW_1^{-1} \end{pmatrix} H_{\text{SR}}. \quad (2.20)$$

Consequently, $H_{\text{SR}} \hat{z} = \lambda \hat{z}$ implies (1.1) with $x = UW_1^{-1} \hat{x}$ and $y = VW_2^{-1} \hat{y}$, where $\hat{z} = \begin{pmatrix} \hat{y} \\ \hat{x} \end{pmatrix}$ conformably partitioned.

Proof. Equations in (2.14) hold for some K_R and M_R . Thus

$$U^T K U = (U^T V) K_R = W_1^T W_2 K_R \quad \text{and} \quad V^T M V = (V^T U) M_R = W_2^T W_1 M_R,$$

which gives

$$W_1^{-T} U^T K U W_1^{-1} = W_2 K_R W_1^{-1} \quad \text{and} \quad W_2^{-T} V^T M V W_2^{-1} = W_1 M_R W_2^{-1}. \quad (2.21)$$

Now by (2.14) and (2.21), we have the identity (2.20) since

$$\begin{aligned} K(UW_1^{-1}) &= V K_R W_1^{-1} = (VW_2^{-1})(W_2 K_R W_1^{-1}) \\ &= (VW_2^{-1})(W_1^{-T} U^T K U W_1^{-1}). \end{aligned}$$

and

$$\begin{aligned} M(VW_2^{-1}) &= U M_R W_2^{-1} = UW_1^{-1}(W_1 M_R W_2^{-1}) \\ &= (UW_1^{-1})(W_2^{-T} V^T M V W_2^{-1}). \end{aligned}$$

Finally, apply Theorem 2.5 to conclude the proof. \square

Equations in (2.21) imply that *under the conditions of Theorem 2.6, H_{R} and H_{SR} are similar:*

$$H_{\text{SR}} = \begin{pmatrix} 0 & W_2 K_{\text{R}} W_1^{-1} \\ W_1 M_{\text{R}} W_2^{-1} & 0 \end{pmatrix} = \begin{pmatrix} W_2 & 0 \\ 0 & W_1 \end{pmatrix} H_{\text{R}} \begin{pmatrix} W_2 & 0 \\ 0 & W_1 \end{pmatrix}^{-1}, \quad (2.22)$$

which is not at all obvious from (2.17) and (2.19).

A trivial pair of deflating subspaces $\{\mathcal{U}, \mathcal{V}\}$ is when $\mathcal{U} = \mathcal{V} = \mathbb{R}^n$. In particular, for $U, V \in \mathbb{R}^{n \times n}$ satisfying $U^{\text{T}}V = I_n$, matrices

$$H = \begin{pmatrix} 0 & K \\ M & 0 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 0 & U^{\text{T}}KU \\ V^{\text{T}}MV & 0 \end{pmatrix} \quad (2.23)$$

have the same eigenvalues. In fact, the two matrices in (2.23) are similar because of (2.18) and for the current case

$$\begin{pmatrix} V & \\ & U \end{pmatrix}^{-1} = \begin{pmatrix} U^{\text{T}} & \\ & V^{\text{T}} \end{pmatrix}.$$

REMARK 2.2. For this subsection, our default assumption on K and M is not required, except for Lemma 2.1. \diamond

2.3 Invariance properties of H_{SR}

In the previous subsection, H_{SR} was introduced as a structure-preserving projection of H onto a pair of deflating subspaces $\{\mathcal{U}, \mathcal{V}\}$. But its definition in (2.19) does not require $\{\mathcal{U}, \mathcal{V}\}$ being a pair of deflating subspaces. In fact, it is well-defined so long as $U^{\text{T}}V$ is nonsingular, where $U, V \in \mathbb{R}^{n \times k}$ are the basis matrices of $\mathcal{U}, \mathcal{V} \subset \mathbb{R}^n$, respectively. This observation will become critically important in the second part of this paper [1] where H_{SR} is often defined for a pair of *approximate* deflating subspaces and will play the same role in the LR eigenvalue computation as the Rayleigh quotient matrix does for the symmetric eigenvalue problem.

As we just pointed out, we need the non-singularity assumption on $U^{\text{T}}V$ to define H_{SR} . We note that this assumption is independent of the freedom in choosing basis matrices. Now we present a necessary and sufficient condition in terms of canonical angles between subspaces for this assumption. Recall that the canonical angles between \mathcal{U} and \mathcal{V} are defined to be

$$\arccos \sigma_i, \quad i = 1, 2, \dots, k,$$

where σ_i ($1 \leq i \leq k$) are the singular values of $(U^{\text{T}}U)^{-1/2}U^{\text{T}}V(V^{\text{T}}V)^{-1/2}$ [21]. Furthermore, the angle $\angle(\mathcal{U}, \mathcal{V})$ between \mathcal{U} and \mathcal{V} is defined to be

$$\angle(\mathcal{U}, \mathcal{V}) = \max_i \arccos(\sigma_i) = \arccos(\min_i \sigma_i).$$

Note the canonical angles $\arccos \sigma_i$ and the angle $\angle(\mathcal{U}, \mathcal{V})$ are independent of the choices of basis matrices [21].

Lemma 2.2. *Let $U, V \in \mathbb{R}^{n \times k}$ be basis matrices of $\mathcal{U} \subset \mathbb{R}^n$ and $\mathcal{V} \subset \mathbb{R}^n$, respectively.*

1. $U^{\text{T}}V$ is nonsingular if and only if $\angle(\mathcal{U}, \mathcal{V}) < \pi/2$.
2. If $\angle(\mathcal{U}, \mathcal{V}) < \pi/2$, then $\mathbb{R}^n = \mathcal{U} \oplus \mathcal{V}^{\perp} = \mathcal{V} \oplus \mathcal{U}^{\perp}$, where \mathcal{U}^{\perp} and \mathcal{V}^{\perp} are the orthogonal complements of \mathcal{U} and \mathcal{V} , respectively.

Proof. Use the notations in the definition of $\angle(\mathcal{U}, \mathcal{V})$ above. $U^{\text{T}}V$ is nonsingular if and only if all $1 \geq \sigma_i > 0$ which is equivalent to all $\arccos(\sigma_i) < \pi/2$. This proves item 1.

Suppose $\angle(\mathcal{U}, \mathcal{V}) < \pi/2$ and thus $U^T V$ is nonsingular. Any $x \in \mathbb{R}^n$ can be written as $x = Px + (I - P)x$, where

$$P = U(V^T U)^{-1} V^T. \quad (2.24)$$

Evidently $Px \in \mathcal{U}$. It can be verified that $V^T(I - P) = 0$ which implies $(I - P)x \in \mathcal{V}^\perp$. Hence $\mathbb{R}^n = \mathcal{U} + \mathcal{V}^\perp$. Furthermore, if $x \in \mathcal{U}$ and $x \in \mathcal{V}^\perp$, then

$$x = U\hat{x}, \quad 0 = V^T x = V^T U\hat{x}$$

which implies $\hat{x} = 0$ and so must $x = 0$ because $V^T U$ is nonsingular. This proves $\mathbb{R}^n = \mathcal{U} \oplus \mathcal{V}^\perp$. Similarly $\mathbb{R}^n = \mathcal{V} \oplus \mathcal{U}^\perp$. \square

Each H_{SR} always corresponds uniquely to two subspaces:

$$\mathcal{U} = \text{span}(U), \quad \mathcal{V} = \text{span}(V)$$

that satisfy $\angle(\mathcal{U}, \mathcal{V}) < \pi/2$. On the other hand, two subspaces \mathcal{U} and \mathcal{V} satisfying $\angle(\mathcal{U}, \mathcal{V}) < \pi/2$ lead to (infinitely) many H_{SR} , due to the following two non-unique choices:

$$\left\{ \begin{array}{l} 1. \text{ Factorization } W = W_1^T W_2 \text{ is not unique.} \\ 2. \text{ Basis matrices } U \text{ and } V \text{ are not unique.} \end{array} \right. \quad (2.25)$$

In the next theorem, we present two invariance properties of H_{SR} with respect to these two non-unique choices. The properties are important in speaking about eigenvalue and eigenvector approximations from a pair of approximate deflating subspaces in [1].

Theorem 2.7. *Let $\mathcal{U}, \mathcal{V} \subset \mathbb{R}^n$ be two subspaces of dimension k such that $\angle(\mathcal{U}, \mathcal{V}) < \pi/2$. We have the following invariance properties of H_{SR} .*

1. *The eigenvalues of H_{SR} defined by (2.19) are invariant with respect to any of the non-uniqueness listed in (2.25);*
2. *For any invariant subspace \mathcal{E} of H_{SR} , the subspace*

$$\left\{ \left(\begin{array}{c} V W_2^{-1} \\ U W_1^{-1} \end{array} \right) \hat{z} : \hat{z} \in \mathcal{E} \right\} \quad (2.26)$$

is invariant with respect to any of the non-uniqueness listed in (2.25). By which we mean for any two realizations $H_{\text{SR}}^{(0)}$ and $H_{\text{SR}}^{(1)}$ of H_{SR} and the subspace (2.26) obtained from an invariant subspace \mathcal{E}_0 of $H_{\text{SR}}^{(0)}$, there exists an invariant subspace \mathcal{E}_1 of $H_{\text{SR}}^{(1)}$ which produces the same subspace (2.26). In particular, if \mathcal{E} has dimension 1, this gives an invariance property on the eigenvectors of H_{SR} .

Proof. We first show the invariant properties with respect to different factorizations $W = W_1^T W_2$. To this end, we note that $H_1 \stackrel{\text{def}}{=} H_{\text{SR}}$ with $W = W_1^T W_2$ and $H_0 \stackrel{\text{def}}{=} H_{\text{SR}}$ with $W = I_k^T \cdot W$ are similar:

$$\left(\begin{array}{c} W_1^{-T} \\ W_1 \end{array} \right)^{-1} H_1 \left(\begin{array}{c} W_1^{-T} \\ W_1 \end{array} \right) = \left(\begin{array}{c} W_1^T \\ W_1^{-1} \end{array} \right) H_1 \left(\begin{array}{c} W_1^{-T} \\ W_1 \end{array} \right) = H_0.$$

Next we verify the invariant properties with respect to different choices of basis matrices. To this end, it suffices to verify the invariant properties under the following substitutions:

$$UR \leftarrow U, \quad VS \leftarrow V, \quad W_1 R \leftarrow W_1, \quad W_2 S \leftarrow W_2, \quad (2.27)$$

where $R, S \in \mathbb{R}^{k \times k}$ are nonsingular because we have just proved the properties with respect to different decompositions of W . The verification is straightforward because H_{SR} and

$$\begin{pmatrix} VW_2^{-1} & \\ & UW_1^{-1} \end{pmatrix}$$

do not change under the substitutions (2.27). \square

REMARK 2.3. For this subsection, our default assumption on K and M is not required. \diamond

3 Minimization principles and Cauchy-like interlacing inequalities

We recall three well-known results for a symmetric matrix $A \in \mathbb{R}^{n \times n}$. Denote by θ_i ($1 \leq i \leq n$) A 's eigenvalues in ascending order. The first well-known result is the following minimization principle for A 's smallest eigenvalue θ_1 :

$$\theta_1 = \min_{x \neq 0} \frac{x^T A x}{x^T x}. \quad (3.1)$$

The trace (or subspace) version of (3.1), the second well-known result, is

$$\sum_{i=1}^k \theta_i = \min_{U \in \mathbb{R}^{n \times k}, U^T U = I_k} \text{trace}(U^T A U). \quad (3.2)$$

Furthermore, given any $U \in \mathbb{R}^{n \times k}$ such that $U^T U = I_k$, denote by μ_i ($1 \leq i \leq k$) the eigenvalues of the projection matrix $U^T A U$ in ascending order. We have Cauchy interlacing inequalities – the third well-known result:

$$\theta_i \leq \mu_i \leq \theta_{i+n-k} \quad \text{for } 1 \leq k. \quad (3.3)$$

The proofs of these well-known theoretical results can be found, for example, in [4, 16, 21]. They are crucial to the establishment of efficient numerical methods for the symmetric eigenvalue problem, and largely responsible for why the symmetric eigenvalue problems are regarded as *nice* eigenvalue problems in a wide range of applications.

In this section, we establish analogs of these results mainly for the LR eigenvalue problem (1.1).

3.1 Minimization principles

Theorem 3.1 is an analog of the minimization principle (3.1) for the symmetric matrix A . It is essentially (1.12) due to Tsiper [25, 26] who deduced it from (1.9) due to Thouless [24], except we allow one of K and M to be singular. We note that Theorem 3.2 presents a subspace version of Theorem 3.1. Although Theorem 3.1 is a corollary of Theorem 3.2, we decide to give a short proof anyway because the proof of Theorem 3.2 is long and is deferred to appendix A.

Theorem 3.1. *Suppose that one of K and M is definite. Then we have*

$$\lambda_1 = \inf_{x, y \in \mathbb{D}} \rho(x, y). \quad (3.4)$$

Moreover, “inf” can be replaced by “min” if and only if both K and M are definite. When they are definite, the optimal argument pair (x, y) gives rise to an eigenvector $z = \begin{pmatrix} y \\ x \end{pmatrix}$ of H associated with λ_1 .

Proof. Note $\rho(x, y) \geq 0$ for any x and y . If K is singular, then $\lambda_1 = 0$. Pick $x \neq 0$ such that $Kx = 0$. Then $x^T Mx > 0$ since one of K and M is assumed definite. We have

$$\rho(x, \epsilon x) = |\epsilon| x^T Mx / (2|x^T x|) \rightarrow 0 \quad \text{as } \epsilon \rightarrow 0.$$

This is (3.4) for the case. We now show that ‘‘inf’’ cannot be replaced by ‘‘min’’. Suppose there were x and y such that $x^T y \neq 0$ and $\rho(x, y) = 0$. We note that $\rho(x, y) = 0$ and $x^T y \neq 0$ imply $x^T Kx = y^T My = 0$ which in turn implies $Kx = My = 0$, contradicting that one of K and M is definite.

Suppose K and M are definite. Then $\lambda_1 > 0$ and Equations in (2.5) hold for some nonsingular $Y \in \mathbb{R}^{n \times n}$ and $X = Y^{-T}$. We have

$$\begin{aligned} \min_{x,y} \frac{x^T Kx + y^T My}{2|x^T y|} &= \min_{x,y} \frac{x^T Y \Lambda^2 Y^T x + y^T Y^{-T} Y^{-1} y}{2|x^T Y Y^{-1} y|} \\ &= \min_{\tilde{x}, \tilde{y}} \frac{\tilde{x}^T \Lambda^2 \tilde{x} + \tilde{y}^T \tilde{y}}{2|\tilde{x}^T \tilde{y}|} \\ &\geq \min_{\tilde{x}, \tilde{y}} \frac{2 \sum_i \lambda_i |\tilde{x}_{(i)} \tilde{y}_{(i)}|}{2|\sum_i \tilde{x}_{(i)} \tilde{y}_{(i)}|} \end{aligned} \quad (3.5)$$

$$\geq \lambda_1, \quad (3.6)$$

where $\tilde{x} = Y^T x$ and $\tilde{y} = Y^{-1} y$. Suppose $0 < \lambda_1 = \dots = \lambda_\ell < \lambda_{\ell+1} \leq \dots \leq \lambda_n$. Both equality signs in (3.5) and (3.6) hold if and only if

$$\begin{aligned} \tilde{x}_{(i)} \lambda_i &= \tilde{y}_{(i)} & \text{for } 1 \leq i \leq \ell, \\ \tilde{x}_{(i)} &= \tilde{y}_{(i)} = 0 & \text{for } \ell < i \leq n, \end{aligned}$$

i.e., $\tilde{y} = \Lambda \tilde{x}$ and $\tilde{x}_{(\ell+1:n)} = \tilde{y}_{(\ell+1:n)} = 0$. So for their corresponding optimal argument pair (x, y) ,

$$Kx = KY^{-T} \tilde{x} = KX \tilde{x} = Y \Lambda^2 \tilde{x} = Y \Lambda \tilde{y} = \lambda_1 Y \tilde{y} = \lambda_1 y,$$

and similarly $My = \lambda_1 x$. □

REMARK 3.1. Equation (3.4) is actually true even both K and M are singular (but still positive semi-definite, of course). They are two cases.

1. Both K and M are singular and their kernels are not orthogonal to each other, i.e., there are nonzero vectors x and y such that $Kx = My = 0$ and $x^T y \neq 0$. For such a case, we have

$$\lambda_1 = \min_{x,y} \rho(x, y). \quad (3.7)$$

2. Both K and M are singular but their kernels are orthogonal to each other. For such a case, we have (3.4) but ‘‘inf’’ cannot be replaced by ‘‘min’’. Here is why. Since K is singular, we pick $x \neq 0$ such that $Kx = 0$. Then $Mx \neq 0$ because the kernels of K and M are orthogonal to each other. So $x^T M = (Mx)^T \neq 0$ which says at least one of the columns of M is not orthogonal to x , and take y to be one of such a column. Now we see

$$\rho(x, \epsilon y) = |\epsilon| y^T My / (2|x^T y|) \rightarrow 0 \quad \text{as } \epsilon \rightarrow 0.$$

This gives (3.4) since $\rho(\cdot, \cdot) \geq 0$ always. To see ‘‘inf’’ cannot be replaced by ‘‘min’’, we assume there were x and y such that $x^T y \neq 0$ and $\rho(x, y) = 0$. We note that $\rho(x, y) = 0$ and $x^T y \neq 0$ imply $x^T Kx = y^T My = 0$ which in turn implies $Kx = My = 0$, contradicting the assumption that the kernels of K and M are orthogonal to each other. ◇

Our next theorem – Theorem 3.2 – presents a subspace version of Theorem 3.1. It is the reason we mentioned in section 1 that the expression in (1.14) can be regarded as a *proper subspace version* of the Thouless functional in the form of $\rho(\cdot, \cdot)$.

Theorem 3.2. *Suppose that one of K and M is definite. Then we have*

$$\sum_{i=1}^k \lambda_i = \frac{1}{2} \inf_{\substack{U, V \in \mathbb{R}^{n \times k} \\ U^T V = I_k}} \text{trace}(U^T K U + V^T M V). \quad (3.8)$$

Moreover, “inf” can be replaced by “min” if and only if both K and M are definite. When they are definite and if also $\lambda_k < \lambda_{k+1}$, then for any U and V that attain the minimum, $\{\text{span}(U), \text{span}(V)\}$ is a pair of deflating subspaces of $\{K, M\}$ and the corresponding H_{SR} (and H_{R} , too) has eigenvalues $\pm \lambda_i$ ($1 \leq i \leq k$).

Proof. The proof is long and deferred to appendix A. □

Corollary 3.1. *Suppose that one of K and M is definite. Then*

$$\sum_{i=1}^n \lambda_i = \frac{1}{2} \inf_{\substack{U, V \in \mathbb{R}^{n \times n} \\ U^T V = I_n}} \text{trace}(U^T K U + V^T M V). \quad (3.9)$$

REMARK 3.2. In (3.2) which is for the symmetric eigenvalue problem of A , if $k = n$, then

$$\sum_{i=1}^n \theta_i = \text{trace}(U^T A U), \quad (3.10)$$

regardless of $U \in \mathbb{R}^{n \times n}$ so long as $U^T U = I_n$. There is certainly a strong resemblance between (3.9) and (3.10), but a fundamental difference, too. That is that “inf” has to be there in (3.9). Without “inf”, (3.9) becomes

$$\sum_{i=1}^n \lambda_i \leq \frac{1}{2} \text{trace}(U^T K U + V^T M V) \quad (3.11)$$

for any two $U, V \in \mathbb{R}^{n \times n}$ satisfying $U^T V = I_n$. ◇

Exploiting the close relation through (1.7) between the two different forms of the Thouless functionals $\varrho(\cdot, \cdot)$ and $\rho(\cdot, \cdot)$, we have by Theorem 3.2 the following theorem. It suggests that

$$\frac{1}{2} \text{trace} \left(\begin{pmatrix} U \\ V \end{pmatrix}^T \begin{pmatrix} A & B \\ B & A \end{pmatrix} \begin{pmatrix} U \\ V \end{pmatrix} \right)$$

subject to

$$U, V \in \mathbb{R}^{n \times k}, \quad U^T U - V^T V = 2I_k \quad \text{and} \quad U^T V = V^T U$$

is a *proper subspace version* of the Thouless functional in the form of $\varrho(\cdot, \cdot)$.

Theorem 3.3. *Suppose that A and B are $n \times n$ real symmetric matrices and that $A + B$ and $A - B$ are positive semi-definite and one of them is definite, and $U, V \in \mathbb{R}^{n \times k}$. Then*

$$\sum_{i=1}^k \lambda_i = \frac{1}{2} \inf_{\substack{U^T U - V^T V = 2I_k \\ U^T V = V^T U}} \text{trace} \left(\begin{pmatrix} U \\ V \end{pmatrix}^T \begin{pmatrix} A & B \\ B & A \end{pmatrix} \begin{pmatrix} U \\ V \end{pmatrix} \right). \quad (3.12)$$

Moreover, “inf” can be replaced by “min” if and only if both $A \pm B$ are definite.

Proof. Assume the assignments in (1.6) for K and M . We have by (1.8)

$$\begin{pmatrix} U \\ V \end{pmatrix}^T \begin{pmatrix} A & B \\ B & A \end{pmatrix} \begin{pmatrix} U \\ V \end{pmatrix} = \begin{pmatrix} \hat{V} \\ \hat{U} \end{pmatrix}^T \begin{pmatrix} M & \\ & K \end{pmatrix} \begin{pmatrix} \hat{V} \\ \hat{U} \end{pmatrix} = \hat{U}^T K \hat{U} + \hat{V}^T M \hat{V},$$

where

$$\begin{pmatrix} \hat{V} \\ \hat{U} \end{pmatrix} = J^T \begin{pmatrix} U \\ V \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} U + V \\ U - V \end{pmatrix}.$$

Therefore

$$\inf_{\hat{U}^T \hat{V} = I_k} \text{trace}(\hat{U}^T K \hat{U} + \hat{V}^T M \hat{V}) = \inf_{(U-V)^T (U+V) = 2I_k} \text{trace} \left(\begin{pmatrix} U \\ V \end{pmatrix}^T \begin{pmatrix} A & B \\ B & A \end{pmatrix} \begin{pmatrix} U \\ V \end{pmatrix} \right). \quad (3.13)$$

We claim

$$(U - V)^T (U + V) = 2I_k \quad \Leftrightarrow \quad U^T U - V^T V = 2I_k \text{ and } U^T V = V^T U. \quad (3.14)$$

This is because $(U - V)^T (U + V) = 2I_k$ and its transpose version give

$$U^T U + U^T V - V^T U - V^T V = 2I_k, \quad (3.15a)$$

$$U^T U + V^T U - U^T V - V^T V = 2I_k. \quad (3.15b)$$

Add both equations in (3.15) to get $U^T U - V^T V = 2I_k$ and subtract one from the other to get $U^T V = V^T U$. That the right-hand side in (3.14) implies its left-hand side can be seen from any of the equations in (3.15). Equation (3.12) is now a consequence of Theorem 3.2, (3.13), and (3.14). \square

3.2 Cauchy-like interlacing inequalities

In the following Theorem 3.4, we obtain inequalities that can be regarded as an extension of Cauchy interlacing inequalities (3.3).

Theorem 3.4. *Suppose that one of K and M is definite. Let $U, V \in \mathbb{R}^{n \times k}$ such that $U^T V$ is nonsingular. Write $W = U^T V = W_1^T W_2$, where $W_i \in \mathbb{R}^{k \times k}$ are nonsingular, and define H_{SR} by (2.19). Denote by $\pm \mu_i$ ($1 \leq i \leq k$) the eigenvalues of H_{SR} , where $0 \leq \mu_1 \leq \dots \leq \mu_k$. Then*

$$\lambda_i \leq \mu_i \leq \frac{\sqrt{\min\{\kappa(K), \kappa(M)\}}}{\cos \angle(\mathcal{U}, \mathcal{V})} \lambda_{i+n-k} \quad \text{for } 1 \leq i \leq k, \quad (3.16)$$

where $\mathcal{U} = \text{span}(U)$ and $\mathcal{V} = \text{span}(V)$. Furthermore, if $\lambda_k < \lambda_{k+1}$ and $\lambda_i = \mu_i$ for $1 \leq i \leq k$, then³

1. $\mathcal{U} = \text{span}(X_{(1:k, \cdot)})$ when M is definite, where X is as in Theorem 2.3;
2. $\{\mathcal{U}, \mathcal{V}\}$ is a pair of deflating subspaces of $\{K, M\}$ corresponding to the eigenvalues $\pm \lambda_i$ ($1 \leq i \leq k$) of H when both K and M are definite.

Proof. The proof is long and deferred to appendix A. \square

³A similar statement for the case in which K is definite (but M is semi-definite) can be made, noting that the decompositions in (2.5) no longer hold but similar decompositions exist.

Corollary 3.2. *Suppose that one of $K, M \in \mathbb{R}^{n \times n}$ is definite. Let K_p and M_p be $k \times k$ principal submatrices of K and M , extracted with same row and column indices for both. Denote by $\pm\mu_i$ ($1 \leq i \leq k$) the eigenvalues of $\begin{pmatrix} 0 & K_p \\ M_p & 0 \end{pmatrix}$, where $0 \leq \mu_1 \leq \dots \leq \mu_k$. Then*

$$\lambda_i \leq \mu_i \leq \sqrt{\min\{\kappa(K), \kappa(M)\}} \lambda_{i+n-k} \quad \text{for } 1 \leq i \leq k. \quad (3.17)$$

Proof. Let i_1, i_2, \dots, i_k be the row and column indices of K and M that give K_p and M_p , and let $U = (e_{i_1}, e_{i_2}, \dots, e_{i_k}) \in \mathbb{R}^{n \times k}$. Then $K_p = U^T K U$ and $M_p = U^T M U$. Apply Theorem 3.4 with $V = U$ to conclude the proof. \square

Inequalities (3.16) and (3.17) mirror Cauchy interlacing inequalities (3.3). But the upper bounds on μ_i by (3.16) and (3.17) are more complicated. The following example shows that the factor $[\cos \angle(\mathcal{U}, \mathcal{V})]^{-1}$ in general cannot be removed. Consider

$$K = \begin{pmatrix} \alpha^2 & 0 \\ 0 & \beta^2 \end{pmatrix}, \quad M = I_2, \quad U = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad V = \begin{pmatrix} t \\ 1 \end{pmatrix},$$

where $0 < \alpha < \beta$ and $t = \tan \angle(\mathcal{U}, \mathcal{V})$. Then the positive eigenvalue of H_{SR} is

$$\sqrt{U^T K U V^T M V} = \beta \sqrt{1 + t^2} = \frac{\beta}{\cos \angle(\mathcal{U}, \mathcal{V})}.$$

We suspect that $\sqrt{\min\{\kappa(K), \kappa(M)\}}$ could be removed or at least be replaceable by something that does not depend on the condition numbers, but we have no proof, except a special case as detailed in the following theorem.

Theorem 3.5. *Under the assumptions of Theorem 3.4, if either $\mathcal{U} \subseteq M\mathcal{V}$ when M is definite or $\mathcal{V} \subseteq K\mathcal{U}$ when K is definite, then*

$$\lambda_i \leq \mu_i \leq \lambda_{i+n-k} \quad \text{for } 1 \leq i \leq k. \quad (3.18)$$

Proof. We will prove (3.18), assuming M is definite and $\mathcal{U} \subseteq M\mathcal{V}$. Since M is definite, $\dim(M\mathcal{V})^\perp = n - k$, where $(M\mathcal{V})^\perp$ is the orthogonal complement of $M\mathcal{V}$. Let $V_\perp \in \mathbb{R}^{n \times (n-k)}$ be a basis matrix of $(M\mathcal{V})^\perp$. Then $V_\perp^T M V_\perp = 0$ and also $U^T V_\perp = 0$ because $\mathcal{U} \subseteq M\mathcal{V}$. Let

$$U = \left(U W_1^{-1}, M V_\perp (V_\perp^T M V_\perp)^{-1/2} \right), \quad V = \left(V W_2^{-1}, V_\perp (V_\perp^T M V_\perp)^{-1/2} \right).$$

It can be verified that $U^T V = I_n$ (which implies $V^T U = I_n$ also) and

$$\widehat{M} \stackrel{\text{def}}{=} V^T M V = \begin{pmatrix} W_2^{-T} V^T M V W_2^{-1} & \\ & I_{n-k} \end{pmatrix}.$$

Let $\widehat{K} = U^T K U$. Notice that

$$\text{eig}(\widehat{K} \widehat{M}) = \text{eig}(\widehat{M}^{1/2} \widehat{K} \widehat{M}^{1/2}) = \{\lambda_i^2, i = 1, 2, \dots, n\},$$

where $\text{eig}(\cdot)$ is the set of eigenvalues of a matrix. The $k \times k$ leading principal matrix of $\widehat{M}^{1/2} \widehat{K} \widehat{M}^{1/2}$ is

$$(W_2^{-T} V^T M V W_2^{-1})^{1/2} (W_1^{-T} U^T K U W_1^{-1}) (W_2^{-T} V^T M V W_2^{-1})^{1/2}$$

whose eigenvalues are μ_i^2 , $i = 1, 2, \dots, k$. Apply Cauchy interlacing inequalities [16] to get

$$\lambda_i^2 \leq \mu_i^2 \leq \lambda_{i+n-k}^2 \quad \text{for } 1 \leq i \leq k$$

which yield (3.18). \square

3.3 Minimization principle and deflation

Deflation is a commonly used technique in solving eigenvalue problems. The basic idea is to avoid computing these eigenpairs that have been already computed to a prescribed accuracy, and it is accomplished by orthogonalizing current vectors against all already converged eigenvectors. Return to the symmetric eigenvalue problem for A we discussed at the beginning of this section. Denote by x_i ($1 \leq i \leq n$) the eigenvectors of A corresponding to θ_i . We may assume $x_i^T x_j = 0$ for $i \neq j$. In (3.1), if “min” is restricted to all x that is orthogonal to x_i ($1 \leq i \leq \ell$), then the minimum becomes $\theta_{\ell+1}$. Similarly, if U is restricted to those such that $U^T x_i = 0$ ($1 \leq i \leq \ell$), then the minimum in (3.2) is $\sum_{i=1}^k \theta_{\ell+i}$, and (3.3) becomes $\theta_{\ell+i} \leq \mu_i \leq \theta_{i+n-k}$. The next theorem gives similar results for H .

Theorem 3.6. *Suppose that K and M are symmetric positive definite. Denote by $z_i = \begin{pmatrix} y_i \\ x_i \end{pmatrix}$ ($1 \leq i \leq n$) the eigenvectors of H corresponding to the positive eigenvalues λ_i , respectively, where all $x_i, y_i \in \mathbb{R}^n$. Suppose⁴ that $\langle z_i, z_j \rangle_{\mathcal{S}} \equiv z_i^T \mathcal{S} z_j = 0$ for $i \neq j$. Set $Y_1 = (y_1, y_2, \dots, y_\ell)$ and $X_1 = (x_1, x_2, \dots, x_\ell)$.*

1. *We have*

$$\sum_{i=1}^k \lambda_{\ell+i} = \frac{1}{2} \inf_{\substack{U, V \in \mathbb{R}^{n \times k}, U^T V = I_k \\ U^T Y_1 = 0, V^T X_1 = 0}} \text{trace}(U^T K U + V^T M V). \quad (3.19)$$

If also $\lambda_{\ell+k} < \lambda_{\ell+k+1}$, then for any U and V that attain the minimum, $\{\text{span}(U), \text{span}(V)\}$ is a pair of deflating subspaces of $\{K, M\}$ corresponding to the eigenvalues $\pm \lambda_{\ell+i}$ ($1 \leq i \leq k$) of H .

2. *Let $U, V \in \mathbb{R}^{n \times k}$ such that $U^T V$ is nonsingular, $U^T Y_1 = 0$ and $V^T X_1 = 0$. Write $W = U^T V = W_1^T W_2$, where $W_i \in \mathbb{R}^{k \times k}$ are nonsingular, and define H_{SR} by (2.19). Denote by $\pm \mu_i$ ($1 \leq i \leq k$) the eigenvalues of H_{SR} , where $0 \leq \mu_1 \leq \dots \leq \mu_k$. Then*

$$\lambda_{\ell+i} \leq \mu_i \leq \frac{\sqrt{\min\{\kappa(K), \kappa(M)\}}}{\cos \angle(\mathcal{U}, \mathcal{V})} \lambda_{i+n-k} \quad \text{for } 1 \leq i \leq k. \quad (3.20)$$

If $\lambda_{\ell+k} < \lambda_{\ell+k+1}$ and if $\lambda_{\ell+i} = \mu_i$ for $1 \leq i \leq k$, then $\{\text{span}(U), \text{span}(V)\}$ is a pair of deflating subspaces of $\{K, M\}$ corresponding to the eigenvalues $\pm \lambda_{\ell+i}$ ($1 \leq i \leq k$) of H .

Proof. See appendix A. □

4 Concluding remarks

We have uncovered new minimization principles and Cauchy-like interlacing inequalities for the LR (a.k.a. RPA) eigenvalue problem arising from the calculation of excitation states of many-particle systems in computational quantum chemistry and physics. In addition, we also obtained a structure-preserving projection H_{SR} of H onto a pair of subspaces. The role of H_{SR} for the LR eigenvalue problem (1.1) in many ways is the same as the Rayleigh quotient matrix for the symmetric eigenvalue problem. These new results mirror the three well-known results for the eigenvalue problem of a real symmetric matrix. They lay the foundation for our numerical investigation in the second paper of this sequel where new efficient numerical methods will

⁴By Theorem 2.2, $\langle z_i, z_j \rangle_{\mathcal{S}} \equiv z_i^T \mathcal{S} z_j = 0$ if $\lambda_i \neq \lambda_j$. In the case when λ_i is a multiple eigenvalue, H has an invariant subspace whose dimension is the same as the algebraic multiplicity of λ_i because H is diagonalizable by Theorem 2.3. Item 1. of Theorem 2.2 guarantees that the invariant subspace has an $\langle \cdot, \cdot \rangle_{\mathcal{S}}$ -orthogonal basis.

be devised for computing the first few smallest positive eigenvalues and their corresponding eigenvectors simultaneously.

Although, throughout this paper and its following one, it is assumed both K and M are real matrices, all results are valid for Hermitian positive semi-definite K and M with one of them being definite after minor changes: replacing all \mathbb{R} by \mathbb{C} and all superscripts $(\cdot)^T$ by complex conjugate transposes $(\cdot)^H$.

The right inequalities in Theorem 3.4 and Corollary 3.2 that mirror Cauchy interlacing inequalities for the standard symmetric eigenvalue problem are not as satisfactory as we would like. We demonstrated that the factor $[\cos \angle(\mathcal{U}, \mathcal{V})]^{-1}$ is in general not removable, but the factor $\sqrt{\min\{\kappa(K), \kappa(M)\}}$ could be an artifact of our proof and thus might be removed. No proof has been found yet.

Acknowledgments

We thank Dario Rocca and Giulia Galli for drawing our attentions to the linear response eigenvalue problem.

A Proofs of Theorems 3.2 – 3.6

Lemma A.1. *Let $\omega_i \in \mathbb{R}$ for $1 \leq i \leq n$ be arranged in ascending order, i.e., $\omega_1 \leq \omega_2 \leq \dots \leq \omega_n$, and let $\alpha_i \in \mathbb{R}$ for $1 \leq i \leq n$. Denote by α_i^\downarrow ($i = 1, \dots, n$) the rearrangement of α_i ($i = 1, \dots, n$) in descending order, i.e., $\alpha_1^\downarrow \geq \dots \geq \alpha_n^\downarrow$. Then*

$$\sum_{i=1}^n \omega_i \alpha_i \geq \sum_{i=1}^n \omega_i \alpha_i^\downarrow. \quad (\text{A.1})$$

If (A.1) is an equality and if $\alpha_k > \alpha_{k+1}$ and $\omega_k < \omega_{k+1}$ for some $1 \leq k < n$, then

$$\{\alpha_j^\downarrow, j = 1, \dots, k\} = \{\alpha_j, j = 1, \dots, k\}. \quad (\text{A.2})$$

Proof. Inequality (A.1) is well-known. See, for example, [4, (II.37) on p.49]. We now prove (A.2), under the conditions that (A.1) is an equality, $\alpha_k^\downarrow > \alpha_{k+1}^\downarrow$, and $\omega_k < \omega_{k+1}$. Suppose, to the contrary, that (A.2) did not hold. Then there would exist

$$\begin{aligned} \ell_1 \leq k \text{ and } \ell_2 > k \text{ such that } \alpha_{\ell_1} &= \alpha_{\ell_2}^\downarrow, \\ j_1 \leq k \text{ and } j_2 > k \text{ such that } \alpha_{j_1}^\downarrow &= \alpha_{j_2}. \end{aligned}$$

Since

$$\begin{aligned} \omega_{\ell_1} \alpha_{\ell_1} + \omega_{j_2} \alpha_{j_2} - (\omega_{\ell_1} \alpha_{j_2} + \omega_{j_2} \alpha_{\ell_1}) &= (\alpha_{j_2} - \alpha_{\ell_1})(\omega_{j_2} - \omega_{\ell_1}) \\ &= (\alpha_{j_1}^\downarrow - \alpha_{\ell_2}^\downarrow)(\omega_{j_2} - \omega_{\ell_1}) \\ &\geq (\alpha_k^\downarrow - \alpha_{k+1}^\downarrow)(\omega_{k+1} - \omega_k) \\ &> 0, \end{aligned}$$

we have

$$\sum_{i=1}^n \omega_i \alpha_i = \sum_{i \neq \ell_1, j_2} \omega_i \alpha_i + \omega_{\ell_1} \alpha_{\ell_1} + \omega_{j_2} \alpha_{j_2} > \sum_{i \neq \ell_1, j_2} \omega_i \alpha_i + \omega_{\ell_1} \alpha_{j_2} + \omega_{j_2} \alpha_{\ell_1} \geq \sum_{i=1}^n \omega_i \alpha_i^\downarrow,$$

contradicting that (A.1) is an equality. This proves (A.2). \square

Lemma A.2. Let $U \in \mathbb{R}^{n \times k}$ and $\Omega = \text{diag}(\omega_1, \omega_2, \dots, \omega_n)$, where $\omega_1 \leq \omega_2 \leq \dots \leq \omega_n$. Then

$$\text{trace}(U^T \Omega U) \geq \sum_{i=1}^k \sigma_i^2 \omega_i, \quad (\text{A.3})$$

where σ_i ($i = 1, \dots, k$) are the k singular values of U in descending order, i.e., $\sigma_1 \geq \dots \geq \sigma_k \geq 0$. If (A.3) is an equality, $\omega_k < \omega_{k+1}$, and $\sigma_k > 0$, then $U_{(k+1:n,:)} = 0$, i.e., the last $n - k$ rows of U are zeros.

Proof. Write $\alpha_i = (UU^T)_{(i,i)}$, the i th diagonal entry of UU^T . By Lemma A.1,

$$\text{trace}(U^T \Omega U) = \text{trace}(UU^T \Omega) = \sum_{i=1}^n \omega_i \alpha_i \geq \sum_{i=1}^n \omega_i \alpha_i^\downarrow, \quad (\text{A.4})$$

where α_i^\downarrow ($i = 1, \dots, n$) are defined as in Lemma A.1. Since UU^T is symmetric positive semidefinite, its diagonal entries: α_i ($i = 1, \dots, n$) are majorized by its n eigenvalues: σ_i^2 ($i = 1, \dots, k$) and $\sigma_i^2 = 0$ ($i = k + 1, \dots, n$) [4, (II.14) on p.35], meaning

$$t_j \stackrel{\text{def}}{=} \sum_{i=1}^j \alpha_i^\downarrow \leq s_j \stackrel{\text{def}}{=} \sum_{i=1}^j \sigma_i^2 \text{ for } 1 \leq j \leq n - 1, \text{ and } t_n = s_n. \quad (\text{A.5})$$

Therefore, by [10, Lemma 2.3],

$$\sum_{i=1}^n \omega_i \alpha_i^\downarrow \geq \sum_{i=1}^k \omega_i \sigma_i^2 \quad (\text{A.6})$$

which, combined with (A.4), lead to (A.3). But in order to characterize those matrices U that make (A.3) an equality, we need to look into when (A.6) becomes an equality. To that end, we still have to give a proof of (A.6), despite of [10, Lemma 2.3]. Let $t_0 = s_0 = 0$. We have

$$\begin{aligned} \sum_{i=1}^n \omega_i \alpha_i^\downarrow &= \sum_{i=1}^n \omega_i (t_i - t_{i-1}) \\ &= \omega_n t_n + \sum_{i=1}^{n-1} (\omega_i - \omega_{i+1}) t_i \\ &\geq \omega_n s_n + \sum_{i=1}^{n-1} (\omega_i - \omega_{i+1}) s_i \\ &= \sum_{i=1}^n \omega_i \sigma_i^2 = \sum_{i=1}^k \omega_i \sigma_i^2. \quad (\sigma_i = 0 \text{ for } i > k) \end{aligned} \quad (\text{A.7})$$

This is (A.6).

Now if (A.3) is an equality and if $\omega_k < \omega_{k+1}$, then the equal sign in (A.7) must hold and thus $t_k = s_k$ because $\omega_k - \omega_{k+1} < 0$. It follows from $\sigma_i^2 = 0$ ($i = k + 1, \dots, n$) that $t_k = s_k = \dots = s_n = t_n$; so $\alpha_j^\downarrow = 0$ for $j > k$ by (A.5). Because (A.4) must be an equality, $\alpha_k^\downarrow > 0 = \alpha_{k+1}^\downarrow$ (since $\sigma_k > 0$), and $\omega_k < \omega_{k+1}$, we conclude by Lemma A.1 that (A.2) holds, and thus $\alpha_j = (UU^T)_{(j,j)} = 0$ for $j > k$ which implies

$$(UU^T)_{(i,j)} = 0 \quad \text{for } \max\{i, j\} > k$$

because UU^T is symmetric positive semi-definite. In particular

$$U_{(k+1:n,:)} U_{(k+1:n,:)}^T = (UU^T)_{(k+1:n,k+1:n)} = 0$$

which implies $U_{(k+1:n,:)} = 0$, as expected. \square

Proof of Theorem 3.2. Suppose that M is definite. Equations in (2.5) hold for some non-singular $Y \in \mathbb{R}^{n \times n}$ and $X = Y^{-T}$. We have by (2.5)

$$\begin{aligned} U^T K U + V^T M V &= U^T Y \Lambda^2 Y^T U + V^T X X^T V \\ &= \hat{U}^T \Lambda^2 \hat{U} + \hat{V}^T \hat{V}, \end{aligned} \quad (\text{A.8})$$

where $\hat{U} = Y^T U$ and $\hat{V} = X^T V$. It can be verified that $\hat{U}^T \hat{V} = U^T V$ and that the correspondences between U and \hat{U} and between V and \hat{V} are one-one. Therefore

$$\inf_{U^T V = I_k} \text{trace}(U^T K U + V^T M V) = \inf_{\hat{U}^T \hat{V} = I_k} \text{trace}(\hat{U}^T \Lambda^2 \hat{U} + \hat{V}^T \hat{V}). \quad (\text{A.9})$$

For any given \hat{U} and \hat{V} , denote their singular values, respectively, by α_i ($i = 1, \dots, k$) and β_i ($i = 1, \dots, k$) in descending order. Then by Lemma A.2

$$\text{trace}(\hat{U}^T \Lambda^2 \hat{U} + \hat{V}^T \hat{V}) \geq \sum_{i=1}^k \alpha_i^2 \lambda_i^2 + \sum_{i=1}^k \beta_i^2 \quad (\text{A.10})$$

$$\begin{aligned} &= \sum_{i=1}^k (\alpha_i^2 \lambda_i^2 + \beta_{k-i+1}^2) \\ &\geq 2 \sum_{i=1}^k \alpha_i \beta_{k-i+1} \lambda_i \end{aligned} \quad (\text{A.11})$$

$$\geq 2 \sum_{i=1}^k \lambda_i. \quad (\text{A.12})$$

The last inequality holds because of [8, (3.3.18) on p.178] which says $\alpha_i \beta_{k-i+1}$ is greater or equal to the k th largest singular value of $U^T V = I_k$ which is 1. Combine (A.9) and (A.12) to get

$$\frac{1}{2} \inf_{U^T V = I_k} \text{trace}(U^T K U + V^T M V) \geq \sum_{i=1}^k \lambda_i. \quad (\text{A.13})$$

Now if all $\lambda_i > 0$ (i.e., K is also definite), then it can be seen that picking U and V such that

$$\hat{U} = \begin{pmatrix} \text{diag}(\lambda_1^{-1/2}, \dots, \lambda_k^{-1/2}) \\ 0 \end{pmatrix}, \quad \hat{V} = \begin{pmatrix} \text{diag}(\lambda_1^{1/2}, \dots, \lambda_k^{1/2}) \\ 0 \end{pmatrix},$$

gives $\frac{1}{2} \text{trace}(U^T K U + V^T M V) = \sum_{i=1}^k \lambda_i$ which, together with (A.13), yield (3.8) with “inf” replaced by “min”.

When K is singular, $\lambda_1 = 0$ and (A.11) is always a strict inequality. So

$$\frac{1}{2} \text{trace}(U^T K U + V^T M V) > \sum_{i=1}^k \lambda_i \quad \text{for any } U^T V = I_k. \quad (\text{A.14})$$

Suppose $0 = \lambda_1 = \dots = \lambda_\ell < \lambda_{\ell+1} \leq \dots \leq \lambda_k$. We pick U and V such that

$$\hat{U} = \begin{pmatrix} \begin{array}{c|c} \epsilon^{-1} I_\ell & \\ \hline & \text{diag}(\lambda_{\ell+1}^{-1/2}, \dots, \lambda_k^{-1/2}) \end{array} \\ 0 \end{pmatrix}, \quad \hat{V} = \begin{pmatrix} \begin{array}{c|c} \epsilon I_\ell & \\ \hline & \text{diag}(\lambda_{\ell+1}^{1/2}, \dots, \lambda_k^{1/2}) \end{array} \\ 0 \end{pmatrix}.$$

Then $\frac{1}{2} \text{trace}(U^T KU + V^T MV) = \sum_{i=1}^k \lambda_{i+\ell} \epsilon^2$ which goes to $\sum_{i=1}^k \lambda_i$ as $\epsilon \rightarrow 0$. So we have (3.8) by (A.14), and ‘‘inf’’ cannot be replaced by ‘‘min’’.

Now suppose $0 < \lambda_1$ and $\lambda_k < \lambda_{k+1}$ and suppose that U and V attain the minimum, i.e.,

$$\frac{1}{2} \text{trace}(U^T KU + V^T MV) = \sum_{i=1}^k \lambda_i.$$

For this to happen, all equal signs in (A.10), (A.11), and (A.12) must take place. For the equality sign in (A.10) to take place, by Lemma A.2 we have $\hat{U}_{(k+1:n,:)} = 0$. Partition

$$\hat{U} = \begin{pmatrix} \hat{U}_1 \\ 0 \end{pmatrix}, \quad \hat{V} = \begin{pmatrix} \hat{V}_1 \\ \hat{V}_2 \end{pmatrix}, \quad \hat{U}_1, \hat{V}_1 \in \mathbb{R}^{k \times k}.$$

We claim $\hat{V}_2 = 0$, too. Here is why. For the equality sign in (A.12) to take place, we have $\alpha_i \beta_{k-i+1} = 1$ for $1 \leq i \leq k$. Now $I_k = \hat{U}^T \hat{V} = \hat{U}_1^T \hat{V}_1$ implies $\alpha_i \gamma_{k-i+1} \geq 1$ [8, (3.3.18) on p.178], where γ_i ($i = 1, \dots, k$) are the singular values of \hat{V}_1 in descending order. Since $\hat{V}^T \hat{V} = \hat{V}_1^T \hat{V}_1 + \hat{V}_2^T \hat{V}_2$, we have $\gamma_i \leq \beta_i$ for $1 \leq i \leq k$ and thus

$$1 \leq \alpha_i \gamma_{k-i+1} \leq \alpha_i \beta_{k-i+1} = 1$$

which implies $\gamma_i = \beta_i$ for $1 \leq i \leq k$. So $\hat{V}_2 = 0$. Now use $U = X\hat{U}$ and $V = Y\hat{V}$ to conclude that $\{\text{span}(U), \text{span}(V)\}$ is the pair of deflating subspaces of $\{K, M\}$ corresponding to the eigenvalues $\pm \lambda_i$ ($1 \leq i \leq k$) of H . \square

Proof of Theorem 3.4. Assume that M is definite. Without loss of generality, we may simply assume $U^T V = I_k$ and $W_1 = W_2 = I_k$; otherwise substitutions:

$$U \leftarrow UW_1^{-1}, \quad V \leftarrow VW_2^{-1}, \quad I_k \leftarrow W_1, \quad I_k \leftarrow W_2,$$

will give new U and V with $U^T V = I_k$ and at the same time the same H_{SR} .

Equations in (2.5) hold for some nonsingular $Y \in \mathbb{R}^{n \times n}$ and $X = Y^{-T}$. Then

$$U^T KU = U^T Y A^2 Y^T U = \hat{U}^T A^2 \hat{U}, \tag{A.15a}$$

$$V^T MV = V^T X X^T V = \hat{V}^T \hat{V}, \tag{A.15b}$$

where $\hat{U} = Y^T U$ and $\hat{V} = X^T V$. Still $\hat{U}^T \hat{V} = U^T V = I_k$. Decompose \hat{V} as

$$\tilde{V} \stackrel{\text{def}}{=} Q^T \hat{V} = \begin{pmatrix} \tilde{V}_1 \\ 0 \end{pmatrix}, \quad Q^T Q = I_n, \quad \tilde{V}_1 \text{ nonsingular.} \tag{A.16}$$

This can be proved, for example, using SVD of \hat{V} . Then $\hat{V}^T \hat{V} = \tilde{V}_1^T \tilde{V}_1$. Partition

$$\tilde{U} \stackrel{\text{def}}{=} Q^T \hat{U} = \begin{pmatrix} \tilde{U}_1 \\ \tilde{U}_2 \end{pmatrix}, \quad \tilde{U}_1 \in \mathbb{R}^{k \times k}. \tag{A.17}$$

Then $\hat{U}^T \hat{V} = (Q^T \hat{U})^T Q^T \hat{V} = \tilde{U}_1^T \tilde{V}_1 = I_k$ which implies $\tilde{U}_1^T = \tilde{V}_1^{-1}$. Set

$$A = Q^T A^2 Q, \quad E = \tilde{U}_2 \tilde{V}_1^T \tag{A.18}$$

to get

$$\hat{U}^T A^2 \hat{U} = \tilde{U}^T A \tilde{U}, \quad \tilde{U} \tilde{V}_1^T = \begin{pmatrix} I_k \\ E \end{pmatrix}. \tag{A.19}$$

By Theorem 2.1, μ_i^2 ($1 \leq i \leq k$) are all the eigenvalues of

$$(U^T K U)(V^T M V) = (\hat{U}^T A^2 \hat{U})(\hat{V}^T \hat{V}) = (\tilde{U}^T A \tilde{U})(\tilde{V}_1^T \tilde{V}_1) \quad (\text{A.20})$$

whose eigenvalues are the same as $\tilde{V}_1^T (\tilde{U}^T A \tilde{U}) \tilde{V}_1$, a real symmetric positive semi-definite matrix. Set

$$P = \tilde{U} \tilde{V}_1^T (I_k + E^T E)^{-1/2}.$$

Then $P^T P = I_k$ by (A.19). Denote by ν_i ($1 \leq i \leq k$) the eigenvalues of $P^T A P$ in ascending order. We have

$$\lambda_i^2 \leq \nu_i \leq \lambda_{i+n-k}^2 \quad \text{for } 1 \leq i \leq k \quad (\text{A.21})$$

by Cauchy interlacing theorem [16, 21]. For any $\hat{u} \in \mathbb{R}^k$, letting $u = (I_k + E^T E)^{1/2} \hat{u}$ gives

$$(1 + \|E\|_2^2) \frac{u^T (P^T A P) u}{u^T u} \geq \frac{\hat{u}^T \left[\tilde{V}_1 (\tilde{U}^T A \tilde{U}) \tilde{V}_1^T \right] \hat{u}}{\hat{u}^T \hat{u}} \geq \frac{u^T (P^T A P) u}{u^T u} \quad (\text{A.22})$$

since

$$\hat{u}^T \hat{u} \leq u^T u = \hat{u}^T \hat{u} + \hat{u}^T E^T E \hat{u} \leq (1 + \|E\|_2^2) \hat{u}^T \hat{u}.$$

Denote by $\hat{\mathcal{U}}_i$ and \mathcal{U}_i subspaces of \mathbb{R}^k of dimension i . Using the Courant-Fisher min-max principle (see [16, p.206], [21, p.201]), we have

$$\begin{aligned} \mu_i^2 &= \min_{\hat{\mathcal{U}}_i} \max_{\hat{u} \in \hat{\mathcal{U}}_i} \frac{\hat{u}^T \left[\tilde{V}_1 (\tilde{U}^T A \tilde{U}) \tilde{V}_1^T \right] \hat{u}}{\hat{u}^T \hat{u}} \\ &\geq \min_{\mathcal{U}_i = (I_k + E^T E)^{1/2} \hat{\mathcal{U}}_i} \max_{u \in \mathcal{U}_i} \frac{u^T (P^T A P) u}{u^T u} && \text{(by (A.22))} \\ &= \min_{\mathcal{U}_i} \max_{u \in \mathcal{U}_i} \frac{u^T (P^T A P) u}{u^T u} \\ &= \nu_i \geq \lambda_i^2, && \text{(by (A.21))} \\ \mu_i^2 &\leq (1 + \|E\|_2^2) \min_{\mathcal{U}_i = (I_k + E^T E)^{1/2} \hat{\mathcal{U}}_i} \max_{u \in \mathcal{U}_i} \frac{u^T (P^T A P) u}{u^T u} && \text{(by (A.22))} \\ &= (1 + \|E\|_2^2) \nu_i \\ &\leq (1 + \|E\|_2^2) \lambda_{i+n-k}^2. && \text{(by (A.21))} \end{aligned}$$

It remains to bound $1 + \|E\|_2^2$. We have from (A.16) – (A.19)

$$\begin{aligned} \sqrt{1 + \|E\|_2^2} &= \|\tilde{U} \tilde{V}_1^T\|_2 \leq \|\tilde{U}\|_2 \|\tilde{V}_1^T\|_2 \\ &= \|\hat{U}\|_2 \|\hat{V}\|_2 = \|Y^T U\|_2 \|X^T V\|_2 \\ &\leq \|Y^T\|_2 \|Y^{-1}\|_2 \|U\|_2 \|V\|_2 = \sqrt{\kappa(M)} \|U\|_2 \|V\|_2. \end{aligned} \quad (\text{A.23})$$

In Theorem 2.7, we proved that the eigenvalues of H_{SR} do not change with respect to the choices of basis matrices. Which means, in proving this theorem, we can use H_{SR} constructed from different basis matrices for \mathcal{U} and \mathcal{V} . What we are going to do is to pick new U and V such that the right hand side of (A.23) is

$$\frac{\sqrt{\kappa(M)}}{\cos \angle(\mathcal{U}, \mathcal{V})}.$$

To this end, we compute QR decompositions

$$U = Q_1 R_1, \quad V = Q_2 R_2,$$

where $Q_1, Q_2 \in \mathbb{R}^{n \times k}$ have orthonormal columns. By [21, Theorem 5.2 on p.40], there are orthogonal matrices $P \in \mathbb{R}^{n \times n}$ and $S_1, S_2 \in \mathbb{R}^{k \times k}$ such that

$$PQ_1S_1 = \begin{matrix} & & k & & \\ & & \begin{pmatrix} I \\ 0 \\ 0 \end{pmatrix} & & \\ k & & & & \\ & & n-2k & & \end{matrix}, \quad PQ_2S_2 = \begin{matrix} & & k & & \\ & & \begin{pmatrix} \Gamma \\ \Sigma \\ 0 \end{pmatrix} & & \\ k & & & & \\ & & n-2k & & \end{matrix} \quad \text{if } 2k \leq n, \quad (\text{A.24a})$$

$$PQ_1S_1 = \begin{matrix} & n-k & 2k-n & & \\ & \begin{pmatrix} I & 0 \\ 0 & I \\ 0 & 0 \end{pmatrix} & & & \\ n-k & & & & \\ & & 2k-n & & \\ & & n-k & & \end{matrix}, \quad PQ_2S_2 = \begin{matrix} & n-k & 2k-n & & \\ & \begin{pmatrix} \Gamma & 0 \\ 0 & I \\ \Sigma & 0 \end{pmatrix} & & & \\ n-k & & & & \\ & & 2k-n & & \\ & & n-k & & \end{matrix} \quad \text{if } 2k > n, \quad (\text{A.24b})$$

where $\Gamma = \text{diag}(\gamma_1, \dots, \gamma_\ell)$ and $\Sigma = \text{diag}(\sigma_1, \dots, \sigma_\ell)$, $\ell = k$ or $n - k$, all $\gamma_i, \sigma_i \geq 0$ and $\gamma_i^2 + \sigma_i^2 = 1$. With (A.24), we pick new U and V to be

$$P^T \begin{pmatrix} \Gamma^{-1} \\ 0 \\ 0 \end{pmatrix}, \quad P^T \begin{pmatrix} \Gamma \\ \Sigma \\ 0 \end{pmatrix} \quad \text{if } 2k \leq n,$$

$$P^T \begin{pmatrix} \Gamma^{-1} & 0 \\ 0 & I_{2n-k} \\ 0 & 0 \end{pmatrix}, \quad P^T \begin{pmatrix} \Gamma & 0 \\ 0 & I_{2n-k} \\ \Sigma & 0 \end{pmatrix} \quad \text{if } 2k > n,$$

respectively. These new U and V span the same space as the old U and V and satisfy $U^T V = I_k$ and $\|U\|_2 \|V\|_2 = [\cos \angle(\mathcal{U}, \mathcal{V})]^{-1}$. The proof of (3.16) is completed for the case when M is definite.

Now if $\lambda_k < \lambda_{k+1}$ and $\lambda_i = \mu_i$ for all $i = 1, 2, \dots, k$, then $\nu_i = \lambda_i^2$ for all $i = 1, 2, \dots, k$ since $\mu_i^2 \geq \nu_i \geq \lambda_i^2$. In particular, $\text{trace}(P^T A P) = \sum_{i=1}^k \lambda_i^2$. Apply [10, Theorem 2.2] or [9, Theorem 4] on $-A = Q^T(-A^2)Q$ to conclude that $(QP)_{(1:k,:)}$ is orthogonal and $(QP)_{(k+1:n,:)} = 0$. Write

$$\hat{U} = \begin{pmatrix} \hat{U}_1 \\ \hat{U}_2 \end{pmatrix}, \quad \hat{V} = \begin{pmatrix} \hat{V}_1 \\ \hat{V}_2 \end{pmatrix}, \quad \hat{U}_1, \hat{V}_1 \in \mathbb{R}^{k \times k}, \quad A_1^2 = \text{diag}(\lambda_1^2, \dots, \lambda_k^2).$$

Since $QP = \hat{U} \tilde{V}_1^T (I_k + E^T E)^{-1/2}$ by (A.17), we conclude that $\hat{U}_2 = 0$ and thus $\mathcal{U} = \text{span}(X_{(1:k,:)})$. Use $\hat{U}^T \hat{V} = I_k$ to get $\hat{U}_1^T \hat{V}_1 = I_k$ or equivalently $\hat{U}_1^T = \hat{V}_1^{-1}$. Note, by (A.20),

$$(U^T K U)(V^T M V) = \hat{U}_1^T A_1^2 \hat{U}_1 \hat{V}_1^T \hat{V}_1$$

which has the same eigenvalues as $A_1^2 \hat{U}_1 \hat{V}_1^T \hat{V}_1 \hat{U}_1^T$ which has the same eigenvalues as

$$A_1 \hat{U}_1 \hat{V}_1^T \hat{V}_1 \hat{U}_1^T A_1 = A_1^2 + A_1 \hat{U}_1 \hat{V}_2^T \hat{V}_2 \hat{U}_1^T A_1.$$

Since by assumption the eigenvalues of $(U^T K U)(V^T M V)$ are λ_i^2 ($1 \leq i \leq k$), we have

$$\sum_{i=1}^k \lambda_i^2 = \text{trace}(A_1^2 + A_1 \hat{U}_1 \hat{V}_2^T \hat{V}_2 \hat{U}_1^T A_1) = \sum_{i=1}^k \lambda_i^2 + \text{trace}(A_1 \hat{U}_1 \hat{V}_2^T \hat{V}_2 \hat{U}_1^T A_1)$$

which implies $\text{trace}(A_1 \hat{U}_1 \hat{V}_2^T \hat{V}_2 \hat{U}_1^T A_1) = 0$ and thus if $\lambda_1 > 0$, then $\hat{V}_2 \hat{U}_1^T = 0 \Rightarrow \hat{V}_2 = 0$. Therefore

$$U = X \hat{U} = X_{(1:k,:)} \hat{U}_1, \quad V = Y \hat{V} = Y_{(1:k,:)} \hat{V}_1,$$

as expected. \square

Proof of Theorem 3.6. Equations in (2.5) holds for some nonsingular $Y \in \mathbb{R}^{n \times n}$ and $X = Y^{-T}$. Since the columns of $Z = \begin{pmatrix} YA \\ X \end{pmatrix}$ are the eigenvectors of H corresponding to λ_i ($i = 1, 2, \dots, n$) and the eigenvectors corresponding to a multiple λ_i can be picked as any $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ -orthogonal basis vectors of the associated invariant subspace, we may assume that z_i is parallel to $Z_{(:,i)}$, the i th column of Z . Now for any $U^T X_1 = 0$ and $V^T Y_1 = 0$, $\hat{U}^T A^2 \hat{U}$ and $\hat{V}^T \hat{V}$ in (A.8) and (A.15) become

$$\hat{U}^T A^2 \hat{U} = \hat{U}_2^T A_2^2 \hat{U}_2, \quad \hat{V}^T \hat{V} = \hat{V}_2^T \hat{V}_2,$$

where

$$\hat{U} = \begin{matrix} \ell \\ n-\ell \end{matrix} \begin{pmatrix} 0 \\ \hat{U}_2 \end{pmatrix}, \quad \hat{V} = \begin{matrix} \ell \\ n-\ell \end{matrix} \begin{pmatrix} 0 \\ \hat{V}_2 \end{pmatrix}, \quad A_2 = \text{diag}(\lambda_{\ell+1}, \dots, \lambda_n).$$

The rest of the proof are the same as the corresponding parts in the proofs of Theorems 3.2 and 3.4. \square

References

- [1] Z. Bai and R.-C. Li. Minimization principles for the linear response eigenvalue problem II: Computation. submitted, 2011.
- [2] P. Benner, H. Faßbender, and M. Stoll. A Hamiltonian Krylov-Schur-type method based on the symplectic Lanczos process. Report 09/32, Oxford Centre for Collaborative Applied Mathematics, September 2009.
- [3] 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.*, 78:329–357, 1998.
- [4] R. Bhatia. *Matrix Analysis*. Springer, New York, 1996.
- [5] M. E. Casida. Time-dependent density-functional response theory for molecules. In D. P. Chong, editor, *Recent advances in Density Functional Methods*, pages 155–189, World Scientific, Singapore, 1995.
- [6] B. E. Chi. The eigenvalue problem for collective motion in the random phase approximation. *Nuclear Physics*, A146(2):449–456, 1970.
- [7] M. Gruning, A. Marini, and X. Gonze. Exciton-plasmon states in nanoscale materials: breakdown of the Tamm-Dancoff approximation. *Nano Letters*, 9:2820–2824, 2009.
- [8] R. A. Horn and C. R. Johnson. *Topics in Matrix Analysis*. Cambridge University Press, Cambridge, 1991.
- [9] J. Kovač-Striko and K. Veselić. Some remarks on the spectra of Hermitian matrices. *Linear Algebra Appl.*, 145:221–229, 1991.
- [10] R.-C. Li. Accuracy of computed eigenvectors via optimizing a Rayleigh quotient. *BIT*, 44(3):585–593, 2004.
- [11] 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(6):064114, 2008.

- [12] M. T. Lusk and A. E. Mattsson. High-performance computing for materials design to advance energy science. *MRS Bulletin*, 36:169–174, 2011.
- [13] C. Mehl, V. Mehrmann, and H. Xu. Canonical forms for doubly structured matrices and pencils. *The Electronic Journal of Linear Algebra*, 7:112–151, 2000.
- [14] J. Olsen, H. J. Aa. Jensen, and P. Jørgensen. Solution of the large matrix equations which occur in response theory. *J. Comput. Phys.*, 74(2):265 – 282, 1988.
- [15] G. Onida, L. Reining, and A. Rubio. Electronic excitations: density-functional versus many-body Green’s-function approaches. *Rev. Mod. Phys.*, 74(2):601–659, 2002.
- [16] B. N. Parlett. *The Symmetric Eigenvalue Problem*. SIAM, Philadelphia, 1998.
- [17] 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 Physics A*, 249:249–268, 2001.
- [18] P. Ring and P. Schuck. *The Nuclear Many-Body Problem*. Springer-Verlag, New York, 1980.
- [19] D. Rocca, D. Lu, and G. Galli. *Ab initio* calculations of optical absorption spectra: solution of the Bethe-Salpeter equation within density matrix perturbation theory. *J. Chem. Phys.*, 133:164109–10, 2010.
- [20] Y. Saad, J. R. Chelikowsky, and S. M. Shontz. Numerical methods for electronic structure calculations of materials. *SIAM Rev.*, 52:3–54, 2010.
- [21] G. W. Stewart and Ji-Guang Sun. *Matrix Perturbation Theory*. Academic Press, Boston, 1990.
- [22] 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.*, 109:8218–8824, 1998.
- [23] D. J. Thouless. *The quantum mechanics of many-body systems*. Academic, 1972.
- [24] D.J. Thouless. Vibrational states of nuclei in the random phase approximation. *Nuclear Physics*, 22(1):78 – 95, 1961.
- [25] E. V. Tsiper. Variational procedure and generalized Lanczos recursion for small-amplitude classical oscillations. *JETP Letters*, 70(11):751–755, 1999.
- [26] E. V. Tsiper. A classical mechanics technique for quantum linear response. *J. Phys. B: At. Mol. Opt. Phys.*, 34(12):L401–L407, 2001.
- [27] D. S. Watkins. *The Matrix Eigenvalue Problems: GR and Krylov Subspace Methods*. SIAM, Philadelphia, 2007.