Extensions of Wielandt's Min-max Principles for Positive Semi-Definite Pencils

Xin Liang* Ren-Cang Li†

February 22, 2013

Abstract

There are numerous min-max principles about the eigenvalues of a Hermitian matrix. The most general ones are Wielandt's min-max principles which include the Courant-Fischer min-max principles and the trace minimization principles as special cases. In this paper, various extensions of Wielandt's principles are obtained for a positive semi-definite pencil $A - \lambda B$ by which we mean that A and B are Hermitian and there is a real number λ_0 such that $A - \lambda_0 B$ is positive semi-definite.

Key words. Positive semi-definite Hermitian matrix pencil, Wielandt's min-max principle, Courant-Fischer min-max principle, trace minimization principle, eigenvalue.

AMS subject classifications. Primary 15A18. Secondary 15A22, 15A42.

1 Introduction

Consider Hermitian matrix $A \in \mathbb{C}^{n\times n}$. Denote its eigenvalues by λ_i $(i = 1, 2, \dots, n)$ arranged in the ascending order:

$$\lambda_1 \le \lambda_2 \le \cdots \le \lambda_n. \tag{1.1}$$

There are numerous min-max principles for these eigenvalues. The most general ones are perhaps these of Wielandt[1] (see [15] and also [2, p.67], [13, p.199]):

$$\sum_{j=1}^{k} \lambda_{i_j} = \min_{\substack{\mathcal{X}_1\subset\cdots\subset\mathcal{X}_k \\ \dim \mathcal{X}_j = i_j}} \max_{\substack{X=[x_1,\dots,x_k],\, x_j\in\mathcal{X}_j \\ X^{\mathrm{H}}X=I_k}} \operatorname{trace}(X^{\mathrm{H}}AX), \tag{1.2a}$$

$$\sum_{j=1}^{k} \lambda_{i_j} = \max_{\substack{\mathcal{X}_1\supset\cdots\supset\mathcal{X}_k \\ \operatorname{codim} \mathcal{X}_j = i_j - 1}} \min_{\substack{X=[x_1,\dots,x_k],\, x_j\in\mathcal{X}_j \\ X^{\mathrm{H}}X=I_k}} \operatorname{trace}(X^{\mathrm{H}}AX), \tag{1.2b}$$

---

*School of Mathematical Sciences, Peking University, Beijing, 100871, P. R. China. E-mail: liangxinslm@pku.edu.cn. Supported in part by China Scholarship Council. This author is currently a visiting student at Department of Mathematics, University of Texas at Arlington, Arlington, TX 76019.

†Department of Mathematics, University of Texas at Arlington, P.O. Box 19408, Arlington, TX 76019. E-mail: rcli@uta.edu. Supported in part by NSF grant DMS-1115834.

[1]These equations are equivalent to the ones in [2, 13] by applying (1.2) to $-A$ to go from (1.2) to the ones in the references and vice versa.

where $1 \le i_1 < \cdots < i_k \le n$, I_k is $k \times k$ the identity matrix, and $\mathcal{X}_j$ are subspaces of $\mathbb{C}^n$ subject to the specified dimension constraints. Equations in (1.2) with $k = 1$ gives the Courant-Fischer min-max principles (see [11, p.206], [13, p.201]):

$$\lambda_i = \min_{\dim \mathcal{X}=i} \max_{x\in\mathcal{X}} \frac{x^{\mathrm{H}}Ax}{x^{\mathrm{H}}x}, \quad \lambda_i = \max_{\operatorname{codim} \mathcal{X}=i-1} \min_{x\in\mathcal{X}} \frac{x^{\mathrm{H}}Ax}{x^{\mathrm{H}}x} \tag{1.3}$$

which, in particular, yields

$$\lambda_1 = \min_{x\neq 0} \frac{x^{\mathrm{H}}Ax}{x^{\mathrm{H}}x}, \quad \lambda_n = \max_{x\neq 0} \frac{x^{\mathrm{H}}Ax}{x^{\mathrm{H}}x}. \tag{1.4}$$

Equation (1.2b) with $i_j = j$ for $1 \le j \le k$ gives the usually trace minimization principle:

$$\sum_{i=1}^{k} \lambda_i = \min_{X^{\mathrm{H}}X=I_k} \operatorname{trace}(X^{\mathrm{H}}AX). \tag{1.5}$$

Wielandt's min-max principles in (1.2) (and thus (1.3) – (1.5) as well) can be straightforwardly extended to the *generalized eigenvalue problem* for a Hermitian matrix pencil $A-\lambda B$, where A, $B \in \mathbb{C}^{n\times n}$ are Hermitian and B positive definite, by exploiting the equivalence between this generalized eigenvalue problem with the standard eigenvalue problem for $B^{-1/2}AB^{-1/2}$. The outcomes are two equations that are the same as these in (1.2) with $X^{\mathrm{H}}X = I_k$ replaced by $X^{\mathrm{H}}BX = I_k$.

In general, a regular Hermitian pencil can have nonreal eigenvalues and non-semi-simple eigenvalues (those corresponding Jordan blocks of order 2×2 or higher). Thus we cannot expect every eigenvalue of a regular Hermitian pencil admits a characterization like (1.3) or gets involved in trace-like min-max principles. Nonetheless, some of them do and there have been a few non-straightforward (as oppose to the ones we mentioned above for the case when B is positive definite) extensions of the principles (1.2) – (1.5).

1. For the Courant-Fischer min-max principle (1.3), extensions are available for a regular Hermitian pencil (i.e., $\det(A - \lambda B) \not\equiv 0$) with nonsingular B [4, 6, 8, 16] and with possibly singular B [3, 9]. In the extensions, only certain real semi-simple eigenvalues admit a characterization like (1.3). Earlier ones appeared in [12, 14].

2. For the trace minimization principle (1.5), extensions were obtained for a positive semi-definite Hermitian pencil (to be defined in Definition 2.1 below) with nonsingular B [5] and with possibly singular B [7] (including possibly $A - \lambda B$ a singular pencil).

3. For Wielandt's min-max principles (1.2), Nakić and Veselić [10] obtained an extension for a regular Hermitian pencil with nonsingular B and whose *none-cancelled* real eigenvalues are all semi-simple[2].

In this paper, we will extend the Wielandt min-max principles (1.2), too, as in [10], but to a positive semi-definite Hermitian pencil with B possibly singular and as well as $A - \lambda B$ being possibly singular.

[2]The reader is cautioned about a few inaccurate/incorrect statements/equations in [10], however.

The rest of this paper is organized as follows. Main results are stated in section 2, with proofs given in section 3. A brief conclusion is given in section 4.

Notation. Throughout this paper, $\mathbb{C}^{n\times m}$ is the set of all $n \times m$ complex matrices, $\mathbb{C}^n = \mathbb{C}^{n\times 1}$, and $\mathbb{C} = \mathbb{C}^1$. $\mathbb{R}$ is the set of all real numbers. 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 jth column. $X^{\rm H}$ is the conjugate transpose of a vector or matrix. $A \succ 0$ ($A \succeq 0$) means that A is Hermitian positive (semi-)definite, and $A \prec 0$ ($A \preceq 0$) if $-A \succ 0$ ($-A \succeq 0$).

2 Main Results

We will use integer triplet $(i_+(H), i_0(H), i_-(H))$ for the inertia of a Hermitian matrix H, where $i_+(H)$, $i_0(H)$, and $i_-(H)$ are the number of positive, zero, and negative eigenvalues of H, respectively.

Definition 2.1 ([7]). *$A - \lambda B$ is a* Hermitian pencil *of order n if both A, $B \in \mathbb{C}^{n\times n}$ are Hermitian. $A-\lambda B$ is a* positive (semi-)definite matrix pencil *of order n if it is a Hermitian pencil of order n and if there exists $\lambda_0 \in \mathbb{R}$ such that $A - \lambda_0 B$ is positive (semi-)definite.*

According to this definition, a Hermitian pencil $A - \lambda B$ with $B \succ 0$, including $B = I$ in particular, is a positive definite matrix pencil because $A - \lambda_0 B \succ 0$ for all $\lambda_0 < 0$ sufficiently large in magnitude.

In [12, 14], min-max characterizations were obtained for the eigenvalues of a positive semi-definite matrix $C \in \mathbb{C}^{n\times n}$ in an indefinite-inner product space with the indefinite-inner product $[x, y]_B := y^{\rm H}Bx$, where $B \in \mathbb{C}^{n\times n}$ is Hermitian, nonsingular, and indefinite. It turns out that such eigenvalue problems fall into the category of the ones for positive (semi-)definite matrix pencils of Definition 2.1. In fact, that C is positive semi-definite in $[\,\cdot\,,\cdot\,]_B$ is equivalent to $BC = C^{\rm H}B \succeq 0$. Note that the eigenvalue problem $Cx = \lambda x$ is the same as $BCx = \lambda Bx$ for which $BC - \lambda B$ is a positive semi-definite matrix pencil with $\lambda_0 = 0$. Therefore the results of this paper indirectly apply to such a problem $Cx = \lambda x$.

As defined by Definition 2.1, a positive semi-definite matrix pencil $A - \lambda B$ can be either regular ($\det(A - \lambda B) \not\equiv 0$ for all λ) or singular ($\det(A - \lambda B) \equiv 0$ for all λ). To cover both regular and singular pencils, as in [7] we say $\mu \neq \infty$ is a *finite eigenvalue* of $A - \lambda B$ if

$$\operatorname{rank}(A - \mu B) < \max_{\lambda\in\mathbb{C}} \operatorname{rank}(A - \lambda B), \tag{2.1}$$

and $x \in \mathbb{C}^n$ is a corresponding *eigenvector* if $0 \neq x \notin \mathcal{N}(A) \cap \mathcal{N}(B)$ satisfies

$$Ax = \mu Bx, \tag{2.2}$$

or equivalently, $0 \neq x \in \mathcal{N}(A - \mu B)\backslash(\mathcal{N}(A) \cap \mathcal{N}(B))$, where $\mathcal{N}(X) = \{x : Xx = 0\}$ denotes X's null space.

Throughout the rest of this paper, by default $A-\lambda B$ *is a positive semi-definite matrix pencil of order n, and*

$$\lambda_0 \in \mathbb{R} \textit{ such that } A - \lambda_0 B \succeq 0.$$

Let

$$n_+ := i_+(B), \quad n_- := i_-(B), \quad n_0 := n - n_+ - n_-.$$

It is known (see Lemma 3.1 below) that $A-\lambda B$ has n_++n_- finite eigenvalues all of which are real. Denote them by $\lambda_i^{\pm}$ arranged in the order:

$$\lambda_{n_-}^- \le \cdots \le \lambda_1^- \le \lambda_1^+ \le \cdots \le \lambda_{n_+}^+. \tag{2.3}$$

Also known is $\lambda_i^- \le \lambda_0 \le \lambda_j^+$ for all i and j.

Given $0 \le k_+ \le n_+$ and $0 \le k_- \le n_-$, set

$$k = k_+ + k_-, \quad J_k = \begin{bmatrix} I_{k_+} & \\ & -I_{k_-} \end{bmatrix}.$$

By convention, $J_k = I_{k_+}$ if $k_- = 0$ and likewise $J_k = -I_{k_-}$ if $k_+ = 0$.

We will state three theorems. The one in the most general form is Theorem 2.2 whose special cases for particular Φ and nonnegative integers $k_\pm$ give the other two theorems. The reason we still state them separately is because Theorem 2.1 gives the most natural extensions of the Courant-Fischer min-max principles in (1.3) in the sense that they are for individual eigenvalues and Theorem 2.3 gives the most natural extensions of Wiedlant's min-max principles in (1.2) in the sense that they are for sums of eigenvalues. Also our proof of Theorem 2.2 uses Theorem 2.1.

For the case of a regular Hermitian pencil $A-\lambda B$ (i.e., $\det(A-\lambda B) \not\equiv 0$), Theorem 2.1 is a special case of the ones considered in [3, 9]. For a diagonalizable positive semi-definite Hermitian pencil $A-\lambda B$ with nonsingular B, Theorem 2.1 was implied in [6, 16]. Recall that a positive semi-definite Hermitian pencil $A-\lambda B$ can be possibly a singular pencil; so the condition of Theorem 2.1 does not exclude a singular pencil $A-\lambda B$ which has not been considered before, not to mention that B may possibly be singular.

Theorem 2.1. *We have*[3]*, for* $1 \le i \le n_+$,

$$\lambda_i^+ = \sup_{\substack{\mathcal{X} \\ \operatorname{codim}\mathcal{X}=i-1}} \inf_{\substack{x\in\mathcal{X} \\ x^{\rm H}Bx=1}} x^{\rm H}Ax = \sup_{\substack{\mathcal{X} \\ \operatorname{codim}\mathcal{X}=i-1}} \inf_{\substack{x\in\mathcal{X} \\ x^{\rm H}Bx>0}} \frac{x^{\rm H}Ax}{x^{\rm H}Bx}, \tag{2.4a}$$

$$\lambda_i^+ = \inf_{\substack{\mathcal{X} \\ \dim\mathcal{X}=i}} \sup_{\substack{x\in\mathcal{X} \\ x^{\rm H}Bx=1}} x^{\rm H}Ax = \inf_{\substack{\mathcal{X} \\ \dim\mathcal{X}=i}} \sup_{\substack{x\in\mathcal{X} \\ x^{\rm H}Bx>0}} \frac{x^{\rm H}Ax}{x^{\rm H}Bx}, \tag{2.4b}$$

and for $1 \le i \le n_-$,

$$\lambda_i^- = -\sup_{\substack{\mathcal{X} \\ \operatorname{codim}\mathcal{X}=i-1}} \inf_{\substack{x\in\mathcal{X} \\ x^{\rm H}Bx=-1}} x^{\rm H}Ax = \inf_{\substack{\mathcal{X} \\ \operatorname{codim}\mathcal{X}=i-1}} \sup_{\substack{x\in\mathcal{X} \\ x^{\rm H}Bx<0}} \frac{x^{\rm H}Ax}{x^{\rm H}Bx}, \tag{2.4c}$$

$$\lambda_i^- = -\inf_{\substack{\mathcal{X} \\ \dim\mathcal{X}=i}} \sup_{\substack{x\in\mathcal{X} \\ x^{\rm H}Bx=-1}} x^{\rm H}Ax = \sup_{\substack{\mathcal{X} \\ \dim\mathcal{X}=i}} \inf_{\substack{x\in\mathcal{X} \\ x^{\rm H}Bx<0}} \frac{x^{\rm H}Ax}{x^{\rm H}Bx}. \tag{2.4d}$$

In particular, setting $i=1$ *in* (2.4) *gives*

$$\lambda_1^+ = \inf_{x^{\rm H}Bx>0} \frac{x^{\rm H}Ax}{x^{\rm H}Bx}, \quad \lambda_1^- = \sup_{x^{\rm H}Bx<0} \frac{x^{\rm H}Ax}{x^{\rm H}Bx}. \tag{2.5}$$

[3]It is understood that each of the first supremums/infimums in (2.4a) – (2.4d) are taken over all subspaces $\mathcal{X} \subseteq \mathbb{C}^n$ in which there is a vector x satisfying the specified constraint. This understanding will be adopted from now on for all supremums/infimums over subspaces followed by vectors in the subspaces satisfying certain constraints.

To state the next theorem, we define

$$\lambda_+ = \sup_{x^{\mathrm{H}}Bx>0} \frac{x^{\mathrm{H}}Ax}{x^{\mathrm{H}}Bx}, \quad \lambda_- = \inf_{x^{\mathrm{H}}Bx<0} \frac{x^{\mathrm{H}}Ax}{x^{\mathrm{H}}Bx},$$

and intervals

$$\mathscr{I}_+ = \begin{cases} [\lambda_0, \lambda_+], & \text{if } \lambda_+ < \infty, \\ [\lambda_0, \infty), & \text{otherwise,} \end{cases} \quad \mathscr{I}_- = \begin{cases} [-\lambda_0, -\lambda_-], & \text{if } \lambda_- > -\infty, \\ [-\lambda_0, \quad \infty), & \text{otherwise.} \end{cases} \tag{2.6}$$

For any $X \in \mathbb{C}^{n\times k}$ such that the inertia of $X^{\mathrm{H}}BX$ is $(k_+, 0, k_-)$, denote the eigenvalues of $X^{\mathrm{H}}(A-\lambda B)X$ (which is a positive semi-definite matrix pencil of order k) by $\lambda_{i,X}^{\pm}$ arranged as

$$\lambda_{k_-,X}^- \le \cdots \le \lambda_{1,X}^- \le \lambda_{1,X}^+ \le \cdots \le \lambda_{k_+,X}^+. \tag{2.7}$$

Theorem 2.2. *Let $1 \le i_1 < \cdots < i_{k_+} \le n_+$ and $1 \le j_1 < \cdots < j_{k_-} \le n_-$. Suppose that*

$$\Phi : \underbrace{\mathscr{I}_+ \times \cdots \times \mathscr{I}_+}_{k_+} \times \underbrace{\mathscr{I}_- \times \cdots \times \mathscr{I}_-}_{k_-} \to \mathbb{R} \tag{2.8}$$

is a continuous function which is non-decreasing in each of its arguments. Then

$$\sup_{\substack{\mathcal{X}_1^+\supset\cdots\supset\mathcal{X}_{k_+}^+ \\ \operatorname{codim}\mathcal{X}_p^+=i_p-1 \\ \mathcal{X}_1^-\supset\cdots\supset\mathcal{X}_{k_-}^- \\ \operatorname{codim}\mathcal{X}_q^-=j_q-1}} \inf_{\substack{x_p\in\mathcal{X}_p^+,\, y_q\in\mathcal{X}_q^- \\ X_+=[x_1,\dots,x_{k_+}] \\ X_-=[y_1,\dots,y_{k_-}] \\ X=[X_+,X_-] \\ \text{subject to (2.10)}}} \Phi(\lambda_{1,X}^+, \dots, \lambda_{k_+,X}^+, -\lambda_{1,X}^-, \dots, -\lambda_{k_-,X}^-)$$

$$= \Phi(\lambda_{i_1}^+, \dots, \lambda_{i_{k_+}}^+, -\lambda_{j_1}^-, \dots, -\lambda_{j_{k_-}}^-), \tag{2.9a}$$

$$\inf_{\substack{\mathcal{X}_1^+\subset\cdots\subset\mathcal{X}_{k_+}^+ \\ \dim\mathcal{X}_p^+=i_p \\ \mathcal{X}_1^-\subset\cdots\subset\mathcal{X}_{k_-}^- \\ \dim\mathcal{X}_q^-=j_q}} \sup_{\substack{x_p\in\mathcal{X}_p^+,\, y_q\in\mathcal{X}_q^- \\ X_+=[x_1,\dots,x_{k_+}] \\ X_-=[y_1,\dots,y_{k_-}] \\ X=[X_+,X_-] \\ \text{subject to (2.10)}}} \Phi(\lambda_{1,X}^+, \dots, \lambda_{k_+,X}^+, -\lambda_{1,X}^-, \dots, -\lambda_{k_-,X}^-)$$

$$= \Phi(\lambda_{i_1}^+, \dots, \lambda_{i_{k_+}}^+, -\lambda_{j_1}^-, \dots, -\lambda_{j_{k_-}}^-), \tag{2.9b}$$

where the constraint (2.10) *is*

$$\textit{either} \quad X^{\mathrm{H}}BX = J_k, \tag{2.10a}$$

$$\textit{or} \quad X_+^{\mathrm{H}}BX_+ = I_{k_+} \ \textit{and} \ X_-^{\mathrm{H}}BX_- = -I_{k_-}. \tag{2.10b}$$

In particular, setting $i_p = p$ and $j_q = q$ for all p and q in (2.9b) *gives*

$$\inf_{\substack{X_+=[x_1,\dots,x_{k_+}] \\ X_-=[y_1,\dots,y_{k_-}] \\ \text{subject to (2.10)}}} \Phi(\lambda_{1,X}^+, \dots, \lambda_{k_+,X}^+, -\lambda_{1,X}^-, \dots, -\lambda_{k_-,X}^-)$$

$$= \Phi(\lambda_1^+, \dots, \lambda_{k_+}^+, -\lambda_1^-, \dots, -\lambda_{k_-}^-). \tag{2.11}$$

Also equations (2.9) *and* (2.11) *remain true with all $\lambda_{i,X}^+$ replaced by λ_{i,X_+}^+ (i.e., the eigenvalues of $X_+^{\mathrm{H}}AX_+$) and all $\lambda_{j,X}^-$ replaced by λ_{j,X_-}^- (i.e., the eigenvalues of $X_-^{\mathrm{H}}AX_-$).*

REMARK **2.1.** 1. Theorem 2.2 is reminiscent of [1, Theorem 2.3] (see also [2, Exercise III.3.6 on p.68]) which is for the case $B = I$ and where a similar function Φ is defined on the product of an interval that contains the eigenvalues of A. Here, however, one ends of the intervals $\mathscr{I}_{\pm}$ may be infinite. This is because λ^+_{i,X_+} and $\lambda^+_{i,X}$ may not be bounded from above and λ^-_{i,X_-} and $\lambda^-_{i,X}$ may not be bounded from below when B is indefinite. Their unboundedness is also reflected in the Cauchy-type interlacing inequalities [7, Theorem 2.2].

2. Constraints in (2.10) implies that the inertia of $X^{\mathrm{H}}BX$ is $(k_+, 0, k_-)$. This is evident for (2.10a). Suppose (2.10b) and let $X_+^{\mathrm{H}}BX_- = U\Sigma V^{\mathrm{H}}$ be the SVD of $X_+^{\mathrm{H}}BX_-$. Then

$$\begin{bmatrix} I_{k_+} & \\ -\Sigma^{\mathrm{H}} & I_{k_-} \end{bmatrix} \begin{bmatrix} U & \\ & V \end{bmatrix} X^{\mathrm{H}}BX \begin{bmatrix} U & \\ & V \end{bmatrix}^{\mathrm{H}} \begin{bmatrix} I_{k_+} & -\Sigma \\ & I_{k_-} \end{bmatrix} = \begin{bmatrix} I_{k_+} & \\ & -I_{k_-} - \Sigma^{\mathrm{H}}\Sigma \end{bmatrix}$$

whose inertia (which is the same as that of $X^{\mathrm{H}}BX$) is $(k_+, 0, k_-)$. On the other hand, that the inertia of $X^{\mathrm{H}}BX$ is $(k_+, 0, k_-)$ does not implies (2.10), but it does implies that there is a nonsingular $Z \in \mathbb{C}^{k\times k}$ such that $\widehat{X}^{\mathrm{H}}B\widehat{X} = J_k$, where $\widehat{X} = XZ$.

3. In [1, Theorem 2.3] and [2, Exercise III.3.6 on p.68], Φ is assumed to be *permutation-invariant*. This assumption is not necessary if eigenvalues are sorted and placed in their respective argument positions.

4. Choice $k_+ = 0$ or $k_- = 0$ is allowed. These choices produce versions involving only the pos-type or neg-type eigenvalues, but not both.

5. The constraint $X^{\mathrm{H}}BX = J_k$ can be decomposed into a set of three: $X_{\pm}^{\mathrm{H}}BX_{\pm} = \pm I_{k_{\pm}}$ and $X_+^{\mathrm{H}}BX_- = 0$. Conversely the three constraints can be merged into one: $X^{\mathrm{H}}BX = J_k$. What (2.10) says that in taking the infimums in (2.9a) and (2.11) and the supremum in (2.9b) whether to enforce $X_+^{\mathrm{H}}BX_- = 0$ or not does not make a difference so long as $X_{\pm}^{\mathrm{H}}BX_{\pm} = \pm I_{k_{\pm}}$ are enforced.

Theorem 2.3. *Let* $1 \le i_1 < \cdots < i_{k_+} \le n_+$ *and* $1 \le j_1 < \cdots < j_{k_-} \le n_-$. *Then*

$$\sup_{\substack{\mathcal{X}_1^+ \supset \cdots \supset \mathcal{X}_{k_+}^+ \\ \operatorname{codim} \mathcal{X}_p^+ = i_p - 1 \\ \mathcal{X}_1^- \supset \cdots \supset \mathcal{X}_{k_-}^- \\ \operatorname{codim} \mathcal{X}_q^- = j_q - 1}} \inf_{\substack{x_p \in \mathcal{X}_p^+,\, y_q \in \mathcal{X}_q^- \\ X_+ = [x_1,\ldots,x_{k_+}] \\ X_- = [y_1,\ldots,y_{k_-}] \\ X = [X_+, X_-] \\ \text{subject to (2.10)}}} \operatorname{trace}(X^{\mathrm{H}}AX) = \sum_{p=1}^{k_+} \lambda^+_{i_p} - \sum_{q=1}^{k_-} \lambda^-_{j_q}, \tag{2.12a}$$

$$\inf_{\substack{\mathcal{X}_1^+ \subset \cdots \subset \mathcal{X}_{k_+}^+ \\ \dim \mathcal{X}_p^+ = i_p \\ \mathcal{X}_1^- \subset \cdots \subset \mathcal{X}_{k_-}^- \\ \dim \mathcal{X}_q^- = j_q}} \sup_{\substack{x_p \in \mathcal{X}_p^+,\, y_q \in \mathcal{X}_q^- \\ X_+ = [x_1,\ldots,x_{k_+}] \\ X_- = [y_1,\ldots,y_{k_-}] \\ X = [X_+, X_-] \\ \text{subject to (2.10)}}} \operatorname{trace}(X^{\mathrm{H}}AX) = \sum_{p=1}^{k_+} \lambda^+_{i_p} - \sum_{q=1}^{k_-} \lambda^-_{j_q}. \tag{2.12b}$$

In particular, setting $i_p = p$ and $j_q = q$ for all p and q in (2.12b) *gives*

$$\inf_{\substack{X_+=[x_1,\dots,x_{k_+}]\\ X_-=[y_1,\dots,y_{k_-}]\\ \text{subject to (2.10)}}} \operatorname{trace}(X^{\mathrm{H}}AX) = \sum_{i=1}^{k_+}\lambda_i^+ - \sum_{i=1}^{k_-}\lambda_i^-. \tag{2.13}$$

Proof. It is a consequence of Theorem 2.2 with

$$\Phi(\alpha_1,\dots,\alpha_{k_+},\beta_1,\dots,\beta_{k_-}) = \alpha_1+\cdots+\alpha_{k_+}+\beta_1+\cdots+\beta_{k_-} \tag{2.14}$$

and that $\operatorname{trace}(X^{\mathrm{H}}AX) = \operatorname{trace}(X_+^{\mathrm{H}}AX_+) + \operatorname{trace}(X_-^{\mathrm{H}}AX_-)$. □

Equation (2.13) was obtained in [7] under $X^{\mathrm{H}}BX = J_k$, i.e., (2.10a), and also in [5] under $X^{\mathrm{H}}BX = J_k$ and nonsingular B.

Theorem 2.3 for $k_+ \cdot k_- = 0$ (i.e., one of k_+ and k_- is zero) seems to be a special case of [10, Theorem 6] restricted to a positive semi-definite pencil, but there are many substantial differences: [10, Theorem 6] assumes that all (non-cancelled) eigenvalues are semi-simple, as well as that B must be nonsingular. This is not needed here, and in fact, not only B may be singular but also $A-\lambda B$ may be a singular pencil. Also [10, Theorem 6] does not have one like (2.12b).

3 Proofs

The following lemma, proved in [7], will play a major role in the rest of our proofs.

Lemma 3.1 ([7, Lemma 3.8])**.** *Let $A - \lambda B$ be a positive semi-definite matrix pencil of order n, and suppose that $\lambda_0 \in \mathbb{R}$ such that $A - \lambda_0 B \succeq 0$.*

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

$$W^{\mathrm{H}}AW = \begin{array}{c} \scriptstyle n_1 \\ \scriptstyle r-n_1 \\ \scriptstyle n-r \end{array}\!\!\overset{\begin{array}{ccc}\scriptstyle n_1 & \scriptstyle r-n_1 & \scriptstyle n-r\end{array}}{\begin{bmatrix} \Lambda_1 & & \\ & \Lambda_0 & \\ & & \Lambda_\infty \end{bmatrix}}, \quad W^{\mathrm{H}}BW = \begin{array}{c} \scriptstyle n_1 \\ \scriptstyle r-n_1 \\ \scriptstyle n-r \end{array}\!\!\overset{\begin{array}{ccc}\scriptstyle n_1 & \scriptstyle r-n_1 & \scriptstyle n-r\end{array}}{\begin{bmatrix} \Omega_1 & & \\ & \Omega_0 & \\ & & 0 \end{bmatrix}}, \tag{3.1}$$

where $r = \operatorname{rank}(B) = n_+ + n_-$, and

(a) *$\Lambda_1 = \operatorname{diag}(s_1\alpha_1,\dots,s_{n_1}\alpha_{n_1})$, $\Omega_1 = \operatorname{diag}(s_1,\dots,s_{n_1})$, $s_i = \pm 1$, and $\Lambda_1 - \lambda_0\Omega_1 \succ 0$,*

(b) *$\Lambda_0 = \operatorname{diag}(\Lambda_{0,1},\dots,\Lambda_{0,m+m_0})$ and $\Omega_0 = \operatorname{diag}(\Omega_{0,1},\dots,\Omega_{0,m+m_0})$ with*

$$\begin{aligned} \Lambda_{0,i} &= t_i\lambda_0, & \Omega_{0,i} &= t_i = \pm 1, & &\text{for } 1 \le i \le m, \\ \Lambda_{0,i} &= \begin{bmatrix} 0 & \lambda_0 \\ \lambda_0 & 1 \end{bmatrix}, & \Omega_{0,i} &= \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, & &\text{for } m+1 \le i \le m+m_0, \end{aligned}$$

(c) *$\Lambda_\infty = \operatorname{diag}(\alpha_{r+1},\dots,\alpha_n) \succeq 0$ with $\alpha_i \in \{1,0\}$ for $r+1 \le i \le n$.*

The representations in (3.1) *are uniquely determined by* $A-\lambda B$*, up to a simultaneous permutation of the corresponding* 1×1 *and* 2×2 *diagonal block pairs* $(s_i\alpha_i, s_i)$ *for* $1\le i\le n_1$, $(\Lambda_{0,i}, \Omega_{0,i})$ *for* $1\le i\le m+m_0$, *and* $(\alpha_i, 0)$ *for* $r+1\le i\le n$.

2. $A-\lambda B$ *has* $n_+ + n_-$ *finite eigenvalues all of which are real. Denote these finite eigenvalues by* $\lambda_i^\pm$ *and arrange them in the order of* (2.3)*:*

$$\lambda_{n_-}^- \le \cdots \le \lambda_1^- \le \lambda_1^+ \le \cdots \le \lambda_{n_+}^+. \tag{2.3}$$

Write $m=m_+ + m_-$, *where* m_+ *is the number of those* 1×1 *diagonal blocks in* Λ_0 *with* $t_i=1$ *and* m_- *is that of those with* $t_i=-1$. *The respective sources of these finite eigenvalues are*

source 1. *the* 1×1 *block pairs* $(\Lambda_{0,j}, \Omega_{0,j})$ *with* $t_j=-1$ *produce* $\lambda_i^-=\lambda_0$ *for* $1\le i\le m_-$*;*

source 2. *the* 1×1 *block pairs* $(\Lambda_{0,j}, \Omega_{0,j})$ *with* $t_j=+1$ *produce* $\lambda_i^+=\lambda_0$ *for* $1\le i\le m_+$*;*

source 3. *the* 2×2 *block pairs* $(\Lambda_{0,m+i}, \Omega_{0,m+i})$ *for* $1\le i\le m_0$ *produce* $\lambda_{m_-+i}^-=\lambda_0$ *and* $\lambda_{m_++i}^+=\lambda_0$*;*

source 4. *the diagonal matrix pair* (Λ_1, Ω_1) *produces* $\lambda_i^\pm$ *(according to* $s_j=\pm 1$*) for* $m_0+m_\pm\le i\le n_\pm$.

Each eigenvalue from sources other than **source 3** *has an eigenvector* x *that satisfies* $x^{\mathrm H}Bx=+1$ *for* λ_i^+ *and* $x^{\mathrm H}Bx=-1$ *for* λ_j^-, *while for* **source 3**, *each pair* $(\lambda_{m_-+i}^-, \lambda_{m_++i}^+)$ *of eigenvalues shares one eigenvector* x *that satisfies* $x^{\mathrm H}Bx=0$. *To be more specific than* (2.3), *we can order these finite eigenvalues as*

$$\begin{aligned}\lambda_{n_-}^- \le \cdots \le \lambda_{m_0+m_-+1}^- < \underbrace{\lambda_0=\cdots=\lambda_0}_{m_0} = \underbrace{\lambda_0=\cdots=\lambda_0}_{m_-} = \\ = \underbrace{\lambda_0=\cdots=\lambda_0}_{m_+} = \underbrace{\lambda_0=\cdots=\lambda_0}_{m_0} < \lambda_{m_0+m_++1}^+ \le \cdots \le \lambda_{n_+}^+.\end{aligned} \tag{3.2}$$

In particular $\lambda_i^-=\lambda_0$ *for* $1\le i\le m_0+m_-$ *and* $\lambda_i^+=\lambda_0$ *for* $1\le i\le m_0+m_+$.

***Proof* of Theorem 2.1**. Equations (2.4c) and (2.4d) are derivable from (2.4a) and (2.4b) applied to the matrix pencil $A-\lambda(-B)$ respectively; so we'll prove (2.4a) and (2.4b) only.

For (2.4a), first consider the case when all eigenvalues of $A-\lambda B$ are semi-simple.

Denote the right-hand side of (2.4a) by σ_i^+. From Lemma 3.1, under the semi-simple condition, there exists a nonsingular $W\in\mathbb{C}^{n\times n}$ such that

$$\begin{aligned} W^{\mathrm H}AW &= \operatorname{diag}(\lambda_1^+,\ldots,\lambda_{n_+}^+, -\lambda_1^-,\ldots,-\lambda_{n_-}^-, \lambda_1^0,\ldots,\lambda_{n_0}^0), \\ W^{\mathrm H}BW &= \operatorname{diag}(\underbrace{1,\ldots,1}_{n_+}, \underbrace{-1,\ldots,-1}_{n_-}, \underbrace{0,\ldots,0}_{n_0}), \end{aligned}$$

where $\lambda_i^\pm$ are as in (2.3) and $\lambda_i^0\ge 0$. Write

$$W=[u_1,\ldots,u_{n_+}, v_1,\ldots,v_{n_-}, w_1,\ldots,w_{n_0}]. \tag{3.3}$$

Then for any scalar λ,

$$u_k^{\mathrm{H}}(A-\lambda B)u_k = \lambda_k^+ - \lambda, \quad k=1,\dots,n_+, \tag{3.4a}$$
$$v_k^{\mathrm{H}}(A-\lambda B)v_k = \lambda - \lambda_k^-, \quad k=1,\dots,n_-, \tag{3.4b}$$
$$w_k^{\mathrm{H}}(A-\lambda B)w_k = \lambda_k^0 \ge 0, \quad k=1,\dots,n_0. \tag{3.4c}$$

Choose $\hat{\mathcal{X}} = \mathrm{span}\{u_i,\dots,u_{n_+},v_1,\dots,v_{n_-},w_1,\dots,w_{n_0}\}$ whose $\mathrm{codim}\,\hat{\mathcal{X}} = i-1$. Since $x^{\mathrm{H}}(A-\lambda_i^+B)x \ge 0$ for any $x \in \hat{\mathcal{X}}$, and $u_i^{\mathrm{H}}(A-\lambda_i^+B)u_i = 0$, we have

$$\inf_{\substack{x\in\hat{\mathcal{X}}\\ x^{\mathrm{H}}Bx=1}} x^{\mathrm{H}}Ax = \lambda_i^+ \quad\Rightarrow\quad \sigma_i^+ \ge \lambda_i^+.$$

We shall now prove $\sigma_i^+ \le \lambda_i^+$ also and thus $\sigma_i^+ = \lambda_i^+$.

Let $\mathcal{X}$ be any subspace with $\mathrm{codim}\,\mathcal{X} = i-1$. Define $\mathcal{Y} = \mathrm{span}\{u_1,\dots,u_i\}$. Then $x^{\mathrm{H}}Bx > 0$ and $x^{\mathrm{H}}(A-\lambda_i^+B)x \le 0$ for $0 \ne x \in \mathcal{Y}$. Since $\dim\mathcal{X}+\dim\mathcal{Y} = n+1$, $\mathcal{X}\cap\mathcal{Y} \ne \emptyset$. For any $y \in \mathcal{X}\cap\mathcal{Y}$ with $y^{\mathrm{H}}By = 1$, we have

$$\inf_{\substack{x\in\mathcal{X}\\ x^{\mathrm{H}}Bx=1}} x^{\mathrm{H}}Ax \le \inf_{\substack{x\in\mathcal{X}\cap\mathcal{Y}\\ x^{\mathrm{H}}Bx=1}} x^{\mathrm{H}}Ax \le y^{\mathrm{H}}Ay \le \lambda_i^+ \quad\Rightarrow\quad \sigma_i^+ \le \lambda_i^+,$$

as expected.

For the general case, by Lemma 3.1 and (3.2), there exists a nonsingular $W \in \mathbb{C}^{n\times n}$ such that

$$W^{\mathrm{H}}AW = \mathrm{diag}(\underbrace{A_0,\dots,A_0}_{m_0},A_1), \qquad W^{\mathrm{H}}BW = \mathrm{diag}(\underbrace{B_0,\dots,B_0}_{m_0},B_1),$$

where $A_0 = \begin{bmatrix} 0 & \lambda_0 \\ \lambda_0 & 1\end{bmatrix}$, $B_0 = \begin{bmatrix} 0 & 1 \\ 1 & 0\end{bmatrix}$, and

$$\begin{aligned} A_1 = \mathrm{diag}(&\lambda_1^+,\dots,\lambda_{m_+}^+,\lambda_{m_++m_0+1}^+,\dots,\lambda_{n_+}^+,\\ &-\lambda_1^-,\dots,-\lambda_{m_-}^-,-\lambda_{m_-+m_0+1}^-,\dots,-\lambda_{n_-}^-,\\ &\lambda_1^0,\dots,\lambda_{n_0}^0),\\ B_1 = \mathrm{diag}(&\underbrace{1,\dots,1}_{n_+-m_0},\underbrace{-1,\dots,-1}_{n_--m_0},\underbrace{0,\dots,0}_{n_0}), \end{aligned}$$

$\lambda_i^0 \ge 0$, and $\lambda_i^\pm$ are as in (2.3) and (3.2). Let $\varepsilon > 0$ and denote by $\lambda_i^\pm(\varepsilon)$ the finite eigenvalues of the perturbed pencil $(A+\varepsilon P)-\lambda B$, where

$$P = W^{-\mathrm{H}}\,\mathrm{diag}(\underbrace{P_0,\dots,P_0}_{m_0},0)W^{-1}, \quad P_0 = \begin{bmatrix} 1 & 0 \\ 0 & 0\end{bmatrix}. \tag{3.5}$$

It is easy to see $(A+\varepsilon P)-\lambda_0 B \succeq 0$, i.e., the perturbed pencil is still a positive semi-definite pencil, and the finite eigenvalues of $(A+\varepsilon P)-\lambda B$ in the order of (3.2) are

$$\lambda_i^+(\varepsilon) = \lambda_i^+, \qquad i=1,\dots,m_+ \text{ and } m_++m_0+1,\dots,n_+,$$

$$
\begin{aligned}
\lambda_i^+(\varepsilon) &= \lambda_0 + \sqrt{\varepsilon}, \quad & i &= m_+ + 1, \dots, m_+ + m_0, \\
\lambda_i^-(\varepsilon) &= \lambda_i^-, & i &= 1, \dots, m_- \text{ and } m_- + m_0 + 1, \dots, n_-, \\
\lambda_i^-(\varepsilon) &= \lambda_0 - \sqrt{\varepsilon}, & i &= m_- + 1, \dots, m_- + m_0.
\end{aligned}
$$

Thus as $\varepsilon \to 0^+$, $\lambda_i^+(\varepsilon) \to \lambda_i^+$ $(1 \le i \le n_+)$ and $\lambda_i^-(\varepsilon) \to \lambda_i^-$ $(1 \le i \le n_-)$. Apply the result we just proved for the semi-simple case to get

$$
\lambda_i^+(\varepsilon) = \sup_{\substack{\mathcal{X} \\ \operatorname{codim} \mathcal{X} = i-1}} \inf_{\substack{x \in \mathcal{X} \\ x^{\mathrm{H}} B x = 1}} x^{\mathrm{H}} (A + \varepsilon P) x.
$$

Now letting $\varepsilon \to 0^+$ leads to (2.4a).

For (2.4b), again the general case can be handled through the limiting argument. For the semi-simple case, a similar proof to that of (2.4a) can devised. Here is the outline. Denote the right-hand side of (2.4b) by τ_i^+. Choose $\hat{\mathcal{X}} = \operatorname{span}\{u_1, \dots, u_i\}$ to produce $\tau_i^+ \le \lambda_i^+$, and then choose $\mathcal{Y} = \operatorname{span}\{u_i, \dots, u_{n_+}, v_1, \dots, v_{n_-}, w_1, \dots, w_{n_0}\}$ to get $\tau_i^+ \ge \lambda_i^+$. □

Lemma 3.2. *Let $\mathcal{M}_1 \subset \cdots \subset \mathcal{M}_k$ and $\mathcal{N}_1 \supset \cdots \supset \mathcal{N}_k$ be subspaces of $\mathbb{C}^n$ such that $\dim \mathcal{M}_j = i_j$ and $\operatorname{codim} \mathcal{N}_j = i_j - 1$ for $j = 1, \dots, k$, where $1 \le i_1 < \cdots < i_k \le n$. Then there exist linearly independent $x_j \in \mathcal{M}_j$ and linearly independent $y_j \in \mathcal{N}_j$ such that*

$$
\operatorname{span}\{x_1, \dots, x_k\} = \operatorname{span}\{y_1, \dots, y_k\}. \tag{3.6}
$$

Moreover, if $\mathcal{M}_k$ is equipped with an inner product, then $x_1, \dots, x_k$ can be made orthonormal with respect to the inner product; likewise if $\mathcal{N}_1$ is equipped with an inner product, then $y_1, \dots, y_k$ can be made orthonormal with respect to the inner product.

Proof. By [1, Theorem 2.1] due to A. Horn, there exist linearly independent $x_j \in \mathcal{M}_j$ and linearly independent $y_j \in \mathcal{N}_j$ such that (3.6) holds. Finally if either $\mathcal{M}_k$ or $\mathcal{N}_1$ or both are equipped with certain inner products, we can use the Gram-Schmidt orthogonalization process on x_i from x_1 to x_k with respect to the given inner product on $\mathcal{M}_k$, and/or on y_i backward from y_k to y_1 with respect to the given inner product on $\mathcal{N}_1$ to see that $x_1, \dots, x_k$ and/or $y_1, \dots, y_k$ can be made orthonormal with respect to corresponding inner products. □

Remark **3.1.** Lemma 3.2 subtly differs from [1, Corollary 2.2] in that here we allow $\mathcal{M}_k$ and $\mathcal{N}_1$ to be equipped with different inner products or only one of them is equipped with an inner product but the other one isn't. In using Lemma 3.2 later, interesting inner products on $\mathcal{M}_k$ or $\mathcal{N}_1$ are

1. $\langle x, y \rangle = \ \ y^{\mathrm{H}} B x$, provided $x^{\mathrm{H}} B x > 0$ for any $0 \ne x \in \mathcal{M}_k$ or $\mathcal{N}_1$;
2. $\langle x, y \rangle = -y^{\mathrm{H}} B x$, provided $x^{\mathrm{H}} B x < 0$ for any $0 \ne x \in \mathcal{M}_k$ or $\mathcal{N}_1$.

When either inner product is definable, we call it *B-inner product* and say any set of orthonormal vectors with respect to the inner product *B-orthonormal* vectors.

Proof **of Theorem 2.2**. We will first prove the last statement of the theorem: (2.9) and (2.11) with all $\lambda_{i,X}^+$ replaced by λ_{i,X_+}^+ and all $\lambda_{j,X}^-$ by λ_{j,X_-}^-.

For (2.9a), first consider the case when all of the finite eigenvalues of $A - \lambda B$ are semi-simple. Adopt the notations in the proof of Theorem 2.1, and also introduce, for any given $Z = [z_1, z_2, \dots, z_m]$,

$$\mathcal{T}_{j,Z} := \operatorname{span}\{z_j, \dots, z_m\},\ \mathcal{S}_{j,Z} := \operatorname{span}\{z_1, \dots, z_j\} \quad \text{for } j = 1, 2, \dots, m,$$

and in particular, write $\mathcal{T}_Z = \mathcal{T}_{1,Z}$ and $\mathcal{S}_Z = \mathcal{S}_{m,Z}$ (thus $\mathcal{T}_Z = \mathcal{S}_Z$).

Choose

$$\begin{aligned}
\hat{\mathcal{X}}_p^+ &= \operatorname{span}\{u_{i_p}, \dots, u_{n_+}, v_1, \dots, v_{n_-}, w_1, \dots, w_{n_0}\},\\
\hat{\mathcal{X}}_q^- &= \operatorname{span}\{v_{j_q}, \dots, v_{n_-}, u_1, \dots, u_{n_+}, w_1, \dots, w_{n_0}\}.
\end{aligned}$$

Evidently, $\hat{\mathcal{X}}_1^+ \supset \cdots \supset \hat{\mathcal{X}}_{k_+}^+, \hat{\mathcal{X}}_1^- \supset \cdots \supset \hat{\mathcal{X}}_{k_-}^-$ and $\operatorname{codim} \hat{\mathcal{X}}_p^+ = i_p - 1, \operatorname{codim} \hat{\mathcal{X}}_q^- = j_q - 1$. By (3.4), $x_p^{\mathrm{H}}(A - \lambda_{i_p}^+ B)x_p \geq 0$ for any $x_p \in \hat{\mathcal{X}}_p^+$, and $u_{i_p}^{\mathrm{H}}(A - \lambda_{i_p}^+ B)u_{i_p} = 0$. Similarly, $y_q^{\mathrm{H}}(A - \lambda_{j_q}^- B)y_q \geq 0$ for any $y_q \in \hat{\mathcal{X}}_q^-$, and $v_{j_q}^{\mathrm{H}}(A - \lambda_{j_q}^- B)v_{j_q} = 0$. Therefore

$$\min_{\substack{x_p \in \hat{\mathcal{X}}_p^+ \\ x_p^{\mathrm{H}} B x_p = 1}} x_p^{\mathrm{H}} A x_p = \lambda_{i_p}^+, \qquad \min_{\substack{y_q \in \hat{\mathcal{X}}_q^- \\ y_q^{\mathrm{H}} B y_q = -1}} y_q^{\mathrm{H}} A y_q = -\lambda_{j_q}^-.$$

For any $X_+ = [x_1, \dots, x_{k_+}]$ with $x_p \in \hat{\mathcal{X}}_p^+$ for $p = 1, \dots, k_+$ and $X_- = [y_1, \dots, y_{k_-}]$ with $y_q \in \hat{\mathcal{X}}_q^-$ for $q = 1, \dots, k_-$ and $X = [X_+, X_-]$ subject to (2.10). $X_+^{\mathrm{H}}(A - \lambda B)X_+$ and $X_-^{\mathrm{H}}(A - \lambda B)X_-$ are positive semi-definite pencils. Let their eigenvalues be

$$\lambda_{1,X_+}^+ \leq \cdots \leq \lambda_{k_+,X_+}^+, \quad \lambda_{k_-,X_-}^- \leq \cdots \leq \lambda_{1,X_-}^-,$$

respectively. By Theorem 2.1 and noticing $\mathcal{T}_{p,X_+} \subset \hat{\mathcal{X}}_p^+$ and $\mathcal{T}_{q,X_-} \subset \hat{\mathcal{X}}_q^-$, we have[4], for $p = 1, \dots, k_+$ and $q = 1, \dots, k_-$,

$$\begin{aligned}
\lambda_{p,X_+}^+ &= \sup_{\substack{\mathcal{X} \subset \mathcal{T}_{X_+} \\ \dim \mathcal{X} = k_+ - p + 1}} \inf_{\substack{x \in \mathcal{X} \\ x^{\mathrm{H}} B x = 1}} x^{\mathrm{H}} A x \\
&\geq \inf_{\substack{x \in \mathcal{T}_{p,X_+} \\ x^{\mathrm{H}} B x = 1}} x^{\mathrm{H}} A x \\
&\geq \inf_{\substack{x \in \hat{\mathcal{X}}_p^+ \\ x^{\mathrm{H}} B x = 1}} x^{\mathrm{H}} A x = \lambda_{i_p}^+, && \text{(3.7a)} \\
-\lambda_{q,X_-}^- &= \sup_{\substack{\mathcal{X} \subset \mathcal{T}_{X_-} \\ \dim \mathcal{X} = k_- - q + 1}} \inf_{\substack{y \in \mathcal{X} \\ y^{\mathrm{H}} B y = -1}} y^{\mathrm{H}} A y \\
&\geq \inf_{\substack{y \in \mathcal{T}_{q,X_-} \\ y^{\mathrm{H}} B y = -1}} y^{\mathrm{H}} A y \\
&\geq \inf_{\substack{y \in \hat{\mathcal{X}}_q^- \\ y^{\mathrm{H}} B y = -1}} y^{\mathrm{H}} A y = -\lambda_{j_q}^-. && \text{(3.7b)}
\end{aligned}$$

[4]Both (3.7) and (3.10) below also follow from [7, Theorem 2.1]. But a (different and simpler) proof is presented here for completeness.

Since $\Phi(\,\cdot\,)$ is non-decreasing in each of its arguments,

$$\Phi(\lambda^+_{1,X_+},\dots,\lambda^+_{k_+,X_+},-\lambda^-_{1,X_-},\dots,-\lambda^-_{k_-,X_-}) \ge \Phi(\lambda^+_{i_1},\dots,\lambda^+_{i_{k_+}},-\lambda^-_{j_1},\dots,\lambda^-_{j_{k_-}}). \tag{3.8}$$

Since $X=[X_+,X_-]$ is arbitrary, we have

$$\inf_{\substack{x_p\in\hat{\mathcal{X}}^+_p,\,y_q\in\hat{\mathcal{X}}^-_q\\ X_+=[x_1,\dots,x_{k_+}]\\ X_-=[y_1,\dots,y_{k_-}]\\ X=[X_+,X_-]\\ \text{subject to (2.10)}}} \Phi(\lambda^+_{1,X_+},\dots,\lambda^+_{k_+,X_+},-\lambda^-_{1,X_-},\dots,-\lambda^-_{k_-,X_-})$$
$$\ge \Phi(\lambda^+_{i_1},\dots,\lambda^+_{i_{k_+}},-\lambda^-_{j_1},\dots,-\lambda^-_{j_{k_-}}),$$

and thus

$$\sup_{\substack{\mathcal{X}^+_1\supset\cdots\supset\mathcal{X}^+_{k_+}\\ \operatorname{codim}\mathcal{X}^+_p=i_p-1\\ \mathcal{X}^-_1\supset\cdots\supset\mathcal{X}^-_{k_-}\\ \operatorname{codim}\mathcal{X}^-_q=j_q-1}} \inf_{\substack{x_p\in\mathcal{X}^+_p,\,y_q\in\mathcal{X}^-_q\\ X_+=[x_1,\dots,x_{k_+}]\\ X_-=[y_1,\dots,y_{k_-}]\\ X=[X_+,X_-]\\ \text{subject to (2.10)}}} \Phi(\lambda^+_{1,X_+},\dots,\lambda^+_{k_+,X_+},-\lambda^-_{1,X_-},\dots,-\lambda^-_{k_-,X_-})$$
$$\ge \Phi(\lambda^+_{i_1},\dots,\lambda^+_{i_{k_+}},-\lambda^-_{j_1},\dots,-\lambda^-_{j_{k_-}}). \tag{3.9}$$

On the other hand, let $\mathcal{X}^+_p$ for $p=1,\dots,k_+$ and $\mathcal{X}^-_q$ for $q=1,\dots,k_-$ be any subspaces that satisfy the assumptions. Define $\mathcal{Y}^+_p=\operatorname{span}\{u_1,\dots,u_{i_p}\}$ and $\mathcal{Y}^-_q=\operatorname{span}\{v_1,\dots,v_{j_q}\}$. Obviously $\dim\mathcal{Y}^+_p=i_p$ and $\dim\mathcal{Y}^-_q=j_q$. By Lemma 3.2, there exists B-orthonormal sets $\{s_1,\dots,s_{k_+}\}$ and $\{t_1,\dots,t_{k_-}\}$, where $s_p\in\mathcal{Y}^+_p$ and $t_q\in\mathcal{Y}^-_q$, such that

$$\mathcal{T}_{X_+}=\operatorname{span}\{x_1,\dots,x_{k_+}\}=\operatorname{span}\{s_1,\dots,s_{k_+}\}=\mathcal{S}_{Y_+},$$
$$\mathcal{T}_{X_-}=\operatorname{span}\{y_1,\dots,y_{k_-}\}=\operatorname{span}\{t_1,\dots,t_{k_-}\}=\mathcal{S}_{Y_-},$$

where $x_p\in\mathcal{X}^+_p$ for $p=1,\dots,k_+$ and $y_q\in\mathcal{X}^-_q$ for $q=1,\dots,k_-$, and $Y_+=[s_1,\dots,s_{k_+}]$ and $Y_-=[t_1,\dots,t_{k_-}]$ satisfy $Y^{\mathrm{H}}_\pm BY_\pm=\pm I_{k_\pm}$ and $Y^{\mathrm{H}}_+BY_-=0$. Let $X=[X_+,X_-]$, $X_+=[x_1,\dots,x_{k_+}]$, and $X_-=[y_1,\dots,y_{k_-}]$. Then $X^{\mathrm{H}}_+BX_-=0$. We may also assume $X^{\mathrm{H}}_\pm BX_\pm=\pm I_{k_\pm}$. Otherwise since $\mathcal{T}_{X_+}=\mathcal{S}_{Y_+}$ and $Y^{\mathrm{H}}_+BY_+=I_{k_+}$, we have $x^{\mathrm{H}}Bx>0$ for any $0\ne x\in\mathcal{T}_{X_+}$ and thus B induces a B-inner product in $\mathcal{T}_{X_+}$. Now perform the Gram-Schmidt orthogonalization process on x_p from x_{k_+} to x_1 to lead to k_+ new x_p with the desired property.

$Y^{\mathrm{H}}_+(A-\lambda B)Y_+$ and $Y^{\mathrm{H}}_-(A-\lambda B)Y_-$ which are the projections of $A-\lambda B$ onto $\mathcal{S}_{Y_+}$ and $\mathcal{S}_{Y_-}$, respectively, are positive semi-definite pencils. In fact, $s^{\mathrm{H}}Bs>0$ for all $0\ne s\in\mathcal{S}_{Y_+}$, and $t^{\mathrm{H}}Bt<0$ for all $0\ne t\in\mathcal{S}_{Y_-}$. Let the eigenvalues of the projections be

$$\lambda^+_{1,Y_+}\le\cdots\le\lambda^+_{k_+,Y_+},\quad \lambda^-_{k_-,Y_-}\le\cdots\le\lambda^-_{1,Y_-}.$$

Since $\mathcal{T}_{X_+}=\mathcal{S}_{Y_+}$ and $\mathcal{T}_{X_-}=\mathcal{S}_{Y_-}$, we have

$$\lambda^+_{p,Y_+}=\lambda^+_{p,X_+}\text{ for }p=1,\dots,k_+\text{ and }\lambda^-_{q,Y_-}=\lambda^-_{q,X_-}\text{ for }q=1,\dots,k_-,$$

where $\lambda^+_{1,X_+}\le\cdots\le\lambda^+_{k_+,X_+}$ and $\lambda^-_{k_-,X_-}\le\cdots\le\lambda^-_{1,X_-}$ are the eigenvalues of $X^{\mathrm{H}}_+(A-\lambda B)X_+$ and those of $X^{\mathrm{H}}_-(A-\lambda B)X_-$, respectively. By (3.4), $s^{\mathrm{H}}_p(A-\lambda^+_{i_p}B)s_p\le 0$ for any

$s_p \in \mathcal{Y}_p^+$ and $u_{i_p}^{\rm H}(A-\lambda_{i_p}^+ B)u_{i_p}=0$, and similarly, $t_q^{\rm H}(A-\lambda_{j_q}^- B)t_q \le 0$ for any $t_q \in \mathcal{Y}_q^-$ and $v_{j_q}^{\rm H}(A-\lambda_{j_q}^- B)v_{j_q}=0$. Therefore

$$\max_{\substack{s_p\in\mathcal{Y}_p^+\\ s_p^{\rm H}Bs_p=1}} s_p^{\rm H}As_p=\lambda_{i_p}^+, \qquad \max_{\substack{t_q\in\mathcal{Y}_q^-\\ t_q^{\rm H}Bt_q=-1}} t_q^{\rm H}At_q=-\lambda_{j_q}^-.$$

By Theorem 2.1 and noticing $\mathcal{S}_{p,Y_+}\subseteq\mathcal{Y}_p^+$ and $\mathcal{S}_{q,Y_-}\subseteq\mathcal{Y}_q^-$, we have, for $p=1,\dots,k_+$ and $q=1,\dots,k_-$,

$$\begin{aligned}
\lambda_{p,X_+}^+ = \lambda_{p,Y_-}^+ &= \inf_{\substack{\mathcal{Y}\subset\mathcal{S}_{Y_+}\\ \dim\mathcal{Y}=p}} \sup_{\substack{s\in\mathcal{Y}\\ s^{\rm H}Bs=1}} s^{\rm H}As \\
&\le \sup_{\substack{s\in\mathcal{S}_{p,Y_+}\\ s^{\rm H}Bs=1}} s^{\rm H}As \\
&\le \sup_{\substack{s\in\mathcal{Y}_p^+\\ s^{\rm H}Bs=1}} s^{\rm H}As=\lambda_{i_p}^+, \qquad (3.10a)\\
-\lambda_{q,X_-}^- = -\lambda_{q,Y_-}^- &= \inf_{\substack{\mathcal{Y}\subset\mathcal{S}_{Y_-}\\ \dim\mathcal{Y}=q}} \sup_{\substack{t\in\mathcal{Y}\\ t^{\rm H}Bt=-1}} t^{\rm H}At \\
&\le \sup_{\substack{t\in\mathcal{S}_{q,Y_-}\\ t^{\rm H}Bt=-1}} t^{\rm H}At \\
&\le \sup_{\substack{t\in\mathcal{Y}_q^-\\ t^{\rm H}Bt=-1}} t^{\rm H}At=-\lambda_{j_q}^-. \qquad (3.10b)
\end{aligned}$$

Since $\Phi(\,\cdot\,)$ is non-decreasing in each of its arguments,

$$\Phi(\lambda_{1,X_+}^+,\dots,\lambda_{k_+,X_+}^+,-\lambda_{1,X_-}^-,\dots,-\lambda_{k_-,X_-}^-) \le \Phi(\lambda_{i_1}^+,\dots,\lambda_{i_{k_+}}^+,-\lambda_{j_1}^-,\dots,\lambda_{j_{k_-}}^-) \qquad (3.11)$$

which yields

$$\inf_{\substack{x_p\in\mathcal{X}_p^+,\,y_q\in\mathcal{X}_q^-\\ X_+=[x_1,\dots,x_{k_+}]\\ X_-=[y_1,\dots,y_{k_-}]\\ X=[X_+,X_-]\\ \text{subject to (2.10a)}}} \Phi(\lambda_{1,X_+}^+,\dots,\lambda_{k_+,X_+}^+,-\lambda_{1,X_-}^-,\dots,-\lambda_{k_-,X_-}^-)$$
$$\le \Phi(\lambda_{i_1}^+,\dots,\lambda_{i_{k_+}}^+,-\lambda_{j_1}^-,\dots,-\lambda_{j_{k_-}}^-)$$

which is also valid with (2.10a) replaced by (2.10b) because (2.10a) implies (2.10b). Since $\mathcal{X}_p^+$ and $\mathcal{X}_q^-$ are arbitrary, we have

$$\sup_{\substack{\mathcal{X}_1^+\supset\cdots\supset\mathcal{X}_{k_+}^+\\ \operatorname{codim}\mathcal{X}_p^+=i_p-1\\ \mathcal{X}_1^-\supset\cdots\supset\mathcal{X}_{k_-}^-\\ \operatorname{codim}\mathcal{X}_q^-=j_q-1}} \inf_{\substack{x_p\in\mathcal{X}_p^+,\,y_q\in\mathcal{X}_q^-\\ X_+=[x_1,\dots,x_{k_+}]\\ X_-=[y_1,\dots,y_{k_-}]\\ X=[X_+,X_-]\\ \text{subject to (2.10)}}} \Phi(\lambda_{1,X_+}^+,\dots,\lambda_{k_+,X_+}^+,-\lambda_{1,X_-}^-,\dots,-\lambda_{k_-,X_-}^-)$$

$$\leq \Phi(\lambda_{i_1}^+, \ldots, \lambda_{i_{k_+}}^+, -\lambda_{j_1}^-, \ldots, -\lambda_{j_{k_-}}^-). \quad (3.12)$$

Combine (3.9) and (3.12) to get (2.9a).

For the general case, let $\varepsilon > 0$ and denote by $\lambda_i^{\pm}(\varepsilon)$ the finite eigenvalues of the perturbed pencil $(A+\varepsilon P)-\lambda B$, where P is defined by (3.5) in the proof of Theorem 2.1. Since all of the finite eigenvalues of the perturbed pencil are semi-simple,

$$\sup_{\substack{\mathcal{X}_1^+\supset\cdots\supset\mathcal{X}_{k_+}^+ \\ \operatorname{codim}\mathcal{X}_p^+=i_p-1 \\ \mathcal{X}_1^-\supset\cdots\supset\mathcal{X}_{k_-}^- \\ \operatorname{codim}\mathcal{X}_q^-=j_q-1}} \inf_{\substack{x_p\in\mathcal{X}_p^+,\, y_q\in\mathcal{X}_q^- \\ X_+=[x_1,\ldots,x_{k_+}] \\ X_-=[y_1,\ldots,y_{k_-}] \\ X=[X_+,X_-] \\ \text{subject to (2.10)}}} \Phi(\lambda_{1,X_+}^+(\varepsilon), \ldots, \lambda_{k_+,X_+}^+(\varepsilon), -\lambda_{1,X_-}^-(\varepsilon), \ldots, -\lambda_{k_-,X_-}^-(\varepsilon))$$

$$= \Phi(\lambda_{i_1}^+(\varepsilon), \ldots, \lambda_{i_{k_+}}^+(\varepsilon), -\lambda_{j_1}^-(\varepsilon), \ldots, -\lambda_{j_{k_-}}^-(\varepsilon)).$$

Now let $\varepsilon \to 0^+$ to get (2.9a).

For (2.9b), the difficult part is also about the semi-simple case since the general case can be handled by the limiting argument. A proof similar to what we did above for (2.9a) still works: choosing $\hat{\mathcal{X}}_p^+ = \operatorname{span}\{u_1, \ldots, u_{i_p}\}$ and $\hat{\mathcal{X}}_p^- = \operatorname{span}\{v_1, \ldots, v_{j_q}\}$ will lead to that the left-hand side is no bigger than its right-hand side, and choosing

$$\mathcal{Y}_p^+ = \operatorname{span}\{u_{i_p}, \ldots, u_{n_+}, v_1, \ldots, v_{n_-}, w_1, \ldots, w_{n_0}\},$$
$$\mathcal{Y}_q^- = \operatorname{span}\{v_{j_q}, \ldots, v_{n_-}, u_1, \ldots, u_{n_+}, w_1, \ldots, w_{n_0}\}$$

will give the opposite.

Now we return to (2.9) and (2.11) as they are by pointing out how to modify the above arguments. For the paragraph leading to (3.9) whose validity is due to (3.7), we need to establish a similar equation to (3.9) for the current purpose. To that end, we need something similar to (3.7). This is (3.13) below. With the same X_+ and X_- there, noting that $X^{\mathrm{H}}(A-\lambda B)X$ and $[X_-, X_+]^{\mathrm{H}}(A-\lambda B)[X_-, X_+]$ have the same finite eigenvalues, we have

$$\begin{aligned}
\lambda_{p,X}^+ = \lambda_{p,[X_-,X_+]}^+ &= \sup_{\substack{\mathcal{X}\subset\mathcal{T}_{[X_-,X_+]} \\ \dim\mathcal{X}=k-p+1}} \inf_{\substack{x\in\mathcal{X} \\ x^{\mathrm{H}}Bx=1}} x^{\mathrm{H}}Ax \\
&\geq \inf_{\substack{x\in\mathcal{T}_{k_-+p,[X_-,X_+]} \\ x^{\mathrm{H}}Bx=1}} x^{\mathrm{H}}Ax \\
&\geq \inf_{\substack{x\in\hat{\mathcal{X}}_p^+ \\ x^{\mathrm{H}}Bx=1}} x^{\mathrm{H}}Ax = \lambda_{i_p}^+, &\text{(3.13a)} \\
-\lambda_{q,X}^- &= \sup_{\substack{\mathcal{X}\subset\mathcal{T}_X \\ \dim\mathcal{X}=k-q+1}} \inf_{\substack{y\in\mathcal{X} \\ y^{\mathrm{H}}By=-1}} y^{\mathrm{H}}Ay \\
&\geq \inf_{\substack{y\in\mathcal{T}_{k_++q,X} \\ y^{\mathrm{H}}By=-1}} y^{\mathrm{H}}Ay \\
&\geq \inf_{\substack{y\in\hat{\mathcal{X}}_q^- \\ y^{\mathrm{H}}By=-1}} y^{\mathrm{H}}Ay = -\lambda_{j_q}^-. &\text{(3.13b)}
\end{aligned}$$

For the paragraph leading to (3.12) whose validity is due to (3.10), we need to establish a similar equation to (3.12) for the current purpose, too. To this end, we need something similar to (3.10). This is (3.14) below. With the same $X_\pm$, $Y_\pm$, $X = [X_+, X_-]$, and $Y = [Y_+, Y_-]$ there, we have

$$\begin{aligned}
\lambda^+_{p,X} = \lambda^+_{p,Y} = \lambda^+_{p,[Y_-,Y_+]} &= \inf_{\substack{\mathcal{Y}\subset\mathcal{S}_{[Y_-,Y_+]}\\ \dim\mathcal{Y}=p}} \sup_{\substack{s\in\mathcal{Y}\\ s^{\rm H}Bs=1}} s^{\rm H}As \\
&\le \sup_{\substack{s\in\mathcal{S}_{k_-+p,[Y_-,Y_+]}\\ s^{\rm H}Bs=1}} s^{\rm H}As \\
&\le \sup_{\substack{s\in\mathcal{Y}_p^+\\ s^{\rm H}Bs=1}} s^{\rm H}As = \lambda^+_{i_p}, \qquad (3.14a)\\
-\lambda^-_{q,X} = -\lambda^-_{q,Y} &= \inf_{\substack{\mathcal{Y}\subset\mathcal{S}_Y\\ \dim\mathcal{Y}=q}} \sup_{\substack{t\in\mathcal{Y}\\ t^{\rm H}Bt=-1}} t^{\rm H}At \\
&\le \sup_{\substack{t\in\mathcal{S}_{k_++q,Y}\\ t^{\rm H}Bt=-1}} t^{\rm H}At \\
&\le \sup_{\substack{t\in\mathcal{Y}_q^-\\ t^{\rm H}Bt=-1}} t^{\rm H}At = -\lambda^-_{j_q}. \qquad (3.14b)
\end{aligned}$$

The rest is straightforward. □

4 Concluding Remarks

We have obtained various extensions of the well-known Wielandt's min-max principles [15], including the Courant-Fischer min-max principles, for positive semi-definite matrix pencils. Similar attempts were made previously in [10] for general Hermitian matrix pencils, but the positive semi-definite matrix pencils of our interest are not a subset of the matrix pencils studied in [10]. Also our extensions in the most general form are stated in terms of any continuous multi-variable function that is non-decreasing in each of its arguments.

We focus on positive semi-definite matrix pencils because extensions for such matrix pencils are more mathematically elegant (and thus perhaps more natural). Unlike in [10], here there are no complex eigenvalues, as well as the so-called *cancelled* eigenvalues introduced in [3, 4], to exclude.

Opposite to the concept of a *positive semi-definite matrix pencil*, naturally, is that of a *negative semi-definite matrix pencil* $A - \lambda B$ by which we mean that A and B are Hermitian and there is a real λ_0 such that $A - \lambda_0 B$ is negative semi-definite. Evidently, if $A - \lambda B$ is a negative semi-definite matrix pencil, then $-(A - \lambda B) = (-A) - \lambda(-B)$ is a positive semi-definite matrix pencil because $(-A) - \lambda_0(-B) = -(A - \lambda_0 B)$. Applying what we have developed so far for positive semi-definite matrix pencils leads to various similar results for negative semi-definite matrix pencils. Detail is omitted.

In general, the roles of A and B in a positive semi-definite matrix pencil $A - \lambda B$ with an indefinite B are not interchangeable in the sense that the semi-definiteness of $A - \lambda B$

doesn't implies that of $B - \lambda A$. Suppose $A - \lambda B$ is a positive semi-definite matrix pencil and B is indefinite. We claim that

1. *if the only* $\lambda_0 \in \mathbb{R}$ *that makes* $A - \lambda_0 B \succeq 0$ *is* $\lambda_0 = 0$*, then* $B - \lambda A$ *is neither a positive nor negative semi-definite matrix pencil*;

2. $B - \lambda A$ *is a positive or negative semi-definite matrix pencil if and only if there is a* nonzero $\lambda_0 \in \mathbb{R}$ *such that* $A - \lambda_0 B \succeq 0$.

To see item 1, we notice $\lambda_1^- = \lambda_1^+ = 0$ by Lemma 3.1. Suppose, to the contrary, that $B - \lambda A$ is a positive (negative) semi-definite matrix pencil and then there exists $\mu_0 \in \mathbb{R}$ such that $B - \mu_0 A$ is positive (negative) semi-definite. First this $\mu_0 \neq 0$ since B is indefinite by assumption, and so

$$B - \mu_0 A = -\mu_0[A - (1/\mu_0)B]$$

which implies $A - (1/\mu_0)B$ is positive or negative definite, depending on the sign of μ_0. By Lemma 3.1, $\lambda_1^- \leq 1/\mu_0 \leq \lambda_1^+$, implying $1/\mu_0 = 0$. That's not possible since μ_0 must be finite. For item 2, the necessity of $A - \lambda_0 B \succeq 0$ for some $0 \neq \lambda_0 \in \mathbb{R}$ is implied by item 1. For sufficiency, we notice that if $\lambda_0 \neq 0$, then

$$B - (1/\lambda_0)A = -(1/\lambda_0)(A - \lambda_0 B)$$

is positive semi-definite if $\lambda_0 < 0$ or negative semi-definite if $\lambda_0 > 0$.

References

[1] A.R. Amir-Moéz. Extreme properties of eigenvalues of a Hermitian transformation and singular values of the sum and product of linear transformations. *Duke Math. J.*, 23:463–476, 1956.

[2] R. Bhatia. *Matrix Analysis.* Graduate Texts in Mathematics, vol. 169. Springer, New York, 1996.

[3] Paul Binding, Branko Najman, and Qiang Ye. A variational principle for eigenvalues of pencils of Hermitian matrices. *Integr. Eq. Oper. Theory*, 35:398–422, 1999.

[4] Paul Binding and Qiang Ye. Variational principles for indefinite eigenvalue problems. *Linear Algebra Appl.*, 218:251–262, 1995.

[5] J. Kovač-Striko and K. Veselić. Trace minimization and definiteness of symmetric pencils. *Linear Algebra Appl.*, 216:139–158, 1995.

[6] P. Lancaster and Q. Ye. Variational properties and Rayleigh quotient algorithms for symmetric matrix pencils. *Oper. Theory: Adv. Appl.*, 40:247–278, 1989.

[7] Xin Liang, Ren-Cang Li, and Zhaojun Bai. Trace minimization principles for positive semi-definite pencils. *Linear Algebra Appl.*, 438:3085–3106, 2013.

[8] B. Najman and Q. Ye. A minimax characterization of eigenvalues of Hermitian pencils. *Linear Algebra Appl.*, 144:217–230, 1991.

[9] B. Najman and Q. Ye. A minimax characterization of eigenvalues of Hermitian pencils II. *Linear Algebra Appl.*, 191:183–197, 1993.

[10] I. Nakić and K. Veselić. Wielandt and Ky-Fan theorem for matrix pairs. *Linear Algebra Appl.*, 369:77–93, 2003.

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

[12] R. S. Phillips. A minimax characterization for the eigenvalues of a positive symmetric operator in a space with an indefinite metric. *J. Fac. Sci. Univ. Tokyo Sect. I*, 17:51–59, 1970.

[13] G. W. Stewart and Ji-Guang Sun. *Matrix Perturbation Theory.* Academic Press, Boston, 1990.

[14] B. Textorius. Minimaxprinzipe zur Bestimmung der Eigenwerte J-nichtnegativer Operatoren. *Math. Scand.*, 35:105–114, 1974.

[15] H. Wielandt. An extremum property of sums of eigenvalues. *Proc. Amer. Math. Soc.*, 6:106–110, 1955.

[16] Qiang Ye. *Variational principles and numerical algorithms for symmetric matrix pencils.* PhD thesis, University of Calgary, Calgary, Canada, 1989.