Linear Algebra and its Applications 687 (2024) 8–37

Contents lists available at ScienceDirect

Linear Algebra and its Applications

journal homepage: www.elsevier.com/locate/laa

On generalizing trace minimization principles, II

Xin Liang [a,b,*], Ren-Cang Li [c]

[a] *Yau Mathematical Sciences Center, Tsinghua University, Beijing 100084, China*
[b] *Yanqi Lake Beijing Institute of Mathematical Sciences and Applications, Beijing 101408, China*
[c] *Department of Mathematics, University of Texas at Arlington, Arlington, TX 76019-0408, USA*

A R T I C L E I N F O

Article history:
Received 25 March 2023
Received in revised form 21 January 2024
Accepted 26 January 2024
Available online 1 February 2024
Submitted by V. Mehrmann

MSC:
15A18
15A22
15A42

Keywords:
Trace minimization
Positive semidefinite matrix pair
Hermitian matrix pair
Eigenvalue

A B S T R A C T

This paper is concerned with establishing a trace minimization principle for two Hermitian matrix pairs. Specifically, we will answer the question: when is $\inf_X \operatorname{tr}(\widehat{A}X^{\mathrm{H}}AX)$ subject to $\widehat{B}X^{\mathrm{H}}BX = I$ (the identity matrix of apt size) finite? Sufficient and necessary conditions are obtained and, when the infimum is finite, an explicit formula for it is established in terms of the finite eigenvalues of the matrix pairs. Our results extend Fan's trace minimization principle (1949) for a Hermitian matrix, a minimization principle of Kovač-Striko and Veselić (1995) for a Hermitian matrix pair, and most recent ones by the authors and their collaborators for a Hermitian matrix pair and a Hermitian matrix.

* Corresponding author at: Yau Mathematical Sciences Center, Tsinghua University, Beijing 100084, China.

E-mail addresses: liangxinslm@tsinghua.edu.cn (X. Liang), rcli@uta.edu (R.-C. Li).

https://doi.org/10.1016/j.laa.2024.01.015

1. Introduction

Various trace minimization principles have served as the theoretical foundations for computing eigenvalues of special kinds of matrix pairs and played important roles in numerical linear algebra [1–11]. Fan's trace minimization principle [12] [13, p.248] is perhaps the earliest and the most well-known one:

$$\min_{X^{\mathrm{H}}X=I_k} \operatorname{tr}(X^{\mathrm{H}}AX) = \sum_{i=1}^{k} \lambda_i,$$

where $A \in \mathbb{C}^{n\times n}$ is Hermitian with its eigenvalues denoted by $\lambda_1 \le \lambda_2 \le \cdots \le \lambda_n$, $\operatorname{tr}(\,\cdot\,)$ takes the trace of a matrix, and I_k is the $k\times k$ identity matrix. Moreover, any minimizer $X_{\min}$ is an orthonormal basis matrix of the invariant subspace of A associated with its eigenvalues $\lambda_1,\dots,\lambda_k$. It has since been generalized to many broader cases:

1. The most straightforward generalization is $\min\limits_{X^{\mathrm{H}}BX=I_k} \operatorname{tr}(X^{\mathrm{H}}AX)$ for a Hermitian matrix pair (A, B), where B is positive definite (see, e.g., [10]).
2. For a Hermitian matrix pair (A, B) with indefinite and possibly singular B, in [14–16] $\inf\limits_{X^{\mathrm{H}}BX=J_k} \operatorname{tr}(X^{\mathrm{H}}AX)$ is investigated, where J_k is a $k\times k$ diagonal matrix with diagonal entries ±1. It is shown that the infimum is finite if and only if (A, B) is a *positive semidefinite matrix pair*, by which we mean there exists $\lambda_0 \in \mathbb{R}$ such that $A - \lambda_0 B$ is positive semidefinite.
3. From the perspective of optimization, in [17] $\min\limits_{X^{\mathrm{H}}X=I_k} \operatorname{tr}(\widehat{A}X^{\mathrm{H}}AX)$ is analyzed, where both A and $\widehat{A}$ are Hermitian matrices.
4. More recently, the authors of [18] investigated two more general cases:

   (a) $\min\limits_{X^{\mathrm{H}}BX=I_k} \operatorname{tr}(\widehat{A}X^{\mathrm{H}}AX)$ where A, B and $\widehat{A}$ are Hermitian matrices and B is positive definite;

   (b) $\inf\limits_{X^{\mathrm{H}}BX=J_k} \operatorname{tr}(\widehat{A}X^{\mathrm{H}}AX)$ where $J_k = \begin{bmatrix} I_{k_+} & \\ & -I_{k_-} \end{bmatrix}$ and $\widehat{A} = \begin{bmatrix} \widehat{A}_+ & \\ & \widehat{A}_- \end{bmatrix}$ have the same block-diagonal structure, $\widehat{A}_\pm$ are of size $k_\pm \times k_\pm$, A, B and $\widehat{A}$ are Hermitian matrices, and (A, B) is a positive semidefinite matrix pair.

Our goal in this paper, as a continuation of [18], is to investigate yet an even more general case:

$$\inf_{\widehat{B}X^{\mathrm{H}}BX=I_k} \operatorname{tr}(\widehat{A}X^{\mathrm{H}}AX), \tag{1.1}$$

where A, B, $\widehat{A}$ and $\widehat{B}$ are Hermitian of apt sizes. It is understood that the infimum in (1.1) is taken over all X of apt sizes subject to $\widehat{B}X^{\mathrm{H}}BX = I_k$. Evidently, both problems in items 4(a) and 4(b) above are special cases of (1.1). Our main result relates this infimum to the eigenvalues of two matrix pairs (A, B) and $(\widehat{A}, \widehat{B})$.

The rest of this paper is organized as follows. We review the basics about a positive semidefinite matrix pair in section 2, and state our main result of this in paper in section 3. The proof of the main result spreads in the next two sections: section 4 deals with the simple case when $n = \widehat{n}$ and both pairs (A, B) and $(\widehat{A}, \widehat{B})$ are diagonalizable[1] while section 5 handles the main result in its generality, with the help of the result for the simple case. We draw our concluding remarks in section 6.

Notation Throughout this paper, $\mathbb{C}^{n\times m}$ is the set of $n \times m$ complex matrices, $\mathbb{C}^n = \mathbb{C}^{n\times 1}$, and $\mathbb{C} = \mathbb{C}^1$, and their real counterparts are denoted similarly by replacing $\mathbb{C}$ with $\mathbb{R}$. By $\mathcal{U}_n, \mathcal{P}_n, \mathcal{D}_n \in \mathbb{C}^{n\times n}$, denote the sets of unitary, permutation, diagonal matrices, respectively (and by $\mathcal{U}, \mathcal{P}, \mathcal{D}$ if their sizes are clear from the context); and, by $\mathcal{P}^u_n, \mathcal{D}^+_n \in \mathbb{C}^{n\times n}$, denote the set of permutation matrices in structure but with nonzero entries being any unit complex number $\mathrm{e}^{\mathrm{i}\theta}$ and that of diagonal matrices with nonnegative diagonal entries, respectively (and by $\mathcal{P}^u, \mathcal{D}^+$ if their sizes are clear from the context). I_n (or simply I if its dimension is clear from the context) is the $n \times n$ identity matrix.

For a matrix $X \in \mathbb{C}^{m\times n}$,

$$\mathcal{N}(X) = \{\boldsymbol{x} \in \mathbb{C}^n : X\boldsymbol{x} = 0\}, \quad \mathcal{R}(X) = \{X\boldsymbol{x} : \boldsymbol{x} \in \mathbb{C}^n\}$$

are the null space and the range of X (also known as the column space of X), respectively. X^{T} and X^{H} are the transpose and the conjugate transpose of a vector or matrix, respectively. $A \succ 0$ ($A \succeq 0$) means that A is Hermitian positive (semi)definite, and $A \prec 0$ ($A \preceq 0$) if $-A \succ 0$ ($-A \succeq 0$). The eigenvalues of an $n \times n$ Hermitian matrix A are written, according to either increasing or decreasing order, as

$$\lambda^{\uparrow}_1(A) \le \lambda^{\uparrow}_2(A) \le \cdots \le \lambda^{\uparrow}_n(A), \quad \text{or } \lambda^{\downarrow}_1(A) \ge \lambda^{\downarrow}_2(A) \ge \cdots \ge \lambda^{\downarrow}_n(A),$$

respectively. Hence $\lambda^{\uparrow}_i(A) = \lambda^{\downarrow}_{n-i+1}(A)$.

Other notational convention will be introduced as they appear for the first time.

2. Preliminaries on a positive semidefinite matrix pair

We review some of related concepts and results about a positive semidefinite matrix pair (A, B) [15].

Given Hermitian $B \in \mathbb{C}^{n\times n}$, the *inertia* of B is the integer triplet $(\mathfrak{i}_+(B), \mathfrak{i}_0(B), \mathfrak{i}_-(B))$, meaning B has $\mathfrak{i}_+(B)$ positive, $\mathfrak{i}_0(B)$ zero, and $\mathfrak{i}_-(B)$ negative eigenvalues, respectively. Necessarily

[1] A Hermitian matrix pair (A, B) is *diagonalizable* if it admits the decomposition $A = Y^{\mathrm{H}} \Lambda Y$ and $B = Y^{\mathrm{H}} J Y$ where Y is nonsingular and Λ and J are diagonal.

$$r := \operatorname{rank}(B) = \mathfrak{i}_+(B) + \mathfrak{i}_-(B).$$

Consider an $n \times n$ matrix pair (A, B). We say that $\mu \neq \infty$ is a *finite eigenvalue* of (A, B) if

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

As in [19,20], for a finite eigenvalue μ of (A, B), letting

$$\mathcal{N}_\mu(A, B) = \left\{ \boldsymbol{x}(\mu) \;\middle|\; \begin{array}{l} \boldsymbol{x}(\lambda) \in \mathbb{C}^n \text{ is a vector-valued polyno-} \\ \text{mial in } \lambda \text{ such that } (A - \lambda B)\boldsymbol{x}(\lambda) \equiv 0 \end{array} \right\},$$

we say that $\boldsymbol{x} \in \mathbb{C}^n$ is a corresponding *eigenvector* if $0' \neq \boldsymbol{x}'$, where $\boldsymbol{x}'$, $0' \in \mathcal{N}(A - \mu B)/\mathcal{N}_\mu(A, B)$ (the quotient space of $\mathcal{N}(A - \mu B)$ over $\mathcal{N}_\mu(A, B)$) are the elements in the quotient space that contain $\boldsymbol{x}$ and 0, respectively. The geometric multiplicity of μ is the dimension of $\mathcal{N}(A - \mu B)/\mathcal{N}_\mu(A, B)$. Together $(\mu, \boldsymbol{x})$ is called a finite *eigenpair* of (A, B). The infinite eigenvalue (eigenpair), if any, of (A, B), is defined through the eigenvalue 0, if any, of (B, A).

Definition 2.1 *([15,14]).* For A, $B \in \mathbb{C}^{n \times n}$, matrix pair (A, B) is a *Hermitian matrix pair* of order n if both A and B are Hermitian, and a *positive (semi)definite matrix pair* of order n if it is a Hermitian matrix pair and if there exists $\lambda_0 \in \mathbb{R}$ such that $A - \lambda_0 B$ is positive (semi)definite, in notation, $(A, B) \succ 0$ $(\succeq 0)$. $(A, B) \prec 0$ $(\preceq 0)$ if $(-A, -B) \succ 0$ $(\succeq 0)$.

Given a finite eigenpair $(\mu, \boldsymbol{x})$ of a Hermitian matrix pair (A, B), we say μ is an eigenvalue of positive type if $\boldsymbol{x}^{\mathrm{H}} B \boldsymbol{x} > 0$ and of negative type if $\boldsymbol{x}^{\mathrm{H}} B \boldsymbol{x} < 0$.

Let (A, B) be a positive semidefinite matrix pair of order n and $\lambda_0 \in \mathbb{R}$ be as in the definition such that $A - \lambda_0 B \succeq 0$. It is known [15, Lemma 3.8] that (A, B) has only $r = \operatorname{rank}(B)$ finite eigenvalues and all of them are real. Denote these finite eigenvalues, according to either increasing or decreasing order, by

$$\lambda_1^{-\uparrow}(A, B) \le \cdots \le \lambda_{\mathfrak{i}_-(B)}^{-\uparrow}(A, B) \le \lambda_1^{+\uparrow}(A, B) \le \cdots \le \lambda_{\mathfrak{i}_+(B)}^{+\uparrow}(A, B),$$

or

$$\lambda_1^{+\downarrow}(A, B) \ge \cdots \ge \lambda_{\mathfrak{i}_+(B)}^{+\downarrow}(A, B) \ge \lambda_1^{-\downarrow}(A, B) \ge \cdots \ge \lambda_{\mathfrak{i}_-(B)}^{-\downarrow}(A, B).$$

Here, eigenvalues $\lambda_j^{+\uparrow}(A, B)$ (and $\lambda_j^{+\downarrow}(A, B)$ too) are those of positive type (according to their associated eigenvectors $\boldsymbol{x}$ that make $\boldsymbol{x}^{\mathrm{H}} B \boldsymbol{x} > 0$), whereas $\lambda_j^{-\uparrow}(A, B)$ (and $\lambda_j^{-\downarrow}(A, B)$ too) are those of negative type (according to their associated eigenvectors $\boldsymbol{x}$ that make $\boldsymbol{x}^{\mathrm{H}} B \boldsymbol{x} < 0$).

Since both $\{\lambda_i^{-\uparrow}(A,B), \lambda_j^{+\uparrow}(A,B)\}_{i,j}$ and $\{\lambda_i^{-\downarrow}(A,B), \lambda_j^{+\downarrow}(A,B)\}_{i,j}$ are the same set of finite eigenvalues of (A,B), we have

$$\lambda_i^{-\uparrow}(A,B) = \lambda^{-\downarrow}_{\mathfrak{i}_-(B)-i+1}(A,B), \quad \lambda_j^{+\downarrow}(A,B) = \lambda^{+\uparrow}_{\mathfrak{i}_+(B)-j+1}(A,B)$$

for $1 \le i \le \mathfrak{i}_-(B)$ and $1 \le j \le \mathfrak{i}_+(B)$. It has been proved [15] that for all i, j,

$$\lambda_i^{-\uparrow}(A,B) \le \lambda_0 \le \lambda_j^{+\uparrow}(A,B), \quad \lambda_j^{+\downarrow}(A,B) \ge \lambda_0 \ge \lambda_i^{-\downarrow}(A,B).$$

As a consequence, eigenvalues of positive type are no smaller than those of negative type, i.e.,

$$\lambda_j^{+\uparrow}(A,B) - \lambda_i^{-\uparrow}(A,B) \ge 0, \quad \lambda_j^{+\downarrow}(A,B) - \lambda_i^{-\downarrow}(A,B) \ge 0.$$

There is an important comment about the types of the eigenvalues of this matrix pair $(A,B) \succeq 0$. When λ_0 in the definition is an eigenvalue, there is a possibility that (A,B) may have 2-by-2 Jordan block pairs associated with eigenvalues λ_0 (see Remark 5.1 later):

$$\left(\begin{bmatrix} 0 & \lambda_0 \\ \lambda_0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \right), \tag{2.1}$$

which corresponds to one eigenvector $\boldsymbol{x}$ with $\boldsymbol{x}^{\mathrm{H}} B \boldsymbol{x} = 0$. Each of such Jordan block pairs brings two copies of λ_0 as eigenvalues. In [15,16], one copy is artificially regarded as of positive type while the other of negative type. Although seemingly artificial, it can be justified by perturbing the first block in the pair (2.1) to $\begin{bmatrix} \varepsilon & \lambda_0 \\ \lambda_0 & 1 \end{bmatrix}$ for $\varepsilon > 0$ and letting $\varepsilon \to 0^+$. The perturbation breaks the two copies of λ_0 into $\lambda_0 + \sqrt{\varepsilon}$ of positive type and $\lambda_0 - \sqrt{\varepsilon}$ of negative type. Any other eigenvalues different from λ_0 are all associated with Jordan block pairs of 1-by-1. It can be seen that if $(A,B) \succeq 0$ does have a 2-by-2 Jordan block pair (2.1), then $\lambda_1^{-\downarrow}(A,B) = \lambda_1^{+\uparrow}(A,B) = \lambda_0$. In view of this discussion, we conclude that

$$A - \lambda_0 B \succeq 0 \quad \text{for any } \lambda_0 \in [\lambda_1^{-\downarrow}(A,B), \lambda_1^{+\uparrow}(A,B)]. \tag{2.2}$$

In fact, if $\lambda_1^{-\downarrow}(A,B) < \lambda_1^{+\uparrow}(A,B)$, then (A,B) can only have 1-by-1 Jordan block pairs.

Similar statements can be made about the eigenvalues of a negative semidefinite matrix pair.

3. Main result

Once again, we are interested in a minimization principle for

$$\inf_{\widehat{B} X^{\mathrm{H}} B X = I_{\widehat{n}}} \operatorname{tr}(\widehat{A} X^{\mathrm{H}} A X), \tag{3.1}$$

where A, $B \in \mathbb{C}^{n\times n}$ and $\widehat{A}$, $\widehat{B} \in \mathbb{C}^{\widehat{n}\times\widehat{n}}$ are all Hermitian matrices, and $\widehat{n} \leq n$. As we pointed out earlier, the infimum in (3.1) is taken over all $X \in \mathbb{C}^{n\times\widehat{n}}$ subject to $\widehat{B}X^{\rm H}BX = I_{\widehat{n}}$. Henceforward the notation k in section 1 is changed to $\widehat{n}$ to align with our overall notation structure.

The constraint $\widehat{B}X^{\rm H}BX = I_{\widehat{n}}$ necessarily implies that both $\widehat{B}$ and $X^{\rm H}BX$ are nonsingular and also

$$\mathfrak{i}_+(\widehat{B}) = \mathfrak{i}_+(\widehat{B}^{-1}) = \mathfrak{i}_+(X^{\rm H}BX) \leq \mathfrak{i}_+(B), \quad \mathfrak{i}_-(\widehat{B}) \leq \mathfrak{i}_-(B). \tag{3.2}$$

Before stating our main result on (3.1), we introduce a new notion on Hermitian matrix triplet $(B, \widehat{A}, \widehat{B})$, which we need to express our conditions for the infimum in (3.1) to be finite.

Definition 3.1. Given a Hermitian positive semidefinite pair $(\widehat{A}, \widehat{B})$ and a Hermitian matrix B, the triplet $(B, \widehat{A}, \widehat{B})$ is said *proper* if one of the following statements holds, where the *proper index pair* $(\mathfrak{d}_+(\widehat{B}), \mathfrak{d}_-(\widehat{B}))$ is defined along the way:

(i) $\mathfrak{i}_+(B) = \mathfrak{i}_+(\widehat{B})$ and $\mathfrak{i}_-(B) = \mathfrak{i}_-(\widehat{B})$, in which case $(\mathfrak{d}_+(\widehat{B}), \mathfrak{d}_-(\widehat{B})) = (0, 0)$;

(ii) $\mathfrak{i}_+(B) = \mathfrak{i}_+(\widehat{B})$, $\mathfrak{i}_-(B) > \mathfrak{i}_-(\widehat{B})$ and $\lambda_1^{+\uparrow}(\widehat{A}, \widehat{B}) \geq 0$, in which case $\mathfrak{d}_+(\widehat{B}) = 0$ and $\mathfrak{d}_-(\widehat{B})$ is the number of positive $\lambda_j^-(\widehat{A}, \widehat{B})$, $1 \leq j \leq \mathfrak{i}_-(B)$;

(iii) $\mathfrak{i}_+(B) > \mathfrak{i}_+(\widehat{B})$, $\mathfrak{i}_-(B) = \mathfrak{i}_-(\widehat{B})$ and $\lambda_1^{-\downarrow}(\widehat{A}, \widehat{B}) \leq 0$, in which case $\mathfrak{d}_-(\widehat{B}) = 0$ and $\mathfrak{d}_+(\widehat{B})$ is the number of negative $\lambda_j^+(\widehat{A}, \widehat{B})$, $1 \leq j \leq \mathfrak{i}_+(B)$;

(iv) $\mathfrak{i}_+(B) > \mathfrak{i}_+(\widehat{B})$, $\mathfrak{i}_-(B) > \mathfrak{i}_-(\widehat{B})$ and $\lambda_1^{-\downarrow}(\widehat{A}, \widehat{B}) \leq 0 \leq \lambda_1^{+\uparrow}(\widehat{A}, \widehat{B})$, in which case $(\mathfrak{d}_+(\widehat{B}), \mathfrak{d}_-(\widehat{B})) = (0, 0)$.

Here the dependency of $(\mathfrak{d}_+(\widehat{B}), \mathfrak{d}_-(\widehat{B}))$ on B and $\widehat{A}$ is suppressed for clarity. The triplet $(B, \widehat{A}, \widehat{B})$ is said *improper* if it is not proper.

In light of (3.2), the properness in Definition 3.1 simply imposes $\lambda_1^{+\uparrow}(\widehat{A}, \widehat{B}) \geq 0$ or $\lambda_1^{-\downarrow}(\widehat{A}, \widehat{B}) \leq 0$ or both, depending on which inequalities or both in (3.2) are strict or not.

As a corollary of our discussions at the end of section 2, the condition $\lambda_1^{-\downarrow}(\widehat{A}, \widehat{B}) \leq 0 \leq \lambda_1^{+\uparrow}(\widehat{A}, \widehat{B})$ in the case (iv) in the definition is the same as $\widehat{A} \succeq 0$.

Theorem 3.1. *Given four Hermitian matrices $A, B \in \mathbb{C}^{n\times n}$, $\widehat{A}, \widehat{B} \in \mathbb{C}^{\widehat{n}\times\widehat{n}}$ where $n \geq \widehat{n}$, suppose that $\widehat{A} \neq 0$, $A \neq \mu B$ for any $\mu \in \mathbb{R}$, and $\widehat{A} \neq \widehat{\mu}\widehat{B}$ for any $\widehat{\mu} \in \mathbb{R}$ when $n = \widehat{n}$. Then*

$$\inf_{\widehat{B}X^{\rm H}BX = I_{\widehat{n}}} \operatorname{tr}(\widehat{A}X^{\rm H}AX) > -\infty,$$

i.e., the infimum is finite, if and only if one of the following two cases occurs:

(i) *both* (A, B) *and* $(\widehat{A}, \widehat{B})$ *are positive semidefinite pairs and* $(B, \widehat{A}, \widehat{B})$ *is proper;*
(ii) *both* (A, B) *and* $(\widehat{A}, \widehat{B})$ *are negative semidefinite pairs and* $(-B, -\widehat{A}, -\widehat{B})$ *is proper.*

Moreover, in the first case, we have[2]

$$\begin{aligned}
&\inf_{\widehat{B}X^{\mathrm{H}}BX=I_{\widehat{n}}} \operatorname{tr}(\widehat{A}X^{\mathrm{H}}AX) \\
&= \sum_{i=1}^{\mathfrak{i}_+(\widehat{B})-\mathfrak{d}_+(\widehat{B})} \lambda_i^{+\downarrow}(\widehat{A}, \widehat{B})\, \lambda_i^{+\uparrow}(A, B) + \sum_{i=1}^{\mathfrak{d}_+(\widehat{B})} \lambda_i^{+\uparrow}(\widehat{A}, \widehat{B})\, \lambda_i^{+\downarrow}(A, B) \\
&\quad + \sum_{j=1}^{\mathfrak{d}_-(\widehat{B})} \lambda_j^{-\downarrow}(\widehat{A}, \widehat{B})\, \lambda_j^{-\uparrow}(A, B) + \sum_{j=1}^{\mathfrak{i}_-(\widehat{B})-\mathfrak{d}_-(\widehat{B})} \lambda_j^{-\uparrow}(\widehat{A}, \widehat{B})\, \lambda_j^{-\downarrow}(A, B). \qquad (3.3)
\end{aligned}$$

Similarly, for the second case, the formula for the infimum can be gotten by applying (3.3) *to matrix pairs* $(-A, -B)$ *and* $(-\widehat{A}, -\widehat{B})$*. The infimum can be attained, when* $(\widehat{A}, \widehat{B})$ *and* (A, B) *are diagonalizable.*

The three excluded cases in the conditions of the theorem are not particularly interesting:

(1) if $\widehat{A} = 0$, then $\operatorname{tr}(\widehat{A}X^{\mathrm{H}}AX) \equiv 0$ for any X;
(2) if $A = \mu B$ for some $\mu \in \mathbb{R}$, then any X such that $\widehat{B}X^{\mathrm{H}}BX = I_{\widehat{n}}$ yields $X^{\mathrm{H}}BX = \widehat{B}^{-1}$, and hence

$$\operatorname{tr}(\widehat{A}X^{\mathrm{H}}AX) = \mu \operatorname{tr}(\widehat{A}X^{\mathrm{H}}BX) \equiv \mu \operatorname{tr}(\widehat{A}\widehat{B}^{-1});$$

(3) if $\widehat{A} = \widehat{\mu}\widehat{B}$ for some $\widehat{\mu} \in \mathbb{R}$ when $n = \widehat{n}$, then any X such that $\widehat{B}X^{\mathrm{H}}BX = I_{\widehat{n}} = I_n$ which yields $X\widehat{B}X^{\mathrm{H}} = B^{-1}$ and hence

$$\operatorname{tr}(\widehat{A}X^{\mathrm{H}}AX) = \widehat{\mu} \operatorname{tr}(\widehat{B}X^{\mathrm{H}}AX) = \widehat{\mu} \operatorname{tr}(X\widehat{B}X^{\mathrm{H}}A) \equiv \widehat{\mu} \operatorname{tr}(B^{-1}A).$$

We will comment on the attainability of the infimum in (3.3) in a moment.

The proof of this theorem spreads out in the next two sections: first for a special case in section 4 and then for the general case of the theorem in section 5. One common step is to simplify the infimum by transforming (A, B) and $(\widehat{A}, \widehat{B})$ into their canonical forms (Λ, J) and $(\widehat{\Lambda}, \widehat{J})$, respectively:

$$A = Y^{\mathrm{H}}\Lambda Y,\ B = Y^{\mathrm{H}}JY, \quad \text{and} \quad \widehat{A} = \widehat{Y}^{\mathrm{H}}\widehat{\Lambda}\widehat{Y},\ \widehat{B} = \widehat{Y}^{\mathrm{H}}\widehat{J}\widehat{Y}, \qquad (3.4)$$

where $Y \in \mathbb{C}^{n\times n}$, $\widehat{Y} \in \mathbb{C}^{\widehat{n}\times\widehat{n}}$ are nonsingular. Exactly, what the canonical forms (Λ, J) and $(\widehat{\Lambda}, \widehat{J})$ look like is not important for the time being, but will be given in Lemma 5.1

[2] We adopt the convention $\sum_{i=1}^{0}(\,\cdot\,) \equiv 0$.

when we need them in section 5. In particular, (A, B) is diagonalizable if Λ and J are diagonal. We have by (3.4)

$$\begin{aligned} \operatorname{tr}(\widehat{A}X^{\mathrm{H}}AX) &= \operatorname{tr}(\widehat{Y}^{\mathrm{H}}\widehat{\Lambda}\widehat{Y}X^{\mathrm{H}}Y^{\mathrm{H}}\Lambda YX) \\ &= \operatorname{tr}(\widehat{\Lambda}\widehat{Y}X^{\mathrm{H}}Y^{\mathrm{H}}\Lambda YX\widehat{Y}^{\mathrm{H}}) \\ &= \operatorname{tr}(\widehat{\Lambda}\widetilde{X}^{\mathrm{H}}\Lambda\widetilde{X}), \end{aligned}$$

where $\widetilde{X} = YX\widehat{Y}^{\mathrm{H}}$. Notice also that $\widehat{B}X^{\mathrm{H}}BX = I_{\widehat{n}}$ can be turned into

$$I_{\widehat{n}} = \widehat{B}X^{\mathrm{H}}BX = \widehat{Y}^{\mathrm{H}}\widehat{J}\widehat{Y}X^{\mathrm{H}}Y^{\mathrm{H}}JYX = \widehat{Y}^{\mathrm{H}}\widehat{J}\widetilde{X}^{\mathrm{H}}J\widetilde{X}\widehat{Y}^{-\mathrm{H}} \Leftrightarrow \widehat{J}\widetilde{X}^{\mathrm{H}}J\widetilde{X} = I_{\widehat{n}},$$

with the same $\widetilde{X} = YX\widehat{Y}^{\mathrm{H}}$ as above. Hence

$$\inf_{\widehat{B}X^{\mathrm{H}}BX=I_{\widehat{n}}} \operatorname{tr}(\widehat{A}X^{\mathrm{H}}AX) = \inf_{\widehat{J}\widetilde{X}^{\mathrm{H}}J\widetilde{X}=I_{\widehat{n}}} \operatorname{tr}(\widehat{\Lambda}\widetilde{X}^{\mathrm{H}}\Lambda\widetilde{X}) = \inf_{\widehat{J}X^{\mathrm{H}}JX=I_{\widehat{n}}} \operatorname{tr}(\widehat{\Lambda}X^{\mathrm{H}}\Lambda X). \tag{3.5}$$

We now comment on the attainability of the infimum in (3.3) when Λ, J, $\widehat{\Lambda}$, and $\widehat{J}$ are diagonal. Suppose that both (A, B) and $(\widehat{A}, \widehat{B})$ are also positive semidefinite pairs, and $\widehat{B}$ is nonsingular (because of $\widehat{B}X^{\mathrm{H}}BX = I_{\widehat{n}}$). The other case when both pairs are nonnegative semidefinite can be handled in the same way. Since any singularity in B can also be eliminated by a congruence transformation (see Remark 5.1 later), we may also assume that B is nonsingular as well. So we can write

$$J = \begin{bmatrix} I_{n_+} & \\ & -I_{n_-} \end{bmatrix}, \quad \Lambda = \begin{array}{c} \\ {\scriptstyle n_+} \\ {\scriptstyle n_-} \end{array}\!\!\begin{array}{c} \begin{array}{cc} {\scriptstyle n_+} & {\scriptstyle n_-} \end{array} \\ \begin{bmatrix} \Lambda_+ & \\ & -\Lambda_- \end{bmatrix} \end{array}, \tag{3.6a}$$

$$\widehat{J} = \begin{bmatrix} I_{\widehat{n}_+} & \\ & -I_{\widehat{n}_-} \end{bmatrix}, \quad \widehat{\Lambda} = \begin{array}{c} \\ {\scriptstyle \widehat{n}_+} \\ {\scriptstyle \widehat{n}_-} \end{array}\!\!\begin{array}{c} \begin{array}{cc} {\scriptstyle \widehat{n}_+} & {\scriptstyle \widehat{n}_-} \end{array} \\ \begin{bmatrix} \widehat{\Lambda}_+ & \\ & -\widehat{\Lambda}_- \end{bmatrix} \end{array}, \tag{3.6b}$$

where $n_+ + n_- = n$, $\widehat{n}_+ + \widehat{n}_- = \widehat{n}$, and Λ, $\widehat{\Lambda}$ are real diagonal matrices. It can be seen that

$$\begin{aligned} \operatorname{eig}(A, B) &= \operatorname{eig}(\Lambda, J) = \operatorname{eig}(\Lambda_+) \cup \operatorname{eig}(\Lambda_-), \\ \operatorname{eig}(\widehat{A}, \widehat{B}) &= \operatorname{eig}(\widehat{\Lambda}, \widehat{J}) = \operatorname{eig}(\widehat{\Lambda}_+) \cup \operatorname{eig}(\widehat{\Lambda}_-), \end{aligned}$$

where and, in what follows, $\operatorname{eig}(\,\cdot\,)$ and $\operatorname{eig}(\,\cdot\,,\,\cdot\,)$ are the spectrum of a matrix and that of a matrix pair, respectively. Each eigenvalue $\lambda^+ \in \operatorname{eig}(\Lambda_+)$ is of positive type, i.e., $\boldsymbol{x}_+^{\mathrm{H}}B\boldsymbol{x}_+ > 0$ for its associated eigenvector $\boldsymbol{x}_+$, and each eigenvalue $\lambda^- \in \operatorname{eig}(\Lambda_-)$ is of positive type, i.e., $\boldsymbol{x}_-^{\mathrm{H}}B\boldsymbol{x}_- < 0$ for its associated eigenvector $\boldsymbol{x}_-$. The same can be said about $(\widehat{A}, \widehat{B})$. For $\widetilde{X} = P_{(:,1:\widehat{n})}\widehat{P}^{\mathrm{T}}$ where $P \in \mathcal{P}_n$ and $P_{(:,1:\widehat{n})}$ stands for the first $\widehat{n}$ columns of P, and $\widehat{P} \in \mathcal{P}_{\widehat{n}}$, we have

$$\operatorname{tr}(\widehat{\Lambda}\widetilde{X}^{\mathrm{H}}\Lambda\widetilde{X}) = \operatorname{tr}\big(\big[\widehat{P}^{\mathrm{T}}\widehat{\Lambda}\widehat{P}\big]\big[P_{(:,1:\widehat{n})}^{\mathrm{T}}\Lambda P_{(:,1:\widehat{n})}\big]\big).$$

Hence with $\widetilde{X} = P_{(:,1:\widehat{n})}\widehat{P}^{\mathrm{T}}$, $\operatorname{tr}(\widehat{\Lambda}\widetilde{X}^{\mathrm{H}}\Lambda\widetilde{X})$ is the sum of products between the diagonal entries of $\widehat{\Lambda}$, i.e., the eigenvalues of $(\widehat{A}, \widehat{B})$, and some of the those of Λ, the eigenvalues of (A, B). Certainly, there is $\widetilde{X}_{\mathrm{opt}}$ that can be explicitly constructed to give the right-hand side of (3.3). Observe that each product there is for two eigenvalues of the same type: positive or negative, and hence $P_{(:,1:\widehat{n})}^{\mathrm{T}}JP_{(:,1:\widehat{n})} = \widehat{P}^{\mathrm{T}}\widehat{J}\widehat{P}$ for that particular $\widetilde{X}_{\mathrm{opt}}$, yielding

$$[\widehat{P}^{\mathrm{T}}\widehat{J}\widehat{P}][P_{(:,1:\widehat{n})}^{\mathrm{T}}JP_{(:,1:\widehat{n})}] = I_{\widehat{n}} \;\Rightarrow\; \widehat{J}\widetilde{X}_{\mathrm{opt}}^{\mathrm{H}}J\widetilde{X}_{\mathrm{opt}} = I_{\widehat{n}}.$$

Hence $\widetilde{X}_{\mathrm{opt}}$ yields the second infimum in (3.5). Finally, $X_{\mathrm{opt}} = Y^{-1}\widetilde{X}_{\mathrm{opt}}\widehat{Y}^{-\mathrm{H}}$ yields the infimum in (3.3).

4. The simple case

In this section, we prove Theorem 3.1 for the simple case: $n = \widehat{n}$, *and both pairs* (A, B) *and* $(\widehat{A}, \widehat{B})$ *are diagonalizable*, namely, we have (3.4) with (3.6) and also $n_+ + n_- = n$, $\widehat{n}_+ = n_+$, $\widehat{n}_- = n_-$, where Λ, $\widehat{\Lambda}$ are real diagonal matrices. Note that, necessarily, $\widehat{B}$, B, and X are nonsingular because of constraint $\widehat{B}X^{\mathrm{H}}BX = I_{\widehat{n}} = I_n$, and that, by the Sylvester inertia law, $\mathfrak{i}_{\pm}(\widehat{B}) = \mathfrak{i}_{\pm}(\widehat{B}^{-1}) = \mathfrak{i}_{\pm}(B)$ upon noticing $X^{\mathrm{H}}BX = \widehat{B}^{-1}$.

We have $J = \widehat{J} = J^{-1}$ and hence the last infimum in (3.5) becomes

$$\inf_{X^{\mathrm{H}}JX=J} \operatorname{tr}(\widehat{\Lambda}X^{\mathrm{H}}\Lambda X). \tag{4.1}$$

When $J = \pm I_n$, both (A, B) and $(\widehat{A}, \widehat{B})$ are positive semidefinite pairs because $A - \lambda_0 B \succ 0$ and $\widehat{A} - \lambda_0\widehat{B} \succ 0$ for any $\lambda_0 < 0$ with sufficiently large $|\lambda_0|$ if $J = I_n$ or for any sufficiently large $\lambda_0 \in \mathbb{R}$ if $J = -I_n$. Also (A, B) and $(\widehat{A}, \widehat{B})$ are negative semidefinite pairs, too, because $(-A) - \lambda_0(-B) \succ 0$ and $(-\widehat{A}) - \lambda_0(-\widehat{B}) \succ 0$ for any sufficiently large $\lambda_0 \in \mathbb{R}$ if $J = I_n$ or for any $\lambda_0 < 0$ with sufficiently large $|\lambda_0|$ if $J = -I_n$.

As for (4.1), the case when[3] $J = \pm I_n$ has been resolved in the literature, e.g., [13, Theorem 4.3.53] as stated in the next lemma.

Lemma 4.1 *([13, Theorem 4.3.53]). Given Hermitian matrices* $A_i = U_i\Lambda_iU_i^{\mathrm{H}} \in \mathbb{C}^{n\times n}$ *with* $U_i \in \mathcal{U}_n$, $\Lambda_i \in \mathcal{D}_n$ *for* $i = 0, 1$, *we have*

$$\min_{V\in\mathcal{U}_n} \operatorname{tr}(A_0VA_1V^{\mathrm{H}}) = \min_{V\in\mathcal{P}_n} \operatorname{tr}(\Lambda_0V\Lambda_1V^{\mathrm{H}}) = \sum_{i=1}^{n} \lambda_i^{\downarrow}(A_0)\lambda_i^{\uparrow}(A_1).$$

[3] When $J = -I_n$, $X^{\mathrm{H}}JX = J$ becomes $X^{\mathrm{H}}X = I_n$, the same as for $J = I_n$.

Lemma 4.1 can be proved by using an important result on doubly stochastic matrices, namely the Birkhoff theorem. A matrix $Y \in \mathbb{R}^{n\times n}$ is *doubly stochastic* if entrywise $Y \geq 0$, and $Y\mathbf{1}_n = \mathbf{1}_n$ and $\mathbf{1}_n^{\mathrm{T}} Y = \mathbf{1}_n^{\mathrm{T}}$ where $\mathbf{1}_n \in \mathbb{R}^n$ is the vector of all ones. The Birkhoff theorem says that a doubly stochastic matrix is a convex combination of permutation matrices. Next, we will use this theorem to prove a result in Lemma 4.3, related to Lemma 4.1.

Lemma 4.2 *([21]). Let $X = [x_{ij}] \in \mathbb{C}^{n\times n}$ and $Y = [|x_{ij}|^2] \in \mathbb{R}^{n\times n}$. Then there exist doubly stochastic matrices $Y_1, Y_2 \in \mathbb{R}^{n\times n}$ such that entrywise*

$$[\sigma_{\min}(X)]^2\, Y_1 \leq Y \leq [\sigma_{\max}(X)]^2\, Y_2,$$

where $\sigma_{\min}(X)$ and $\sigma_{\max}(X)$ are the smallest and largest singular values of X, respectively.

Lemma 4.3. *Given positive semidefinite matrices $A_i = U_i \varLambda_i U_i^{\mathrm{H}} \in \mathbb{C}^{n\times n}$ with $U_i \in \mathcal{U}_n$, $\varLambda_i \in \mathcal{D}_n$ for $i = 0, 1$, we have*

$$\begin{aligned}
\operatorname{tr}(A_0 X^{\mathrm{H}} A_1 X) &\leq [\sigma_{\max}(X)]^2 \max_{V\in\mathcal{P}_n} \operatorname{tr}(\varLambda_0 V \varLambda_1 V^{\mathrm{H}})) \\
&= [\sigma_{\max}(X)]^2 \sum_{i=1}^{n} \lambda_i^{\downarrow}(A_0)\lambda_i^{\downarrow}(A_1), \\
\operatorname{tr}(A_0 X^{\mathrm{H}} A_1 X) &\geq [\sigma_{\min}(X)]^2 \min_{V\in\mathcal{P}_n} \operatorname{tr}(\varLambda_0 V \varLambda_1 V^{\mathrm{H}})) \\
&= [\sigma_{\min}(X)]^2 \sum_{i=1}^{n} \lambda_i^{\downarrow}(A_0)\lambda_i^{\uparrow}(A_1),
\end{aligned}$$

Proof. It can be seen that

$$\operatorname{tr}(A_0 X^{\mathrm{H}} A_1 X) = \operatorname{tr}(U_0 \varLambda_0 U_0^{\mathrm{H}} X^{\mathrm{H}} U_1 \varLambda_1 U_1^{\mathrm{H}} X) = \operatorname{tr}(\varLambda_0 [U_1^{\mathrm{H}} X U_0]^{\mathrm{H}} \varLambda_1 [U_1^{\mathrm{H}} X U_0]).$$

Write $U_1^{\mathrm{H}} X U_0 = [x_{ij}]$ which has the same singular values as X and let $Y = [|x_{ij}|^2]$. We get

$$\operatorname{tr}(A_0 X^{\mathrm{H}} A_1 X) = \sum_{i,j=1}^{n} \lambda_j(A_0)\lambda_i(A_1)|x_{ij}|^2.$$

Notice that all $\lambda_j(A_0)$, $\lambda_i(A_1) \geq 0$. Now use Lemma 4.2 and the Birkhoff theorem to complete the proof, following the standard technique that has been used frequently in the matrix eigenvalue perturbation theory [22,23]. □

The key tool to analyze the infimum in (4.1) is the structure of matrix $X \in \mathbb{C}^{n\times n}$ satisfying $X^{\mathrm{H}} J X = J$. Such matrix X is said *J-unitary* in literature.

Lemma 4.4 *([24, Example 6.3]). Let $J = \operatorname{diag}(I_{n_+}, -I_{n_-})$ and $n = n_+ + n_-$. A matrix $X \in \mathbb{C}^{n\times n}$ satisfies $X^{\mathrm{H}}JX = J$ if and only if it is of the form*

$$X = \begin{bmatrix} (I_{n_+} + WW^{\mathrm{H}})^{1/2} & W \\ W^{\mathrm{H}} & (I_{n_-} + W^{\mathrm{H}}W)^{1/2} \end{bmatrix} \begin{bmatrix} V_+ & \\ & V_- \end{bmatrix}, \tag{4.2}$$

where $V_+ \in \mathcal{U}_{n_+}$, $V_- \in \mathcal{U}_{n_-}$, and $W \in \mathbb{C}^{n_+\times n_-}$.

Lemma 4.4 can be found in [25,14], where (4.2) is called a (*hyperbolic*) *polar decomposition* of X. In what follows, we will limit our consideration to the case $n_+ \geq n_- \geq 1$, and the other case $1 \leq n_+ < n_-$ can be handled in a similar way.

A direct consequence of Lemma 4.4, through the SVD of W, is Lemma 4.5 below, in which (4.3) is the so-called *ChSh decomposition* of a J-unitary matrix X, an analogue of the CS decomposition of a unitary matrix [26].

Lemma 4.5 *(ChSh Decomposition). Let $J = \operatorname{diag}(I_{n_+}, -I_{n_-})$ and $n = n_+ + n_-$, where $n_+ \geq n_-$. A matrix $X \in \mathbb{C}^{n\times n}$ is J-unitary if and only if it is of the form*

$$\begin{aligned} X &= \begin{bmatrix} U_+ & \\ & U_- \end{bmatrix} \begin{bmatrix} I_{n_+-n_-} & & \\ & (I_{n_-} + \Sigma^2)^{1/2} & \Sigma \\ & \Sigma & (I_{n_-} + \Sigma^2)^{1/2} \end{bmatrix} \begin{bmatrix} V_+ & \\ & V_- \end{bmatrix} \\ &= \begin{bmatrix} U_+ & \\ & U_- \end{bmatrix} \begin{bmatrix} (I_{n_+} + \widetilde{\Sigma}\widetilde{\Sigma}^{\mathrm{H}})^{1/2} & \widetilde{\Sigma} \\ \widetilde{\Sigma} & (I_{n_-} + \widetilde{\Sigma}^{\mathrm{H}}\widetilde{\Sigma})^{1/2} \end{bmatrix} \begin{bmatrix} V_+ & \\ & V_- \end{bmatrix}, \end{aligned} \tag{4.3}$$

where $U_+, V_+ \in \mathcal{U}_{n_+}$ and $U_-, V_- \in \mathcal{U}_{n_-}$, $\widetilde{\Sigma} = \begin{bmatrix} 0 \\ \Sigma \end{bmatrix} \in \mathbb{R}^{n_+\times n_-}$ with $\Sigma \in \mathbb{R}^{n_-\times n_-}$ being diagonal and having nonnegative diagonal entries.

Lemma 4.6. *Let (Λ, J) be as in (3.6a) where $n_\pm \geq 1$ and Λ is real diagonal. Then*

(i) *$(\Lambda, J) \succeq 0$ if and only if $\lambda_i^+ - \lambda_j^- \geq 0$ for any $\lambda_i^+ \in \operatorname{eig}(\Lambda_+), \lambda_j^- \in \operatorname{eig}(\Lambda_-)$;*
(ii) *$(\Lambda, J) \preceq 0$ if and only if $\lambda_i^+ - \lambda_j^- \leq 0$ for any $\lambda_i^+ \in \operatorname{eig}(\Lambda_+), \lambda_j^- \in \operatorname{eig}(\Lambda_-)$.*

Proof. If $(\Lambda, J) \succeq 0$, then there exists $\lambda_0 \in \mathbb{R}$ such that $\Lambda - \lambda_0 J \succeq 0$, i.e., $\lambda_j^- \leq \lambda_0 \leq \lambda_i^+$ for any $\lambda_i^+ \in \operatorname{eig}(\Lambda_+), \lambda_j^- \in \operatorname{eig}(\Lambda_-)$ and thus $\lambda_i^+ - \lambda_j^- \geq 0$. On the other hand if $\lambda_i^+ - \lambda_j^- \geq 0$ for any $\lambda_i^+ \in \operatorname{eig}(\Lambda_+), \lambda_j^- \in \operatorname{eig}(\Lambda_-)$, then

$$\max\{\lambda_j^- \,:\, \lambda_j^- \in \operatorname{eig}(\Lambda_-)\} \leq \min\{\lambda_i^+ \,:\, \lambda_i^+ \in \operatorname{eig}(\Lambda_+)\}$$

and hence any λ_0 that lies between the maximum and minimum in this inequality makes $\Lambda - \lambda_0 J \succeq 0$. This proves item (i). For item (ii), by definition, $(\Lambda, J) \preceq 0$ if and only if $(-\Lambda, -J) \succeq 0$, and we then can use item (i). □

With Lemma 4.5, we have

$$\begin{aligned}
&\inf_{X^{\mathrm H}JX=J} \operatorname{tr}(\widehat{\Lambda}X^{\mathrm H}\Lambda X) \\
&= \inf_{\substack{0\preceq\Sigma\in\mathcal{D}_{n_-}\\ U_\pm,V_\pm\in\mathcal{U}_{n_\pm}}} \operatorname{tr}\left(\begin{bmatrix} V_+\widehat{\Lambda}_+V_+^{\mathrm H} & \\ & -V_-\widehat{\Lambda}_-V_-^{\mathrm H}\end{bmatrix}\begin{bmatrix} I & & \\ & (I+\Sigma^2)^{1/2} & \Sigma \\ & \Sigma & (I+\Sigma^2)^{1/2}\end{bmatrix}\times\right. \\
&\qquad \left.\begin{bmatrix} U_+^{\mathrm H}\Lambda_+U_+ & \\ & -U_-^{\mathrm H}\Lambda_-U_-\end{bmatrix}\begin{bmatrix} I & & \\ & (I+\Sigma^2)^{1/2} & \Sigma \\ & \Sigma & (I+\Sigma^2)^{1/2}\end{bmatrix}\right). \qquad (4.4)
\end{aligned}$$

By Lemma 4.6, if (Λ, J) and $(\widehat{\Lambda}, J)$ are not both positive semidefinite pairs, or not both negative definite pairs, then[4] there exist $\widehat{\lambda}_{\widehat{i}}^+ \in \operatorname{eig}(\widehat{\Lambda}_+)$, $\widehat{\lambda}_{\widehat{j}}^- \in \operatorname{eig}(\widehat{\Lambda}_-)$, $\lambda_i^+ \in \operatorname{eig}(\Lambda_+)$, and $\lambda_j^- \in \operatorname{eig}(\Lambda_-)$ with

$$(\widehat{\lambda}_{\widehat{i}}^+ - \widehat{\lambda}_{\widehat{j}}^-)(\lambda_i^+ - \lambda_j^-) < 0.$$

We now restrict Σ, $U_\pm$, and $V_\pm$ in (4.4) to special ones and doing so will increase the infimum there. Specifically, we let $\Sigma = \sigma\boldsymbol{e}_1\boldsymbol{e}_1^{\mathrm T}$ where σ is free to vary and $\boldsymbol{e}_1$ is the first column of I of apt size, and let $V_\pm$ and $U_\pm$ as products of permutation matrices

$$V_+ = P_{2+}\widehat{P}_{1+}^{\mathrm H}, \quad V_- = P_{2-}\widehat{P}_{1-}^{\mathrm H}, \quad U_+ = P_{1+}P_{2+}^{\mathrm H}, \quad U_+ = P_{1-}P_{2-}^{\mathrm H}$$

such that

$$\begin{gathered}
\big[\widehat{P}_{1+}^{\mathrm H}\widehat{\Lambda}_+\widehat{P}_{1+}\big]_{(1,1)} = \widehat{\lambda}_{\widehat{i}}^+, \quad \big[P_{1+}^{\mathrm H}\Lambda_+P_{1+}\big]_{(1,1)} = \lambda_i^+, \\
\big[\widehat{P}_{1-}^{\mathrm H}\widehat{\Lambda}_-\widehat{P}_{1-}\big]_{(1,1)} = \widehat{\lambda}_{\widehat{j}}^-, \quad \big[P_{1-}^{\mathrm H}\Lambda_-P_{1-}\big]_{(1,1)} = \lambda_j^-, \\
\big[P_{2+}^{\mathrm H}\operatorname{diag}\big(I,(I+\Sigma^2)^{1/2}\big)P_{2+}\big]_{(1,1)} = (1+\sigma^2)^{1/2}, \\
\big[P_{2-}^{\mathrm H}(I+\Sigma^2)^{1/2}P_{2-}\big]_{(1,1)} = (1+\sigma^2)^{1/2},
\end{gathered}$$

where $[\cdots]_{(1,1)}$ is the $(1,1)$st entry of a matrix. We get from (4.4)

$$\begin{aligned}
&\inf_{X^{\mathrm H}JX=J} \operatorname{tr}(\widehat{\Lambda}X^{\mathrm H}\Lambda X) \\
&\le \inf_{\sigma>0} \operatorname{tr}\left(\begin{bmatrix} \widehat{\lambda}_{\widehat{i}}^+ & & & \\ & * & & \\ & & -\widehat{\lambda}_{\widehat{j}}^- & \\ & & & * \end{bmatrix}\begin{bmatrix} (1+\sigma^2)^{1/2} & & \sigma & \\ & I & & 0 \\ \sigma & & (1+\sigma^2)^{1/2} & \\ & 0 & & I\end{bmatrix}\times\right.
\end{aligned}$$

[4] Besides the condition just mentioned, this claim also requires the condition given in the theorem: $A \neq \mu B$ for any $\mu \in \mathbb{R}$ and $\widehat{A} \neq \widehat{\mu}\widehat{B}$ for any $\widehat{\mu} \in \mathbb{R}$, which is equivalent to $\Lambda \neq \mu J$ for any $\mu \in \mathbb{R}$ and $\widehat{\Lambda} \neq \widehat{\mu} J$ for any $\widehat{\mu} \in \mathbb{R}$ because of (3.4). Otherwise if $\Lambda = \lambda_0 J$ for some $\lambda_0 \in \mathbb{R}$, then $(\Lambda, J) \succeq 0$ and $\lambda_i^+ - \lambda_j^- = 0$ for any $\lambda_i^+ \in \operatorname{eig}(\Lambda_+)$ and $\lambda_j^- \in \operatorname{eig}(\Lambda_-)$ and hence $(\widehat{\lambda}_{\widehat{i}}^+ - \widehat{\lambda}_{\widehat{j}}^-)(\lambda_i^+ - \lambda_j^-) = 0$, regardless whether $(\widehat{\Lambda}, J) \succeq 0$ or not.

$$\begin{bmatrix} \lambda_i^+ & & & \\ & * & & \\ & & -\lambda_j^- & \\ & & & * \end{bmatrix} \begin{bmatrix} (1+\sigma^2)^{1/2} & & \sigma & \\ & I & & 0 \\ \sigma & & (1+\sigma^2)^{1/2} & \\ & 0 & & I \end{bmatrix} \Bigg)$$

$$\begin{aligned}
&= \inf_{\sigma>0} \operatorname{tr} \left(\begin{bmatrix} \widehat{\lambda}_{\widehat{i}}^+ & \\ & -\widehat{\lambda}_{\widehat{j}}^- \end{bmatrix} \begin{bmatrix} (1+\sigma^2)^{1/2} & \sigma \\ \sigma & (1+\sigma^2)^{1/2} \end{bmatrix} \times \right. \\
&\qquad\qquad \left. \begin{bmatrix} \lambda_i^+ & \\ & -\lambda_j^- \end{bmatrix} \begin{bmatrix} (1+\sigma^2)^{1/2} & \sigma \\ \sigma & (1+\sigma^2)^{1/2} \end{bmatrix} \right) + (\text{constant}) \\
&= \inf_{\sigma>0} \left(\widehat{\lambda}_{\widehat{i}}^+ - \widehat{\lambda}_{\widehat{j}}^- \right)(\lambda_i^+ - \lambda_j^-)\, \sigma^2 + (\text{constant}) \\
&= -\infty. \qquad (4.5)
\end{aligned}$$

Suppose now that (Λ, J) and $(\widehat{\Lambda}, J)$ are both positive semidefinite or both negative semidefinite. Since we can switch to considering $(-\Lambda, -J)$ and $(-\widehat{\Lambda}, -J)$ instead if both (Λ, J) and $(\widehat{\Lambda}, J)$ are negative semidefinite, it suffices to consider the case when both (Λ, J) and $(\widehat{\Lambda}, J)$ are positive semidefinite, which we now assume. Then there exist $\lambda_0, \widehat{\lambda}_0 \in \mathbb{R}$ such that $\Lambda - \lambda_0 J \succeq 0$ and $\widehat{\Lambda} - \widehat{\lambda}_0 J \succeq 0$. Consider first the case $\lambda_0 = \widehat{\lambda}_0 = 0$ for which both $\widehat{\Lambda}$ and Λ are positive semidefinite, i.e., $\Lambda_+,\ \widehat{\Lambda}_+ \succeq 0$ and $-\Lambda_-,\ -\widehat{\Lambda}_- \succeq 0$. We have

$$\begin{aligned}
&\inf_{X^{\mathrm{H}} J X = J} \operatorname{tr}(\widehat{\Lambda} X^{\mathrm{H}} \Lambda X) \\
&= \inf_{\substack{0 \preceq \Sigma \in \mathcal{D}_{n_-} \\ U_\pm, V_\pm \in \mathcal{U}_{n_\pm}}} \operatorname{tr} \left(\begin{bmatrix} V_+ \widehat{\Lambda}_+ V_+^{\mathrm{H}} & \\ & -V_- \widehat{\Lambda}_- V_-^{\mathrm{H}} \end{bmatrix} \begin{bmatrix} (I+\widetilde{\Sigma}\widetilde{\Sigma}^{\mathrm{H}})^{1/2} & \widetilde{\Sigma} \\ \widetilde{\Sigma} & (I+\widetilde{\Sigma}^{\mathrm{H}}\widetilde{\Sigma})^{1/2} \end{bmatrix} \times \right. \\
&\qquad \left. \begin{bmatrix} U_+^{\mathrm{H}} \Lambda_+ U_+ & \\ & -U_-^{\mathrm{H}} \Lambda_- U_- \end{bmatrix} \begin{bmatrix} (I+\widetilde{\Sigma}\widetilde{\Sigma}^{\mathrm{H}})^{1/2} & \widetilde{\Sigma} \\ \widetilde{\Sigma} & (I+\widetilde{\Sigma}^{\mathrm{H}}\widetilde{\Sigma})^{1/2} \end{bmatrix} \right) \\
&= \inf_{\substack{0 \preceq \Sigma \in \mathcal{D}_{n_-} \\ U_\pm, V_\pm \in \mathcal{U}_{n_\pm}}} \Bigg[\underbrace{\operatorname{tr}\left(V_+ \widehat{\Lambda}_+ V_+^{\mathrm{H}} (I + \widetilde{\Sigma}\widetilde{\Sigma}^{\mathrm{H}})^{1/2} U_+^{\mathrm{H}} \Lambda_+ U_+ (I + \widetilde{\Sigma}\widetilde{\Sigma}^{\mathrm{H}})^{1/2} \right)}_{=:\tau_1} \\
&\qquad + \underbrace{\operatorname{tr}\left(V_+ \widehat{\Lambda}_+ V_+^{\mathrm{H}} \widetilde{\Sigma}^{\mathrm{H}} U_-^{\mathrm{H}} [-\Lambda_-] U_- \widetilde{\Sigma} \right)}_{=:\tau_2} \\
&\qquad + \underbrace{\operatorname{tr}\left(V_- [-\widehat{\Lambda}_-] V_-^{\mathrm{H}} \widetilde{\Sigma} U_+^{\mathrm{H}} \Lambda_+ U_+ \widetilde{\Sigma}^{\mathrm{H}} \right)}_{=:\tau_3} \\
&\qquad + \underbrace{\operatorname{tr}\left(V_- [-\widehat{\Lambda}_-] V_-^{\mathrm{H}} (I + \widetilde{\Sigma}^{\mathrm{H}}\widetilde{\Sigma})^{1/2} U_-^{\mathrm{H}} [-\Lambda_-] U_- (I + \widetilde{\Sigma}^{\mathrm{H}}\widetilde{\Sigma})^{1/2} \right)}_{=:\tau_4} \Bigg] \\
&\ge \inf_{\substack{0 \preceq \Sigma \in \mathcal{D}_{n_-} \\ U_+, V_+ \in \mathcal{U}_{n_+}}} \tau_1 + \inf_{\substack{0 \preceq \Sigma \in \mathcal{D}_{n_-} \\ U_- \in \mathcal{U}_{n_-}, V_+ \in \mathcal{U}_{n_+}}} \tau_2 + \inf_{\substack{0 \preceq \Sigma \in \mathcal{D}_{n_-} \\ U_+ \in \mathcal{U}_{n_+}, V_- \in \mathcal{U}_{n_-}}} \tau_3 + \inf_{\substack{0 \preceq \Sigma \in \mathcal{D}_{n_-} \\ U_-, V_- \in \mathcal{U}_{n_-}}} \tau_4.
\end{aligned}$$

Next we bound these four infima from below. By Lemma 4.3 (with $X = (I + \widetilde{\Sigma}\widetilde{\Sigma}^{\mathrm{H}})^{1/2}$), we have

$$\begin{aligned}\inf_{\substack{0\preceq\Sigma\in\mathcal{D}_{n_-}\\ U_+,V_+\in\mathcal{U}_{n_+}}} \tau_1 &\ge \inf_{U_\pm,V_\pm\in\mathcal{P}_n} \operatorname{tr}(V_+\widehat{\Lambda}_+V_+^{\mathrm{H}}U_+^{\mathrm{H}}\Lambda_+U_+)\\ &= \sum_{i=1}^{n_+}\lambda_i^{\downarrow}(\widehat{\Lambda}_+)\lambda_i^{\uparrow}(\Lambda_+),\end{aligned}$$

and, again by Lemma 4.3 (with $X = \left[\begin{smallmatrix}0\\ \Sigma\end{smallmatrix}\right]$),

$$\begin{aligned}\inf_{\substack{0\preceq\Sigma\in\mathcal{D}_{n_-}\\ U_-\in\mathcal{U}_{n_-},V_+\in\mathcal{U}_{n_+}}} \tau_2 &\ge \inf_{\substack{0\preceq\Sigma\in\mathcal{D}_{n_-}\\ \widetilde{U}_-,V_+\in\mathcal{U}_{n_+}}} \operatorname{tr}(V_+\widehat{\Lambda}_+V_+^{\mathrm{H}}\left[\begin{smallmatrix}0\\ \Sigma\end{smallmatrix}\right]\widetilde{U}_-^{\mathrm{H}}\left[\begin{smallmatrix}0 & \\ & -\Lambda_-\end{smallmatrix}\right]\widetilde{U}_-\left[\begin{smallmatrix}0\\ \Sigma\end{smallmatrix}\right])\\ &\ge 0.\end{aligned}$$

Similarly, we can bound τ_3 and τ_4 from below. Put all together to get

$$\inf_{X^{\mathrm{H}}JX=J} \operatorname{tr}(\widehat{\Lambda}X^{\mathrm{H}}\Lambda X) \ge \sum_{i=1}^{n_+}\lambda_i^{\downarrow}(\widehat{\Lambda}_+)\lambda_i^{\uparrow}(\Lambda_+) + \sum_{j=1}^{n_-}\lambda_j^{\downarrow}(\widehat{\Lambda}_-)\lambda_j^{\uparrow}(\Lambda_-).$$

Since the right-hand side is achieved by $\operatorname{tr}(\widehat{\Lambda}X^{\mathrm{H}}\Lambda X)$ at $\Sigma = 0$, and $U_\pm, V_\pm \in \mathcal{P}_{n_\pm}$ such that the diagonal values of $\Lambda_\pm$ and $\widehat{\Lambda}_\pm$ are in the increasing and decreasing order, respectively, we conclude that

$$\min_{X^{\mathrm{H}}JX=J} \operatorname{tr}(\widehat{\Lambda}X^{\mathrm{H}}\Lambda X) = \sum_{i=1}^{n_+}\lambda_i^{\downarrow}(\widehat{\Lambda}_+)\lambda_i^{\uparrow}(\Lambda_+) + \sum_{j=1}^{n_-}\lambda_j^{\downarrow}(\widehat{\Lambda}_-)\lambda_j^{\uparrow}(\Lambda_-). \tag{4.6}$$

We now claim (4.6) remains valid for the case when at least one of λ_0 and $\widehat{\lambda}_0$ is not 0. With $X^{\mathrm{H}}JX = J$, we have

$$\begin{aligned}&\operatorname{tr}(\widehat{\Lambda}X^{\mathrm{H}}\Lambda X)\\ &= \operatorname{tr}(\widehat{\Lambda}X^{\mathrm{H}}[\Lambda - \lambda_0 J]X) + \lambda_0\operatorname{tr}(\widehat{\Lambda}J)\\ &= \operatorname{tr}([\widehat{\Lambda} - \widehat{\lambda}_0 J]X^{\mathrm{H}}[\Lambda - \lambda_0 J]X) + \widehat{\lambda}_0\operatorname{tr}(JX^{\mathrm{H}}[\Lambda - \lambda_0 J]X) + \lambda_0\operatorname{tr}(\widehat{\Lambda}J)\\ &= \operatorname{tr}([\widehat{\Lambda} - \widehat{\lambda}_0 J]X^{\mathrm{H}}[\Lambda - \lambda_0 J]X) + \widehat{\lambda}_0\operatorname{tr}((JX)^{-1}[\Lambda - \lambda_0 J]X) + \lambda_0\operatorname{tr}(\widehat{\Lambda}J)\\ &= \operatorname{tr}([\widehat{\Lambda} - \widehat{\lambda}_0 J]X^{\mathrm{H}}[\Lambda - \lambda_0 J]X) + \widehat{\lambda}_0\operatorname{tr}(J\Lambda) - \widehat{\lambda}_0\lambda_0\operatorname{tr}(I) + \lambda_0\operatorname{tr}(\widehat{\Lambda}J),\end{aligned} \tag{4.7}$$

where only the first term varies with X. Since

$$\Lambda - \lambda_0 J = \begin{bmatrix}\Lambda_+ - \lambda_0 I_{n_+} & \\ & -(\Lambda_- - \lambda_0 I_{n_-})\end{bmatrix} \succeq 0,$$

$$\widehat{\Lambda}-\widehat{\lambda}_0J=\begin{bmatrix}\widehat{\Lambda}_+-\widehat{\lambda}_0I_{n_+} & \\ & -(\widehat{\Lambda}_--\widehat{\lambda}_0I_{n_-})\end{bmatrix}\succeq 0,$$

Similarly to (4.6), we can get, for the first term in (4.7),

$$\begin{aligned}&\min_{X^{\mathrm{H}}JX=J}\operatorname{tr}([\widehat{\Lambda}-\widehat{\lambda}_0J]X^{\mathrm{H}}[\Lambda-\lambda_0J]X)\\&=\sum_{i=1}^{n_+}[\lambda_i^{\downarrow}(\widehat{\Lambda}_+)-\widehat{\lambda}_0][\lambda_i^{\uparrow}(\Lambda_+)-\lambda_0]+\sum_{j=1}^{n_-}[\lambda_j^{\downarrow}(\widehat{\Lambda}_-)-\widehat{\lambda}_0][\lambda_j^{\uparrow}(\Lambda_-)-\lambda_0].\end{aligned}\tag{4.8}$$

Plug this expression into (4.7) to yield (4.6) for the case when at least one of λ_0 and $\widehat{\lambda}_0$ is not 0.

We summarize what we just proved into Lemma 4.7.

Lemma 4.7. *Given Hermitian matrix pairs (A,B) and $(\widehat{A},\widehat{B})$ with nonsingular $B,\widehat{B}\in\mathbb{C}^{n\times n}$, suppose that both pairs are diagonalizable and that $A\neq\mu B$ for any $\mu\in\mathbb{R}$, and $\widehat{A}\neq\widehat{\mu}\widehat{B}$ for any $\widehat{\mu}\in\mathbb{R}$. Then*

$$\inf_{\widehat{B}X^{\mathrm{H}}BX=I_n}\operatorname{tr}(\widehat{\Lambda}X^{\mathrm{H}}\Lambda X)>-\infty$$

if and only if either both (A,B) and $(\widehat{A},\widehat{B})$ are positive semidefinite pairs or both are negative semidefinite pairs. Moreover, in the first case, i.e., when both (A,B) and $(\widehat{A},\widehat{B})$ are positive semidefinite pairs,

$$\min_{\widehat{B}X^{\mathrm{H}}BX=I_{\widehat{n}}}\operatorname{tr}(\widehat{A}X^{\mathrm{H}}AX)=\sum_{i=1}^{\mathrm{i}_+(B)}\lambda_i^{+\downarrow}(\widehat{A},\widehat{B})\lambda_i^{+\uparrow}(A,B)+\sum_{j=1}^{\mathrm{i}_-(B)}\lambda_j^{-\downarrow}(\widehat{A},\widehat{B})\lambda_j^{-\uparrow}(A,B).\tag{4.9}$$

A similar expression for the infimum for the case when both (A,B) and $(\widehat{A},\widehat{B})$ are negative semidefinite pairs can be gotten by applying (4.9) to $(-A,-B)$ and $(-\widehat{A},-\widehat{B})$.

Lemma 4.7 is a special case of Theorem 3.1, and it with $B=\widehat{B}=I_n$ yields Lemma 4.1.

5. The general case

In this section we prove Theorem 3.1 in its generality. We will assume that B is indefinite, except in Remark 5.2 later where we will comment on how the case when B is positive or negative semidefinite can be handled in a simpler way.

We still have the decompositions in (3.4) and simplification in (3.5), with (Λ,J) to be specified as in Lemma 5.1 and similarly for $(\widehat{\Lambda},\widehat{J})$.

Lemma 5.1 *([27, Theorem 5.1]). Let p be a positive integer and*

$$K_p(\tau) = \begin{bmatrix} & & & & \tau \\ & & & \tau & 1 \\ & & \iddots & 1 & \\ & \iddots & \iddots & & \\ \tau & 1 & & & \\ \tau \;\; 1 & & & & \end{bmatrix}_{p\times p}, \qquad F_p = \begin{bmatrix} & & & & 1 \\ & & & 1 & \\ & & \iddots & & \\ & \iddots & & & \\ & 1 & & & \\ 1 & & & & \end{bmatrix}_{p\times p}.$$

Any Hermitian matrix pair (A, B) is congruent to (Λ, J) in the sense that

$$A = Y^{\mathrm{H}} \Lambda Y,\ B = Y^{\mathrm{H}} J Y,$$

where Y is nonsingular and Λ and J are block-diagonal matrices with corresponding diagonal blocks coming from block pairs of types:

T-o: $(0, 0)$,
T-s$(2p+1)$: $\left(K_{2p+1}(0), \begin{bmatrix} & & F_p \\ & 0 & \\ F_p & & \end{bmatrix}\right)$,
T-∞(p): $(\eta F_p, \eta K_p(0))$ *with* $\eta \in \{\pm 1\}$, *associated with an infinite eigenvalue,*
T-c(p): $(\begin{bmatrix} 0 & K_p(\alpha + \mathrm{i}\beta) \\ K_p(\alpha - \mathrm{i}\beta) & \end{bmatrix}, F_{2p})$, *associated with a pair of conjugate complex eigenvalues* $\alpha \pm \mathrm{i}\beta$ *with* $\alpha \in \mathbb{R}, \beta > 0$,
T-r(p): $(\eta K_p(\alpha), \eta F_p)$ *with* $\eta \in \{\pm 1\}$, *associated with a finite real eigenvalue* α.

Moreover, (Λ, J) is unique up to a simultaneous permutation of the corresponding diagonal block pairs.

Although Lemma 5.1 lists five possible types of block pairs that each of (Λ, J) and $(\widehat{\Lambda}, \widehat{J})$ may contain, we can quickly exclude some types of block pairs from (Λ, J) and $(\widehat{\Lambda}, \widehat{J})$, based on the conditions of Theorem 3.1.

- $(\widehat{\Lambda}, \widehat{J})$ possibly contains block pairs of types **T-c(p)** and **T-r(p)** only. This is because $\widehat{B}$ is nonsingular and so is $\widehat{J}$, and hence block pairs of type **T-o**, **T-s$(2p+1)$**, or **T-∞(p)** do not show up in pair $(\widehat{\Lambda}, \widehat{J})$. For that reason, we will have $\widehat{J}^{-1} = \widehat{J}$ and hence constraint $\widehat{J} X^{\mathrm{H}} J X = I_{\widehat{n}}$ is equivalent to $X^{\mathrm{H}} J X = \widehat{J}$. It follows from (3.5) that

$$\inf_{\widehat{B} X^{\mathrm{H}} B X = I_{\widehat{n}}} \operatorname{tr}(\widehat{A} X^{\mathrm{H}} A X) = \inf_{\widehat{J} X^{\mathrm{H}} J X = I_{\widehat{n}}} \operatorname{tr}(\widehat{\Lambda} X^{\mathrm{H}} \Lambda X) = \inf_{X^{\mathrm{H}} J X = \widehat{J}} \operatorname{tr}(\widehat{\Lambda} X^{\mathrm{H}} \Lambda X). \tag{5.1}$$

In the rest of this section, we will investigate the last infimum in (5.1).

- We can also exclude block pairs of type **T-o** from (Λ, J). In fact, if (Λ, J) contains block pairs of type **T-o**, then we can write $\Lambda = \begin{bmatrix} \Lambda_r & \\ & 0 \end{bmatrix}$, $J = \begin{bmatrix} J_r & \\ & 0 \end{bmatrix}$, and partition $X = \begin{bmatrix} X_r \\ X_s \end{bmatrix}$ accordingly to get

$$\inf_{X^{\mathrm{H}} J X = \widehat{J}} \mathrm{tr}(\widehat{\Lambda} X^{\mathrm{H}} \Lambda X) = \inf_{X_r^{\mathrm{H}} J_r X_r = \widehat{J}} \mathrm{tr}(\widehat{\Lambda} X_r^{\mathrm{H}} \Lambda_r X_r),$$

which falls into the case that (Λ, J) contains no block pair of type **T-o**.

In summary, possible types of block pairs in (Λ, J) and $(\widehat{\Lambda}, \widehat{J})$ to consider henceforward are

$$(\Lambda, J): \quad \textbf{T-s}(2p+1),\ \textbf{T-}\infty(p),\ \textbf{T-c}(p),\ \textbf{T-r}(p); \tag{5.2a}$$

$$(\widehat{\Lambda}, \widehat{J}): \quad \textbf{T-c}(p),\ \textbf{T-r}(p). \tag{5.2b}$$

In our later analysis, we will also replace any block pair of type **T-c**(p) with

$$\left(\begin{bmatrix} K_p(\alpha) & -\mathrm{i}\beta F_p \\ \mathrm{i}\beta F_p & -K_p(\alpha) \end{bmatrix}, \begin{bmatrix} F_p & \\ & -F_p \end{bmatrix} \right). \tag{5.3}$$

This is because they are congruent:

$$\frac{1}{\sqrt{2}} \begin{bmatrix} I & I \\ I & -I \end{bmatrix} \cdot \begin{bmatrix} 0 & K_p(\alpha + \mathrm{i}\beta) \\ K_p(\alpha - \mathrm{i}\beta) & \end{bmatrix} \cdot \frac{1}{\sqrt{2}} \begin{bmatrix} I & I \\ I & -I \end{bmatrix} = \begin{bmatrix} K_p(\alpha) & -\mathrm{i}\beta F_p \\ \mathrm{i}\beta F_p & -K_p(\alpha) \end{bmatrix},$$
$$\frac{1}{\sqrt{2}} \begin{bmatrix} I & I \\ I & -I \end{bmatrix} \cdot \begin{bmatrix} & F_p \\ F_p & \end{bmatrix} \cdot \frac{1}{\sqrt{2}} \begin{bmatrix} I & I \\ I & -I \end{bmatrix} = \begin{bmatrix} F_p & \\ & -F_p \end{bmatrix}.$$

Remark 5.1. When (A, B) is positive semidefinite, possible block pairs in its canonical form are considerably limited [15, Lemma 3.8]. In fact, if $A - \lambda_0 B \succeq 0$ for some $\lambda_0 \in \mathbb{R}$, then its canonical form possibly contains (0,0) of type **T-o**, $(\eta K_1(\alpha), \eta F_1)$ of type **T-r(1)** such that $\eta(\alpha - \lambda_0) \geq 0$, $(K_2(\lambda_0), F_2)$ of type **T-r(2)**, and $(1, 0)$ of type **T-**∞**(1)**.

The next lemma will be used in subsection 5.1 to reduce the case $n > \widehat{n}$ to the case $n = \widehat{n}$. It may be of interest in its own and it also sheds light on why Definition 3.1 reads the way it is.

Lemma 5.2. *Let $B \in \mathbb{C}^{n \times n}$, $\widehat{A}$, $\widehat{B} \in \mathbb{C}^{\widehat{n} \times \widehat{n}}$ be Hermitian matrices. Suppose that B and $\widehat{B}$ are nonsingular, and $\widehat{n}_\pm := \mathfrak{i}_\pm(\widehat{B}) \leq n_\pm := \mathfrak{i}_\pm(B)$, and let*

$$\widetilde{A} = \begin{bmatrix} \widehat{A} & \\ & 0 \end{bmatrix} \in \mathbb{C}^{n \times n}, \quad J_c = \begin{bmatrix} I_{n_+ - \widehat{n}_+} & \\ & -I_{n_- - \widehat{n}_-} \end{bmatrix}, \quad \widetilde{B} = \begin{bmatrix} \widehat{B} & \\ & J_c \end{bmatrix}.$$

(i) *If* $(\widetilde{A}, \widetilde{B}) \succeq 0$, *then* $(\widehat{A}, \widehat{B}) \succeq 0$ *and* $(B, \widehat{A}, \widehat{B})$ *is proper; conversely, if* $(\widehat{A}, \widehat{B}) \succeq 0$ *and* $(B, \widehat{A}, \widehat{B})$ *is proper, then* $(\widetilde{A}, \widetilde{B}) \succeq 0$.

(ii) *If* $(\widetilde{A}, \widetilde{B}) \preceq 0$, *then* $(\widehat{A}, \widehat{B}) \preceq 0$ *and* $(-B, -\widehat{A}, -\widehat{B})$ *is proper; conversely, if* $(\widehat{A}, \widehat{B}) \preceq 0$ *and* $(-B, -\widehat{A}, -\widehat{B})$ *is proper, then* $(\widetilde{A}, \widetilde{B}) \preceq 0$.

Proof. We will prove item (i) only. Item (ii) becomes item (i) by simply considering $(-\widehat{A}, -\widehat{B})$ instead. No proof is necessary if $\widehat{n} = n$. Suppose that $\widehat{n} < n$. There are three subcases to consider: (1) both $\widehat{n}_\pm < n_\pm$, (2) $\widehat{n}_+ < n_+$ and $\widehat{n}_- = n_-$, and (3) $\widehat{n}_+ = n_+$ and $\widehat{n}_- < n_-$.

Consider subcase (1). Suppose that $(\widetilde{A}, \widetilde{B}) \succeq 0$, i.e., $\widetilde{A} - \lambda_0 \widetilde{B} \succeq 0$. Then $\widehat{A} - \lambda_0 \widehat{B} \succeq 0$ and $-\lambda_0 J_c \succeq 0$, implying $(\widehat{A}, \widehat{B}) \succeq 0$ and $\lambda_0 = 0$. That $\lambda_0 = 0$ implies that $\lambda_1^{-\downarrow}(\widehat{A}, \widehat{B}) \le 0 \le \lambda_1^{+\uparrow}(\widehat{A}, \widehat{B})$ and thus $(B, \widehat{A}, \widehat{B})$ is proper. Conversely, if $(\widehat{A}, \widehat{B}) \succeq 0$ and $(B, \widehat{A}, \widehat{B})$ is proper, then by Definition 3.1, we find that $\lambda_1^{-\downarrow}(\widehat{A}, \widehat{B}) \le 0 \le \lambda_1^{+\uparrow}(\widehat{A}, \widehat{B})$ and hence $\widehat{A} - \lambda_0 \widehat{B} \succeq 0$ for $\lambda_0 = 0$, i.e., $\widehat{A} \succeq 0$, and hence $\widetilde{A} - \lambda_0 \widetilde{B} \succeq 0$, i.e., $(\widetilde{A}, \widetilde{B}) \succeq 0$.

Consider subcase (2). Suppose that $(\widetilde{A}, \widetilde{B}) \succeq 0$, i.e., $\widetilde{A} - \lambda_0 \widetilde{B} \succeq 0$. Then $\widehat{A} - \lambda_0 \widehat{B} \succeq 0$ and $-\lambda_0 J_c \succeq 0$, implying $(\widehat{A}, \widehat{B}) \succeq 0$ and $\lambda_0 \le 0$. Hence $\lambda_1^{-\downarrow}(\widehat{A}, \widehat{B}) \le \lambda_0 \le 0$ and $(B, \widehat{A}, \widehat{B})$ is proper. Conversely, if $(\widehat{A}, \widehat{B}) \succeq 0$ and $(B, \widehat{A}, \widehat{B})$ is proper, then by Definition 3.1, we find that $\lambda_1^{-\downarrow}(\widehat{A}, \widehat{B}) \le 0$. By (2.2), $\widehat{A} - \lambda_0 \widehat{B} \succeq 0$ for some $\lambda_0 \le 0$ and hence $-\lambda_0 J_c \succeq 0$ and $\widetilde{A} - \lambda_0 \widetilde{B} \succeq 0$, i.e., $(\widetilde{A}, \widetilde{B}) \succeq 0$.

Subcase (3) can be handled in the same way as handling subcase (2). □

We now prove Theorem 3.1 in an order of increasing complexity of (Λ, J) and $(\widehat{\Lambda}, \widehat{J})$ in terms of possible combinations of block pairs of types listed in (5.2), and hence conclude the proof at the end.

5.1. *Involving block pairs of type* **T-r**(1), **T-c**(1) *only*

In this case, we have

$$J = \begin{bmatrix} I_{n_+} & \\ & -I_{n_-} \end{bmatrix}, \quad \widehat{J} = \begin{bmatrix} I_{\widehat{n}_+} & \\ & -I_{\widehat{n}_-} \end{bmatrix}.$$

Recall $\widehat{n}_+ \le n_+$ and $\widehat{n}_- \le n_-$ by (3.2). Let $J_c = \begin{bmatrix} I_{n_+ - \widehat{n}_+} & \\ & -I_{n_- - \widehat{n}_-} \end{bmatrix}$. For any X such that $X^{\mathrm{H}} J X = \widehat{J}$, we can complement X to a square matrix $\widetilde{X} = [X \quad X_c]$ such that $\widetilde{X}^{\mathrm{H}} J \widetilde{X} = \mathrm{diag}(\widehat{J}, J_c)$ and then $(\widetilde{X}P)^{\mathrm{H}} J (\widetilde{X}P) = J$ upon permuting the columns of $\widetilde{X}$ by some permutation matrix P. This is guaranteed by Lemma 5.3 below that can be found in many classical monographs, e.g., [28,29].

Lemma 5.3 *([24, Corollary 5.12]). Let* $J = \mathrm{diag}(I_{n_+}, -I_{n_-})$ *and* $n = n_+ + n_-$. *Any set of vectors* $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k$ *satisfying* $\boldsymbol{u}_i^{\mathrm{H}} J \boldsymbol{u}_j = \pm\delta_{ij}$ *for* $i, j = 1, \ldots, k$ *can be complemented to a basis* $\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n\}$ *of* $\mathbb{C}^n$ *satisfying* $\boldsymbol{u}_i^{\mathrm{H}} J \boldsymbol{u}_j = \pm\delta_{ij}$ *for* $i, j = 1, \ldots, n$, *where* δ_{ij} *is* *the Kronecker delta which is* 1 *for* $i = j$ *and* 0 *otherwise, and the numbers of* 1 *and* -1 *among* $\boldsymbol{u}_i^{\mathrm{H}} J \boldsymbol{u}_i$ *for* $1 \le i \le n$ *are* n_+ *and* n_-*, respectively.*

Set

$$\widetilde{\Lambda} = \begin{bmatrix} \widehat{\Lambda} & \\ & 0 \end{bmatrix} \in \mathbb{C}^{n \times n}, \quad \widetilde{J} = \begin{bmatrix} \widehat{J} & \\ & J_c \end{bmatrix}. \tag{5.4}$$

It can be seen that

$$\inf_{\widetilde{X}^{\mathrm{H}} J \widetilde{X} = \widetilde{J}} \operatorname{tr}(\widetilde{\Lambda} \widetilde{X}^{\mathrm{H}} \Lambda \widetilde{X}) = \inf_{\substack{X^{\mathrm{H}} J X = \widehat{J} \\ X_c^{\mathrm{H}} J X = 0, X_c^{\mathrm{H}} J X_c = J_c}} \operatorname{tr}(\widehat{\Lambda} X^{\mathrm{H}} \Lambda X) = \inf_{X^{\mathrm{H}} J X = \widehat{J}} \operatorname{tr}(\widehat{\Lambda} X^{\mathrm{H}} \Lambda X).$$

This and Lemma 5.2 show that we can consider (Λ, J) and $(\widetilde{\Lambda}, \widetilde{J})$ instead, for which $n = \widehat{n}$.

In the rest of this subsection, we will assume $n = \widehat{n}$. We consider three subcases:

(1) Only block pairs of type **T-r**(1) are involved;
(2) $\widehat{\Lambda} = \widehat{\mu} \widehat{J}$ for some $\widehat{\mu} \in \mathbb{R}$;
(3) Besides possibly block pairs of type **T-r**(1), at least one block pair of type **T-c**(1) is also involved and $\widehat{\Lambda} \neq \widehat{\mu} \widehat{J}$ for any $\widehat{\mu} \in \mathbb{R}$.

Subcase (1) has already been taken care of in section 4. Subcase (2) falls into the excluded cases of the theorem: $\widehat{A} \neq \widehat{\mu} \widehat{B}$ for any $\widehat{\mu} \in \mathbb{R}$ if $n = \widehat{n}$ to begin with, i.e., without the expansions in (5.4), or if with the expansions then $0 = \widehat{\mu} J_c \Rightarrow \widehat{\mu} = 0$, yielding $\widehat{A} = 0$.

We now turn our attention to subcase (3). Now J, $\widehat{J} \in \mathbb{C}^{n \times n}$ are nonsingular, and $\mathrm{i}_{\pm}(J) = \mathrm{i}_{\pm}(\widehat{J})$. Notice that the direct sum of pairs of type **T-r**(1) is a diagonal pair, and each block pair of type **T-c**(1) can be turned into (5.3) for $p = 1$ by a congruent transformation. Thus we can assume

$$\Lambda = \begin{bmatrix} \Lambda_+^c & -\mathrm{i}\Omega^c \\ \mathrm{i}\Omega^c & -\Lambda_-^c \end{bmatrix}, \; J = \begin{bmatrix} I_{n_+} & \\ & -I_{n_-} \end{bmatrix}, \; \widehat{\Lambda} = \begin{bmatrix} \widehat{\Lambda}_+^c & -\mathrm{i}\widehat{\Omega}^c \\ \mathrm{i}\widehat{\Omega}^c & -\widehat{\Lambda}_-^c \end{bmatrix}, \; \widehat{J} = J,$$

where $\Lambda_\pm^c, \widehat{\Lambda}_\pm^c \in \mathcal{D}$, and $\Omega^c, \widehat{\Omega}^c \in \mathbb{R}^{n_+ \times n_-}$ are leading diagonal matrices with nonnegative diagonal entries. As a result,

$$\inf_{X^{\mathrm{H}} J X = \widehat{J}} \operatorname{tr}(\widehat{\Lambda} X^{\mathrm{H}} \Lambda X) = \inf_{\substack{\Sigma \in \mathcal{D}^+ \\ U_\pm, V_\pm \in \mathcal{U}}} \operatorname{tr}(\widehat{\Lambda} X^{\mathrm{H}} \Lambda X) \le \inf_{\substack{\Sigma \in \mathcal{D}^+ \\ U_\pm, V_\pm \in \mathcal{P}^u}} \operatorname{tr}(\widehat{\Lambda} X^{\mathrm{H}} \Lambda X).$$

In a way similar to that in (4.5), we will select concrete $U_\pm, V_\pm \in \mathcal{P}^u$ to establish a necessary condition such that the infimum is not $-\infty$.

First we consider the case $n = \widehat{n} = 2$. Note that

$$\begin{bmatrix} \sqrt{1+\sigma^2} & \sigma \\ \sigma & \sqrt{1+\sigma^2} \end{bmatrix} \begin{bmatrix} \alpha & \\ & -\alpha \end{bmatrix} \begin{bmatrix} \sqrt{1+\sigma^2} & \sigma \\ \sigma & \sqrt{1+\sigma^2} \end{bmatrix} = \begin{bmatrix} \alpha & \\ & -\alpha \end{bmatrix}.$$

There are three mutually exclusive subcases:

(i) both $(\widehat{\Lambda}, \widehat{J})$ and (Λ, J) are block pairs of type **T-c(1)**. We have

$$\begin{aligned}
&\inf_{\substack{\Sigma\in\mathcal{D}^+\\ U_\pm,V_\pm\in\mathcal{P}^u}} \operatorname{tr}(\widehat{\Lambda}X^{\mathrm{H}}\Lambda X)\\
&= \inf_{\substack{\sigma\geq 0\\ \theta,\widehat{\theta}\in[0,2\pi)}} \operatorname{tr}\left(\begin{bmatrix} \widehat{\alpha} & -\mathrm{i}\widehat{\beta}\mathrm{e}^{\mathrm{i}\widehat{\theta}} \\ \mathrm{i}\widehat{\beta}\mathrm{e}^{-\mathrm{i}\widehat{\theta}} & -\widehat{\alpha}\end{bmatrix}\begin{bmatrix} \sqrt{1+\sigma^2} & \sigma \\ \sigma & \sqrt{1+\sigma^2}\end{bmatrix}\begin{bmatrix} \alpha & -\mathrm{i}\beta\mathrm{e}^{\mathrm{i}\theta} \\ \mathrm{i}\beta\mathrm{e}^{-\mathrm{i}\theta} & -\alpha\end{bmatrix}\begin{bmatrix} \sqrt{1+\sigma^2} & \sigma \\ \sigma & \sqrt{1+\sigma^2}\end{bmatrix}\right)\\
&= \inf_{\substack{\sigma\geq 0\\ \theta,\widehat{\theta}\in[0,2\pi)}} \beta\widehat{\beta}\left[(1+\sigma^2)(\mathrm{e}^{\mathrm{i}(\theta-\widehat{\theta})}+\mathrm{e}^{\mathrm{i}(\widehat{\theta}-\theta)}) - \sigma^2(\mathrm{e}^{\mathrm{i}(\theta+\widehat{\theta})}+\mathrm{e}^{-\mathrm{i}(\widehat{\theta}+\theta)})\right] + 2\alpha\widehat{\alpha}\\
&= \inf_{\substack{\sigma\geq 0\\ \theta,\widehat{\theta}\in[0,2\pi)}} 2\beta\widehat{\beta}\left[(1+\sigma^2)\cos(\theta-\widehat{\theta}) - \sigma^2\cos(\theta+\widehat{\theta})\right] + 2\alpha\widehat{\alpha}\\
&= \inf_{\substack{\sigma\geq 0\\ \theta,\widehat{\theta}\in[0,2\pi)}} 2\beta\widehat{\beta}\left[\cos(\theta-\widehat{\theta}) + 2\sigma^2\sin\theta\sin\widehat{\theta}\right] + 2\alpha\widehat{\alpha}\\
&= -\infty;
\end{aligned}$$

(ii) $(\widehat{\Lambda}, \widehat{J})$ is a block pair of type **T-c(1)** and (Λ, J) consists of two pairs of type **T-r(1)**. We have

$$\begin{aligned}
&\inf_{\substack{\Sigma\in\mathcal{D}^+\\ U_\pm,V_\pm\in\mathcal{P}^u}} \operatorname{tr}(\widehat{\Lambda}X^{\mathrm{H}}\Lambda X)\\
&= \inf_{\substack{\sigma\geq 0\\ \widehat{\theta}\in[0,2\pi)}} \operatorname{tr}\left(\begin{bmatrix} \widehat{\alpha} & -\mathrm{i}\widehat{\beta}\mathrm{e}^{\mathrm{i}\widehat{\theta}} \\ \mathrm{i}\widehat{\beta}\mathrm{e}^{-\mathrm{i}\widehat{\theta}} & -\widehat{\alpha}\end{bmatrix}\begin{bmatrix} \sqrt{1+\sigma^2} & \sigma \\ \sigma & \sqrt{1+\sigma^2}\end{bmatrix}\begin{bmatrix} \lambda_+ & \\ & -\lambda_-\end{bmatrix}\begin{bmatrix} \sqrt{1+\sigma^2} & \sigma \\ \sigma & \sqrt{1+\sigma^2}\end{bmatrix}\right)\\
&= \inf_{\substack{\sigma\geq 0\\ \widehat{\theta}\in[0,2\pi)}} (\lambda_+-\lambda_-)\widehat{\beta}\mathrm{i}(\mathrm{e}^{-\mathrm{i}\widehat{\theta}}-\mathrm{e}^{\mathrm{i}\widehat{\theta}})\sigma\sqrt{1+\sigma^2} + \widehat{\alpha}(\lambda_+ + \lambda_-)\\
&= \inf_{\substack{\sigma\geq 0\\ \widehat{\theta}\in[0,2\pi)}} 2(\lambda_+-\lambda_-)\widehat{\beta}\sigma\sqrt{1+\sigma^2}\sin\widehat{\theta} + \widehat{\alpha}(\lambda_+ + \lambda_-)\\
&= -\infty,
\end{aligned}$$

because $\lambda_+ \neq \lambda_-$; otherwise $\Lambda = \lambda_+ J$ which has been excluded from subcase (3) above;

(iii) (Λ, J) is a block pair of type **T-c(1)** and $(\widehat{\Lambda}, \widehat{J})$ consists of two pairs of type **T-r(1)**. This is similar to subcase (ii) we just considered, with the same conclusion: the infimum is $-\infty$.

Consider, in general, $n = \widehat{n} > 2$ and at least one block pair of type **T-c(1)** is contained in (Λ, J) or $(\widehat{\Lambda}, \widehat{J})$ or both. Suppose for the moment that $(\widehat{\Lambda}, \widehat{J})$ contains a block pair of type **T-c**(1) with $\widehat{\alpha}$, $\widehat{\beta} \in \mathbb{R}$ and $\widehat{\beta} > 0$. With the same reasoning we have employed in (4.5), picking $\varSigma = \sigma \boldsymbol{e}_1 \boldsymbol{e}_1^{\mathrm{T}}$ and suitable permutation matrices $V_{\pm}, U_{\pm}$ of apt sizes, we get

$$
\begin{aligned}
&\inf_{\substack{\varSigma \in \mathcal{D}^+ \\ U_{\pm}, V_{\pm} \in \mathcal{P}^u}} \operatorname{tr}(\widehat{\varLambda} X^{\mathrm{H}} \varLambda X) \\
&= \inf_{\substack{\varSigma \in \mathcal{D}^+ \\ U_{\pm}, V_{\pm} \in \mathcal{P}^u}} \operatorname{tr}\left(\begin{bmatrix} V_+ \widehat{\varLambda}_+^c V_+^{\mathrm{H}} & -\mathrm{i} V_+ \widehat{\varOmega}^c V_-^{\mathrm{H}} \\ \mathrm{i} V_- \widehat{\varOmega}^c V_+^{\mathrm{H}} & -V_- \widehat{\varLambda}_- V_-^{\mathrm{H}} \end{bmatrix} \begin{bmatrix} (I + \widetilde{\varSigma}\widetilde{\varSigma}^{\mathrm{H}})^{1/2} & \widetilde{\varSigma} \\ \widetilde{\varSigma} & (I + \widetilde{\varSigma}^{\mathrm{H}}\widetilde{\varSigma})^{1/2} \end{bmatrix} \times \right. \\
&\qquad \left. \begin{bmatrix} U_+^{\mathrm{H}} \varLambda_+^c U_+ & -\mathrm{i} U_+^{\mathrm{H}} \varOmega^c U_- \\ \mathrm{i} U_-^{\mathrm{H}} \varOmega^{cH} U_+ & -U_-^{\mathrm{H}} \varLambda_-^c U_- \end{bmatrix} \begin{bmatrix} (I + \widetilde{\varSigma}\widetilde{\varSigma}^{\mathrm{H}})^{1/2} & \widetilde{\varSigma} \\ \widetilde{\varSigma} & (I + \widetilde{\varSigma}^{\mathrm{H}}\widetilde{\varSigma})^{1/2} \end{bmatrix} \right) \\
&\le \inf_{\sigma > 0} \operatorname{tr}\left(\begin{bmatrix} \widehat{\alpha} & & -\mathrm{i}\widehat{\beta}\mathrm{e}^{\mathrm{i}\widehat{\theta}} & \\ & * & & \\ \mathrm{i}\widehat{\beta}\mathrm{e}^{-\mathrm{i}\widehat{\theta}} & & -\widehat{\alpha} & \\ & & & * \end{bmatrix} \begin{bmatrix} (1+\sigma^2)^{1/2} & & \sigma & \\ & I & & 0 \\ \sigma & & (1+\sigma^2)^{1/2} & \\ & 0 & & I \end{bmatrix} \times \right. \\
&\qquad \left. \begin{bmatrix} + & & + & \\ & * & & \\ + & & + & \\ & & & * \end{bmatrix} \begin{bmatrix} (1+\sigma^2)^{1/2} & & \sigma & \\ & I & & 0 \\ \sigma & & (1+\sigma^2)^{1/2} & \\ & 0 & & I \end{bmatrix} \right) \\
&\qquad\qquad \left(\text{where } \left[\begin{smallmatrix} + & + \\ + & + \end{smallmatrix}\right] \text{ is either } \begin{bmatrix} \alpha & -\mathrm{i}\beta\mathrm{e}^{\mathrm{i}\theta} \\ \mathrm{i}\beta\mathrm{e}^{-\mathrm{i}\theta} & -\alpha \end{bmatrix} \text{ or } \begin{bmatrix} \lambda_i & \\ & -\lambda_j \end{bmatrix} \right) \\
&= -\infty.
\end{aligned}
$$

In summary, the infimum is $-\infty$ as long as block pairs of type **T-c**(1) are involved, while if only **T-r**(1) block pairs are involved, it is turned into the case already considered in section 4.

5.2. Involving block pairs of types **T-r**(1)*,* **T-c**(1)*, and* **T-∞**(1)

It suffices to consider the case that at least one pair of type **T-∞**(1) is contained in $(\varLambda, J)$ because the case of involving block pairs of types **T-r**(1) and **T-c**(1) has already been dealt with in subsection 5.1 and our discussion prior to the subsection excludes any possibility that $(\widehat{\varLambda}, \widehat{J})$ may contain any block pair of type **T-∞**(1). We write

$$
\varLambda = \begin{bmatrix} \varLambda_r & \\ & \varLambda_\infty \end{bmatrix}, \quad J = \begin{bmatrix} J_r & \\ & 0 \end{bmatrix}, \quad X = \begin{bmatrix} X_r \\ X_\infty \end{bmatrix}, \tag{5.5}
$$

where $\varLambda_\infty$ is diagonal with diagonal entries ± 1, to get

$$
\begin{aligned}
\inf_{X^{\mathrm{H}} J X = \widehat{J}} \operatorname{tr}(\widehat{\varLambda} X^{\mathrm{H}} \varLambda X) &= \inf_{X_r^{\mathrm{H}} J_r X_r = \widehat{J}} \operatorname{tr}\left(\widehat{\varLambda}[X_r^{\mathrm{H}} \varLambda_r X_r + X_\infty^{\mathrm{H}} \varLambda_\infty X_\infty]\right) \\
&= \inf_{X_r^{\mathrm{H}} J_r X_r = \widehat{J}} \operatorname{tr}(\widehat{\varLambda} X_r^{\mathrm{H}} \varLambda_r X_r) + \inf_{X_\infty} \operatorname{tr}(\widehat{\varLambda} X_\infty^{\mathrm{H}} \varLambda_\infty X_\infty).
\end{aligned} \tag{5.6}
$$

Consider the second term in (5.6), which is an infimum over X_∞ without any constraint. Without loss of generality, we may assume that $\widehat{\varLambda}$ is real diagonal; otherwise, since $\widehat{\Lambda}$ is Hermitian, we let $\widehat{\Lambda} = Q\widetilde{\Lambda}Q^{\mathrm{H}}$ where Q is an orthogonal matrix and $\widetilde{\Lambda}$ is diagonal, and we get

$$\mathrm{tr}(\widehat{\Lambda}X_\infty^{\mathrm{H}}\Lambda_\infty X_\infty) = \mathrm{tr}(Q\widetilde{\Lambda}Q^{\mathrm{H}}X_\infty^{\mathrm{H}}\Lambda_\infty X_\infty) = \mathrm{tr}(\widetilde{\Lambda}(X_\infty Q)^{\mathrm{H}}\Lambda_\infty(X_\infty Q)).$$

Let $\widehat{\Lambda} = \mathrm{diag}(\widehat{\lambda}_1, \dots, \widehat{\lambda}_{\widehat{n}})$ and $\Lambda_\infty = \mathrm{diag}(\lambda_{\infty,1}, \dots, \lambda_{\infty,t})$ where $t \geq 1$. We have

$$\inf_{X_\infty} \mathrm{tr}(\widehat{\Lambda}X_\infty^{\mathrm{H}}\Lambda_\infty X_\infty) = \inf_{X_\infty} \sum_{i,j} \widehat{\lambda}_j \lambda_{\infty,i}|x_{\infty;ij}|^2,$$

where we have written $X_\infty = [x_{\infty;ij}]$. Since X_∞ is arbitrary, each $|x_{\infty;ij}|^2 \geq 0$ can be made 0 or arbitrarily large. Hence

$$\inf_{X_\infty} \mathrm{tr}(\widehat{\Lambda}X_\infty^{\mathrm{H}}\Lambda_\infty X_\infty) > -\infty \quad \text{if and only if all } \widehat{\lambda}_j \lambda_{\infty,i} \geq 0,$$

in which case, the infimum is 0. Notice that $\lambda_{\infty,i} = \pm 1$. There are three possible situations for all $\widehat{\lambda}_j \lambda_{\infty,i} \geq 0$:

1. all $\widehat{\lambda}_j = 0$ if both ± 1 appear among all $\lambda_{\infty,i}$;
2. all $\widehat{\lambda}_j \geq 0$ if all $\lambda_{\infty,i} = 1$;
3. all $\widehat{\lambda}_j \leq 0$ if all $\lambda_{\infty,i} = -1$.

The first situation is not allowed because it implies $\widehat{\Lambda} = 0$ and hence $\widehat{A} = 0$ which is excluded to begin with. Therefore, we conclude

$$\inf_{X_\infty} \mathrm{tr}(\widehat{\Lambda}X_\infty^{\mathrm{H}}\Lambda_\infty X_\infty) > -\infty \Leftrightarrow \text{ either } \widehat{\Lambda} \succeq 0,\ \Lambda_\infty = I, \text{ or } \widehat{\Lambda} \preceq 0,\ \Lambda_\infty = -I. \tag{5.7}$$

Consider now the first term in (5.6), which falls into the case in subsection 5.1. In light of (5.7), to see when

$$\inf_{X_r^{\mathrm{H}} J_r X_r = \widehat{J}} \mathrm{tr}(\widehat{\Lambda}X_r^{\mathrm{H}}\Lambda_r X_r) > -\infty \tag{5.8}$$

and what the infimum is, it suffices to investigate what will happen when either $\widehat{\Lambda} \succeq 0$, $\Lambda_\infty = I$, or $\widehat{\Lambda} \preceq 0$, $\Lambda_\infty = -I$. We have the following:

1. Suppose $\widehat{\Lambda} \succeq 0$, $\Lambda_\infty = I$. Then $(\widehat{\Lambda}, \widehat{J}) \succeq 0$ and, by the result of subsection 5.1, (5.8) holds if and only if $(\Lambda_r, J_r) \succeq 0$ and $(J_r, \widehat{\Lambda}, \widehat{J})$ is proper, which is the same as that $(\Lambda, J) \succeq 0$ and $(J, \widehat{\Lambda}, \widehat{J})$ is proper because of (5.5) and $\Lambda_\infty = I$.
2. Suppose $\widehat{\Lambda} \preceq 0$, $\Lambda_\infty = -I$. Then $(\widehat{\Lambda}, \widehat{J}) \preceq 0$ and, by the result of subsection 5.1, (5.8) holds if and only if $(\Lambda_r, J_r) \preceq 0$ and $(-J_r, -\widehat{\Lambda}, -\widehat{J})$ is proper, which is the same as that $(\Lambda, J) \preceq 0$ and $(-J, -\widehat{\Lambda}, -\widehat{J})$ is proper because of (5.5) and $\Lambda_\infty = -I$.

5.3. Involving block pairs of types $\mathbf{T\text{-}r}(p)$ with $p \le 2$, $\mathbf{T\text{-}c}(1)$, and $\mathbf{T\text{-}\infty}(1)$

It suffices to consider there are some block pairs of type $\mathbf{T\text{-}r}(2)$ in the mix; otherwise the situation has already been taken care of in subsection 5.2. Let $\varepsilon > 0$ be arbitrary tiny, and perturb each block pair $(\eta K_2(\alpha), \eta F_2)$ of type $\mathbf{T\text{-}r}(2)$ according to

$$K_2(\alpha) \to K_2(\alpha) + \varepsilon \boldsymbol{e}_1 \boldsymbol{e}_1^{\mathrm{T}}, \tag{5.9}$$

which turns the block pair $(\eta K_2(\alpha), \eta F_2)$ to two block pairs $\mathbf{T\text{-}r}(1)$ with eigenvalues $\alpha \pm \sqrt{\varepsilon}$, respectively, and both are continuous in ε and go to α as $\varepsilon \to 0^+$. As a result, both $\varLambda$ and $\widehat{\varLambda}$ are possibly perturbed to $\varLambda_\varepsilon$ and $\widehat{\varLambda}_\varepsilon$, satisfying

$$\varLambda_\varepsilon \begin{cases} \equiv \varLambda, & \text{if no block pair of type } \mathbf{T\text{-}r}(2) \text{ in } (\varLambda, J), \\ \to \varLambda, & \text{as } \varepsilon \to 0^+. \end{cases}$$

The same holds true for $(\widehat{\varLambda}_\varepsilon, \widehat{J})$. Consider now $(\varLambda_\varepsilon, J)$ and $(\widehat{\varLambda}_\varepsilon, \widehat{J})$ in which only block pairs of types $\mathbf{T\text{-}r}(1)$, $\mathbf{T\text{-}c}(1)$, and $\mathbf{T\text{-}\infty}(1)$ are possibly involved. It is important to note that both J and $\widehat{J}$ are not perturbed, leaving $\mathfrak{i}_\pm(J)$ and $\mathfrak{i}_\pm(\widehat{J})$ unaffected. Note that, for any $\alpha \in \mathbb{R}$, $\big(K_2(\alpha), F_2\big) \succeq 0$ but $\big(K_2(\alpha), F_2\big) \not\succ 0$.

Lemma 5.4. *Given $\varepsilon > 0$, $\big(K_2(\alpha) + \varepsilon \boldsymbol{e}_1 \boldsymbol{e}_1^{\mathrm{T}}\big) - \lambda_0 F_2 \succeq 0$ if and only if $\alpha - \sqrt{\varepsilon} \le \lambda_0 \le \alpha + \sqrt{\varepsilon}$.*

Proof. Notice that

$$K_2(\alpha) + \varepsilon \boldsymbol{e}_1 \boldsymbol{e}_1^{\mathrm{T}} - \lambda_0 F_2 = \begin{bmatrix} \varepsilon & \alpha - \lambda_0 \\ \alpha - \lambda_0 & 1 \end{bmatrix}.$$

Since $\varepsilon > 0$, the matrix is positive semidefinite if and only if its determinant $\varepsilon - (\alpha - \lambda_0)^2 \ge 0$. □

The next lemma is stated in terms of $(\varLambda, J)$. It is clearly valid if $(\varLambda, J)$ is replaced with $(\widehat{\varLambda}, \widehat{J})$.

Lemma 5.5. *Suppose that $(\varLambda, J)$ is a direct sum of block pairs of types $\mathbf{T\text{-}r}(p)$ with $p \le 2$, $\mathbf{T\text{-}c}(1)$, and $\mathbf{T\text{-}\infty}(1)$ and that each block pair of type $\mathbf{T\text{-}r}(2)$ is perturbed according to (5.9) where $\varepsilon > 0$.*

(a) *If there is a positive sequence $\{\varepsilon_i\}_{i=1}^\infty$ converging to 0, i.e., $0 < \varepsilon_i \to 0$ as $i \to \infty$, such that $(\varLambda_{\varepsilon_i}, J) \succeq 0$ for all i, then $(\varLambda, J) \succeq 0$, and $(\varLambda, J)$ can only contain block pairs $\big(\eta K_2(\alpha), \eta F_2\big)$ with $\eta = 1$ and the same α for all block pairs of type $\mathbf{T\text{-}r}(2)$, in which case $\varLambda - \lambda_0 J \succeq 0$ with $\lambda_0 = \alpha$ and only with $\lambda_0 = \alpha$.*
Conversely, if $(\varLambda, J) \succeq 0$, then $(\varLambda_\varepsilon, J) \succeq 0$.

(b) *If there is a positive sequence* $\{\varepsilon_i\}_{i=1}^{\infty}$ *converging to* 0*, i.e.,* $0 < \varepsilon_i \to 0$ *as* $i \to \infty$*, such that* $(\Lambda_{\varepsilon_i}, J) \preceq 0$ *for all* i*, then* $(\Lambda, J) \preceq 0$*, and* (Λ, J) *can only contain block pairs* $\big(\eta K_2(\alpha), \eta F_2\big)$ *with* $\eta = -1$ *and the same* α *for all block pairs of type* **T-r(2)***, in which case* $\Lambda - \lambda_0 J \preceq 0$ *with* $\lambda_0 = \alpha$ *and only with* $\lambda_0 = \alpha$*.*
Conversely, if $(\Lambda, J) \preceq 0$*, then* $(\Lambda_\varepsilon, J) \preceq 0$*.*

Proof. We will only prove item (a). The same argument with minor modifications can be used to prove item (b).

Suppose that $(\Lambda_{\varepsilon_i}, J) \succeq 0$ for all i, which means that for each i there exists μ_i such that $\Lambda_{\varepsilon_i} - \mu_i J \succeq 0$. By [15, Lemma 3.8], $|\mu_i|$ can be taken no bigger than the absolute values of the finite eigenvalues of $(\Lambda_{\varepsilon_i}, J)$. Under the perturbation, the finite eigenvalues of matrix pairs $(\Lambda_{\varepsilon_i}, J)$ are uniformly bounded because they converge to the finite eigenvalues of (Λ, J). Hence $\{\mu_i\}_{i=1}^{\infty}$ is bounded and thus has a convergent subsequence $\{\mu_i\}_{i\in\mathbb{I}}$, say converging to λ_0, where $\mathbb{I}$ is an infinite subset of $\{1, 2, \dots, \}$. Letting $\mathbb{I} \ni i \to \infty$ in $\Lambda_{\varepsilon_i} - \mu_i J \succeq 0$ yields $\Lambda - \lambda_0 J \succeq 0$.

If (Λ, J) ever contains a block pair $\big(\eta K_2(\alpha), \eta F_2\big)$, then we will have

$$\eta(K_2(\alpha) + \varepsilon_i \boldsymbol{e}_1 \boldsymbol{e}_1^{\mathrm{T}} - \mu_i F_2) = \eta \begin{bmatrix} \varepsilon_i & \alpha - \mu_i \\ \alpha - \mu_i & 1 \end{bmatrix} \succeq 0$$

for all i, which implies $\eta = 1$ and $\alpha - \sqrt{\varepsilon_i} \le \mu_i \le \alpha + \sqrt{\varepsilon_i}$. Letting $i \to \infty$ yields $\mu_i \to \alpha$. If (Λ, J) also contains another block pair $\big(\widetilde{\eta} K_2(\widetilde{\alpha}), \widetilde{\eta} F_2\big)$ of the same type. Using the same argument as we just did, we find $\widetilde{\eta} = 1$ and also $\mu_i \to \widetilde{\alpha}$ yielding $\widetilde{\alpha} = \alpha$.

Conversely, suppose that $(\Lambda, J) \succeq 0$. If no block pair of type **T-r(2)** is involved in (Λ, J), then $\Lambda_\varepsilon \equiv \Lambda$ and hence no proof is necessary. If (Λ, J) does contain block pairs of **T-r(2)**, then these block pairs must be $\big(K_2(\alpha), F_2\big)$ with the same α. Therefore the only λ_0 that makes $\Lambda - \lambda_0 J \succeq 0$ is $\lambda_0 = \alpha$ which also makes $\big(K_2(\alpha) + \varepsilon \boldsymbol{e}_1 \boldsymbol{e}_1^{\mathrm{T}}\big) - \lambda_0 F_2 \succeq 0$ for any $\varepsilon > 0$. By the way how Λ is perturbed to Λ_ε, we find $\Lambda_\varepsilon - \lambda_0 J \succeq 0$. □

By the results of subsections 5.1 and 5.2, we conclude that

$$\inf_{X^{\mathrm{H}} J X = \widehat{J}} \operatorname{tr}(\widehat{\Lambda}_\varepsilon X^{\mathrm{H}} \Lambda_\varepsilon X) > -\infty \tag{5.10}$$

if and only if one of the following two scenarios occurs:

(1) both (Λ_ε, J) and $(\widehat{\Lambda}_\varepsilon, \widehat{J})$ are positive semidefinite pairs and $(J, \widehat{\Lambda}_\varepsilon, \widehat{J})$ is proper;
(2) both (Λ_ε, J) and $(\widehat{\Lambda}_\varepsilon, \widehat{J})$ are negative semidefinite pairs and $(-J, -\widehat{\Lambda}_\varepsilon, -\widehat{J})$ is proper.

Let $\{\varepsilon_i\}_{i=1}^{\infty}$ be a positive sequence that converges to 0, i.e., $0 < \varepsilon_i \to 0$ as $i \to \infty$. Since there are only two scenarios here, there is a subsequence $\{\varepsilon_i\}_{i\in\mathbb{I}}$ such that one of the two scenarios holds true for all $i \in \mathbb{I}$. In the case when

$$\text{for all } i \in \mathbb{I}, \text{ both } (\Lambda_{\varepsilon_i}, J) \succeq 0,\ (\widehat{\Lambda}_{\varepsilon_i}, \widehat{J}) \succeq 0, \text{ and } (J, \widehat{\Lambda}_{\varepsilon_i}, \widehat{J}) \text{ is proper,}$$

we have both $(\Lambda, J) \succeq 0$, $(\widehat{\Lambda}, \widehat{J}) \succeq 0$, and $(J, \widehat{\Lambda}, \widehat{J})$ is proper, as a consequence of Lemma 5.5. Similarly, we can conclude that if

$$\text{for all } i \in \mathbb{I}, \text{ both } (\Lambda_{\varepsilon_i}, J) \preceq 0,\ (\widehat{\Lambda}_{\varepsilon_i}, \widehat{J}) \preceq 0, \text{ and } (-J, -\widehat{\Lambda}_{\varepsilon_i}, -\widehat{J}) \text{ is proper,}$$

then both $(\Lambda, J) \preceq 0$, $(\widehat{\Lambda}, \widehat{J}) \preceq 0$, and $(-J, -\widehat{\Lambda}, -\widehat{J})$ is proper.

With either scenario, the infimum in (5.10) has a closed formula as in (3.3), or it applied to $(-\Lambda_\varepsilon, -J)$ and $(-\widehat{\Lambda}_\varepsilon, -\widehat{J})$. Because of the continuity of these eigenvalues with respect to ε, the limit of the infimum exists as $\varepsilon \to 0^+$. Since the perturbation does not affect $\mathfrak{i}_\pm(J)$ and $\mathfrak{i}_\pm(\widehat{J})$ at all, the limit takes the same form as (3.3), or it applied to $(-\Lambda, -J)$ and $(-\widehat{\Lambda}, -\widehat{J})$.

5.4. Involving block pairs of all possible types in (5.2)

In this subsection, we will allow all block pairs of types in (5.2) to possibly appear in (Λ, J) and $(\widehat{\Lambda}, \widehat{J})$. Block pairs of types in

type	**T-s**$(2p+1)$	**T-c**(p)	**T-r**(p)	**T-**$\infty(p)$
p	$p \geq 1$	$p \geq 2$	$p \geq 3$	$p \geq 2$

(5.11)

remain to be included for considerations, as we have already considered **T-r**(p) with $p \leq 2$, **T-c**(1), and **T-**$\infty(1)$,

Notice that a positive/negative semidefinite matrix pair does not contain any block pair of these types in (5.11) in its canonical form (see Remark 5.1). In what follows, we will show that

$$\inf_{X^{\mathrm{H}} J X = \widehat{J}} \operatorname{tr}(\widehat{\Lambda} X^{\mathrm{H}} \Lambda X) = -\infty \tag{5.12}$$

if any block pair of these types in (5.11) is contained in either (Λ, J) or $(\widehat{\Lambda}, \widehat{J})$ or both, besides **T-r**(p) with $p \leq 2$, **T-c**(1), and **T-**$\infty(1)$. The idea is to perturb (Λ, J) and/or $(\widehat{\Lambda}, \widehat{J})$ to (Λ_ε, J) and/or $(\widehat{\Lambda}_\varepsilon, \widehat{J})$ such that

1. $\Lambda_\varepsilon \to \Lambda$ and $\widehat{\Lambda}_\varepsilon \to \widehat{\Lambda}$ as $\varepsilon \to 0$,
2. for sufficiently tiny $\varepsilon > 0$, the canonical forms of (Λ_ε, J) and $(\widehat{\Lambda}_\varepsilon, \widehat{J})$ contain block pairs of types **T-r**(1), **T-c**(1), and **T-**$\infty(1)$ only, and that has been investigated in subsection 5.2, and either

$$\inf_{X^{\mathrm{H}} J X = \widehat{J}} \operatorname{tr}(\widehat{\Lambda}_\varepsilon X^{\mathrm{H}} \Lambda_\varepsilon X) = -\infty, \tag{5.13}$$

or

$$\lim_{\varepsilon \to 0} \inf_{X^{\mathrm{H}} J X = \widehat{J}} \operatorname{tr}(\widehat{\Lambda}_\varepsilon X^{\mathrm{H}} \Lambda_\varepsilon X) = -\infty.$$

Hence, we justify our claim (5.12) for the case of interest.

Specifically, we perturb the *first* block elements in block pairs of the types in (5.11) as follows:

$$K_{2p+1}(0) = \begin{bmatrix} & & K_p(0) \\ & 0 & \boldsymbol{e}_1^{\mathrm{T}} \\ K_p(0) & \boldsymbol{e}_1 & \end{bmatrix} \to \begin{bmatrix} & & K_p(\mathrm{i}\varepsilon) \\ & \varepsilon & \boldsymbol{e}_1^{\mathrm{T}} \\ K_p(-\mathrm{i}\varepsilon) & \boldsymbol{e}_1 & \end{bmatrix},$$

$$\begin{bmatrix} 0 & K_p(\alpha+\mathrm{i}\beta) \\ K_p(\alpha+\mathrm{i}\beta) & \end{bmatrix} \to \begin{bmatrix} 0 & K_p(\alpha+\mathrm{i}\beta)+\varepsilon\boldsymbol{e}_1\boldsymbol{e}_1^{\mathrm{T}} \\ K_p(\alpha+\mathrm{i}\beta)+\varepsilon\boldsymbol{e}_1\boldsymbol{e}_1^{\mathrm{T}} & \end{bmatrix},$$

$$F_p \to F_p + \varepsilon\boldsymbol{e}_1\boldsymbol{e}_1^{\mathrm{T}},$$

$$K_p(\alpha) \to K_p(\alpha) + \varepsilon\boldsymbol{e}_1\boldsymbol{e}_1^{\mathrm{T}}.$$

We restrict $\varepsilon > 0$, except for **T-**$\boldsymbol{\infty}$**(2)**, for which $\varepsilon < 0$ is also allowed.

(i) A block pair of type **T-c(**p**)** with $p \geq 2$ and eigenvalues $\alpha \pm \mathrm{i}\beta$ generates p block pairs of type **T-c(**1**)** with eigenvalues

$$\alpha \pm \mathrm{i}\beta \pm \varepsilon^{1/p} \exp(\mathrm{i}\frac{2\pi j}{p}), j = 0, \dots, p-1.$$

Among them there are conjugate complex eigenvalues.

(ii) A block pair of type **T-s(**$2p+1$**)** with $p \geq 1$ generates a block pair of type **T-**$\boldsymbol{\infty}$**(**1**)** and a block pair of type **T-c(**p**)** with a pair of conjugate complex eigenvalues, and eventually generates a block pair of type **T-**$\boldsymbol{\infty}$**(**1**)** and $2p$ block pairs of type **T-c(**1**)** with eigenvalues some of which are conjugate complex eigenvalues.

(iii) A block pair of type **T-r(**p**)** with $p \geq 3$ and eigenvalues α generates p block pairs of type **T-c(**1**)** or **T-r(**1**)** with eigenvalues

$$\alpha + \varepsilon^{1/p} \exp(\mathrm{i}\frac{2\pi j}{p}), j = 0, \dots, p-1.$$

Among them there are conjugate complex eigenvalues.

(iv) A block pair of type **T-**$\boldsymbol{\infty}$**(**p**)** with $p \geq 2$ generates a block pair of type **T-**$\boldsymbol{\infty}$**(**1**)** and $p-1$ block pairs of type **T-c(**1**)** or **T-r(**1**)** with eigenvalues

$$\varepsilon^{-1/(p-1)} \exp(\mathrm{i}\frac{2\pi j}{p-1}), j = 0, \dots, p-2.$$

Among them there are conjugate complex eigenvalues if $p \geq 3$.

After perturbations, (Λ_ε, J) and $(\widehat{\Lambda}_\varepsilon, \widehat{J})$ themselves are no longer in their canonical forms as the ones specified in Lemma 5.1. But they can be turned into their canonical forms, in which only possible block pairs of types **T-c(**1**)**, **T-r(**1**)**, and **T-**$\boldsymbol{\infty}$**(**1**)** show up. When any one of (i), (ii), (iii), and (iv) with $p > 2$ occurs, we will have at least one block pair of type

T-c(1) in the canonical forms, and hence (5.13) holds by the results in subsection 5.2, which implies (5.12).

It remains to consider (iv) with $p = 2$ and only block pairs of type **T-∞(2)**, besides **T-r**(p) with $p \leq 2$ and **T-∞(1)**, can show up. We exclude any block pair of type **T-c(1)** because if such a block pair exists, we will have, after perturbations, (5.13) and hence (5.12). Note that block pair of type **T-∞(2)** can only be contained in (Λ, J) according to (5.2), while $(\widehat{\Lambda}, \widehat{J})$ contains possibly block pairs of type **T-r**(p) with $p \leq 2$. Without needing to perturb any block pair of type **T-r(2)** in $(\widehat{\Lambda}, \widehat{J})$, if any, Lemma 5.6 below shows that (5.12) holds.

Lemma 5.6. *If* (Λ, J) *contains a block pair of type* **T-∞(2)***, then* (5.12) *holds.*

Proof. We perturb any block pair of type **T-∞(2)** in (Λ, J) as

$$\eta(F_2, K_2(0)) \to \eta\left(\begin{bmatrix} \varepsilon & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}\right)$$
$$\sim \eta\left(\begin{bmatrix} \varepsilon & 0 \\ 0 & -\frac{1}{\varepsilon} \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}\right) \sim \left(\begin{bmatrix} \operatorname{sign}(\eta\varepsilon) & 0 \\ 0 & -\frac{1}{\varepsilon} \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & \eta \end{bmatrix}\right),$$

where "∼"stands for "is congruent to". Without loss of generality, we may assume that $\eta(F_2, K_2(0))$ is the last block pair in (Λ, J). As a result,

$$\begin{aligned}
(\Lambda, J) &= \left(\begin{bmatrix} \Lambda_r & 0 \\ 0 & \eta F_2 \end{bmatrix}, \begin{bmatrix} J_r & 0 \\ 0 & \eta K_2(0) \end{bmatrix}\right) \\
\to (\Lambda_\varepsilon, J) &= \left(\begin{bmatrix} \Lambda_r & 0 \\ 0 & \eta(F_2 + \varepsilon \boldsymbol{e}_1 \boldsymbol{e}_1^{\mathrm{T}}) \end{bmatrix}, \begin{bmatrix} J_r & 0 \\ 0 & \eta K_2(0) \end{bmatrix}\right) \\
\sim (\underline{\Lambda}_\varepsilon, \underline{J}) &= \left(\begin{bmatrix} \Lambda_r & 0 & 0 \\ 0 & -\frac{1}{\varepsilon} & 0 \\ 0 & 0 & \operatorname{sign}(\eta\varepsilon) \end{bmatrix}, \begin{bmatrix} J_r & 0 & 0 \\ 0 & \eta & 0 \\ 0 & 0 & 0 \end{bmatrix}\right) \\
&=: \left(\begin{bmatrix} \underline{\Lambda}_{\varepsilon;r} & 0 \\ 0 & \operatorname{sign}(\eta\varepsilon) \end{bmatrix}, \begin{bmatrix} \underline{J}_r & 0 \\ 0 & 0 \end{bmatrix}\right),
\end{aligned}$$

the canonical form of (Λ_ε, J). Similarly to (5.6), we have

$$\begin{aligned}
&\inf_{X^{\mathrm{H}} J X = \widehat{J}} \operatorname{tr}(\widehat{\Lambda} X^{\mathrm{H}} \Lambda_\varepsilon X) \\
&= \inf_{X^{\mathrm{H}} \underline{J} X = \widehat{J}} \operatorname{tr}(\widehat{\Lambda} X^{\mathrm{H}} \underline{\Lambda}_\varepsilon X) \\
&= \inf_{X_r^{\mathrm{H}} \underline{J}_r X_r = \widehat{J}} \operatorname{tr}(\widehat{\Lambda} X_r^{\mathrm{H}} \underline{\Lambda}_{\varepsilon;r} X_r) + \inf_{X_\infty} \operatorname{tr}(\widehat{\Lambda} X_\infty^{\mathrm{H}} \operatorname{sign}(\eta\varepsilon) X_\infty).
\end{aligned} \tag{5.14}$$

As in our argument after (5.6), we will consider the last infimum in (5.14). For that purpose we may assume $\widehat{\Lambda}$ is diagonal. Because the freedom in making either $\varepsilon > 0$ or

$\varepsilon < 0$, we can show the infimum over X_∞ is $-\infty$, unless $\widehat{\Lambda} = 0$ which is excluded in Theorem 3.1. Hence we have (5.13) and hence (5.12). □

Summarizing what we have done so far leads to the main result in Theorem 3.1.

Remark 5.2. So far, we have been assumed that B is indefinite. We now comment on the proof for the case when B is positive or negative semidefinite. It suffices to consider the case $B \succeq 0$, because when $B \preceq 0$, we can consider the infimum of interest for $(-A, -B)$ and $(-\widehat{A}, -\widehat{B})$, instead. Suppose that $B \succeq 0$. Then $\widehat{B} \succ 0$ because $\widehat{B}$ is always nonsingular and $\mathfrak{i}_+(\widehat{B}) \le \mathfrak{i}_+(B)$ and $\mathfrak{i}_-(\widehat{B}) \le \mathfrak{i}_-(B)$ by (3.2). We again transform matrix pairs (A, B) and $(\widehat{A}, \widehat{B})$ to their canonical forms as in Lemma 5.1. We will still have (5.1) but with fewer possible types of block pairs to consider in (Λ, J) and $(\widehat{\Lambda}, \widehat{J})$ than those in (5.2). Specifically,

$$(\Lambda, J): \ \mathbf{T\text{-}\infty(}p\mathbf{)} \text{ with } p \le 2, \ \mathbf{T\text{-}r(}1\mathbf{)}; \quad (\widehat{\Lambda}, \widehat{J}): \ \mathbf{T\text{-}r(}1\mathbf{)}.$$

Also $\widehat{J} = I_{\widehat{n}}$ always. If no block pair of type **T-∞(2)** shows up in (Λ, J), then it falls into a special situation of subsection 5.2 where, though under the scope of B being indefinite, no argument there relies on that. If, however, **T-∞(2)** is involved, we can use Lemma 5.6.

6. Concluding remarks

We have established a trace minimization principle for two Hermitian matrix pairs (A, B) and $(\widehat{A}, \widehat{B})$:

$$\inf_{\widehat{B}X^{\mathrm{H}}BX = I_{\widehat{n}}} \operatorname{tr}(\widehat{A}X^{\mathrm{H}}AX), \tag{6.1}$$

where A, $B \in \mathbb{C}^{n\times n}$ and $\widehat{A}$, $\widehat{B} \in \mathbb{C}^{\widehat{n}\times\widehat{n}}$ are all Hermitian, and $\widehat{n} \le n$. It is the most general one up to date, encompassing Fan's trace minimization principle [12] (for $\widehat{A} = \widehat{B} = I_{\widehat{n}}$ and $B = I_n$) and its straightforward extension (for $\widehat{A} = \widehat{B} = I_{\widehat{n}}$ and positive definite B), and most recent ones [14,15,18] reviewed in section 1. In those recent investigations, the notion of positive semidefinite matrix pair was introduced: a Hermitian matrix pair (A, B) is *positive* (*negative*) *semidefinite* if there exists $\lambda_0 \in \mathbb{R}$ such that $A - \lambda_0 B$ is positive (negative) semidefinite.

For investigating (6.1), we introduced yet another notion for a Hermitian matrix triplet $(B, \widehat{A}, \widehat{B})$ being *proper* in Definition 3.1. We showed that the infimum in (6.1) is finite if and only if either both (A, B) and $(\widehat{A}, \widehat{B})$ are positive semidefinite pairs and $(B, \widehat{A}, \widehat{B})$ is proper, or both are negative semidefinite pairs and $(-B, -\widehat{A}, -\widehat{B})$ is proper, assuming $\widehat{A} \ne 0$, $A \ne \mu B$ for any $\mu \in \mathbb{R}$, and $\widehat{A} \ne \widehat{\mu}\widehat{B}$ for any $\widehat{\mu} \in \mathbb{R}$ when $n = \widehat{n}$. A closed formula for the infimum is given in terms of the finite eigenvalues of the two semidefinite matrix pairs.

In [18, Example 3.1], the following example (in the notation in this paper):

$$\mu = 2,\ A = \begin{bmatrix} 1 & \\ & \mu \end{bmatrix},\ B = \begin{bmatrix} 1 & \\ & -1 \end{bmatrix},\ \widehat{B} = B,$$

$$\sigma = \frac{\sqrt{18 - 6\sqrt{2}}}{6},\ \Omega = \begin{bmatrix} 1 & \\ & 1/4 \end{bmatrix},\ Q = \begin{bmatrix} \sqrt{1-\sigma^2} & -\sigma \\ \sigma & \sqrt{1-\sigma^2} \end{bmatrix},\ \widehat{A} = Q^{\mathrm{H}}\Omega Q,$$

was given to demonstrate that the infimum in (6.1) may not be any sum of the products between the eigenvalues of $\widehat{A}$ and some of the ones of (A, B), as a justification for an assumption of [18, Theorem 3.2]. This now can be well explained by our Theorem 3.1 in this paper, i.e., it is the eigenvalues of $(\widehat{A}, \widehat{B})$, not $\widehat{A}$ alone, that should appear in the infimum. For the example, the eigenvalues of $(\widehat{A}, \widehat{B})$ and of (A, B) are

$$\widehat{\lambda}_1^+ = \frac{1}{2}\sqrt{2},\ \widehat{\lambda}_1^- = -\frac{1}{4}\sqrt{2}, \quad \text{and} \quad \lambda_1^+ = 1,\ \lambda_1^- = -2,$$

respectively. Hence the infimum, by Theorem 3.1, is $\widehat{\lambda}_1^+\lambda_1^+ + \widehat{\lambda}_1^-\lambda_1^- = \sqrt{2}$.

Declaration of competing interest

None.

Data availability

No data was used for the research described in the article.

Acknowledgements

X. Liang is supported in part by NSFC 12371380 and 12071332. R.-C. Li is supported in part by NSF DMS-2009689.

References

[1] Z. Bai, R.-C. Li, Minimization principle for linear response eigenvalue problem, I: theory, SIAM J. Matrix Anal. Appl. 33 (4) (2012) 1075–1100.
[2] Z. Bai, R.-C. Li, Minimization principles for the linear response eigenvalue problem II: computation, SIAM J. Matrix Anal. Appl. 34 (2) (2013) 392–416.
[3] Z. Bai, R.-C. Li, Minimization principles and computation for the generalized linear response eigenvalue problem, BIT Numer. Math. 54 (1) (2014) 31–54.
[4] Z. Bai, R.-C. Li, W.-W. Lin, Linear response eigenvalue problem solved by extended locally optimal preconditioned conjugate gradient methods, Sci. China Math. 59 (8) (2016) 1443–1460.
[5] A.V. Knyazev, Toward the optimal preconditioned eigensolver: locally optimal block preconditioned conjugate gradient method, SIAM J. Sci. Comput. 23 (2) (2001) 517–541.
[6] D. Kressner, M.M. Pandur, M. Shao, An indefinite variant of LOBPCG for definite matrix pencils, Numer. Algorithms (2013) 1–23.
[7] R.-C. Li, Accuracy of computed eigenvectors via optimizing a Rayleigh quotient, BIT 44 (3) (2004) 585–593.

[8] R.-C. Li, Rayleigh quotient based optimization methods for eigenvalue problems, in: Z. Bai, W. Gao, Y. Su (Eds.), Matrix Functions and Matrix Equations, in: Series in Contemporary Applied Mathematics, vol. 19, World Scientific, Singapore, 2015, pp. 76–108.
[9] X. Liang, R.-C. Li, The hyperbolic quadratic eigenvalue problem, Forum Math. Sigma 3 (2015) 1–93, https://doi.org/10.1017/fms.2015.14.
[10] A.H. Sameh, J.A. Wisniewski, A trace, minimization algorithm for the generalized eigenvalue problem, SIAM J. Numer. Anal. 19 (6) (1982) 1243–1259.
[11] N.T. Son, P.-A. Absil, B. Gao, T. Stykel, Computing symplectic eigenpairs of symmetric positive-definite matrices via trace minimization and Riemannian optimization, SIAM J. Matrix Anal. Appl. 42 (4) (2021) 1732–1757, https://doi.org/10.1137/21M1390621.
[12] K. Fan, On a theorem of Weyl concerning eigenvalues of linear transformations. I, Proc. Natl. Acad. Sci. USA 35 (11) (1949) 652–655.
[13] R.A. Horn, C.R. Johnson, Matrix Analysis, 2nd edition, Cambridge University Press, New York, NY, 2013.
[14] J. Kovač-Striko, K. Veselić, Trace minimization and definiteness of symmetric pencils, Linear Algebra Appl. 216 (1995) 139–158.
[15] X. Liang, R.-C. Li, Z. Bai, Trace minimization principles for positive semi-definite pencils, Linear Algebra Appl. 438 (2013) 3085–3106.
[16] X. Liang, R.-C. Li, Extensions of Wielandt's min-max principles for positive semi-definite pencils, Linear Multilinear Algebra 62 (8) (2014) 1032–1048.
[17] H. Liu, A.M.-C. So, W. Wu, Quadratic optimization with orthogonality constraint: explicit Łojasiewicz exponent and linear convergence of retraction-based line-search and stochastic variance-reduced gradient methods, Math. Program., Ser. A 178 (1–2) (2019) 215–262.
[18] X. Liang, L. Wang, L.-H. Zhang, R.-C. Li, On generalizing trace minimization principles, Linear Algebra Appl. 656 (2023) 483–509.
[19] F.M. Dopico, V. Noferini, Root polynomials and their role in the theory of matrix polynomials, Linear Algebra Appl. 584 (2020) 37–78.
[20] V. Noferini, P. Van Dooren, Root vectors of polynomial and rational matrices: theory and computation, Linear Algebra Appl. 656 (2023) 510–540.
[21] L. Elsner, S. Friedland, Singular values, doubly stochastic matrices, and applications, Linear Algebra Appl. 220 (1995) 161–169.
[22] A.J. Hoffman, H.W. Wielandt, The variation of the spectrum of a normal matrix, Duke Math. J. 20 (1953) 37–39.
[23] R.-C. Li, Norms of certain matrices with applications to variations of the spectra of matrices and matrix pencils, Linear Algebra Appl. 182 (1993) 199–234.
[24] K. Veselić, Damped Oscillations of Linear Systems, Lecture Notes in Mathematics, vol. 2023, Springer, Heidelberg, Berlin, 2011.
[25] K. Veselić, A Jacobi eigenreduction algorithm for definite matrix pairs, Numer. Math. 64 (1993) 241–269.
[26] G.W. Stewart, J.-G. Sun, Matrix Perturbation Theory, Academic Press, Boston, 1990.
[27] P. Lancaster, L. Rodman, Canonical forms for Hermitian matrix pairs under strict equivalence and congruence, SIAM Rev. 47 (3) (2005) 407–443.
[28] A. Mal'cev, Foundation of Linear Algebra, Freeman, 1963.
[29] I. Gohberg, P. Lancaster, L. Rodman, Indefinite Linear Algebra and Applications, Birkhäuser, Basel, Switzerland, 2005.