

Absolute variation of Ritz values, principal angles and spectral spread

Pedro Massey, Demetrio Stojanoff and Sebastián Zárate *

Centro de Matemática, FCE-UNLP, La Plata and IAM-CONICET, Argentina

Abstract

Let A be a $d \times d$ complex self-adjoint matrix, $\mathcal{X}, \mathcal{Y} \subset \mathbb{C}^d$ be k -dimensional subspaces and let X be a $d \times k$ complex matrix whose columns form an orthonormal basis of \mathcal{X} ; that is, X is an isometry whose range is the subspace \mathcal{X} . We construct a $d \times k$ complex matrix Y_r whose columns form an orthonormal basis of \mathcal{Y} and obtain sharp upper bounds for the singular values $s(X^*AX - Y_r^*AY_r)$ in terms of submajorization relations involving the principal angles between \mathcal{X} and \mathcal{Y} and the spectral spread of A . We apply these results to obtain sharp upper bounds for the absolute variation of the Ritz values of A associated with the subspaces \mathcal{X} and \mathcal{Y} , that partially confirm conjectures by Knyazev and Argentati.

AMS subject classification: 42C15, 15A60.

Keywords: principal angles, Ritz values, spectral spread, majorization.

Contents

1	Introduction	2
2	Preliminaries	4
3	On principal submatrices, angles and spectral spread	5
3.1	On the variation of principal submatrices	5
3.2	An application: on the absolute variation of Ritz values	7
4	Proof of the main results	10
4.1	Proof of Theorem 3.2	10
4.2	Proof of Theorem 3.3	15
5	Appendix	20
5.1	Majorization theory in matrix analysis	20
5.2	Spectral spread of self-adjoint matrices	21

*Partially supported by CONICET (PIP 0152 CO), FONCyT (PICT-2015-1505) and UNLP (11X829)
e-mail addresses: massey@mate.unlp.edu.ar , demetrio@mate.unlp.edu.ar , seb4.zarate@gmail.com

1 Introduction

The problem of computing eigenvalues and invariant subspaces of self-adjoint matrices is ubiquitous in applications of linear algebra and numerical analysis. For example, given a $d \times d$ complex positive semidefinite matrix A , we could be interested in computing a low rank matrix approximation, which is an essential tool to deal with large and/or sparse self-adjoint matrices (see [17, 18, 19]); in this case, optimal approximations of A of rank k are constructed in terms of invariant subspaces associated with the k largest eigenvalues of A . In general, there is no universal effective solution to these problems ; there are several fundamental algorithmic methods for the computation of *approximate* eigenvalues and invariant subspaces, and the best suited such method typically depends on the context. One common fundamental aspect of these algorithms is the assessment of the quality of a subspace (and its associated Ritz values), as a candidate for an approximate invariant subspace (and approximate eigenvalues). Indeed, let X be a $d \times k$ complex matrix whose columns form an orthonormal basis of a k -dimensional subspace \mathcal{X} . Then, \mathcal{X} is an A -invariant subspace if and only if the so-called residual of A at X , given by $R_X = AX - X(X^*AX)$, is the zero matrix. In this case, the Ritz values of A associated with \mathcal{X} that is, the eigenvalues $\lambda(X^*AX) = (\lambda_j(X^*AX))_{j=1}^k$ (counting multiplicities and arranged non-increasingly) of the $k \times k$ submatrix X^*AX , are eigenvalues of A . In general, if the residual is *small*, then we consider \mathcal{X} an approximate invariant subspace and its associated Ritz values as approximate eigenvalues of A .

There are some other, rather indirect, measures of the quality of a subspace \mathcal{X} as a possible invariant subspace of the complex $d \times d$ self-adjoint matrix A , based on the local behavior of the Ritz values of A associated with subspaces \mathcal{Y} that are close to \mathcal{X} (an example of this phenomenon is described at the beginning of Section 3.2). This fact has been one motivation for the study of the so-called absolute variation of Ritz values (see [1, 3, 8, 9, 10, 11, 15, 16, 20, 21, 22]). In this context, we are interested in getting upper bounds for (some measure of) the vector

$$|\lambda(X^*AX) - \lambda(Y^*AY)| := (|\lambda_j(X^*AX) - \lambda_j(Y^*AY)|)_{j=1}^k \in \mathbb{R}_{\geq 0}^k, \quad (1)$$

where X and Y are $d \times k$ isometries with ranges $R(X) = \mathcal{X}$ and $R(Y) = \mathcal{Y}$. By continuity of eigenvalues, the absolute variation of Ritz values is controlled by the distance between the subspaces \mathcal{X} and \mathcal{Y} ; as a vector valued measure of distance between the subspaces \mathcal{X} and \mathcal{Y} , we consider the principal angles between \mathcal{X} and \mathcal{Y} , denoted $\Theta(\mathcal{X}, \mathcal{Y}) = (\theta_j)_{j=1}^k \in [0, \pi/2]^k$, that are the angles whose cosines are the singular values of X^*Y arranged in non-decreasing order and counting multiplicities.

On the other hand, it turns out that the absolute variation of Ritz values is also controlled by the spread of the eigenvalues of A (independently of the subspaces \mathcal{X} and \mathcal{Y}); notice that in the extreme case in which $A = aI_d$ is a multiple of the identity matrix, or equivalently when the spread of the eigenvalues of A is zero, then $X^*AX = aI_k = Y^*AY$ so the variation of Ritz values is also zero. As a vector valued measure of the spread of the eigenvalues of A , Knyazev and Argentati (see [10]) have suggested to consider the so-called *spectral spread of* A , denoted $\text{Spr}^+(A)$, given by

$$\text{Spr}^+(A) = (\lambda_j(A) - \lambda_{d-j+1}(A))_{j=1}^h \in (\mathbb{R}_{\geq 0}^h)^\dagger,$$

where $h = \lfloor \frac{d}{2} \rfloor$ (integer part). In this context, they have conjectured the following (autonomous) *a priori* upper bounds (see [10, Conjecture 2.1]): given two $d \times k$ isometries X

and Y then

$$|\lambda(X^*AX) - \lambda(Y^*AY)| \prec_w \sin(\Theta(\mathcal{X}, \mathcal{Y})) \text{Spr}^+(A), \quad (2)$$

where multiplication is performed entry-wise and \prec_w denotes *submajorization*, which is a pre-order relation between real vectors (see Section 2 for details). Moreover, if in addition the subspace \mathcal{X} is A -invariant then they have also conjectured that

$$|\lambda(X^*AX) - \lambda(Y^*AY)| \prec_w \sin(\Theta(\mathcal{X}, \mathcal{Y}))^2 \text{Spr}^+(A). \quad (3)$$

There has been important progress in this direction, and there are several related results and numerical evidence supporting these conjectures (see [8, 10, 13]).

Based on Lidskii's additive inequality, in order to bound the absolute variation of Ritz values we can look for upper bounds of the singular values

$$s(X^*AX - Y^*AY) \in \mathbb{R}^k, \quad (4)$$

where $s(Z) \in \mathbb{R}^d$ denotes the vector of singular values of a $d \times d$ complex matrix Z , counting multiplicities and arranged in non-decreasing order. As opposed to the variation of Ritz values in Eq. (1), the singular values in Eq. (4) actually depend on the particular choices of partial isometries X and Y with fixed ranges $R(X) = \mathcal{X}$ and $R(Y) = \mathcal{Y}$. Motivated by geometric insights, given the isometry $X \in \mathcal{M}_{d,k}(\mathbb{C})$ with range $R(X) = \mathcal{X}$ and the subspace \mathcal{Y} as above, we will choose an explicit (and convenient) $Y = Y_r$ and obtain upper bounds for the singular values in Eq. (4) in terms of submajorization relations involving the principal angles between \mathcal{X} and \mathcal{Y} and the spectral spread of A (see Section 3.1 for a detailed description of our main results regarding this problem). In a sense, our approach corresponds to the study of the stability of the *restricted* submatrix extraction process $(A, X, \mathcal{Y}) \mapsto Y_r^*AY_r$ around the fixed data (X, A) , where $Y_r = Y_r(X, \mathcal{Y})$ has an explicit dependence that we describe in detail in Section 4.

Once we have obtained such upper bounds we can apply Lidskii's additive inequality and obtain the following upper bounds for the absolute variation of Ritz values:

$$|\lambda(X^*AX) - \lambda(Y^*AY)| \prec_w \Theta(\mathcal{X}, \mathcal{Y}) \text{Spr}^+(A), \quad (5)$$

where multiplication is performed entry-wise, and if in addition the subspace \mathcal{X} is an A -invariant subspace:

$$|\lambda(X^*AX) - \lambda(Y^*AY)| \prec_w \Theta(\mathcal{X}, \mathcal{Y})^2 \text{Spr}^+(A). \quad (6)$$

Although our results do not settle the conjectures in their original form, our upper bounds in Eqs. (5) and (6) are comparable with those conjectured in Eqs. (2) and (3) for the general and the invariant case, specially for perturbations Y of X , since $\lim_{\theta \rightarrow 0^+} \frac{\sin(\theta)}{\theta} = 1$. Moreover, we include a family of examples that show that our vector valued upper bounds are sharp (see Section 3.2): explicitly, we show that there exist selections of $d \times k$ isometries $Y(t) \neq X$ for $t \in (0, 1)$ such that

$$\lim_{t \rightarrow 0^+} Y(t) = X \quad \text{and} \quad \lim_{t \rightarrow 0^+} \frac{|\lambda(X^*AX) - \lambda(Y(t)^*AY(t))|}{\Theta(\mathcal{X}, \mathcal{Y}(t)) \text{Spr}^+(A)} = (1, \dots, 1) \in \mathbb{R}^k.$$

The previous facts suggest that the upper bounds in Eqs. (5) and (6) could be a useful tool to deal with the case where Y is a perturbation of X .

The paper is organized as follows. In section 2 we introduce the notation and terminology used throughout the paper. In Section 3 we state our main results; indeed, in Section 3.1 we state our results on the stability of the restricted submatrix extraction process $(A, X, \mathcal{Y}) \mapsto Y_r^* A Y_r$ (for $Y_r = Y_r(X, \mathcal{Y})$ explicitly constructed) around the fixed data (X, A) , in terms of the principal angles $\Theta(\mathcal{X}, \mathcal{Y})$ and the spectral spread $\text{Spr}^+(A)$. In Section 3.2 we apply the previous results to obtain upper bounds for the absolute variation of Ritz values. In Section 4 we develop the proofs of the results in Section 3.1. Our approach is based on bounding the singular values of $X^* A X - Y_r^* A Y_r$ in terms of the integral of the singular values of the derivative $\gamma'(t)$ of a smooth curve $\gamma(t)$ joining $X^* A X$ with $Y_r^* A Y_r$. An essential part of our argument relies on the construction of a convenient curve $\gamma(t)$. Then, we apply recent results from [14] for the spectral spread of self-adjoint matrices to bound the singular values of the derivative $s(\gamma'(t))$. We have included a short Appendix (Section 5) with some facts related to majorization theory and the spectral spread of self-adjoint matrices, that are used throughout the paper.

2 Preliminaries

Throughout our work we use the following.

Notation and terminology. We let $\mathcal{M}_{d,k}(\mathbb{C})$ be the space of complex $d \times k$ matrices and write $\mathcal{M}_{d,d}(\mathbb{C}) = \mathcal{M}_d(\mathbb{C})$ for the algebra of $d \times d$ complex matrices. We denote by $\mathcal{H}(d) \subset \mathcal{M}_d(\mathbb{C})$ the real subspace of Hermitian (self-adjoint) matrices, by $i \cdot \mathcal{H}(d)$ the space of skew-Hermitian matrices and by $\mathcal{M}_d(\mathbb{C})^+$, the cone of positive semi-definite matrices. Also, $\mathcal{U}(d) \subset \mathcal{M}_d(\mathbb{C})$ denotes the group of unitary matrices. Given $1 \leq k \leq d$ we denote $\mathcal{I}(k, d)$ the set of isometries $X \in \mathcal{M}_{d,k}(\mathbb{C})$ i.e. such that $X^* X = I_k$; notice that $\mathcal{I}(k, d)$ can be identified in a natural way with the complex Stiefel manifold. Moreover, if $\mathcal{X} \subset \mathbb{C}^d$ is a k -dimensional subspace of \mathbb{C}^d we let $\mathcal{I}_{\mathcal{X}}(k, d)$ denote those isometries $X \in \mathcal{M}_{d,k}(\mathbb{C})$ with range $R(X) = \mathcal{X}$.

For $d \in \mathbb{N}$, let $\mathbb{I}_d = \{1, \dots, d\}$. Given a vector $x \in \mathbb{C}^d$ we denote by $D_x = \text{diag}(x_1, \dots, x_d)$ the diagonal matrix in $\mathcal{M}_d(\mathbb{C})$ whose main diagonal is x . Given $x = (x_i)_{i \in \mathbb{I}_d} \in \mathbb{R}^d$ we denote by $x^\downarrow = (x_i^\downarrow)_{i \in \mathbb{I}_d}$ the vector obtained by rearranging the entries of x in non-increasing order. We also use the notation $(\mathbb{R}^d)^\downarrow = \{x \in \mathbb{R}^d : x = x^\downarrow\}$ and $(\mathbb{R}_{\geq 0}^d)^\downarrow = \{x \in \mathbb{R}_{\geq 0}^d : x = x^\downarrow\}$. For $r \in \mathbb{N}$, we let $\mathbb{1}_r = (1, \dots, 1) \in \mathbb{R}^r$.

Given a matrix $A \in \mathcal{H}(d)$ we denote by $\lambda(A) = (\lambda_i(A))_{i \in \mathbb{I}_d} \in (\mathbb{R}^d)^\downarrow$ the eigenvalues of A counting multiplicities and arranged in non-increasing order. For $B \in \mathcal{M}_d(\mathbb{C})$, $s(B) = \lambda(|B|) \in (\mathbb{R}_{\geq 0}^d)^\downarrow$ denotes the singular values of B , i.e. the eigenvalues of $|B| = (B^* B)^{1/2} \in \mathcal{M}_d(\mathbb{C})^+$. Arithmetic operations with vectors are performed entry-wise i.e., in case $x = (x_i)_{i \in \mathbb{I}_k}$, $y = (y_i)_{i \in \mathbb{I}_k} \in \mathbb{C}^k$ then $x + y = (x_i + y_i)_i$, $x y = (x_i y_i)_i$ and (assuming that $y_i \neq 0$, for $i \in \mathbb{I}_k$) $x/y = (x_i/y_i)_i$, where these vectors all lie in \mathbb{C}^k . Moreover, if we assume further which $x, y \in \mathbb{R}^k$ then we write $x \leq y$ whenever $x_i \leq y_i$, for $i \in \mathbb{I}_k$. Finally, given a function $f : I \rightarrow \mathbb{R}$ defined on $I \subseteq \mathbb{R}$ and $x = (x_i)_{i \in \mathbb{I}_k} \in I^k$ then we set $f(x) = (f(x_i))_{i \in \mathbb{I}_k} \in \mathbb{R}^k$. \triangle

Next we recall the notion of majorization between vectors, which will play a central role throughout our work.

Definition 2.1. Let $x, y \in \mathbb{R}^k$. We say that x is *submajorized* by y , and write $x \prec_w y$, if

$$\sum_{i=1}^r x_i^\downarrow \leq \sum_{i=1}^r y_i^\downarrow \quad \text{for every } r \in \mathbb{I}_k.$$

If $x \prec_w y$ and $\text{tr } x \stackrel{\text{def}}{=} \sum_{i=1}^k x_i = \text{tr } y$, then we say that x is *majorized* by y , and write $x \prec y$. \triangle

We point out that (sub)majorization is a preorder relation in \mathbb{R}^k that plays a central role in matrix analysis (see Section 5.1).

Remark 2.2. Let $x \in \mathbb{R}_{\geq 0}^k$ and $y \in \mathbb{R}_{\geq 0}^h$ be two vector with *non-negative* entries (of different sizes). We extend the notion of submajorization, sum and the product between x and y in the following sense: Let 0_n denotes the zero vector of \mathbb{R}^n and $\ell := \max\{h, k\}$.

1. We say that x is submajorized by y if

$$x \prec_w y \quad \text{if} \quad \begin{cases} (x, 0_{h-k}) \prec_w y & \text{for } k < h \\ x \prec_w (y, 0_{k-h}) & \text{for } h < k \end{cases}, \quad (7)$$

2. Similarly we define $x + y$ and $xy \in \mathbb{R}_{\geq 0}^\ell$, adding zeros to the right to get two vectors with equal size. \triangle

3 On principal submatrices, angles and spectral spread

In this section we state our main results. Indeed, in Section 3.1 we state our results on the stability of the restricted submatrix extraction process $(A, X, \mathcal{Y}) \mapsto Y_r^* A Y_r$, where $Y_r = Y_r(X, \mathcal{Y})$ is defined as $Y_r = U X$ for a direct rotation U from \mathcal{X} to \mathcal{Y} . In this case, we bound the vector of singular values $s(X^* A X - Y_r^* A Y_r)$ (counting multiplicities and arranged non-increasingly) in terms of the principal angles $\Theta(\mathcal{X}, \mathcal{Y})$ and the spectral spread $\text{Spr}^+(A)$. In Section 3.2 below we apply the previous results to obtain upper bounds for the absolute variation of Ritz values.

3.1 On the variation of principal submatrices

Remark 3.1. We begin by recalling some of the notions involved in the statements of our main results. Given k -dimensional subspaces $\mathcal{X}, \mathcal{Y} \subset \mathbb{C}^d$, we denote by $\mathcal{I}_{\mathcal{X}}(k, d)$ the set of isometries $X \in \mathcal{M}_{d, k}(\mathbb{C})$ with range $R(X) = \mathcal{X}$ (similarly $\mathcal{I}_{\mathcal{Y}}(k, d)$), and we consider the *principal angles* between \mathcal{X} and \mathcal{Y} , denoted $\Theta(\mathcal{X}, \mathcal{Y}) = (\theta_j)_{j \in \mathbb{I}_k} \in [0, \pi/2]^k$, given by

$$\cos(\Theta) = (\cos(\theta_j))_{j \in \mathbb{I}_k} = (s_{k-j+1}(X^* Y))_{j \in \mathbb{I}_k}, \quad (8)$$

where $X \in \mathcal{I}_{\mathcal{X}}(k, d)$ and $Y \in \mathcal{I}_{\mathcal{Y}}(k, d)$. By construction, $\Theta(\mathcal{X}, \mathcal{Y}) = \Theta(\mathcal{X}, \mathcal{Y})^\perp$. On the other hand we also consider *direct rotations* between subspaces, introduced by Davis and Kahan in [4] (also see [5]); for the purposes of this work, it is convenient to describe such a concept in a direct and suitable way. In order to do this, we introduce a series of subspaces naturally associated with \mathcal{X} and \mathcal{Y} , that decompose \mathbb{C}^d into mutually orthogonal components; then we describe the direct rotations from \mathcal{X} onto \mathcal{Y} in terms of block representations with respect to the previous orthogonal decomposition of \mathbb{C}^d (see Eq. (10)). Indeed, we consider the generic part in the decomposition of \mathbb{C}^d in terms of the two subspaces \mathcal{X} and \mathcal{Y} (see Section 4.1), defined as the subspace

$$\mathcal{G} \stackrel{\text{def}}{=} [(\mathcal{X} \cap \mathcal{Y}) \oplus (\mathcal{X} \cap \mathcal{Y}^\perp) \oplus (\mathcal{X}^\perp \cap \mathcal{Y}) \oplus (\mathcal{X}^\perp \cap \mathcal{Y}^\perp)]^\perp. \quad (9)$$

Denote by $p = \dim \mathcal{X} \cap \mathcal{Y}^\perp = \dim \mathcal{X}^\perp \cap \mathcal{Y}$, $r = \dim \mathcal{X} \cap \mathcal{G} = \dim \mathcal{X}^\perp \cap \mathcal{G} = \dim \mathcal{Y} \cap \mathcal{G}$,
 $\mathcal{S}_1 = (\mathcal{X} \cap \mathcal{Y}^\perp) \oplus (\mathcal{X} \cap \mathcal{G})$, $\mathcal{S}_2 = (\mathcal{X}^\perp \cap \mathcal{Y}) \oplus (\mathcal{X}^\perp \cap \mathcal{G})$ and $\mathcal{S}_3 = (\mathcal{X}^\perp \cap \mathcal{Y}) \oplus (\mathcal{Y} \cap \mathcal{G})$.
Notice that $\mathcal{S}_1 \subseteq \mathcal{X}$, $\mathcal{S}_2 \subseteq \mathcal{X}^\perp$ and $\mathcal{S}_3 \subseteq \mathcal{Y}$. Consider the angles $\Theta' = \Theta(\mathcal{S}_1, \mathcal{S}_3) \in [0, \pi/2]^{p+r}$,
and the diagonal matrices

$$C = \text{diag}(\cos(\Theta')) \quad \text{and} \quad S = \text{diag}(\sin(\Theta')) \in \mathcal{M}_{p+r}(\mathbb{C})^+.$$

Finally, we say that a unitary $U \in \mathcal{U}(d)$ is a direct rotation from \mathcal{X} onto \mathcal{Y} if there exists \mathcal{B} an orthonormal basis (ONB) of \mathbb{C}^d obtained by juxtaposition of ONB's for $\mathcal{X} \cap \mathcal{Y}$, \mathcal{S}_1 , \mathcal{S}_2 and $\mathcal{X}^\perp \cap \mathcal{Y}^\perp$ such that the block matrix representation of U with respect to \mathcal{B} and these subspaces is

$$\begin{pmatrix} I & 0 & 0 & 0 \\ 0 & C & -S & 0 \\ 0 & S & C & 0 \\ 0 & 0 & 0 & I \end{pmatrix} \begin{matrix} \mathcal{X} \cap \mathcal{Y} \\ \mathcal{S}_1 \\ \mathcal{S}_2 \\ \mathcal{X}^\perp \cap \mathcal{Y}^\perp \end{matrix} \quad (10)$$

Moreover, we remark that in this case, if $s = \dim \mathcal{X} \cap \mathcal{Y}$, then

$$\Theta' = \left(\frac{\pi}{2} \mathbf{1}_p, \theta_1, \dots, \theta_r \right) \quad \text{and} \quad \Theta(\mathcal{X}, \mathcal{Y}) = (\Theta', 0_s). \quad (11)$$

Notice that since $\mathcal{X}, \mathcal{Y} \subset \mathbb{C}^d$ and $\dim \mathcal{X} = \dim \mathcal{Y}$, a direct rotation U between \mathcal{X} and \mathcal{Y} always exists. This relies on the fact that $p = \dim \mathcal{X} \cap \mathcal{Y}^\perp = \dim \mathcal{X}^\perp \cap \mathcal{Y}$ (see [4] for more details). On the other hand, recall that given a self-adjoint matrix $A \in \mathcal{H}(d)$ then the spectral spread of A , denoted $\text{Spr}^+(A)$, is given by

$$\text{Spr}^+(A) = (\text{Spr}_j(A))_{j \in \mathbb{I}_h} = (\lambda_j(A) - \lambda_j^\uparrow(A))_{j \in \mathbb{I}_h} \in (\mathbb{R}_{\geq 0}^h)^\downarrow,$$

where $h = \lfloor \frac{d}{2} \rfloor$ (integer part). See section 5.2 for more information about this notion. \triangle

In the following formulae we operate among vectors with non-negative entries of different sizes, using the notation given in Remark 2.2. The proofs of Theorems 3.2 and 3.3 below are developed in Section 4.

Theorem 3.2. *Let $A, B \in \mathcal{H}(d)$ and let $\mathcal{X}, \mathcal{Y} \subset \mathbb{C}^d$ be k -dimensional subspaces. Let $U = U(\mathcal{X}, \mathcal{Y})$ be a direct rotation of \mathcal{X} onto \mathcal{Y} and $\Theta = \Theta(\mathcal{X}, \mathcal{Y})^\downarrow \in [0, \pi/2]^k$ the principal angles between \mathcal{X} and \mathcal{Y} defined in (8). Given $X \in \mathcal{I}_X(k, d)$, if we let*

$$Y_r = Y_r(X, U) \stackrel{\text{def}}{=} UX \in \mathcal{I}_Y(k, d)$$

then we have that

$$s(X^* A X - Y_r^* B Y_r) \prec_w s(A - B) + \Theta(\mathcal{X}, \mathcal{Y})^\downarrow \left(\frac{\text{Spr}^+(A) + \text{Spr}^+(B)}{2} \right), \quad (12)$$

where submajorization, sums and products are as in Remark 2.2. \square

Notice that we have considered the more general situation of two self-adjoint matrices $A, B \in \mathcal{H}(d)$. In the next section, we apply this result in the particular case $A = B$ to obtain upper bounds for the absolute variation of the Ritz values in the self-adjoint case.

In the following result we deal with the so-called invariant case, and obtain a stronger upper bound for small perturbations \mathcal{Y} of \mathcal{X} .

Theorem 3.3. *With the same hypothesis and notation of Theorem 3.2, assume further that the k -dimensional subspace \mathcal{X} is A -invariant. Then we have that*

$$s(X^* A X - Y_r^* A Y_r) \prec_w \Theta^2(\mathcal{X}, \mathcal{Y})^\downarrow \text{Spr}^+(A), \quad (13)$$

where submajorization and products are as in Remark 2.2. \square

In the next section, we apply Theorems 3.2 and 3.3 and obtain upper bounds for the absolute variation of the Ritz values in the self-adjoint case.

3.2 An application: on the absolute variation of Ritz values

As already mentioned in the Introduction, one motivation for the study of absolute variation of Ritz values of $A \in \mathcal{H}(d)$ around a subspace \mathcal{X} in \mathbb{C}^d , is that such variation provides an indirect measure of the quality of \mathcal{X} as a possible invariant subspace of A . As an example of this phenomenon, we recall the following inequality, recently obtained in [13]: let $P_{\mathcal{X}+\mathcal{Y}}$ denote the orthogonal projection onto the subspace $\mathcal{X} + \mathcal{Y}$ of \mathbb{C}^d , let $R_X = AX - X(X^*AX)$ and $R_Y = AY - Y(Y^*AY)$ denote the residuals of A at X and Y , let $s(P_{\mathcal{X}+\mathcal{Y}} R_X), s(P_{\mathcal{X}+\mathcal{Y}} R_Y) \in \mathbb{R}^k$ denote the vectors of singular values of $P_{\mathcal{X}+\mathcal{Y}} R_X$ and $P_{\mathcal{X}+\mathcal{Y}} R_Y$ (counting multiplicities and arranged non-increasingly); if \mathcal{X} and \mathcal{Y} are in a acute relative position (i.e. so that $\theta_j < \pi/2$, for $j \in \mathbb{I}_k$) then

$$|\lambda(X^*AX) - \lambda(Y^*AY)| \prec_w [s(P_{\mathcal{X}+\mathcal{Y}} R_X) + s(P_{\mathcal{X}+\mathcal{Y}} R_Y)] (\tan(\theta_j))_{j=1}^k. \quad (14)$$

In case \mathcal{X} and \mathcal{Y} are close to each other, then the residuals of A at X and Y are comparable; moreover, since the tangents $\tan(\theta_j) \approx \theta_j$ are comparable, for $j \in \mathbb{I}_k$, then conclude that the inequality in Eq. (14) provides an upper bound for the first order absolute variation of the Ritz values of A around \mathcal{X} , in terms of the residual R_X (where we measured the distance between subspaces in terms of the principal angles). Hence, if R_X is small, then this first order variation (that can be tested numerically) is also small.

We are interested in obtaining *autonomous* upper bounds for the absolute variation of Ritz values of $A \in \mathcal{H}(d)$ associated with the k -dimensional subspaces \mathcal{X} and \mathcal{Y} , in terms of the principal angles between \mathcal{X} and \mathcal{Y} , and the spectral spread of the matrix A . In this context, Knyazev and Argentati have conjectured (see [10, Conjecture 2.1]) the upper bounds in Eq. (2) (for the general subspaces) and (3) (for an A -invariant subspace \mathcal{X}). The following results partially confirm these conjectures.

Theorem 3.4. *Let $A \in \mathcal{H}(d)$ and let $X, Y \in \mathcal{I}(k, d)$, with ranges $\mathcal{X} = R(X)$ and $\mathcal{Y} = R(Y)$ such that $\dim \mathcal{X} = \dim \mathcal{Y} = k$. If $\Theta = \Theta(\mathcal{X}, \mathcal{Y})$ then:*

1. *We have that*

$$|\lambda(X^*AX) - \lambda(Y^*AY)| \prec_w \Theta \text{Spr}^+(A). \quad (15)$$

2. *If we further assume that \mathcal{X} is A -invariant, we have that*

$$|\lambda(X^*AX) - \lambda(Y^*AY)| \prec_w \Theta^2 \text{Spr}^+(A). \quad (16)$$

Proof. Let $U \in \mathcal{U}(d)$ be a direct rotation from \mathcal{X} onto \mathcal{Y} and let $Y_r = U X$ be as in Theorem 3.2. Since $R(Y_r) = R(Y)$ then $V = Y_r^* Y \in \mathcal{U}(k)$, and we get that

$$\lambda(Y^*AY) = \lambda(V Y_r^* A Y_r V^*) = \lambda(Y_r^* P_{\mathcal{Y}} A P_{\mathcal{Y}} Y_r) = \lambda(Y_r^* A Y_r) \in \mathbb{R}^k.$$

By Lidskii's inequality (Theorem 5.2) and Theorem 3.2 we have that

$$\begin{aligned} |\lambda(X^*AX) - \lambda(Y^*AY)| &= |\lambda(X^*AX) - \lambda(Y_r^*AY_r)| \\ &\prec_w |\lambda(X^*AX - Y_r^*AY_r)|^\downarrow \\ &= s(X^*AX - Y_r^*AY_r) \prec_w \Theta \operatorname{Spr}^+(A). \end{aligned}$$

Item 2 follows from a similar argument, using Theorem 3.3. \square

Remark 3.5. The only difference between the Conjectures (2) and (3) and our main result Theorem 3.4 relies in the slightly bigger upper bounds Θ (resp. Θ^2) instead of $\sin \Theta$ (resp. $\sin^2 \Theta$). These numbers are asymptotically close when the angles $\Theta \rightarrow 0^+$, and they can be globally compared with a constant $\pi/2$.

Nevertheless, Conjectures (2) and (3) are supported by large computational experimentation. In order to explore the content of our main results we consider the following examples, where we show that both Conjectures (2) and (3) and Theorem 3.4 are sharp inequalities. \triangle

Example 3.6. Consider $a > b > 0$ and let

$$A = \begin{pmatrix} 0 & 0 & a & 0 \\ 0 & 0 & 0 & b \\ a & 0 & 0 & 0 \\ 0 & b & 0 & 0 \end{pmatrix} \in \mathcal{H}(4).$$

If $\{e_1, e_2, e_3, e_4\}$ denotes the canonical basis of \mathbb{C}^4 , we let $\mathcal{X} = \operatorname{Span}\{e_1, e_2\}$. On the other hand, given $\theta \in [0, \pi/2]$ let

$$f_1(\theta) = \cos(\theta) e_1 + \sin(\theta) e_3 \quad \text{and} \quad f_2(\theta) = \cos(\theta) e_2 + \sin(\theta) e_4. \quad (17)$$

Then, we set $\mathcal{Y}(\theta) = \operatorname{Span}\{f_1(\theta), f_2(\theta)\}$, for $\theta \in [0, \pi/2]$. It is straightforward to show that $\Theta(\mathcal{X}, \mathcal{Y}(\theta)) = (\theta, \theta)$. Then, it follows that the isometries $X, Y_r(\theta)$ as in Theorem 3.2 (associated with the subspaces \mathcal{X} and $\mathcal{Y}(\theta)$) are given by

$$X = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{pmatrix} \quad \text{and} \quad Y_r(\theta) = \begin{pmatrix} \cos(\theta) & 0 \\ 0 & \cos(\theta) \\ \sin(\theta) & 0 \\ 0 & \sin(\theta) \end{pmatrix}. \quad (18)$$

Direct computations with the matrices described above show that

$$X^*AX = 0 \quad \text{and} \quad Y_r(\theta)^*AY_r(\theta) = \begin{pmatrix} a \sin(2\theta) & 0 \\ 0 & b \sin(2\theta) \end{pmatrix}$$

Hence, in this case we get that for $\theta \in [0, \pi/2]$,

$$|\lambda(X^*AX) - \lambda(Y_r(\theta)^*AY_r(\theta))| = |\lambda(X^*AX - Y_r(\theta)^*AY_r(\theta))| = \sin(2\theta) (a, b)$$

so in particular, Lidskii's inequality holds with equality in this case. On the other hand, it turns out that $\lambda(A) = (a, b, -b, -a) \in (\mathbb{R}^4)^\downarrow$ which shows that $\operatorname{Spr}^+(A) = 2(a, b)$. Therefore, the inequalities in Theorems 3.2 (with $A = B$) and 3.4 (item 1.) become

$$\sin(2\theta) (a, b) \prec_w 2(\theta, \theta) (a, b) = 2\theta (a, b),$$

where the submajorization relation above is equivalent to the inequalities

$$\sin(2\theta) a \leq 2\theta a \quad \text{and} \quad \sin(2\theta) (a + b) \leq 2\theta (a + b)$$

which are sharp; this last claim can be seen by considering $\theta \rightarrow 0^+$. Notice that we further get that

$$\lim_{\theta \rightarrow 0^+} \frac{|\lambda(X^*AX) - \lambda(Y_r(\theta)^*AY_r(\theta))|}{\Theta(\mathcal{X}, \mathcal{Y}(\theta)) \text{Spr}^+(A)} = \mathbb{1}_2$$

where we have considered the entry-wise quotient of the vectors. This last fact shows that our upper bound for the absolute variation of Ritz values is sharp. \triangle

Example 3.7. Consider $a > b > 0$ and let

$$A = \begin{pmatrix} a & 0 & 0 & 0 \\ 0 & b & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \in \mathcal{H}(4).$$

We consider the canonical basis $\{e_1, e_2, e_3, e_4\}$ of \mathbb{C}^4 , and we let $\mathcal{X} = \text{Span}\{e_1, e_2\}$. Notice that in this case \mathcal{X} is an A -invariant subspace. On the other hand, given $\theta \in [0, \pi/2]$ let $\mathcal{Y}(\theta) = \text{Span}\{f_1(\theta), f_2(\theta)\}$, where $f_j(\theta)$ are as in Eq. (17). Hence, as in the previous example, we have that $\Theta(\mathcal{X}, \mathcal{Y}(\theta)) = (\theta, \theta)$. In this case, the isometries $X, Y_r(\theta)$ as in Theorem 3.3 are given by Eq. (18). Direct computations with the matrices described above show that

$$X^*AX = \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix} \quad \text{and} \quad Y_r(\theta)^*AY_r(\theta) = \begin{pmatrix} a \cos^2(\theta) & 0 \\ 0 & b \cos^2(\theta) \end{pmatrix}$$

Hence, in this case we get that for $\theta \in [0, \pi/2]$,

$$|\lambda(X^*AX) - \lambda(Y_r(\theta)^*AY_r(\theta))| = |\lambda(X^*AX - Y_r(\theta)^*AY_r(\theta))| = \sin^2(\theta) (a, b)$$

so in particular, Lidskii's inequality holds with equality in this case. On the other hand, it turns out that $\lambda(A) = (a, b, 0, 0) \in (\mathbb{R}^4)^\downarrow$ which shows that $\text{Spr}^+(A) = (a, b)$. Therefore, the inequalities in Theorems 3.3 and 3.4 (item 2.) become

$$\sin^2(\theta) (a, b) \prec_w (\theta, \theta)^2 (a, b) = \theta^2 (a, b),$$

where the submajorization relation above is equivalent to the inequalities

$$\sin^2(\theta) a \leq \theta^2 a \quad \text{and} \quad \sin^2(\theta) (a + b) \leq \theta^2 (a + b)$$

which are sharp; this last claim can be seen by considering $\theta \rightarrow 0^+$. Notice that we further get that

$$\lim_{\theta \rightarrow 0^+} \frac{|\lambda(X^*AX) - \lambda(Y_r(\theta)^*AY_r(\theta))|}{\Theta(\mathcal{X}, \mathcal{Y}(\theta))^2 \text{Spr}^+(A)} = \mathbb{1}_2,$$

where we have considered the entry-wise quotient of the vectors. This last fact shows that our upper bound for the absolute variation of Ritz values is sharp. \triangle

As a final comment, notice that it's natural to wonder whether the estimates from Theorems 3.2 and 3.3 can be used to obtain estimates for the distance between the compressions (that we can think of as lower rank approximations of A) given by $P_{\mathcal{X}} A P_{\mathcal{X}}$ and $P_{\mathcal{Y}} A P_{\mathcal{Y}}$. It turns out that this is not the case. Indeed, in the trivial case in which $A = I$ then we have that $P_{\mathcal{X}} I P_{\mathcal{X}} - P_{\mathcal{Y}} I P_{\mathcal{Y}} = P_{\mathcal{X}} - P_{\mathcal{Y}}$; but we have that $\text{Spr}^+(I) = 0$ is the zero vector, so there is no hope in obtaining an upper bound for $P_{\mathcal{X}} A P_{\mathcal{X}} - P_{\mathcal{Y}} A P_{\mathcal{Y}}$ in terms $\text{Spr}^+(A)$ in general. Also, in general there is no dependence of $P_{\mathcal{X}} A P_{\mathcal{X}} - P_{\mathcal{Y}} A P_{\mathcal{Y}}$ in terms of $\Theta(\mathcal{X}, \mathcal{Y})^2$ when \mathcal{X} is A -invariant (again, take $A = I$ to see this).

4 Proof of the main results

In this section we present complete proofs of our main results. Our approach is based on some geometric arguments and inequalities for the spectral spread recently obtained in [14].

4.1 Proof of Theorem 3.2

Throughout this section we adopt the notation and terminology in Theorem 3.2. Hence, we consider:

1. $A, B \in \mathcal{H}(d)$;
2. $\mathcal{X}, \mathcal{Y} \subset \mathbb{C}^d$, k -dimensional subspaces and $X \in \mathcal{I}_{\mathcal{X}}(k, d)$;
3. a direct rotation $U = U(\mathcal{X}, \mathcal{Y}) \in \mathcal{U}(d)$, of \mathcal{X} onto \mathcal{Y} .
4. $Y_r = Y_r(X, \mathcal{Y}) \stackrel{\text{def}}{=} UX \in \mathcal{I}_{\mathcal{Y}}(k, d)$.

Our approach to prove Theorem 3.2 is as follows: we will consider smooth curves

$$L(\cdot) : [0, 1] \rightarrow \mathcal{H}(d) \quad \text{and} \quad Y_r(\cdot) : [0, 1] \rightarrow \mathcal{I}(k, d) \quad (19)$$

such that $L(0) = A$, $L(1) = B$, $Y_r(0) = X$ and $Y_r(1) = Y_r$. Then, we consider the smooth curve

$$\gamma : [0, 1] \rightarrow \mathcal{H}(k) \quad \text{given by} \quad \gamma(t) = Y_r(t)^* L(t) Y_r(t) \quad \text{for} \quad t \in [0, 1]. \quad (20)$$

Once we have constructed $\gamma(\cdot)$ we will apply the following result. Recall that given $Z \in \mathcal{M}_d(\mathbb{C})$, $s(Z) \in \mathbb{R}^d$ denotes the vector of singular values, counting multiplicities and arranged in non-increasing order.

Proposition 4.1. *Let $\gamma : [0, 1] \rightarrow \mathcal{M}_{m,n}(\mathbb{C})$ be a smooth curve such that $\gamma(0) = C$ and $\gamma(1) = D$. Then*

$$s(D - C) \prec_w \int_0^1 s(\gamma'(t)) dt.$$

Proof. First notice that by the fundamental theorem of calculus we have that

$$D - C = \gamma(1) - \gamma(0) = \int_0^1 \gamma'(t) dt. \quad (21)$$

On the other hand, for $n \in \mathbb{N}$ consider the regular partition $\{t_0 = 0 < t_1 < \dots < t_n = 1\}$ of $[0, 1]$ so that $t_j = \frac{j}{n}$, for $j \in \{0\} \cup \mathbb{I}_n$ and $\Delta_j = t_j - t_{j-1} = \frac{1}{n} = \Delta_n$, for $j \in \mathbb{I}_n$. Then,

$$\int_0^1 \gamma'(t) dt = \lim_{n \rightarrow \infty} \sum_{j=1}^n \gamma'(t_j) \Delta_n, \quad \int_0^1 s(\gamma'(t)) dt = \lim_{n \rightarrow \infty} \sum_{j=1}^n s(\gamma'(t_j)) \Delta_n \quad (22)$$

where we have used that the curves $\gamma'(t) \in \mathcal{M}_{m,n}(\mathbb{C})$ and $s(\gamma'(t)) \in \mathbb{R}^d$, for $t \in [0, 1]$, are continuous (by hypothesis and by the continuity of singular values). By Weyl's additive inequality (see Theorem 5.1) we have that

$$s\left(\sum_{j=1}^n \gamma'(t_j) \Delta_n\right) \prec_w \sum_{j=1}^n s(\gamma'(t_j)) \Delta_n \quad (23)$$

The result now follows from Eqs. (21), (22), (23), the continuity of singular values and the following fact: if $(x_n)_{n \in \mathbb{N}}$ and $(y_n)_{n \in \mathbb{N}} \in \mathbb{R}^d$ are sequences that converge to x and $y \in \mathbb{R}^d$ respectively, and such that $x_n \prec_w y_n$ for $n \in \mathbb{N}$, then $x \prec_w y$. \square

Although the arguments considered in the previous paragraphs are valid for any choice of smooth curves $L(\cdot)$ and $Y_r(\cdot)$ as in Eq. (19), we are interested in choices that lead to better upper bounds. Thus, we are interested in those curves $L(\cdot)$ and $Y_r(\cdot)$ for which the associated curve $\gamma(\cdot)$ in Eq. (20) is *minimal* in a certain sense. We point out that we will not study the corresponding minimality problem, but rather we will choose $L(\cdot)$ and $Y_r(\cdot)$ that have separately minimal properties, and use these choices to build $\gamma(\cdot)$.

For $L(\cdot)$ there is a natural choice, namely the line segment joining A and B , i.e.

$$L(t) = (1 - t)A + tB \quad \text{for } t \in [0, 1]. \quad (24)$$

Next we construct $Y_r(t)$, based on the notion of direct rotation as developed in [4]; hence, we consider the following notions related to the direct rotation of \mathcal{X} onto \mathcal{Y} as in Remark 3.1. Indeed, given these subspaces we have the orthogonal decomposition (see [6])

$$\mathbb{C}^d = (\mathcal{X} \cap \mathcal{Y}) \oplus (\mathcal{X} \cap \mathcal{Y}^\perp) \oplus \mathcal{G} \oplus (\mathcal{X}^\perp \cap \mathcal{Y}) \oplus (\mathcal{X}^\perp \cap \mathcal{Y}^\perp).$$

Here $\mathcal{G} \subseteq \mathbb{C}^d$ stands for the generic part of the pair of subspaces \mathcal{X} and \mathcal{Y} (see Eq. (9)). We point out that some of these subspaces can be null. It turns out that for our purposes, we can assume further (see the proof of Theorem 3.2 below) that

$$\mathcal{X}^\perp \cap \mathcal{Y}^\perp = \{0\}.$$

Recall that we can decompose the generic part $\mathcal{G} = \mathcal{X} \cap \mathcal{G} \oplus \mathcal{X}^\perp \cap \mathcal{G}$ into two subspaces such that $\dim \mathcal{X} \cap \mathcal{G} = \dim \mathcal{X}^\perp \cap \mathcal{G} = r$. Therefore, in our case we get the decomposition

$$\mathbb{C}^d = (\mathcal{X} \cap \mathcal{Y}) \oplus \mathcal{S}_1 \oplus \mathcal{S}_2, \quad (25)$$

where

$$\mathcal{S}_1 = (\mathcal{X} \cap \mathcal{Y}^\perp) \oplus (\mathcal{X} \cap \mathcal{G}) \quad \text{and} \quad \mathcal{S}_2 = (\mathcal{X}^\perp \cap \mathcal{Y}) \oplus (\mathcal{X}^\perp \cap \mathcal{G}).$$

Since $\dim \mathcal{X} = \dim \mathcal{Y}$ we conclude that $\dim \mathcal{X} \cap \mathcal{Y}^\perp = \dim \mathcal{X}^\perp \cap \mathcal{Y} = p$ so, in particular, $\dim \mathcal{S}_1 = \dim \mathcal{S}_2 = r + p$.

Let U be a direct rotation of \mathcal{X} onto \mathcal{Y} . In this case, there exist orthonormal bases of \mathcal{S}_1 and \mathcal{S}_2 and diagonal positive semidefinite matrices

$$C = \text{diag}(\cos(\Theta')) \quad \text{and} \quad S = \text{diag}(\sin(\Theta')) \in \mathcal{M}_{p+r}(\mathbb{C})^+$$

where $\Theta' = (\frac{\pi}{2} \mathbf{1}_p, \theta_1, \dots, \theta_r)$ denote the principal angles between \mathcal{S}_1 and $(\mathcal{Y} \cap \mathcal{G}) \oplus (\mathcal{X}^\perp \cap \mathcal{Y})$, such that

$$U = \begin{pmatrix} I & 0 & 0 \\ 0 & C & -S \\ 0 & S & C \end{pmatrix}, \quad (26)$$

where the block matrix representation is with respect to the ONB obtained by juxtaposition of the ONB's of $\mathcal{X} \cap \mathcal{Y}$, \mathcal{S}_1 and \mathcal{S}_2 . Here $(\theta_1, \dots, \theta_r) \in (0, \pi/2)^r$ denote the principal angles between $\mathcal{X} \cap \mathcal{G}$ and $\mathcal{Y} \cap \mathcal{G}$. We now define

$$U(t) = \begin{pmatrix} I & 0 & 0 \\ 0 & C(t) & -S(t) \\ 0 & S(t) & C(t) \end{pmatrix} \in \mathcal{U}(d), \quad (27)$$

where $S(\cdot), C(\cdot) : [0, 1] \rightarrow \mathcal{H}(p+r)$ are given by

$$S(t) = \text{diag}(\sin(t\Theta')) \quad \text{and} \quad C(t) = \text{diag}(\cos(t\Theta')) \in \mathcal{M}_{p+r}(\mathbb{C})^+. \quad (28)$$

Then $U(0) = I$ and $U(1) = U$. Now that we have explicitly constructed $U(t)$ we define

$$Y_r(t) = U(t) X = \begin{pmatrix} I & 0 \\ 0 & C(t) \\ 0 & S(t) \end{pmatrix} \quad \text{with} \quad X = \begin{pmatrix} I & 0 \\ 0 & I \\ 0 & 0 \end{pmatrix}, \quad (29)$$

where the matrix block representation above is with respect to the decomposition $\mathbb{C}^d = (\mathcal{X} \cap \mathcal{Y}) \oplus \mathcal{S}_1 \oplus \mathcal{S}_2$, as before.

In this case, we have that

$$\gamma(t) = Y_r(t)^* L(t) Y_r(t) = X^* (U(t)^* L(t) U(t)) X \quad \text{for} \quad t \in [0, 1]. \quad (30)$$

Notice that by taking a derivative,

$$\gamma'(t) = X^* U(t)^* L'(t) U(t) X + X^* ((U'(t))^* L(t) U(t) + U(t)^* L(t) U'(t)) X. \quad (31)$$

The first term in Eq. (31) can be dealt with using Lidskii's inequality in a simple way. Thus, we are left with the analysis of the second term. Using basic trigonometric identities we get that $U(t+h) = U(t)U(h)$ for $t, h, t+h \in [0, 1]$; hence $U'(t) = U(t)U'(0)$ and similarly $(U'(t))^* = -U'(0)U(t)^*$. Therefore, if we let

$$A(t) = U(t)^* L(t) U(t) \quad (32)$$

then we have that

$$X^* ((U'(t))^* L(t) U(t) + U(t)^* L(t) U'(t)) X = X^* (A(t) U'(0) - U'(0) A(t)) X. \quad (33)$$

By taking a derivative in Eq. (27) we get that

$$U'(0) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -D_{\Theta'} \\ 0 & D_{\Theta'} & 0 \end{pmatrix} \in i \cdot \mathcal{H}(d).$$

where $D_{\Theta'}$ is the diagonal matrix with main diagonal $\Theta' = (\frac{\pi}{2} \mathbf{1}_p, \theta_1, \dots, \theta_r)$ defined above, and $i \cdot \mathcal{H}(d)$ denotes the space of skew-Hermitian matrices.

Consider the block matrix representation of $A(t) = (A_{ij}(t))_{i,j=1}^3$ with respect to the decomposition in Eq. (25). A direct computation, using the previous block matrix representations, shows that

$$X^* (A(t) U'(0) - U'(0) A(t)) X = 2 \text{Re} \left(\begin{pmatrix} 0 & A_{13}(t) \\ 0 & A_{23}(t) \end{pmatrix} \begin{pmatrix} 0 \\ D_{\Theta'} \end{pmatrix} \right). \quad (34)$$

Remark 4.2. With the previous notation, using Weyl's multiplicative inequality (see Theorem 5.1) we conclude that

$$s(X^* (A(t) U'(0) - U'(0) A(t)) X) \prec_w 2 \Theta s \left(\begin{pmatrix} A_{13}(t) \\ A_{23}(t) \end{pmatrix} \right), \quad (35)$$

where we have used that $\Theta = (\Theta', 0_s) \in (\mathbb{R}_{\geq 0}^k)^\downarrow$ and item 2 in Remark 2.2. The previous facts suggest to develop a bound for the singular values of the anti-diagonal block of A according to the representation

$$\left(\begin{array}{cc|c} A_{11}(t) & A_{12}(t) & A_{13}(t) \\ A_{21}(t) & A_{22}(t) & A_{23}(t) \\ \hline A_{31}(t) & A_{32}(t) & A_{33}(t) \end{array} \right) \quad (36)$$

(that formally corresponds to the decomposition $\mathbb{C}^d = (\mathcal{X} \cap \mathcal{Y} \oplus \mathcal{S}_1) \oplus \mathcal{S}_2$, where \mathcal{S}_1 and \mathcal{S}_2 are described after Eq. (25)) in terms of spectral properties of the matrix $A(t) \in \mathcal{H}(d)$ (defined in Eq. (32)). \triangle

Next, we describe a bound of the singular values of the anti-diagonal block as in Eq. (36) in terms of the spectral spread of $A(t)$.

Theorem 4.3 (See [14]). *Let $H = \begin{bmatrix} H_{11} & H_{12} \\ H_{12}^* & H_{22} \end{bmatrix} \begin{matrix} \mathbb{C}^k \\ \mathbb{C}^r \end{matrix} \in \mathcal{H}(k+r)$ be arbitrary. Then*

$$2s(H_{12}) \prec_w \text{Spr}^+(H). \quad (37)$$

□

We can now prove our first main result. We will re-write the statement for the reader's convenience.

Theorem 3.2. Let $A, B \in \mathcal{H}(d)$ and let $\mathcal{X}, \mathcal{Y} \subset \mathbb{C}^d$ be k -dimensional subspaces. Let $U = U(\mathcal{X}, \mathcal{Y})$ be a direct rotation of \mathcal{X} onto \mathcal{Y} and $\Theta = \Theta(\mathcal{X}, \mathcal{Y})^\downarrow \in [0, \pi/2]^k$ the principal angles between \mathcal{X} and \mathcal{Y} defined in (8). Given $X \in \mathcal{I}_{\mathcal{X}}(k, d)$, if we let

$$Y_r = Y_r(X, U) \stackrel{\text{def}}{=} UX \in \mathcal{I}_{\mathcal{Y}}(k, d)$$

then we have that

$$s(X^* A X - Y_r^* B Y_r) \prec_w s(A - B) + \Theta(\mathcal{X}, \mathcal{Y})^\downarrow \left(\frac{\text{Spr}^+(A) + \text{Spr}^+(B)}{2} \right),$$

where submajorization, sums and products are as in Remark 2.2.

Proof. Assume first that $\mathcal{X}^\perp \cap \mathcal{Y}^\perp = \{0\}$. Let $\gamma(\cdot) : [0, 1] \rightarrow \mathcal{H}(k)$ be defined as in Eq. (30), where $Y_r(\cdot)$ is defined as in Eq. (29). Notice that by construction γ is a smooth curve such that $\gamma(0) = X^* A X$ and $\gamma(1) = Y_r^* B Y_r$. By Proposition 4.1 we have that

$$s(X^* A X - Y_r^* B Y_r) \prec_w \int_0^1 s(\gamma'(t)) dt. \quad (38)$$

Using Eq. (31) and Weyl's inequality for singular values (see Theorem 5.1), we have that

$$\begin{aligned} s(\gamma'(t)) &\prec_w s(X^* U(t)^* L'(t) U(t) X) \\ &+ s(X^* ((U'(t))^* L(t) U(t) + U(t)^* L(t) U'(t)) X), \end{aligned} \quad (39)$$

where $L(t) = (1-t)A + tB$, for $t \in [0, 1]$ as before. Hence $L'(t) = B - A$ and therefore

$$s(X^* U(t)^* L'(t) U(t) X) = s(X^* U(t)^* (B - A) U(t) X) \prec_w s(A - B),$$

where we have used Remark 5.4. The previous facts together with item 3. in Lemma 5.3 imply that

$$\int_0^1 s(X^* U(t)^* L'(t) U(t) X) dt \prec_w s(A - B). \quad (40)$$

Moreover, using Eqs. (33), (34), (35) and Theorem 4.3 (also see item 5. in Lemma 5.3) we get that

$$s(X^* ((U'(t))^* L(t) U(t) + U(t)^* L(t) U'(t)) X) \prec_w \Theta \text{Spr}^+(A(t)). \quad (41)$$

Notice that $A(t)$ and $L(t)$ are unitary conjugates, so they have the same spectral spread. Hence, by Proposition 5.6 we get that

$$\text{Spr}^+(A(t)) = \text{Spr}^+(L(t)) \prec (1 - t) \text{Spr}^+(A) + t \text{Spr}^+(B). \quad (42)$$

By Eqs. (41), (42) and item 3. in Lemma 5.3,

$$\begin{aligned} & \int_0^1 s(X^* ((U'(t))^* L(t) U(t) + U(t)^* L(t) U'(t)) X) dt \prec_w \\ & \Theta \left(\text{Spr}^+(A) \int_0^1 (1 - t) dt + \text{Spr}^+(B) \int_0^1 t dt \right) = \frac{1}{2} \Theta (\text{Spr}^+(A) + \text{Spr}^+(B)). \end{aligned}$$

Using this last submajorization relation, together with Eq. (39) and Eq. (40) we get that

$$\int_0^1 s(\gamma'(t)) dt \prec_w s_j(A - B)_{j=1}^k + \Theta \left(\frac{\text{Spr}^+(A) + \text{Spr}^+(B)}{2} \right), \quad (43)$$

where we have used the first part of item 3. in Lemma 5.3. The result now follows from Eq. (38) and (43).

In case $\mathcal{X}^\perp \cap \mathcal{Y}^\perp \neq \{0\}$, we consider the subspace $\mathcal{Z} = \mathcal{X} + \mathcal{Y} \subset \mathbb{C}^d$ and $Z \in \mathcal{I}_{\mathcal{Z}}(m, d)$, where $m = \dim \mathcal{Z}$. Let $A' = Z^* A Z$, $B' = Z^* B Z \in \mathcal{H}(m)$ and let $X' = Z^* X \in \mathcal{I}(k, m)$. If we let $\mathcal{X}' = R(X')$ and $\mathcal{Y}' = R(Z^* Y_r)$ denote the subspaces that are the ranges of X' and $Z^* Y_r$ in \mathbb{C}^m we get that $(\mathcal{X}')^\perp \cap (\mathcal{Y}')^\perp = \{0\}$. Moreover, since $(X')^* Y' = (X^* Z)(Z^* Y_r) = X^* P_{\mathcal{Z}} Y_r = X^* Y_r$ (where $P_{\mathcal{Z}} \in \mathcal{H}(d)$ denotes the orthogonal projection onto $\mathcal{Z} \subset \mathbb{C}^d$) we conclude that $\Theta(\mathcal{X}', \mathcal{Y}') = \Theta(\mathcal{X}, \mathcal{Y}) \in \mathbb{R}^k$, by definition of principal angles. By the first part of the proof, we now have that

$$s((X')^* A' X' - (Y_r')^* B' Y_r') \prec_w s_j(A' - B')_{j=1}^k + \Theta \left(\frac{\text{Spr}^+(A') + \text{Spr}^+(B')}{2} \right), \quad (44)$$

where $Y_r' = U' X'$, and U' is a direct rotation of \mathcal{X}' onto \mathcal{Y}' . But notice that $Z^* U Z$ is a direct rotation from \mathcal{X}' onto \mathcal{Y}' so we can take $Y_r' = Z^* U Z Z^* X = Z^* Y_r$. Hence,

$$\begin{aligned} (X')^* A' X' &= X^* P_{\mathcal{Z}} A P_{\mathcal{Z}} X = X^* A X \quad \text{and} \\ (Y_r')^* B' Y_r' &= Y_r^* P_{\mathcal{Z}} B P_{\mathcal{Z}} Y_r = Y_r^* B Y_r \end{aligned}$$

Therefore,

$$s((X')^* A' X' - (Y_r')^* B' Y_r') = s(X^* A X - Y_r^* B Y_r). \quad (45)$$

On the other hand, using Remark 5.4 (or the interlacing inequalities) we now see that

$$\begin{aligned} s(A' - B') &= s(Z^* (A - B) Z) \prec_w s(A - B) \quad \text{and} \\ \text{Spr}^+(A') &\prec_w \text{Spr}^+(A), \quad \text{Spr}^+(B') \prec_w \text{Spr}^+(B). \end{aligned}$$

Thus, in this case the result follows from the previous submajorization relations together with Eq. (44) and Eq. (45). \square

4.2 Proof of Theorem 3.3

Notation 4.4. Throughout this section we adopt the notation and terminology in Theorem 3.3. Hence, we consider:

1. $A \in \mathcal{H}(d)$;
2. $\mathcal{X}, \mathcal{Y} \subset \mathbb{C}^d$, k -dimensional subspaces such that \mathcal{X} is A -invariant;
3. a direct rotation $U = U(\mathcal{X}, \mathcal{Y}) \in \mathcal{U}(d)$, from \mathcal{X} onto \mathcal{Y} .
4. $Y_r = Y_r(\mathcal{X}, \mathcal{Y}) \stackrel{\text{def}}{=} UX \in \mathcal{I}_{\mathcal{Y}}(k, d)$.
5. We denote by $\Theta = \Theta(\mathcal{X}, \mathcal{Y})^\downarrow \in [0, \pi/2]^k$ the principal angles between \mathcal{X} and \mathcal{Y} . \triangle

Our approach to prove the result is similar to that in the previous section. We first prove Theorem 3.3 in the particular case that $k = \dim \mathcal{X} = \dim \mathcal{Y} \leq d/2$ (see Propositions 4.6 and 4.7 below). Then, we reduce the general case to the previous particular case. In what follows we consider the decomposition

$$\mathbb{C}^d = \mathcal{X} \oplus \mathcal{X}^\perp \quad \text{assuming that} \quad \mathcal{X} \cap \mathcal{Y} = \{0\} \quad \text{and} \quad \mathcal{X}^\perp \cap \mathcal{Y}^\perp = \{0\}.$$

Notice that in this case we have that

$$A = \begin{pmatrix} A_{11} & 0 \\ 0 & A_{22} \end{pmatrix} \quad \text{i.e.} \quad A_{12} = A_{21}^* = 0. \quad (46)$$

Given a direct rotation U from \mathcal{X} onto \mathcal{Y} we consider the block representation

$$U(t) = \begin{pmatrix} C(t) & -S(t) \\ S(t) & C(t) \end{pmatrix},$$

where $S(\cdot), C(\cdot) : [0, 1] \rightarrow \mathcal{H}(r)$ are given by

$$S(t) = \text{diag}(\sin(t\Theta)) \quad \text{and} \quad C(t) = \text{diag}(\cos(t\Theta)) \in \mathcal{M}_{p+r}(\mathbb{C})^+.$$

Here $\Theta = \Theta(\mathcal{X}, \mathcal{Y})^\downarrow = (\frac{\pi}{2} \mathbf{1}_p, \theta_1, \dots, \theta_r) \in [0, \pi/2]^k$ where, as before,

$$p = \dim \mathcal{X} \cap \mathcal{Y}^\perp = \dim \mathcal{X}^\perp \cap \mathcal{Y} \quad \text{and} \quad r = \dim \mathcal{X} \ominus (\mathcal{X} \cap \mathcal{Y}^\perp) = k - p.$$

In this case, we define

$$Y_r(t) = U(t) X = \begin{pmatrix} C(t) \\ S(t) \end{pmatrix} \quad \text{with} \quad X = \begin{pmatrix} I \\ 0 \end{pmatrix}, \quad (47)$$

where the matrix block representation is with respect to the decomposition $\mathbb{C}^d = \mathcal{X} \oplus \mathcal{X}^\perp$, as before. In this case, we set

$$\gamma(t) = Y_r(t)^* A Y_r(t) = X (U(t)^* A U(t)) X = X^* A(t) X \quad \text{for} \quad t \in [0, 1], \quad (48)$$

where $A(t) = U(t)^* A U(t)$. By taking derivatives,

$$\gamma'(t) = X^* A'(t) X = X^* (A(t) U'(0) - U'(0) A(t)) X, \quad (49)$$

where we have used that $U(t+h) = U(t)(h)$, for $t, h, t+h \in [0, 1]$, so that $U'(t) = U(t)U'(0)$ and similarly $(U'(t))^* = -U'(0)U(t)^*$. In a similar way as in Eq. (34), using the previous block matrix representations and setting $A(t) := (A_{ij}(t))_{i,j=1}^2$, we have that

$$\gamma'(t) = X^* (A(t)U'(0) - U'(0)A(t))X = 2 \operatorname{Re}(A_{12}(t)D_\Theta). \quad (50)$$

Now, using the fact that \mathcal{X} is an A -invariant subspace (so that A has the block representation in Eq. (46)), if we consider the block matrix representation $A(t) = U(t)^*AU(t) = (A_{ij}(t))_{i,j=1}^2$ then, for $t \in [0, 1]$

$$A_{12}(t) = S(t)A_{22}C(t) - C(t)A_{11}S(t). \quad (51)$$

Equation (51) shows that \mathcal{X} might not be $A(t)$ -invariant, for $t \in (0, 1]$; nevertheless, based on the fact that \mathcal{X} is A -invariant, we show that it is possible to obtain a convenient upper bound for the growth of the singular values $\gamma'(t)$, for $t \in (0, 1]$. Indeed, using Eq. (51) we see that

$$2 \operatorname{Re}(A_{12}(t)D_\Theta) = 2 \operatorname{Re}(S(t)A_{22}C(t) - C(t)A_{11}S(t))D_\Theta$$

Moreover, since $S(t)$, $C(t)$ and D_Θ commute with each other, then

$$2 \operatorname{Re}(A_{12}(t)D_\Theta) = 2 \operatorname{Re}(S(t)(A_{22}D_\Theta - D_\Theta A_{11})C(t)). \quad (52)$$

Using Weyl's multiplicative inequality (item 3 of Theorem 5.1) and that given $E \in \mathcal{M}_k(\mathbb{C})$, then $s(\operatorname{Re}(E)) \prec_w s(E)$, we get from Eqs. (50) and (52) that

$$s(\gamma'(t)) = 2s(\operatorname{Re}(A_{12}(t)D_\Theta)) \prec_w 2t\Theta s(A_{22}D_\Theta - D_\Theta A_{11}), \quad (53)$$

where we have used that $s(S(t)) = \lambda(S(t)) = \sin(t\Theta) \leq t\Theta$ and that $0 \leq C(t) \leq I$.

We will also consider the following inequality for the singular values of the so-called generalized commutators.

Theorem 4.5 (See [14]). Let $A_1, A_2 \in \mathcal{H}(k)$ be arbitrary and let $D \in \mathcal{M}_k(\mathbb{C})$. Then

$$s(A_1D - DA_2) \prec_w s(D) \operatorname{Spr}^+(A_1 \oplus A_2). \quad (54)$$

Proposition 4.6. *With the notation and terminology of Notation 4.4, assume that $\mathcal{X} \cap \mathcal{Y} = \{0\}$. Then we have that*

$$s(X^*AX - Y_r^*AY_r) \prec_w \Theta^2 \operatorname{Spr}^+ A, \quad (55)$$

Proof. We first assume that $\mathcal{X}^\perp \cap \mathcal{Y}^\perp = \{0\}$. Under this hypothesis, we consider the notation and terminology introduced previously in this section. In particular, we consider the smooth curve $\gamma(t)$ for $t \in [0, 1]$ introduced in Eq. (48). Since $\gamma(0) = X^*AX$ and $\gamma(1) = Y_r^*AY_r$ then, Proposition 4.1 shows that

$$s(X^*AX - Y_r^*AY_r) \prec_w \int_0^1 s(\gamma'(t)) dt. \quad (56)$$

By Eq. (53) above, we have that

$$s(\gamma'(t)) \prec_w 2t\Theta s(A_{22}D_\Theta - D_\Theta A_{11}), \quad (57)$$

By Theorem 4.5 we get

$$s(A_{22} D_{\Theta} - D_{\Theta} A_{11}) \prec_w \Theta \operatorname{Spr}^+(A_{11} \oplus A_{22}) = \Theta \operatorname{Spr}^+(A), \quad (58)$$

since $A = A_{11} \oplus A_{22}$ by hypothesis (\mathcal{X} is A -invariant). Therefore, by Eqs. (57) and (58),

$$s(\gamma'(t)) \prec_w 2t \Theta^2 \operatorname{Spr}^+(A) \stackrel{(56)}{\implies} s(X^* A X - Y_r^* A Y_r) \prec_w \Theta^2 \operatorname{Spr}^+ A,$$

where we have used items 5. and 3. in Lemma 5.3. Finally, in case $\mathcal{X}^{\perp} \cap \mathcal{Y}^{\perp} \neq \{0\}$ we can argue by reduction to the case in which $\mathcal{X}^{\perp} \cap \mathcal{Y}^{\perp} = \{0\}$, in the same way as in the last part of the proof of Theorem 3.2, and prove the inequality in Eq. (55) (without the restriction $\mathcal{X}^{\perp} \cap \mathcal{Y}^{\perp} = \{0\}$). The details are left to the reader. \square

Proposition 4.7. *With the notation and terminology of Notation 4.4, assume that $\mathcal{X}, \mathcal{Y} \subset \mathbb{C}^d$ be k -dimensional subspaces with $k \leq d/2$. Then we have that*

$$s(X^* A X - Y_r^* A Y_r) \prec_w \Theta^2 \operatorname{Spr}^+ A, \quad (59)$$

Proof. Assume that $k = \dim \mathcal{X} = \dim \mathcal{Y} \leq d/2$ and that \mathcal{X} and \mathcal{Y} are arbitrary subspaces such that \mathcal{X} is A -invariant. If $\mathcal{X} \cap \mathcal{Y} = \{0\}$ then we conclude that Eq. (59) holds in this case, by Proposition 4.6. In case $\mathcal{X} \cap \mathcal{Y} \neq \{0\}$ we consider the following

Claim: for every $t \in [0, 1]$ there exist a k -dimensional subspace $\mathcal{Y}(t) \subset \mathbb{C}^d$ and a direct rotation $W(t) \in \mathcal{U}(d)$ from \mathcal{X} onto $\mathcal{Y}(t)$ in such a way that $\mathcal{X} \cap \mathcal{Y}(t) = \{0\}$ for $t \in (0, 1]$,

$$\lim_{t \rightarrow 0^+} P_{\mathcal{Y}(t)} = P_{\mathcal{Y}} \in \mathcal{H}(d) \quad \text{and} \quad \lim_{t \rightarrow 0^+} W(t) = U.$$

Let us assume for a moment that the claim is true. Let $Y_{r,t} = W(t) X$, for $t \in [0, 1]$, and notice that $\lim_{t \rightarrow 0^+} Y_{r,t} = U X = Y_r$ (as in the statement of the theorem). Since $\mathcal{X} \cap \mathcal{Y}(t) = \{0\}$ for $t \in (0, 1]$, by Proposition 4.6 we have that

$$s(X^* A X - Y_{r,t}^* A Y_{r,t}) \prec_w \Theta(\mathcal{X}, \mathcal{Y}(t))^2 \operatorname{Spr}^+(A) \quad \text{for } t \in (0, 1].$$

By continuity of singular values (and hence of principal angles) we conclude that

$$\begin{aligned} s(X^* A X - Y_r^* A Y_r) &= \lim_{t \rightarrow 0^+} s(X^* A X - Y_{r,t}^* A Y_{r,t}) \prec_w \lim_{t \rightarrow 0^+} \Theta(\mathcal{X}, \mathcal{Y}(t))^2 \operatorname{Spr}^+(A) \\ &= \Theta(\mathcal{X}, \mathcal{Y})^2 \operatorname{Spr}^+(A). \end{aligned}$$

Proof of the claim. In case $\dim(\mathcal{X} \cap \mathcal{Y}) = s \geq 1$, we notice that

$$\dim \mathcal{X}^{\perp} \cap \mathcal{Y}^{\perp} = \dim \mathcal{X}^{\perp} + \dim \mathcal{Y}^{\perp} - \dim(\mathcal{X} \cap \mathcal{Y})^{\perp} \geq s.$$

On the other hand, we have the decompositions

$$\mathcal{X} = (\mathcal{X} \cap \mathcal{Y}) \oplus (\mathcal{X} \cap \mathcal{Y}^{\perp}) \oplus (\mathcal{X} \cap \mathcal{G}) \quad \text{and} \quad \mathcal{Y} = (\mathcal{X} \cap \mathcal{Y}) \oplus (\mathcal{X}^{\perp} \cap \mathcal{Y}) \oplus (\mathcal{Y} \cap \mathcal{G}),$$

where \mathcal{G} denotes the generic part of the pair of subspaces \mathcal{X} and \mathcal{Y} (see Eq. (9)). Let $\Theta' = (\frac{\pi}{2} \mathbf{1}_p, \theta_1, \dots, \theta_r)$ denote the principal angles between $\mathcal{S}_1 := (\mathcal{X} \cap \mathcal{Y}^{\perp}) \oplus (\mathcal{X} \cap \mathcal{G})$ and $(\mathcal{X}^{\perp} \cap \mathcal{Y}) \oplus (\mathcal{Y} \cap \mathcal{G})$, where $p = \dim \mathcal{X} \cap \mathcal{Y}^{\perp} = \dim \mathcal{X}^{\perp} \cap \mathcal{Y}$ and $r = \dim \mathcal{X} \cap \mathcal{G} = \dim \mathcal{Y} \cap \mathcal{G}$; then, $\Theta = \Theta(\mathcal{X}, \mathcal{Y}) = (\Theta', 0_s) \in [0, \pi/2]^k$, where $k = p + r + s$.

Let $U \in \mathcal{U}(d)$ be a direct rotation from \mathcal{X} onto \mathcal{Y} . Then there exists \mathcal{B} an ONB for \mathbb{C}^d , compatible with the decomposition

$$\mathbb{C}^d = (\mathcal{X} \cap \mathcal{Y}) \oplus \mathcal{S}_1 \oplus \mathcal{S}_2 \oplus (\mathcal{X}^\perp \cap \mathcal{Y}^\perp), \quad (60)$$

where $\mathcal{S}_2 = (\mathcal{X}^\perp \cap \mathcal{Y}) \oplus (\mathcal{X}^\perp \cap \mathcal{G})$, such that

$$U = \begin{pmatrix} I & 0 & 0 & 0 \\ 0 & C & -S & 0 \\ 0 & S & C & 0 \\ 0 & 0 & 0 & I \end{pmatrix}, \quad (61)$$

where $C = \text{diag}(\cos(\Theta'))$ and $S = \text{diag}(\sin(\Theta')) \in \mathcal{M}_{p+r}(\mathbb{C})^+$

are diagonal positive semidefinite matrices.

Let $\{v_j\}_{j \in \mathbb{I}_s}$ be the vectors in \mathcal{B} that form an ONB of $\mathcal{X} \cap \mathcal{Y}$ and let $\{w_j\}_{j \in \mathbb{I}_s}$ be vectors in \mathcal{B} that form an ONS in $\mathcal{X}^\perp \cap \mathcal{Y}^\perp$ (here we are using that $\dim \mathcal{X}^\perp \cap \mathcal{Y}^\perp \geq s$). We now set $u_j(t) = \cos(t)v_j + \sin(t)w_j$, for $j \in \mathbb{I}_s$, and let $\mathcal{S}(t)$ be the subspace generated by the ONS $\{u_j(t)\}_{j \in \mathbb{I}_s}$, for $t \in [0, 1]$. Also, let

$$\mathcal{Y}(t) = \mathcal{S}(t) \oplus (\mathcal{X}^\perp \cap \mathcal{Y}) \oplus (\mathcal{Y} \cap \mathcal{G}) \quad \text{for } t \in [0, 1].$$

By construction, $\dim \mathcal{Y}(t) = k$, and for $0 < t \leq \theta_r$ we have that

$$\Theta(\mathcal{X}, \mathcal{Y}(t)) = (\theta_j(t))_{j \in \mathbb{I}_k} = (\Theta', t \mathbf{1}_s) \in (0, \pi/2]^k.$$

In particular, $\mathcal{X} \cap \mathcal{Y}(t) = \{0\}$ for $0 < t < \theta_r$. It is clear that $\lim_{t \rightarrow 0^+} P_{\mathcal{Y}(t)} = P_{\mathcal{Y}}$.

Consider the orthogonal decomposition

$$\mathbb{C}^d = \mathcal{Z}_1 \oplus \mathcal{Z}_2 \oplus \mathcal{T} \quad (62)$$

where $\mathcal{T} = (\mathcal{X}^\perp \cap \mathcal{Y}^\perp) \ominus \mathcal{S}(\frac{\pi}{2})$,

$$\mathcal{Z}_1 = (\mathcal{X} \cap \mathcal{Y}) \oplus (\mathcal{X} \cap \mathcal{Y}^\perp) \oplus (\mathcal{X} \cap \mathcal{G}) \quad \text{and} \quad \mathcal{Z}_2 = (\mathcal{X}^\perp \cap \mathcal{Y}) \oplus (\mathcal{X}^\perp \cap \mathcal{G}) \oplus \mathcal{S}(\frac{\pi}{2}).$$

We now construct the block matrix $W(t)$ in terms of the decomposition in Eq. (62), given by

$$W(t) = \begin{pmatrix} C_1(t) & -S_1(t) & 0 \\ S_1(t) & C_1(t) & 0 \\ 0 & 0 & I \end{pmatrix} \quad \text{for } t \in [0, 1], \quad (63)$$

where $C_1(t) = \text{diag}(\cos(t) \mathbf{1}_s, \cos(\Theta'))$ and $S_1(t) = \text{diag}(\sin(t) \mathbf{1}_s, \sin(\Theta'))$. Then, $W(\cdot)$ is a continuous function such that $W(t)$ is a direct rotation from \mathcal{X} onto $\mathcal{Y}(t)$, for $t \in [0, 1]$, and such that $W(0) = U$ (compare Eqs. (61) and (63)). The claim follows from these remarks. \square

We can now prove our second main result. We will re-write the statement with all its notation and terminology for the reader's convenience.

Theorem 3.3. Let $A \in \mathcal{H}(d)$ and let $\mathcal{X}, \mathcal{Y} \subset \mathbb{C}^d$ be k -dimensional subspaces such that \mathcal{X} is A -invariant. Let $U = U(\mathcal{X}, \mathcal{Y})$ be a direct rotation of \mathcal{X} onto \mathcal{Y} and $\Theta = \Theta(\mathcal{X}, \mathcal{Y})^\downarrow \in$

$[0, \pi/2]^k$ the principal angles between \mathcal{X} and \mathcal{Y} defined in (8). Given $X \in \mathcal{I}_{\mathcal{X}}(k, d)$, if we let $Y_r = Y_r(X, U) = UX \in \mathcal{I}_{\mathcal{Y}}(k, d)$ then we have that

$$s(X^*AX - Y_r^*AY_r) \prec_w \Theta^2(\mathcal{X}, \mathcal{Y})^\downarrow \text{Spr}^+(A), \quad (64)$$

where submajorization and products are as in Remark 2.2.

Proof. In case $k \leq d/2$, the result follows from Proposition 4.7. On the other hand, if $k > d/2$ we embed the subspaces \mathcal{X}, \mathcal{Y} in $\mathbb{C}^{d'} = \mathbb{C}^d \oplus \mathbb{C}^{(d'-d)}$ for some $k \leq d'/2$ and get $\mathcal{X}' = \mathcal{X} \oplus 0_{(d'-d)}$ and $\mathcal{Y}' = \mathcal{Y} \oplus 0_{(d'-d)}$. Notice that in this case $\Theta(\mathcal{X}', \mathcal{Y}') = (\Theta(\mathcal{X}, \mathcal{Y}), 0_{(d'-d)})$.

Similarly, we can embed X, A and U and get $X' = \begin{pmatrix} X \\ 0 \end{pmatrix} \in \mathcal{I}_{\mathcal{X}'}(k, d')$,

$$A' = \begin{pmatrix} A & 0 \\ 0 & \lambda_h(A)I \end{pmatrix} \in \mathcal{H}(d') \quad \text{and} \quad U' = \begin{pmatrix} U & 0 \\ 0 & I \end{pmatrix} \in \mathcal{U}(d'),$$

where $h = \lfloor \frac{d'+1}{2} \rfloor$ (integer part). In this case U' is a direct rotation of \mathcal{X}' onto \mathcal{Y}' such that

$$Y'_r = U'X' = \begin{pmatrix} UX \\ 0 \end{pmatrix} = \begin{pmatrix} Y_r \\ 0 \end{pmatrix}.$$

Now it is straightforward to check that

$$(X')^*A'X' - (Y'_r)^*A'Y'_r = \begin{pmatrix} X^*AX - Y_r^*AY_r & \\ & 0 \end{pmatrix}.$$

Hence, by Proposition 4.7 we get that

$$s(X^*AX - Y_r^*AY_r) = s((X')^*A'X' - (Y'_r)^*A'Y'_r) \prec_w \Theta(\mathcal{X}', \mathcal{Y}')^2 \text{Spr}^+(A').$$

Notice that, by construction,

$$\text{Spr}^+(A') = (\text{Spr}^+(A), 0_{h'}) \implies \Theta(\mathcal{X}', \mathcal{Y}')^2 \text{Spr}^+(A') = \Theta(\mathcal{X}, \mathcal{Y})^2 \text{Spr}^+(A),$$

where $h' = \lfloor \frac{d'}{2} \rfloor - \lfloor \frac{d}{2} \rfloor \geq 1$. □

Remark 4.8 (Final comments). Consider the notation of Theorem 3.4. The reader could wonder why is it that our bounds do not coincide with the bounds in Eqs. (2) and (3). Now that we have described our techniques in detail we can give our opinion on this issue. We believe that the fact that our bounds depend on the $\Theta(\mathcal{X}, \mathcal{Y})$, which from a geometric point of view corresponds to the arc length between subspaces, is a consequence of our choice of curves that lead to the construction of $\gamma(t)$ as in Eqs. (30) and (48), based on direct rotations. On the other hand $\sin(\Theta(\mathcal{X}, \mathcal{Y}))$, which from a geometric point of view corresponds to the chordal length between subspaces, seems to be associated with shorter curves $\tilde{\gamma}(t)$. Still, the importance of our choice $\gamma(t)$ is that its derivative $\gamma'(t)$ leads to singular values inequalities that allow to reduce the problem to the infinitesimal level (these inequalities were obtained in [14], motivated by the problems in the present paper). For example, notice that the shortest curve $\mu(t) = (1-t)X^*AX - tY^*AY$, $t \in [0, 1]$, joining X^*AX and Y^*AY does not provide such reduction, since $\mu'(t) = Y^*AY - X^*AX$, for $t \in [0, 1]$. Nevertheless, the geometric technique considered in here is rather flexible in several ways, and we plan to keep working in these problems in the future. △

5 Appendix

Here we collect several results about majorization and the spectral spread of self-adjoint matrices, used throughout our work.

5.1 Majorization theory in matrix analysis

There are many fundamental results in matrix theory that are stated in terms of submajorization relations. Below, we mention only those results related to the content of this work; for a detailed exposition on majorization theory see [2, Chapters II and III], [7, Chapter 3] and [12, Chapter 9].

The first result deals with submajorization relations between singular values of arbitrary matrices in $\mathcal{M}_d(\mathbb{C})$ (see [2] p. 35, p. 74 and p. 94).

Theorem 5.1. Let $C, D \in \mathcal{M}_d(\mathbb{C})$. Then,

1. Weyl's additive inequality for singular values : $s(C + D) \prec_w s(C) + s(D)$;
2. $s(\operatorname{Re}(C)) \prec_w s(C)$;
3. Weyl's multiplicative inequality for singular values: $s(CD) \prec_w s(C) s(D)$;

□

Theorem 5.2. ([2], *Thm.III.4.1*) Let $C, D \in \mathcal{H}(d)$. Then,

1. Lidskii's additive inequality: $\lambda(C) - \lambda(D) \prec \lambda(C - D)$;
2. Lidskii's additive inequality for singular values: $|\lambda(C) - \lambda(D)| \prec_w s(C - D)$;

□

In the next result we describe several elementary but useful properties of majorization and submajorization between real vectors.

Lemma 5.3. Let $x, y, z, w \in \mathbb{R}^k$. Then,

1. $x^\downarrow + y^\uparrow \prec x + y \prec x^\downarrow + y^\downarrow$;
2. If $x \prec_w y$ and $y, z \in (\mathbb{R}^k)^\downarrow$ then $x + z \prec_w y + z$;
3. If $z, w \in (\mathbb{R}^k)^\downarrow$, $x \prec z$ and $y \prec w$ then $x + y \prec z + w$. Moreover, if $x(\cdot), z(\cdot) : [0, 1] \rightarrow \mathbb{R}^k$ are continuous functions such that $x(t) \prec_w z(t) = (z(t))^\downarrow$ for $t \in [0, 1]$, then

$$\int_0^1 x(t) dt \prec_w \int_0^1 z(t) dt.$$

If we assume further that $x, y, z \in \mathbb{R}_{\geq 0}^k$ then,

4. $x^\downarrow y^\uparrow \prec_w xy \prec_w x^\downarrow y^\downarrow$;
5. If $x \prec_w y$ and $y, z \in (\mathbb{R}_{\geq 0}^k)^\downarrow$ then $xz \prec_w yz$.

□

Proof. A proof of all these facts can be found in [2] with the exception of the second assertion in item 3, that we now prove. Indeed, for $n \in \mathbb{N}$ consider the regular partition $\{t_0 = 0 < t_1 < \dots < t_n = 1\}$ of $[0, 1]$ so that $t_j = \frac{j}{n}$ for $j \in \{0\} \cup \mathbb{I}_n$ and $\Delta_j = t_j - t_{j-1} = \frac{1}{n} = \Delta_n$, for $j \in \mathbb{I}_n$. By the first part of item 3. we have that, for $n \in \mathbb{N}$,

$$\sum_{j=1}^n x(t_j) \Delta_n \prec_w \sum_{j=1}^n z(t_j) \Delta_n.$$

Since $x(\cdot), z(\cdot)$ are continuous functions, we have that

$$\int_0^1 x(t) dt = \lim_n \sum_{j=1}^n x(t_j) \Delta_n \prec_w \lim_n \sum_{j=1}^n z(t_j) \Delta_n = \int_0^1 z(t) dt,$$

where we have used the following fact: if $(u_n)_{n \in \mathbb{N}}$ and $(v_n)_{n \in \mathbb{N}} \in \mathbb{R}^d$ are sequences that converge to u and $v \in \mathbb{R}^d$ respectively, and such that $u_n \prec_w v_n$ for $n \in \mathbb{N}$, then $u \prec_w v$. \square

Remark 5.4. Let $x, y \in \mathbb{R}^k$. If $x \leq y$ then,

$$x^\downarrow \leq y^\downarrow \quad \text{and} \quad x \prec_w y.$$

In particular, if $A, P \in \mathcal{M}_d(\mathbb{C})$ are such that P is a projection then $s(PAP) \leq s(A)$ so that $s(PAP) \prec_w s(A)$. \triangle

Proposition 5.5. Let $1 \leq k < d$ and let $E \in \mathcal{M}_{k, (d-k)}(\mathbb{C})$. Then

$$\hat{E} = \begin{pmatrix} 0 & E \\ E^* & 0 \end{pmatrix} \in \mathcal{H}(d) \quad \text{and} \quad \lambda(\hat{E}) = (s(E), -s(E^*))^\downarrow \in (\mathbb{R}_{\geq 0}^d)^\downarrow. \quad (65)$$

\square

5.2 Spectral spread of self-adjoint matrices

In this section we describe a Weyl's type inequality for the spectral spread. Recall that given $A \in \mathcal{H}(d)$ then the spectral spread of A is

$$\text{Spr}^+(A) = (\text{Spr}_j(A))_{j \in \mathbb{I}_h} = (\lambda_j(A) - \lambda_j^\uparrow(A))_{j \in \mathbb{I}_h} \in (\mathbb{R}_+^h)^\downarrow.$$

where $h = \lfloor \frac{d}{2} \rfloor$ (integer part).

Proposition 5.6. Let $A, B \in \mathcal{H}(d)$. Then

$$\text{Spr}^+(A + B) \prec \text{Spr}^+(A) + \text{Spr}^+(B).$$

Proof. By Lidskii's additive inequality

$$\lambda(A + B) \prec \lambda(A) + \lambda(B) \implies -\lambda(A + B) \prec -\lambda(A) - \lambda(B).$$

By item 3 of Lemma 5.3, since $-\lambda^\uparrow(A) - \lambda^\uparrow(B) \in (\mathbb{R}^d)^\downarrow$, then

$$\begin{aligned} \lambda(A + B) - \lambda^\uparrow(A + B) &\prec \lambda(A) + \lambda(B) - \lambda^\uparrow(A) - \lambda^\uparrow(B) \\ &= (\lambda(A) - \lambda^\uparrow(A)) + (\lambda(B) - \lambda^\uparrow(B)) \in (\mathbb{R}^d)^\downarrow. \end{aligned}$$

Let $h = \lfloor \frac{d}{2} \rfloor$ (integer part). If $r \in \mathbb{I}_h$,

$$\begin{aligned}
\sum_{i=1}^r \text{Spr}_i^+(A+B) &= \sum_{i=1}^r \lambda_i(A+B) - \lambda_i^\uparrow(A+B) \\
&\leq \sum_{i=1}^r (\lambda_i(A) - \lambda_i^\uparrow(A)) + (\lambda_i(B) - \lambda_i^\uparrow(B)) \\
&= \sum_{i=1}^r \text{Spr}_i^+(A) + \text{Spr}_i^+(B).
\end{aligned} \tag{66}$$

□

References

- [1] M.E. Argentati, A.V. Knyazev, C.C. Paige, I. Panayotov, Bounds on changes in Ritz values for a perturbed invariant subspace of a Hermitian matrix. *SIAM J. Matrix Anal. Appl.* 30 (2008), no. 2, 548-559.
- [2] R. Bhatia, *Matrix analysis*, 169, Springer-Verlag, New York, 1997.
- [3] N. Bosner, Z. Drmač, Subspace gap residuals for Rayleigh-Ritz approximations. *SIAM J. Matrix Anal. Appl.* 31 (2009), no. 1, 54-67.
- [4] C. Davis, W.M. Kahan, The rotation of eigenvectors by a perturbation. III. *SIAM J. Numer. Anal.* 7 1970 1-46.
- [5] G.H. Golub, C.F. Van Loan, *Matrix computations*. Fourth edition. Johns Hopkins Studies in the Mathematical Sciences. Johns Hopkins University Press, 2013.
- [6] P.R. Halmos, Two subspaces. *Trans. Amer. Math. Soc.* 144 1969 381-389.
- [7] R.A. Horn, C.R. Johnson, *Topics in Matrix analysis*, Cambridge University Press, Cambridge, 1991.
- [8] A.V. Knyazev, M.E. Argentati, On proximity of Rayleigh quotients for different vectors and Ritz values generated by different trial subspaces. *Linear Algebra Appl.* 415 (2006), no. 1, 82-95.
- [9] A.V. Knyazev, M.E. Argentati, Majorization for changes in angles between subspaces, Ritz values, and graph Laplacian spectra. *SIAM J. Matrix Anal. Appl.* 29 (2006/07), no. 1, 15-32.
- [10] A.V. Knyazev, M.E. Argentati, Rayleigh-Ritz majorization error bounds with applications to FEM. *SIAM J. Matrix Anal. Appl.* 31 (2009), no. 3, 1521-1537.
- [11] C.-K. Li, R.-C. Li, A note on eigenvalues of perturbed Hermitian matrices. *Linear Algebra Appl.* 395 (2005), 183-190.
- [12] A.W. Marshall, I. Olkin, B.C. Arnold, *Inequalities: theory of majorization and its applications*. Second edition. Springer Series in Statistics. Springer, New York, 2011.
- [13] P.G. Massey, D. Stojanoff, S. Zárate, Majorization bounds for Ritz values of self-adjoint matrices. *SIAM J. Matrix Anal. Appl.* 41 (2020), no. 2, 554-572.
- [14] P. Massey, D. Stojanoff, S. Zárate, The spectral spread of Hermitian matrices, *Linear Algebra Appl.* 616 (2021), 19-44.
- [15] R. Mathias, Quadratic residual bounds for the Hermitian eigenvalue problem. *SIAM J. Matrix Anal. Appl.* 19 (1998), no. 2, 541-550.
- [16] E. Ovtchinnikov, Cluster robust error estimates for the Rayleigh-Ritz approximation. II. Estimates for eigenvalues. *Linear Algebra Appl.* 415 (2006), no. 1, 188-209.
- [17] B.N. Parlett, *The symmetric eigenvalue problem*. Corrected reprint of the 1980 original, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1998.
- [18] Y. Saad, *Numerical Methods for Large Eigenvalue Problems*, Revised Edition, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2011.
- [19] G.W. Stewart, J. Sun, *Matrix perturbation theory*, Computer Science and Scientific Computing, Academic Press Inc., Boston, MA, 1990.

- [20] Z. Teng, L. Lu, R.-C. Li, Cluster-robust accuracy bounds for Ritz subspaces. *Linear Algebra Appl.* 480 (2015), 11-26.
- [21] P. Zhu, M.E. Argentati, A.V. Knyazev, Bounds for the Rayleigh quotient and the spectrum of self-adjoint operators. *SIAM J. Matrix Anal. Appl.* 34 (2013), no. 1, 244-256.
- [22] P. Zhu, A.K. Knyazev, Rayleigh-Ritz majorization error bounds of mixed type. *SIAM J. Matrix Anal. Appl.* 38 (2017), no. 1, 30-49.