MAJORIZATION BOUNDS FOR RITZ VALUES OF SELF-ADJOINT MATRICES *

PEDRO MASSEY†, DEMETRIO STOJANOFF†, AND SEBASTIÁN ZÁRATE†

Abstract. A priori, a posteriori, and mixed type upper bounds for the absolute change in Ritz values of self-adjoint matrices in terms of submajorization relations are obtained. Some of our results prove recent conjectures by Knyazev, Argentati, and Zhu, which extend several known results for one dimensional subspaces to arbitrary subspaces. In addition, we improve Nakatsukasa's version of the $\tan\Theta$ theorem of Davis and Kahan. As a consequence, we obtain new quadratic a posteriori bounds for the absolute change in Ritz values.

Key words. Principal angles, Rayleigh quotients, Ritz values, majorization.

AMS subject classifications. 42C15, 15A60.

1. Introduction. The study of sensitivity of Ritz values of Rayleigh quotients of self-adjoint matrices (i.e. the changes in the eigenvalues of compressions of a self-adjoint matrix) is a well established and active research field in applied mathematics [1, 3, 8, 9, 10, 11, 13, 15, 18, 19, 20, 21]. Explicitly, given a $d\times d$ complex self-adjoint matrix A and isometries X, Y of size $d\times k$, with ranges $\mathcal{X}$ and $\mathcal{Y}$ respectively, we are interested in computing upper and lower bounds for

$$|\lambda(\rho(X))-\lambda(\rho(Y))| = (\,|\lambda_i(\rho(X))-\lambda_i(\rho(Y))|\,)_{i\in\mathbb{I}_k}\in\mathbb{R}^k_{\geq 0}$$

where $\rho(X)=X^*A\,X$, $\rho(Y)=Y^*A\,Y$ are $k\times k$ complex self-adjoint matrices known as Rayleigh quotients (RQ) of A, and $\lambda(\rho(X))$, $\lambda(\rho(Y))\in\mathbb{R}^k$ are the eigenvalues (counting multiplicities and arranged in non-increasing order) also known as Ritz values.

Typically, the bounds for the absolute change in the Ritz values are obtained in terms of the residuals $R_X = AX - X\,\rho(X)$ and $R_Y = AY - Y\,\rho(Y)$ or in terms of the principal angles between subspaces (PABS) denoted by $\Theta(\mathcal{X},\mathcal{Y})\in[0,\pi/2]^k$. Upper bounds are classified according to which parameters are used to bound the change in Ritz values (see [19]). Indeed, the *a priori* bounds are those obtained in terms of PABS; the *a posteriori* bounds are those obtained in terms of (singular values of) residuals while the *mixed type* bounds are obtained in terms of both PABS and residuals. It is worth pointing out that PABS appearing in a priori bounds may not be readily available in practice. On the other hand, a posteriori bounds are based on computable singular values of residual matrices. Moreover, bounds based on residuals (i.e. both a posteriori and mixed type) are particularly convenient in case one of the spaces, say $\mathcal{X}$, is A-invariant (as in this case $R_X=0$), as opposed to (autonomous) a priori bounds.

The abstract matrix analysis formulation of the sensitivity problem stated above makes it possible to apply this theory in a variety of different research areas such as: graph matching [9] in terms of spectral analysis of the graphs; signal distinction in signal processing, where Ritz values serve as harmonic signature to differentiate

*Revised version submitted to the editors on September, 2019.

Funding: Partially supported by CONICET (PICT ANPCyT 1505/15) and Universidad Nacional de La Plata (UNLP 11X829)

†Centro de Matemática, FCE-UNLP and IAM-CONICET, Buenos Aires, Argentina (massey@mate.unlp.edu.ar, demetrio@mate.unlp.edu.ar, seb4.zarate@gmail.com).

subspaces; finite element methods (FEM) [8], for approximation of subspaces corresponding to fundamental modes; of course, matrix analysis, e.g. for bounds for eigenvalues after matrix additive perturbations. Also, bounds for changes in Ritz values play a central role in the analysis of algorithms for simultaneous approximation of eigenvalues based on Rayleigh-Ritz methods (see [16, 17] and the references therein). By now, the role of submajorization in obtaining bounds for the change of Ritz values (recognized in the seminal paper [9]) is well known; this partial pre-order relation is a powerful tool in this context, as bounds in terms of submajorization imply a whole family of inequalities with respect to unitarily invariant norms and with respect to the class of non-decreasing convex functions ([12]).

In this work we obtain a priori, a posteriori and mixed type upper bounds for the absolute change in Ritz values of self-adjoint matrices in terms of submajorization. Some of our results prove recent conjectures from [8, 19, 20] which extend several known results for one dimensional subspaces to arbitrary subspaces. In addition, we improve Nakatsukasa's version of the $\tan\Theta$ theorem [14] of Davis and Kahan [4]. We have included some (rather simple) examples to establish comparisons with previous work (for a detailed exposition of the context, previous work, our results and some applications, see Section 3). We will consider further applications of the results herein elsewhere.

The paper is organized as follows. In Section 2 we introduce preliminary results in majorization theory and principal angles between subspaces. In Section 3 we develop our main results; our approach to obtain these results is based on methods from abstract matrix analysis, so we delay the proofs of some technical results until an appendix section. Section 3 is divided in three subsections: in Section 3.1 we prove a mixed type upper bound for the change of the Ritz values that is conjectured in [20] and show that this bound is sharp. We have also included some comments with a comparison of our results with previous works and with future applications of the results of this subsection. In Section 3.2 we establish a link between the results from Section 3.1 and an a priori upper bound for Ritz values conjectured from [8]. Although the results in this section are not sharp, they can be applied in quite general situations and they capture the order of approximation conjectured in [8]. In Section 3.3 we revisit Nakatsukasa's version of the $\tan\Theta$ theorem of Davis and Kahan and obtain an improved version of this result; we include an example that shows that this new version of the $\tan\Theta$ theorem is sharp in cases in which the classical result is not. As an application, we obtain improved quadratic a posteriori error bounds for Ritz values. The paper ends with an Appendix (Section 4) in which we include a detailed background on majorization theory and present the proofs of some technical results needed in Section 3.

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 set of self-adjoint matrices and by $\mathcal{M}_d(\mathbb{C})^+$, the cone of positive semi-definite matrices. Also, $\mathcal{G}l(d)\subset\mathcal{M}_d(\mathbb{C})$ and $\mathcal{U}(d)$ denote the groups of invertible and unitary matrices respectively, and $\mathcal{G}l(d)^+=\mathcal{G}l(d)\cap\mathcal{M}_d(\mathbb{C})^+$. On the other hand, given a subspace $\mathcal{Z}\subset\mathbb{C}^d$, we let $\mathcal{L}(\mathcal{Z})$ denote the space of linear operators acting on $\mathcal{Z}$.

For $d\in\mathbb{N}$, let $\mathbb{I}_d=\{1,\ldots,d\}$. Given a vector $x\in\mathbb{C}^d$ we denote by D_x 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 $(\mathbb{R}^d)^\downarrow = \{x \in \mathbb{R}^d \ : \ x = x^\downarrow\}$, $R_{\geq 0} = \{x \in \mathbb{R} : x \geq 0\}$ and $(\mathbb{R}^d_{\geq 0})^\downarrow = \{x \in \mathbb{R}^d_{\geq 0} \ : \ x = x^\downarrow\}$. For $r \in \mathbb{N}$, we let $\mathbb{1}_r = (1, \ldots, 1) \in \mathbb{R}^r$.

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 $B \in \mathcal{M}_d(\mathbb{C})$ we let $s(B) = \lambda(|B|)$ denote the singular values of B, i.e. the eigenvalues of $|B| = (B^*B)^{1/2} \in \mathcal{M}_d(\mathbb{C})^+$. We use the abbreviation ONB for "orthonormal basis".

Arithmetic operations with vectors are performed entry-wise i.e., in case $x = (x_i)_{i\in\mathbb{I}_k}$ and $y = (y_i)_{i\in\mathbb{I}_k} \in \mathbb{C}^k$ then $x + y = (x_i + y_i)_i$ and, following the notational convention of the principal references on these matters,

$$x\,y = (x_i\, y_i)_i \qquad \text{and (assuming that } y_i \neq 0, \text{ for } i \in \mathbb{I}_k) \qquad \frac{x}{y} = (x_i/y_i)_i \ ,$$

where these vectors all lie in $\mathbb{C}^k$. Moreover, if we assume further that $x, y \in \mathbb{R}^k$ then we write $x \leq y$ whenever $x_i \leq y_i$, for $i \in \mathbb{I}_k$. $\triangle$

Next we recall the notion of majorization between vectors, that 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

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

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

There are many fundamental results in matrix theory that are stated in terms of submajorization relations. In what follows, we mention some elementary properties of submajorization that we will need in Section 3 (for detailed expositions on majorization theory, including proofs of the results mentioned below, see [2, 6, 12]). We will consider some further properties and results on majorization theory in Section 4. Given $f : [a\,,\, b] \to \mathbb{R}$, where $[a\,,\, b] \subset \mathbb{R}$ is an interval, and $z = (z_i)_{i\in\mathbb{I}_k} \in [a\,,\, b]^k$ we denote $f(z) = (f(z_i))_{i\in\mathbb{I}_k} \in \mathbb{R}^k$.

Remark 2.2. *Let $[a\,,\, b] \subset \mathbb{R}$ be an interval and let $f : [a\,,\, b] \to \mathbb{R}$ be a convex function. Then,*

1. *if $x, y \in [a\,,\, b]^k$ satisfy $x \prec y$ then $f(x) \prec_w f(y)$.*
2. *If $x, y \in [a\,,\, b]^k$ only satisfy $x \prec_w y$ but f is further non-decreasing in $[a\,,\, b]$, then $f(x) \prec_w f(y)$.*

$\triangle$

Definition 2.3. A norm N in $\mathcal{M}_d(\mathbb{C})$ is **unitarily invariant** (briefly u.i.n.) if $N(UAV) = N(A)$, for every $A \in \mathcal{M}_d(\mathbb{C})$ and $U, V \in \mathcal{U}(d)$. $\triangle$

Well known examples of u.i.n. are the spectral norm $\|\cdot\|_{sp}$ and the Schatten p-norms $\|\cdot\|_p$, for $p \geq 1$.

Remark 2.4. It is well known that (sub)majorization relations between singular values of matrices are intimately related with inequalities with respect to u.i.n's. Indeed, given $A, B \in \mathcal{M}_d(\mathbb{C})$ the following statements are equivalent:

1. For every u.i.n. N in $\mathcal{M}_d(\mathbb{C})$ we have that $N(A) \leq N(B)$.
2. $s(A) \prec_w s(B)$. $\triangle$

Principal Angles Between Subspaces. Let $\mathcal{X}$, $\mathcal{Y} \subset \mathbb{C}^d$ denote subspaces, with $\dim \mathcal{X} = h$ and $\dim \mathcal{Y} = k$. Let $X \in \mathcal{M}_{d,h}$ and $Y \in \mathcal{M}_{d,k}$ be such that their columns form orthonormal bases of $\mathcal{X}$ and $\mathcal{Y}$ respectively. Then, the principal angles between $\mathcal{X}$ and $\mathcal{Y}$, denoted $\pi/2 \geq \Theta_1(\mathcal{X}, \mathcal{Y}) \geq \ldots \geq \Theta_m(\mathcal{X}, \mathcal{Y}) \geq 0$ where $m = \min\{h, k\}$ - are determined by

$$\cos(\Theta_{m-i+1}(\mathcal{X}, \mathcal{Y})) = s_i(X^*Y) \quad \text{for} \quad i \in \mathbb{I}_m \,.$$

We further write $\Theta(\mathcal{X}, \mathcal{Y}) = (\Theta_i(\mathcal{X}, \mathcal{Y}))_{i \in \mathbb{I}_m} \in (\mathbb{R}^m)^{\downarrow}$ for the vector of principal angles between $\mathcal{X}$ and $\mathcal{Y}$. Principal angles are a useful tool in describing the relative position and several geometric and metric aspects related with the subspaces $\mathcal{X}$ and $\mathcal{Y}$ in $\mathbb{C}^d$ (see [4, 5] and the references therein).

3. Main results.

In this section we develop our main results. The section is divided in three parts; first we prove [20, Conjecture 2.1] which establishes a mixed type bound for the error in the (absolute) change of the Ritz values. In the second part, we establish connections between the mixed type bounds of the first section and some a priori bounds for the change of Ritz values conjectured in [8, 10]. Finally we take a closer look at Nakatsukasa's $\tan \Theta$ theorem under relaxed conditions from [14] and obtain an improved version of this result. As a consequence we obtain quadratic a posteriori error bounds for the change of the Ritz values that improve several known bounds. Our approach to obtain these results is based on methods from abstract matrix analysis, so we delay the proofs of some technical results until Section 4, where we have also included several classical results of this area that we will refer to in this section.

We begin by introducing the following

Setting 3.1. Throughout this section we consider the following notation and terminology:

1. $\mathcal{X}$, $\mathcal{Y} \subset \mathbb{C}^d$ denote two subspaces of dimension k. We fix X, $Y \in \mathcal{M}_{d,k}(\mathbb{C})$ such that their columns form orthonormal bases of $\mathcal{X}$ and $\mathcal{Y}$, respectively.
2. $\Theta(\mathcal{X}\,,\,\mathcal{Y}) \in (\mathbb{R}^k_{\geq 0})^{\downarrow}$ denotes the vector of principal angles between the subspaces $\mathcal{X}$ and $\mathcal{Y}$; in this case,
$$\cos(\Theta^{\uparrow}(\mathcal{X}\,,\,\mathcal{Y})) = s(X^*Y) = (s_1(X^*Y), \ldots, s_k(X^*Y)) \in (\mathbb{R}^k_{\geq 0})^{\downarrow}.$$
3. For a (fixed) self-adjoint $A \in \mathcal{H}(d)$ we set $\rho(X) = X^*AX \in \mathcal{M}_k(\mathbb{C})$, $R_X = AX - X\rho(X) \in \mathcal{M}_{d,k}(\mathbb{C})$ and similarly $\rho(Y)$ and R_Y for Y. Notice that
$$R_X = AX - XX^*AX = AX - P_{\mathcal{X}}AX = P_{\mathcal{X}^\perp}AX \in \mathcal{M}_{d,k}(\mathbb{C})\,,$$
where $P_{\mathcal{X}} \in \mathcal{M}_d(\mathbb{C})$ denotes the orthogonal projection onto $\mathcal{X}$ and $\mathcal{X}^\perp$ denotes the orthogonal complement of $\mathcal{X}$. We consider similar notation and identities for $\mathcal{Y}$.
4. Let $X_\perp \in \mathcal{M}_{d\,,\,d-k}(\mathbb{C})$ be such that its columns form an ONB of $\mathcal{X}^\perp$. Then, the matrix $(X, X_\perp) \in \mathcal{U}(d)$ and we get
$$\tilde{A} = (X, X_\perp)\ \ A\ \ (X, X_\perp)^* = \begin{pmatrix} \rho(X) & R_X^*\,X_\perp \\ X_\perp^*\,R_X & \rho(X_\perp) \end{pmatrix}.$$

Note that, since $R_X = (I - P_{\mathcal{X}})\, R_X$, then $s(R_X) = s(X_\perp^* \, R_X)$, so that we can think of R_X (up to an isometric factor) as the $(2,1)$-block of $\tilde{A}$, in the block matrix representation of (the unitary conjugate of A) $\tilde{A}$ as above. $\triangle$

3.1. Rayleigh-Ritz majorization error bounds of the mixed type. We adopt Setting 3.1; moreover, in this subsection we further assume that $\mathcal{X}$ and $\mathcal{Y}$ are such that $\Theta_1(\mathcal{X}\,,\,\mathcal{Y}) < \frac{\pi}{2}$ that is, that $X^*Y \in \mathcal{G}l(k)$ is invertible.

Our first result concerns a submajorization error bound for the distance of eigenvalue lists of self-adjoint matrices:

THEOREM 3.2. *Let $C,\, D \in \mathcal{H}(k)$ and let $T \in \mathcal{G}l(k)$. Then,*

$$|\lambda(C) - \lambda(D)| \prec_w s(T^{-1})\ s(CT - TD). \tag{3.1}$$

Proof. See the Appendix (Section 4). ∎

The following result is [20, Conjecture 2.1] (see also Corollary 3.4 below).

THEOREM 3.3. *Under Setting 3.1, if $\Theta_1(\mathcal{X}\,,\,\mathcal{Y}) < \frac{\pi}{2}$ then*

$$|\lambda(\rho(X)) - \lambda(\rho(Y))| \prec_w \frac{s(P_{\mathcal{Y}}\ R_X) + s(P_{\mathcal{X}}\ R_Y)}{\cos(\Theta(\mathcal{X}\,,\,\mathcal{Y}))} \quad \textit{and} \tag{3.2}$$

$$|\lambda(\rho(X)) - \lambda(\rho(Y))| \prec_w [s(P_{\mathcal{X}+\mathcal{Y}}\ R_X) + s(P_{\mathcal{X}+\mathcal{Y}}\ R_Y)]\ \tan(\Theta(\mathcal{X}\,,\,\mathcal{Y}))\,. \tag{3.3}$$

Proof. Set $T = X^*Y$ and notice that, since $\Theta_1(\mathcal{X},\mathcal{Y}) < \frac{\pi}{2}$, $T \in \mathcal{M}_k(\mathbb{C})$ is invertible. Using Theorem 3.2 we get that

$$|\lambda(\rho(X)) - \lambda(\rho(Y))| \prec_w s(T^{-1})\ s(\rho(X)T - T\rho(Y))\,, \tag{3.4}$$

where $\rho(X) = X^*AX$, $\rho(Y) = Y^*AY \in \mathcal{H}(k)$. By construction we have that

$$s(T^{-1}) = \frac{1}{\cos(\Theta(\mathcal{X},\mathcal{Y}))} \in (\mathbb{R}^k_{>0})^{\downarrow}\,. \tag{3.5}$$

Arguing as in [20, Thm 4.1] we notice that

$$\begin{aligned}
\rho(X)T - T\rho(Y) &= X^*A\,XX^*Y - X^*YY^*A\,Y = X^*A\,P_{\mathcal{X}}Y - X^*P_{\mathcal{Y}}A\,Y \\
&= X^*A\,(I - P_{\mathcal{X}^\perp})Y - X^*(I - P_{\mathcal{Y}^\perp})A\,Y \\
&= X^*A\,Y - X^*A\,P_{\mathcal{X}^\perp}Y - X^*A\,Y + X^*P_{\mathcal{Y}^\perp}A\,Y = -X^*A\,P_{\mathcal{X}^\perp}Y + X^*P_{\mathcal{Y}^\perp}A\,Y\,.
\end{aligned}$$

Using that $s(C) = s(C^*)$ for $C \in \mathcal{M}_k(\mathbb{C})$, we see that

$$s(X^*A\,P_{\mathcal{X}^\perp}Y) = s(Y^*P_{\mathcal{X}^\perp}A\,X) = s(P_{\mathcal{Y}}P_{\mathcal{X}^\perp}A\,X) = s(P_{\mathcal{Y}}R_X) \in (\mathbb{R}^k_{\geq 0})^{\downarrow}\,.$$

Analogously $s(X^*P_{\mathcal{Y}^\perp}A\,Y) = s(P_{\mathcal{X}}R_Y)$. The previous facts together with the sub-additivity property of taking singular values (item 1 in Theorem 4.1) imply that

$$s(\rho(X)T - T\rho(Y)) = s(-X^*A\,P_{\mathcal{X}^\perp}Y + X^*P_{\mathcal{Y}^\perp}A\,Y) \prec_w s(P_{\mathcal{X}}R_Y) + s(P_{\mathcal{Y}}R_X)\,. \tag{3.6}$$

Now, if we apply (3.5) and (3.6) to (3.4), together with item 4 in Lemma 4.3, we get (3.2).

In order to show (3.3) we point out that by [20, Lemma 4.1] we get that

$$s(P_{\mathcal{X}} R_Y) \prec_w s(P_{\mathcal{X}+\mathcal{Y}}\, R_Y)\sin(\Theta(\mathcal{X},\mathcal{Y}))\,. \tag{3.7}$$

Since the entries of these vectors are ordered downwards, by Lemma 4.3 we deduce that

$$s(P_{\mathcal{X}} R_Y) + s(P_{\mathcal{Y}} R_X) \prec_w \big(s(P_{\mathcal{X}+\mathcal{Y}}\, R_Y) + s(P_{\mathcal{X}+\mathcal{Y}}\, R_X)\big)\,\sin(\Theta(\mathcal{X},\mathcal{Y}))\ . \tag{3.8}$$

Hence, using (3.2) and (3.8) together with Lemma 4.3 we see that (3.3) holds. ∎

The fact that (3.2) implies (3.3) was already observed in [20]; we have included the proof of this fact for the benefit of the reader.

Corollary 3.4. *Consider Setting 3.1 and assume that* $\Theta_1(\mathcal{X}\,,\,\mathcal{Y}) < \frac{\pi}{2}$. *If we further assume that* $\mathcal{X}$ *is* A*-invariant then*

$$|\lambda(\rho(X)) - \lambda(\rho(Y))| \prec_w \frac{s(P_{\mathcal{X}}\ R_Y)}{\cos(\Theta(\mathcal{X}\,,\,\mathcal{Y}))} \quad \textit{and} \tag{3.9}$$

$$|\lambda(\rho(X)) - \lambda(\rho(Y))| \prec_w s(P_{\mathcal{X}+\mathcal{Y}}\ R_Y)\ \tan(\Theta(\mathcal{X}\,,\,\mathcal{Y}))\,. \tag{3.10}$$

Proof. In case $\mathcal{X}$ is A-invariant notice that $R_X = 0$. The result now follows from Theorem 3.3. ∎

It is natural to wonder whether we can improve the bounds in the previous results. As shown in the following example, the submajorization bounds in Theorem 3.3 and Corollary 3.4 are *sharp*.

Example 3.5. Let $\lambda = (a,b,c,d) \in \mathbb{R}^4$, where $a < b < c < d$, and consider $A \in \mathcal{H}(4)$ given by $A = D_\lambda$, i.e. A is the diagonal matrix with main diagonal λ.

Let $\mathcal{X}$ be the A-invariant subspace $\mathcal{X} = \operatorname{span}\{e_1,\, e_2\}$ spanned by the first two elements of the canonical basis of $\mathbb{C}^4$. For $\theta \in (0,\pi/2)$ let $f_\theta = \cos\theta\, e_2 + \sin\theta\, e_3$ and set $\mathcal{Y}_\theta = \operatorname{span}\{e_1,\, f_\theta\}$. Then, the principal angles are given by $\Theta(\mathcal{X},\mathcal{Y}_\theta) = (\theta, 0)$. Let

$$X = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{pmatrix} \quad , \quad X_\perp = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad Y_\theta = \begin{pmatrix} 1 & 0 \\ 0 & \cos\theta \\ 0 & \sin\theta \\ 0 & 0 \end{pmatrix}\,.$$

It is straightforward to check that $\lambda(X^*AX) = (b,a)$ and that $\lambda(Y_\theta^* AY_\theta) = (b\,\cos^2\theta + c\,\sin^2(\theta), a)$. Again, simple computations show that

$$R_{Y_\theta} = \begin{pmatrix} 0 & 0 \\ 0 & (b-c)\,\cos\theta\,\sin^2\theta \\ 0 & (c-b)\,\cos^2\theta\,\sin\theta \\ 0 & 0 \end{pmatrix} \quad , \quad P_{\mathcal{X}}\, R_{Y_\theta} = \begin{pmatrix} 0 & 0 \\ 0 & (b-c)\,\cos\theta\,\sin^2\theta \\ 0 & 0 \\ 0 & 0 \end{pmatrix}\,.$$

Hence, $s(P_{\mathcal{X}}\, R_{Y_\theta}) = ((c-b)\cos\theta\,\sin^2\theta, 0)$. Now,

$$|\lambda(X^*A\,X) - \lambda((Y_\theta)^*A\,Y_\theta)| = ((c-b)\,\sin^2\theta, 0)\ , \tag{3.11}$$

$$\frac{s(P_{\mathcal{X}}\ R_{Y_\theta})}{\cos(\Theta(\mathcal{X}\,,\,\mathcal{Y}_\theta))} = ((c-b)\,\sin^2\theta, 0)\,. \tag{3.12}$$

That is, (3.9) in Corollary 3.4 becomes an equality in this case. This also shows that (3.2) is sharp, since (3.9) above is a particular case (when $\mathcal{X}$ is A-invariant). Notice that $\mathcal{X}+\mathcal{Y}_\theta = \operatorname{span}\{e_1, e_2, e_3\}$. Therefore, since $P_{\mathcal{X}+\mathcal{Y}_\theta} R_{Y_\theta} = R_{Y_\theta}$ and $s(R_{Y_\theta}) = ((c-b)\cos\theta\,\sin\theta, 0)$,

$$s(P_{\mathcal{X}+\mathcal{Y}_\theta}\ R_{Y_\theta})\ \tan(\Theta(\mathcal{X}\,,\,\mathcal{Y}_\theta)) = ((c-b)\,\sin^2\theta, 0)\,. \tag{3.13}$$

By (3.11) and (3.13) we now see that (3.10) in Corollary 3.4 becomes an equality in this case. This also shows that (3.3) is sharp, since (3.10) above is a particular case (when $\mathcal{X}$ is A-invariant). △

Remark 3.6 (Relations between our work and previous results). In the vector case, that is when $\mathcal{X}$ and $\mathcal{Y}$ are one dimensional spaces, Theorem 3.3 implies the upper bounds in [19, Theorem 3.7], which is one of the main results of that work (see also Corollary 3.24 and Remark 3.25).

In [20] Knyazev and Zhu obtained several bounds for the absolute change of the Ritz values. Using Setting 3.1, the authors show (see [20, Theorem 4.2 and Corollary 4.4]) that

$$|\lambda(\rho(X)) - \lambda(\rho(Y))|^2 \prec_w \frac{\{s(P_{\mathcal{Y}}\ R_X) + s(P_{\mathcal{X}}\ R_Y)\}^2}{\cos^2(\Theta(\mathcal{X}\,,\,\mathcal{Y}))} \quad \text{and} \tag{3.14}$$

$$|\lambda(\rho(X)) - \lambda(\rho(Y))|^2 \prec_w \{s(P_{\mathcal{X}+\mathcal{Y}}\ R_X) + s(P_{\mathcal{X}+\mathcal{Y}}\ R_Y)\}^2\ \tan^2(\Theta(\mathcal{X}\,,\,\mathcal{Y}))\,. \tag{3.15}$$

Using the fact that $f:\mathbb{R}_{\geq 0}\to\mathbb{R}_{\geq 0}$ given by $f(x)=x^2$ is an increasing and convex function, Remark 2.2 shows that (3.14) and (3.15) follow from (3.2) and (3.3) from Theorem 3.3. Similarly, using that $\cos\Theta_1(\mathcal{X},\mathcal{Y}) = \cos\Theta_{\max}(\mathcal{X},\mathcal{Y}) \leq \cos\Theta_i(\mathcal{X},\mathcal{Y})$, for $i\in\mathbb{I}_k$, we get that Theorem 3.3 implies [20, Theorems 4.1, 4.3].

In [20] the authors show that their results can be applied in several situations such as: first order and quadratic a posteriori majorization bounds; bounds for eigenvalues after matrix additive perturbations. The previous remarks show that our bounds can also be applied in these settings. Moreover, Theorem 3.3 allows to formalize the arguments related with bounds for eigenvalues after matrix additive perturbations, and in particular with bounds for eigenvalues after discarding off-diagonal blocks from [20, Section 5] (see the detailed discussion there). △

The bounds in Theorem 3.3 can be used to perform a detailed analysis and obtain better convergence rates for iterative algorithms related with the Rayleigh-Ritz method (see [16, 17, 21]). We will consider such applications elsewhere.

3.2. Applications: a priori majorization error bounds for Ritz values. In this section we establish a link between the majorization error bounds of the mixed type obtained in the previous section and some a priori majorization error bounds considered in [8, 10].

Definition 3.7. *Let $A\in\mathcal{H}(d)$ and let $\mathcal{Z}\subset\mathbb{C}^d$ be a subspace with $\dim\mathcal{Z}=p$. We define the (spectral) spread of A relative to $\mathcal{Z}$, denoted $\operatorname{Spr}(A\,,\,\mathcal{Z})$, given by*

$$\operatorname{Spr}(A,\mathcal{Z}) = \lambda(A_{\mathcal{Z}}) - \lambda^{\uparrow}(A_{\mathcal{Z}}) = (\lambda_i(A_{\mathcal{Z}}) - \lambda_{p-i+1}(A_{\mathcal{Z}}))_{i\in\mathbb{I}_p} \in (\mathbb{R}^p)^{\downarrow}\,,$$

where $A_{\mathcal{Z}} = P_{\mathcal{Z}}\,A|_{\mathcal{Z}} \in \mathcal{L}(\mathcal{Z})$ is a self-adjoint operator (defined in the obvious way). In case $\mathcal{Z}=\mathbb{C}^d$, we write $\operatorname{Spr}(A,\mathbb{C}^d)=\operatorname{Spr}(A)$. △

REMARK 3.8. Let $A \in \mathcal{H}(d)$ and let $\mathcal{X}, \mathcal{Y} \subset \mathbb{C}^d$ with $\dim(\mathcal{X}) = \dim(\mathcal{Y}) = k$. Denote by $p = \dim \mathcal{X} + \mathcal{Y}$. In what follows we consider the vector

$$\operatorname{Spr}(A, \mathcal{X} + \mathcal{Y}) \sin(\Theta(\mathcal{X}, \mathcal{Y})) = (\,(\lambda_i(A_{\mathcal{X}+\mathcal{Y}}) - \lambda_{p-i+1}(A_{\mathcal{X}+\mathcal{Y}})) \sin(\Theta_i(\mathcal{X}, \mathcal{Y})\,)\,)_{i \in \mathbb{I}_k} .$$

We point out that this vector has non-negative entries, which are arranged in non-increasing order (in particular, $\sin(\Theta_i(\mathcal{X}, \mathcal{Y})) = 0$ whenever $\lambda_i(A_{\mathcal{X}+\mathcal{Y}}) - \lambda_{p-i+1}(A_{\mathcal{X}+\mathcal{Y}}) < 0$, for $i \in \mathbb{I}_k$); hence, $\operatorname{Spr}(A, \mathcal{X} + \mathcal{Y}) \sin(\Theta(\mathcal{X}, \mathcal{Y})) \in (\mathbb{R}^k_{\geq 0})^{\downarrow}$ (see [20]). This fact becomes relevant for the conjectures posed in (3.16) and (3.17) below. △

REMARK 3.9 (A priori error bounds for changes of Ritz values: conjectures and previous work). Let $A \in \mathcal{H}(d)$ and let $\mathcal{X}, \mathcal{Y} \subset \mathbb{C}^d$ with $\dim(\mathcal{X}) = \dim(\mathcal{Y}) = k$. In [8] the authors conjectured that, in general, the following submajorization bound for the Ritz values holds:

$$|\lambda(\rho(X)) - \lambda(\rho(Y))| \prec_w \operatorname{Spr}(A, \mathcal{X} + \mathcal{Y}) \sin(\Theta(\mathcal{X}, \mathcal{Y})) . \tag{3.16}$$

Moreover, in case $\mathcal{X}$ is A-invariant, the authors conjectured that

$$|\lambda(\rho(X)) - \lambda(\rho(Y))| \prec_w \operatorname{Spr}(A, \mathcal{X} + \mathcal{Y}) \sin(\Theta(\mathcal{X}, \mathcal{Y}))^2 . \tag{3.17}$$

These conjectures are natural extensions of results from [10] (that were obtained for $k = 1$). Although [8, Conjecture 2.1.] claims the validity of (3.16) and (3.17) for arbitrary subspaces $\mathcal{X}$ and $\mathcal{Y}$ such that $\dim \mathcal{X} = \dim \mathcal{Y}$, such bounds would become relevant in the particular case when the subspace $\mathcal{Y}$ is a (small) perturbation of the subspace $\mathcal{X}$. In this case, the validity of (3.16) and (3.17) would reveal the different orders of approximation of $\rho(X)$ by $\rho(Y)$ in terms of PABS as well as in terms of the spectral spread of A (i.e. when considering A as well as $\mathcal{X}$ and $\mathcal{Y}$ as variables). Notice that these results would have immediate applications in the study of numerical stability and convergence of iterative methods related with the Rayleigh-Ritz type algorithms.

In [8, Theorem 2.1.] the authors showed that, in general,

$$|\lambda(\rho(X)) - \lambda(\rho(Y))| \prec_w (\lambda_{\max}(A_{\mathcal{X}+\mathcal{Y}}) - \lambda_{\min}(A_{\mathcal{X}+\mathcal{Y}})) \sin(\Theta(\mathcal{X}, \mathcal{Y})) , \tag{3.18}$$

while, in case $\mathcal{X}$ is A-invariant,

$$|\lambda(\rho(X)) - \lambda(\rho(Y))| \prec_w (\lambda_{\max}(A_{\mathcal{X}+\mathcal{Y}}) - \lambda_{\min}(A_{\mathcal{X}+\mathcal{Y}})) \sin(\Theta(\mathcal{X}, \mathcal{Y}))^2 , \tag{3.19}$$

where $A_{\mathcal{X}+\mathcal{Y}} = P_{\mathcal{X}+\mathcal{Y}} \ A|_{\mathcal{X}+\mathcal{Y}} \in \mathcal{L}(\mathcal{X} + \mathcal{Y})$; moreover, in [8, Theorem 2.2.] they showed that in the particular case in which $\mathcal{X}$ is the A-invariant subspace corresponding to the k largest eigenvalues of A, then

$$0 \leq \lambda(\rho(X)) - \lambda(\rho(Y)) \prec_w (\lambda_i(A_{\mathcal{X}+\mathcal{Y}}) - \lambda_{\min}(A_{\mathcal{X}+\mathcal{Y}}))_{i \in \mathbb{I}_k} \sin(\Theta(\mathcal{X}, \mathcal{Y}))^2 . \tag{3.20}$$

Notice that, (3.20) is a stronger bound than that in (3.19); yet, it is weaker than the bound conjectured in (3.17), since $\operatorname{Spr}_i(A, \mathcal{X} + \mathcal{Y}) \leq \lambda_i(A_{\mathcal{X}+\mathcal{Y}}) - \lambda_{\min}(A_{\mathcal{X}+\mathcal{Y}})$, for $i \in \mathbb{I}_k$. △

In what follows we apply Theorem 3.3 and obtain some results related with the conjectures from [8] described in (3.16) and (3.17). In order to obtain these results, we take a closer look at the quantity $s(P_{\mathcal{X}} \, R_Y)$ for arbitrary $\mathcal{X}$ and $\mathcal{Y}$, as well as in the case where $\mathcal{X}$ is A-invariant.

Proposition 3.10. *Let $A \in \mathcal{H}(d)$ and let $\mathcal{X}, \mathcal{Y} \subset \mathbb{C}^d$ with $\dim(\mathcal{X}) = \dim(\mathcal{Y}) = k$. Then*

$$s(P_{\mathcal{X}}\, R_Y) \prec_w \operatorname{Spr}(A, \mathcal{X} + \mathcal{Y})\, \sin(\Theta(\mathcal{X}, \mathcal{Y}))\,. \tag{3.21}$$

Proof. See the Appendix (Section 4). ∎

Theorem 3.11. *Let $A \in \mathcal{H}(d)$, $\mathcal{X}, \mathcal{Y} \subset \mathbb{C}^d$ subspaces, $\dim(\mathcal{X}) = \dim(\mathcal{Y}) = k$. If $\Theta_1(\mathcal{X}, \mathcal{Y}) < \frac{\pi}{2}$, then*

$$|\lambda(\rho(X)) - \lambda(\rho(Y))| \prec_w \frac{2\operatorname{Spr}(A, \mathcal{X} + \mathcal{Y})\, \sin(\Theta(\mathcal{X}, \mathcal{Y}))}{\cos(\Theta(\mathcal{X}\,,\, \mathcal{Y}))}\,. \tag{3.22}$$

Proof. Theorem 3.3 establishes that

$$|\lambda(\rho(X)) - \lambda(\rho(Y))| \prec_w \frac{s(P_{\mathcal{X}} R_Y) + s(P_{\mathcal{Y}} R_X)}{\cos(\Theta(\mathcal{X}, \mathcal{Y}))}\,.$$

Proposition (3.10) together with Lemma 4.3 imply that

$$\frac{s(P_{\mathcal{X}} R_Y) + s(P_{\mathcal{Y}} R_X)}{\cos(\Theta(\mathcal{X}, \mathcal{Y}))} \prec_w \frac{2\operatorname{Spr}(A, \mathcal{X} + \mathcal{Y})\, \sin(\Theta(\mathcal{X}, \mathcal{Y}))}{\cos(\Theta(\mathcal{X}\,,\, \mathcal{Y}))}\,.$$

The result follows from combining these last two inequalities. ∎

The next result illustrates the quadratic dependance of $s(P_{\mathcal{X}} R_Y)$ from $\sin(\Theta(\mathcal{X}, \mathcal{Y}))$ in case $\mathcal{X}$ is A-invariant.

Proposition 3.12. *Let $A \in \mathcal{H}(d)$,$\mathcal{X}, \mathcal{Y} \subset \mathbb{C}^d$ subspaces with $\dim(\mathcal{X}) = \dim(\mathcal{Y}) = k$. Assume that $\mathcal{X}$ is A-invariant. Then,*

$$s(P_{\mathcal{X}} R_Y) \prec_w 2\ (\lambda_i(A_{\mathcal{X}+\mathcal{Y}}) - \lambda_{\min}(A_{\mathcal{X}+\mathcal{Y}}))_{i \in \mathbb{I}_k}\ \sin^2(\Theta(\mathcal{X}, \mathcal{Y}))\,. \tag{3.23}$$

Proof. See the Appendix (Section 4). ∎

Theorem 3.13. *Let $A \in \mathcal{H}(d)$, $\mathcal{X}, \mathcal{Y} \subset \mathbb{C}^d$ subspaces, $\dim(\mathcal{X}) = \dim(\mathcal{Y}) = k$, and assume that $\mathcal{X}$ is A-invariant. If $\Theta_1(\mathcal{X}, \mathcal{Y}) < \frac{\pi}{2}$, then*

$$|\lambda(\rho(X)) - \lambda(\rho(Y))| \prec_w \frac{2\ (\lambda_i(A_{\mathcal{X}+\mathcal{Y}}) - \lambda_{\min}(A_{\mathcal{X}+\mathcal{Y}}))_{i \in \mathbb{I}_k}\ \sin^2(\Theta(\mathcal{X}, \mathcal{Y}))}{\cos(\Theta(\mathcal{X}\,,\, \mathcal{Y}))}\,. \tag{3.24}$$

Proof. The result follows from Corollary 3.4 and Proposition 3.12 with an argument similar to that in the proof of Theorem 3.11 above. ∎

Corollary 3.14. *Let $A \in \mathcal{H}(d)$, $\mathcal{X}, \mathcal{Y} \subset \mathbb{C}^d$ subspaces, $\dim(\mathcal{X}) = \dim(\mathcal{Y}) = k$. If $\Theta_1(\mathcal{X}, \mathcal{Y}) < \frac{\pi}{2}$, then*

$$|\lambda(\rho(X)) - \lambda(\rho(Y))| \prec_w \frac{2}{\cos(\Theta_1(\mathcal{X}, \mathcal{Y}))}\ \operatorname{Spr}(A, \mathcal{X} + \mathcal{Y})\, \sin(\Theta(\mathcal{X}, \mathcal{Y}))\,.$$

If we assume further that $\mathcal{X}$ is A-invariant, then

$$|\lambda(\rho(X)) - \lambda(\rho(Y))| \prec_w \frac{2}{\cos(\Theta_1(\mathcal{X}, \mathcal{Y}))}\ (\lambda_i(A_{\mathcal{X}+\mathcal{Y}}) - \lambda_{\min}(A_{\mathcal{X}+\mathcal{Y}}))_{i \in \mathbb{I}_k}\ \sin^2(\Theta(\mathcal{X}, \mathcal{Y}))\,.$$

We end this section with some remarks concerning the relations among Theorems 3.11 and 3.13, Corollary 3.14 and the conjectured bounds in (3.16) and (3.17). As already mentioned in Remark 3.9, the bounds in (3.16) and (3.17) would be particularly relevant in case $\mathcal{Y}$ is a (small) perturbation of $\mathcal{X}$ or, in other terms, in case that $\mathcal{X}$ and $\mathcal{Y}$ are close subspaces (e.g. $\Theta_1(\mathcal{X},\mathcal{Y})$ is small). In order to simplify the discussion, let us assume that $\Theta_1(\mathcal{X},\mathcal{Y}) \leq \pi/4$. We point out that this assumption holds in a number of significant situations (see for example [20, Section 5.2.]). In this case, if $A \in \mathcal{H}(d)$ then Corollary 3.14 implies that

$$|\lambda(\rho(X)) - \lambda(\rho(Y))| \prec_w (2\sqrt{2})\ \mathrm{Spr}(A, \mathcal{X}+\mathcal{Y})\ \sin(\Theta(\mathcal{X},\mathcal{Y}))\,. \tag{3.25}$$

Hence, under the present assumptions ($\Theta_1(\mathcal{X},\mathcal{Y}) \leq \pi/4$), the upper bound in (3.25) has the conjectured order of approximation (when considering A as well as the subspaces $\mathcal{X}$ and $\mathcal{Y}$ as variables), up to the constant factor $2\sqrt{2}$.

If we further assume that $\mathcal{X}$ is A-invariant then by the same result we get that

$$|\lambda(\rho(X)) - \lambda(\rho(Y))| \prec_w (2\sqrt{2})\ (\lambda_i(A_{\mathcal{X}+\mathcal{Y}}) - \lambda_{\min}(A_{\mathcal{X}+\mathcal{Y}}))_{i\in\mathbb{I}_k}\ \sin^2(\Theta(\mathcal{X},\mathcal{Y}))\,. \tag{3.26}$$

Again, the upper bound in (3.26) has the conjectured order of approximation (when considering A as well as the subspaces $\mathcal{X}$ and $\mathcal{Y}$ as variables), up to the constant factor $2\sqrt{2}$. Moreover, notice that this bound holds for an arbitrary A-invariant subspace $\mathcal{X}$ (as opposed the bound in (3.20) from [8] that is shown to hold for special choices of A-invariant subspaces $\mathcal{X}$).

3.3. The $\tan\Theta$ theorem revisited: improved quadratic a posteriori error bounds. In this section we revisit Nakatsukasa's extension of Davis-Kahan's $\tan(\theta)$ theorem. Our motivation is the study of an improved version of this result conjectured in [20] (see Corollary 3.22 below). We first recall the separation hypothesis for Nakatsukasa's result. As before, in this section we adopt Setting 3.1.

Definition 3.15. Let $A \in \mathcal{H}(d)$ and let $\mathcal{X},\, \mathcal{Y} \subset \mathbb{C}^d$ be subspaces with $\dim \mathcal{X} = \dim \mathcal{Y} = k$, such that $\mathcal{X}$ is A-invariant. Let $[X, X_\perp]$, $[Y, Y_\perp] \in \mathcal{U}(d)$ be unitary matrices such that the columns of (the $d\times k$ matrices) X and Y form ONB's of $\mathcal{X}$ and $\mathcal{Y}$ respectively. Given $\delta > 0$ we say that $(A\,,\,\mathcal{X}\,,\,\mathcal{Y}\,,\,\delta)$ satisfies the Davis-Kahan-Nakatsukasa (DKN) separation property if there exist $a \leq b$ such that

1. $\lambda_i(X_\perp^* A X_\perp) = \lambda_i(P_{\mathcal{X}^\perp}\, A\, P_{\mathcal{X}^\perp}) \in [a, b]$, for $i \in \mathbb{I}_{d-k}$;
2. $\lambda_i(Y^* A Y) = \lambda_i(P_{\mathcal{Y}}\, A\, P_{\mathcal{Y}}) \in (\infty, a-\delta] \cup [b+\delta, \infty)$, for $i \in \mathbb{I}_k$. △

Next we state Nakatsukasa's $\tan\Theta$ theorem under relaxed conditions.

Theorem 3.16 ([14]). *Let $A \in \mathcal{H}(d)$, $\mathcal{X}, \mathcal{Y} \subset \mathbb{C}^d$ and let $\delta > 0$ be such that $(A\,,\,\mathcal{X}\,,\,\mathcal{Y}\,,\,\delta)$ satisfies the DKN separation property. Then, $\Theta_1(\mathcal{X},\mathcal{Y}) < \pi/2$ and*

$$\delta\, \|\tan(\Theta(\mathcal{X}\,,\,\mathcal{Y}))\| \leq \|R_Y\|\,,$$

for every unitarily invariant norm $\|\cdot\|$. Equivalently, $\delta\, \tan(\Theta(\mathcal{X}\,,\,\mathcal{Y})) \prec_w s(R_Y)$.

Remark 3.17. Theorem 3.16 requires the knowledge of the full matrix A in order to bound the (norm of the) vector $\tan(\Theta(\mathcal{X},\mathcal{Y}))$ from above. Instead, it would be interesting to bound the vector $\tan(\Theta(\mathcal{X},\mathcal{Y}))$ from above (only) in terms of the self-adjoint operator $A_{\mathcal{X}+\mathcal{Y}} = P_{\mathcal{X}+\mathcal{Y}} A|_{\mathcal{X}+\mathcal{Y}} \in \mathcal{L}(\mathcal{X}+\mathcal{Y})$ (defined in the obvious way). In the next result we show that the $\tan\Theta$ theorem mentioned above allow to obtain such a result. Moreover, we will also see that it is possible to describe separation hypothesis for $(A_{\mathcal{X}+\mathcal{Y}},\, \mathcal{X},\, \mathcal{Y})$, that are more general than the DKN separation hypothesis for $(A,\, \mathcal{X},\, \mathcal{Y})$, for which the $\tan\Theta$ theorem holds; arguing in terms of interlacing inequalities, we can show that these separation hypotheses on $A_{\mathcal{X}+\mathcal{Y}}$ provide better separation constants than the DKN separation hypotheses on the matrix A. $\triangle$

We formalize the content of the previous remark - with a small variation on the notation - in the following result. First, we recall some facts related with the relative position of two subspaces.

REMARK 3.18. Let $\mathcal{X},\, \mathcal{Y} \subset \mathbb{C}^d$ be two subspaces with $\dim \mathcal{X} = \dim \mathcal{Y} = k$. Consider the mutually orthogonal subspaces

$$\mathcal{H}_{00} = \mathcal{X}^\perp \cap \mathcal{Y}^\perp \ ,\ \mathcal{H}_{10} = \mathcal{X} \cap \mathcal{Y}^\perp \ ,\ \mathcal{H}_{01} = \mathcal{X}^\perp \cap \mathcal{Y} \ ,\ \mathcal{H}_{11} = \mathcal{X} \cap \mathcal{Y} \ ,$$

and $\mathcal{H}_g = \mathbb{C}^d \ominus (\mathcal{H}_{00} \oplus \mathcal{H}_{10} \oplus \mathcal{H}_{01} \oplus \mathcal{H}_{11})$ which is called the *generic part* of the pair $(\mathcal{X}, \mathcal{Y})$. Each of these five (possible zero) subspaces reduces each projection $P_\mathcal{X}$ and $P_\mathcal{Y}$. Moreover, the subspaces $\mathcal{X}_g = \mathcal{X} \cap \mathcal{H}_g$ and $\mathcal{Y}_g = \mathcal{Y} \cap \mathcal{H}_g$ are in *generic position* so that $\mathcal{H}_g = \mathcal{X}_g + \mathcal{Y}_g$. For details of this well known construction and several fundamental results see [5]. $\triangle$

THEOREM 3.19. *Let $A \in \mathcal{H}(d)$, and let $\mathcal{X},\, \mathcal{Y} \subset \mathbb{C}^d$ be such that $\dim \mathcal{X} = \dim \mathcal{Y} = k$. Let $A_{\mathcal{X}+\mathcal{Y}} = S^*AS \in \mathcal{H}(p)$, where $S \in \mathcal{M}_{d,p}(\mathbb{C})$ is such that its columns form an ONB for $\mathcal{X} + \mathcal{Y}$. Then,*

1. *If $\delta > 0$ is such that $(A\,,\, \mathcal{X}\,,\, \mathcal{Y}\,,\, \delta)$ satisfies the DKN separation property then there exists $\delta' \geq \delta$ such that $(A_{\mathcal{X}+\mathcal{Y}}\,,\, S^*\mathcal{X}\,,\, S^*\mathcal{Y}\,,\, \delta')$ satisfies the DKN separation property.*
2. *If $\delta' > 0$ is such that $(A_{\mathcal{X}+\mathcal{Y}}\,,\, S^*\mathcal{X}\,,\, S^*\mathcal{Y}\,,\, \delta')$ satisfies the DKN separation property, then*

$$\delta' \,\|\tan(\Theta(\mathcal{X}\,,\, \mathcal{Y}))\| \leq \|A_{\mathcal{X}+\mathcal{Y}}\, Y_S - Y_S\, (Y_S^* A_{\mathcal{X}+\mathcal{Y}}\, Y_S)\| = \|P_{\mathcal{X}+\mathcal{Y}}\, R_Y\| \tag{3.27}$$

for every unitarily invariant norm $\|\cdot\|$, where $Y_S = S^*Y \in \mathcal{M}_{p,k}(\mathbb{C})$.

Proof. We first show item 1 Let $X,\, Y \in \mathcal{M}_{d,k}(\mathbb{C})$ be such that their columns form orthonormal bases of $\mathcal{X}$ and $\mathcal{Y}$, respectively. By hypothesis, there exist $a \leq b$ such that: for $i \in \mathbb{I}_{d-k}$ and $j \in \mathbb{I}_k$ we have that

$$\lambda_i(X_\perp^* A X_\perp) \in [a, b] \quad \text{ and } \quad \lambda_j(Y^*AY) \in (\infty, a-\delta] \cup [b+\delta, \infty)\,,$$

where $X_\perp \in \mathcal{M}_{d,d-k}(\mathbb{C})$ is such that its columns for an ONB for $\mathcal{X}^\perp$. Let $\mathcal{Z} = \mathcal{X} + \mathcal{Y}$ and notice that $S \in \mathcal{M}_{d,p}(\mathbb{C})$ is an isometry from $\mathbb{C}^p$ onto $\mathcal{Z}$. Moreover, the matrix $S^*A\,S \in \mathcal{H}(p)$. Similarly, $X_S = S^*X$, $Y_S = S^*Y \in \mathcal{M}_{p,k}$ are isometries from $\mathbb{C}^k$ onto $S^*\mathcal{X}$, $S^*\mathcal{Y} \subseteq \mathbb{C}^p$, respectively. Consider the mutually orthogonal subspaces

$$\mathcal{H}_{11} = \mathcal{X} \cap \mathcal{Y} \quad , \quad \mathcal{X}_g = \mathcal{H}_g \cap \mathcal{X} \quad \text{ and } \quad \mathcal{X}_{g^\perp} = \mathcal{H}_g \ominus \mathcal{X}_g\,,$$

where $\mathcal{H}_g$ is the subspace of $\mathbb{C}^d$ corresponding to the generic part of the pair $(\mathcal{X}\,,\, \mathcal{Y})$ (see Remark 3.18). By Theorem 3.16 we have that $\Theta_1(\mathcal{X}, \mathcal{Y}) < \pi/2$ so then, $\mathcal{X}^\perp \cap \mathcal{Y} = \{0\} = \mathcal{X} \cap \mathcal{Y}^\perp$. Thus,

$$\mathcal{X} = \mathcal{H}_{11} \oplus \mathcal{X}_g \quad , \quad \mathcal{Z} = \mathcal{H}_{11} \oplus \mathcal{X}_g \oplus \mathcal{X}_{g^\perp} \quad \text{ and } \quad \mathcal{X}_{g^\perp} = \mathcal{Z} \ominus \mathcal{X}\,.$$

Let $X' \in M_{d,(p-k)}(\mathbb{C})$ be such that its columns form an orthonormal basis of $\mathcal{X}_{g^\perp} \subset \mathcal{X}^\perp$. Then, $X'_S = S^*\, X' \in \mathcal{M}_{p,(p-k)}(\mathbb{C})$ is an isometry from $\mathbb{C}^{p-k}$ onto $S^*\mathcal{X}_{g^\perp} =$

$(S^*\mathcal{X})^\perp \subseteq \mathbb{C}^p$. To check the DKN separation property for $(A_{\mathcal{X}+\mathcal{Y}}\,,\, S^*\mathcal{X}\,,\, S^*\mathcal{Y})$ we consider the eigenvalues of

$$(X'_S)^*(S^*A\,S)\,X'_S = (X')^*\,S\,S^*\,A\,S\,S^*\,X' = (X')^*\,A\,X' \in \mathcal{H}(p-k)\,,$$

since $SS^* = P_{\mathcal{Z}} \in \mathcal{M}_d(\mathbb{C})$, $P_{\mathcal{Z}}\,X' = X'$ and $(X')^*\,P_{\mathcal{Z}} = (X')^*$. Hence, we now see that

$$\lambda_i((X'_S)^*(S^*A\,S)\,X'_S) = \lambda_i(P_{\mathcal{X}_{g^\perp}}\,A\,P_{\mathcal{X}_{g^\perp}}) \quad \text{for} \quad i \in \mathbb{I}_{p-k}\,.$$

Since $\mathcal{X}_{g^\perp} \subset \mathcal{X}^\perp$ we have that $P_{\mathcal{X}_{g^\perp}}\,A\,P_{\mathcal{X}_{g^\perp}}$ is a compression of $P_{\mathcal{X}^\perp}\,A\,P_{\mathcal{X}^\perp}$. Using the interlacing inequalities for compressions of self-adjoint matrices (see [2]), we get that if $\lambda_i((P_{\mathcal{X}^\perp}\,A\,P_{\mathcal{X}^\perp})) \in [a,b]$, for $i \in \mathbb{I}_{d-k}$, then

$$\lambda_i(P_{\mathcal{X}_{g^\perp}}\,A\,P_{\mathcal{X}_{g^\perp}}) \in [a,b] \quad \text{for} \quad i \in \mathbb{I}_{p-k}\,. \tag{3.28}$$

On the other hand, notice that

$$Y_S^*\,(S^*A\,S)\,Y_S = Y^*P_{\mathcal{Z}}A\,P_{\mathcal{Z}}\,Y = Y^*A\,Y$$

since, as before, $SS^* = P_{\mathcal{Z}}$, $P_{\mathcal{Z}}Y = Y$ and $Y^*P_{\mathcal{Z}} = Y^*$. Therefore, we get that

$$\lambda_i(Y_S^*\,(S^*A\,S)\,Y_S) = \lambda_i(Y^*A\,Y) \in (\infty, a-\delta] \cup [b+\delta, \infty) \quad \text{for} \quad i \in \mathbb{I}_k\,. \tag{3.29}$$

Item 1 now follows from (3.28) and (3.29) and the fact that $S^*\mathcal{X} \subseteq \mathbb{C}^p$ is, by construction, an $A_{\mathcal{X}+\mathcal{Y}}$-invariant subspace.

In order to show item 2, we fix a unitarily invariant norm $\|\cdot\|$. Using that $\mathcal{X}$, $\mathcal{Y} \subset \mathcal{Z}$ and the fact that S^* is an isometry from $\mathcal{Z}$ onto $\mathbb{C}^p$, we see that $\Theta(\mathcal{X},\mathcal{Y}) = \Theta(S^*\mathcal{X}, S^*\mathcal{Y})$. Then, an application of Nakatsukasa's $\tan\Theta$ theorem (Theorem 3.16) to the self-adjoint matrix $S^*AS \in \mathcal{H}(p)$ and subspaces $S^*\mathcal{X}$, $S^*\mathcal{Y} \subseteq \mathbb{C}^p$ shows that

$$\delta'\,\|\tan(\Theta(\mathcal{X}\,,\,\mathcal{Y}))\| \le \|A_{\mathcal{X}+\mathcal{Y}}\,Y_S - Y_S\,(Y_S^*A_{\mathcal{X}+\mathcal{Y}}\,Y_S)\,\|\,,$$

where $Y_S = S^*Y \in \mathcal{M}_{p,k}$ is an isometry from $\mathbb{C}^k$ onto $S^*\mathcal{Y}$. We notice that

$$\begin{aligned} A_{\mathcal{X}+\mathcal{Y}}\,Y_S - Y_S\,(Y_S^*A_{\mathcal{X}+\mathcal{Y}}\,Y_S) &= S^*A\,S\,S^*Y - S^*Y\,(Y^*S(S^*A\,S)S^*Y) \\ &= S^*\,(A\,Y - Y\,(Y^*A\,Y))\,, \end{aligned}$$

where we have used that $SS^* = P_{\mathcal{Z}}$, $P_{\mathcal{Z}}\,Y = Y$ and $Y^*\,P_{\mathcal{Z}} = Y^*$. Hence, it follows that

$$\|A_{\mathcal{X}+\mathcal{Y}}\,Y_S - Y_S\,(Y_S^*A_{\mathcal{X}+\mathcal{Y}}\,Y_S)\| = \|P_{\mathcal{Z}}\,(A\,Y - Y\,(Y^*A\,Y))\| = \|P_{\mathcal{X}+\mathcal{Y}}\,R_Y\|\,.$$

□

Remark 3.20. With the notation of Theorem 3.19 and using Remark 2.4, (3.27) is equivalent to the majorization relation

$$\delta'\,\tan(\Theta(\mathcal{X}\,,\,\mathcal{Y}) \prec_w s(A_{\mathcal{X}+\mathcal{Y}}\,Y_S - Y_S\,(Y_S^*A_{\mathcal{X}+\mathcal{Y}}\,Y_S)) = s(P_{\mathcal{X}+\mathcal{Y}}\;R_Y)$$

in terms of the separation constant δ' for $A_{\mathcal{X}+\mathcal{Y}} = S^*A\,S$, $S^*\mathcal{X}$ and $S^*\mathcal{Y}$. △

Consider the notation in Theorem 3.19. Let $\delta > 0$ be such that $(A\,,\,\mathcal{X}\,,\,\mathcal{Y}\,,\,\delta)$ satisfies the DKN separation property. Given a unitarily invariant norm $\|\cdot\|$, Theorem 3.16 allows to bound $\|\tan\Theta(\mathcal{X},\mathcal{Y})\|$ from above by

$$\|\tan\Theta(\mathcal{X},\mathcal{Y})\| \leq \frac{\|R_Y\|}{\delta}\,. \tag{3.30}$$

On the other hand, by item 2 in Theorem 3.19 there exists $\delta' \geq \delta > 0$ such that $(A_{\mathcal{X}+\mathcal{Y}}\,,\,S^*\mathcal{X}\,,\,S^*\mathcal{Y}\,,\,\delta')$ satisfies the DKN separation property, so that we get the upper bound

$$\|\tan\Theta(\mathcal{X},\mathcal{Y})\| \leq \frac{\|P_{\mathcal{X}+\mathcal{Y}}\,R_Y\|}{\delta\,'}\,. \tag{3.31}$$

Since $\|P_{\mathcal{X}+\mathcal{Y}}\,R_Y\| \leq \|R_Y\|$ and $\delta \leq \delta\,'$, we immediately see that the upper bound in (3.31) improves the classical bound in (3.30). In order to compare these two bounds in some more detail, let us consider the following

Example 3.21. Let $\tilde\lambda = (a,b,d,c) \in \mathbb{R}^4$, where $a < b < c < d$, and let $\tilde A \in \mathcal{H}(4)$ be given by $\tilde A = D_{\tilde\lambda}$. For the purposes of this example, we consider the real parameter $c \in (b,d)$ as variable (while a, b, d are fixed).

Let $\mathcal{X}$, $\mathcal{Y}_\theta \subset \mathbb{C}^4$ be as in Example 3.5 i.e. $\mathcal{X} = \operatorname{span}\{e_1,\,e_2\}$ and $\mathcal{Y}_\theta = \operatorname{span}\{e_1,\,f_\theta\}$. Recall that $\Theta(\mathcal{X},\mathcal{Y}_\theta) = (\theta,0)$. In particular, $\tan\Theta(\mathcal{X},\mathcal{Y}_\theta) = (\tan\theta,0)$ in this case.

It is clear that $\mathcal{X}+\mathcal{Y}_\theta = \operatorname{span}\{e_1,\,e_2,\,e_3\}$. Let

$$X = \begin{pmatrix} 1 & 0\\ 0 & 1\\ 0 & 0\\ 0 & 0\end{pmatrix} \quad , \quad X_\perp = \begin{pmatrix} 0 & 0\\ 0 & 0\\ 1 & 0\\ 0 & 1\end{pmatrix} \quad \text{and} \quad Y_\theta = \begin{pmatrix} 1 & 0\\ 0 & \cos\theta\\ 0 & \sin\theta\\ 0 & 0\end{pmatrix}.$$

Then, we have that $\lambda(Y_\theta^*\tilde A Y_\theta) = (b\,\cos^2\theta + d\,\sin^2(\theta), a)$, while $\lambda(X_\perp^*\tilde A\,X_\perp) = (d,c)$. Therefore, if we let $\theta_0(c) = \theta_0 = \arcsin\left(\sqrt{\frac{c-b}{d-b}}\right)$ and set

$$\delta_\theta = c - (b\,\cos^2\theta + d\,\sin^2\theta) > 0 \quad \text{for} \quad 0 < \theta < \theta_0\,,$$

then $(\tilde A,\mathcal{X},\mathcal{Y}_\theta,\delta_\theta)$ satisfies the DKN separation property, and δ_θ is the optimal (largest) separation constant and the separation property holds only for $0 < \theta < \theta_0$ in this case. Again, simple computations show that $s(R_{Y_\theta}) = ((d-b)\cos\theta\,\sin\theta, 0)$.

Now, (3.30) obtained from Theorem 3.16 becomes

$$\tan\theta \leq \frac{(d-b)\cos\theta\,\sin\theta}{c-(b\,\cos^2\theta + d\,\sin^2\theta)} \quad \text{for} \quad 0 < \theta < \theta_0\,. \tag{3.32}$$

Notice that $\lim_{c\to b^+}\theta_0 = 0$ i.e., the range of θ for which we can apply the bound in (3.32) tend to become small. In the limit case in which $b = c$ (i.e. multiple eigenvalues) we can not apply the bound (3.32) (the separation constant in this case is $\delta_0 = 0$). Finally, if we consider the limit case in which θ becomes small, then the upper bound is comparable with the upper bound $(\frac{d-b}{c-b})\,\tan\theta$ $(>\tan\theta)$.

On the other hand, $\mathcal{X}+\mathcal{Y}_\theta \ominus \mathcal{X} = \mathbb{C}\,e_3$, the subspace spanned by e_3. In this case, if we let $X' = (0,0,1,0)^t$, it is clear that $\lambda((X_S')^*\tilde A\,X_S') = d$. Therefore, if we let

$\delta'_\theta = d - (b\cos^2\theta + d\sin^2\theta) = (d-b)\cos^2\theta > 0$, for $\theta \in (0,\pi/2)$, we get that $(\tilde{A}_{\mathcal{X}+\mathcal{Y}_\theta}, S^*\mathcal{X}, S^*\mathcal{Y}_\theta, \delta'_\theta)$ satisfies the DKN separation property, where $S \in \mathcal{M}_{4,3}(\mathbb{C})$ is the matrix whose columns are the first three elements in the canonical basis. In this case we have that

$$\frac{s_1(P_{\mathcal{X}+\mathcal{Y}_\theta}\, R_{Y_\theta})}{\delta'_\theta} = \frac{(d-b)\cos\theta\,\sin\theta}{(d-b)\,\cos^2\theta} = \tan\theta\,,$$

and hence, the upper bound in (3.31) coincides with $\tan\theta$ (where $\tan\Theta(\mathcal{X},\mathcal{Y}_\theta) = (\tan\theta, 0)$) i.e. the upper bound is sharp. Notice that the bound is applicable for every $\theta \in (0,\pi/2)$. △

The following result was conjectured in [20].

Corollary 3.22. *Let $A \in \mathcal{H}(d)$, $\mathcal{X}$, $\mathcal{Y} \subset \mathbb{C}^d$ and $\delta > 0$ be such that $(A\,,\mathcal{X}\,,\mathcal{Y}\,,\delta)$ satisfies the DKN separation property. Then,*

$$\delta\,\|\tan(\Theta(\mathcal{X}\,,\,\mathcal{Y}))\| \leq \|P_{\mathcal{X}+\mathcal{Y}}\ R_Y\|\,.$$

for every unitarily invariant norm $\|\cdot\|$.

Proof. Let $S \in \mathcal{M}_{d,p}(\mathbb{C})$ be such that its columns form an ONB for $\mathcal{X}+\mathcal{Y}$. By item 1 in Theorem 3.19, there exists $\delta' \geq \delta$ such that $(S^*A\,S\,,\ S^*\mathcal{X}\,,\ S^*\mathcal{Y}\,,\ \delta')$ satisfies the DKN separation property. By item 2 of the same result, we have that

$$\delta\,\|\tan(\Theta(\mathcal{X}\,,\,\mathcal{Y}))\| \leq \delta'\,\|\tan(\Theta(\mathcal{X}\,,\,\mathcal{Y}))\| \leq \|P_{\mathcal{X}+\mathcal{Y}}\ R_Y\|\,.$$

Finally, we get the following quadratic a posteriori error bound for the simultaneous approximation of eigenvalues of A by the Ritz values corresponding to Rayleigh quotients for which a DKN separation property holds.

Theorem 3.23. *Let $A \in \mathcal{H}(d)$, $\mathcal{X}$, $\mathcal{Y} \subset \mathbb{C}^d$ and $\delta > 0$ be such that $(A\,,\,\mathcal{X}\,,\,\mathcal{Y}\,,\,\delta)$ satisfies the DKN separation property. Then, for every unitarily invariant norm $\|\cdot\|$ we have that*

$$\|\lambda(\rho(X)) - \lambda(\rho(Y))\| \leq \frac{\|P_{\mathcal{X}+\mathcal{Y}}\,R_Y\|^2}{\delta}\,.$$

Proof. This is a consequence of Corollary 3.4 and Theorem 3.19. ∎

Theorem 3.23 allows to obtain the following extension of [19, Theorem 5.3] (see Remark 3.25 below) which is a quadratic a posteriori majorization error bound for simultaneous approximation of consecutive eigenvalues.

Corollary 3.24. *Let $A \in \mathcal{H}(d)$ and let $\mathcal{Y} \subset \mathbb{C}^d$ be such that:*

1. *$\lambda_1(Y^*AY) < \lambda_j(A)$, where $j \in \mathbb{I}_{d-k}$ is the smallest such index;*
2. *$\lambda_i(Y^*AY) \geq \lambda_{i+j}(A)$, for $i \in \mathbb{I}_k$.*

Let $\mathcal{U}$ be the A-invariant space spanned by the eigenvectors associated with $\lambda_i(A)$, for $1 \leq i \leq j$, and set $\mathcal{X} = (I - P_U)\mathcal{Y}$. If $\eta = \lambda_j(A) - \lambda_1(Y^*AY) > 0$ then

$$\|(\lambda_{i+j}(A))_{i\in\mathbb{I}_k} - \lambda(\rho(Y))\| \leq \frac{\|P_{\mathcal{X}+\mathcal{Y}}\,R_Y\|^2}{\eta}\,,$$

for every unitarily invariant norm $\|\cdot\|$.

Proof. Let $\mathcal{V} = \mathcal{U} + \mathcal{Y}$ and notice that $\mathcal{U} \cap \mathcal{Y} = \{0\}$; hence, $p = \dim \mathcal{V} = \dim \mathcal{U} + k$ i.e. $j = \dim \mathcal{U} = p - k$. Moreover, $\mathcal{V} \ominus \mathcal{U} = (I - P_{\mathcal{U}})\mathcal{Y} = \mathcal{X}$; then, in particular, $\dim \mathcal{X} = \dim \mathcal{Y}$ and $\mathcal{V} \ominus \mathcal{X} = \mathcal{U}$. Also notice that $\Theta_1(\mathcal{X}, \mathcal{Y}) < \pi/2$ or otherwise, we would have that $\mathcal{U} \cap \mathcal{Y} \neq \{0\}$, since $\mathcal{V} \ominus \mathcal{X} = \mathcal{U}$.

Let $V \in \mathcal{M}_{d,p}(\mathbb{C})$ be such that its columns form an ONB of $\mathcal{V}$ and set $A_V = V^* A V \in \mathcal{H}(p)$. Similarly, let $X, Y \in \mathcal{M}_{d,k}(\mathbb{C})$, $U \in \mathcal{M}_{d,p-k}(\mathbb{C})$ be such that their columns form ONB's of $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{U}$ respectively; set $X_V = V^* X$, $Y_V = V^* Y \in \mathcal{M}_{p,k}(\mathbb{C})$ and $U_V = V^* U \in \mathcal{M}_{p,p-k}(\mathbb{C})$. Then, the columns of U_V span $\mathcal{U}_V \subset \mathbb{C}^p$ an A-invariant space of A_V. In particular, the columns of X_V span $\mathcal{X}_V \subset \mathbb{C}^p$ which is also an A-invariant space of A_V. In this case $\mathcal{X}_V^\perp = \mathcal{U}_V$ and $\Theta_1(\mathcal{X}_V, \mathcal{Y}_V) = \Theta_1(\mathcal{X}, \mathcal{Y}) < \pi/2$, where $\mathcal{Y}_V \subset \mathbb{C}^p$ is the space spanned by the columns of Y_V. Notice that, by construction $\lambda_i(Y_V^* A_V\, Y_V) = \lambda_i(Y^* A\, Y)$, for $i \in \mathbb{I}_k$. Since $\mathcal{X} \subset \mathcal{U}^\perp$ by the interlacing inequalities for compressions of self-adjoint matrices and item 2 above, we get that for $i \in \mathbb{I}_k$,

$$\lambda_i(X_V^* A_V\, X_V) = \lambda_i(X^* A\, X) \leq \lambda_i(A_{U_\perp}) = \lambda_{j+i}(A) \leq \lambda_i(Y_V^* A_V\, Y_V)\,, \tag{3.33}$$

where $U_\perp \in \mathcal{M}_{d,d-j}(\mathbb{C})$ is such that its columns for an ONB for $\mathcal{U}^\perp$. On the other hand, by hypothesis $(A_V, \mathcal{X}_V, \mathcal{Y}_V, \eta)$ satisfies the DKN separation property (recall that $\mathcal{X}_V^\perp = \mathcal{U}_V$). Hence, by Theorem 3.23 we conclude that

$$\|\lambda(X_V^* A_V\, X_V) - \lambda(Y_V^* A_V\, Y_V)\| \leq \frac{\|\, P_{\mathcal{X}_V + \mathcal{Y}_V}\, (A_V\, Y_V - Y_V\, (Y_V^* A_V\, Y_V))\,\|^2}{\eta}\,. \tag{3.34}$$

By (3.33) we get that

$$|(\lambda_{i+j}(A))_{i \in \mathbb{I}_k} - \lambda(Y_V^* A_V\, Y_V)| \prec_w |\lambda(X_V^* A_V\, X_V) - \lambda(Y_V^* A_V\, Y_V)|\,.$$

On the other hand, arguing as in the proof of Theorem 3.19 we see that

$$\|\, P_{\mathcal{X}_V + \mathcal{Y}_V}\, (A_V\, Y_V - Y_V\, (Y_V^* A_V\, Y_V))\,\| = \|\, P_{\mathcal{X} + \mathcal{Y}}\, R_Y\,\|\,.$$

The result follows from these last facts together with (3.34) and Remark 2.4. ∎

Remark 3.25. We mention that the hypothesis in item 1 in Corollary 3.24 is that there exists an eigenvalue β of A such that $\lambda_1(Y^* A Y) < \beta$. Indeed, in this case we can apply the interlacing inequalities and get that $\lambda_i(Y^* A Y) \geq \lambda_{d-k+i}(A)$, for $i \in \mathbb{I}_k$. Therefore, $\beta = \lambda_j(A)$ for some $1 \leq j \leq d - k$.

The hypothesis in item 2 is rather restrictive and difficult to check in general. Nevertheless, we mention two cases in which the hypotheses in Corollary 3.24 can be easily checked:

1. In case the hypothesis in item 1 holds for $j = d - k$, by the interlacing inequalities we have
$$\lambda_i(Y^* A Y) \geq \lambda_{i+d-k}(A) \quad \text{for} \quad i \in \mathbb{I}_k\,,$$
so the hypothesis in item 2 automatically hold.
2. In case $k = 1$ that is, if $\mathcal{Y} = \mathbb{C}\, y$ for a unit norm vector $y \in \mathbb{C}^d$, the hypotheses become the existence of $j \in \mathbb{I}_{d-1}$ such that $\lambda_{j+1}(A) \leq \langle A\, y,\, y\rangle < \lambda_j(A)$; then, Corollary 3.24 implies that
$$0 \leq \langle Ay,\, y\rangle - \lambda_{j+1}(A) \leq \frac{\|P_{\mathcal{X}+\mathcal{Y}}(Ay - \langle Ay,\, y\rangle\, y)\|}{\lambda_j(A) - \langle Ay,\, y\rangle}\,,$$

where $\mathcal{X} = \mathbb{C}\, x$, for $x = (I - P_U)y \in \mathbb{C}^d$; this is [19, Theorem 5.3]. As explained in [19], Corollary 3.24 encodes several known bounds related with eigenvalue estimation even when $k = 1$. △

4. Appendix. Here we collect several and well known results about majorization, used throughout our work. The first result deals with submajorization relations between singular values of arbitrary matrices in $\mathcal{M}_d(\mathbb{C})$. For detailed proofs of these results and general references in majorization theory see [2, 6, 12]. For $A \in \mathcal{M}_d(\mathbb{C})$ we denote by $\mathrm{re}(A) = \frac{A+A^*}{2} \in \mathcal{H}(d)$.

Theorem 4.1. Let $C,\, D \in \mathcal{M}_d(\mathbb{C})$. Then,
1. $s(C + D) \prec_w s(C) + s(D)$; (Lidskii's additive property)
2. $s(\mathrm{re}(C)) \prec_w s(C)$;
3. $s(CD) \prec_w s(C)\, s(D)$; (Lidskii's multiplicative property)
4. If we assume that $CD \in \mathcal{H}(d)$ then $s(CD) \prec_w s(\mathrm{re}(DC))$.

For hermitian matrices we have the following majorization relations

Theorem 4.2. Let $C,\, D \in \mathcal{H}(d)$. Then,
1. $\lambda(C) - \lambda(D) \prec \lambda(C - D) \prec \lambda(C) - \lambda^{\uparrow}(D)$;
2. $|\lambda(C) - \lambda(D)| \prec_w s(C - D)$;
3. Let $\mathcal{P} = \{P_j\}_{j=1}^r$ be a system of projections (i.e. they are mutually orthogonal projections on $\mathbb{C}^d$ such that $\sum_{i=1}^r P_i = I$). If $C_{\mathcal{P}}(C) = \sum_{i=1}^r P_i C P_i$, then $\lambda(C_{\mathcal{P}}(C)) \prec \lambda(C)$.

In the next result we describe elementary but useful properties of (sub)majorization between real vectors.

Lemma 4.3. Let $x,\, y,\, z \in \mathbb{R}^k$. 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 $x + z \prec_w y + z$;

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

3. $x^{\downarrow}\, y^{\uparrow} \prec_w x\, y \prec_w x^{\downarrow}\, y^{\downarrow}$;
4. If $x \prec_w y$ and $y,\, z \in (\mathbb{R}^k_{\geq 0})^{\downarrow}$ then $x\, z \prec_w y\, z$.

Proposition 4.4. Let $1 \leq k < d$ and let $E \in \mathbb{M}_{k,(d-k)}(\mathbb{C})$. Then

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

Theorem 4.5 ([8]). Let $\mathcal{X},\, \mathcal{Y} \subset \mathbb{C}^d$ be such that $\dim(\mathcal{X}) = \dim(\mathcal{Y}) = k$. Then

$$\lambda(P_{\mathcal{X}} P_{\mathcal{Y}^\perp} P_{\mathcal{X}}) = s(P_{\mathcal{X}} P_{\mathcal{Y}^\perp} P_{\mathcal{X}}) = s^2(P_{\mathcal{Y}} P_{\mathcal{X}^\perp}) = s^2(P_{\mathcal{X}^\perp} P_{\mathcal{Y}}) = (\sin^2(\Theta(\mathcal{X}, \mathcal{Y})), 0_{d-k}).$$

Notice that item 2 below is Theorem 3.2 from Section 3.

Theorem 4.6. *Let $C,\, D \in \mathcal{H}(k)$. Then,*
1. *if $T \in \mathcal{G}l(k)^+$, then $s(C - D) \prec_w s(T^{-1})\, s(CT - TD)$.*
2. *if $T \in \mathcal{G}l(k)$, then $|\lambda(C) - \lambda(D)| \prec_w s(T^{-1})\, s(CT - TD)$.*

Proof. We first show item 1 Since T is positive and invertible, using Theorem 4.2 (item 3) we get that

$$\begin{aligned} s(C - D) &= s(CT^{\frac12} T^{-\frac12} - T^{-\frac12} T^{\frac12} D)) = s(T^{-\frac12}(T^{\frac12} C T^{\frac12} - T^{\frac12} D T^{\frac12}) T^{-\frac12}) \\ &\prec_w s(T^{-\frac12})^2\, s(T^{\frac12} C T^{\frac12} - T^{\frac12} D T^{\frac12}) = s(T^{-1})\, s(T^{\frac12} (C - D)\, T^{\frac12})\,. \end{aligned}$$

By Theorem 4.1 (items 2 and 4) and the fact that $\mathrm{re}(DT) = \mathrm{re}(TD)$ we obtain that

$$s(T^{\frac{1}{2}}(C-D)T^{\frac{1}{2}}) \prec_w s(\mathrm{re}[(C-D)T]) = s(\mathrm{re}[CT-TD]) \prec_w s(CT-TD), \tag{4.1}$$

By the previous inequalities and Lemma 4.3 we see that

$$s(C-D) \prec_w s(T^{-1})\, s(CT-TD)\,. \tag{4.2}$$

In order to show item 2, consider a representation of T given by $T = U\Sigma V^*$, where $U, V \in \mathcal{U}(k)$ are unitary matrices and $\Sigma \in \mathcal{M}_k(\mathbb{C})$ is the diagonal matrix with main diagonal $s(T) \in \mathbb{R}^k_{\geq 0}$ (notice that such representation follows from the SVD decomposition of T); note that Σ is definite positive and invertible. Using item 2 in Theorem 4.2 and (the already proved) item 1 of the statement we get

$$\begin{aligned}|\lambda(C)-\lambda(D)| &= |\lambda(U^*CU)-\lambda(V^*DV)| \prec_w s(U^*CU-V^*DV)\\ &\prec_w s(\Sigma^{-1})\, s(U^*CU\Sigma-\Sigma V^*DV) = s(T^{-1})\, s(U^*(CT-TD)V)\\ &= s(T^{-1})\, s(CT-TD)\,.\end{aligned}$$

□

In what follows we re-state and prove two propositions of Section 3.2.

Proposition 3.10. Let $A \in \mathcal{H}(d)$ and let $\mathcal{X}, \mathcal{Y} \subset \mathbb{C}^d$ with $\dim(\mathcal{X}) = \dim(\mathcal{Y}) = k$. Then

$$s(P_\mathcal{X}\, R_Y) \prec_w \mathrm{Spr}(A, \mathcal{X}+\mathcal{Y})\, \sin(\Theta(\mathcal{X},\mathcal{Y}))\,. \tag{4.3}$$

Proof. We begin with a simple reduction argument. In order to describe this reduction it will be convenient to consider matrices in terms of the linear operators that they induce. Hence, given $A \in \mathcal{M}_d(\mathbb{C})$, we consider $A \in \mathcal{L}(\mathbb{C}^d)$ (defined in the obvious way). The advantage in considering $A \in \mathcal{L}(\mathbb{C}^d)$ is that we can get different block matrix representations of A (considered as an operator) with respect to orthogonal decompositions $\mathbb{C}^d = \mathcal{V} \oplus \mathcal{V}^\perp$ for a (proper) subspace $\mathcal{V} \subset \mathbb{C}^d$, in the usual manner. We now proceed as follows: Let $\mathcal{Z} = \mathcal{X}+\mathcal{Y}$ with $\dim \mathcal{Z} = p$, and consider the matrix representations with respect to the decomposition $\mathbb{C}^d = \mathcal{Z} \oplus \mathcal{Z}^\perp$:

$$P_\mathcal{X} = \begin{pmatrix} P^\mathcal{X} & 0 \\ 0 & 0 \end{pmatrix} \ ,\ P_\mathcal{Y} = \begin{pmatrix} P^\mathcal{Y} & 0 \\ 0 & 0 \end{pmatrix} \quad \text{and} \quad A = \begin{pmatrix} A_\mathcal{Z} & * \\ * & * \end{pmatrix},$$

where $P^\mathcal{X}$, $P^\mathcal{Y}$, $A_\mathcal{Z} = P_\mathcal{Z} A|_\mathcal{Z} \in \mathcal{L}(\mathcal{Z})$ are self-adjoint operators. In this case we have

$$P_\mathcal{X}\,(A\,P_\mathcal{Y} - P_\mathcal{Y}\,A\,P_\mathcal{Y}) = \begin{pmatrix} P^\mathcal{X}\,(A_\mathcal{Z}\,P^\mathcal{Y} - P^\mathcal{Y}\,A_\mathcal{Z}\,P^\mathcal{Y}) & 0 \\ 0 & 0 \end{pmatrix}.$$

On the other hand, a simple calculation show that

$$(s(P_\mathcal{X} R_Y), 0_{d-k}) = s(P_\mathcal{X}\,(A\,P_\mathcal{Y} - P_\mathcal{Y}\,A\,P_\mathcal{Y})) \in (\mathbb{R}^d_{\geq 0})^\downarrow\,.$$

Hence, $(s(P_\mathcal{X} R_Y), 0_{p-k}) = s(P^\mathcal{X}\,(A_\mathcal{Z}\,P^\mathcal{Y} - P^\mathcal{Y}\,A_\mathcal{Z}\,P^\mathcal{Y})) = s(P^\mathcal{X}(I_\mathcal{Z} - P^\mathcal{Y})\,A_\mathcal{Z}\,P^\mathcal{Y})$. Thus, we can assume further that $\mathbb{C}^d = \mathcal{Z} = \mathcal{X}+\mathcal{Y}$ and show that

$$(s(P_\mathcal{X}\,R_Y), 0_{d-k}) = s(P_\mathcal{X}\,(P_{\mathcal{Y}^\perp}\,A\,P_\mathcal{Y})) \prec_w (\mathrm{Spr}(A)\,\sin(\Theta(\mathcal{X},\mathcal{Y})), 0_{d-k})\,. \tag{4.4}$$

Now using item 3 of Theorem 4.1 (Lidskii's multiplicative property),

$$s(P_\mathcal{X} P_{\mathcal{Y}^\perp} A P_\mathcal{Y}) = s(P_\mathcal{X} P_{\mathcal{Y}^\perp} P_{\mathcal{Y}^\perp} A P_\mathcal{Y}) \prec_w s(P_\mathcal{X} P_{\mathcal{Y}^\perp})\, s(P_{\mathcal{Y}^\perp} A P_\mathcal{Y}). \tag{4.5}$$

First noticing that by Theorem 4.5, we have that $s(P_{\mathcal{X}} P_{\mathcal{Y}^\perp}) = (\sin(\Theta(\mathcal{X},\mathcal{Y})), 0_{d-k})$. On the other hand, consider the matrix representation induced by the decomposition $\mathbb{C}^d = \mathcal{Y} \oplus \mathcal{Y}^\perp$:

$$A = \begin{pmatrix} A_{11} & A_{21}^* \\ A_{21} & A_{22} \end{pmatrix} \text{ and set } A_1 := \begin{pmatrix} A_{11} & 0 \\ 0 & A_{22} \end{pmatrix} \ , \ A_2 := \begin{pmatrix} 0 & A_{21}^* \\ A_{21} & 0 \end{pmatrix} . \tag{4.6}$$

Then, we have that $A = A_1 + A_2$. Now, A_1 is a pinching of A (associated with the system of projections $\{P_{\mathcal{Y}}\ ,\ P_{\mathcal{Y}^\perp}\}$) so $\lambda(A_1) \prec \lambda(A)$ so then

$$-\lambda^\uparrow(A_1) \prec -\lambda^\uparrow(A) . \tag{4.7}$$

Using Lidskii's additive property for $A_2 = A - A_1$ (see item 1 in Theorem 4.2)

$$\lambda(A_2) \prec \lambda(A) - \lambda^\uparrow(A_1) . \tag{4.8}$$

Combining (4.7) and (4.8), we obtain

$$\lambda(A_2) \prec \lambda(A) - \lambda^\uparrow(A) = \mathrm{Spr}(A) \in \mathbb{R}^d . \tag{4.9}$$

By Proposition 4.4, we get that $\lambda(A_2) = (s(A_{21}), -s(A_{21}^*))^\downarrow$; in particular, $s(A_{21}) = (\lambda_i(A_2))_{i\in\mathbb{I}_k}$. Now, $s(P_{\mathcal{Y}^\perp} A P_{\mathcal{Y}}) = (s(A_{21}), 0_{d-k})$; thus, we see that

$$s(P_{\mathcal{Y}^\perp} A P_{\mathcal{Y}}) = (s(A_{21}), 0_{d-k}) = ((\lambda_i(A_2))_{i\in\mathbb{I}_k}, 0_{d-k}) \prec_w ((\mathrm{Spr}_i(A))_{i\in\mathbb{I}_k}, 0_{d-k}) , \tag{4.10}$$

where $\mathrm{Spr}(A) = (\mathrm{Spr}_i(A))_{i\in\mathbb{I}_d}$. Using (4.5) and (4.10) together with Lemma 4.3 we finally get that

$$s(P_{\mathcal{X}} P_{\mathcal{Y}^\perp} A P_{\mathcal{Y}}) \prec_w (\mathrm{Spr}(A)\ \sin(\Theta(\mathcal{X},\mathcal{Y})), 0_{d-k}) \in (\mathbb{R}^d_{\geq 0})^\downarrow .$$

Now the result follows from the last submajorization relation, by considering the first k entries of both vectors. ∎

Proposition 3.12. Let $A \in \mathcal{H}(d)$, $\mathcal{X}, \mathcal{Y} \subset \mathbb{C}^d$ subspaces with $\dim(\mathcal{X}) = \dim(\mathcal{Y}) = k$. Assume that $\mathcal{X}$ is A-invariant. Then,

$$s(P_{\mathcal{X}} R_Y) \prec_w 2\ (\lambda_i(A_{\mathcal{X}+\mathcal{Y}}) - \lambda_{\min}(A_{\mathcal{X}+\mathcal{Y}}))_{i\in\mathbb{I}_k}\ \sin^2(\Theta(\mathcal{X},\mathcal{Y})). \tag{4.11}$$

Proof. Arguing as in the proof of Proposition 3.10, we can assume further that $\mathbb{C}^d = \mathcal{X} + \mathcal{Y}$. With this assumption, we consider first the case where $A \in \mathcal{M}_d(\mathbb{C})^+$ and show that

$$s(P_{\mathcal{X}} R_Y) \prec_w 2\,(\lambda_i(A))_{i\in\mathbb{I}_k}\ \sin^2(\Theta(\mathcal{X},\mathcal{Y})) . \tag{4.12}$$

Indeed, the A-invariance of $\mathcal{X}$, allows us to write $A = P_{\mathcal{X}} A P_{\mathcal{X}} + P_{\mathcal{X}^\perp} A P_{\mathcal{X}^\perp}$. With this decomposition in mind using the fact that $(s(P_{\mathcal{X}} R_Y), 0_{d-k}) = s(P_{\mathcal{X}} P_{\mathcal{Y}^\perp} A P_{\mathcal{Y}})$, we have that

$$\begin{aligned} s(P_{\mathcal{X}} P_{\mathcal{Y}^\perp} A\, P_{\mathcal{Y}}) &= s(P_{\mathcal{X}} P_{\mathcal{Y}^\perp} P_{\mathcal{X}} A\, P_{\mathcal{X}} P_{\mathcal{Y}} + P_{\mathcal{X}} P_{\mathcal{Y}^\perp} P_{\mathcal{X}^\perp} A\, P_{\mathcal{X}^\perp} P_{\mathcal{Y}}) \\ &\prec_w s(P_{\mathcal{X}} P_{\mathcal{Y}^\perp} P_{\mathcal{X}} A\, P_{\mathcal{X}} P_{\mathcal{Y}}) + s(P_{\mathcal{X}} P_{\mathcal{Y}^\perp} A\, P_{\mathcal{X}^\perp} P_{\mathcal{Y}}) \overset{\text{def}}{=} M \ . \end{aligned}$$

Using item 3 of Theorem 4.1 (Lidskii's multiplicative property), the fact that $0_d \leq s(P_{\mathcal{X}}\, P_{\mathcal{Y}}) \leq \mathbb{1}_d$ and Theorem 4.5, we get

$$\begin{aligned} M &\prec_w s(P_{\mathcal{X}} P_{\mathcal{Y}^\perp} P_{\mathcal{X}})\, s(A) + s(P_{\mathcal{X}} P_{\mathcal{Y}^\perp})\, s(A)\, s(P_{\mathcal{X}^\perp} P_{\mathcal{Y}}) \\ &\prec_w 2\,\lambda(A)\,(\sin^2(\Theta(\mathcal{X},\mathcal{Y})), 0_{d-k}) \in (\mathbb{R}^d_{\geq 0})^\downarrow\,, \end{aligned}$$

since $A \in \mathcal{M}_d(\mathbb{C})^+$ is positive semi-definite. The result now follows from the previous facts.

In general, for $A \in \mathcal{H}(d)$ consider the auxiliary matrix $\tilde{A} = A - \lambda_{\min(A)}\, I \in \mathcal{M}_d(\mathbb{C})^+$. Notice that

$$R_Y(\tilde{A}) = \tilde{A}\,Y - Y(Y^*\tilde{A}\,Y) = A\,Y - Y(Y^*A\,Y) = R_Y\,,$$

and $\lambda(\tilde{A}) = \lambda(A) - \lambda_{\min(A)}\,\mathbb{1}_d$. The result now follows from these facts and from (4.12) applied to $\tilde{A}$. ∎

Acknowledgment. We would like to thank the reviewers of the manuscript for providing several useful comments that helped us improve the presentation of the results herein.

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] Bhatia, R., 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] P.R. Halmos, Two subspaces. Trans. Amer. Math. Soc. 144 1969 381-389.
[6] R.A. Horn, C.R: and Johnson, Matrix analysis. Second Edition, Cambridge University Press, Cambridge, 2012.
[7] A. V. Knyazev and M. E. Argentati, Principal angles between subspaces in an A-based scalar product: algorithms and perturbation estimates. SIAM J. Sci. Comput. 23 (2002), no. 6, 2008-2040.
[8] 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.
[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, 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.
[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] R. Mathias, Quadratic residual bounds for the Hermitian eigenvalue problem. SIAM J. Matrix Anal. Appl. 19 (1998), no. 2, 541-550.
[14] Nakatsukasa, Y., The tan θ theorem with relaxed conditions. Linear Algebra Appl. 436 (2012), no. 5, 1528-1534.
[15] E. Ovtchinnikov, Cluster robust error estimates for the Rayleigh-Ritz approximation. II. Estimates for eigenvalues. Linear Algebra Appl. 415 (2006), no. 1, 188-209.
[16] B.N. Parlett, The symmetric eigenvalue problem. Corrected reprint of the 1980 original, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1998.
[17] G.W. Stewart, J. Sun, Matrix perturbation theory, Computer Science and Scientific Computing, Academic Press Inc., Boston, MA, 1990.

[18] Z. Teng, L. Lu, R.-C. Li, Cluster-robust accuracy bounds for Ritz subspaces. Linear Algebra Appl. 480 (2015), 11-26.
[19] 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.
[20] P. Zhu, A.K. Knyazev, Rayleigh-Ritz majorization error bounds of mixed type. SIAM J. Matrix Anal. Appl. 38 (2017), no. 1, 30-49.
[21] P. Zhu, Angles between subspaces and the Rayleigh-Ritz method, Ph.D thesis, University of Colorado Denver, 2012.