Λ-admissible subspaces of self-adjoint matrices

Francisco Arrieta Zuccalli and Pedro Massey [*]
Centro de Matemática, FCE-UNLP, La Plata and IAM-CONICET, Argentina

Abstract

Given a self-adjoint matrix A and an index h such that $\lambda_h(A)$ lies in a cluster of eigenvalues of A, we introduce the novel class of Λ-admissible subspaces of A of dimension h. First, we show that the low-rank approximation of the form $P_{\mathcal{T}} A P_{\mathcal{T}}$, for a subspace $\mathcal{T}$ that is close to any Λ-admissible subspace of A, has nice properties. Then, we prove that some well-known iterative algorithms (such as the Subspace Iteration Method, or the Krylov subspace method) produce subspaces that become arbitrarily close to Λ-admissible subspaces. We obtain upper bounds for the distance between subspaces obtained by the Rayleigh-Ritz method applied to A and the class of Λ-admissible subspaces. We also find upper bounds for the condition number of the (set-valued) map computing the class of Λ-admissible subspaces of A. Finally, we include numerical examples that show the advantage of considering this new class of subspaces in the clustered eigenvalue setting.

AMS subject classification: 42C15, 15A60.
Keywords: principal angles, Ritz values, Rayleigh quotients, majorization.

1 Introduction

One central task in numerical linear algebra is to compute *nice* self-adjoint low-rank approximations $\tilde{A}$ of $n \times n$ self-adjoint matrices A. Even when A has a rich inner structure, we can be interested in reducing the computational cost of manipulations of A. By *nice*, we mean low-rank approximations of A that provide good approximations of its extreme eigenvalues and are such that the approximation error $\|A - \tilde{A}\|$ is close to $\min\{\|A - B\| :\ B \text{ is self-adjoint, } \operatorname{rk}(B) \leq h\}$ (see [10]).

One standard approach for the computation of such nice low-rank approximations is the construction of suitable subspaces $\mathcal{T} \subset \mathbb{C}^n$ with $\dim \mathcal{T} = h$ so that $\tilde{A}$ becomes the compression of A to $\mathcal{T}$. That is, $\tilde{A} = P_{\mathcal{T}}\, A\, P_{\mathcal{T}}$, where $P_{\mathcal{T}}$ denotes the orthogonal projection onto $\mathcal{T}$. If we assume further that A is positive semidefinite and that we are interested in the approximation of the largest h eigenvalues $\lambda_1 \geq \ldots \geq \lambda_h \geq 0$ of A, then we typically seek subspaces $\mathcal{T}$ that are close (with respect to some angular metric) to dominant eigenspaces, i.e. subspaces $\mathcal{X}_h$ spanned by orthonormal systems $\{x_1, \ldots, x_h\}$ of eigenvectors of A corresponding to $\lambda_1 \geq \ldots \geq \lambda_h$. In this case, the closer that $\mathcal{T}$ and $\mathcal{X}_h$ are, the better $P_{\mathcal{T}} A P_{\mathcal{T}}$ is as a low-rank approximation of A.

There are several numerical methods for the computation of subspaces $\mathcal{T}$ as above (e.g. when A is positive semidefinite as before); among them, we mention the Subspace Iteration Method (SIM) and the Krylov subspace method (see [16]). The convergence analysis of the subspaces $\mathcal{T}$ constructed by these numerical methods to the dominant eigenspace $\mathcal{X}_h$, obtained in the deterministic setting (intimately related to our present approach), reveals that the inverse of the eigen-gap $(\lambda_h - \lambda_{h+1})^{-1}$ plays a key role [1, 3, 5, 17, 21]. Explicitly, the smaller the inverse of the eigen-gap is, the faster

---

[*] Partially supported by CONICET (PICT ANPCyT 1505/15) and Universidad Nacional de La Plata (UNLP 11X829) e-mail addresses: farrieta@mate.unlp.edu.ar , massey@mate.unlp.edu.ar

these numerical methods converge. Thus, these results provide theoretical foundations for the application of methods such as the SIM and the Krylov method when the eigen-gap is significant. Moreover, theoretical results about the condition number of the computation of h-dimensional dominant eigenspaces also reveal the role played by the inverse of the eigen-gap $(\lambda_h - \lambda_{h+1})^{-1}$ in the approximation of these dominant eigenspaces (see [19, 20]).

Our main motivation for this work is the introduction of new tools and results that allow us to deal with the analysis of the quality of subspaces $\mathcal{T}$, with $\dim \mathcal{T} = h$, so that $P_{\mathcal{T}} A P_{\mathcal{T}}$ is a nice low-rank approximation of A, in case λ_h lies in a cluster of eigenvalues. Thus, our work can be regarded as complementary to the previous literature. To put our work in context, consider the case where $\lambda_h - \lambda_{h+1} = 0$. Notice that this case can not be analyzed as a limiting case when $\lambda_h - \lambda_{h+1}$ tends to zero: the main reason for this seems to be that, in the limit $\lambda_h - \lambda_{h+1} = 0$, instead of getting a uniquely-defined dominant eigenspace $\mathcal{X}_h$ of dimension h, we get infinitely many. At this point, it is essential to notice if $\mathcal{T}$ is sufficiently close to *any* dominant subspace of A, then $P_{\mathcal{T}} A\, P_{\mathcal{T}}$ will be a nice low-rank approximation of A. This last fact suggests a shift of perspective in the analysis, so that given a suitable candidate subspace $\mathcal{T}$ with $\dim \mathcal{T} = h$, we should be interested in bounding from above the angular distance between $\mathcal{T}$ and the class of h-dimensional dominant subspaces of A. Indeed, if we measure the distance between h-dimensional subspaces $\mathcal{S}$ and $\mathcal{T}$ in terms of the sine of the largest principal angle [13, 14, 18] we should be interested in computing an upper bound for

$$\inf\{\sin \theta_{\max}(\mathcal{X}, \mathcal{T}) \ : \ \mathcal{X} \ \text{ is a dominant subspace for } A,\ \dim(\mathcal{X}) = h \ \}\,.$$

Notice that bounding the quantity above is a non-trivial problem: while $\sin \theta_{\max}(\mathcal{X}, \mathcal{T})$ can be large for some dominant subspace $\mathcal{X}$ of A, $\sin \theta_{\max}(\tilde{\mathcal{X}}, \mathcal{T})$ can be quite small for some other dominant subspace $\tilde{\mathcal{X}}$ of A. In turn, this shift of perspective induces what we call a proximity analysis to the class of dominant subspaces, rather than a convergence analysis to a fixed dominant subspace. In this work, we pursue this novel perspective of proximity analysis.

Our approach actually allows us to consider a more flexible setting: assume that λ_h lies in a cluster of eigenvalues: in this case, we consider (enveloping) indices $1 \leq j < h < k \leq \operatorname{rank}(A)$ such that a cluster of eigenvalues of A is formed by $\lambda_{j+1} \geq \ldots \geq \lambda_k$ in such a way that the spread of the cluster $\delta = \lambda_{j+1} - \lambda_k$ is small and the eigen-gaps $\lambda_j - \lambda_{j+1}$ and $\lambda_k - \lambda_{k+1}$ are significant (notice that we are allowed to choose indices, instead of requiring that there is an eigen-gap at the prescribed index h). In this setting, we introduce the class of Λ-admissible subspaces of A, given by

$$\Lambda\text{-adm}_h(A) = \{\mathcal{S} : \ \dim \mathcal{S} = h\,,\ \mathcal{X}_j \subset \mathcal{S} \subset \mathcal{X}_k\}$$

where $\mathcal{X}_j$ and $\mathcal{X}_k$ are the uniquely defined dominant eigenspaces of A of dimensions j and k respectively (notice that if $\lambda_h = \lambda_{h+1}$, then the class of h-dimensional dominant subspaces of A coincides with $\Lambda\text{-adm}_h(A)$ when we choose $j = \max\{\ell : \lambda_\ell > \lambda_h\}$ and $k = \max\{\ell : \ \lambda_h = \lambda_\ell\}$). We show that any h-dimensional Λ-admissible subspace $\mathcal{S}$ of A provides a nice low-rank approximation of A, in the sense that the approximation error $\|A - P_{\mathcal{S}} A P_{\mathcal{S}}\|$ is close to the minimal error (for an arbitrary unitarily invariant norm $\|\cdot\|$) and the first h eigenvalues of $P_{\mathcal{S}} A P_{\mathcal{S}}$ are close to those of A, up to an additive constant that depends on the spread of the cluster (the smaller the spread is, the more accurate the estimates are). On the other hand, we obtain some first quantitative results showing that h-dimensional subspaces $\mathcal{T}$ that are close to some element of $\Lambda\text{-adm}_h(A)$ will also induce nice low-rank approximations $P_{\mathcal{T}} A P_{\mathcal{T}}$ of A. We will consider a detailed quantitative analysis of the quality of $P_{\mathcal{T}} A P_{\mathcal{T}}$ as a low-rank approximation of A (in the clustered eigenvalue setting) in terms of the distance to $\Lambda\text{-adm}_h(A)$ elsewhere.

Our first main result computes an informative upper bound for the distance

$$d(\Lambda\text{-adm}_h(A), \mathcal{T}) = \inf\{\sin \theta_{\max}(\mathcal{S}, \mathcal{T}) \ : \ \mathcal{S} \in \Lambda\text{-adm}_h(A) \ \}\,.$$

This upper bound allows us to tackle several related problems: we consider subspaces of the form $\phi(A)(\mathcal{W})$, for polynomials ϕ and initial subspaces $\mathcal{W}$, with $\dim \mathcal{W} = r$ for some $h \leq r < k$ (which play a central role in the analysis of some iterative methods such as SIM or the Krylov subspace method); we show that, under some natural (generic) hypotheses, $\phi(A)(\mathcal{W})$ contains h-dimensional subspaces that become arbitrarily close to the class $\Lambda\text{-adm}_h(A)$. Thus, our results show that several well-known numerical methods can be used to construct approximations of Λ-admissible subspaces. On the other hand, motivated by some recent results by Nakatsukasa ([11, 12]), we obtain upper bounds for the distance between the class $\Lambda\text{-adm}_h(A)$ and subspaces that are constructed using the Rayleigh-Ritz method applied to A and a trial subspace $\mathcal{Q}$ with $\dim \mathcal{Q} = r$, for some $h \leq r < k$ (recall that $\lambda_{j+1} \geq \ldots \geq \lambda_k$ denote the clustered eigenvalues); notice that in this setting, previous results in the literature can only take advantage of the eigen-gap $\lambda_j > \lambda_{j+1}$, that corresponds to the analysis of subspaces of dimension $j < h$. Finally, we complement the previous results by defining a suitable condition number for the computation of the classes of Λ-admissible subspaces (in the context of set-valued functions). Our results regarding the last two topics show that both the inverse of the eigen-gaps $(\lambda_j - \lambda_{j+1})^{-1}$ and $(\lambda_k - \lambda_{k+1})^{-1}$ play a key role in our analysis. Hence, the ideal scenario for the application of our results is when we can find enveloping indices $j < h < k$ such that the eigen-gaps $\lambda_j - \lambda_{j+1}$ and $\lambda_k - \lambda_{k+1}$ are significant, and the spread of the cluster $\delta = \lambda_{j+1} - \lambda_k$ is small. Our perspective is in the vein of [4] in the sense that we can obtain nice approximations $\tilde{A}$ of A with $\text{rk}(\tilde{A}) = h$ (in terms of Λ-admissible subspaces) even when $\lambda_h = \lambda_{h+1}$.

The paper is organized as follows. In Section 2 we describe notations and notions used throughout the paper. In Section 3 we introduce Λ-admissible and consider their basic approximation properties. In Section 4 we present our main results: first, we obtain an upper bound for the distance between a generic subspace $\mathcal{T} \subset \mathbb{K}^n$ with $\dim \mathcal{T} \geq h$ and the class $\Lambda\text{-adm}_h(A)$. Then, we show that many well-known iterative algorithms used for computing approximations of dominant eigenspaces, produce subspaces that become close to $\Lambda\text{-adm}_h(A)$. We also obtain upper bounds for the distance between subspaces constructed using the Rayleigh-Ritz method using a trial subspace and the class $\Lambda\text{-adm}_h(A)$. We complement the previous analysis with upper bounds for the sensitivity of the computation of Λ-admissible subspaces. In Section 5 we present several numerical examples that test our main results and compare them with previous results in the literature; these numerical examples show the advantages of considering the class $\Lambda\text{-adm}_h(A)$ in the clustered eigenvalue setting. Finally, in Section 6 we present the proofs of all the previous results.

2 Preliminaries

Here, we include some definitions, notations, and well-known results necessary in our work.

2.1 General notation

Throughout this work, $\mathbb{K}$ denotes the field of real or complex numbers. We let $\mathbb{K}^{n\times m}$ be the space of $n \times m$ matrices with entries in $\mathbb{K}$, and we use I and 0 to denote the identity and null matrices, whose sizes will be clear from context. The Grassmannian of ℓ-dimensional subspaces of $\mathbb{K}^n$ will be denoted as $\mathcal{G}_{n,\ell} = \{\mathcal{S} \subset \mathbb{K}^n : \dim \mathcal{S} = \ell\}$. We will use $\|\cdot\|$ to denote an unitarily invariant norm (briefly u.i.n.) in $\mathbb{K}^{n\times n}$ (i.e. a norm such that $\|UAV\| = \|A\|$, for every $A \in \mathbb{K}^{n\times n}$ and unitary or orthogonal matrices U, $V \in \mathbb{K}^{n\times n}$). For example, the operator (or spectral) and Frobenius norms, denoted by $\|\cdot\|_2$ and $\|\cdot\|_F$ respectively, are u.i.n.'s. If $\mathcal{V} \subset \mathbb{K}^n$ is a subspace, we let $P_{\mathcal{V}} \in \mathbb{K}^{n\times n}$ denote the orthogonal projection onto $\mathcal{V}$.

For $A \in \mathbb{K}^{m\times n}$, we denote it's Moore–Penrose pseudo-inverse by $A^\dagger$. Among other basic properties, we use the fact that $AA^\dagger = P_{R(A)}$ and $A^\dagger A = P_{\ker(A)^\perp} = P_{R(A^*)}$, where $R(A)$ denotes the subspace of $\mathbb{K}^m$ spanned by the columns of A.

Given a vector $x \in \mathbb{K}^n$, we denote by $\operatorname{diag}(x) \in \mathbb{K}^{n\times n}$ the diagonal matrix whose main diagonal is x. If $x = (x_i)_{1\leq i\leq n} \in \mathbb{R}^n$ we denote by $x^\downarrow = (x_i^\downarrow)_{1\leq i\leq n}$ the vector obtained by rearranging the entries of x in non-increasing order. We also use the notation $(\mathbb{R}^n)^\downarrow = \{x \in \mathbb{R}^n \ : \ x = x^\downarrow\}$. If $x, y \in \mathbb{R}^n$ then x is submajorized by y, denoted $x \prec_w y$, if $\sum_{i=1}^k (x^\downarrow)_i \leq \sum_{i=1}^k (y^\downarrow)_i$ for $1 \leq k \leq n$. If $x \prec_w y$ and $\sum_{i=1}^n x_i = \sum_{i=1}^n y_i$, we say that x is majorized by y, and write $x \prec y$. For a detailed account on majorization theory see R. Bhatia's book [2]. Given an Hermitian matrix $A \in \mathbb{K}^{n\times n}$ we denote by $\lambda(A) = (\lambda_i(A))_{1\leq i\leq n} \in (\mathbb{R}^n)^\downarrow$ the eigenvalues of A, counting multiplicities and arranged in non-increasing order. Finally, we denote by $\operatorname{rk}(A)$ the rank of A.

2.2 Principal angles between subspaces

Let $\mathcal{S}, \mathcal{T} \subset \mathbb{K}^n$ be two subspaces such that $\dim \mathcal{S} = s \leq t = \dim \mathcal{T}$. Let $S \in \mathbb{K}^{n\times s}$ and $T \in \mathbb{K}^{n\times t}$ have orthonormal columns and ranges given by $\mathcal{S}$ and $\mathcal{T}$, respectively. Following [18, Section I.5.2], we define the principal (also called canonical) angles between the subspaces $\mathcal{S}$ and $\mathcal{T}$, denoted $0 \leq \theta_1(\mathcal{S},\mathcal{T}) \leq \ldots \leq \theta_s(\mathcal{S},\mathcal{T}) \leq \frac{\pi}{2}$, determined by the identities $\cos(\theta_i(\mathcal{S},\mathcal{T})) = \sigma_i(S^*T)$, for $1 \leq i \leq s$. These can also be determined in terms of the identities

$$\sin(\theta_{s-i+1}(\mathcal{S},\mathcal{T})) = \sigma_i((I - TT^*)S) = \sigma_i((I - TT^*)SS^*) = \sigma_i((I - P_\mathcal{T})P_\mathcal{S}) \ , \quad 1 \leq i \leq s \qquad (1)$$

Following [18] we let $\theta(\mathcal{S},\mathcal{T}) = (\theta_1(\mathcal{S},\mathcal{T}), \ldots, \theta_s(\mathcal{S},\mathcal{T}))$ denote the vector of principal angles and $\Theta(\mathcal{S},\mathcal{T}) = \operatorname{diag}(\theta(\mathcal{S}, \mathcal{T}))$ denote the diagonal matrix with the principal angles in its main diagonal. In case $s = 0$ (i.e. $\mathcal{S} = \{0\}$) we define $\theta(\mathcal{S},\mathcal{T}) = \Theta(\mathcal{S},\mathcal{T}) = 0 \in \mathbb{R}$. It turns out that $\|\sin\Theta(\mathcal{S},\mathcal{T})\|_{2,F} = \|(I - P_\mathcal{T})P_\mathcal{S}\|_{2,F}$ are (scalar measures of the angular) distances between $\mathcal{S}$ and $\mathcal{T}$ (see [18, Section II.4.1]).

With the previous notation, if $\mathcal{S}' \subset \mathcal{S}$ and $\mathcal{T} \subset \mathcal{T}'$ are subspaces with $\dim \mathcal{S}' = s'$, then

$$\Theta(\mathcal{S},\mathcal{T}') \leq \Theta(\mathcal{S},\mathcal{T}) \quad \text{and} \quad \theta_{s'+1-i}(\mathcal{S}', \mathcal{T}) \leq \theta_{s+1-i}(\mathcal{S}, \mathcal{T}) \quad \text{for} \quad 1 \leq i \leq s' \qquad (2)$$

which follow from Eq. (1). Using Equation (2), the fact that $\sin x$ is an increasing in $[0, \frac{\pi}{2}]$ and the monotonicity of the u.i.n.'s, we get that

$$\big\|\Theta(\mathcal{S},\mathcal{T}')\big\|, \ \big\|\Theta(\mathcal{S}',\mathcal{T})\big\| \leq \|\Theta(\mathcal{S},\mathcal{T})\| \ , \quad \big\|\sin\Theta(\mathcal{S},\mathcal{T}')\big\|, \ \big\|\sin\Theta(\mathcal{S}',\mathcal{T})\big\| \leq \|\sin\Theta(\mathcal{S},\mathcal{T})\| \, . \qquad (3)$$

We will refer to Equations (2) and (3) as the *monotonicity of the principal angles.*

Remark 2.1. It is well-known that the following conditions are equivalent:

1. $\Theta(\mathcal{S}, \mathcal{T}) < \frac{\pi}{2} I$ (all the angles between $\mathcal{S}$ and $\mathcal{T}$ are strictly smaller the $\pi/2$);
2. $\mathcal{S} \cap \mathcal{T}^\perp = \{0\}$ ($\mathcal{S}$ has no non zero vectors that are orthogonal to $\mathcal{T}$);
3. $\operatorname{rk}(T^*S) = \operatorname{rk}(S) = s$ (that is, T^*S is full column rank or S^*T is full row rank).

We will use these conditions interchangeably throughout the text.

2.3 Dominant eigenspaces of Hermitian matrices

In what follows, we recall the class of dominant eigenspaces of arbitrary Hermitian matrices, and some of their basic properties. We consider the following

Notation 2.2. Let $\mathbb{K}=\mathbb{R}$ or $\mathbb{K}=\mathbb{C}$. In this work, we consider:

1. A self-adjoint matrix $A\in\mathbb{K}^{n\times n}$; we let $\lambda(A)=(\lambda_1,\ldots,\lambda_n)\in(\mathbb{R}^n)^\downarrow$ denote the vector of eigenvalues of A, counting multiplicities and arranged in non-increasing order i.e $\lambda_1\geq\ldots\geq\lambda_n$.

2. An eigendecomposition (or unitary diagonalization) of A given by $A=X\,\Lambda X^*$, where $\Lambda=\operatorname{diag}(\lambda(A))$ and $X\in\mathbb{K}^{n\times n}$ is a unitary matrix whose columns are denoted by $x_1,\ldots,x_n\in\mathbb{K}^n$, respectively. By construction, $Ax_i=\lambda_i\,x_i$, for $1\leq i\leq n$. Also, given $1\leq\ell\leq n$ we set

$$\mathcal{X}_\ell:=\overline{\{x_1,\ldots,x_\ell\}}\subset\mathbb{K}^n\,.$$

Definition 2.3. *Consider Notation 2.2 and let $\mathcal{S}\subset\mathbb{K}^n$ be such that $\dim\mathcal{S}=h$. We say that $\mathcal{S}$ is an h-dimensional dominant eigenspace of A if there exists an orthonormal basis $\{s_1,\ldots,s_h\}$ of $\mathcal{S}$ such that $As_i=\lambda_i\,s_i$, for $1\leq i\leq h$.*

Remark 2.4 (Dominant eigenspaces and eigendecompositions)**.** Consider Notation 2.2. Then, given $1\leq\ell\leq n$, we see that $\mathcal{X}_\ell$ is an ℓ-dimensional dominant eigenspace of A. Conversely, if $\mathcal{S}$ is an ℓ-dimensional dominant eigenspace of A then there exists an eigendecomposition $A=\tilde{X}\Lambda\tilde{X}^*$ such that $\tilde{X}$ is a unitary matrix with columns $\tilde{x}_1,\ldots,\tilde{x}_n$ and such that $\mathcal{S}=\overline{\{\tilde{x}_1,\ldots,\tilde{x}_\ell\}}$.

It is well known that if $\ell<n$ and $\lambda_\ell>\lambda_{\ell+1}$, then the dominant eigenspace for A of dimension ℓ is uniquely determined; hence, in this case, $\mathcal{X}_\ell$ does not depend on the particular choice of eigendecomposition $X\Lambda X^*$ being considered for A.

Remark 2.5 (Dominant eigenspaces and eigenvalue approximations)**.** Consider Notation 2.2. Notice that h-dimensional dominant eigenspaces $\mathcal{X}_h\subset\mathbb{K}^n$ of A are exactly those subspaces that allow us to get the exact h largest eigenvalues $\lambda_1\geq\ldots\geq\lambda_h$ of A by applying the Rayleigh-Ritz method, i.e. by considering the eigenvalues of the compression $P_{\mathcal{X}_h}AP_{\mathcal{X}_h}$. Hence, given an h-dimensional subspace $\tilde{\mathcal{X}}\subset\mathbb{K}^n$ that is close to $\mathcal{X}_h$, by continuity of the eigenvalues, we get that the eigenvalues of the compression $P_{\tilde{\mathcal{X}}}AP_{\tilde{\mathcal{X}}}$ should also be close to the exact eigenvalues $\lambda_1\geq\ldots\geq\lambda_h$ of A. We can even get quantitative bounds on the error of the approximation in terms of the principal angles between $\mathcal{X}_h$ and $\tilde{\mathcal{X}}$. See for example [4, Thm. 6].

Remark 2.6 (Dominant eigenspaces and low-rank approximations)**.** Consider Notation 2.2 and assume further that $1\leq h\leq n$ is such that $\lambda_h\geq 0$. Then, dominant eigenspaces allow us to construct optimal low-rank positive approximations of A as follows: if we let $A_h=P_{\mathcal{X}_h}AP_{\mathcal{X}_h}$ then $A_h\geq 0$ and it satisfies that

$$\|A-A_h\|\leq\|A-C\|\quad\text{for}\quad C\geq 0\ ,\ \operatorname{rk}(C)\leq h\,,\tag{4}$$

where $\|\cdot\|$ denotes an arbitrary u.i.n. Indeed, using Lidskii's inequality for self-adjoint matrices [2, Corollary III.4.2], we get that $((\lambda_i(A)-\lambda_i(C))_{i=1}^h,(\lambda_i(A))_{i=h+1}^n)=\lambda(A)-\lambda(C)\prec\lambda(A-C)\in\mathbb{R}^n$, where we used that $\operatorname{rk}(C)\leq h$. On the other hand, $\lambda(A-A_h)=(\lambda_{h+1},\ldots,\lambda_n,0,\ldots,0)^\downarrow\in\mathbb{R}^n$, by construction. Hence, we see that $|\lambda(A-A_h)|\prec_w|\lambda(A)-\lambda(C)|\prec_w|\lambda(A-C)|$, where we used that $f(x)=|x|$, for $x\in\mathbb{R}$, is a convex function and the properties of vector majorization in $\mathbb{R}^n$. The previous submajorization relation together with the properties of unitarily invariant norms imply Eq. (4). We call A_h the *truncated eigendecomposition* of A (corresponding to the index h).

3 Introducing Λ-admissible subspaces

In this section, we introduce the notion of Λ-admissible subspaces of a self-adjoint matrix A, which are suitable substitutes for dominant eigenspaces in the clustered eigenvalue setting.

Consider Notation 2.2. Hence, $A \in \mathbb{K}^{n\times n}$ is a self-adjoint matrix, with eigenvalue list $\lambda(A) = (\lambda_1, \dots, \lambda_n) \in (\mathbb{R}^n)^\downarrow$ and set $\lambda_0 := \infty$. For a target dimension $1 \le h \ll n$, we consider any fixed (enveloping) indices $0 \le j < h < k \le \mathrm{rk}(A)$ such that

$$\lambda_j > \lambda_{j+1} \quad \text{and} \quad \lambda_k > \lambda_{k+1}\,. \tag{5}$$

Consider an eigendecomposition $A = X\Lambda X^*$ of A; in this case, the unique dominant eigenspaces of A of dimensions j and k are given by $\mathcal{X}_j$ and $\mathcal{X}_k$, respectively; in case $j = 0$ we let $\mathcal{X}_0 = \{0\}$.

We introduce the following notion that will play a central role in our work.

Definition 3.1. *Let $A \in \mathbb{K}^{n\times n}$ be self-adjoint and let $0 \le j < h < k \le \mathrm{rk}(A)$ be such that they satisfy Eq.* (5)*. Given an h-dimensional subspace $\mathcal{S} \subset \mathbb{K}^n$, we say that $\mathcal{S}$ is Λ-admissible for A if*

$$\mathcal{X}_j \subset \mathcal{S} \subset \mathcal{X}_k\,.$$

Formally, the notion of Λ-admissible subspace depends on several parameters (e.g. j, h and k). Nevertheless, we will avoid this level of formalism and follow the (rather informal) description given in Definition 3.1 above throughout this section.

With the previous notation, in the case that $j = \max\{\ell : \lambda_\ell > \lambda_h\}$ and $k = \max\{\ell : \lambda_h = \lambda_\ell\}$ (i.e. $\lambda_j > \lambda_{j+1} = \ldots = \lambda_h = \ldots = \lambda_k > \lambda_{k+1}$) Definitions 2.3 and 3.1 coincide. That is, the classes of h-dimensional dominant eigenspaces and h-dimensional Λ-admissible subspaces of A coincide.

The motivation for the introduction of the class of Λ-admissible subspaces of a Hermitian matrix is related to the motivation for the introduction of left and right admissible subspaces of rectangular matrices in [8]. Yet, the present setting and the general approach to dealing with Λ-admissible subspaces differ from those considered in [8].

Remark 3.2 (When to - and why - consider Λ-admissible subspaces)**.** Let $A \in \mathbb{K}^{n\times n}$ be a self-adjoint matrix and assume that we are interested in approximating its h largest eigenvalues, $\lambda_1 \ge \ldots \ge \lambda_h$. Assume further that λ_h lies in a cluster of eigenvalues: ideally, there exist (enveloping) indices $0 \le j < h < k$ such that the eigen-gaps $\lambda_j > \lambda_{j+1}$ and $\lambda_k > \lambda_{k+1}$ are significant, and $\delta := \lambda_{j+1} - \lambda_k \approx 0$. In this case, $\lambda_{j+1} \ge \ldots \ge \lambda_k$ is a cluster of eigenvalues that includes λ_h. In these situations, we will see that h-dimensional Λ-admissible subspaces $\mathcal{S}$ of A (and then, by continuity, h-dimensional subspaces $\mathcal{T}$ that are close to these) allow us to compute good approximations of $\lambda_1 \ge \ldots \ge \lambda_h$ and (under further assumptions) nice low-rank approximations of A.

Furthermore, we will also show that some well-known iterative methods allow us to compute subspaces that are arbitrarily close to the class of Λ-admissible subspaces. This last fact indicates that there are advantages in considering the (typically infinite) class of Λ-admissible subspaces instead of the (generically uniquely determined) eigenspace $\mathcal{X}_h$; indeed, the sensitivity of computing $\mathcal{X}_h$ depends on $(\lambda_h - \lambda_{h+1})^{-1} \ge \delta^{-1}$ which is assumed to be large is the present setting, while the sensitivity of computing the class of Λ-admissible subspaces depends on $(\lambda_j - \lambda_{j+1})^{-1} + (\lambda_k - \lambda_{k+1})^{-1}$ which is assumed to be much smaller. We consider a detailed analysis of these facts in Section 4.4. We also present some numerical examples in Section 5 showing the advantages of considering Λ-admissible subspaces in this setting.

Remark 3.3 (Λ-admissible subspaces and eigenvalue approximation)**.** Consider Notation 2.2. It turns out that Λ-admissible subspaces of A allow us to compute approximations of its eigenvalues, e.g. using the Rayleigh-Ritz method. Indeed, consider the notation and assumptions from Remark 3.2, and further assume that $\mathcal{S} \subset \mathbb{K}^n$ is an h-dimensional Λ-admissible subspace of A. Recall that in this case $\mathcal{X}_j$ and $\mathcal{X}_k$ are the unique dominant eigenspaces of A with dimensions j and k, respectively. Then, the Ritz values of A corresponding to $\mathcal{S}$ satisfy that

$$\lambda_i(P_{\mathcal{S}} A P_{\mathcal{S}}) = \lambda_i(P_{\mathcal{S}} A P_{\mathcal{S}}|_{\mathcal{X}_j}) = \lambda_i \quad \text{for} \quad 1 \le i \le j \quad \text{and}$$

$$\lambda_{i+k-h} \le \lambda_i(P_{\mathcal{S}}AP_{\mathcal{S}}) = \lambda_{i-j}(P_{\mathcal{S}}AP_{\mathcal{S}}|_{\mathcal{X}_k \ominus \mathcal{X}_j}) \le \lambda_i \quad \text{for} \quad j+1 \le i \le h\,.$$

To prove the assertions above, notice that $P_{\mathcal{S}}AP_{\mathcal{S}} = P_{\mathcal{S}}(P_{\mathcal{X}_k}AP_{\mathcal{X}_k})P_{\mathcal{S}}$. Since the eigenvalues of the matrix $P_{\mathcal{X}_k}AP_{\mathcal{X}_k}\big|_{\mathcal{X}_k}$ restricted to the subspace $\mathcal{X}_k$ are $\lambda_1 \ge \ldots \ge \lambda_k$, the interlacing inequalities [2, Corollary III.1.5] imply that $\lambda_{i+k-h} = \lambda_{i+k-h}(P_{\mathcal{X}_k}AP_{\mathcal{X}_k}) \le \lambda_i(P_{\mathcal{S}}AP_{\mathcal{S}}) \le \lambda_i(P_{\mathcal{X}_k}AP_{\mathcal{X}_k}) = \lambda_i$, for $1 \le i \le h$, and the inequality for the first j eigenvalues becomes an equality since $\mathcal{X}_j \subseteq \mathcal{S}$. Finally,

$$0 \le \lambda_i - \lambda_i(P_{\mathcal{S}}AP_{\mathcal{S}}) \le \lambda_i - \lambda_{i+k-h} \le \delta \quad \text{for} \quad j+1 \le i \le h\,.$$

Hence, the smaller the spread of the eigenvalues in the cluster is, the more accurate the approximation of the first h eigenvalues of A by the Ritz values becomes.

Remark 3.4 (Λ-admissible subspaces and low-rank approximations)**.** It turns out that, under some additional hypotheses, Λ-admissible subspaces also induce low-rank approximations of a self-adjoint matrix that have several nice properties. To see this, consider the notation and setting from Remark 3.2, and assume that $\lambda_k \ge 0$. In this case, the truncated eigendecomposition $A_h = P_{\mathcal{X}_h}AP_{\mathcal{X}_h}$ is the optimal positive semi-definite low-rank approximation of A, since $\lambda_h \ge \lambda_k \ge 0$ (see Remark 2.6). Let $\mathcal{S} \subset \mathbb{K}^n$ be an h-dimensional Λ-admissible subspace of A. Then, the low-rank approximation $P_{\mathcal{S}}AP_{\mathcal{S}}$ obtained by compressing A onto $\mathcal{S}$ is positive semi-definite and it satisfies the following:

$$\|A - A_h\| \le \|A - P_{\mathcal{S}}AP_{\mathcal{S}}\| \le \|A - A_h\| + \|\mathrm{diag}(\underbrace{\delta, \ldots, \delta}_{k-j}, 0, \ldots, 0)\|\,, \tag{6}$$

where $\delta = \lambda_{j+1} - \lambda_k$ is the spread of the cluster and $\|\cdot\|$ is a unitarily invariant norm. In particular, $\|A - P_{\mathcal{S}}AP_{\mathcal{S}}\|_2 \le \|A - A_h\|_2 + \delta = \lambda_{h+1} + \delta$. The proof of Eq. (6) is developed in Section 6.

Remark 3.5 (Stability of Λ-admissible subspaces under the action of A)**.** Consider the notation and assumptions from Remark 3.2. It is well known that dominant eigenspaces $\mathcal{X}_h$ of a Hermitian matrix $A \in \mathbb{K}^{n\times n}$ are invariant under the action of A i.e. $P_{\mathcal{X}_h}A - AP_{\mathcal{X}_h} = 0$. On the other hand, Λ-admissible subspaces are (typically) not invariant under the action of A; nevertheless, we show that they are *approximately invariant*, in the sense that when $\mathcal{S} \subset \mathbb{K}^n$ is an h-dimensional Λ-admissible subspace of A, we have that $\|P_{\mathcal{S}}A - AP_{\mathcal{S}}\|_2 \le \lambda_{j+1} - \lambda_k = \delta$ (which we assume is small). To see this, let $A = X\Lambda X^*$ be an eigendecomposition of A and consider an auxiliary matrix $\tilde{A} = X\tilde{\Lambda}X^*$ with $\tilde{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_j, \tilde{\lambda}, \ldots, \tilde{\lambda}, \lambda_{k+1}, \ldots, \lambda_n)$ where $\tilde{\lambda} = \frac{\lambda_{j+1}+\lambda_k}{2}$. Notice that

$$|\lambda(A - \tilde{A})| = (0, \ldots, 0, |\lambda_{j+1} - \tilde{\lambda}|, \ldots, |\lambda_k - \tilde{\lambda}|, 0, \ldots, 0)^{\downarrow} \prec_w (\underbrace{\frac{\lambda_{j+1} - \lambda_k}{2}, \ldots, \frac{\lambda_{j+1} - \lambda_k}{2}}_{k-j}, 0, \ldots, 0)$$

which implies that $\|A - \tilde{A}\|_2 = \frac{\lambda_{j+1}-\lambda_k}{2}$. Moreover, by construction, $\mathcal{S}$ is $\tilde{A}$-invariant; so, in particular $P_{\mathcal{S}}A - AP_{\mathcal{S}} = P_{\mathcal{S}}A - P_{\mathcal{S}}\tilde{A} + \tilde{A}P_{\mathcal{S}} - AP_{\mathcal{S}} = P_{\mathcal{S}}(A - \tilde{A}) - (A - \tilde{A})P_{\mathcal{S}}$, which implies that

$$\|P_{\mathcal{S}}A - AP_{\mathcal{S}}\|_2 \le 2\|A - \tilde{A}\|_2 = \lambda_{j+1} - \lambda_k = \delta\,.$$

The previous estimate also shows that the norm of the residual satisfies

$$\|AP_{\mathcal{S}} - P_{\mathcal{S}}AP_{\mathcal{S}}\|_2 = \|(I - P_{\mathcal{S}})AP_{\mathcal{S}}\|_2 = \|(I - P_{\mathcal{S}})(AP_{\mathcal{S}} - P_{\mathcal{S}}A)\|_2 \le \delta\,.$$

Thus, the smaller the spread δ is, the more stable under the action of A the Λ-admissible subspace $\mathcal{S}$ is (as the norm $\|P_{\mathcal{S}}A - AP_{\mathcal{S}}\|_2$ becomes smaller). In case that $\delta = 0$, it is actually invariant.

In the perfect cluster setting $\delta = \lambda_{j+1} - \lambda_k = 0$ we have noted that the class of Λ-admissible subspaces coincides with the class of dominant eigenspaces of A (see the comments after Definition 3.1). In this special case, Remarks 3.3, 3.4 and 3.5 recover some known facts about the dominant eigenspaces of Hermitian matrices. Indeed, compare Remarks 3.3 and 2.5 or Remarks 3.4 and 2.6.

These remarks show that (under some natural assumptions) Λ-admissible subspaces of a matrix A are nice subspaces to tackle simultaneously eigenvalue approximations and low-rank approximations of A, in the context of clustered eigenvalues; indeed, the estimates for the approximation of eigenvalues and for the approximation error $\|A - A_\ell\|$ imply that Λ-admissible subspaces induce low-rank approximations that fulfill the paradigm proposed in [10, Section 2.2] (and adopted in recent works [5, 17]) for qualitatively good low-rank approximations. Continuity arguments imply that low-rank approximations $P_{\mathcal{T}} A P_{\mathcal{T}}$ induced by h-dimensional subspaces $\mathcal{T}$ that are close to some Λ-admissible subspace (in the terms discussed in Remark 3.4) are also (qualitatively) nice. To get some first quantitative estimates, consider the notation in Remark 3.2. By [9, Theorem 3.3] we get

$$\max\{|\lambda_i(P_{\mathcal{S}} A P_{\mathcal{S}}) - \lambda_i(P_{\mathcal{T}} A P_{\mathcal{T}})| \ : \ 1 \le i \le h\} \le \tan(\theta_{\max}(\mathcal{S},\mathcal{T}))(2\,\delta + \sin(\theta_{\max}(\mathcal{S},\mathcal{T}))\,\|A\|_2)\,,$$

where we used that $\|P_{\mathcal{T}}(I - P_{\mathcal{S}})AP_{\mathcal{S}}\|_2 \le \sin(\theta_{\max}(\mathcal{S},\mathcal{T}))\,\|AP_{\mathcal{S}} - P_{\mathcal{S}}AP_{\mathcal{S}}\|_2 \le \sin(\theta_{\max}(\mathcal{S},\mathcal{T}))\,\delta$ (by Remark 3.5), and $\|P_{\mathcal{S}}(I - P_{\mathcal{T}})AP_{\mathcal{T}}\| \le \sin(\theta_{\max}(\mathcal{S},\mathcal{T}))\,\|(I - P_{\mathcal{T}})A\|_2$ together with $\|(I - P_{\mathcal{T}})A\|_2 \le \|(I - P_{\mathcal{S}})A\|_2 + \sin(\theta_{\max}(\mathcal{S},\mathcal{T}))\|A\|_2$ and $\|(I - P_{\mathcal{S}})A\|_2 = \|(I - P_{\mathcal{S}})(A - P_{\mathcal{S}}AP_{\mathcal{S}})\|_2 \le \delta$. Hence, by Remark 3.3 we get that

$$0 \le \lambda_i(A) - \lambda_i(P_{\mathcal{T}} A P_{\mathcal{T}}) \le \tan(\theta_{\max}(\mathcal{S},\mathcal{T}))(2\,\delta + \sin(\theta_{\max}(\mathcal{S},\mathcal{T}))\,\|A\|_2) \quad \text{for} \quad 1 \le i \le j;$$

$$0 \le \lambda_i(A) - \lambda_i(P_{\mathcal{T}} A P_{\mathcal{T}}) \le \delta + \tan(\theta_{\max}(\mathcal{S},\mathcal{T}))(2\,\delta + \sin(\theta_{\max}(\mathcal{S},\mathcal{T}))\,\|A\|_2) \quad \text{for} \quad j+1 \le i \le h\,.$$

Assume further that $\lambda_k \ge 0$; it is straightforward to check that $\|A - P_{\mathcal{T}}AP_{\mathcal{T}}\|_2 \le \|A - P_{\mathcal{S}}AP_{\mathcal{S}}\|_2 + \|P_{\mathcal{S}}AP_{\mathcal{S}} - P_{\mathcal{T}}AP_{\mathcal{T}}\|_2$ and $\|P_{\mathcal{S}}AP_{\mathcal{S}} - P_{\mathcal{T}}AP_{\mathcal{T}}\|_2 \le 2\sin(\theta_{\max}(\mathcal{S},\mathcal{T}))\,\|A\|_2$; thus, by Remark 3.4,

$$\|A - P_{\mathcal{T}}AP_{\mathcal{T}}\|_2 \le \|A - A_h\|_2 + \delta + 2\sin(\theta_{\max}(\mathcal{S},\mathcal{T}))\,\|A\|_2\,,$$

where $A_h = P_{\mathcal{X}_h} A P_{\mathcal{X}_h}$ is the truncated eigendecomposition (see Remark 2.6). We will consider a more detailed quantitative analysis of the quality of the low-rank approximation $P_{\mathcal{T}}AP_{\mathcal{T}}$ elsewhere.

4 Main results

In this Section, we present our main results. In Section 4.1 we obtain an upper bound for the distance between a generic h-dimensional subspace $\mathcal{T} \subset \mathbb{K}^n$ and the class of h-dimensional Λ-admissible subspaces of A. This result plays a key role in our work. In Section 4.2, we show that some of the well-known iterative algorithms for computing approximations of dominant eigenspaces also produce subspaces that become close to Λ-admissible subspaces; numerical examples show that the speed at which these iterative subspaces become close to Λ-admissible subspaces is much faster than the speed at which these iterative subspaces become close to dominant eigenspaces when λ_h lies in a cluster of eigenvalues. In Section 4.3 we obtain upper bounds for the distance between subspaces constructed using the Rayleigh-Ritz method and the class of Λ-admissible subspaces. Finally, in Section 4.4, we obtain upper bounds for the sensitivity of the computation of Λ-admissible subspaces. The proofs of these results are presented in Section 6.

Next, we include some of the notation used throughout this section.

Notation 4.1. Let $\mathbb{K} = \mathbb{R}$ or $\mathbb{K} = \mathbb{C}$. We consider:

1. A self-adjoint or Hermitian matrix $A \in \mathbb{K}^{n\times n}$ with an eigendecomposition given by $A = X\,\Lambda X^*$. The eigenvalues of A are given by $\lambda(A) = (\lambda_1, \ldots, \lambda_n) \in (\mathbb{R}^n)^\downarrow$, counting multiplicities and arranged non-increasingly and we set $\lambda_0 := \infty$. The columns of X are denoted by $x_1, \ldots, x_n \in \mathbb{K}^n$, respectively. We further consider $\mathcal{X}_l = \overline{\{x_1, \ldots, x_l\}}$, for $1 \le l \le n$ and $\mathcal{X}_0 = \{0\}$.

2. For a target dimension $1 \le h \ll n$, we consider (enveloping) indices $0 \le j < h < k \le \operatorname{rank}(A)$ such that $\lambda_j > \lambda_{j+1}$ and $\lambda_k > \lambda_{k+1}$. Notice that $\mathcal{X}_j$ and $\mathcal{X}_k$ do not depend on the choice of X in the present setting. We use the indices j and k for the Λ-admissible subspaces of A and set

$$\begin{aligned}\Lambda\text{-adm}_h(A) &:= \{\,\mathcal{S} :\ \mathcal{S} \text{ is an } h\text{-dimensional } \Lambda\text{-admissible space of } A\ \} \\ &= \{\,\mathcal{S} \in \mathcal{G}_{n,h} : \mathcal{X}_j \subseteq \mathcal{S} \subseteq \mathcal{X}_k\}\,.\end{aligned} \tag{7}$$

4.1 A bound for the distance to the class of Λ-admissible subspaces

Consider Notation 4.1. If $\mathcal{T} \subset \mathbb{K}^n$ is an h-dimensional subspace such that $\|\sin\Theta(\mathcal{S},\mathcal{T})\|$ is small for *some* h-dimensional Λ-admissible space of A then Remark 3.3 and continuity arguments show that $\mathcal{T}$ can be used to derive approximate eigenvalues that are close to the exact eigenvalues $\lambda_1, \ldots, \lambda_h$ of A. This fact induces us to consider upper bounds for the distance

$$d(\Lambda\text{-adm}_h(A),\, \mathcal{T}) = \inf\{\, \|\sin\Theta(\mathcal{S},\mathcal{T})\| :\ \mathcal{S} \in \Lambda\text{-adm}_h(A)\}\,, \tag{8}$$

where $\|\cdot\|$ is a u.i.n. Notice that since $h < k$, obtaining upper bounds for $d(\Lambda\text{-adm}_h(A),\, \mathcal{T})$ is a non-trivial problem, since we can find $\mathcal{S},\, \mathcal{S}' \in \Lambda\text{-adm}_h(A)$ such that $\theta_h(\mathcal{S},\mathcal{S}') = \pi/2$.

We point out that $\mathcal{S} \in \Lambda\text{-adm}_h(A)$ if and only if $\mathcal{S} = \mathcal{X}_j \oplus \mathcal{D}$ where $\mathcal{D} \subseteq \mathcal{X}_k \ominus \mathcal{X}_j := \mathcal{X}_k \cap \mathcal{X}_j^\perp$ is such that $\dim \mathcal{D} = h - j$. In particular, $\Lambda\text{-adm}_h(A)$ is a compact subset of $\mathcal{G}_{n,h}$ (endowed with the topology associated with the metric $\|\sin\Theta(\mathcal{T}_1,\, \mathcal{T}_2)\|$ for $\mathcal{T}_1,\, \mathcal{T}_2 \in \mathcal{G}_{n,h}$ and any u.i.n.) and thus the infimum in Eq. (8) is actually a minimum.

The next result describes a useful upper bound for the distance in Eq. (8) in a generic case and plays a key role in the rest of our work (see Section 6).

Theorem 4.2. *Consider Notation 4.1. Let* $\mathcal{T} \subset \mathbb{K}^n$ *be such that* $\Theta(\mathcal{X}_k,\mathcal{T}) < \frac{\pi}{2}\, I_{\min\{t,k\}}$. *Then, for every unitarily invariant norm* $\|\cdot\|$ *we have that*

$$d(\Lambda\text{-}adm_h(A),\, \mathcal{T}) \le \|\sin\Theta(\mathcal{X}_j,\mathcal{T})\| + \|\sin\Theta(\mathcal{X}_k,\mathcal{T})\|\,.$$

Proof. See Section 6.1 ☐

Consider the notation in Theorem 4.2. This result will allow us to address the problem of finding upper bounds for the distance to the class $\Lambda\text{-adm}_h(A)$ by splitting the problem and searching for upper bounds for the distance to the (under the current hypothesis) uniquely determined subspaces $\mathcal{X}_j$ and $\mathcal{X}_k$. We point out that the problem of finding an upper bound for $\|\sin\Theta(\mathcal{X}_j,\mathcal{T})\|$, e.g. for subspaces $\mathcal{T}$ obtained from iterative methods, has been widely studied [1, 5, 17, 21] (recall that $\dim(\mathcal{T}) \ge h \ge j$) and thus we can employ ideas and tools previously developed for this case. On the other hand, the study of $\|\sin\Theta(\mathcal{X}_k,\mathcal{T})\|$ is more delicate, since depending on the situation, we may find ourselves in the case where $\dim(\mathcal{T}) \ge k$ (where once again we can refer to the literature) or in the case that $\dim(\mathcal{T}) < k$ (where the mentioned works do not apply).

Remark 4.3. Consider the notation in Theorem 4.2 and assume further that $\dim(\mathcal{T}) = h$. Given $\mathcal{S}' \in \Lambda\text{-adm}_h(A)$, the monotonicity of the principal angles implies that for every u.i.n. $\|\cdot\|$ we have

$$\|\sin\Theta(\mathcal{X}_j,\mathcal{T})\|,\ \|\sin\Theta(\mathcal{X}_k,\mathcal{T})\| \le \left\|\sin\Theta(\mathcal{S}',\mathcal{T})\right\|\,.$$

The previous fact together with Theorem 4.2 show that $d(\Lambda\text{-adm}_h(A),\, \mathcal{T})$ is small (i.e., $\mathcal{T}$ is close to some element of $\Lambda\text{-adm}_h(A)$) if and only if $\mathcal{T}$ is close to both $\mathcal{X}_j$ and $\mathcal{X}_k$. Moreover, if $\mathcal{T} \subseteq \mathcal{X}_k$, then $\|\sin\Theta(\mathcal{X}_j,\mathcal{T})\| = d(\Lambda\text{-adm}_h(A),\, \mathcal{T})$ and if $\mathcal{X}_j \subseteq \mathcal{T}$, then $\|\sin\Theta(\mathcal{X}_k,\mathcal{T})\| = d(\Lambda\text{-adm}_h(A),\, \mathcal{T})$.

4.2 Computable approximations of Λ-admissible subspaces

As pointed out at the beginning of Section 4.1, if $\mathcal{T} \subset \mathbb{K}^n$ is an h-dimensional subspace such that $\|\sin\Theta(\mathcal{S},\mathcal{T})\|$ is small for some $\mathcal{S} \in \Lambda\text{-adm}_h(A)$, then $\mathcal{T}$ can be used to derive approximate eigenvalues that are close to the exact eigenvalues $\lambda_1,\ldots,\lambda_h$ of A. Hence, it is important to obtain algorithms that can compute (or construct) in an effective way such subspaces $\mathcal{T}$: the following result is a first step towards the construction of such subspaces (also see Remark 4.5 below).

Theorem 4.4. *Consider Notation 4.1. Let* $\mathcal{W} \subset \mathbb{K}^n$ *be such that* $\dim \mathcal{W} = r$, *with* $h \le r < k$. *Let* $0 \le p_1, p_2$ *be such that* $p_1 + p_2 \le r - h$ *and assume that*

$$\mathcal{W}^\perp \cap \overline{\{x_1,\ldots,x_{j+p_1},x_{k+1},\ldots,x_{k+p_2}\}} = \{0\} \quad \textit{and} \quad \mathcal{W} \cap \overline{\{x_1,\ldots,x_k\}}^\perp = \{0\}\,. \tag{9}$$

Then, there exists $\mathcal{H}_p \subset \mathcal{W}$ *such that* $\dim \mathcal{H}_p = r - (p_1+p_2) \ge h$, $\mathcal{H}_p^\perp \cap \mathcal{X}_j = \{0\}$, $\mathcal{H}_p \cap \mathcal{X}_k^\perp = \{0\}$ *such that for every polynomial* $\phi \in \mathbb{K}[x]$ *with* $\phi(\Lambda_k)$ *invertible, and every u.i.n.* $\|\cdot\|$

$$\begin{aligned} d(\Lambda\text{-}adm_h(A),\phi(A)(\mathcal{W})) \quad &\le \quad \|\phi(\Lambda_j)^{-1}\|_2\ \|\phi(\Lambda_{j+p_1,\perp})\|_2\ \|\tan(\Theta(\mathcal{X}_j,\,\mathcal{H}_p))\| + \\ &\qquad \|\phi(\Lambda_k)^{-1}\|_2\ \|\phi(\Lambda_{k+p_2,\perp})\|_2\ \|\tan(\Theta(\mathcal{X}_k,\,\mathcal{H}_p))\|\,. \end{aligned}$$

If $j=0$ *(respectively if* $k = \operatorname{rk}(A)$ *and* $\phi(0)=0$*) then the first (respectively the second) term in the right-hand side of the inequality above should be omitted.*

Proof. See Section 6.2. ☐

Consider the notation of Theorem 4.4. Notice that the assumptions in Eq. (9) are generic conditions. Also, the conditions $\mathcal{H}_p^\perp \cap \mathcal{X}_j = \{0\}$, $\mathcal{H}_p \cap \mathcal{X}_k^\perp = \{0\}$ imply that $\tan(\Theta(\mathcal{X}_j,\,\mathcal{H}_p))$ and $\tan(\Theta(\mathcal{X}_k,\,\mathcal{H}_p))$ are well defined. We also point out that since $j \le j+p_1 \le r+j-h < k$ we have $\|\Lambda_{j+p_1,\perp}\|_2 = \lambda_{j+p_1+1} \in [\lambda_k,\,\lambda_{j+1}]$. Thus, in the context described in Remark 3.2, we have that $\lambda_{j+p_1} \approx \lambda_j$. One should keep this in mind while reading the following remarks.

Remark 4.5. Consider the notation from Theorem 4.4 and assume further that A is a positive semi-definite matrix. We let $\phi(x) = x^q$ for some $q \ge 1$. In this case $\phi(A) = A^q$; hence, if we let $\mathcal{W} \subset \mathbb{K}^n$ then $\phi(A)(\mathcal{W}) = A^q(\mathcal{W})$ is the q-th iteration of the Subspace Iteration Method (SIM).

In this case, we have $\|\Lambda_j^{-1}\|_2 = \lambda_j^{-1}$, $\|\Lambda_{j+p_1,\perp}\|_2 = \lambda_{j+p_1+1}$, $\|\Lambda_k^{-1}\|_2 = \lambda_k^{-1}$ and $\|\Lambda_{k+p_2,\perp}\|_2 = \lambda_{k+p_2+1}$. By Theorem 4.4, there exist $\mathcal{H}_p \subset \mathcal{V}$ and $\mathcal{S} \in \Lambda\text{-adm}_h(A)$ such that

$$\|\sin\Theta(\mathcal{S}\,,\,A^q(\mathcal{W}))\|_2 \le \left(\frac{\lambda_{j+p_1+1}}{\lambda_j}\right)^q \|\tan\Theta(\mathcal{X}_j,\,\mathcal{H}_p)\|_2 + \left(\frac{\lambda_{k+p_2+1}}{\lambda_k}\right)^q \|\tan\Theta(\mathcal{X}_k,\,\mathcal{W})\|_2\,, \tag{10}$$

where we have chosen $\|\cdot\| = \|\cdot\|_2$ to be the spectral norm. We remark that our analysis takes advantage of the oversampling parameter $r-h$; indeed, notice that the decay of the upper bound in Eq. (10) is related to the quotients $\frac{\lambda_{j+p_1+1}}{\lambda_j} < 1$ and $\frac{\lambda_{k+p_2+1}}{\lambda_k} < 1$, making use of the decay of the eigenvalue list, rather than merely of the eigen-gaps $\lambda_j > \lambda_{j+1}$ and $\lambda_k > \lambda_{k+1}$ corresponding to consecutive indices. Notice that the upper bound in Eq. (10) becomes arbitrarily small for large enough $q \ge 0$. On the other hand, we remark the fact that the subspace $\mathcal{S}$ depends on q; hence, the previous analysis is not a convergence analysis but rather a proximity analysis between the subspaces $A^q(\mathcal{W})$ and the class $\Lambda\text{-adm}_h(A)$ of h-dimensional Λ-admissible subspaces of A.

Remark 4.6. Consider the notation from Theorem 4.4. For $1 \le q$ we can consider the Krylov subspace $K_q(A,\mathcal{W}) := \mathcal{W} + A(\mathcal{W}) + \ldots + A^q(\mathcal{W}) \subset \mathbb{K}^n$. In this case, if $\phi(x) \in \mathbb{K}[x]$ is any polynomial with degree at most q then $\phi(A)(\mathcal{W}) \subset K_q(A, \mathcal{W})$. If we further assume that $\phi(\Lambda_k)$ is invertible then, by the monotonicity of principal angles and Theorem 4.4 we get that

$$\begin{aligned} d(\Lambda\text{-adm}_h(A), K_q(A, \mathcal{W})) \ \leq \ & \|\phi(\Lambda_j)^{-1}\|_2 \ \|\phi(\Lambda_{j+p_1,\perp})\|_2 \ \|\tan(\Theta(\mathcal{X}_j,\, \mathcal{H}_p))\| \, + \\ & \|\phi(\Lambda_k)^{-1}\|_2 \ \|\phi(\Lambda_{k+p_2,\perp})\|_2 \ \|\tan(\Theta(\mathcal{X}_k,\, \mathcal{W}))\| \, . \end{aligned}$$

It is well known that Chebyshev's polynomials play a key role in constructing convenient polynomials $\phi(x)$ that induce relevant upper bounds for the angles above [1, 5, 10].

As a final comment, we mention the recent works [22, 23], where the authors obtain cluster robust bounds for Ritz vectors and Ritz values of self-adjoint matrices. In these works the authors consider an abstract block iteration, similar to the transformation $\mathcal{W} \mapsto \phi(A)(\mathcal{W})$ considered in Theorem 4.4 above. Nevertheless, the general setting in [22, 23] differs from ours; indeed, the authors assume (using the notation from Theorem 4.4) that $\{\lambda_1, \ldots, \lambda_r\}$ and $\{\lambda_{r+1}, \ldots, \lambda_n\}$ are disjoint, and the accuracy of Ritz vectors is measured in terms of the distance to the dominant eigenspace $\mathcal{X}_r$. On the other hand, our setting allows that $\lambda_r = \lambda_{r+1}$ (since $h \leq r < k$ and we can have that $\lambda_h = \ldots = \lambda_k$, for example) and the accuracy of Ritz vectors is measured in terms of the distance to the class of Λ-admissible subspaces.

4.3 Rayleigh-Ritz method and Λ-admissible subspaces

We use Notation 4.1 and consider the problem of bounding the distance between subspaces constructed using the Rayleigh-Ritz method and the class of h-dimensional Λ-admissible subspaces of A; our analysis follows the approach developed in [11, 12]. Indeed, consider a trial subspace $\mathcal{Q} \subset \mathbb{K}^n$ with $\dim \mathcal{Q} = r$ such that $h \leq r \ll n$. Following the Rayleigh-Ritz method, we construct the subspace $\widehat{X}_1 \subset \mathbb{K}^n$ spanned by the eigenvectors $\{\widehat{x}_1, \ldots, \widehat{x}_h\}$ corresponding to the h largest eigenvalues $\widehat{\lambda}_1 \geq \ldots \geq \widehat{\lambda}_h$ of the hermitian matrix $P_{\mathcal{Q}}\, A\, P_{\mathcal{Q}}$. In this case, we are interested in getting upper bounds for the distance $d(\Lambda\text{-adm}_h(A), \widehat{X}_1)$, using (the norm of) the residual $(I - P_{\mathcal{Q}})AP_{\mathcal{Q}}$ and information about the separation between eigenvalues of A and eigenvalues of $P_{\mathcal{Q}}\, A\, P_{\mathcal{Q}}$; in case the distance $d(\Lambda\text{-adm}_h(A), \widehat{X}_1)$ is small, the remarks in Section 3 imply that $P_{\widehat{X}_1}\, A\, P_{\widehat{X}_1}$ provide good eigenvalue approximations $\widehat{\lambda}_1 \geq \ldots \geq \widehat{\lambda}_h$ of the h-largest eigenvalues of A. Incidentally, our approach also allows us to get upper bounds for $d(\Lambda\text{-adm}_h(A), \mathcal{Q})$ in similar terms; this last quantity can be considered as a measure of the *quality* of the trial subspace $\mathcal{Q}$, with respect to the class $\Lambda\text{-adm}_h(A)$ (irrespectively of any method), that is of independent interest.

Next, we present a pseudo-code for the Rayleigh-Ritz method and introduce the notation used in the statement of the main result of this subsection.

Algorithm 4.1 (Rayleigh-Ritz method)

```
Require: $A \in \mathbb{K}^{n\times n}$ self-adjoint, $Q \in \mathbb{K}^{n\times r}$, $R(Q) = \mathcal{Q}$, $Q^*Q = I_r$, target rank $1 \leq h \leq r$.
1: Compute the compression $Q^*AQ \in \mathbb{K}^{r\times r}$ and the eigen-decomposition $Q^*AQ = \Omega\widehat{\Lambda}\Omega^*$ with
   $\Omega \in \mathcal{U}(r)$, $\mathrm{d}(\widehat{\Lambda}) = (\widehat{\lambda}_1, \ldots, \widehat{\lambda}_r) \in \mathbb{R}^r$ and $\widehat{\lambda}_1 \geq \ldots \geq \widehat{\lambda}_r$ .
2: Let
   $$\Omega = [\Omega_1 \quad \Omega_2] \quad \text{and} \quad \widehat{\Lambda} = \begin{bmatrix} \widehat{\Lambda}_1 & \\ & \widehat{\Lambda}_2 \end{bmatrix} \quad \text{with} \quad \Omega_1 \in \mathbb{K}^{r\times h} \quad \text{and} \quad \widehat{\Lambda}_1 \in \mathbb{K}^{h\times h}.$$
3: Return: $\widehat{X}_1 := Q\Omega_1 \in \mathbb{K}^{n\times h}$ and $\widehat{\Lambda}_1$.
```

We now describe a convenient inner structure associated with the pair (A, Q) as above. We will need this inner structure in the statement of Theorem 4.8 below.

Notation 4.7. Consider Notation 4.1. Let $Q \in \mathbb{K}^{n\times r}$ have orthonormal columns, let $\mathcal{Q} = R(Q)$ and set a target rank $1 \leq h \leq r$. Apply Algorithm 4.1 and let Ω and $\widehat{\Lambda}$ be as in Step 2. Furthermore, consider the partitions as in Step 3. We set

$$\widehat{X} := Q\Omega = [Q\Omega_1 \quad Q\Omega_2] = [\widehat{X}_1 \quad \widehat{X}_2] = [\widehat{x}_1, \ldots, \widehat{x}_r]\,,$$

where $\widehat{X}_1 = Q\Omega_1 = [\widehat{x}_1, \ldots, \widehat{x}_h] \ \in \mathbb{K}^{n\times h}$ and $\widehat{X}_2 = Q\Omega_2 = [\widehat{x}_{h+1}, \ldots, \widehat{x}_r] \in \mathbb{K}^{n\times(r-h)}$. Notice that $\operatorname{Ran}(\widehat{X}) = \mathcal{Q}$, since Ω is unitary, and $\operatorname{Ran}(\widehat{X}_1)$ is an h-dimensional subspace of $\mathcal{Q}$. In this case, $(\widehat{\Lambda}, \widehat{X}) = ((\widehat{\lambda}_i, \widehat{x}_i))_{i=1}^r$ are Ritz pairs for $1 \leq i \leq r$.

We also consider $\widehat{X}_3 = [\widehat{x}_{r+1}, \ldots, \widehat{x}_n] \in \mathbb{K}^{n\times(n-r)}$ such that $[\widehat{X}_1 \quad \widehat{X}_2 \quad \widehat{X}_3] \in \mathcal{U}(n)$. Taking into account the construction of $\widehat{X}_i$, $i = 1, 2, 3$, we get that

$$\widetilde{A} := [\widehat{X}_1 \quad \widehat{X}_2 \quad \widehat{X}_3]^*\, A\, [\widehat{X}_1 \quad \widehat{X}_2 \quad \widehat{X}_3] = \begin{bmatrix} \widehat{\Lambda}_1 & 0 & R_1^* \\ 0 & \widehat{\Lambda}_2 & R_2^* \\ R_1 & R_2 & A_3 \end{bmatrix} \quad \text{with } \widehat{\Lambda}_1 = \begin{bmatrix} \widehat{\Lambda}_{11} & 0 \\ 0 & \widehat{\Lambda}_{12} \end{bmatrix}, \tag{11}$$

where $\widehat{\Lambda}_{11} \in \mathbb{K}^{j\times j}$, $\widehat{\Lambda}_{12} \in \mathbb{K}^{(h-j)\times(h-j)}$; similarly,

$$R_1 = \widehat{X}_3^* A \widehat{X}_1 = [R_{11}\ R_{12}] \quad , \quad R_2 = \widehat{X}_3^* A \widehat{X}_2 \quad , \quad R := [R_1 \ \ R_2] \quad \text{and} \quad A_3 = \widehat{X}_3^* A \widehat{X}_3$$

where $R_{11} \in \mathbb{K}^{(n-r)\times j}$ and $R_{12} \in \mathbb{K}^{(n-r)\times(r-j)}$. Finally, we define the gaps:

$$\widetilde{\mathrm{Gap}} := \min |\lambda(\widehat{\Lambda}_{11}) - (\lambda_{j+1}, \ldots, \lambda_n)| \ \ , \ \ \widehat{\mathrm{Gap}}(l) := \min |\lambda(\widehat{\Lambda}_l) - (\lambda_{k+1}, \ldots, \lambda_n)| \ , \ l = 1,\, 2\,,$$

and $\mathrm{Gap}_i := \min |(\lambda_1, \ldots, \lambda_i) - \lambda(A_3)|$ for $i = j,\, k$.

Theorem 4.8. *Consider Notation 4.7 and let Λ-$adm_\ell(A)$ denote the class of ℓ-dimensional Λ-admissible subspaces of A. Then,*

$$d(\Lambda\text{-}adm_h(A), R(\widehat{X}_1)) \leq \frac{\|R_{11}\|}{\widetilde{Gap}} + \frac{\|R_1\|}{\widehat{\mathrm{Gap}}(1)}\,. \tag{12}$$

Regarding the subspace $R(\widehat{X}) = \mathcal{Q}$: If $k \leq r$ and $\dim(P_{R(\widehat{X})}(\mathcal{X}_k)) \geq h$ (generic case) then

$$d(\Lambda\text{-}adm_h(A), R(\widehat{X})) \leq \|R\|\,(\frac{1}{\mathrm{Gap}_j} + \frac{1}{\mathrm{Gap}_k})\,. \tag{13}$$

If $r < k$ then

$$d(\Lambda\text{-}adm_r(A), R(\widehat{X})) \leq \frac{\|R\|}{\mathrm{Gap}_j} + \frac{\|R_1\|}{\widehat{\mathrm{Gap}}(1)} + \frac{\|R_2\|}{\widehat{\mathrm{Gap}}(2)}\,, \tag{14}$$

where the third term in Eq. (14) *should be omitted if $r = h$.*

Proof. See Section 6.3. □

Remark 4.9. Consider Notation 4.7 and assume that the gaps $\lambda_j - \lambda_{j+1}$ and $\lambda_k - \lambda_{k+1}$ are significant, even with respect to the spread $\lambda_{j+1} - \lambda_k$ of the cluster (i.e. $\delta\,(\lambda_\ell - \lambda_{\ell+1})^{-1} \ll 1$ for $\ell = j,\, k$). Below we elaborate on the fact that if the trial subspace $R(\widehat{X})$ is sufficiently nice for eigenvalue approximation, then all the gaps appearing in the inequalities in Theorem 4.8 would be significant, even if $\lambda_h - \lambda_{h+1} \approx 0$. Indeed:

1. For the gaps in Eq. (12), if $\lambda_i \approx \widehat{\lambda}_i$ for $1 \leq i \leq h$, we would get $\widetilde{\mathrm{Gap}} \approx \lambda_j - \lambda_{j+1}$ and $\widehat{\mathrm{Gap}}(1) \approx \lambda_h - \lambda_{k+1} \geq \lambda_k - \lambda_{k+1}$.

2. For the gaps in Eq. (13), where we are assuming that $k \leq r$, if $\lambda_i \approx \widehat{\lambda}_i$ for $1 \leq i \leq r$ and $\lambda_{\max}(A_3) \leq \lambda_{k+1}$, we would get $\mathrm{Gap}_j \approx (\lambda_j - \lambda_{j+1}) + \delta + (\lambda_k - \lambda_{k+1})$ and $\mathrm{Gap}_k \approx \lambda_k - \lambda_{k+1}$.

3. For the gaps in Eq. (14), where we are assuming that $k > r$, if $\lambda_i \approx \widehat{\lambda}_i$ for $1 \leq i \leq r$ and $\lambda_{\max}(A_3) \leq \lambda_{r+1}$, we would get $\mathrm{Gap}_j \approx \lambda_j - \lambda_{r+1} \geq \lambda_j - \lambda_{j+1}$, $\widehat{\mathrm{Gap}}(1) \approx \lambda_h - \lambda_{k+1} \geq \lambda_k - \lambda_{k+1}$ and $\widehat{\mathrm{Gap}}(2) \approx \lambda_r - \lambda_{k+1} \geq \lambda_k - \lambda_{k+1}$.

These facts are to be compared with the *good gap* and *bad gap* appearing in the upper bound obtained in [12] (see the discussion in [12, Section 2]).

We remark that if the upper bound in Eq. (12) is zero, then $R(\widehat{X}_1) \in \Lambda\text{-adm}_h(A)$, even when $\|R\| > 0$. On the other hand, by [12, Theorem 5.1] we have that

$$\sin \theta_{\max}(\mathcal{X}_h, R(\widehat{X}_1)) \leq \frac{\|R\|_2}{\mathrm{Gap}} \sqrt{1 + \frac{\|R_2\|_2^2}{\mathrm{gap}^2}}\,,$$

where $\mathrm{Gap} = \min |\lambda(\Lambda_1) - \lambda(A_3)| > 0$ and $\mathrm{gap} = \min |\lambda(\Lambda_1) - \lambda(\widehat{\Lambda}_2)| > 0$. For this upper bound to be zero, it is required that $R = 0$ (i.e., the range $R(\widehat{X})$ is an eigenspace of A).

There are situations in which $\sin \theta_{\max}(\mathcal{X}_h, R(\widehat{X}_1))$ is large and yet $R(\widehat{X}_1)$ is close to some (generic) $\mathcal{S} \in \Lambda\text{-adm}_h(A)$ (see Figures 1-3 in Section 5). In these cases, for the operator norm, we expect $\|R_{11}\|_2 \approx 0$ and $\|R_1\|_2$ to be close to $\|AP_{\mathcal{S}} - P_{\mathcal{S}}AP_{\mathcal{S}}\|_2 \approx \delta$ for this generic $\mathcal{S}$ (see Remark 3.5). Hence, the upper bound in Eq. (12) would become approximately $(\lambda_k - \lambda_{k+1})^{-1}\,\delta$. Since the spread $\delta > 0$ is fixed, this upper bound would not become arbitrarily small (i.e., would not reflect the proximity between $R(\widehat{X}_1)$ and $\mathcal{S}$). Nevertheless, this upper bound is informative in our present setting (see the numerical examples in Section 5). In contrast, bounds as in [12] cannot be informative when $\sin \theta_{\max}(\mathcal{X}_h, R(\widehat{X}_1))$ is large.

On the other hand, if $\sin \theta_{\max}(\mathcal{X}_h, R(\widehat{X}_1))$ is small, the upper bound provided by Eq. (12) avoids the influence of the gap $\lambda_h - \lambda_{h+1} \leq \delta$ and provides informative (sharp) estimates (see Figures 4-5 in Section 5). Finally notice that when the upper bound in [12, Section 2] is informative (i.e., $\|R\|$ is small) then the upper bounds in Eqs. (13) and (14) are also informative.

Consider Notation 4.7. We end this section by pointing out that our approach allows us to obtain upper bounds for the distance between the class $\Lambda\text{-adm}_h(A)$ and the subspaces spanned by $\{\widehat{x}_{\sigma(1)}, \ldots, \widehat{x}_{\sigma(h)}\}$, where $\sigma : \{1, \ldots, h\} \to \{1, \ldots, r\}$ is any injective function (that is, we have not made essential use of the fact that $\widehat{X}_1$ is spanned by the eigenvectors of $P_{\mathcal{Q}}AP_{\mathcal{Q}}$, corresponding to its largest eigenvalues). Nevertheless, for definiteness, we have made this specific choice for $\widehat{X}_1$ (see Section 5 for some numerical examples that test Theorem 4.8 above).

4.4 On the sensitivity of Λ-admissible subspaces computation

In this subsection, we turn our attention to the sensitivity of the computation of the class of Λ-admissible subspaces of a Hermitian matrix, measured in terms of a suitably defined condition number. The condition number of computing left and right singular subspaces has been treated in the pioneering work of Sun [19] and has been recently revisited in [20], where the exact value of the condition number is obtained. Using different approaches, these works show that given a matrix $M \in \mathbb{K}^{n\times n}$ with singular values $\sigma_1 \geq \ldots \geq \sigma_n$, the condition number for the computation of its h-dimensional singular subspace associated with the largest singular values (with respect to the Frobenius norm) is [20, Theorem 1]

$$\max_{\substack{1\leq i\leq h \\ h+1\leq l\leq n}} \frac{1}{|\sigma_i - \sigma_l|} \sqrt{\frac{\sigma_i^2 + \sigma_l^2}{(\sigma_i + \sigma_l)^2}}\,,$$

which is essentially $(\sigma_h - \sigma_{h+1})^{-1}$ since the fraction inside the square root is bounded between 1/2 and 1. The techniques and results from [19, 20] can be used to obtain upper bounds for the condition number of the computation of dominant eigenspaces of self-adjoint matrices. Nevertheless, for completeness, we include a simple upper bound (see Proposition 4.10 below) that will allow us to compare it with our upper bound derived for Λ-admissible subspaces (see Theorem 4.12 below).

We first consider the condition number of the dominant eigenspaces of self-adjoint matrices. To describe the necessary context, we follow Notation 4.1. Let

$$\mathcal{B} := \{B \in \mathbb{K}^{n\times n} \, : \, B = B^* \, , \ \lambda_h(B) > \lambda_{h+1}(B)\} \, ,$$

which is exactly the set of self-adjoint matrices such that the h-dimensional dominant eigenspace is uniquely defined. Notice that this is an open subset of the set of Hermitian $n \times n$ matrices, with the topology induced by a fixed u.i.n. denoted by $\|\cdot\|$. Indeed, given $B \in \mathcal{B}$, let $r_B := (\lambda_h(B) - \lambda_{h+1}(B))/2 > 0$. By Weyl's inequality for eigenvalues [2, Section III.2], if $C = C^* \in \mathbb{K}^{n\times n}$ is such that $\|B - C\|_2 < r_B$, we have that

$$\lambda_h(C) \geq \lambda_h(B) - \|B - C\|_2 > \lambda_{h+1}(B) + \|B - C\|_2 \geq \lambda_{h+1}(C) \, , \tag{15}$$

so $C \in \mathcal{B}$. Hence, $\mathcal{B}$ is a natural domain in which to study the sensitivity of h-dimensional dominant eigenspaces of self-adjoint matrices. Consider $g : \mathcal{B} \to \mathcal{G}_{n,\,h}$ such that $g(B) = \mathcal{X}_h(B)$ is the uniquely determined h-dimensional dominant eigenspace of $B \in \mathcal{B}$. For the fixed u.i.n. $\|\cdot\|$, we consider the associated metric in $\mathcal{G}_{n,\,h}$ given by $(\mathcal{T}_1, \mathcal{T}_2) \mapsto \|\sin\Theta(\mathcal{T}_1, \mathcal{T}_2)\|$ for $\mathcal{T}_1, \mathcal{T}_2 \in \mathcal{G}_{n,\,h}$.

A precise and general definition of condition numbers of maps between metric spaces, along with many examples, can be found in [15]. In the next result, we apply the general notion of condition number in our particular setting. Our approach to obtain an upper bound for the condition number of g is essentially based on Davis-Kahan's sine Theorem from [3].

Proposition 4.10. *For $B \in \mathcal{B}$ and a u.i.n. $\| \ \|$, we have that*

$$\kappa[g](B) := \lim_{\epsilon\to 0} \sup_{\substack{C\in\mathcal{B} \\ 0<\|B-C\|<\epsilon}} \frac{\|\sin\Theta(\mathcal{X}_h(B), \mathcal{X}_h(C))\|}{\|B - C\|} \leq \frac{1}{\lambda_h(B) - \lambda_{h+1}(B)} \, .$$

Proof. See Section 6.4 □

Example 4.11. Although the upper bound in Proposition 4.10 does not coincide with the exact value of $\kappa[g](B)$, this upper bound is sharp for the operator (or spectral) norm $\|\cdot\|_2$. Let $\mathcal{S}, \mathcal{T} \in \mathcal{G}_{n,\,h}$ and for some $\lambda > 0$ set $B = \lambda \, P_{\mathcal{S}}$ and $C = \lambda \, P_{\mathcal{T}}$. In this case, $B, C \in \mathcal{B}$, $g(B) = \mathcal{X}_h(B) = \mathcal{S}$ and $g(C) = \mathcal{X}_h(C) = \mathcal{T}$; moreover, $\|B - C\|_2 = \lambda \, \| \sin\Theta(\mathcal{S}, \mathcal{T})\|_2$ and $\lambda_h(B) - \lambda_{h+1}(B) = \lambda$. Hence,

$$\frac{\|\sin\Theta(\mathcal{X}_h(B), \mathcal{X}_h(C))\|}{\|B - C\|} = \frac{\| \sin\Theta(\mathcal{S}, \mathcal{T})\|_2}{\lambda \, \| \sin\Theta(\mathcal{S}, \mathcal{T})\|_2} = \frac{1}{\lambda} = \frac{1}{\lambda_h(B) - \lambda_{h+1}(B)} \, .$$

Since we can construct subspaces $\mathcal{T}$ arbitrarily close to $\mathcal{S}$ (or equivalently, orthogonal projection $P_{\mathcal{T}}$ arbitrarily close to $P_{\mathcal{S}}$), we conclude that in this case $\kappa[g](B) = (\lambda_h(B) - \lambda_{h+1}(B))^{-1}$.

In what follows, we consider a possible adaptation of these ideas to Λ-admissible subspaces. Let $0 \leq j < h < k < n$ and set

$$\mathcal{A} = \{B \in \mathbb{K}^{n\times n} \, : \, B = B^* \, , \ \lambda_j(B) > \lambda_{j+1}(B) \text{ and } \lambda_k(B) > \lambda_{k+1}(B)\} \, , \tag{16}$$

which is an open subset of the set of Hermitian $n \times n$ matrices. Indeed, given $B \in \mathcal{A}$, let $r_B := \min\{\lambda_j(B) - \lambda_{j+1}(B) \, , \ \lambda_k(B) - \lambda_{k+1}(B)\}/2$. By Weyl's inequality for eigenvalues, if $C \in \mathbb{K}^{n\times n}$ is also Hermitian and such that $\|B - C\| < r_B$, then $C \in \mathcal{A}$. Notice that if $B \in \mathcal{A}$, then the class Λ-$\mathrm{adm}_h(B)$ (with respect to the indices j and k) is well defined, whether $\lambda_h(B) = \lambda_{h+1}(B)$ or not. It turns out that $\mathcal{A}$ is a natural domain for studying the sensitivity of the class of Λ-admissible subspaces. Indeed, consider $f : \mathcal{A} \to \mathcal{P}(\mathcal{G}_{n,\,h})$ given by

$$f(B) = \Lambda\text{-}\mathrm{adm}_h(B) = \{\mathcal{S} \in \mathcal{G}_{n,\,h} \,:\, \mathcal{X}_j(B) \subseteq \mathcal{S} \subseteq \mathcal{X}_k(B)\}\,,$$

for all $B \in \mathcal{A}$. We are interested in choosing some metrics for the domain and codomain of f and computing (or at least finding an upper bound of) the associated condition number.

Let us consider a u.i.n. $\|\cdot\|$ for the domain of f and for its codomain the Hausdorff distance associated to the metric in $\mathcal{G}_{n,\,h}$ given by $(\mathcal{T}_1,\, \mathcal{T}_2) \mapsto \|\sin\Theta(\mathcal{T}_1,\, \mathcal{T}_2)\|$ for $\mathcal{T}_1,\, \mathcal{T}_2 \in \mathcal{G}_{n,\,h}$. This distance will be denoted as d_H and is given by

$$d_H(\mathfrak{C},\, \mathfrak{D}) = \max\left\{ \sup_{\mathcal{C}\in\mathfrak{C}} \inf_{\mathcal{D}\in\mathfrak{D}} \|\sin(\Theta(\mathcal{C},\, \mathcal{D}))\|\,,\, \sup_{\mathcal{D}\in\mathfrak{D}} \inf_{\mathcal{C}\in\mathfrak{C}} \|\sin(\Theta(\mathcal{C},\, \mathcal{D}))\| \right\} \text{ for } \mathfrak{C},\, \mathfrak{D} \subseteq \mathcal{G}_{n,\,h}\,.$$

The next result provides an upper bound for the sensitivity of the computation of the class of Λ-admissible subspaces a self-adjoint matrices in terms of the eigen-gaps at indices j and k.

Theorem 4.12. *Given A as in Notation 4.1 consider $f : \mathcal{A} \to \mathcal{P}(\mathcal{G}_{n,\,h})$ as above and fix an u.i.n. $\|\cdot\|$. Then, the condition number $\kappa[f](A)$ of f at A satisfies that*

$$\kappa[f](A) := \lim_{\epsilon\to 0} \sup_{\substack{B\in\mathcal{A} \\ \|A-B\|<\epsilon}} \frac{d_H(\Lambda\text{-}adm_h(A),\, \Lambda\text{-}adm_h(B))}{\|A-B\|} \leq \frac{1}{\lambda_j - \lambda_{j+1}} + \frac{1}{\lambda_k - \lambda_{k+1}}\,.$$

If $j = 0$, then the first term in the right-hand side of the inequality above should be omitted.

Proof. See Section 6.4 ☐

We remark that the upper bound for the condition number of computing the class of Λ-admissible subspaces in Theorem 4.12 is sharp for the operator norm (see Example 6.14 in Section 6.4).

Remark 4.13. Here we point out a consequence of Theorem 4.12. Thus, consider the notation in Theorem 4.12 and set some $\delta_0 > 0$. Let $\epsilon_0 > 0$ be such that

$$\sup_{\substack{B\in\mathcal{A} \\ \|A-B\|<\epsilon_0}} \frac{d_H(\Lambda\text{-}\mathrm{adm}_h(A),\, \Lambda\text{-}\mathrm{adm}_h(B))}{\|A-B\|} < \kappa[f](A) + \delta_0\,.$$

Then, for every $C \in \mathcal{A}$ such that $\|A - C\| < \epsilon_0$ and $\lambda_h(C) > \lambda_{h+1}(C)$ we have

$$\begin{aligned} \inf_{\mathcal{S}_A\in\Lambda\text{-}\mathrm{adm}_h(A)} \|\sin\Theta(\mathcal{S}_A,\, \mathcal{X}_h(C))\| &\leq \sup_{\mathcal{S}_C\in\Lambda\text{-}\mathrm{adm}_h(C)} \inf_{\mathcal{S}_A\in\Lambda\text{-}\mathrm{adm}_h(A)} \|\sin\Theta(\mathcal{S}_A,\, \mathcal{S}_C)\| \\ &\leq \frac{d_H(\Lambda\text{-}\mathrm{adm}_h(A),\, \Lambda\text{-}\mathrm{adm}_h(C))}{\|A-C\|} \quad \|A-C\| \\ &\leq (\kappa[f](A) + \delta_0)\,\|A-C\|\,. \end{aligned}$$

This can be interpreted as follows: fix A as in Notation 4.1 and let $C \in \mathbb{K}^{n\times n}$ be any matrix which satisfies $C \in \mathcal{A}$ and $\lambda_h(C) > \lambda_{h+1}(C)$ (where the last condition is generic). Then, if $\|A - C\|$ is small enough, there must be some h-dimensional Λ-admissible subspace of A close to $\mathcal{X}_h(C)$. This implies that even though the subspaces $\mathcal{X}_h(C)$ might be very different for different choices of C, each of them is close to *some* element of Λ-$\mathrm{adm}_h(A)$; hence, all such subspaces $\mathcal{X}_h(C)$ can be used to produce nice low-rank approximations of A. In particular, assume further that $\lambda_h(A) = \lambda_{h+1}(A)$ and let $j = \max\{\ell : \lambda_\ell(A) > \lambda_h(A)\}$, and $h < k = \max\{\ell : \lambda_h(A) = \lambda_\ell(A)\}$. Then, for each matrix C as above, the subspace $\mathcal{X}_h(C)$ is close to some h-dimensional dominant subspace of A.

5 Numerical examples

In this section we consider some numerical examples related to the results in Section 4 for concrete choices of the parameters involved. These have been performed in Python (version 3.13.5) mainly using numpy (version 2.3.1) and scipy (version 1.16) packages with machine precision of 2.22×10^{-16}.

We keep using the notation from Section 4. We fix $n := 3000$ and construct a symmetric and positive semi-definite matrix $A \in \mathbb{R}^{3000\times 3000}$ with an eigendecomposition $A = X\,\Lambda X^*$, where the eigenvalues of A are given by $\lambda(A) = (\lambda_1, \dots, \lambda_{3000}) \in (\mathbb{R}^{3000})^{\downarrow}$ and the columns of X are denoted by $x_1, \dots, x_{3000} \in \mathbb{R}^{3000}$, respectively. We further consider $\mathcal{X}_l = \overline{\{x_1, \dots, x_l\}}$, for $1 \le l \le 3000$.

The eigenvalues of A are chosen to include a cluster, spanning indices $j = 5$ to $k = 30$. With the target dimension fixed at $h = 10$, the eigenvalue λ_{10} falls within this cluster, whose spread is given by $\delta := \lambda_6 - \lambda_{30}$. We use the indices $j = 5$ and $k = 30$ for defining $\Lambda\text{-adm}_{10}(A)$ and introduce the convenient parameter

$$0 < \gamma := \min\{\lambda_5 - \lambda_6\,,\ \lambda_{30} - \lambda_{31}\} \implies (\lambda_5 - \lambda_6)^{-1}\,,\ (\lambda_{30} - \lambda_{31})^{-1} \le \gamma^{-1}\,.$$

In most of our examples, the eigenvalues have an exponential decay approximately of the form $10 * \exp(-0.01 * i)$ for $i \notin [5, 30]$, outside the cluster. The final numerical example involves a matrix A whose eigenvalues decay linearly outside the cluster, interpolating the values $\lambda_1 \approx 10$ and $\lambda_{3000} = 1$, which corresponds to a much challenging numerical setting.

Regarding the initial subspaces, we fix $r := 20$ and draw samples $W \in \mathbb{R}^{3000\times 20}$ from a standard Gaussian random matrix. Notice that $h = 10 \le r = 20 < k = 30$. In all the plots below, the x-axis denotes the number of iterations of the given iterative method applied to the initial subspace and A (SIM method for Figures 1-3 or Krylov method for Figures 4-5). For the rest of this section, we will abbreviate $\sin\theta_{\max}(\mathcal{S}, \mathcal{T})$ as $d(\mathcal{S}, \mathcal{T})$ for two subspaces $\mathcal{S}$ and $\mathcal{T}$. All upper bounds for the sines of principal angles are truncated at the value 1 (since clearly $\sin(\theta) \le 1$, for $\theta \in [0, \pi/2]$).

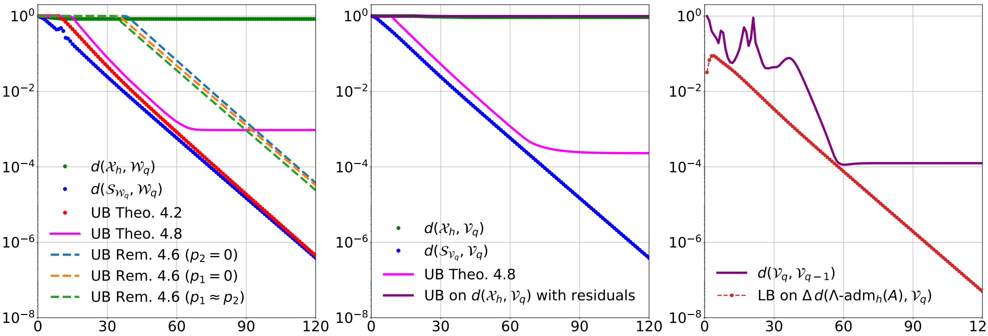

Figure 1: Exponential decay model with $\delta \approx 10^{-3}$ and $\gamma \approx 1$

Figures 1-3: we apply the Subspace Iteration Method (SIM) to A i.e., we set $R(W) = \mathcal{W}_0$ and compute $\mathcal{W}_q = R(A^q W) = A^q(\mathcal{W}_0)$ for different values of q. The decay of $\lambda(A)$ outside the cluster is exponential, exhibiting different spreads δ and gaps γ, as indicated in the captions.

First plot in Figures 1-3: we compute the values $d(\mathcal{X}_{10}, \mathcal{W}_q)$ for different values of q. The plots show that these values do not decrease at a sufficiently fast speed. Notice that the classical strategy for the convergence analysis of the SIM is to bound from above $d(\mathcal{X}_{10}, \mathcal{W}_q)$; since the actual values of $d(\mathcal{X}_{10}, \mathcal{W}_q)$ stay close to $\sin(1) \approx 0.84$ in these numerical examples, this approach does not seem convenient for the present setting.

Following the proof of Theorem 4.2 (see Section 6.1 below) we construct a convenient $\mathcal{S}_{\mathcal{W}_q} \in \Lambda\text{-adm}_{10}(A)$ to obtain numerical estimates for $d(\Lambda\text{-adm}_{10}(A),\, \mathcal{W}_q) \le d(\mathcal{S}_{\mathcal{W}_q}, \mathcal{W}_q)$ and to test the

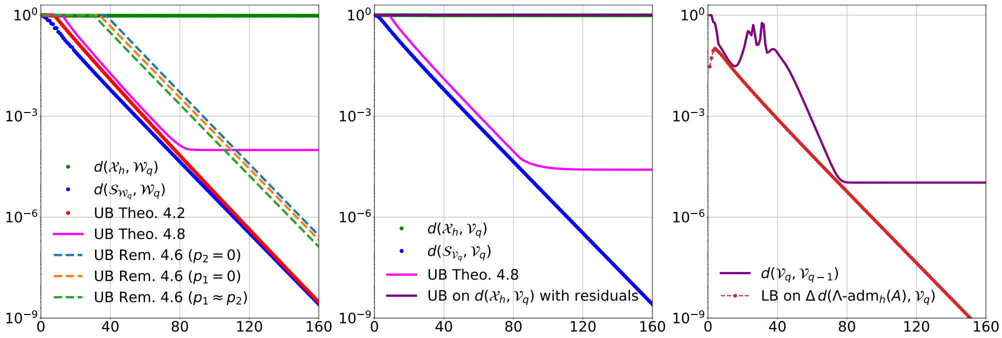

Figure 2: Exponential decay model with $\delta \approx 10^{-4}$ and $\gamma \approx 1$

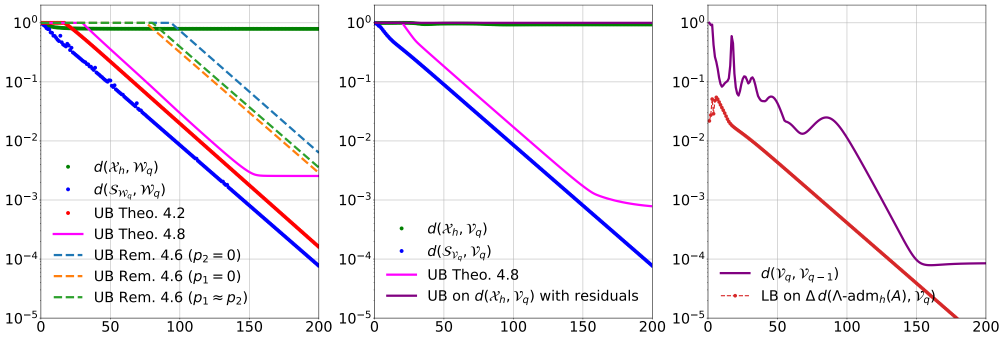

Figure 3: Exponential decay model with $\delta \approx 10^{-3}$ and $\gamma \approx 0.4$

inequality $d(\mathcal{S}_{\mathcal{W}_q}, \mathcal{W}_q) \leq d(\mathcal{X}_5, \mathcal{W}_q) + d(\mathcal{X}_{30}, \mathcal{W}_q)$ obtained in the same proof. We also plot the upper bounds $d(\Lambda\text{-adm}_{10}(A), \mathcal{W}_q)$ obtained from Theorem 4.4 (see Remark 4.5), for different choices of p_1 and p_2 (so that $p_1 + p_2 \approx r - h = 20 - 10 = 10$). In these numerical examples, the previously mentioned distances decay exponentially (though the upper bounds from Theorem 4.4 tend to take some iterations before becoming informative). The speed at which the corresponding values decay has a dependence on the gaps (i.e. on γ): the larger the gaps are, the faster these values decay.

Finally, we apply the upper bound for $d(\Lambda\text{-adm}_{20}(A), \mathcal{W}_q)$ (i.e. setting $\mathcal{Q} = \mathcal{W}_q$) in Eq. (14) from Theorem 4.8. Since the values of $d(\mathcal{X}_{10}, \mathcal{W}_q)$ stay close to 1 (see Remark 4.9), our approach in this setting is limited by the lower threshold $0 < \gamma^{-1} \cdot \delta$ (the values of our upper bounds stabilize around this quantity). On the other hand, the numerical examples show that during an initial number of iterations, the upper bound in Eq. (14) outperforms the upper bounds from Theorem 4.4. A possible reason for this behavior is that the proof of Theorem 4.4 includes several arguments considering *worst case scenarios*, while the upper bound in Eq. (14) involves the computation of some concrete values associated with (partitions of) A. In the long run, Theorem 4.4 seems to provide better estimates (as seen in Figures 1-3). Thus, these results can be considered complementary.

Second plot in Figures 1-3: we apply the Rayleigh-Ritz method to the matrix A and the subspaces $\mathcal{W}_q$, and compute the 10-dimensional dominant eigenspace of the compression $P_{\mathcal{W}_q} A P_{\mathcal{W}_q}$, denoted by $\mathcal{V}_q$, for different numbers of iterations q. First, we compute the distance $d(\mathcal{X}_{10}, \mathcal{V}_q)$. In these numerical examples, these values remain close to 1; hence, the classical approach of bounding them from above (as a measure of the quality of $\mathcal{V}_q$) does not seem a good strategy in our present (clustered) setting. We also compute the upper bounds for $d(\mathcal{X}_{10}, \mathcal{V}_q)$ obtained in [12], which cannot be informative, since they provide upper bounds for $d(\mathcal{X}_{10}, \mathcal{V}_q) \approx 1$. On the other hand, we computed two upper bounds for the distance $d(\Lambda\text{-adm}_{10}(A), \mathcal{V}_q)$. We computed the upper bound in Eq. (12) from Theorem 4.8 (in this setting $\mathcal{Q} = \mathcal{W}_q$ so $\mathcal{V}_q = R(\widehat{X}_1)$) and we also applied the constructive approach in the proof of Theorem 4.2 to produce a convenient $\mathcal{S}_{\mathcal{V}_q} \in \Lambda\text{-adm}_{10}(A)$ and plotted the values of $d(\mathcal{S}_{\mathcal{V}_q}, \mathcal{V}_q)$ for different values of q. Since distance $d(\mathcal{X}_{10}, \mathcal{V}_q)$ does not decrease sufficiently fast as a function of q, the upper bound in Eq. (12) suffers from the threshold $0 < \gamma^{-1} \cdot \delta$ (which can be explained using an argument analogous to that in Remark 4.9). Notice that despite the behavior of the upper bound in Eq. (12), the values $d(\mathcal{S}_{\mathcal{V}_q}, \mathcal{V}_q)$ decay exponentially.

Third plot in Figures 1-3: the curves in these plots correspond to a rather heuristic analysis, as follows: consider $\{\mathcal{V}_q\}_q$ as a finite sequence of points (i.e. a walk) in the Grassmannian $\mathcal{G}_{3000,\,10}(\mathbb{R})$. The length between consecutive points is given by $d(\mathcal{V}_{q-1}, \mathcal{V}_q)$. If when moving from $\mathcal{V}_{q-1}$ to $\mathcal{V}_q$ the points become closer to $\mathcal{X}_{10}$ then the reduction of the distances $d(\mathcal{X}_{10}, \mathcal{V}_{q-1}) - d(\mathcal{X}_{10}, \mathcal{V}_q)$ is at most the length of the step $d(\mathcal{V}_{q-1}, \mathcal{V}_q)$. Formally,

$$d(\mathcal{X}_{10},\, \mathcal{V}_{q-1}) - d(\mathcal{X}_{10},\, \mathcal{V}_q) \leq d(\mathcal{V}_{q-1},\, \mathcal{V}_q)\,. \tag{17}$$

Similarly, the decrease in the distance to the class of Λ-admissible subspaces also satisfies

$$d(\Lambda\text{-adm}_{10}(A),\, \mathcal{V}_{q-1}) - d(\Lambda\text{-adm}_{10}(A),\, \mathcal{V}_q) \leq d(\mathcal{V}_{q-1},\, \mathcal{V}_q)\,. \tag{18}$$

Unfortunately, it is not possible to compute the exact value of the distances to the class of Λ-admissible subspaces considered above; nevertheless, since $\dim \mathcal{V}_q = 10$ for every q, it is possible to obtain the following lower bound: for any $\mathcal{S} \in \Lambda\text{-adm}_{10}(A)$:

$$d(\Lambda\text{-adm}_{10}(A),\, \mathcal{V}_{q-1}) - d(\Lambda\text{-adm}_{10}(A),\, \mathcal{V}_q) \geq \max\{d(\mathcal{X}_5,\, \mathcal{V}_{q-1}),\, d(\mathcal{X}_{30},\, \mathcal{V}_{q-1})\} - d(\mathcal{S},\, \mathcal{V}_q)\,. \tag{19}$$

We present a formal verification of these inequalities in Remark 5.1. In a *metric sense*, the approach based on the class $\Lambda\text{-adm}_{10}(A)$ is (numerically) efficient if the difference $d(\Lambda\text{-adm}_{10}(A),\, \mathcal{V}_{q-1}) - d(\Lambda\text{-adm}_{10}(A),\, \mathcal{V}_q)$ is (uniformly) proportional to the length of the step $d(\mathcal{V}_{q-1},\, \mathcal{V}_q)$.

We plot the values of $d(\mathcal{V}_{q-1},\, \mathcal{V}_q)$ for different values of q: in all numerical examples these values have two regimes into which they have quite different behaviors. In an initial number of iterations, the values decay exponentially, until they reach a value of the order of $10^{-1}\,\delta$; then, the values stabilize around $10^{-1}\,\delta$. That is, after an initial number of iterations where the values decay, then the length of the step $d(\mathcal{V}_{q-1},\, \mathcal{V}_q)$ becomes essentially constant. In particular, the (numerical, computed) sequence $\{\mathcal{V}_q\}$ does not converge to any fixed subspace (for practical purposes). On the other hand, these last facts are compatible with the situation in which, after an initial number of iterations, the subspaces $\mathcal{V}_q$ lie in a spiral trajectory that is getting closer to *the class* $\Lambda\text{-adm}_{10}(A)$. During the initial number of iterations in which the values of $d(\mathcal{V}_{q-1},\, \mathcal{V}_q)$ decay exponentially, the values of the lower bound in Eq. (19) are quite close to those of $d(\mathcal{V}_{q-1},\, \mathcal{V}_q)$. Afterwards, the values of the lower bound in Eq. (19) keep decaying exponentially, which is consistent with the fact that $\mathcal{V}_q$ keeps getting closer to $\Lambda\text{-adm}_{10}(A)$ as the number q of iterations increases. We also point out that the inequality from Eq (17) was numerically tested and it does not provide a sharp estimation. Indeed, the values $d(\mathcal{X}_{10},\, \mathcal{V}_{q-1}) - d(\mathcal{X}_{10},\, \mathcal{V}_q)$ oscillated between being a few order of magnitude lower than $d(\mathcal{V}_{q-1},\, \mathcal{V}_q)$ and being negative.

Figures 4-5: we apply the Krylov Subspace Method to A i.e., we consider the matrix W, and compute the Krylov space $\mathcal{K}_q = R([A^q W \ \ldots \ AW\ W]) \subset \mathbb{R}^n$ for different values of q. Both examples correspond to matrices A whose eigenvalues have spread $\delta \approx 10^{-3}$ and gaps $\gamma \approx 0.4$, but different decays outside the cluster: in Figure 4 we consider an exponential decay model, while in Figure 5 we consider a (more challenging) linear decay model.

Notice that, due to our choice of parameters, the dimension of the Krylov subspaces rapidly surpasses the enveloping index k (indeed, $\dim \mathcal{K}_1$ is expected to be $2 \times r = 40 > 30 = k$). Also, estimations from Eqs. (13) and (14) in Theorem 4.8 prove useful when one has a subspace $\mathcal{Q}$ which has an associated residual R with small norm (which can be interpreted as $\mathcal{Q}$ being close to being invariant). Since one does not expect the increasing sequence $\mathcal{K}_q$ to produce these small residuals, we adopt the following approach in our analysis: we apply the Rayleigh-Ritz method to the matrix A and the subspaces $\mathcal{K}_q$, and compute both the 20-dimensional and the 10-dimensional dominant eigenspaces of the compression $P_{\mathcal{K}_q} A P_{\mathcal{K}_q}$, denoted by $\mathcal{W}_q$ and $\mathcal{V}_q$ respectively, for different numbers of iterations q (notice that $\mathcal{V}_q \subseteq \mathcal{W}_q \subseteq \mathcal{K}_q$). Then, as before, we consider $\mathcal{Q} = \mathcal{W}_q$ and $\mathcal{V}_q = R(\widehat{X}_1)$.

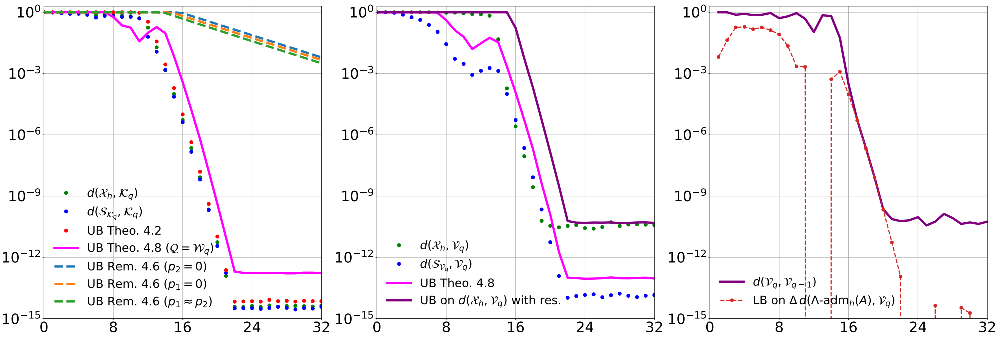

Figure 4: Exponential decay model with $\delta \approx 10^{-3}$ and $\gamma \approx 1$

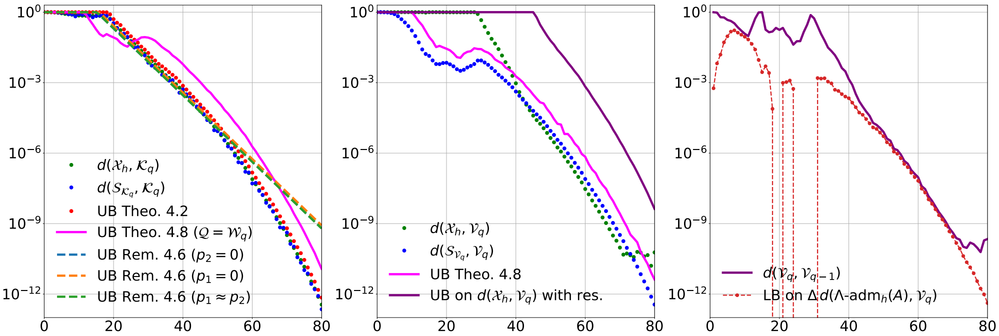

Figure 5: Linear decay model with $\delta \approx 10^{-3}$ and $\gamma \approx 1$

First plot in Figures 4-5: as in the first plot of the previous figures, we compute $d(\mathcal{X}_{10}, \mathcal{K}_q)$ and we follow the proof of Theorem 4.2 (see Section 6.1 below) to construct a convenient $\mathcal{S}_{\mathcal{K}_q} \in \Lambda\text{-adm}_{10}(A)$ and obtain numerical estimates for $d(\Lambda\text{-adm}_{10}(A), \mathcal{K}_q) \le d(\mathcal{S}_{\mathcal{W}_q}, \mathcal{K}_q) \le d(\mathcal{X}_5, \mathcal{K}_q) + d(\mathcal{X}_{30}, \mathcal{K}_q)$. For these examples, we got $\dim \mathcal{K}_1 \ge 30$ and thus $d(\Lambda\text{-adm}_{10}(A),\, \mathcal{K}_q) \le d(\mathcal{X}_h, \mathcal{K}_q) \le d(\mathcal{X}_{30},\, \mathcal{K}_q)$. Since $\lambda_{30} - \lambda_{31}$ is significant, these distances decrease exponentially.

We also plot the upper bounds $d(\Lambda\text{-adm}_{10}(A), \mathcal{K}_q)$ obtained from Theorem 4.4 (see Remark 4.6), for different choices of p_1 and p_2 (so that $p_1 + p_2 \approx r - h = 20 - 10 = 10$). As mentioned in Remark 4.6, in case of the Krylov sequence Theorem 4.4 could potentially provide many different useful upper bounds for the decay of $d(\Lambda\text{-adm}_{10}(A), \mathcal{K}_q)$ (via the choice of different polynomials ϕ, provided that $\phi(\Lambda_k)$ is invertible). Here, we intend to take advantage of the well known separation properties of the Chebyshev polynomials [1, 5, 10] to produce informative bounds. Concretely, for a previously fixed pair p_1, p_2 as we described, let $\phi_\ell(x) = T_\ell(\frac{x - \lambda_{3000}}{\lambda_{30+p_2+1} - \lambda_{3000}})$, where T_ℓ denotes the Chebyshev polynomial of the first kind of degree ℓ. Using the properties of T_ℓ listed in [1, Section 5] one can notice, for example, that ϕ_ℓ has degree ℓ, $\phi_\ell(\lambda_{30+p_2+1}) = 1$ and that ϕ_ℓ is monotonically increasing in $[\lambda_{30+p_2+1}, \infty)$. These facts imply that $\phi_\ell(\Lambda_k)$ is invertible, and thus we can compute the associated bounds from 4.6 exploiting the properties of T_ℓ listed in [1, Section 5].

Finally, we apply the upper bound for $d(\Lambda\text{-adm}_{20}(A), \mathcal{W}_q)$ (i.e. setting $\mathcal{Q} = \mathcal{W}_q$) in Eq. (14) from Theorem 4.8. Since the values of $d(\mathcal{X}_{10}, \mathcal{W}_q)$ decrease exponentially as the number of iterations increases, the upper bound in Eq. (14) does not suffer from the threshold $\gamma^{-1}\cdot\delta$; these estimates typically outperform the estimates from Theorem 4.4 above in these examples, until they reach a threshold (and stabilize). Also, it can be easily shown that $d(\Lambda\text{-adm}_{20}(A), \mathcal{K}_q) \leq d(\Lambda\text{-adm}_{20}(A), \mathcal{W}_q)$, which is why we include this curve here.

Second plot in Figures 4-5: first, we compute the distance $d(\mathcal{X}_{10}, \mathcal{V}_q)$ and apply the constructions in the proof of Theorem 4.2 to produce $\mathcal{S}_{\mathcal{V}_q} \in \Lambda\text{-adm}_{10}(A)$ and compute $d(\mathcal{S}_{\mathcal{V}_q}, \mathcal{V}_q)$. In these numerical examples, both of these values decay exponentially. We then set $\mathcal{Q} = \mathcal{W}_q$ and $\mathcal{V}_q = R(\widehat{X}_1)$ and compute the upper bounds for these Ritz spaces: The upper bound obtained in [12] for $d(\mathcal{X}_h, \mathcal{V}_q)$ and the upper bound for $d(\Lambda\text{-adm}_{10}(A), \mathcal{V}_q)$ in Eq. (12) from Theorem 4.8.

We remark that the (discrepancy) quotient between the upper bound derived from [12] and the upper bound in Eq. (12) is $\delta^{-1} \approx 10^3$; indeed, by [12, Theorem 5.1], $\|\sin(\mathcal{X}_{10}, \widehat{X}_1)\|_2 \leq \frac{\|R\|_2}{\text{Gap}}(1 + \frac{\|R_2\|^2}{\text{gap}^2})$, where $\text{Gap} = \min|\lambda(\Lambda_1) - \lambda(A_3)| \approx \lambda_{10} - \lambda_{31} \geq \lambda_{30} - \lambda_{31}$, $\text{gap} = \min|\lambda(\Lambda_1) - \lambda(\widehat{\Lambda}_2)| \approx \delta = 10^{-3}$ and $\|R\|_2 \approx \|R_2\|_2$, in these examples. Thus, the upper bound for $d(\mathcal{X}_h, \mathcal{V}_q)$ in [12, Theorem 5.1] is affected by the spread of the cluster δ (by the factor δ^{-1}), while our estimates for $d(\Lambda\text{-adm}_{10}(A), \mathcal{V}_q)$ are not.

Third plot in Figures 4-5: the curves in these third plots correspond to the same heuristic analysis as that described for the Third plot in Figures 1-3.

Remark 5.1 (Proofs of Equations (17)-(19))**.** Given $\mathcal{T} \in \mathcal{G}_{3000,\,10}$, the triangle inequality for $d(\mathcal{T}, \mathcal{V}) = \sin(\theta_{\max}(\mathcal{T}, \mathcal{V}))$ (see [13, 14]) implies that $|\ d(\mathcal{T},\, \mathcal{V}_{q-1}) - d(\mathcal{T},\, \mathcal{V}_q)\ | \leq d(\mathcal{V}_{q-1},\, \mathcal{V}_q)$, and taking $\mathcal{T} = \mathcal{X}_{10}$ gives us Eq. (17). Next, given $q \geq 1$ consider a subspace $\mathcal{S}_q \in \Lambda\text{-adm}_{10}(A)$ such that $d(\Lambda\text{-adm}_{10}(A),\, \mathcal{V}_q) = d(\mathcal{S}_q,\, \mathcal{V}_q)$. Then, by the triangle inequality once more,

$$d(\Lambda\text{-adm}_{10}(A),\, \mathcal{V}_{q-1}) - d(\Lambda\text{-adm}_{10}(A),\, \mathcal{V}_q) \leq d(\mathcal{S}_q,\, \mathcal{V}_{q-1}) - d(\mathcal{S}_q,\, \mathcal{V}_q) \leq d(\mathcal{V}_{q-1},\, \mathcal{V}_q)\,,$$

which gives Eq. (18). Finally, Eq. (19) is a simple consequence of Remark 4.3.

6 Several proofs

In this section, we include the proofs of several results considered in the previous sections. We first recall the general notation that we have used so far

Notation 6.1. Let $\mathbb{K} = \mathbb{R}$ or $\mathbb{K} = \mathbb{C}$. We consider:

1. A self-adjoint matrix $A \in \mathbb{K}^{n\times n}$ with an eigendecomposition given by $A = X\,\Lambda X^*$. The eigenvalues of A are given by $\lambda(A) = (\lambda_1, \ldots, \lambda_n) \in (\mathbb{R}^n)^{\downarrow}$, counting multiplicities and arranged non-increasingly and we set $\lambda_0 := \infty$. The columns of X are denoted by $x_1, \ldots, x_n \in \mathbb{K}^n$, respectively. We further consider $\mathcal{X}_l = \overline{\{x_1, \ldots, x_l\}}$, for $1 \leq l \leq n$.

2. For a target dimension $1 \leq h$, we consider (enveloping) indices $0 \leq j < h < k$ such that $\lambda_j > \lambda_{j+1}$ and $\lambda_k > \lambda_{k+1}$. Notice that $\mathcal{X}_j$ and $\mathcal{X}_k$ are unique and independent of X in the present setting. We use the indices j and k for the Λ-admissible subspaces of A and set

$$\Lambda\text{-adm}_h(A) = \{\,\mathcal{S} : \mathcal{S} \text{ is an } h\text{-dimensional } \Lambda\text{-admissible space of } A\,\}\,.$$

Proof of Eq. (6) *in Remark 3.4.* consider Notation 6.1 and assume further that $\lambda_k \geq 0$. Let $\mathcal{S} \in \Lambda\text{-adm}_h(A)$: then, the low-rank approximation $P_\mathcal{S} A P_\mathcal{S}$ obtained by compressing A onto $\mathcal{S}$ satisfies that it is positive semi-definite and

$$\|A - P_\mathcal{S} A P_\mathcal{S}\| \leq \|A - A_h\| + \|\text{diag}(\underbrace{\delta, \ldots, \delta}_{k-j}, 0, \ldots, 0)\|\,, \tag{20}$$

where $\delta = \lambda_{j+1} - \lambda_k \geq 0$ is the spread of the cluster and $\|\cdot\|$ denotes a u.i.n. In particular, $\|A - P_{\mathcal{S}}AP_{\mathcal{S}}\|_2 \leq \|A - A_h\|_2 + \delta = \lambda_{h+1} + \delta$. To prove Eq. (20) we first show that

$$(|\lambda_i(A - P_{\mathcal{S}}AP_{\mathcal{S}})|)_{i=1}^n \prec_w (\underbrace{0,\dots,0}_{j},\, |\lambda_{j+1}|,\dots,|\lambda_{k-h+j}|\,,\, \underbrace{\delta,\dots,\delta}_{h-j}, |\lambda_{k+1}|,\dots,|\lambda_n|)^{\downarrow}\,. \tag{21}$$

Indeed, consider the representation $A - P_{\mathcal{S}}AP_{\mathcal{S}} = P_{\mathcal{X}_k}AP_{\mathcal{X}_k} - P_{\mathcal{S}}AP_{\mathcal{S}} + (I - P_{\mathcal{X}_k})A(I - P_{\mathcal{X}_k})$. Since $\mathcal{X}_j \subset \mathcal{S} \subset \mathcal{X}_k$ then $P_{\mathcal{X}_j}P_{\mathcal{S}} = P_{\mathcal{X}_j}$ and hence,

$$\lambda(A - P_{\mathcal{S}}AP_{\mathcal{S}}) = (\underbrace{0,\dots,0}_{j},\, \lambda((P_{\mathcal{X}_k}AP_{\mathcal{X}_k} - P_{\mathcal{S}}AP_{\mathcal{S}})|_{\mathcal{X}_k\ominus\mathcal{X}_j})\,,\, \lambda((I - P_{\mathcal{X}_k})A)|_{\mathcal{X}_k^{\perp}}))^{\downarrow}\,. \tag{22}$$

On the one hand, we can apply Weyl's inequality for selfadjoint matrices and get

$$\begin{aligned} \lambda((P_{\mathcal{X}_k}AP_{\mathcal{X}_k} - P_{\mathcal{S}}AP_{\mathcal{S}})|_{\mathcal{X}_k\ominus\mathcal{X}_j}) &\prec \lambda(P_{\mathcal{X}_k}AP_{\mathcal{X}_k}|_{\mathcal{X}_k\ominus\mathcal{X}_j}) - \lambda(P_{\mathcal{S}}AP_{\mathcal{S}}|_{\mathcal{X}_k\ominus\mathcal{X}_j})^{\uparrow} \\ &= (\,(\lambda_{i+j})_{i=1}^{k-h}\,,\, (\lambda_{k-h+j+i} - \lambda_{h-j-i+1}(P_{\mathcal{S}}AP_{\mathcal{S}}|_{\mathcal{X}_k\ominus\mathcal{X}_j}))_{i=1}^{h-j})\,. \end{aligned}$$

Using that $f(x) = |x|$, $x \in \mathbb{R}$ is convex and the properties of majorization [2, Corollary II.3.4],

$$|\lambda((P_{\mathcal{X}_k}AP_{\mathcal{X}_k} - P_{\mathcal{S}}AP_{\mathcal{S}})|_{\mathcal{X}_k\ominus\mathcal{X}_j})| \prec_w (\,(\lambda_{i+j})_{i=1}^{k-h}\,,\, (\,|\,\lambda_{k-h+j+i} - \lambda_{h-i+1}(P_{\mathcal{S}}AP_{\mathcal{S}})\,|\,)_{i=1}^{h-j})\,.$$

Using the interlacing inequalities in Remark 3.3, we see that

$$(\,|\,\lambda_{k-h+j+i} - \lambda_{h-i+1}(P_{\mathcal{S}}AP_{\mathcal{S}})\,|\,)_{i=1}^{h-j} \leq (\,\max\{|\lambda_{k-h+j+i} - \lambda_{k-i+1}|, |\lambda_{k-h+j+i} - \lambda_{h-i+1}|\}\,)_{i=1}^{h-j}\,.$$

Since for $1 \leq i \leq h-j$ we have that $j+1 \leq k-h+j+i$, $h-i+1$, $k-i+1 \leq k$, then

$$\max\{\, |\lambda_{k-h+j+i} - \lambda_{k-i+1}|, |\lambda_{k-h+j+i} - \lambda_{h-i+1}|\,,\ \ 1 \leq i \leq h-j \,\} \leq \lambda_{j+1} - \lambda_k = \delta\,.$$

The previous facts show that

$$|\lambda((P_{\mathcal{X}_k}AP_{\mathcal{X}_k} - P_{\mathcal{S}}AP_{\mathcal{S}})|_{\mathcal{X}_k\ominus\mathcal{X}_j})| \prec_w (\,|\lambda_{j+1}|,\dots,|\lambda_{k-h+j}|\,,\, \underbrace{\delta,\dots,\delta}_{h-j})\,.$$

This last fact, together with Eq. (22) and the properties of block submajorization of vectors with non-negative entries prove Eq. (21). In particular, we get that

$$\|A - P_{\mathcal{S}}AP_{\mathcal{S}}\| \leq \|\mathrm{diag}(\underbrace{0,\dots,0}_{j},\, |\lambda_{j+1}|,\dots,|\lambda_{k-h+j}|\,,\, \underbrace{\delta,\dots,\delta}_{h-j}, |\lambda_{k+1}|,\dots,|\lambda_n|)\|\,.$$

Using that $|\ |\lambda_{j+i}| - |\lambda_{h+i}|\ | \leq |\lambda_{j+i} - \lambda_{h+i}| \leq \delta$, for $1 \leq i \leq k-h$, the triangle inequality for $\|\cdot\|$ and the entry-wise monotonicity of $\|\cdot\|$ for vectors of non-negative entries, we now see that

$$\begin{gathered} \|\mathrm{diag}(\underbrace{0,\dots,0}_{j},\, |\lambda_{j+1}|,\dots,|\lambda_{k-h+j}|\,,\, \underbrace{\delta,\dots,\delta}_{h-j}, |\lambda_{k+1}|,\dots,|\lambda_n|)\| \leq \\ \|\mathrm{diag}(\underbrace{0,\dots,0}_{j}\ \underbrace{\delta,\dots,\delta}_{k-h}\,,\, \underbrace{\delta,\dots,\delta}_{h-j}, \underbrace{0,\dots,0}_{n-k})\| + \|A - A_h\| \end{gathered}$$

since $\|A - A_h\| = \|\mathrm{diag}(\underbrace{0,\dots,0}_{j},\, |\lambda_{h+1}|,\dots,|\lambda_k|\,,\, \underbrace{0,\dots,0}_{h-j}, |\lambda_{k+1}|,\dots,|\lambda_n|)\|$. $\square$

6.1 Proof of Theorem 4.2

We first consider some results that will allow us to prove Theorem 4.2. The following lemma improves [7, Proposition 5.3].

Lemma 6.2. *Let* $(\mathcal{T}',\,\mathcal{T}'')$ *and* $(\mathcal{H}',\,\mathcal{H}'')$ *be pairs of subspaces in* $\mathbb{K}^n$*, such that* $\dim(\mathcal{H}') \le \dim(\mathcal{T}')$, $\dim(\mathcal{H}'') \le \dim(\mathcal{T}'')$ *and* $\mathcal{H}'\perp\mathcal{H}''$*. Let* $\mathcal{H} := \mathcal{H}' \oplus \mathcal{H}''$ *and consider a subspace* $\mathcal{T} \subseteq \mathbb{K}^n$ *such that* $\mathcal{T}',\,\mathcal{T}'' \subseteq \mathcal{T}$ *and* $\dim(\mathcal{H}) \le \dim(\mathcal{T})$*. In this case, we have that*

$$\left\|\sin\Theta(\mathcal{T},\mathcal{H})\right\| \le \left\|\sin\Theta(\mathcal{T}',\mathcal{H}')\right\| + \left\|\sin\Theta(\mathcal{T}'',\mathcal{H}'')\right\|$$

for every u.i.n. $\|\cdot\|$*. Moreover, for the operator and Frobenius norms* $\|\cdot\|_{2,F}$ *we also have that*

$$\|\sin\Theta(\mathcal{T},\mathcal{H})\|_{2,F}^2 \le \|\sin\Theta(\mathcal{T}',\mathcal{H}')\|_{2,F}^2 + \|\sin\Theta(\mathcal{T}'',\mathcal{H}'')\|_{2,F}^2\,.$$

Proof. As usual, we compute the sines of the principal angles in terms of singular values of products of projections. Using that $P_{\mathcal{H}} = P_{\mathcal{H}'} + P_{\mathcal{H}''}$ and the monotony of the principal angles we have that

$$\begin{aligned}\left\|\sin\Theta(\mathcal{T},\mathcal{H})\right\| &= \|(I-P_{\mathcal{T}})(P_{\mathcal{H}'}+P_{\mathcal{H}''})\| \le \|(I-P_{\mathcal{T}})P_{\mathcal{H}'}\| + \|(I-P_{\mathcal{T}})P_{\mathcal{H}''}\| \\ &= \left\|\sin\Theta(\mathcal{T},\mathcal{H}')\right\| + \left\|\sin\Theta(\mathcal{T},\mathcal{H}'')\right\| \le \left\|\sin\Theta(\mathcal{T}',\mathcal{H}')\right\| + \left\|\sin\Theta(\mathcal{T}'',\mathcal{H}'')\right\|.\end{aligned}$$

For the cases of $\|\cdot\|_2$ and $\|\cdot\|_F$, the same steps can be followed, but the first inequality now uses the fact that the ranges of $P_{\mathcal{H}'}(I-P_{\mathcal{T}})$ and $P_{\mathcal{H}''}(I-P_{\mathcal{T}})$ are orthogonal subspaces. □

In the following results, we consider Notation 6.1.

Proposition 6.3. *Let* $\mathcal{T} \subset \mathcal{X}_k$ *be such that* $t := \dim\mathcal{T}$ *satisfies* $h \le t \le k$*. Then, there exists* $\mathcal{S} \in \Lambda$*-adm*$_h(A)$*, such that*

$$\left\|\sin\Theta(\mathcal{S},\mathcal{T})\right\| = \left\|\sin\Theta(\mathcal{X}_j,\mathcal{T})\right\|. \tag{23}$$

Proof. Since all $\mathcal{S} \in \Lambda\text{-adm}_h(A)$ contain $\mathcal{X}_j$, the monotonicity of principal angles implies that $\left\|\sin\Theta(\mathcal{S},\mathcal{T})\right\| \ge \left\|\sin\Theta(\mathcal{X}_j,\mathcal{T})\right\|$. The difficulty lies in choosing $\mathcal{S}$ so that the reverse inequality is also true. Since $\dim(\mathcal{X}_j^\perp \cap \mathcal{T}) \ge t-j$, we can choose a subspace $\mathcal{T}' \subseteq \mathcal{T} \cap \mathcal{X}_j^\perp \subseteq \mathcal{X}_k$ of dimension $h-j$. Consider $\mathcal{S} = \mathcal{X}_j \oplus \mathcal{T}'$. By construction $\mathcal{S} \in \Lambda\text{-adm}_h(A)$ and by Lemma 6.2 we have that

$$\left\|\sin\Theta(\mathcal{S},\mathcal{T})\right\| \le \left\|\sin\Theta(\mathcal{X}_j,(\mathcal{T}\ominus\mathcal{T}'))\right\| + \left\|\sin\Theta(\mathcal{T}',\mathcal{T}')\right\| = \left\|\sin\Theta(\mathcal{X}_j,(\mathcal{T}\ominus\mathcal{T}'))\right\|.$$

Thus, it would be sufficient to prove that $\Theta(\mathcal{X}_j,\,\mathcal{T}\ominus\mathcal{T}') = \Theta(\mathcal{X}_j,\,\mathcal{T})$. To do this, first notice that $\dim(\mathcal{T}\ominus\mathcal{T}') = t-(h-j) = j+(t-h) \ge j$ and thus, the cosines of the angles between $\mathcal{X}_j$ and the subspaces $\mathcal{T}$ and $\mathcal{T}\ominus\mathcal{T}'$ are the first j singular values of $P_{\mathcal{X}_j}P_{\mathcal{T}}$ and $P_{\mathcal{X}_j}P_{\mathcal{T}\ominus\mathcal{T}'}$ respectively, but $P_{\mathcal{X}_j}P_{\mathcal{T}} = P_{\mathcal{X}_j}(P_{\mathcal{T}'}+P_{\mathcal{T}\ominus\mathcal{T}'}) = P_{\mathcal{X}_j}P_{\mathcal{T}\ominus\mathcal{T}'}$, since $\mathcal{T}' \subseteq \mathcal{X}_j^\perp$. □

Remark 6.4. We point out that, in the conditions of Proposition 6.3, subspaces $\mathcal{S}$ can be explicitly constructed as follows. If $T \in \mathbb{K}^{n\times t}$ and $X_j \in \mathbb{K}^{n\times j}$ are matrices with orthonormal columns that span $\mathcal{T}$ and $\mathcal{X}_j$ respectively, then $\dim(\ker(X_j^*T)) \ge t-j$. Thus, we can consider $Z \in \mathbb{K}^{t\times(h-j)}$ with orthonormal columns such that $R(Z) \subset \ker(X_j^*T)$. Then, matrix $TZ \in \mathbb{K}^{n\times(h-j)}$ has orthonormal columns that span a subspace of $\mathcal{T}$ (which would play the role of $\mathcal{T}'$ in the proof of Proposition 6.3) and matrix $S = \begin{bmatrix} X_j & TZ\end{bmatrix} \in \mathbb{K}^{n\times h}$ has orthonormal columns that span the desired subspace.

Lemma 6.5. *Let* $\mathcal{W} \subseteq \mathcal{X}_k$ *be a subspace of dimension* w *and* $\mathcal{T} \subseteq \mathbb{K}^n$ *be any subspace Then,*

1. $P_{\mathcal{X}_k}P_{\mathcal{T}}P_{\mathcal{X}_k} \le P_{\mathcal{V}}$*, where* $\mathcal{V} := P_{\mathcal{X}_k}(\mathcal{T})$*.*

2. $\theta_i(\mathcal{W},\, \mathcal{V}) \leq \theta_i(\mathcal{W},\, \mathcal{T})$ *for* $1 \leq i \leq \min\{w,\, \dim \mathcal{V}\}$*. Thus, for every u.i.n.*

$$\|\sin\Theta(\mathcal{W},\, \mathcal{V})\| \leq \|\sin\Theta(\mathcal{W},\, \mathcal{T})\| \,.$$

Proof. Notice that $P_{\mathcal{X}_k} P_{\mathcal{T}} P_{\mathcal{X}_k} = (P_{\mathcal{X}_k} P_{\mathcal{T}})(P_{\mathcal{X}_k} P_{\mathcal{T}})^*$, which tells us that

$$\operatorname{Ran}(P_{\mathcal{X}_k} P_{\mathcal{T}} P_{\mathcal{X}_k}) = \operatorname{Ran}((P_{\mathcal{X}_k} P_{\mathcal{T}})(P_{\mathcal{X}_k} P_{\mathcal{T}})^*) = \operatorname{Ran}(P_{\mathcal{X}_k} P_{\mathcal{T}}) = P_{\mathcal{X}_k}(\mathcal{T})$$

and we also have $\|P_{\mathcal{X}_k} P_{\mathcal{T}} P_{\mathcal{X}_k}\|_2 \leq 1$ by the sub-multiplicativity of the operator norm. These facts imply the first item. Now, for $1 \leq i \leq \min\{w,\, \dim P_{\mathcal{X}_k}(\mathcal{T})\}$, by the first item we have that

$$\begin{aligned} \cos^2(\theta_i(\mathcal{T},\, \mathcal{W})) = \sigma_i^2(P_{\mathcal{T}} P_{\mathcal{W}}) = \lambda_i(P_{\mathcal{W}} P_{\mathcal{T}} P_{\mathcal{W}}) &= \lambda_i(P_{\mathcal{W}}(P_{\mathcal{X}_k} P_{\mathcal{T}} P_{\mathcal{X}_k}) P_{\mathcal{W}}) \\ \leq \lambda_i(P_{\mathcal{W}} P_{\mathcal{V}} P_{\mathcal{W}}) = \sigma_i^2(P_{\mathcal{V}} P_{\mathcal{W}}) &= \cos^2(\theta_i(\mathcal{V},\, \mathcal{W})) \end{aligned}$$

which implies the first assertion of the second item. Since we are comparing the complete list of angles between $\mathcal{W}$ and $\mathcal{V} = P_{\mathcal{X}_k}(\mathcal{T})$ this implies that $(\theta(\mathcal{W},\, \mathcal{V}), 0, \ldots, 0)$ is (entry-wise) smaller than $\theta(\mathcal{W},\, \mathcal{T})$, which in turn implies the last assertion of the second item by the monotonicity of the function $\sin(x)$ and the properties of u.i.n. (see [2, Theorem IV.2.2]). □

Proof of Theorem 4.2. In what follows we show that there exists $\mathcal{S} \in \Lambda\text{-adm}_h(A)$, such that

$$\|\sin\Theta(\mathcal{S}, \mathcal{T})\| \leq \|\sin\Theta(\mathcal{X}_j, \mathcal{T})\| + \|\sin\Theta(\mathcal{X}_k, \mathcal{T})\| \,.$$

The result then follows from this last fact. We first consider $\mathcal{T} \subset \mathbb{K}^n$ such that $t := \dim \mathcal{T}$ satisfies $h \leq t \leq k$ and $\Theta(\mathcal{X}_k, \mathcal{T}) < \frac{\pi}{2} I_t$. The hypothesis on $\mathcal{T}$ implies that $\dim(P_{\mathcal{X}_k}(\mathcal{T})) = t$. By Proposition 6.3 there exists $\mathcal{S}_1 \in \Lambda\text{-adm}_t(A)$ such that $\|\sin\Theta(\mathcal{S}_1, P_{\mathcal{X}_k}(\mathcal{T}))\| = \|\sin\Theta(\mathcal{X}_j, P_{\mathcal{X}_k}(\mathcal{T}))\|$. Recall that such subspaces were constructed as $\mathcal{X}_j \oplus \mathcal{W}_1$ where $\dim \mathcal{W}_1 = t - j$ and $\mathcal{W}_1 \subseteq \mathcal{X}_j^\perp \cap P_{\mathcal{X}_k}(\mathcal{T})$ (see the proof of Proposition 6.3). Now, set $\mathcal{S} := \mathcal{X}_j \oplus \mathcal{W} \subseteq \mathcal{S}_1$ for some $h-j$-dimensional subspace $\mathcal{W}$ of $\mathcal{W}_1$. By construction, $\mathcal{S} \in \Lambda\text{-adm}_h(A)$. By combining the monotonicity of principal angles with the triangular inequality for angular metrics [13, 14] and Lemma 6.5 we can estimate that

$$\begin{aligned} \|\sin\Theta(\mathcal{S}, \mathcal{T})\| &\leq \|\sin\Theta(\mathcal{S}_1, \mathcal{T})\| \leq \|\sin\Theta(\mathcal{S}_1, P_{\mathcal{X}_k}(\mathcal{T}))\| + \|\sin\Theta(P_{\mathcal{X}_k}(\mathcal{T}), \mathcal{T})\| \\ &= \|\sin\Theta(\mathcal{X}_j, P_{\mathcal{X}_k}(\mathcal{T}))\| + \|\sin\Theta(\mathcal{X}_k, \mathcal{T})\| \\ &\leq \|\sin\Theta(\mathcal{X}_j, \mathcal{T})\| + \|\sin\Theta(\mathcal{X}_k, \mathcal{T})\| \,. \end{aligned}$$

In case $t = \dim \mathcal{T} > k$ then $\Theta(\mathcal{X}_k, \mathcal{T}) < \frac{\pi}{2} I_k$. Hence, $\tilde{T} = P_{\mathcal{T}}(\mathcal{X}_k)$ then $\dim \tilde{\mathcal{T}} = k$ and $\Theta(\tilde{\mathcal{T}}, \mathcal{X}_k) = \Theta(\mathcal{T}, \mathcal{X}_k) < \frac{\pi}{2} I_k$. Moreover, since $P_{\mathcal{T}}(\mathcal{X}_j) \subset \tilde{\mathcal{T}}$ then we also get that $\Theta(\tilde{\mathcal{T}}, \mathcal{X}_j) = \Theta(\mathcal{T}, \mathcal{X}_j)$. Thus, we can apply the previous case to $\tilde{\mathcal{T}}$ and conclude that there exists $\mathcal{S} \in \Lambda\text{-adm}_h(A)$ such that

$$\left\|\sin\Theta(\mathcal{S}, \tilde{\mathcal{T}})\right\| \leq \|\sin\Theta(\mathcal{X}_j, \mathcal{T})\| + \|\sin\Theta(\mathcal{X}_k, \mathcal{T})\| \,.$$

The result now follows from the monotonicity of principal angles i.e. $\|\sin\Theta(\mathcal{S}, \mathcal{T})\| \leq \|\sin\Theta(\mathcal{S}, \tilde{\mathcal{T}})\|$. □

Remark 6.6. Given matrices with orthonormal columns $T \in \mathbb{K}^{n\times t}$, $X_j \in \mathbb{K}^{n\times j}$ and $X_k \in \mathbb{K}^{n\times k}$, that span the subspaces involved in the statement of Theorem 4.2, we can present the subspace $\mathcal{S}$ explicitly as follows. If $t > k$, replace T with $TT^* X_k \in \mathbb{K}^{n\times k}$. Then, compute $X_k X_k^* T$ and take a matrix $Q \in \mathbb{K}^{n\times t}$ with orthonormal columns that span its range, $P_{\mathcal{X}_k}(\mathcal{T})$, and follow Remark 6.4 using Q as the matrix T of said remark to produce a matrix whose range is the desired subspace.

6.2 Proof of Theorem 4.4

To prove Theorem 4.4 we will need some preliminary results. We begin with the following identity for principal angles obtained by Knyazev and Zhu.

Theorem 6.7 ([24]). *Let* $\begin{bmatrix} X & X_\perp \end{bmatrix}$ *be a unitary matrix with* $X \in \mathbb{K}^{n\times \ell}$ *and set* $\mathcal{X} = R(X)$*. Let* $H \in \mathbb{K}^{n\times d}$ *be such that it has orthonormal columns, or such that* $rk(H) = rank(X^*H)$*. If we let* $\mathcal{H} = R(H)$ *then, the positive singular values* $\sigma_+(T)$ *of* $T = X_\perp^* H(X^*H)^\dagger$ *satisfy*

$$\tan\Theta(\mathcal{X},\mathcal{H}) = [\infty,\dots,\infty,\sigma_+(T),0,\dots,0]$$

with $\min(\dim(\mathcal{X}^\perp\cap\mathcal{H}),\dim(\mathcal{X}\cap\mathcal{H}^\perp))$ ∞*'s and* $\dim(\mathcal{X}\cap\mathcal{H})$ *zeros. In case* $rk(H) = rank(X^*H)$*, then* $\min(\dim(\mathcal{X}^\perp\cap\mathcal{H}),\dim(\mathcal{X}\cap\mathcal{H}^\perp)) = 0$*.* □

The next result appears in [1], and complements Theorem 6.7 above.

Theorem 6.8 ([1]). *Let* $\begin{bmatrix} X & X_\perp \end{bmatrix}$ *be a unitary matrix with* $X \in \mathbb{K}^{n\times \ell}$ *and set* $\mathcal{X} = R(X)$*. Let* $H' \in \mathbb{K}^{n\times d'}$ *be such that* $\mathcal{H} = R(H')$*,* $\dim\mathcal{H} \geq \ell$*, and assume that* $\mathcal{X}\cap\mathcal{H}^\perp = \{0\}$*. Then, for every unitarily invariant norm* $\|\cdot\|$ *we have that*

$$\|\tan\Theta(\mathcal{X},\mathcal{H})\| \leq \|X_\perp^* H'(X^*H')^\dagger\|\,. \tag{24}$$

□

The next results are inspired by some results from [6, 21]. In the proof of the next lemma, we will use the following elementary fact from linear algebra: if $\mathcal{S}$, $\mathcal{T} \subseteq \mathbb{K}^n$ are subspaces, then

$$\dim(\mathcal{T}) - \dim(\mathcal{S}) = \dim(\mathcal{T}\cap\mathcal{S}^\perp) - \dim(\mathcal{T}^\perp\cap\mathcal{S})\,. \tag{25}$$

Lemma 6.9. *Let* $\{x_1,\dots,x_n\}$ *be an orthonormal basis of* $\mathbb{K}^n$ *and let* $0 \leq j < h < k \leq n$*. Let* $\mathcal{W} \subset \mathbb{K}^n$ *such that* $\dim\mathcal{W} = r$ *with* $h \leq r < k$*. Let* $0 \leq p_1$*,* p_2 *be such that* $p_1 + p_2 \leq r - h$ *and*

$$\mathcal{W}^\perp \cap \overline{\{x_1,\dots,x_{j+p_1},x_{k+1},\dots,x_{k+p_2}\}} = \{0\} \quad \text{and} \quad \mathcal{W}\cap\overline{\{x_1,\dots,x_k\}}^\perp = \{0\}\,.$$

(notice that the conditions above are generic). Then, there exists $\mathcal{H}_p \subset \mathcal{W}$ *such that*

1. $\dim\mathcal{H}_p = r - (p_1+p_2) \geq h$;
2. $\mathcal{H}_p \subset \overline{\{x_{j+1},\dots,x_{j+p_1}\}}^\perp \cap \overline{\{x_{k+1},\dots,x_{k+p_2}\}}^\perp$;
3. $\mathcal{H}_p^\perp \cap \overline{\{x_1,\dots,x_j\}} = \{0\}$ *and* $\mathcal{H}_p\cap\overline{\{x_1,\dots,x_k\}}^\perp = \{0\}$.

Proof. Notice that by item 1. above we should have $\mathcal{H}_p = \mathcal{W}$ whenever $p_1 = p_2 = 0$; hence we assume that $p_1 + p_2 \geq 1$. Let $\mathcal{X}_j = \overline{\{x_1,\dots,x_j\}}$, $\mathcal{X}_{\text{aux}} = \overline{\{x_{j+1},\dots,x_{j+p_1},\, x_{k+1},\dots,x_{k+p_2}\}}$. Then, one of the assumptions about $\mathcal{W}$ is that $\mathcal{W}^\perp\cap(\mathcal{X}_j\oplus\mathcal{X}_{\text{aux}}) = \{0\}$. Set $\mathcal{H}_p := \mathcal{W}\cap\mathcal{X}_{\text{aux}}^\perp \subseteq \mathcal{W}$. It is clear that the condition in item 2. is fulfilled. To prove the other conditions are met, we rely on Eq. (25). Indeed, $\dim(\mathcal{H}_p) = \dim(\mathcal{W}\cap\mathcal{X}_{\text{aux}}^\perp) = \dim(\mathcal{W}) - \dim(\mathcal{X}_{\text{aux}}) + \dim(\mathcal{W}^\perp\cap\mathcal{X}_{\text{aux}})$, which is equal to $r-(p_1+p_2)$, since $\mathcal{W}^\perp\cap\mathcal{X}_{\text{aux}} = \{0\}$ by the assumptions about $\mathcal{W}$. So, the condition in item 1 is met. Next,

$$\begin{aligned}
\dim(\mathcal{H}_p^\perp\cap\mathcal{X}_j) &= \dim(\mathcal{X}_j) - \dim(\mathcal{H}_p) + \dim(\mathcal{H}_p\cap\mathcal{X}_j^\perp)\\
&= j + p_1 + p_2 - r + \dim(\mathcal{W}\cap\mathcal{X}_j^\perp\cap\mathcal{X}_{\text{aux}}^\perp)\\
&= \dim(\mathcal{X}_j\oplus\mathcal{X}_{\text{aux}}) - \dim(\mathcal{W}) + \dim(\mathcal{W}\cap(\mathcal{X}_j\oplus\mathcal{X}_{\text{aux}})^\perp)\\
&= \dim(\mathcal{W}^\perp\cap(\mathcal{X}_j\oplus\mathcal{X}_{\text{aux}})) = 0\,,
\end{aligned}$$

which implies that the first condition in item 3 is fulfilled. Finally, combining the facts that $\mathcal{H}_p \subset \mathcal{W}$ and $\mathcal{W}\cap\overline{\{x_1,\dots,x_k\}}^\perp = \{0\}$ we get that $\mathcal{H}_p\cap\overline{\{x_1,\dots,x_k\}}^\perp = \{0\}$. □

Remark 6.10. We point out that the subspace $\mathcal{H}_p$ from Lemma 6.9 can be explicitly constructed as follows. With the notations of that lemma, let $W \in \mathbb{K}^{n\times r}$ have orthonormal columns that span $\mathcal{W}$ and consider $F = [x_j, \dots, x_{j+p_1}, x_{k+1}, \dots, x_{k+p_2}]^* W \in \mathbb{K}^{(p_1+p_2)\times r}$. The hypothesis on $\mathcal{W}$ implies that F is full rank. Now, let $Z \in \mathbb{K}^{r\times r-(p_1+p_2)}$ be a matrix with orthonormal columns that span $\ker(F)$. Then, it can be shown that the matrix $WZ \in \mathbb{K}^{n\times(r-p_1-p_2)}$ has orthonormal columns that span a subspace $\mathcal{H}_p$ with the desired properties.

Notice that the conditions in item 3. of the previous lemma imply that the angles between $\mathcal{H}_p$ and the subspaces $\overline{\{x_1, \dots, x_j\}}$ and $\overline{\{x_1, \dots, x_k\}}$ are strictly less than $\pi/2$; thus, the tangents between them are well-defined. Next, we exploit the existence of the subspace $\mathcal{H}_p$.

Lemma 6.11. *Let $A = X\Lambda X^*$ be an eigendecomposition and consider partitions $X = \begin{bmatrix} X_j & X_{j,\perp}\end{bmatrix} = \begin{bmatrix} X_k & X_{k,\perp}\end{bmatrix}$. Let $\mathcal{W} \subset \mathbb{K}^n$ with $\dim \mathcal{W} = r$ and such that $h \le r < k$. Let p_1 and p_2 satisfy the conditions of Lemma 6.9 with respect to indices $1 \le j < h < k$ and the orthonormal basis $\{x_1, \dots, x_n\}$ of $\mathbb{K}^n$, formed by the columns of X. Let $\mathcal{H}_p$ be the subspace from Lemma 6.9. Then, for every polynomial $\phi \in \mathbb{K}[x]$ such that $\phi(\Lambda_k)$ is invertible and for every u.i.n. $\|\cdot\|$ we have that*

$$\begin{aligned} \|\tan\Theta(\mathcal{X}_j,\, \phi(A)(\mathcal{H}_p))\| &\le \|\phi(\Lambda_j)^{-1}\|_2\ \|\phi(\Lambda_{j+p_1,\perp})\|_2\ \|\tan\Theta(\mathcal{X}_j,\, \mathcal{H}_p)\|\,, \\ \|\tan\Theta(\mathcal{X}_k,\, \phi(A)(\mathcal{H}_p))\| &\le \|\phi(\Lambda_k)^{-1}\|_2\ \|\phi(\Lambda_{k+p_2,\perp})\|_2\ \|\tan\Theta(\mathcal{X}_k, \mathcal{H}_p)\|\,. \end{aligned} \tag{26}$$

Proof. Let us consider a matrix $H \in \mathbb{K}^{n\times(r-p_1-p_2)}$ with orthonormal columns and range $\mathcal{H}_p$. The proofs of the inequalities in Eq. (26) differ slightly, since the dimensions of the subspaces involved in the left-hand sides of both inequalities have different relationships. Indeed, by item 3. in Lemma 6.9 we have $\mathcal{H}_p \cap \mathcal{X}_k^\perp = \{0\}$ and hence $\operatorname{rank}(\phi(\Lambda_k)X_k^* H) = \dim \mathcal{H}_p$, since $\phi(\Lambda_k)$ is invertible. Hence,

$$\dim \mathcal{H}_p \ge \operatorname{rank}(\phi(A)(\mathcal{H}_p)) \ge \operatorname{rank}(X_k^*\phi(A)H) = \operatorname{rank}(\phi(\Lambda_k)X_k^* H) = \dim \mathcal{H}_p\,. \tag{27}$$

Thus, the inequalities above are equalities. We now prove the first inequality in Eq. (26). Recall that $j < h \le \dim \mathcal{H}_p$ and that, by item 3 in Lemma 6.9, $\mathcal{X}_j \cap \mathcal{H}_p^\perp = \{0\}$; by Theorem 6.8 we have

$$\|\tan\Theta(\mathcal{X}_j\,,\, \phi(A)(\mathcal{H}_p))\| \le \left\| X_{j,\perp}^*\phi(A)H\ \ (X_j^*\phi(A)H)^\dagger \right\| = \left\| \phi(\Lambda_{j,\perp})X_{j,\perp}^* H\ \ (\phi(\Lambda_j)X_j^* H)^\dagger \right\|.$$

Since $\mathcal{H}_p \subseteq \{x_{j+1}, \dots, x_{j+p_1}\}^\perp$ we have that $X_{j,\perp}^* H = \begin{bmatrix} x_{j+1}, \dots, x_{j+p_1}, x_{j+p_1+1}, \dots, x_n\end{bmatrix}^* H = [0\ \ X_{j+p_1,\perp}^* H]$ and thus,

$$\phi(\Lambda_{j,\perp})X_{j,\perp}^* H = \begin{bmatrix} 0 & 0 \\ 0 & \phi(\Lambda_{j+p_1,\perp}) \end{bmatrix} X_{j,\perp}^* H\,.$$

Notice that as a consequence of the hypotheses $\phi(\Lambda_j)$ is invertible, so we can apply [7, Proposition 6.4] and get that $(\phi(\Lambda_j)X_j^* H)^\dagger = (X_j^* H)^\dagger (\phi(\Lambda_j)P_{R(X_j^* H)}))^\dagger = (X_j^* H)^\dagger\ \phi(\Lambda_j)^{-1}$, since $R(X_j^* H) = \mathbb{K}^j$. Using the previous facts together with the sub-multiplicativity of u.i.n.'s, and Theorem 6.7

$$\begin{aligned} \|\tan\Theta(\mathcal{X}_j\,,\, \phi(A)(\mathcal{H}_p))\| &\le \left\| \begin{bmatrix} 0 & 0 \\ 0 & \phi(\Lambda_{j+p_1,\perp}) \end{bmatrix} X_{j,\perp}^* H\ \ (X_j^* H)^\dagger \phi(\Lambda_j)^{-1} \right\| \\ &\le \|\phi(\Lambda_{j+p_1,\perp})\|_2\ \|\phi(\Lambda_j)^{-1}\|_2\ \left\| X_{j,\perp}^* H\ \ (X_j^* H)^\dagger \right\| \\ &= \|\phi(\Lambda_{j+p_1,\perp})\|_2\ \|\phi(\Lambda_j)^{-1}\|_2\ \|\tan\Theta(\mathcal{X}_k, \mathcal{H}_p)\|\,. \end{aligned}$$

The proof of the second inequality in Eq. (26) follows a similar path. Indeed, by the comments at the beginning of the proof, we can apply Theorem 6.7 case (ii) and get that

$$\|\tan\Theta(\mathcal{X}_k\,,\,\phi(A)(\mathcal{H}_p))\| = \left\|X_{k,\perp}^*\phi(A)H(X_k^*\phi(A)H)^\dagger\right\| = \left\|\phi(\Lambda_{k,\perp})X_{k,\perp}^*H(\phi(\Lambda_k)X_k^*H)^\dagger\right\|.$$

Now, since $\mathcal{H}_p \subseteq \{x_{k+1},\ldots,x_{k+p_2}\}^\perp$ we have $X_{k,\perp}^*H = \left[x_{k+1},\ldots,x_{k+p_2},x_{k+p_2+1},\ldots,v_n\right]^*H = [0\ x_{k+p_2,\perp}^*H]$ and thus,

$$\phi(\Lambda_{k,\perp})X_{k,\perp}^*H = \begin{bmatrix}0 & 0\\ 0 & \phi(\Lambda_{k+p_2,\perp})\end{bmatrix} X_{k,\perp}^*H\,.$$

On the other hand, since $\phi(\Lambda_k)$ is invertible by hypothesis, we can apply [7, Proposition 6.4] to see that $(\phi(\Lambda_k)X_k^*H)^\dagger = (X_k^*H)^\dagger(\phi(\Lambda_k)P_{R(X_k^*H)})^\dagger$. As before, the previous facts together with the hypothesis $\mathcal{H}_p\cap\mathcal{X}_k^\perp = \{0\}$ imply that

$$\begin{aligned}\|\tan\Theta(\mathcal{X}_k\,,\,\phi(A)(\mathcal{H}_p))\| &= \left\|\begin{bmatrix}0 & 0\\ 0 & \phi(\Lambda_{k+p_2,\perp})\end{bmatrix} X_{k,\perp}^*H\ \ (X_k^*H)^\dagger(\phi(\Lambda_k)P_{R(X_k^*H)})^\dagger\right\|\\ &\le \|\phi(\Lambda_{k+p_2,\perp})\|_2\ \|(\phi(\Lambda_k)P_{R(X_k^*H)})^\dagger\|_2\ \left\|X_{k,\perp}^*H\ \ (X_k^*H)^\dagger\right\|\\ &\le \|\phi(\Lambda_{k+p_2,\perp})\|_2\ \|\phi(\Lambda_k)^{-1}\|_2\ \|\tan\Theta(\mathcal{X}_k,\mathcal{H}_p)\|\,.\end{aligned}$$ □

As a consequence of the previous results and Theorem 4.2, we can now present the

Proof of Theorem 4.4. Let $\mathcal{W}\subset\mathbb{K}^n$ with $\dim\mathcal{W}=r$, and let p_1, p_2 satisfy the conditions of Lemma 6.9 with respect to the orthonormal basis $\{x_1,\ldots,x_n\}$ of $\mathbb{K}^n$ formed by the columns of X. Let $\mathcal{H}_p$ be the subspace constructed in Lemma 6.9 and fix a polynomial $\phi\in\mathbb{K}[x]$ such that $\phi(\Lambda_k)$ is invertible. As a consequence of Theorem 4.2 there exists $\mathcal{S}\in\Lambda\text{-adm}_h(A)$ such that, for every u.i.n. $\|\cdot\|$,

$$\|\sin\Theta(\mathcal{S},\phi(A)(\mathcal{H}_p))\| \le \|\sin\Theta(\mathcal{X}_j,\phi(A)(\mathcal{H}_p))\| + \|\sin\Theta(\mathcal{X}_k,\phi(A)(\mathcal{H}_p))\|\,. \tag{28}$$

Using Lemma 6.11 we now get that

$$\begin{aligned}\|\sin\Theta(\mathcal{S},\phi(A)(\mathcal{H}_p))\| \le\ & \|\phi(\Lambda_j)^{-1}\|_2\ \|\phi(\Lambda_{j+p_1,\perp})\|_2\ \|\tan\Theta(\mathcal{X}_j,\,\mathcal{H}_p)\|+\\ & \|\phi(\Lambda_k)^{-1}\|_2\ \|\phi(\Lambda_{k+p_2,\perp})\|_2\ \|\tan\Theta(\mathcal{X}_k,\,\mathcal{H}_p))\|\,.\end{aligned} \tag{29}$$

Arguing as in the proof of Lemma 6.11 (see Eq. (27)) we get that $\operatorname{rank}(\phi(A)(\mathcal{H}_p)) = \dim\mathcal{H}_p \ge h = \dim\mathcal{S}$. By the monotonicity of the principal angles and the inclusion $\phi(A)(\mathcal{H}_p)\subseteq\phi(A)(\mathcal{W})$, we get that $\|\sin\Theta(\mathcal{S},\phi(A)(\mathcal{W}))\| \le \|\sin\Theta(\mathcal{S},\phi(A)(\mathcal{H}_p))\|$. The result now follows from the previous fact and Eq. (29). Finally, notice that if $j=0$ then $\mathcal{X}_j=\{0\}$ and in this case we have that $\Theta(\mathcal{X}_j,\mathcal{H}_p)=0$; if $k=\operatorname{rk}(A)$ and $\phi(0)=0$ then $\mathcal{X}_k=R(A)$ and the second term in the RHS of Eq. (28) is zero, since $\phi(A)(\mathcal{H}_p)\subset R(A)$. □

6.3 Proof of Theorem 4.8

We first recall the general setting: we let $A = X\Lambda X^*$ be an eigendecomposition of A and let $\Lambda\text{-adm}_\ell(A)$ denote the class of ℓ-dimensional Λ-admissible subspaces of A. Let $Q\in\mathbb{K}^{n\times r}$ have orthonormal columns; we consider an eigendecomposition $Q^*AQ = \Omega\widehat{\Lambda}\Omega^*$. Then, we consider:

$$\Omega = [\Omega_1\quad\Omega_2]\quad\text{and}\quad\widehat{\Lambda} = \begin{bmatrix}\widehat{\Lambda}_1 & \\ & \widehat{\Lambda}_2\end{bmatrix}\quad\text{with}\quad\Omega_1\in\mathbb{K}^{r\times h}\quad,\quad\widehat{\Lambda}_1\in\mathbb{K}^{h\times h}\,,$$

and $\widehat{X} := Q\Omega = [Q\Omega_1 \quad Q\Omega_2] = [\widehat{X}_1 \quad \widehat{X}_2]$. Also let $\widehat{X}_3 \in \mathbb{K}^{n\times(n-r)}$ be such that $[\widehat{X}_1 \quad \widehat{X}_2 \quad \widehat{X}_3] \in \mathcal{U}(n)$. We further set:

$$\widetilde{A} := [\widehat{X}_1 \quad \widehat{X}_2 \quad \widehat{X}_3]^*\, A\, [\widehat{X}_1 \quad \widehat{X}_2 \quad \widehat{X}_3] = \begin{bmatrix} \widehat{\Lambda}_1 & 0 & R_1^* \\ 0 & \widehat{\Lambda}_2 & R_2^* \\ R_1 & R_2 & A_3 \end{bmatrix} \quad \text{with } \widehat{\Lambda}_1 = \begin{bmatrix} \widehat{\Lambda}_{11} & 0 \\ 0 & \widehat{\Lambda}_{12} \end{bmatrix}, \tag{30}$$

where $\widehat{\Lambda}_{11} \in \mathbb{K}^{j\times j}$, $\widehat{\Lambda}_{12} \in \mathbb{K}^{(h-j)\times(h-j)}$

$$R_1 = \widehat{X}_3^* A \widehat{X}_1 = [R_{11} \; R_{12}] \quad , \quad R_2 = \widehat{X}_3^* A \widehat{X}_2 \quad , \quad R := [R_1 \;\; R_2] \quad \text{and} \quad A_3 = \widehat{X}_3^* A \widehat{X}_3$$

where $R_{11} \in \mathbb{K}^{(n-r)\times j}$ and $R_{12} \in \mathbb{K}^{(n-r)\times(r-j)}$. Finally, we define the gaps:

$$\widetilde{\mathrm{Gap}} := \min|\lambda(\widehat{\Lambda}_{11}) - (\lambda_{j+1}, \ldots, \lambda_n)| \;\; , \;\; \widehat{\mathrm{Gap}}(l) := \min|\lambda(\widehat{\Lambda}_l) - (\lambda_{k+1}, \ldots, \lambda_n)| \; , \; l = 1, 2\, ,$$

and $\mathrm{Gap}_i := \min|(\lambda_1, \ldots, \lambda_i) - \lambda(A_3)|$ for $i = j,\, k$.

Theorem 6.12. *Consider the previous notation. Then, for every u.i.n* $\|\cdot\|$ *we have that:*

$$\begin{aligned}
\|\sin\Theta(\mathcal{X}_j,\, R(\widehat{X}))\| &\le \frac{\|R\|}{\mathrm{Gap_j}}\,, && (31)\\
\|\sin\Theta(\mathcal{X}_j,\, R(\widehat{X}))\| \le \|\sin\Theta(\mathcal{X}_j,\, R(\widehat{X}_1))\| &\le \frac{\|R_{11}\|}{\widetilde{\mathrm{Gap}}}\,, && (32)\\
\|\sin\Theta(\mathcal{X}_k,\, R(\widehat{X}_1))\| &\le \frac{\|R_1\|}{\widehat{\mathrm{Gap}}(1)}\,, && (33)\\
\|\sin\Theta(\mathcal{X}_k,\, R(\widehat{X}))\| &\le \frac{\|R\|}{\mathrm{Gap}_k} \qquad\qquad \textit{if} \quad k \le r\,, && (34)\\
\|\sin\Theta(\mathcal{X}_k,\, R(\widehat{X}))\| &\le \frac{\|R_1\|}{\widehat{\mathrm{Gap}}(1)} + \frac{\|R_2\|}{\widehat{\mathrm{Gap}}(2)} \quad \textit{if} \quad r < k\,. && (35)
\end{aligned}$$

where the second term in Eq. (35) *should be omitted if* $r = h$.

Proof. The proof follows the outline of [12, Theo. 5.1]. Let L be an (increasingly ordered) subset of $\{1, \ldots, n\}$ with l elements. Denote $E_L \in \mathbb{K}^{n\times l}$ the matrix with columns x_l for every $l \in L$ and $\Lambda_L \in \mathbb{K}^{l\times l}$ the diagonal matrix with diagonal entries λ_l for $l \in L$. With this notation, we have that $AE_L = E_L\Lambda_L$ and, by Eq. (30) we also have that $\tilde{A}[\widehat{X}_1 \quad \widehat{X}_2 \quad \widehat{X}_3]^* E_L = [\widehat{X}_1 \quad \widehat{X}_2 \quad \widehat{X}_3]^* E_L \Lambda_L$. Using the previous identity and equating the corresponding blocks (see Eq. (30)), we get:

$$R_1^*\, \widehat{X}_3^* E_L = \widehat{X}_1^* E_L \Lambda_L - \widehat{\Lambda}_1 \widehat{X}_1^* E_L \quad , \quad R_2^*\, \widehat{X}_3^* E_L = \widehat{X}_2^* E_L \Lambda_L - \widehat{\Lambda}_2 \widehat{X}_2^* E_L\, ,$$

$$R \begin{bmatrix} \widehat{X}_1^* \\ \widehat{X}_2^* \end{bmatrix} E_L = \widehat{X}_3^* E_L \Lambda_L - A_3 \widehat{X}_3^* E_L\, .$$

By the well-known bound for Sylvester's equations (see [19, Ch. V]) we get that

$$\left\|\widehat{X}_1^* E_L\right\| \le \frac{\left\|R_1^*\, \widehat{X}_3^* E_L\right\|}{\min|\Lambda_L - \widehat{\Lambda}_1|}\,, \quad \left\|\widehat{X}_2^* E_L\right\| \le \frac{\left\|R_2^*\, \widehat{X}_3^* E_L\right\|}{\min|\Lambda_L - \widehat{\Lambda}_2|}\,, \quad \left\|\widehat{X}_3^* E_L\right\| \le \frac{\left\|R \begin{bmatrix} \widehat{X}_1^* \\ \widehat{X}_2^* \end{bmatrix} E_L\right\|}{\min|\Lambda_L - \lambda(A_3)|}\,.$$

First, take $L = \{1, \ldots, j\}$. Since $j \leq h \leq r$ we have that

$$\|\sin\Theta(\mathcal{X}_j,\, R(\widehat{X}))\| = \left\|\widehat{X}_3^* E_L\right\| \leq \frac{\left\| R \begin{bmatrix} \widehat{X}_1^* \\ \widehat{X}_2^* \end{bmatrix} E_L \right\|}{\min|\Lambda_L - \lambda(A_3)|} \leq \frac{\|R\|}{\mathrm{Gap_j}}\,,$$

which proves Eq. (31). Now, let us consider the auxiliary partition $\widehat{X}_1 = [\widehat{X}_{11}\, \widehat{X}_{12}]$ where the columns of $\widehat{X}_{11} \in \mathbb{K}^{n\times j}$ correspond to the Ritz vectors associated with the largest j eigenvalues of Q^*AQ; also, consider the partition of R_1 and of $\widehat{\Lambda}$ as before. Following the first part of the proof with this auxiliary decomposition yields

$$\left\|\widehat{X}_{11}^* E_L\right\| \leq \frac{\left\|R_{11}^* \widehat{X}_3^* E_L\right\|}{\min|\Lambda_L - \widehat{\Lambda}_{11}|} \leq \frac{\|R_{11}^*\|}{\min|\Lambda_L - \widehat{\Lambda}_{11}|}\,.$$

By the monotonicity of the principal angles, if we take $L = \{j+1, \ldots, n\}$ we get that

$$\|\sin\Theta(\mathcal{X}_j,\, R(\widehat{X}_1))\| \leq \|\sin\Theta(\mathcal{X}_j,\, R(\widehat{X}_{11}))\| = \left\|\widehat{X}_{11}^* E_L\right\|,$$

which proves Eq. (32). Next, take $L = \{k+1, \ldots, n\}$. Since $h < k$ we have that

$$\|\sin\Theta(\mathcal{X}_k,\, R(\widehat{X}_1))\| = \left\|E_L^* \widehat{X}_1\right\| = \left\|\widehat{X}_1^* E_L\right\| \leq \frac{\left\|R_1^*\, \widehat{X}_3^* E_L\right\|}{\min|\Lambda_L - \widehat{\Lambda}_1|} \leq \frac{\|R_1\|}{\widehat{\mathrm{Gap}}(1)}\,,$$

which is Eq. (33). Finally, to estimate $\left\|\sin\Theta(\mathcal{X}_k\, R(\widehat{X}))\right\|$ we consider separately the cases where $r < k$ or $k \leq r$. In the latter case, we can take $L = \{1, \ldots, k\}$ and use the same strategy as that of the first part of the proof to obtain

$$\|\sin\Theta(\mathcal{X}_k,\, R(\widehat{X}))\| \leq \frac{\|R\|}{\mathrm{Gap}_k}\,,$$

which is Eq. (34). If we assume that $r < k$, and take $L = \{k+1, \ldots, n\}$, then

$$\left\|\sin(\Theta(\mathcal{X}_k,\, R(\widehat{X}))\right\| = \left\|\begin{bmatrix} E_L^* \widehat{X}_1 \\ E_L^* \widehat{X}_2 \end{bmatrix}\right\| \leq \left\|\widehat{X}_1^* E_L\right\| + \left\|\widehat{X}_2^* E_L\right\| \leq \frac{\|R_1\|}{\widehat{\mathrm{Gap}}(1)} + \frac{\|R_2\|}{\widehat{\mathrm{Gap}}(2)}$$

which is Eq. (35). Finally, notice that in the case that $r = h$, there is no matrix $\widehat{X}_2$, so the terms involving it should not be accounted for. □

Proof of Theorem 4.8. By Theorem 6.12 we get that

$$\|\sin\Theta(\mathcal{X}_j,\, R(\widehat{X}_1))\| \leq \frac{\|R_{11}\|}{\widetilde{\mathrm{Gap}}} \quad \text{and} \quad \|\sin\Theta(\mathcal{X}_k, R(\widehat{X}_1))\| \leq \frac{\|R_1\|}{\widehat{\mathrm{Gap}}(1)}\,.$$

By Theorem 4.2, there exists $\mathcal{S}_1 \in \Lambda\text{-adm}_h(A)$ such that

$$\|\sin\Theta(\mathcal{S}_1, R(\widehat{X}_1))\| \leq \frac{\|R_{11}\|}{\widetilde{\mathrm{Gap}}} + \frac{\|R_1\|}{\widehat{\mathrm{Gap}}(1)}\,.$$

Notice that the inequality in Eq. (12) follows from the previous fact. Assume now that $k \leq r$ and that $\dim(P_{R(\widehat{X})}(\mathcal{X}_k)) \geq h$. Then, we can choose an h-dimensional subspace $\mathcal{T}$ such that $P_{R(\widehat{X})}(\mathcal{X}_j) \subseteq \mathcal{T} \subseteq P_{R(\widehat{X})}(\mathcal{X}_k)$. In this case, by the monotonicity of the principal angles, we get that

$$\|\sin\Theta(\mathcal{X}_j, \mathcal{T})\| = \|\sin\Theta(\mathcal{X}_j, R(\widehat{X}))\| \quad \text{and}$$

$$\|\sin\Theta(\mathcal{X}_k, \mathcal{T})\| \le \|\sin\Theta(\mathcal{X}_k, P_{R(\widehat{X})}(\mathcal{X}_k))\| = \|\sin\Theta(\mathcal{X}_k, R(\widehat{X}))\| .$$

Now, by combining Theorem 4.2 with Theorem 6.12 (and monotonicity once more) there exists $\mathcal{S}_2 \in \Lambda\text{-adm}_h(A)$ such that

$$\begin{aligned}\|\sin\Theta(\mathcal{S}_2, R(\widehat{X}))\| &\le \|\sin\Theta(\mathcal{S}_2, \mathcal{T})\| \le \|\sin\Theta(\mathcal{X}_j, \mathcal{T})\| + \|\sin\Theta(\mathcal{X}_k, \mathcal{T})\| \\ &\le \|\sin\Theta(\mathcal{X}_j, R(\widehat{X}))\| + \|\sin\Theta(\mathcal{X}_k, R(\widehat{X}))\| \le \frac{\|R\|}{\mathrm{Gap}_j} + \frac{\|R\|}{\mathrm{Gap}_k} .\end{aligned}$$

Notice that the inequality in Eq. (13) follows from the previous fact.

Assume now that $r < k$. Since $j < h \le r < k$ then (considering r in the index set of the cluster) by Theorem 4.2 and Theorem 6.12 there exists $\mathcal{S}_3 \in \Lambda\text{-adm}_r(A)$ of A such that

$$\|\sin\Theta(\mathcal{S}_3, R(\widehat{X}))\| \le \frac{\|R\|}{\mathrm{Gap}_j} + \frac{\|R_1\|}{\widehat{\mathrm{Gap}}(1)} + \frac{\|R_2\|}{\widehat{\mathrm{Gap}}(2)} .$$

Finally, as before, Eq. (14) follows from the previous inequality. □

6.4 Proof of Theorem 4.12

Let us recall the following notations: $\mathcal{B} = \{B \in \mathbb{K}^{n\times n} : B = B^* ,\ \lambda_h(B) > \lambda_{h+1}(B)\}$ and for $B \in \mathcal{B}$ we have set $r_B = (\lambda_h(B) - \lambda_{h+1}(B))/2 > 0$.

Proof of Proposition 4.10. Fix $B \in \mathcal{B}$. Suppose that $C \in \mathcal{B}$ is such that $0 < \|B - C\|_2 < r_B$ and denote as $\mathcal{X}_h(B)$ and $\mathcal{X}_h(C)$ the h-dimensional dominant eigenspaces of B and C respectively. From Weyl's inequality for eigenvalues, we have that $\lambda_h(C) - \lambda_{h+1}(B) \ge \lambda_h(B) - \lambda_{h+1}(B) - \|B - C\|_2 > 0$. This allows us to apply [2, Theo. VII.3.1] and obtain the following:

$$\|\sin\Theta(\mathcal{X}_h(B),\, \mathcal{X}_h(C))\| \le \frac{\|B - C\|}{\lambda_h(B) - \lambda_{h+1}(B) - \|B - C\|_2} .$$

Since all u.i.n. in $\mathbb{K}^{n\times n}$ are equivalent, this implies

$$\begin{aligned}\kappa[g](B) &= \lim_{\epsilon\to 0} \sup_{\substack{C\in\mathcal{B} \\ 0<\|B-C\|<\epsilon}} \frac{\|\sin\Theta(\mathcal{X}_h(B),\, \mathcal{X}_h(C))\|}{\|B - C\|} \\ &\le \lim_{\epsilon\to 0} \sup_{\substack{C\in\mathcal{B} \\ 0<\|B-C\|_2<\epsilon}} \frac{1}{\lambda_h(B) - \lambda_{h+1}(B) - \|B - C\|_2} = \frac{1}{\lambda_h(B) - \lambda_{h+1}(B)} .\end{aligned}$$ □

Before moving forward to the proof of Theorem 4.12, recall that we have fixed $0 \le j < h < k < n$ and considered $\mathcal{A} = \{B \in \mathbb{K}^{n\times n} : B = B^* ,\ \lambda_j(B) > \lambda_{j+1}(B) \text{ and } \lambda_k(B) > \lambda_{k+1}(B)\}$. If $B \in \mathcal{A}$, the class $\Lambda\text{-adm}_h(B)$ (with respect to the indices j and k) is well defined. Also, for two matrices $A,\ B \in \mathcal{A}$ and a fixed u.i.n. $\|\cdot\|$ we consider the Hausdorff distance between the sets of h-dimensional Λ-admissible subspaces of A and B, given by

$$d_H(\Lambda\text{-adm}_h(A),\, \Lambda\text{-adm}_h(B)) = \max\left\{ \sup_{\mathcal{S}_A} \inf_{\mathcal{S}_B} \|\sin(\Theta(\mathcal{S}_A,\, \mathcal{S}_B))\| \,,\, \sup_{\mathcal{S}_B} \inf_{\mathcal{S}_A} \|\sin(\Theta(\mathcal{S}_A,\, \mathcal{S}_B))\| \right\}$$

where $\mathcal{S}_A \in \Lambda\text{-adm}_h(A)$ and $\mathcal{S}_B \in \Lambda\text{-adm}_h(B)$. The next lemma bounds this distance from above.

Lemma 6.13. *Consider $\mathcal{A}$ as above. For $B \in \mathcal{A}$, let $\mathcal{X}_j(B)$ and $\mathcal{X}_k(B)$ denote its dominant eigenspaces of dimensions j and k, respectively (which are well defined). Then, for $A,\ B \in \mathcal{A}$,*

$$d_H(\Lambda\text{-}adm_h(A),\, \Lambda\text{-}adm_h(B)) \le \|\sin(\Theta(\mathcal{X}_j(A),\, \mathcal{X}_j(B))\| + \|\sin(\Theta(\mathcal{X}_k(A),\, \mathcal{X}_k(B))\| .$$

Proof. Let $\mathcal{S}_B \in \Lambda\text{-adm}_h(B)$. By monotonicity of the principal angles

$$\|\sin(\Theta(\mathcal{S}_B,\, \mathcal{X}_j(A))\| \le \|\sin(\Theta(\mathcal{X}_j(B),\, \mathcal{X}_j(A))\|\ ,\quad \|\sin(\Theta(\mathcal{S}_B,\, \mathcal{X}_k(A))\| \le \|\sin(\Theta(\mathcal{X}_k(B),\, \mathcal{X}_k(A))\|\,.$$

Then, by Theorem 4.2, there exist a subspace $\mathcal{S}_A \in \Lambda\text{-adm}_h(A)$ such that

$$\|\sin(\Theta(\mathcal{S}_B,\, \mathcal{S}_A)\| \le \|\sin(\Theta(\mathcal{X}_j(B),\, \mathcal{X}_j(A))\| + \|\sin(\Theta(\mathcal{X}_k(B),\, \mathcal{X}_k(A))\|\ , \text{ so}$$

$$\inf_{\mathcal{S}_A \in \Lambda\text{-adm}_h(A)} \|\sin(\Theta(\mathcal{S}_B,\, \mathcal{S}_A))\| \le \|\sin(\Theta(\mathcal{X}_j(B),\, \mathcal{X}_j(A))\| + \|\sin(\Theta(\mathcal{X}_k(B),\, \mathcal{X}_k(A))\|\,.$$

Since $\mathcal{S}_B \in \Lambda\text{-adm}_h(B)$ was arbitrary, we get that

$$\sup_{\mathcal{S}_B \in \Lambda\text{-adm}_h(B)}\ \inf_{\mathcal{S}_A \in \Lambda\text{-adm}_h(A)} \|\sin(\Theta(\mathcal{S}_B,\, \mathcal{S}_A))\| \le \|\sin(\Theta(\mathcal{X}_j(B),\, \mathcal{X}_j(A))\| + \|\sin(\Theta(\mathcal{X}_k(B),\, \mathcal{X}_k(A))\|$$

and by changing the roles of A and B, we obtain the other bound needed to prove the statement. □

Proof of Theorem 4.12. First, recall that we are considering a fixed A as in Notation 4.1 and in this context we have set $r_A := \min\{\lambda_j - \lambda_{j+1}\,,\ \lambda_k - \lambda_{k+1}\}/2$. Now, following the ideas of the proof on Proposition 4.10 we get that for every $B \in \mathcal{A}$ such that $0 < \|A - B\| < r_A$ we have

$$\|\sin(\Theta(\mathcal{X}_l(B),\, \mathcal{X}_l)\| \le \frac{\|A - B\|}{\lambda_l - \lambda_{l+1} - \|A - B\|_2} \quad \text{for} \quad l = j,\, k\,.$$

Combining this fact with Lemma 6.13, results in

$$\frac{d_H(\Lambda\text{-adm}_h(A),\, \Lambda\text{-adm}_h(B))}{\|A - B\|} \le \frac{1}{\lambda_j - \lambda_{j+1} - \|A - B\|_2} + \frac{1}{\lambda_k - \lambda_{k+1} - \|A - B\|_2}\,.$$

for all $B \in \mathcal{A}$ with $\|A - B\| < r_A$. Finally, as in the proof of Proposition 4.10, this implies that

$$\begin{aligned}
\kappa[g](A) &= \lim_{\epsilon \to 0} \sup_{\substack{B \in \mathcal{A}\\ 0 < \|A - B\| < \epsilon}} \frac{d_H(\Lambda\text{-adm}_h(A),\, \Lambda\text{-adm}_h(B))}{\|A - B\|}\\
&\le \lim_{\epsilon \to 0} \sup_{\substack{B \in \mathcal{A}\\ 0 < \|A - B\| < \epsilon}} \frac{1}{\lambda_j - \lambda_{j+1} - \|A - B\|_2} + \frac{1}{\lambda_k - \lambda_{k+1} - \|A - B\|_2}\\
&= \frac{1}{\lambda_j - \lambda_{j+1}} + \frac{1}{\lambda_k - \lambda_{k+1}}\,.
\end{aligned}$$

□

Next, we show that the upper bound from Theorem 4.12 can be sharp for the operator norm.

Example 6.14. Assume that $2 \le j < h < k < 2k \le n$. Let $\mathcal{X}$ and $\mathcal{Y}$ be orthogonal subspaces of $\mathbb{K}^n$ of dimensions j and $k - j$ respectively and set $A = \alpha P_{\mathcal{X}} + \beta P_{\mathcal{Y}}$ for some $\alpha > \beta > 0$. Thus, $\lambda_j = \alpha > \lambda_{j+1} = \beta = \lambda_k > \lambda_{k+1} = 0$. The hypotheses on the dimensions of $\mathcal{X}$ and $\mathcal{Y}$ allow us to find another pair of orthogonal subspaces $\mathcal{X}'$ and $\mathcal{Y}'$ of dimensions j and $k - j$ respectively such that $\mathcal{Y} \perp \mathcal{X}'$ and $\mathcal{X} \perp \mathcal{Y}'$ in such a way that $0 < \theta_{\mathcal{X}} := \theta_j(\mathcal{X},\, \mathcal{X}') < \pi/2$ and $0 < \theta_{\mathcal{Y}} := \theta_{k-j}(\mathcal{Y},\, \mathcal{Y}') < \pi/2$ are arbitrarily small. If we let $B = \alpha P_{\mathcal{X}'} + \beta P_{\mathcal{Y}'}$, the orthogonality between $\mathcal{X}$ and $\mathcal{Y}$ implies

$$\begin{aligned}
\|A - B\|_2 &= \|\alpha(P_{\mathcal{X}} - P_{\mathcal{X}'}) + \beta(P_{\mathcal{Y}} - P_{\mathcal{Y}'})\|_2\\
&= \max\{\alpha \|P_{\mathcal{X}} - P_{\mathcal{X}'}\|_2\,;\ \beta \|P_{\mathcal{Y}} - P_{\mathcal{Y}'}\|_2\} = \max\{\alpha \sin(\theta_{\mathcal{X}})\,;\ \beta \sin(\theta_{\mathcal{Y}})\}\,,
\end{aligned}$$

and thus the matrices B constructed this way can be arbitrarily close to A by making $\theta_{\mathcal{X}}$ and $\theta_{\mathcal{Y}}$ small. Next, notice that if $\mathcal{S}_A \in \Lambda\text{-adm}_h(A)$, then $\mathcal{S}_A = \mathcal{X} \oplus \mathcal{D}$ for some $(h-j)$-dimensional subspace $\mathcal{D} \subseteq \mathcal{Y}$. Similarly, if $\mathcal{S}_B \in \Lambda\text{-adm}_h(B)$ then $\mathcal{S}_B = \mathcal{X}' \oplus \mathcal{D}'$ for some $(h-j)$-dimensional subspace $\mathcal{D}' \subseteq \mathcal{Y}'$. By direct computation,

$$\theta(\mathcal{S}_A,\, \mathcal{S}_B) = \theta(\mathcal{X} \oplus \mathcal{D},\, \mathcal{X}' \oplus \mathcal{D}') = (\,\theta(\mathcal{X},\, \mathcal{X}'),\, \theta(\mathcal{D},\, \mathcal{D}')\,)^{\downarrow}$$

and thus, for a fixed $\mathcal{S}_B \in \Lambda\text{-adm}(B)$ where $\mathcal{S}_B = \mathcal{X}' \oplus \mathcal{D}'$ as above we have

$$\begin{aligned}
\inf_{\mathcal{S}_A \in \Lambda\text{-adm}(A)} \|\sin\Theta(\mathcal{S}_A,\, \mathcal{S}_B)\|_2 &= \inf_{\substack{\mathcal{D}\subseteq\mathcal{Y} \\ \dim\mathcal{D}=h-j}} \max\{\|\sin\Theta(\mathcal{X},\, \mathcal{X}')\|_2\,,\ \|\sin\Theta(\mathcal{D},\, \mathcal{D}')\|_2\} \\
&= \max\left\{ \|\sin\Theta(\mathcal{X},\, \mathcal{X}')\|_2\,,\ \inf_{\substack{\mathcal{D}\subseteq\mathcal{Y} \\ \dim\mathcal{D}=h-j}} \|\sin\Theta(\mathcal{D},\, \mathcal{D}')\|_2 \right\} \\
&= \max\{\sin\theta_{\mathcal{X}}\,,\ \|\sin\Theta(P_{\mathcal{Y}}(\mathcal{D}'),\, \mathcal{D}')\|_2\}\,,
\end{aligned}$$

where we have used the fact that $\theta_{\mathcal{Y}} < \pi/2$. Also notice that $\theta(P_{\mathcal{Y}}(\mathcal{D}'),\, \mathcal{D}') = \theta(\mathcal{Y},\, \mathcal{D}')$. Then,

$$\begin{aligned}
d_H(\Lambda\text{-adm}_h(A),\, \Lambda\text{-adm}_h(B)) &= \sup_{\mathcal{S}_B \in \Lambda\text{-adm}(B)}\ \inf_{\mathcal{S}_A \in \Lambda\text{-adm}(A)} \|\sin\Theta(\mathcal{S}_A,\, \mathcal{S}_B)\|_2 \\
&= \sup_{\substack{\mathcal{D}'\subseteq\mathcal{Y}' \\ \dim\mathcal{D}'=h-j}} \max\{\sin\theta_{\mathcal{X}}\,,\ \|\sin\Theta(\mathcal{Y},\, \mathcal{D}')\|_2\} \\
&= \max\left\{ \sin\theta_{\mathcal{X}}\,,\ \sup_{\substack{\mathcal{D}'\subseteq\mathcal{Y}' \\ \dim\mathcal{D}'=h-j}} \|\sin\Theta(\mathcal{Y},\, \mathcal{D}')\|_2 \right\} \\
&= \max\left\{\sin\theta_{\mathcal{X}}\,,\ \sin\theta_{\mathcal{Y}}\right\}\,.
\end{aligned}$$

We now take $0 < \alpha\, \sin\theta_{\mathcal{X}} < \beta\, \sin\theta_{\mathcal{Y}}$ and make $\theta_{\mathcal{X}}$ and $\theta_{\mathcal{Y}}$ become arbitrarily small. The previous facts together with Theorem 4.12 show that

$$\frac{1}{\beta} \le \kappa[f](A) \le \frac{1}{\lambda_j - \lambda_{j+1}} + \frac{1}{\lambda_k - \lambda_{k+1}} = \frac{1}{\alpha - \beta} + \frac{1}{\beta} = \frac{1}{\beta}\ \frac{\alpha}{\alpha - \beta}\,.$$

For example, if we set $\alpha = 12$ and $\beta = 1$ then, $0 \le (\frac{1}{\lambda_j - \lambda_{j+1}} + \frac{1}{\lambda_k - \lambda_{k+1}}) - \kappa[f](A) \le 0.1\,.$

References

[1] F. Arrieta Zuccalli, P. Massey, D. Stojanoff. Block subspace expansions for eigenvalues and eigenvectors approximation (submitted): available at https://arxiv.org/pdf/2411.14578.

[2] R. Bhatia. Matrix analysis. Graduate Texts in Mathematics, 169. Springer-Verlag, New York, 1997.

[3] C. Davis, W.M. Kahan: The rotation of eigenvectors by a perturbation. III. SIAM J. Numer. Anal. 7(1), 1-46 (1970).

[4] P. Drineas, I.C.F. Ipsen. Low-rank matrix approximations do not need a singular value gap. SIAM J. Matrix Anal. Appl. 40 (2019), no. 1, 299-319.

[5] P. Drineas, I.C.F. Ipsen, E. Kontopoulou, M. Magdon-Ismail. Structural convergence results for approximation of dominant subspaces from block Krylov spaces. SIAM J. Matrix Anal. Appl. 39 (2018), no. 2, 567-586.

[6] M. Gu. Subspace iteration randomization and singular value problems. SIAM J. Sci. Comput. 37 (2015), no. 3, A1139-A1173.

[7] P. Massey. Dominant subspace and low-rank approximations from block Krylov subspaces without a prescribed gap. Linear Algebra Appl. 708 (2025), 112-149.

[8] P. Massey. Admissible subspaces and the subspace iteration method. BIT 64 (2024), no. 1, Paper No. 12, 36 pp.

[9] P. Massey, D. Stojanoff, S. Zarate. Majorization bounds for Ritz values of self-adjoint matrices. SIAM J. Matrix Anal. Appl. 41 (2020), no. 2, 554-572.

[10] Cameron Musco and Christopher Musco. Randomized Block Krylov Methods for Stronger and Faster Approximate Singular Value Decomposition. 2015. url: https://arxiv.org/abs/1504.05477.

[11] Y. Nakatsukasa. Accuracy of singular vectors obtained by projection-based SVD methods. BIT 57 (2017), no. 4, 1137-1152.

[12] Y. Nakatsukasa. Sharp error bounds for Ritz vectors and approximate singular vectors. Math. Comp. 89 (2020), no. 324, 1843.1866.

[13] Y.A. Neretin. On Jordan angles and the triangle inequality in Grassmann manifolds. Geom. Dedicata 86 (2001), no. 1-3, 81-92.

[14] L. Qiu, Y. Zhang, C.K. Li. Unitarily invariant metrics on the Grassmann space. SIAM J. Matrix Anal. Appl. 27 (2005), no. 2, 507-531.

[15] J.R. Rice. A theory of condition. SIAM J. Numer. Anal. 3 (1966), 287-310.

[16] Y. Saad. Numerical methods for large eigenvalue problems. Classics in Applied Mathematics, 66. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2011.

[17] A.K. Saibaba. Randomized subspace iteration: analysis of canonical angles and unitarily invariant norms. SIAM J. Matrix Anal. Appl. 40 (2019), no. 1, 23-48.

[18] G.W. Stewart, Ji-G. Sun. Matrix perturbation theory. Computer Science and Scientific Computing. Academic Press, Inc., Boston, MA, 1990.

[19] Ji-G. Sun. Perturbation analysis of singular subspaces and deflating subspaces. Numer. Math. 73 (1996), no. 2, 235-263.

[20] N. Vannieuwenhoven. The condition number of singular subspaces, revisited. Linear Algebra Appl. 687 (2024), 157-180.

[21] S. Wang, Z. Zhang, T. Zhang. Improved analyses of the randomized power method and block Lanczos method. arXiv preprint arXiv:1508.06429, 2015.

[22] M. Zhou, M. Argentati, A.V. Knyazev, K. Neymeyr. Sharp majorization-type cluster robust bounds for block filters and eigensolvers. SIAM J. Matrix Anal. Appl. 44 (2023), no. 4, 1852-1878.

[23] M. Zhou, A.V. Knyazev, K. Neymeyr. Angle-free cluster-robust Ritz value bounds for restarted block eigensolvers. Numer. Linear Algebra Appl. 32 (2025), no. 1, Paper No. e2607, 17 pp.

[24] P. Zhu, A.V. Knyazev. Angles between subspaces and their tangents. J. Numer. Math. 21 (2013), no. 4, 325-340.