Dominant subspace and low-rank approximations from block Krylov subspaces without a gap

Pedro Massey*

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

Abstract

In this work we obtain results related to the approximation of h-dimensional dominant subspaces and low rank approximations of matrices $\mathbf{A} \in \mathbb{K}^{m\times n}$ (where $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$) in case there is no singular gap, i.e. if $\sigma_h = \sigma_{h+1}$ (where $\sigma_1 \geq \ldots \geq \sigma_p \geq 0$ denote the singular values of $\mathbf{A}$, and $p = \min\{m, n\}$). In order to do this, we describe in a convenient way the class of h-dimensional right (respectively left) dominant subspaces. Then, we show that starting with a matrix $\mathbf{X} \in \mathbb{K}^{n\times r}$ with $r \geq h$ satisfying a compatibility assumption with some h-dimensional right dominant subspace, block Krylov methods produce arbitrarily good approximations for both problems mentioned above. Our approach is based on recent work by Drineas, Ipsen, Kontopoulou and Magdon-Ismail on approximation of structural left dominant subspaces; but instead of exploiting a singular gap at h (which is zero in this case) we exploit the nearest existing singular gaps.

Key words. Singular value decomposition, principal angles, dominant subspaces, low rank approximation.

AMS subject classifications. 15A18, 15A42, 65F15.

Contents

*e-mail address: massey@mate.unlp.edu.ar

1 Introduction

Low rank matrix approximation is a central problem in numerical linear algebra (see [19]). It is well known that truncated singular value decompositions (SVD) of a matrix $\mathbf{A} \in \mathbb{K}^{m\times n}$ (for $\mathbb{K} = \mathbb{R}$ or $\mathbb{K} = \mathbb{C}$) produce optimal solutions to this problem ([2, 10, 14, 19]). Indeed, let $\mathbf{A} = \mathbf{U\Sigma V}^*$ be a SVD and let $\sigma_1 \geq \ldots \geq \sigma_p \geq 0$ be the singular values of $\mathbf{A}$, where $p = \min\{m, n\}$. Given $1 \leq h \leq \operatorname{rank}(\mathbf{A})$, recall that the truncated SVD of $\mathbf{A}$ is given by $\mathbf{A}_h = \mathbf{U}_h\mathbf{\Sigma}_h\mathbf{V}_h^*$, where the columns of $\mathbf{U}_h$, and $\mathbf{V}_h$ are the top h columns of $\mathbf{U}$, and $\mathbf{V}$ respectively, and $\mathbf{\Sigma}_h$ is the diagonal matrix with main diagonal given by $\sigma_1, \ldots, \sigma_h$. In this case, we have that $\|\mathbf{A} - \mathbf{A}_h\|_{2,F} \leq \|\mathbf{A} - \mathbf{B}\|_{2,F}$ for every $\mathbf{B} \in \mathbb{K}^{m\times n}$ with $\operatorname{rank}(\mathbf{B}) \leq h$, where $\|\cdot\|_{2,F}$ stands for spectral and Frobenius norms respectively. Nevertheless, it is well known that (in general) computation of SVD of a matrix is expensive. In turn, this last fact is one the motivations for the efficient numerical computation of approximations of truncated SVD of matrices [5, 11, 12, 18, 19, 21, 22].

A closer look at this optimal approximations show that they can be described as $\mathbf{A}_h = \mathbf{P}_h\mathbf{A}$, where $\mathbf{P}_h \in \mathbb{K}^{m\times m}$ is the orthogonal projection onto the subspace $\mathcal{U}_h$, spanned by the top h columns of $\mathbf{U}$. Hence, one of the main strategies for computing low rank approximations is the computation of h-dimensional subspaces $\mathcal{S}' \subset \mathbb{K}^m$ that, in a sense, approximate the subspaces $\mathcal{U}_h$.

There are several methods for efficient computation of low rank approximations based on the construction of convenient h-dimensional subspaces $\mathcal{S}'$ (equivalently, orthonormal sets of h vectors). Among others, implementations of the power and block Krylov methods have become very popular. The applications of these methods are based on deterministic and randomized approaches. Randomized methods [6, 7, 11, 12, 18] are typically based on a random $n \times r$ matrix $\mathbf{X}$ (a test matrix) and consider the random subspace $R(\mathbf{X}) \subset \mathbb{K}^n$ given by the range of $\mathbf{X}$. One of the advantages of this approach is that it is possible to prove that, with high probability, $\mathbf{X}$ satisfies compatibility assumptions with the structure of $\mathbf{A}$, regardless of the particular choice of $\mathbf{A}$. Thus, when the power or block Krylov methods are applied, it is possible to extract in an efficient way subspaces $\mathcal{S}' \subset \mathbb{K}^m$ such that $\mathbf{PA}$ are good low rank approximates of $\mathbf{A}$, where $\mathbf{P}$ denotes the orthogonal projection onto $\mathcal{S}'$. Moreover, low rank approximations of the form $\mathbf{PA}$ have the advantage of being numerically stable under several types of perturbations [4].

Yet, the range $R(\mathbf{PA}) \subset \mathbb{K}^m$ might actually not be close to the subspaces $\mathcal{U}_h \subset \mathbb{K}^m$; here, the distance between subspaces is measured in terms of the principal angles between them. Indeed, in case $\sigma_h > \sigma_{h+1}$ (so that the subspace $\mathcal{U}_h$ is uniquely determined) the angular distance between $\mathcal{U}_h$ and $\mathcal{S}'$ is bounded by the approximation error divided by σ_h (which is expected to be small, by the nature of our problem). On the other hand, if our choice $\mathcal{S}'$ is close to $\mathcal{U}_h$ then we can bound the approximation error in approximating $\mathbf{A}$ by $\mathbf{PA}$ by the sum of the optimal approximation error (i.e. the error in approximating $\mathbf{A}$ by $\mathbf{A}_h$) plus the spectral norm $\|\mathbf{A}\|_2$ times the angular distance between $\mathcal{S}'$ and $\mathcal{U}_h$ (for both estimates see [4]).

Thus, in order to derive low rank approximations that also share some other features with $\mathbf{A}$, it seems natural to consider subspaces $\mathcal{S}'$ that are close to the subspaces $\mathcal{U}_h$. Moreover, these subspaces can be used to construct approximated truncated SVD and are also relevant in the study of principal component analysis [15]. But, as opposed to the low rank approximation problem, there is an obstruction to consider the approximation of the subspaces $\mathcal{U}_h$, namely that they are not uniquely determined unless there is a singular gap $\sigma_h > \sigma_{h+1}$. In case there is a singular gap, then $\mathcal{U}_h$ - also called left dominant subspace - has structural relations with $\mathbf{A}$, and there are several positive results (both deterministic and randomized). Indeed, subspaces $\mathcal{S}'$ that are close to $\mathcal{U}_h$ can be obtained by the power and block Krylov methods and an initial matrix $\mathbf{X} \in \mathbb{K}^{n\times r}$ for $r \geq h$, that verifies some compatibility assumptions in terms of $\mathbf{V}_h^*$ [5, 21]. In case there is no singular gap, there are also positive results related to approximation of $\mathcal{U}_h$ in terms of an initial matrix $\mathbf{X} \in \mathbb{K}^{n\times r}$ for $r \geq h$ large enough (also satisfying compatibility conditions with $\mathbf{A}$) [11].

In this work, we adopt a deterministic approach and adapt some of the main ideas of [5], to deal with the approximation of $\mathcal{U}_h$, in case there is no singular gap (i.e. $\sigma_h = \sigma_{h+1}$). In order to do this, we consider an initial matrix $\mathbf{X} \in \mathbb{K}^{n\times r}$ that verifies some compatibility assumptions with $\mathbf{A}$, that can always be achieved with $r = h$ (i.e. for a minimal choice of r). Notice that in this case $\mathcal{U}_h$ (that is, the subspace spanned by the top h columns of $\mathbf{U}$, that corresponds to a SVD of $\mathbf{A}$) is not uniquely determined; indeed, it is possible to parametrize this class of subspaces (for all possible SVD's) in terms of the Grassmannian manifold of subspaces of a fixed dimension h' inside a d-dimensional space, with $1 \leq h' < d$. We call these spaces h-dimensional left dominant subspaces of $\mathbf{A}$ (in case there is a singular gap, then this class reduces to a uniquely well-determined left dominant subspace $\mathcal{U}_h$).

Our approach is based on enclosing $\sigma_j > \sigma_h = \sigma_k > \sigma_{k+1}$ in such a way that $j < h$ is the largest index ℓ such that $\sigma_\ell < \sigma_h$ (if such index exists or otherwise $j = 0$) and $h \leq k$ is the largest index ℓ such that $\sigma_h = \sigma_\ell$. Hence, by construction, we have singular gaps $\sigma_j > \sigma_{j+1} = \sigma_h$ and $\sigma_h = \sigma_k > \sigma_{k+1}$. Our analysis is based on both these singular gaps; indeed, these gaps appear explicitly in the upper bounds related to the convergence analysis of block Krylov methods. Nevertheless, these gaps and indexes are not related to the assumptions on the starting matrix $\mathbf{X}$; indeed, we only ask $\mathbf{X}$ to satisfy that $R(\mathbf{V}_h^*\mathbf{X}) = R(\mathbf{V}_h^*)$, for some SVD of $\mathbf{A}$. As before, in case there is no singular gap, then there is a (continuum) class of such matrices $\mathbf{V}_h^*$; we consider this as an advantage, since we are allowed to consider any such matrix to test our assumptions. On the other hand, since our assumptions are based on non-structural choices $\mathbf{V}_h^*$, the analysis requires the use of some adapted tools, tailored for our present setting. Still, we prove that in the previous case, we get arbitrarily good h-dimensional approximations of left and right h-dominant subspaces in terms of block Krylov subspaces. Moreover, we show that block Krylov spaces can be used to compute (in terms of a proto-algorithm) low rank approximations of $\mathbf{A}$ so that the approximation errors are arbitrarily close to the optimal approximation errors corresponding to truncated SVD as above. We hope that our approach can provide with some new insights on the relations between low rank matrix approximations and dominant subspace approximations, in terms of this extended notion of dominant subspace.

The paper is organized as follows. In Section 2 we first give a formal description of the class of dominant subspaces of a matrix (whether there is a singular gap or not) and describe some of the main problems considered in this work. In Sections 2.2 and 2.3 we state our main results without proofs. In Section 3 we include some remarks and comments related to our approach. In Section 4 we recall some tools from matrix analysis and some of the main results in [5]. In Section 5 we present the proof of the results described in Section 2; some of these proof require some technical facts that we consider in Section 6 (Appendix).

2 Main results

In this section we first describe the class of dominant subspaces of a matrix and the main problems considered in this work. Then, we state our main results related to dominant subspace approximations and low rank matrix approximations in terms of block Krylov subspaces. The proofs of these results are considered in Section 5.

2.1 Setting the context and problems

We begin with a formal description of the class of dominant subspaces of a matrix, without assuming a singular gap. Then, we describe the context and main problems considered in this work.

Let $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^*$ be a full SVD for $\mathbf{A} \in \mathbb{K}^{m\times n}$, where $\mathbb{K} = \mathbb{R}$ or $\mathbb{K} = \mathbb{C}$, $\mathbf{\Sigma} \in \mathbb{R}^{m\times n}$ and $\mathbf{U} \in \mathbb{K}^{m\times m}$ and $\mathbf{V} \in \mathbb{K}^{n\times n}$ are unitary (orthogonal when $\mathbb{K} = \mathbb{R}$) matrices. In this case $\mathbf{\Sigma}$ is a diagonal matrix, with diagonal entries given by the singular values $\sigma_1 := \sigma_1(\mathbf{A}) \geq \ldots \geq \sigma_p := \sigma_p(\mathbf{A}) \geq 0$, where $p = \min\{m, n\}$. In what follows we let $\mathbf{u}_j$ (respectively $\mathbf{v}_j$) denote the columns of $\mathbf{U}$ (respectively of $\mathbf{V}$).

Given $1 \leq h \leq m$, we define the subspaces $\mathcal{U}_h = \mathrm{Span}\{\mathbf{u}_1, \ldots, \mathbf{u}_h\} \subset \mathbb{K}^m$; similarly, if $1 \leq h \leq n$, we let $\mathcal{V}_h = \mathrm{Span}\{\mathbf{v}_1, \ldots, \mathbf{v}_h\} \subset \mathbb{K}^n$. Notice that in case $1 \leq h \leq \mathrm{rank}(\mathbf{A})$ and $\sigma_h > \sigma_{h+1}$, then the spaces $\mathcal{U}_h$ and $\mathcal{V}_h$ do not depend on our particular choice of SVD for $\mathbf{A}$.

Let $\mathcal{S}' \subset \mathbb{K}^m$ be a subspace of dimension $1 \leq h \leq \mathrm{rank}(\mathbf{A}) \leq p$. We say that $\mathcal{S}'$ is a *left dominant subspace* for $\mathbf{A}$ if $\mathcal{S}'$ admits an orthonormal basis $\{\mathbf{w}_1, \ldots, \mathbf{w}_h\}$ such that $\mathbf{A}\mathbf{A}^*\mathbf{w}_i = \sigma_i^2\, \mathbf{w}_i$, for $1 \leq i \leq h$. Equivalently, $\mathcal{S}'$ is a left dominant subspace for $\mathbf{A}$ if the h largest singular values of $\mathbf{P}_{\mathcal{S}'}\mathbf{A}$ are $\sigma_1 \geq \ldots \geq \sigma_h$. Hence, in this case we have that

$$\|\mathbf{P}_{\mathcal{S}'}\mathbf{A} - \mathbf{A}\| \leq \|\mathbf{Q}\mathbf{A} - \mathbf{A}\|$$

for every projection $\mathbf{Q} \in \mathbb{K}^{m\times m}$ with $\mathrm{rank}(\mathbf{Q}) = h$ and every unitarily invariant norm; that is, $\mathbf{P}_{\mathcal{S}'}\mathbf{A}$ is an optimal low-rank approximation of $\mathbf{A}$ (see [2, Section IV.3]).

On the other hand, we say that $\mathcal{S} \subset \mathbb{K}^n$ is a *right dominant subspace* for $\mathbf{A}$ if $\mathcal{S}$ admits an orthonormal basis $\{\mathbf{z}_1, \ldots, \mathbf{z}_h\}$ such that $\mathbf{A}^*\mathbf{A}\mathbf{z}_i = \sigma_i^2\mathbf{z}_i$, for $1 \leq i \leq h$. Similar remarks apply also to right dominant subspaces. It is interesting to notice that the class of h-dimensional left dominant subspaces of $\mathbf{A}$ coincides with the class of h-dimensional right dominant subspaces of $\mathbf{A}^*$; in what follows we will make use of this fact.

Consider the previous notation. As mentioned above, in case $\sigma_h > \sigma_{h+1}$ then the left (respectively right) dominant space for $\mathbf{A}$ of dimension h is uniquely determined and given by $\mathcal{U}_h$ (respectively by $\mathcal{V}_h$). On the other hand, if $\sigma_h = \sigma_{h+1}$ then we have a continuum class of h-dimensional left dominant subspaces: indeed, let $0 \leq j = j(h) < h < k = k(h)$ be given by

$$j(h) = \max\{0 \leq j < h \ : \ \sigma_j > \sigma_h\}$$

where we set $\sigma_0 = \infty$ and

$$k = k(h) = \max\{1 \leq j \leq \mathrm{rank}(\mathbf{A}) \ : \ \sigma_j = \sigma_h\}\,.$$

If we further let $\mathcal{U}_0 = \{0\}$ then, it is straightforward to check that an h-dimensional subspace $\mathcal{S}'$ is a left dominant subspace for $\mathbf{A}$ if and only if there exists an $(h-j)$-dimensional subspace $\mathcal{U} \subset \mathcal{U}_k \ominus \mathcal{U}_j \ (= \mathcal{U}_k \cap \mathcal{U}_j^\perp)$ such that

$$\mathcal{S}' = \mathcal{U}_j \oplus \mathcal{U} \, .$$

Therefore, we have a natural parametrization of h-dimensional left dominant subspaces in terms of the Grassmann manifold of $(h-j)$-dimensional subspaces of $\mathcal{U}_k \ominus \mathcal{U}_j$. We point out that the proof of Theorem 2.1 below is based on this simple fact.

In what follows we use that if $\mathcal{S}'$ is a left dominant subspace of dimension $h \geq 1$ then there exists a full SVD, $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^*$ such that $\mathcal{S}' = \mathcal{U}_h$, i.e. the subspace spanned by the top h columns of $\mathbf{U}$; and a similar fact also holds for right dominant subspaces.

Consider $\mathbf{A}$ as above and a matrix $\mathbf{X} \in \mathbb{K}^{n\times r}$. From $\mathbf{A}$ and $\mathbf{X}$ we construct the block Krylov space $\mathcal{K}_q$, for $1 \leq q$, that is

$$\mathcal{K}_q = \mathcal{K}_q(\mathbf{A}, \mathbf{X}) = R(\ \mathbf{AX} \quad (\mathbf{AA}^*)\mathbf{AX} \quad \ldots \quad (\mathbf{AA}^*)^q\mathbf{AX}\ ) \subset \mathbb{K}^m \, , \tag{1}$$

where $R(\mathbf{B})$ denotes the range of a matrix $\mathbf{B}$. We are interested in obtaining h-dimensional subspaces of $\mathcal{K}_q$ that are close to h-dimensional left dominant subspaces for $\mathbf{A}$. Moreover, in this case we are interested in computing low rank approximations of $\mathbf{A}$ from appropriate subspaces of $\mathcal{K}_q$. It is clear that we must impose some kind of compatibility assumption on $\mathbf{A}$ and $\mathbf{X}$ in order to achieve this goal.

The previous problems have been considered in several research papers, both based on deterministic as well as probabilistic methods, taking into consideration numerical implementations in some cases. In this work we adopt a deterministic approach. In case the matrix $\mathbf{A}$ verifies a singular gap condition i.e. $\sigma_h > \sigma_{h+1}$ these problems have been recently solved in [5] (see also the references therein). Thus, in our work we focus on the case in which there is no singular gap (i.e. $\sigma_h = \sigma_{h+1}$). In what follows we show that under a compatibility assumption between (the range of) $\mathbf{X}$ and *some* h-dimensional right dominant subspace of $\mathbf{A}$, the block Krylov methods provide with good approximations of dominant subspaces; in this context, proximity between subspaces is measured in terms of the principal angles between subspaces. Further, we revisit an algorithmic procedure considered in [5] and show that this process constructs arbitrarily good low rank approximations of $\mathbf{A}$ based on block Krylov subspaces $\mathcal{K}_q$, for sufficiently large $q \geq 1$.

2.2 Approximation of dominant subspaces by block Krylov spaces

Notation and terminology. As before, let $\mathbf{A} \in \mathbb{K}^{m\times n}$ with singular values $\sigma_1 \geq \ldots \geq \sigma_p$, for $p = \min\{m, n\}$. Given $1 \leq h \leq \operatorname{rank}(\mathbf{A}) \leq p$, we let $0 \leq j(h) < h$ be given by

$$j = j(h) = \max\{0 \leq j < h \ : \ \sigma_j > \sigma_h\}$$

where we set $\sigma_0 = \infty$ and

$$k = k(h) = \max\{1 \leq j \leq \operatorname{rank}(\mathbf{A}) \, : \, \sigma_j = \sigma_h\} \, .$$

Since $h \leq \operatorname{rank}(\mathbf{A})$, we get that $\sigma_k > 0$. As mentioned in the preceding section, from now on we focus on the case when $h < k$ (i.e. when $\sigma_h = \sigma_{h+1}$). In case $1 \leq k < \operatorname{rank}(\mathbf{A}) \leq p$ then we consider the partitioning

$$\mathbf{\Sigma} = \begin{pmatrix} \mathbf{\Sigma}_k & \\ & \mathbf{\Sigma}_{k,\perp} \end{pmatrix} \ , \quad \mathbf{U} = \begin{pmatrix} \mathbf{U}_k & \mathbf{U}_{k,\perp} \end{pmatrix} \ , \quad \mathbf{V} = \begin{pmatrix} \mathbf{V}_k & \mathbf{V}_{k,\perp} \end{pmatrix} \, . \tag{2}$$

Notice that in case $\operatorname{rank}(\mathbf{A}) \leq k \leq p$ then we can also consider a partition as in Eq. (2), but we get that $\mathbf{\Sigma}_{k,\perp} = 0$ and the partition, as far as we are concerned, becomes trivial.

Given $\mathbf{X} \in \mathbb{K}^{n\times r}$ we say that $(\mathbf{A}, \mathbf{X})$ is *h-compatible* if there is an h-dimensional right dominant subspace $\mathcal{S} \subset \mathbb{K}^n$ for $\mathbf{A}$, with

$$\mathbf{\Theta}(\mathcal{S}, R(\mathbf{X})) < \frac{\pi}{2}\,\mathbf{I}_h\,,$$

where $\mathbf{\Theta}(\mathcal{S}, R(\mathbf{X})) \in \mathbb{R}^{h\times h}$ denotes a diagonal matrix, with diagonal entries given by the principal angles between $\mathcal{S}$ and $R(\mathbf{X})$ (see Section 4.2). Equivalently, $(\mathbf{A}, \mathbf{X})$ is h-compatible if $\dim(\mathbf{X}^*\mathcal{S}) = h$, for some h-dimensional right dominant subspace $\mathcal{S}$.

In what follows we fix $1 \leq h \leq \operatorname{rank}(\mathbf{A}) \leq p = \min\{m, n\}$ and we let let $j = j(h) < h \leq k = k(h)$ be defined as above. Also, we fix $\mathbf{X} \in \mathbb{K}^{n\times r}$ such that $(\mathbf{A}, \mathbf{X})$ is h-compatible.

We can now state our main results. We begin with the next technical result that will allow us to show that block Krylov methods produce good approximations of right and left dominant subspaces.

Theorem 2.1. *Let $\phi(x)$ be a polynomial of degree at most $2q+1$ with odd powers only, such that $\phi(\sigma_1) \geq \ldots \geq \phi(\sigma_k) > 0$. Let $\mathcal{K}_q = \mathcal{K}_q(\mathbf{A}, \mathbf{X})$ denote the block Krylov subspace. Then, there exists an h-dimensional left dominant subspace $\mathcal{S}'$ for $\mathbf{A}$ such that*

$$\begin{aligned}\|\sin \mathbf{\Theta}(\mathcal{K}_q, \mathcal{S}')\|_{2,F} \quad &\leq \quad 4\,\|\sin \mathbf{\Theta}(R(\mathbf{V}_k^*\mathbf{X}), \mathbf{V}_k^*\mathcal{V}_j)\|_{2,F} + \\ &\qquad \|\phi(\mathbf{\Sigma}_{k,\perp})\|_2 \|\phi(\mathbf{\Sigma}_k)^{-1}\|_2 \|\mathbf{V}_{k,\perp}^*\mathbf{X}(\mathbf{V}_k^*\mathbf{X})^\dagger\|_{2,F}\,.\end{aligned}$$

Moreover, we have the inequality

$$\mathbf{\Theta}(R(\mathbf{V}_k^*\mathbf{X}), \mathbf{V}_k^*\mathcal{V}_j) \leq \mathbf{\Theta}(R(\mathbf{X}), \mathcal{V}_j)\,.$$

In case $j = 0$ (respectively $k = \operatorname{rank}(\mathbf{A})$) the first term (respectively the second term) should be omitted in the previous upper bound.

Proof. See section 5.1. □

Remark 2.2. Consider the notation and conditions in Theorem 2.1. In case $k = h$, i.e. $\sigma_h > \sigma_{h+1}$ then $\mathcal{S} = \mathcal{V}_h$ is uniquely determined. On the other hand, the fact that $\mathbf{\Theta}(\mathcal{S}, R(\mathbf{X})) < \frac{\pi}{2}\,\mathbf{I}_h$ implies that $R(\mathbf{V}_k^*\mathbf{X}) = \mathbb{K}^k = R(\mathbf{V}_k^*)$. Hence, $\mathbf{\Theta}(R(\mathbf{V}_k^*\mathbf{X}), \mathbf{V}_k^*\mathcal{V}_j) = 0$ and we recover [5, Theorem 2.1] (or see Theorem 4.1 below) from the first estimate in Theorem 2.1. △

In what follows, given a matrix $\mathbf{Z}$ we let $\mathbf{Z}^\dagger$ denote its Moore-Penrose pseudo-inverse. We further consider the notation

$$\gamma_k = \frac{\sigma_k - \sigma_{k+1}}{\sigma_{k+1}} > 0 \quad \text{and} \quad \Delta(\mathbf{W}, q, k)_{2,F} = 4\,\frac{\|\mathbf{V}_{k,\perp}^*\mathbf{W}(\mathbf{V}_k^*\mathbf{W})^\dagger\|_{2,F}}{2^{(2q+1)\,\min\{\sqrt{\gamma_k}\,,\,1\}}}\,, \tag{3}$$

where $\mathbf{W} \in \mathbb{K}^{n\times \ell}$, for some $\ell \geq 1$ and $q \geq 1$.

Corollary 2.3. *Let $\mathcal{K}_q = \mathcal{K}_q(\mathbf{A}, \mathbf{X})$ denote the block Krylov subspace. Then, there exists an h-dimensional left dominant subspace $\mathcal{S}'$ for $\mathbf{A}$ such that*

$$\|\sin \mathbf{\Theta}(\mathcal{K}_q, \mathcal{S}')\|_{2,F} \quad \leq \quad 4\,\|\sin \mathbf{\Theta}(R(\mathbf{X}), \mathcal{V}_j)\|_{2,F} + \Delta(\mathbf{X}, q, k)_{2,F}\,\frac{\sigma_{k+1}}{\sigma_k}\,.$$

In case $j = 0$ (respectively $k = \operatorname{rank}(\mathbf{A})$) the first term (respectively the second term) should be omitted in the previous upper bound.

Proof. See section 5.1. □

Theorem 2.1 and Corollary 2.3 involve the matrix $\mathbf{U}_{k,\perp}^*\mathbf{Y}(\mathbf{U}_k^*\mathbf{Y})^\dagger$ for an isometry $\mathbf{Y}$ and a (partitioned) unitary matrix $(\mathbf{U}_k\ \ \mathbf{U}_{k,\perp})$. This type of matrix has already appeared in previous analysis of (randomized) algorithms (see [5, 7, 8, 9, 17]). In [23] there is a detailed description of the relation between the singular values of this matrix and the tangents of the principal angles between associated subspaces. The following result contains a similar description to that given in [5, 23].

Proposition 2.4. *Let $(\mathbf{U}_k\ \ \mathbf{U}_{k,\perp}) \in \mathbb{K}^{m\times m}$ be a unitary matrix and let $\mathbf{Y}$ be an isometry into $\mathbb{K}^m$. Let $\mathcal{Y} = R(\mathbf{Y})$, $\mathcal{U}_k = R(\mathbf{U}_k) \subset \mathbb{K}^m$ and let $\tilde{\mathcal{Y}} = \mathcal{Y} \ominus (\mathcal{Y} \cap \mathcal{U}_k^\perp)$. Then, $\dim(\tilde{\mathcal{Y}}) \leq k$ and we have that*

$$\|\mathbf{U}_{k,\perp}^*\mathbf{Y}(\mathbf{U}_k^*\mathbf{Y})^\dagger\|_{2,F} = \|\tan\mathbf{\Theta}(\tilde{\mathcal{Y}},\mathcal{U}_k)\|_{2,F}\,. \tag{4}$$

Proof. See Section 5.3. □

Remark 2.5. Consider the notation in Corollary 2.3. As a consequence of Proposition 2.4 we get an upper bound for the angles $\mathbf{\Theta}(\mathcal{K}_q,\mathcal{S}')$ in terms of the angles $\mathbf{\Theta}(R(\mathbf{X}),\mathcal{V}_j)$ and $\mathbf{\Theta}(\tilde{\mathcal{K}}_q,\mathcal{U}_k)$, where $\tilde{\mathcal{K}}_q = \mathcal{K}_q \ominus (\mathcal{K}_q \cap \mathcal{U}_k^\perp)$. △

The following result can be regarded as a convenient algorithmic augmentation process of the initial subspace $R(\mathbf{X}) = \mathcal{X} \subset \mathbb{K}^n$; that is, we begin with $\mathbf{X}$ that satisfies a compatibility assumption with *some* h-dimensional right dominant subspace of $\mathbf{A}$ and we construct an associated subspace $\mathcal{K}_{q,t}^* \subset \mathbb{K}^n$ that is (arbitrarily) close to an h-dimensional *right* dominant subspace (also see Remark 2.8 below). We remark that this result plays a central role in the construction of approximate left dominant subspaces and low-rank approximations of $\mathbf{A}$ from block Krylov methods in Theorems 2.9 and 2.11 below.

For the next result we consider the notation in Eq. (3); we further introduce

$$\Delta(\mathbf{X},q,j)_{2,F} = 4\,\frac{\|\mathbf{V}_{j,\perp}^*\mathbf{X}(\mathbf{V}_j^*\mathbf{X})^\dagger\|_{2,F}}{2^{(2q+1)\,\min\{\sqrt{\gamma_j}\,,\,1\}}}\quad,\quad \Delta^*(\mathbf{Y},t,k)_{2,F} = 4\,\frac{\|\mathbf{U}_{k,\perp}^*\mathbf{Y}(\mathbf{U}_k^*\mathbf{Y})^\dagger\|_{2,F}}{2^{(2t+1)\,\min\{\sqrt{\gamma_k}\,,\,1\}}}\,, \tag{5}$$

where $\gamma_j = \frac{\sigma_j-\sigma_{j+1}}{\sigma_{j+1}} > 0$, $\mathbf{Y} \in \mathbb{K}^{m\times\ell}$, for some $\ell \geq 1$ and q, $t \geq 1$.

Theorem 2.6. *Let $\mathcal{K}_q = \mathcal{K}_q(\mathbf{A},\mathbf{X}) \subset \mathbb{K}^m$ and let $\mathbf{Y}_q$ be such that $\mathbf{Y}_q\mathbf{Y}_q^*$ is the orthogonal projection onto $\mathcal{K}_q$. For $t \geq 1$ we let*

$$\mathcal{K}_{q,t}^* = R((\mathbf{A}^*\mathbf{A})\mathbf{X}) + R((\mathbf{A}^*\mathbf{A})^2\mathbf{X}) + \ \cdots + R((\mathbf{A}^*\mathbf{A})^{q+t+1}\mathbf{X}) \subset \mathbb{K}^n\,.$$

Then, there exists an h-dimensional right dominant subspace $\tilde{\mathcal{S}}$ for $\mathbf{A}$ such that

$$\|\sin\mathbf{\Theta}(\mathcal{K}_{q,t}^*,\tilde{\mathcal{S}})\|_{2,F} \leq 4\,\Delta(\mathbf{X},q,j)_{2,F}\,\frac{\sigma_{j+1}}{\sigma_j} + \Delta^*(\mathbf{Y}_q,t,k)_{2,F}\,\frac{\sigma_{k+1}}{\sigma_k}\,. \tag{6}$$

In case $j = 0$ (respectively $k = rank(\mathbf{A})$) the first term (respectively the second term) should be omitted in the previous upper bound.

Proof. See section 5.2. □

With the notation of Theorem 2.6, it seem useful to obtain a uniform upper bound for the norms $\|\mathbf{U}_{k,\perp}^*\mathbf{Y}_q(\mathbf{U}_k^*\mathbf{Y}_q)^\dagger\|_{2,F}$, at least for $q \geq \tilde{q}$, where $\tilde{q}$ is some fixed number. The following result shows that we can obtain such a uniform upper bound.

In what follows we let $\#\{\sigma_1,\ldots,\sigma_h\}$ denote the number of *different* singular values of $\mathbf{A}$ between σ_1 and σ_h.

Theorem 2.7. *Let $q_0 = \#\{\sigma_1, \ldots, \sigma_h\} - 1 < h$, for $h \leq rank(\mathbf{A})$. If $\mathbf{Y}_q$ is an isometry such that $\mathbf{Y}_q\mathbf{Y}_q^*$ is the orthogonal projection onto $\mathcal{K}_q = \mathcal{K}_q(\mathbf{A}, \mathbf{X})$ then*

$$\|\mathbf{U}_{k,\perp}^* \mathbf{Y}_{q'} (\mathbf{U}_k^* \mathbf{Y}_{q'})^\dagger\|_{2,F} \leq \|\mathbf{U}_{k,\perp}^* \mathbf{Y}_q (\mathbf{U}_k^* \mathbf{Y}_q)^\dagger\|_{2,F} \quad \textit{for} \quad q_0 \leq q \leq q' .$$

Proof. See Section 5.3. □

Remark 2.8. Consider the notation in Theorems 2.6 and 2.7. Recall that $q_0 = \#\{\sigma_1, \ldots, \sigma_h\} - 1 < h$. In this case, as a consequence of Theorem 2.7 we have that

$$\Delta^*(\mathbf{Y}_q, t, k)_{2,F} \leq \Delta^*(\mathbf{Y}_{q_0}, t, k)_{2,F} \quad \text{for} \quad q \geq q_0 \quad \text{and} \quad t \geq 1 .$$

Therefore,

$$4\,\Delta(\mathbf{X}, q, j)_{2,F}\, \frac{\sigma_{j+1}}{\sigma_j} + \Delta^*(\mathbf{Y}_q, t, k)_{2,F}\, \frac{\sigma_{k+1}}{\sigma_k} \leq 4\,\Delta(\mathbf{X}, q, j)_{2,F}\, \frac{\sigma_{j+1}}{\sigma_j} + \Delta^*(\mathbf{Y}_{q_0}, t, k)_{2,F}\, \frac{\sigma_{k+1}}{\sigma_k} .$$

This last fact shows that the quantity to the left in the previous inequality (that is the upper bound for $\|\sin \mathbf{\Theta}(\mathcal{K}_{q,t}^*, \tilde{\mathcal{S}})\|_{2,F}$ in Theorem 2.6) becomes arbitrarily small for $q \geq q_0$ and $t \geq 1$ sufficiently large. △

Using the correspondence between left and right dominant subspaces of $\mathbf{A}$ we can derive the existence of (arbitrarily good) approximates of left dominant subspaces obtained from the block Krylov method in case there is no singular gap.

In the next result we consider the notation in Eqs. (3) and (5).

Theorem 2.9. *Given q, $t \geq 1$, consider the block Krylov subspace $\mathcal{K}_{q+t+1} = \mathcal{K}_{q+t+1}(\mathbf{A}, \mathbf{X}) \subset \mathbb{K}^m$. Then, there exists an h-dimensional left dominant subspace $\hat{\mathcal{S}}$ for $\mathbf{A}$ such that*

$$\|\sin \mathbf{\Theta}(\mathcal{K}_{q+t+1}, \hat{\mathcal{S}})\|_{2,F} \leq 4\,\Delta^*(\mathbf{AX}, q, j)_{2,F}\, \frac{\sigma_{j+1}}{\sigma_j} + \Delta(\mathbf{W}_q, t, k)_{2,F}\, \frac{\sigma_{k+1}}{\sigma_k} , \tag{7}$$

where $\mathbf{W}_q$ is such that $\mathbf{W}_q\mathbf{W}_q^*$ is the orthogonal projection onto $\mathcal{K}_q(\mathbf{A}^*, \mathbf{AX}) \subset \mathbb{K}^n$. In case $j = 0$ (respectively $k = rank(\mathbf{A})$) the first term (respectively the second term) should be omitted in the previous upper bound.

Proof. Since the pair $(\mathbf{A}, \mathbf{X})$ is h-compatible, there exists an h-dimensional right dominant subspace $\mathcal{S} \subset \mathbb{K}^n$ for $\mathbf{A}$, such that $\dim(\mathbf{X}^*\mathcal{S}) = h$. Set $\mathbf{Z} = \mathbf{AX}$ and let $\mathcal{S}' = \mathbf{A}\mathcal{S} \subset \mathbb{K}^m$. Hence, $\mathcal{S}'$ is a left dominant subspace for $\mathbf{A}$ and then, a right dominant subspace of $\mathbf{A}^*$ with $\dim \mathcal{S}' = h$. Moreover, $\mathbf{Z}^*\mathcal{S}' = \mathbf{X}^*\mathbf{A}^*\mathbf{A}\mathcal{S} = \mathbf{X}^*\mathcal{S}$, since $\mathbf{A}^*\mathbf{A}\mathcal{S} = \mathcal{S}$. In particular, $\dim \mathbf{Z}^*\mathcal{S}' = h$ and hence $\mathbf{\Theta}(R(\mathbf{Z}), \mathcal{S}') < \frac{\pi}{2}\,\mathbf{I}_h$. Therefore, we can apply Theorem 2.6 to $\mathbf{A}^*$ and $\mathbf{Z}$; in this case, we consider the (auxiliary) subspace

$$\mathcal{K}_{q,t}^*(\mathbf{Z}) = R((\mathbf{AA}^*)\mathbf{Z}) + R((\mathbf{AA}^*)^2\mathbf{Z}) + \cdots + R((\mathbf{AA}^*)^{q+t+1}\mathbf{Z}) \subset \mathbb{K}^n .$$

It is clear that $\mathcal{K}_{q,t}^*(\mathbf{Z}) \subset \mathcal{K}_{q+t+1}$. Then, by Theorem 2.6 there exists an h-dimensional right dominant $\hat{\mathcal{S}}$ for $\mathbf{A}^*$ (and therefore a left dominant subspace for $\mathbf{A}$) such that

$$\begin{aligned} \|\sin \mathbf{\Theta}(\mathcal{K}_{q+t+1}, \hat{\mathcal{S}})\|_{2,F} &\leq \|\sin \mathbf{\Theta}(\mathcal{K}_{q,t}^*(\mathbf{Z}), \hat{\mathcal{S}})\|_{2,F} \\ &\leq 4\,\Delta^*(\mathbf{AX}, q, j)_{2,F}\, \frac{\sigma_{j+1}}{\sigma_j} + \Delta(\mathbf{W}_q, t, k)_{2,F}\, \frac{\sigma_{k+1}}{\sigma_k} , \end{aligned}$$

where we used that if $\mathbf{A} = \mathbf{U\Sigma V}^*$ is a SVD for $\mathbf{A}$ then $\mathbf{A}^* = \mathbf{V\Sigma U}^*$ is a SVD for $\mathbf{A}^*$. □

Remark 2.10. Consider the notation in Theorem 2.9. Let $q_0 = \#\{\sigma_1, \ldots, \sigma_h\} - 1 < h$, where $\#\{\sigma_1, \ldots, \sigma_h\}$ denotes the number of different singular values of $\mathbf{A}^*$ (or equivalently of $\mathbf{A}$) between σ_1 and σ_h. Then, by Theorem 2.7 we get that

$$\Delta(\mathbf{W}_q, t, k)_{2,F} \leq 4 \, \frac{\|\mathbf{V}_{k,\perp}^* \mathbf{W}_{q_0} (\mathbf{V}_k^* \mathbf{W}_{q_0})^\dagger\|_{2,F}}{2^{(2t+1)\min\{\sqrt{\gamma_k}\,,\,1\}}} \quad \text{for} \quad q \geq q_0 \quad \text{and} \quad t \geq 1\,.$$

This last fact shows that the upper bound in Eq. (7) can be made arbitrarily small for sufficiently large q, $t \geq 1$. In this case, if we let $\{\mathbf{v}_1, \ldots, \mathbf{v}_h\} \subset \mathcal{K}_{q+t+1}$ be the principal vectors corresponding to the pair of subspaces $\mathcal{K}_{q+t+1}$ and $\hat{\mathcal{S}}$ (see Section 4.2), then we get that $\|\sin \mathbf{\Theta}(\mathcal{T}, \hat{\mathcal{S}})\|_{2,F}$ can be made arbitrarily small, where $\mathcal{T} = \operatorname{span}\{\mathbf{v}_1, \ldots, \mathbf{v}_h\} \subset \mathcal{K}_{q+t+1}$. △

2.3 Low rank approximations from block Krylov methods

We point out that the upper bound in Theorem 2.9 can be made arbitrarily small (see Remark 2.10) and therefore the corresponding block Krylov subspace contains (arbitrarily good) approximate left dominant subspaces. Still, the previous results do not provide a practical method to compute such approximate dominant subspaces and the corresponding low rank approximations. In this section we revisit an algorithmic scheme considered in [5] that is a practical way to construct such low rank approximations, even without assuming a singular gap. Our approach to deal with this problem is based on approximate right dominant subspaces of a matrix $\mathbf{A}$; indeed, we follow arguments from [21]; in this sense, our technique differs from that in [5].

We first recall the proto-algorithm for low-rank approximation of $\mathbf{A}$ from $\mathcal{K}_q$ from [5].

Algorithm 2.1 Proto-algorithm

Require: $\mathbf{A} \in \mathbb{K}^{m\times n}$, starting guess $\mathbf{X} \in \mathbb{K}^{n\times r}$;
Target: rank $h \leq$ rank $(\mathbf{A})$;
For $\ell \geq 1$ set $\mathbf{K}_\ell = (\mathbf{AX} \quad (\mathbf{AA}^*)\mathbf{AX} \quad \cdots \quad (\mathbf{AA}^*)^\ell \mathbf{AX}) \in \mathbb{K}^{m\times \ell(r+1)}$;
Set the Block dimension $\ell \geq 1$ such that $\dim R(\mathbf{K}_\ell) = d \geq h$.
Ensure: $\hat{\mathbf{U}} \in \mathbb{K}^{m\times h}$ with orthonormal columns
1: Compute an orthonormal basis $\mathbf{U}_K \in \mathbb{K}^{m\times d}$ for $R(\mathbf{K}_\ell)$.
2: Set $\mathbf{W} = \mathbf{U}_K^* \mathbf{A} \in \mathbb{K}^{d\times n}$ and assume that $\operatorname{rank}(\mathbf{W}) \geq h$.
3: Compute $\mathbf{U}_{W,h} \in \mathbb{K}^{d\times h}$ isometry, such that $R(\mathbf{U}_{W,h})$ is a left dominant subspace of $\mathbf{W}$.
4: Return $\hat{\mathbf{U}}_h = \mathbf{U}_K\, \mathbf{U}_{W,h} \in \mathbb{K}^{m\times h}$.

Once the algorithm stops, we describe the output matrix in terms of its columns, $\hat{\mathbf{U}}_h = (\hat{\mathbf{u}}_1, \ldots, \hat{\mathbf{u}}_h) \in \mathbb{K}^{m\times h}$. In this case we set

$$\hat{\mathbf{U}}_i = (\hat{\mathbf{u}}_1, \ldots, \hat{\mathbf{u}}_i) \in \mathbb{K}^{m\times i}\ , \quad \text{for} \quad 1 \leq i \leq h\,. \tag{8}$$

In what follows we fix $1 \leq h \leq \operatorname{rank}(\mathbf{A}) \leq p = \min\{m, n\}$ and we let let $j = j(h) < h \leq k = k(h)$ be defined as in the begining of Section 2.2, and we consider the notation used so far; in particular, $\mathbf{X} \in \mathbb{K}^{n\times r}$ is such that $(\mathbf{A}, \mathbf{X})$ is h-compatible, and we consider the expressions defined in Eqs. (3) and (5). Further, given $1 \leq i \leq h$ we let $\mathbf{A}_i \in \mathbb{K}^{m\times n}$ denote a best rank-i approximation of $\mathbf{A}$ (so that $\|\mathbf{A} - \mathbf{A}_i\|_2 = \sigma_{i+1}$).

Theorem 2.11. *Consider the block Krylov subspace $\mathcal{K}_q = \mathcal{K}_q(\mathbf{A},\mathbf{X}) \subset \mathbb{K}^m$ and let $\mathbf{Y}_q$ be such that $\mathbf{Y}_q\mathbf{Y}_q^*$ is the orthogonal projection onto $\mathcal{K}_q$. Let $q, t \geq 1$ be such that*

$$4\,\Delta(\mathbf{X},q,j)_2\,\frac{\sigma_{j+1}}{\sigma_j} + \Delta^*(\mathbf{Y}_q,t,k)_2\,\frac{\sigma_{k+1}}{\sigma_k} \leq \frac{1}{\sqrt{2}}\,. \tag{9}$$

Set the rank parameter to $q+t+1$ in Algorithm 2.1. Then, for every $1 \leq i \leq h$ we have that

$$\|\mathbf{A} - \hat{\mathbf{U}}_i\hat{\mathbf{U}}_i^*\mathbf{A}\|_{2,F} \leq \|\mathbf{A} - \mathbf{A}_i\|_{2,F} + \delta_i \tag{10}$$

where

$$\delta_i := \sqrt{2}\,\|\mathbf{A}-\mathbf{A}_i\|_2\,[\,4\,\Delta(\mathbf{X},q,j)_F\,\frac{\sigma_{j+1}}{\sigma_j} + \Delta^*(\mathbf{Y}_q,t,k)_F\,\frac{\sigma_{k+1}}{\sigma_k}\,]\,)\,.$$

In case $j = 0$ (respectively $k = rank(\mathbf{A})$) the first term (respectively the second term) in the expression for δ_i should be omitted.

Proof. See Section 5.4. □

Remark 2.12. Consider the notation in Theorem 2.11. We point out that the hypothesis in Eq. (9) holds for sufficiently large q, $t \geq 1$ (see Remark 2.8). In a similar way, we see that the upper bound in Eq. (10) becomes arbitarily close to $\|\mathbf{A} - \mathbf{A}_i\|_{2,F}$, for sufficiently large q, $t \geq 1$.

On the other hand, the hypothesis considered in Eq. (9) is rather arbitrary. More generally, if we assume that the expression to the left in Eq. (9) is bounded from above by $\sin(\theta_0)$ for some $0 < \theta_0 < \pi/2$, then we conclude that Eq. (10) holds with δ_i replace by $\delta_i(\theta_0) = \frac{\delta_i}{\sqrt{2}\,\cos(\theta_0)}$. This can be seen by inspection of the proof of this result (see Section 5.4). Notice that the statement of Theorem 2.11 above corresponds to $\theta_0 = \pi/4$. △

3 Comments and remarks

Our present work deals with two different (yet related) topics: dominant subspace approximation and low rank matrix approximation. On the one hand, there is a vast research literature related to low rank approximation, both from a deterministic and randomized approach, taking into account singular gaps, or disregarding this gaps.

We point out that our approach is deterministic in nature, and does not assume a singular gap. Moreover, our techniques decompose the problem of low rank approximation into two sub-problems (see Section 5): Assuming that $\mathbf{A}$, $\mathbf{X}$ and $j < h \leq k$ are as in Theorem 2.11

- We first apply Drineas, Ipsen, Kontopoulou and Magdon-Ismail theory from [5] to construct a block Krylov space $\mathcal{K}_q = \mathcal{K}_q(\mathbf{A},\mathbf{X})$ (for appropriate q), that now has a strong compatibility with the (uniquely determined) left dominant j-dimensional subspace $\mathcal{U}_j$ of $\mathbf{A}$ (see the proof of Theorem 2.6).
- Then we construct the auxiliary subspace $\mathcal{K}^*_{q,t} = \mathcal{K}_t(\mathbf{A}^*,\mathbf{Y}_q)$, where $\mathbf{Y}_q\mathbf{Y}_q^*$ is the orthogonal projection onto $\mathcal{K}_q$, that has a strong compatibility with *some* h-dimensional right dominant subspace of $\mathbf{A}$.

The low rank approximation is now obtained from the best Frobenius approximation of $\mathbf{A}$ from $\mathbf{A}(\mathcal{K}_{q,t}^*) \subseteq \mathcal{K}_{q+t+1}$, as described in Algorithm 2.1. This factorization of the analysis is reflected in the upper bound obtained in Theorem 2.11 for the convergence of the method. This approach suggest possible numerical implementations including enlarging and re-starting techniques, by which we construct low rank approximations of dimension h from approximate dominant subspaces of dimension j, for $j < h$. Such numerical implementations (including randomized initial matrices) would also have to deal with efficiency and stability; these matter are beyond the scope of our present work.

On the other hand, the results related to dominant subspace approximation without singular gaps (in terms of initial compatible matrices $\mathbf{X} \in \mathbb{K}^{n\times r}$ with $r = h$) seem to be new; indeed, we first consider a convenient description of the class of h-dimensional dominant subspaces in case there is no singular gap. We then deal with the approximation of left and right dominant subspaces, adapting some of the main techniques of [5] to this more general setting. In case there is no singular gap, then an interesting problem arises: namely, that there is no uniquely determined target subspace to approximate; in order to deal with his last fact, we are required to develop some geometric arguments related to subspace approximation. Our approach also decomposes this problem into two sub-problems. As before, this suggest possible numerical implementations that consider enlarging and re-starting techniques to build high quality approximations of dominant subspaces from high quality approximation of dominant subspaces of lower dimensions. We believe that this type of factorization of the analysis can also be of interest to deal with randomized methods. We plan to consider this analysis elsewhere.

4 Auxiliary results

In this section we recall some well known notions and tools that we need in the sequel. Then, we include some of the main results from [5] that play a central role in our work.

4.1 Matrix functions for the block Krylov spaces

Consider $\mathbf{A} \in \mathbb{K}^{m\times n}$ and let $\mathbf{A} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^*$ be a SVD of $\mathbf{A}$; let $\mathbf{X} \in \mathbb{K}^{n\times r}$ and let $\mathcal{K}_q = \mathcal{K}_q(\mathbf{A}, \mathbf{X})$ denote the Krylov space constructed in terms of $\mathbf{A}$ and $\mathbf{X}$ as in Eq. (1). Notice that the elements in $\mathcal{K}_q$ can be described in terms of the elements of the range of matrices $\psi(\mathbf{A}\mathbf{A}^*)\mathbf{A}\mathbf{X} \in \mathbb{K}^{m\times r}$, where $\psi(x) \in \mathbb{K}[x]$ is a polynomial of degree at most q. In terms of SVD of $\mathbf{A}$, we get that

$$\psi(\mathbf{A}\mathbf{A}^*)\mathbf{A}\mathbf{X} = \mathbf{U}\psi(\boldsymbol{\Sigma}^2)\boldsymbol{\Sigma}\mathbf{V}^*\mathbf{X} = \mathbf{U}\phi(\boldsymbol{\Sigma})\mathbf{V}^*\mathbf{X}$$

where $\phi(x) = x\,\psi(x^2) \in \mathbb{K}[x]$ is a polynomial of degree at most $2q+1$ with odd powers only and its represents a generalized matrix function (see [1, 13]). Here $\boldsymbol{\Sigma} = \operatorname{diag}(\sigma_1, \ldots, \sigma_p) \in \mathbb{R}^{m\times n}$, where $p = \min\{m, n\}$; hence, $\phi(\boldsymbol{\Sigma}) = \operatorname{diag}(\phi(\sigma_1), \ldots, \phi(\sigma_p)) \in \mathbb{K}^{m\times n}$. In this case we note

$$\boldsymbol{\Phi} := \mathbf{U}\phi(\boldsymbol{\Sigma})\mathbf{V}^*\mathbf{X} \in \mathbb{K}^{m\times r}\,, \tag{11}$$

so by the previous facts, $R(\boldsymbol{\Phi}) \subset \mathcal{K}_q$. We will further consider similar notions related to convenient block decompositions of the SVD of $\mathbf{A}$.

Assume now that the pair $(\mathbf{A}, \mathbf{X})$ is h-compatible for some $1 \le h \le \operatorname{rank}(\mathbf{A})$; then, there exists an h-dimensional right dominant subspace $\mathcal{S}$ of $\mathbf{A}$ such that $\boldsymbol{\Theta}(R(\mathbf{X}), \mathcal{S}) < \frac{\pi}{2}\mathbf{I}_h$.

Consider a SVD of $\mathbf{A}$ as above, in such a way that $\mathcal{S} = \mathcal{V}_h$. In this case, $R(\mathbf{V}_h^*\mathbf{X}) = R(\mathbf{V}_h^*)$ and therefore $\dim R(\mathbf{\Phi}) \geq h$, where $\mathbf{\Phi}$ is defined as in Eq. (11).

4.2 Principal angles and vectors between subspaces

Let $\mathcal{S}, \mathcal{T} \subset \mathbb{K}^n$ be two subspaces of dimensions s and t respectively. Let $\mathbf{S} \in \mathbb{K}^{n\times s}$ and $\mathbf{T} \in \mathbb{K}^{n\times t}$ be isometries such that $R(\mathbf{S}) = \mathcal{S}$ and $R(\mathbf{T}) = \mathcal{T}$. Following [10], we define the principal angles between $\mathcal{S}$ and $\mathcal{T}$, denoted

$$0 \leq \theta_1(\mathcal{S},\mathcal{T}) \leq \ldots \leq \theta_k(\mathcal{S},\mathcal{T}) \leq \frac{\pi}{2} \quad \text{where} \quad k = \min\{s,\, t\}\,,$$

determined by the identities $\cos(\theta_j(\mathcal{S},\mathcal{T})) = \sigma_j(\mathbf{S}^*\mathbf{T})$, for $1 \leq j \leq k$; in this case the roles of $\mathbf{S}$ and $\mathbf{T}$ are symmetric. If we assume that $s \leq t$ (so $k = s$) the principal angles can be also determined in terms of the identities

$$\sin(\theta_{s-j+1}(\mathcal{S},\mathcal{T})) = \sigma_j((\mathbf{I}-\mathbf{T}\mathbf{T}^*)\mathbf{S}) = \sigma_j((\mathbf{I}-\mathbf{T}\mathbf{T}^*)\mathbf{S}\mathbf{S}^*) = \sigma_j((\mathbf{I}-\mathbf{P}_\mathcal{T})\mathbf{P}_\mathcal{S}) \tag{12}$$

for $1 \leq j \leq s$, where $\mathbf{P}_\mathcal{H} \in \mathbb{K}^{n\times n}$ denotes the orthogonal projection onto a subspace $\mathcal{H} \subset \mathbb{K}^n$; it is worth noticing that in this case the roles of $\mathbf{S}$ and $\mathbf{T}$ (equivalently the roles of $\mathbf{P}_\mathcal{T}$ and $\mathbf{P}_\mathcal{T}$) are not symmetric (unless $s = t$). Principal angles can be considered as a vector valued measure of the distance between the subspaces $\mathcal{S}$ and $\mathcal{T}$.

Following [20] we let $\mathbf{\Theta}(\mathcal{S},\mathcal{T}) = \operatorname{diag}(\theta_1(\mathcal{S},\mathcal{T}),\ldots,\theta_s(\mathcal{S},\mathcal{T}))$ denote the diagonal matrix with the principal angles in its main diagonal. In particular,

$$\|\sin\mathbf{\Theta}(\mathcal{S},\mathcal{T})\|_{2,F} = \|(\mathbf{I}-\mathbf{P}_\mathcal{T})\mathbf{P}_\mathcal{S}\|_{2,F}$$

are scalar measures of the (angular) distance between $\mathcal{S}$ and $\mathcal{T}$ (see [10, 20]).

We mention some properties of the principal angles between subspaces that we will need in what follows. With the previous notation, we point out that if $\mathcal{S}' \subset \mathcal{S}$ and $\mathcal{T} \subset \mathcal{T}'$ are subspaces with dimensions s' and t' respectively, then (recall that $s = \dim\mathcal{S} \leq \dim\mathcal{T} = t$)

$$\|\mathbf{\Theta}(\mathcal{S},\mathcal{T}')\|_{2,F} \leq \|\mathbf{\Theta}(\mathcal{S},\mathcal{T})\|_{2,F} \quad , \quad \|\sin\mathbf{\Theta}(\mathcal{S},\mathcal{T}')\|_{2,F} \leq \|\sin\mathbf{\Theta}(\mathcal{S},\mathcal{T})\|_{2,F}$$

and similarly

$$\|\mathbf{\Theta}(\mathcal{S}',\mathcal{T})\|_{2,F} \leq \|\mathbf{\Theta}(\mathcal{S},\mathcal{T})\|_{2,F} \quad , \quad \|\sin\mathbf{\Theta}(\mathcal{S}',\mathcal{T})\|_{2,F} \leq \|\sin\mathbf{\Theta}(\mathcal{S},\mathcal{T})\|_{2,F}\,,$$

which follow from Eq. (12). On the other hand, $\dim\mathcal{S}^\perp = n - s \geq n - t = \dim\mathcal{T}^\perp$ and therefore,

$$\sin(\theta_{(n-t)-j+1}(\mathcal{S}^\perp,\mathcal{T}^\perp)) = \sigma_j((\mathbf{I}-\mathbf{P}_{\mathcal{S}^\perp})\mathbf{P}_{\mathcal{T}^\perp}) = \sigma_j(\mathbf{P}_\mathcal{S}(\mathbf{I}-\mathbf{P}_\mathcal{T}))\;,\quad 1 \leq j \leq n-t\,.$$

By comparing the previous identity with Eq. (12), if $\theta_1(\mathcal{S},\mathcal{T}),\ldots,\theta_d(\mathcal{S},\mathcal{T}) > 0$ are the positive angles between $\mathcal{S}$ and $\mathcal{T}$ (for some $0 \leq d \leq \min\{s, n-t\}$) then these coincide with the positive angles between $\mathcal{S}^\perp$ and $\mathcal{T}^\perp$ i.e.

$$\theta_j(\mathcal{S},\mathcal{T}) = \theta_j(\mathcal{S}^\perp,\mathcal{T}^\perp) \quad \text{for} \quad 1 \leq j \leq d\,. \tag{13}$$

Notice that as a consequence of Eq. (13) we get that

$$\|\mathbf{\Theta}(\mathcal{S},\mathcal{T})\|_{2,F} = \|\mathbf{\Theta}(\mathcal{S}^\perp,\mathcal{T}^\perp)\|_{2,F} \quad , \quad \|\sin\mathbf{\Theta}(\mathcal{S},\mathcal{T})\|_{2,F} = \|\sin\mathbf{\Theta}(\mathcal{S}^\perp,\mathcal{T}^\perp)\|_{2,F}\,. \tag{14}$$

In what follows we shall also make use of the principal vectors associated to the subspaces $\mathcal{S}$ and $\mathcal{T}$: indeed, by construction of the principal angles, we get that there exist orthonormal systems $\{\mathbf{u}_1,\ldots,\mathbf{u}_s\}\subset\mathcal{S}$ and $\{\mathbf{v}_1,\ldots,\mathbf{v}_s\}\subset\mathcal{T}$ such that

$$\langle \mathbf{u}_i,\mathbf{v}_j\rangle=\delta_{ij}\,\cos(\theta_j(\mathcal{S},\mathcal{T}))\quad\text{for}\quad 1\le i,\,j\le s\,,$$

where δ_{ij} denotes Kronecker's delta function. We say that $\{\mathbf{u}_1,\ldots,\mathbf{u}_s\}$ and $\{\mathbf{v}_1,\ldots,\mathbf{v}_s\}$ are the *principal vectors (directions)* associated with the subspaces $\mathcal{S}$ and $\mathcal{T}$. Notice that the previous facts imply, in particular, that the subspaces $\mathcal{S}_j=\operatorname{Span}\{\mathbf{u}_1,\ldots,\mathbf{u}_j\}\subset\mathcal{S}$ and $\mathcal{T}_j=\operatorname{Span}\{\mathbf{v}_1,\ldots,\mathbf{v}_j\}\subset\mathcal{T}$ are such that $\mathbf{\Theta}(\mathcal{S}_j,\mathcal{T}_j)=\operatorname{diag}(\theta_1(\mathcal{S},\mathcal{T}),\ldots,\theta_j(\mathcal{S},\mathcal{T}))\in\mathbb{R}^{j\times j}$, for $1\le j\le s$. Moreover, if $\tilde{\mathcal{S}}\subset\mathcal{S}$ and $\tilde{\mathcal{T}}\subset\mathcal{T}$ are two j-dimensional subspaces then, it follows that $\mathbf{\Theta}(\mathcal{S}_j,\mathcal{T}_j)\le\mathbf{\Theta}(\tilde{\mathcal{S}},\tilde{\mathcal{T}})$; that is, $\mathcal{S}_j$ and $\mathcal{T}_j$ are j-dimensional subspaces of $\mathcal{S}$ and $\mathcal{T}$ respectively, that are at minimal angular (vector valued) length.

4.3 On the DIKM-I theory with singular gaps

In [5] P. Drineas, I.C.F. Ipsen, E.M. Kontopoulou and M. Magdon-Ismail merged a series of techniques, tools and arguments that lead to structural results related to the approximation of dominant subspaces from block Krylov spaces in the presence of a singular gap. That work has a deep influence in our present work; indeed, we shall follow some of the lines developed in that work, that we refer to as the *DIKM-I theory.* Of course, at some points we have to departure from those arguments to deal with the no-singular-gap case. Next we include some of the features of the DIKM-I theory that we need in what follows.

In this section we keep using the notation considered so far ($\mathbf{A}\in\mathbb{K}^{m\times n}$, its SVD, the associated partitions as in Eq. (2) and so on).

Theorem 4.1 ([5]). Assume that $\sigma_k>\sigma_{k+1}$, let $\phi(x)$ be a polynomial of degree at most $2q+1$ with odd powers only, such that $\phi(\mathbf{\Sigma}_k)$ is non-singular. If $\tilde{\mathbf{X}}\in\mathbb{K}^{n\times r}$ is such that $\operatorname{rank}(\mathbf{V}_k^*\tilde{\mathbf{X}})=k$ (so $r\ge k$) and $\tilde{\mathcal{K}}_q=\mathcal{K}_q(\mathbf{A},\tilde{\mathbf{X}})$ then

$$\|\sin\mathbf{\Theta}(\tilde{\mathcal{K}}_q,\mathcal{U}_k)\|_{2,F}\le\|\phi(\mathbf{\Sigma}_{k,\perp})\|_2\,\|\phi(\mathbf{\Sigma}_k)^{-1}\|_2\,\|\mathbf{V}_{k,\perp}^*\tilde{\mathbf{X}}(\mathbf{V}_k^*\tilde{\mathbf{X}})^\dagger\|_{2,F}\,.$$

If, in addition, $\tilde{\mathbf{X}}$ has orthonormal or linearly independent columns, then

$$\|\mathbf{V}_{k,\perp}^*\tilde{\mathbf{X}}(\mathbf{V}_k^*\tilde{\mathbf{X}})^\dagger\|_{2,F}=\|\tan\mathbf{\Theta}(R(\tilde{\mathbf{X}}),\mathcal{V}_k)\|_{2,F}\,.$$

□

The arguments involved in the proof of the Theorem 4.1 played a central role in [5]. In the next result we make use of Algorithm 2.1 with the rank parameter set to $k\le\operatorname{rank}(\mathbf{A})$ and input matrices $\mathbf{A}\in\mathbb{K}^{m\times n}$, $\tilde{\mathbf{X}}\in\mathbb{K}^{n\times r}$ such that $R(\mathbf{V}_k^*\tilde{\mathbf{X}})=k$. In this case we describe the output matrix in terms of its columns $\hat{\mathbf{U}}_k=(\hat{\mathbf{u}}_1,\ldots,\hat{\mathbf{u}}_k)$. We also consider the matrices $\hat{\mathbf{U}}_i=(\hat{\mathbf{u}}_1,\ldots,\hat{\mathbf{u}}_i)\in\mathbb{K}^{m\times i}$, for $1\le i\le k$.

Theorem 4.2 ([5]). Assume that $\sigma_k>\sigma_{k+1}$, let $\phi(x)$ be a polynomial of degree at most $2q+1$ with odd powers only, such that $\phi(\mathbf{\Sigma}_k)$ is non-singular and $\phi(\sigma_i)\ge\sigma_i$ for $1\le i\le k$. Then for $1\le i\le k$,

$$\begin{aligned}\|\mathbf{A}-\hat{\mathbf{U}}_i\hat{\mathbf{U}}_i^*\mathbf{A}\|_{2,F}&\le\|\mathbf{A}-\mathbf{A}_i\|_{2,F}+\Delta\\ \sigma_i-\Delta&\le\|\hat{\mathbf{u}}_i^*\mathbf{A}\|_2\le\sigma_i\end{aligned}$$

where $\mathbf{A}_i \in \mathbb{K}^{m\times n}$ denotes a best rank-i approximation of $\mathbf{A}$ and

$$\Delta = \|\phi(\mathbf{\Sigma}_{k,\perp})\|_2 \, \|\mathbf{V}_{k,\perp}^*\mathbf{X}(\mathbf{V}_k^*\mathbf{X})^\dagger\|_F \, .$$

□

The following result from [5] complements Theorems 4.1 and 4.2 above, in the sense that it implies that the upper bounds in those theorems can be made arbitrarily small. This result corresponds to a generalization of the Chebyshev-based gap-amplifying polynomials developed in [18] and [21].

Lemma 4.3 ([5])**.** Assume that $k < \text{rank}(\mathbf{A})$, so that $\sigma_k > \sigma_{k+1} > 0$, and let

$$\gamma_k = \frac{\sigma_k - \sigma_{k+1}}{\sigma_{k+1}} > 0 \, .$$

Then, there exists a polynomial $\phi(x)$ of degree at most $2q+1$ with odd powers only, such that

$$\begin{aligned} \phi(\sigma_1) \geq \ldots \geq \phi(\sigma_k) \quad , \quad \phi(\sigma_i) \geq \sigma_i > 0 \quad &, \qquad \text{for} \quad 1 \leq i \leq k \, , \\ \text{and} \quad |\phi(\sigma_i)| \leq \frac{4\sigma_{k+1}}{2^{(2q+1)\min\{\sqrt{\gamma_k}\,,\,1\}}} \quad &, \qquad \text{for} \quad i \geq k+1 \, . \end{aligned}$$

Hence,

$$\|\phi(\mathbf{\Sigma}_k)^{-1}\|_2 \leq \sigma_k^{-1} \quad \text{and} \quad \|\phi(\mathbf{\Sigma}_{k,\perp})\|_2 \leq \frac{4\sigma_{k+1}}{2^{(2q+1)\min\{\sqrt{\gamma_k}\,,\,1\}}} \, .$$

□

We point out that the inequalities $\phi(\sigma_1) \geq \ldots \geq \phi(\sigma_k)$ in the lemma above are a consequence of the super-linear growth for large input values (i.e. in this case for $x \geq \sigma_{k+1}$) of the gap amplifying Chebyshev polynomials (see [5]).

5 Proofs of the main results

In this section we present detailed proofs of our main results. Some of our arguments make use of some basic facts from matrix analysis, that we develop in Section 6 (Appendix).

5.1 Proof of Theorem 2.1 and Corollary 2.3

We begin this section with a proof of Theorem 2.1. We present our arguments divided into steps.

Step 1: adapting the DIKM-I theory to the present context. Let $\mathbf{A} \in \mathbb{K}^{m\times n}$. For a $1 \leq h \leq \text{rank}(\mathbf{A})$ we let $0 \leq j = j(h) < h \leq k = k(h)$ be defined as in the beginning of Section 2.2. By construction $\sigma_j > \sigma_{j+1} = \sigma_h = \sigma_k$. We first assume that $1 \leq j$ and $k < \text{rank}(\mathbf{A}) \leq p = \min\{m, n\}$. Since $k < \text{rank}(\mathbf{A})$ then $\sigma_h = \sigma_k > \sigma_{k+1} > 0$.

Let $\mathbf{X} \in \mathbb{K}^{n\times r}$ be such $\dim(\mathcal{X}) = s \geq h$ and such that there exists a right dominant subspace $\mathcal{S} \subset \mathbb{K}^n$ of dimension h such $\mathbf{\Theta}(\mathcal{S}, \mathcal{X}) < \pi/2\,\mathbf{I}_h$, where $\mathcal{X} = R(\mathbf{X}) \subset \mathbb{K}^n$ denotes the range of $\mathbf{X}$. Consider $\mathbf{A} = \mathbf{U\Sigma V}^*$ a full SVD. We now consider the partitioning

$$\mathbf{\Sigma} = \begin{pmatrix} \mathbf{\Sigma}_k & \\ & \mathbf{\Sigma}_{k,\perp} \end{pmatrix} \; , \quad \mathbf{U} = \begin{pmatrix} \mathbf{U}_k & \mathbf{U}_{k,\perp} \end{pmatrix} \; , \quad \mathbf{V} = \begin{pmatrix} \mathbf{V}_k & \mathbf{V}_{k,\perp} \end{pmatrix} \, . \tag{15}$$

It is worth to notice that $\mathbf{\Sigma}_k$, $R(\mathbf{U}_k) = \mathcal{U}_k$ and $R(\mathbf{V}_k) = \mathcal{V}_k$ do not depend on the particular choice of SDV of $\mathbf{A}$; also notice that the partition is also well defined since $k < \text{rank}(\mathbf{A})$. Let $\phi(x)$ be a polynomial of degree $2q+1$ with odd powers only, such that $\phi(\sigma_1) \geq \ldots \geq \phi(\sigma_k) > 0$; hence $\phi(\mathbf{\Sigma}_k)$ is invertible.

Step 2: applying the DIKM-I theory to the adapted model. We let $\mathcal{K}_q = \mathcal{K}_q(\mathbf{A}, \mathbf{X})$ denote the block Krylov subspace and let $\mathbf{P}_q \in \mathbb{K}^{m\times m}$ denote the orthogonal projection onto $\mathcal{K}_q$. Notice that if we let $\mathbf{\Phi} \in \mathbb{K}^{m\times r}$ be as in Eq. (11) then $R(\mathbf{\Phi}) \subset \mathcal{K}_q$. Consider for now an arbitrary h-dimensional subspace $\mathcal{S}' \subset \mathbb{K}^m$. Then

$$\|\sin \mathbf{\Theta}(\mathcal{K}_q, \mathcal{S}')\|_{2,F} = \|(I - \mathbf{P}_q)\mathbf{P}_{\mathcal{S}'}\|_{2,F} \leq \|(I - \mathbf{\Phi}\mathbf{\Phi}^\dagger)\mathbf{P}_{\mathcal{S}'}\|_{2,F}\,, \tag{16}$$

where we have used that $\dim \mathcal{K}_q \geq \dim R(\mathbf{\Phi}) \geq \dim \mathcal{S}' = h$. We now consider the decomposition $\mathbf{\Phi} = \mathbf{\Phi}_k + \mathbf{\Phi}_{k,\perp}$, where

$$\mathbf{\Phi}_k \equiv \mathbf{U}_k \phi(\mathbf{\Sigma}_k)\mathbf{V}_k^*\mathbf{X} \quad \text{and} \quad \mathbf{\Phi}_{k,\perp} \equiv \mathbf{U}_{k,\perp}\phi(\mathbf{\Sigma}_{k,\perp})\mathbf{V}_{k,\perp}^*\mathbf{X}\,.$$

By [5, Lemma 4.2] (see also [16]) we get that

$$\|(I - \mathbf{\Phi}\mathbf{\Phi}^\dagger)\mathbf{P}_{\mathcal{S}'}\|_{2,F} \leq \|\mathbf{P}_{\mathcal{S}'} - \mathbf{\Phi}\mathbf{B}\|_{2,F} \quad \text{for} \quad \mathbf{B} \in \mathbb{K}^{r\times m}\,.$$

By the previous inequality we get that

$$\|(I - \mathbf{\Phi}\mathbf{\Phi}^\dagger)\mathbf{P}_{\mathcal{S}'}\|_{2,F} \leq \|(I - \mathbf{\Phi}\mathbf{\Phi}_k^\dagger)\mathbf{P}_{\mathcal{S}'}\|_{2,F} \leq \|(I - \mathbf{\Phi}_k\mathbf{\Phi}_k^\dagger)\mathbf{P}_{\mathcal{S}'}\|_{2,F} + \|\mathbf{\Phi}_{k,\perp}\mathbf{\Phi}_k^\dagger\mathbf{P}_{\mathcal{S}'}\|_{2,F}\,. \tag{17}$$

Step 3: dealing with the fact that $R(\mathbf{V}_k^*\mathbf{X}) \neq R(\mathbf{V}_k^*)$. We now consider the two terms to the right of Eq. (17). In our present case, we have to deal with the fact that $R(\mathbf{V}_k^*\mathbf{X}) \neq R(\mathbf{V}_k^*)$ when $h < k$. Indeed, since $\mathbf{\Theta}(\mathcal{S}, \mathcal{X}) < \pi/2\, \mathbf{I}_h$ and $\mathcal{S} \subset R(\mathbf{V}_k)$ we see that if we let

$$\mathcal{W} \equiv R(\mathbf{V}_k^*\mathbf{X}) = \mathbf{V}_k^*\mathcal{X} \subset \mathbb{K}^k$$

then $k \geq \dim(\mathcal{W}) := t \geq h$. Let

$$\mathcal{T} = \phi(\mathbf{\Sigma}_k)\mathcal{W} \subset \mathbb{K}^k\,.$$

Since, by hypothesis, $\phi(\mathbf{\Sigma}_k) \in \mathbb{R}^{k\times k}$ is an invertible matrix then $\dim \mathcal{T} = t$ and

$$\mathbf{\Phi}_k\mathbf{\Phi}_k^\dagger = \mathbf{U}_k\mathbf{P}_{\mathcal{T}}\mathbf{U}_k^*\,. \tag{18}$$

We now consider $\mathcal{H}' = \text{Span}\{\mathbf{e}_1, \ldots, \mathbf{e}_j\} \subset \mathbb{K}^k$, where $j = j(h) \geq 1$ and $\{\mathbf{e}_1, \ldots, \mathbf{e}_k\}$ denotes the canonical basis of $\mathbb{K}^k$; we also consider the principal angles

$$\mathbf{\Theta}(\mathcal{W}, \mathcal{H}') = \text{diag}(\theta_1(\mathcal{W}, \mathcal{H}'), \ldots, \theta_j(\mathcal{W}, \mathcal{H}')) \in \mathbb{R}^{j\times j}\,.$$

By Proposition 6.1 we get that

$$\mathbf{\Theta}(\mathcal{W}, \mathcal{H}') \leq \mathbf{\Theta}(\mathcal{X}, \mathcal{V}_j) < \frac{\pi}{2}\,\mathbf{I}_j\,,$$

since $\mathcal{V}_j \subset R(\mathbf{V}_k)$ and $\mathbf{V}_k^*\mathcal{V}_j = \mathcal{H}'$, and the second inequality above follows from the fact that $\theta_i(\mathcal{X}, \mathcal{V}_j) \leq \theta_i(\mathcal{X}, \mathcal{S}) < \frac{\pi}{2}$, for $1 \leq i \leq j$, since $\mathcal{V}_j \subset \mathcal{S}$ (see Section 4.2).

Step 4: computing the left dominant subspace $\mathcal{S}'$. Let $\{\mathbf{w}_1, \ldots, \mathbf{w}_j\} \subset \mathcal{W}$ and $\{\mathbf{f}_1, \ldots, \mathbf{f}_j\} \subset \mathcal{H}'$ be the principal vectors associated to $\mathcal{W}$ and $\mathcal{H}'$ (as described in Section 4.2). Let $\mathcal{W}' =$

$\operatorname{Span}\{\mathbf{w}_1,\ldots,\mathbf{w}_j\} \subset \mathcal{W}$; in this case, $\mathbf{\Theta}(\mathcal{W},\mathcal{H}') = \mathbf{\Theta}(\mathcal{W}',\mathcal{H}')$, by construction. Consider the subspace $\mathcal{T}' = \phi(\mathbf{\Sigma}_k)\,\mathcal{W}' \subset \mathcal{T}$ so $\dim(\mathcal{T}') = \dim(\mathcal{W}') = j = \dim(\mathcal{H}')$; since $\mathcal{H}'$ is an invariant subspace of $\phi(\mathbf{\Sigma}_k)$ then Proposition 6.2 implies that

$$\|\sin\mathbf{\Theta}(\mathcal{T}',\mathcal{H}')\|_{2,F} \leq \|\sin\mathbf{\Theta}(\mathcal{W}',\mathcal{H}')\|_{2,F} = \|\sin\mathbf{\Theta}(\mathcal{W},\mathcal{H}')\|_{2,F}$$

since $\|\phi(\mathbf{\Sigma}_k)(\mathbf{I}-\mathbf{P}_{\mathcal{H}'})\|_2\,\|\phi(\mathbf{\Sigma}_k)^{-1}\|_2 = 1$, where we used that $\phi(\sigma_i) \geq \phi(\sigma_k) > 0$, for $1 \leq i \leq k$ and that $\phi(\sigma_{j+1}) = \phi(\sigma_k)$.

Let $\mathcal{T}'' = \mathcal{T} \ominus \mathcal{T}'$ so $\dim\mathcal{T}'' = t-j$ and $\mathcal{T}'' \subset (\mathcal{T}')^\perp$. Since $\dim((\mathcal{T}')^\perp) = \dim((\mathcal{H}')^\perp)$, by Eq. (14) we see that

$$\|\mathbf{\Theta}(\mathcal{T}'',(\mathcal{H}')^\perp)\|_{2,F} \leq \|\mathbf{\Theta}((\mathcal{T}')^\perp,(\mathcal{H}')^\perp)\|_{2,F} = \|\mathbf{\Theta}(\mathcal{T}',\mathcal{H}')\|_{2,F} \leq \|\mathbf{\Theta}(\mathcal{W},\mathcal{H}')\|_{2,F}\,.$$

Let $\{\mathbf{y}_1,\ldots,\mathbf{y}_{t-j}\} \subset \mathcal{T}''$ and $\{\mathbf{z}_1,\ldots,\mathbf{z}_{t-j}\} \subset (\mathcal{H}')^\perp$ be the principal vectors associated to $\mathcal{T}''$ and $(\mathcal{H}')^\perp$. Then, if we let $\mathcal{H}'' = \operatorname{Span}\{\mathbf{z}_1,\ldots,\mathbf{z}_{h-j}\}$ we have that $\dim\mathcal{H}'' = h-j$,

$$\|\sin\mathbf{\Theta}(\mathcal{T}'',\mathcal{H}'')\|_{2,F} \leq \|\sin\mathbf{\Theta}(\mathcal{T}'',(\mathcal{H}')^\perp)\|_{2,F} \leq \|\sin\mathbf{\Theta}(\mathcal{W},\mathcal{H}')\|_{2,F}\,.$$

On the one hand, we have that $\mathcal{T} = \mathcal{T}' \oplus \mathcal{T}''$; on the other hand, we have that

$$\mathcal{S}' := \mathbf{U}_k(\mathcal{H}' \oplus \mathcal{H}'') = \mathcal{U}_j \oplus \mathbf{U}_k\mathcal{H}'' \subseteq \mathcal{U}_k \subseteq \mathbb{K}^m$$

is an h-dimensional left dominant subspace of A (see Section 2.1).

Step 5: obtaining some more upper bounds. Since

$$\|\sin\mathbf{\Theta}(\mathcal{T}',\mathcal{H}')\|_{2,F}\,,\ \|\sin\mathbf{\Theta}(\mathcal{T}'',\mathcal{H}'')\|_{2,F} \leq \|\sin\mathbf{\Theta}(\mathcal{W},\mathcal{H}')\|_{2,F}$$

then Proposition 6.3 implies that $\|\sin\mathbf{\Theta}(\mathcal{T}\,,\,\mathcal{H}'\oplus\mathcal{H}'')\|_{2,F} \leq 4\,\|\sin\mathbf{\Theta}(\mathcal{W},\mathcal{H}')\|_{2,F}$ and hence

$$\|(I-\mathbf{\Phi}_k\mathbf{\Phi}_k^\dagger)\mathbf{P}_{\mathcal{S}'}\|_{2,F} = \|\sin\mathbf{\Theta}(\mathbf{U}_k\mathcal{T}\,,\,\mathcal{S}')\|_{2,F} \leq 4\,\|\sin\mathbf{\Theta}(\mathcal{W},\mathcal{H}')\|_{2,F} \tag{19}$$

since $\mathbf{U}_k$ is an isometry and $R(\mathbf{\Phi}_k) = \mathbf{U}_k\,\mathcal{T}$ (see Eq. (18)).

By Proposition 6.4, since $\mathcal{W} = R(\mathbf{V}_k^*\mathbf{X})$,

$$\mathbf{\Phi}_k^\dagger = (\mathbf{V}_k^*\mathbf{X})^\dagger(\mathbf{U}_k\phi(\mathbf{\Sigma}_k)\mathbf{P}_{\mathcal{W}})^\dagger\,.$$

Since $\mathbf{U}_k \in \mathbb{K}^{m\times k}$ has trivial kernel, we get that

$$(\mathbf{U}_k\phi(\mathbf{\Sigma}_k)\mathbf{P}_{\mathcal{W}})^\dagger = (\phi(\mathbf{\Sigma}_k)\mathbf{P}_{\mathcal{W}})^\dagger(\mathbf{U}_k\mathbf{P}_{\mathcal{T}})^\dagger = (\phi(\mathbf{\Sigma}_k)\mathbf{P}_{\mathcal{W}})^\dagger\mathbf{P}_{\mathcal{T}}\mathbf{U}_k^* = (\phi(\mathbf{\Sigma}_k)\mathbf{P}_{\mathcal{W}})^\dagger\mathbf{U}_k^*$$

where we have used Proposition 6.4, that $\mathbf{U}_k\mathbf{P}_{\mathcal{T}}$ is a partial isometry so that $(\mathbf{U}_k\mathbf{P}_{\mathcal{T}})^\dagger = (\mathbf{U}_k\mathbf{P}_{\mathcal{T}})^* = \mathbf{P}_{\mathcal{T}}\mathbf{U}_k^*$ and that $\ker((\phi(\mathbf{\Sigma}_k)\mathbf{P}_{\mathcal{W}})^\dagger)^\perp = \mathcal{T}$. The previous facts show that

$$\mathbf{\Phi}_{k,\perp}\mathbf{\Phi}_k^\dagger\mathbf{P}_{\mathcal{S}'} = \mathbf{U}_{k,\perp}\phi(\mathbf{\Sigma}_{k,\perp})\mathbf{V}_{k,\perp}^*\mathbf{X}(\mathbf{V}_k^*\mathbf{X})^\dagger(\phi(\mathbf{\Sigma}_k)\mathbf{P}_{\mathcal{W}})^\dagger\mathbf{U}_k^*\mathbf{P}_{\mathcal{S}'}$$

so then,

$$\|\mathbf{\Phi}_{k,\perp}\mathbf{\Phi}_k^\dagger\mathbf{P}_{\mathcal{S}'}\|_{2,F} \leq \|\phi(\mathbf{\Sigma}_{k,\perp})\|_2\,\|\phi(\mathbf{\Sigma}_k)^{-1}\|_2\,\|\mathbf{V}_{k,\perp}^*\mathbf{X}(\mathbf{V}_k^*\mathbf{X})^\dagger\|_{2,F}\,. \tag{20}$$

The result now follows from the estimates in Eqs. (16) and (17) together with the bounds in Eqs. (19) and (20).

The cases in which $j=0$ or $k = \operatorname{rank}(\mathbf{A})$ can be dealt with similar arguments. Indeed, notice that if $j=0$ then we can take $\tilde{\mathcal{T}} \subset \mathcal{T}$ such that $\dim\tilde{\mathcal{T}} = h$, and set $\mathcal{S}' = \mathbf{U}_k\tilde{\mathcal{T}}$.

By construction, $\mathcal{S}' \subset R(\mathbf{\Phi}_k)$ is a left dominant subspace of $\mathbf{A}$ (in this case any subspace of $\mathcal{U}_k$ is a dominant subspace of $\mathbf{A}$). Finally, in case $k = \operatorname{rank}(\mathbf{A})$ then $\mathbf{\Sigma}_{k,\perp} = 0$ and then $\phi(\mathbf{\Sigma}_{k,\perp}) = 0$, so that we also get $\mathbf{\Phi}_{k,\perp} = 0$. □

Now we consider a brief proof of Corollary 2.3. Indeed, by Lemma 4.3 we conclude that there exists a polynomial $\phi(x)$ satisfying the hypothesis of Theorem 2.1 and such that

$$\|\phi(\mathbf{\Sigma}_{k,\perp})\|_2 \|\phi(\mathbf{\Sigma}_k)^{-1}\|_2 \|\mathbf{V}_{k,\perp}^* \mathbf{X} (\mathbf{V}_k^* \mathbf{X})^\dagger\|_{2,F} \leq 4 \, \frac{\|\mathbf{V}_{k,\perp}^* \mathbf{X} (\mathbf{V}_k^* \mathbf{X})^\dagger\|_{2,F}}{2^{(2q+1)\min\{\sqrt{\gamma_k}\,,\,1\}}} \, \frac{\sigma_{k+1}}{\sigma_k} \, .$$

The result now follows from the previous inequality and the definition of $\Delta(\mathbf{X}, q, k)_{2,F}$. □

Remark 5.1. Some comments related to the previous proof are in order. We have followed the general lines of the proof of [5, Theorem 2.1]. Nevertheless, the assumption in [5] (i.e., that $R(\mathbf{V}_k^* \mathbf{X}) = R(\mathbf{V}_k^*)$) automatically implies that $\|(I - \mathbf{\Phi}_k \mathbf{\Phi}_k^\dagger) \mathbf{P}_{\mathcal{S}'}\|_{2,F} = 0$ in Eq. (17). Since we are only assuming that the pair $(\mathbf{A}, \mathbf{X})$ is h-compatible, our arguments need to include Steps 3, 4 and the first part of Step 5.

We can now see that the assumption that the pair $(\mathbf{A}, \mathbf{X})$ is h-compatible (for an arbitrary $1 \leq h \leq \operatorname{rank}(\mathbf{A})$) is weaker, at least from the point of view of our present approach, than the *structural* assumption that the pair $(\mathbf{A}, \mathbf{X})$ is k-compatible for an index k such that $\sigma_k > \sigma_{k+1}$, as considered in [5]. △

5.2 Proof of Theorem 2.6

Proof. Let $0 \leq j = j(h) < h \leq k = k(h)$ be defined as in the beginning of Section 2.2. We assume further that $1 \leq j < h < k < \operatorname{rank}(\mathbf{A})$; the cases $j = 0$ or $k = \operatorname{rank}(\mathbf{A})$ can be treated with similar arguments (the details are left to the reader). Notice that $\sigma_h = \sigma_k > 0$, since $h \leq \operatorname{rank}(\mathbf{A})$.

Step 1: applying the DIKM-I theory using the singular gap $\sigma_j > \sigma_{j+1}$. Let $\mathbf{X} \in \mathbb{K}^{n \times r}$ be such that $\dim(\mathcal{X}) = s \geq h$ and

$$\mathbf{\Theta}(\mathcal{S}, \mathcal{X}) < \frac{\pi}{2} \, \mathbf{I}_h \, ,$$

where $\mathcal{X} = R(\mathbf{X}) \subset \mathbb{K}^n$ denotes the range of $\mathbf{X}$ and $\mathcal{S} \subset \mathbb{K}^n$ is an h-dimensional right dominant subspace of $\mathbf{A}$. Consider a full SVD $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^*$ in such a way that $\mathcal{S} = \mathcal{V}_h$.

We can now consider decompositions as in Eq. (15), using the index $j = j(h)$ that is,

$$\mathbf{\Sigma} = \begin{pmatrix} \mathbf{\Sigma}_j & \\ & \mathbf{\Sigma}_{j,\perp} \end{pmatrix} \ , \quad \mathbf{U} = \begin{pmatrix} \mathbf{U}_j & \mathbf{U}_{j,\perp} \end{pmatrix} \ , \quad \mathbf{V} = \begin{pmatrix} \mathbf{V}_j & \mathbf{V}_{j,\perp} \end{pmatrix} \, . \tag{21}$$

As a consequence of our hypothesis, we get that $R(\mathbf{V}_j^* \mathbf{X}) = R(\mathbf{V}_j^*)$ (notice that the subspace $R(\mathbf{V}_j^*)$ is independent of our choice of SVD of $\mathbf{A}$, since $\sigma_j > \sigma_{j+1}$). Hence, we can apply Theorem 4.1 and Lemma 4.3 (that correspond to the DIKM-I theory with singular gaps) and conclude that if we let

$$\gamma_j = \frac{\sigma_j - \sigma_{j+1}}{\sigma_{j+1}} \quad \text{and} \quad \Delta(\mathbf{X}, q, j)_{2,F} = 4 \, \frac{\|\mathbf{V}_{j,\perp}^* \mathbf{X} (\mathbf{V}_j^* \mathbf{X})^\dagger\|_{2,F}}{2^{(2q+1)\min\{\sqrt{\gamma_j}\,,\,1\}}}$$

then, we have that

$$\|\sin \mathbf{\Theta}(\mathcal{K}_q, R(\mathbf{U}_j))\|_{2,F} \leq \Delta(\mathbf{X}, q, j)_{2,F} \, \frac{\sigma_{j+1}}{\sigma_j} \, , \tag{22}$$

where $\mathcal{K}_q = \mathcal{K}_q(\mathbf{A}, \mathbf{X})$ denotes the Krylov space of order $q \geq 1$.

Step 2: applying Theorem 2.1 to $\mathbf{A}^*$. On the other hand, since $\sigma_h > 0$ and $\mathbf{\Theta}(R(\mathbf{V}_h), R(\mathbf{X})) < \pi/2\, \mathbf{I}_h$, we conclude that $R(\mathbf{V}_h^*\mathbf{X}) = R(\mathbf{V}_h^*)$ and therefore $\operatorname{rank}(\mathbf{U}_h^*\mathbf{A}\mathbf{X}) = \operatorname{rank}(\mathbf{\Sigma}_h\mathbf{V}_h^*\mathbf{X}) = \operatorname{rank}(\mathbf{V}_h^*\mathbf{X}) = h$. The previous facts show that

$$\dim \mathcal{K}_q \geq h \quad \text{and} \quad \mathbf{\Theta}(\mathcal{K}_q, R(\mathbf{U}_h)) < \frac{\pi}{2}\, \mathbf{I}_h\,.$$

Let $\mathbf{Y}_q$ denote an isometry such that $\mathbf{Y}_q\mathbf{Y}_q^* \in \mathbb{K}^{m\times m}$ is the orthogonal projection onto $\mathcal{K}_q$. We now consider $\mathcal{K}_{q,t}^* = \mathcal{K}_t(\mathbf{A}^*, \mathbf{Y}_q)$ which is the block Krylov space of order t constructed in terms of $\mathbf{A}^*$ and $\mathbf{Y}_q$. Notice that $\mathcal{S}^* = R(\mathbf{U}_h)$ is an h-dimensional right dominant subspace of $\mathbf{A}^*$ such that $\mathbf{\Theta}(R(\mathbf{Y}_q), \mathcal{S}^*) < \frac{\pi}{2}\, \mathbf{I}_h$. Moreover, the subspace $R(\mathbf{U}_j)$ is a j-dimensional right dominant subspace of $\mathbf{A}^*$, such that $\mathbf{\Theta}(R(\mathbf{Y}_q), R(\mathbf{U}_j)) = \mathbf{\Theta}(\mathcal{K}_q, R(\mathbf{U}_j))$. Hence, we can apply Theorem 2.1 and Lemma 4.3 to the matrices $\mathbf{A}^*$ and $\mathbf{Y}_q$ and conclude that if we let

$$\gamma_k = \frac{\sigma_k - \sigma_{k+1}}{\sigma_{k+1}} > 0 \quad \text{and} \quad \Delta^*(\mathbf{Y}_q, t, k)_{2,F} = 4\, \frac{\|\mathbf{U}_{k,\perp}^*\mathbf{Y}_q(\mathbf{U}_k^*\mathbf{Y}_q)^\dagger\|_{2,F}}{2^{(2t+1)\,\min\{\sqrt{\gamma_k}\,,\,1\}}}$$

then, there exists an h-dimensional left dominant subspace $\tilde{\mathcal{S}}$ of $\mathbf{A}^*$ such that

$$\|\sin\mathbf{\Theta}(\mathcal{K}_{q,t}^*, \tilde{\mathcal{S}})\|_{2,F} \leq 4\,\Delta(\mathbf{X}, q, j)_{2,F}\, \frac{\sigma_{j+1}}{\sigma_j} + \Delta^*(\mathbf{Y}_q, t, k)_{2,F}\, \frac{\sigma_{k+1}}{\sigma_k}\,,$$

where we have also applied Eq. (22). It is clear that $\tilde{\mathcal{S}}$ is an h-dimensional right dominant subspace of $\mathbf{A}$.

Step 3: computing $\mathcal{K}_{q,t}^*$. We end the proof by noticing the following facts: on the one hand, recall that

$$\mathcal{K}_q = R(\mathbf{A}\mathbf{X}) + R((\mathbf{A}\mathbf{A}^*)\mathbf{A}\mathbf{X}) + \cdots + R((\mathbf{A}\mathbf{A}^*)^q\mathbf{A}\mathbf{X}) = R(\mathbf{Y}_q)\,. \tag{23}$$

Similarly, notice that

$$\mathcal{K}_{q,t}^* = R(\mathbf{A}^*\mathbf{Y}_q) + R((\mathbf{A}^*\mathbf{A})\mathbf{A}^*\mathbf{Y}_q) + \cdots + R((\mathbf{A}^*\mathbf{A})^t\mathbf{A}^*\mathbf{Y}_q)\,.$$

If we consider the identity in Eq. (23) and we let $0 \leq \ell \leq t$ then

$$R((\mathbf{A}^*\mathbf{A})^\ell\mathbf{A}^*\mathbf{Y}_q) = R((\mathbf{A}^*\mathbf{A})^{\ell+1}\mathbf{X}) + R((\mathbf{A}^*\mathbf{A})^{\ell+2}\mathbf{X}) + \cdots + R((\mathbf{A}^*\mathbf{A})^{\ell+q+1}\mathbf{X})\,.$$

The previous facts show that

$$\mathcal{K}_{q,t}^* = R((\mathbf{A}^*\mathbf{A})\mathbf{X}) + R((\mathbf{A}^*\mathbf{A})^2\mathbf{X}) + \cdots + R((\mathbf{A}^*\mathbf{A})^{q+t+1}\mathbf{X})\,.$$

□

5.3 Proofs of Proposition 2.4 and Theorem 2.7

Proof of Proposition 2.4. In order to prove the identity in Eq. (4) that is,

$$\|\mathbf{U}_{k,\perp}^*\mathbf{Y}(\mathbf{U}_k^*\mathbf{Y})^\dagger\|_{2,F} = \|\tan\mathbf{\Theta}(\tilde{\mathcal{Y}}, \mathcal{U}_k)\|_{2,F}\,,$$

we have to analyze the strictly positive singular values of $\mathbf{U}_{k,\perp}^*\mathbf{Y}(\mathbf{U}_k^*\mathbf{Y})^\dagger$. On the one hand, we have that

$$s(\mathbf{U}_{k,\perp}^*\mathbf{Y}(\mathbf{U}_k^*\mathbf{Y})^\dagger) = s((\mathbf{U}_{k,\perp}^*\mathbf{Y}(\mathbf{U}_k^*\mathbf{Y})^\dagger)^*)\,.$$

On the other hand, notice that

$$\begin{aligned} s^2((\mathbf{U}_{k,\perp}^*\mathbf{Y}(\mathbf{U}_k^*\mathbf{Y})^\dagger)^*) &= \lambda(\mathbf{U}_{k,\perp}^*\mathbf{Y}(\mathbf{Y}^*\mathbf{U}_k\mathbf{U}_k^*\mathbf{Y})^\dagger\mathbf{Y}^*\mathbf{U}_{k,\perp}) \\ &= \lambda((\mathbf{Y}^*\mathbf{U}_k\mathbf{U}_k^*\mathbf{Y})^\dagger\mathbf{Y}^*\mathbf{U}_{k,\perp}\mathbf{U}_{k,\perp}^*\mathbf{Y}) \\ &= \lambda((\mathbf{Y}^*\mathbf{U}_k\mathbf{U}_k^*\mathbf{Y})^\dagger(\mathbf{Y}^*\mathbf{Y}-\mathbf{Y}^*\mathbf{U}_k\mathbf{U}_k^*\mathbf{Y}))\,. \end{aligned}$$

where we have used that $\mathbf{Z}^\dagger(\mathbf{Z}^\dagger)^*=(\mathbf{Z}^*\mathbf{Z})^\dagger$ and that for matrices $\mathbf{D}$, $\mathbf{E}$, the nonzero eigenvalues of $\mathbf{DE}$ coincide (counting multiplicities) with those of $\mathbf{ED}$. That is, there is some abuse of notation since the vectors above have different sizes in general, so the identities are correct up to some zero entries (which will not modify the norms $\|\mathbf{U}_{k,\perp}^*\mathbf{Y}(\mathbf{U}_k^*\mathbf{Y})^\dagger\|_{2,F}$ in our argument). If we consider the last expression, we now introduce the function

$$f(x)=x^\dagger(1-x) \quad \text{for} \quad x\in[0,1]\,,$$

where $x^\dagger=x^{-1}$ for $x\neq 0$ and $0^\dagger=0$; we conclude that the (strictly) positive singular values of $\mathbf{U}_{k,\perp}^*\mathbf{Y}(\mathbf{U}_k^*\mathbf{Y})^\dagger$ are $f(\lambda_1)^{1/2},\ldots,f(\lambda_t)^{1/2}>0$, where $\lambda_1,\ldots,\lambda_t>0$ are the positive eigenvalues of $\mathbf{Y}^*\mathbf{U}_k\mathbf{U}_k^*\mathbf{Y}$; notice that $t\leq k$ in this case.

Again, we see that $\lambda(\mathbf{Y}^*\mathbf{U}_k\mathbf{U}_k^*\mathbf{Y})=\lambda(\mathbf{Y}\mathbf{Y}^*\mathbf{U}_k\mathbf{U}_k^*\mathbf{Y}\mathbf{Y}^*)=s^2(\mathbf{U}_k\mathbf{U}_k^*\mathbf{Y}\mathbf{Y}^*)=s^2(\mathbf{PQ})$ where $\mathbf{P}$ and $\mathbf{Q}$ denote the orthogonal projections onto $R(\mathbf{U}_k)=\mathcal{U}_k$ and $R(\mathbf{Y})=\mathcal{Y}$ and the equalities are up to some zero entries. Moreover, we let $\mathbf{Q}\wedge\mathbf{P}^\perp\leq\mathbf{Q}$ denote the orthogonal projection onto $\mathcal{Y}\cap\mathcal{U}_k^\perp$. Then, $\mathbf{Q}-\mathbf{Q}\wedge\mathbf{P}^\perp$ is the orthogonal projection onto $\tilde{\mathcal{Y}}=\mathcal{Y}\ominus(\mathcal{Y}\cap\mathcal{U}_k^\perp)$. Notice that

$$\mathbf{PQ}=\mathbf{P}(\mathbf{Q}-\mathbf{Q}\wedge\mathbf{P}^\perp)\,.$$

Since $\ker(\mathbf{P})\cap R(\mathbf{Q}-\mathbf{Q}\wedge\mathbf{P}^\perp)=\{0\}$ we now see that $\operatorname{rank}(\mathbf{P}(\mathbf{Q}-\mathbf{Q}\wedge\mathbf{P}^\perp))=\operatorname{rank}(\mathbf{Q}-\mathbf{Q}\wedge\mathbf{P}^\perp)=\dim\tilde{\mathcal{Y}}=:d$ and the positive singular values of $\mathbf{PQ}=\mathbf{P}(\mathbf{Q}-\mathbf{Q}\wedge\mathbf{P}^\perp)$ correspond $\cos(\mathbf{\Theta}(\mathcal{U}_k,\tilde{\mathcal{Y}}))\in\mathbb{R}^d$. Finally, using the previous notation and facts, we now see that $d=t\leq k$ and $(\lambda_1,\ldots,\lambda_d)=\cos^2(\mathbf{\Theta}(\mathcal{U}_k,\tilde{\mathcal{Y}}))$; therefore, $(f(\lambda_1)^{1/2},\ldots,f(\lambda_d)^{1/2})=\tan(\mathbf{\Theta}(\mathcal{U}_k,\tilde{\mathcal{Y}}))$. □

Proof of Theorem 2.7. We analyze the strictly positive singular values of $\mathbf{U}_{k,\perp}^*\mathbf{Y}_q(\mathbf{U}_k^*\mathbf{Y}_q)^\dagger$ as a function of $q\geq 1$. Arguing as in the proof of Proposition 2.4 we get that

$$s^2(\mathbf{U}_{k,\perp}^*\mathbf{Y}_q(\mathbf{U}_k^*\mathbf{Y}_q)^\dagger)=\lambda((\mathbf{Y}_q^*\mathbf{U}_k\mathbf{U}_k^*\mathbf{Y}_q)^\dagger(\mathbf{Y}_q^*\mathbf{Y}_q-\mathbf{Y}_q^*\mathbf{U}_k\mathbf{U}_k^*\mathbf{Y}_q))\,.$$

where the vectors above have different sizes in general so the identities are correct up to some zero entries, which will not modify the norms $\|\cdot\|_{2,F}$ of these vectors. If we consider the last expression, and take into account that the function

$$f(x)=x^\dagger(1-x) \quad \text{for} \quad x\in[0,1]$$

is decreasing in $(0,1]$, we conclude that the positive singular values $\mathbf{U}_{k,\perp}^*\mathbf{Y}_q(\mathbf{U}_k^*\mathbf{Y}_q)^\dagger$ are non-increasing as a function of the eigenvalues $\lambda(\mathbf{Y}_q^*\mathbf{U}_k\mathbf{U}_k^*\mathbf{Y}_q)$ as long as the rank of $\mathbf{Y}_q^*\mathbf{U}_k\mathbf{U}_k^*\mathbf{Y}_q$ is preserved. Since $\lambda(\mathbf{Y}_q^*\mathbf{U}_k\mathbf{U}_k^*\mathbf{Y}_q)=\lambda(\mathbf{U}_k^*\mathbf{Y}_q\mathbf{Y}_q^*\mathbf{U}_k)$ (where the equality is up to some zero entries) we see that the positive singular values $\mathbf{U}_{k,\perp}^*\mathbf{Y}_q(\mathbf{U}_k^*\mathbf{Y}_q)^\dagger$ are non-increasing as a function of the projection $\mathbf{Y}_q\mathbf{Y}_q^*$, where we consider the operator order between these projections, as long as the rank of $\mathbf{Y}_q^*\mathbf{U}_k\mathbf{U}_k^*\mathbf{Y}_q$ is preserved. It is clear that $\mathcal{K}_q\subseteq\mathcal{K}_{q'}$ whenever $q\leq q'$ and therefore, $\mathbf{Y}_q\mathbf{Y}_q^*\leq\mathbf{Y}_{q'}\mathbf{Y}_{q'}^*$; thus we are left to consider the problem of the stabilization of

$$\operatorname{rank}(\mathbf{Y}_q^*\mathbf{U}_k\mathbf{U}_k^*\mathbf{Y}_q)=\operatorname{rank}(\mathbf{U}_k^*\mathbf{Y}_q\mathbf{Y}_q^*\mathbf{U}_k)=\operatorname{rank}(\mathbf{U}_k^*\mathbf{Y}_q)$$

as a function of q. Since $R(\mathbf{Y}_q) = \mathcal{K}_q$ we consider a convenient description of the elements of this space: in this case, given $\mathbf{u} \in \mathcal{K}_q$ there exists a polynomial $p(x) \in \mathbb{K}[x]$ of degree at most q and $\mathbf{v} \in \mathbb{K}^r$ such that if $\phi(x) = x\, p(x^2) \in \mathbb{K}[x]$ then

$$\mathbf{u} = \mathbf{U}\phi(\mathbf{\Sigma})\mathbf{V}^*\mathbf{X}\mathbf{v}\,,$$

so then

$$\mathbf{U}_k^*\mathbf{u} = \phi(\mathbf{\Sigma}_k)\mathbf{V}_k^*\mathbf{X}\mathbf{v}\,.$$

Let $q_0 = \#\{\sigma_1, \ldots, \sigma_h\} - 1$, where $\#\{\sigma_1, \ldots, \sigma_h\}$ denotes the number of different singular values of $\mathbf{A}$ between σ_1 and σ_h. We can take a partition $\{C_1, \ldots, C_{q_0+1}\}$ of $\{1, \ldots, k\}$ in such a way that $\ell, \tilde{\ell} \in C_t$ if and only if $\sigma_\ell = \sigma_{\tilde{\ell}}$. We select a representative $g(t) \in C_t$, for $1 \le t \le q_0 + 1$. Consider the diagonal orthogonal projections $\mathbf{P}_1, \ldots, \mathbf{P}_{q_0+1} \in \mathbb{K}^{k\times k}$ such that $\mathbf{P}_t$ projects onto the subspace $\operatorname{Span}\{\mathbf{e}_\ell : \ell \in C_t\}$, where $\{\mathbf{e}_1, \ldots, \mathbf{e}_k\}$ denotes the canonical basis of $\mathbb{K}^k$. In this case,

$$\mathbf{U}_k^*\mathbf{u} = \phi(\mathbf{\Sigma}_k)\mathbf{V}_k^*\mathbf{X}\mathbf{v} = \sum_{t=1}^{q_0+1} \phi(\sigma_{g(t)})\mathbf{P}_t\mathbf{V}_k^*\mathbf{X}\mathbf{v} \subseteq \sum_{t=1}^{q_0+1} R(\mathbf{P}_t\mathbf{V}_k^*\mathbf{X}) = \mathcal{R}\,.$$

Hence, $R(\mathbf{U}_k^*\mathbf{Y}_q) \subset \mathcal{R}$. Assume that $q \ge q_0$; for $1 \le t \le q_0 + 1$ let $p_t(x) \in \mathbb{K}[x]$ with degree at most q such that $p_t(\sigma_{g(\ell)}^2) = \delta_{t,\ell}$, for $1 \le \ell \le q_0 + 1$ (for instance, take the Lagrange's polynomials). Let $\phi_t(x) = x\, p_t(x^2)$ and notice that

$$\mathbf{U}_k^*(\mathbf{U}\phi_t(\mathbf{\Sigma})\mathbf{V}^*\mathbf{X}) = \sigma_{g(t)}\mathbf{P}_t\mathbf{V}_k^*\mathbf{X}\,.$$

The previous facts show that

$$R(\mathbf{P}_t\mathbf{V}_k^*\mathbf{X}) \subset R(\mathbf{U}_k^*\mathbf{Y}_q) \quad \text{for} \quad 1 \le t \le q_0 + 1\,,$$

since $\sigma_1, \ldots, \sigma_k > 0$ ($k \le \operatorname{rank}(\mathbf{A})$), and then $R(\mathbf{U}_k^*\mathbf{Y}_q) = \mathcal{R}$, for $q \ge q_0$.

As already explained, the fact that the range $R(\mathbf{U}_k^*\mathbf{Y}_q)$ stabilizes for $q \ge q_0$ implies that the positive singular values $s(\mathbf{U}_{k,\perp}^*\mathbf{Y}_q(\mathbf{U}_k^*\mathbf{Y}_q)^\dagger)$ are (entry-wise) non-increasing functions of $q \ge q_0$. In particular,

$$\|\mathbf{U}_{k,\perp}^*\mathbf{Y}_{q'}(\mathbf{U}_k^*\mathbf{Y}_{q'})^\dagger\|_{2,F} \le \|\mathbf{U}_{k,\perp}^*\mathbf{Y}_q(\mathbf{U}_k^*\mathbf{Y}_q)^\dagger\|_{2,F} \quad \text{for} \quad q_0 \le q \le q'\,.$$

□

5.4 Proof of Theorem 2.11

Proof of Theorem 2.11. In what follows we consider the notation from Theorem 2.11. We let $\mathbf{U}_K \in \mathbb{K}^{m\times\tilde{d}}$ denote the matrix whose columns form an orthonormal basis of the Krylov space $\mathcal{K}_{q+t+1}$ constructed in terms of $\mathbf{A}$ and $\mathbf{X}$; further, we let $\hat{\mathbf{U}}_i$ denote the matrix whose columns are the top i columns of the output of Algorithm 2.1.

Step 1: applying Theorem 2.6. Let $\mathcal{K}_{q,t}^* \subseteq \mathbb{K}^n$ be the subspace defined in Theorem 2.6, that is

$$\mathcal{K}_{q,t}^* = R((\mathbf{A}^*\mathbf{A})\mathbf{X}) + R((\mathbf{A}^*\mathbf{A})^2\mathbf{X}) + \cdots + R((\mathbf{A}^*\mathbf{A})^{q+t+1}\mathbf{X})\,.$$

By Theorem 2.6, there exists an h-dimensional right dominant subspace $\tilde{\mathcal{S}}$ for $\mathbf{A}$ such that

$$\|\sin\mathbf{\Theta}(\mathcal{K}_{q,t}^*, \tilde{\mathcal{S}})\|_{2,F} \le 4\,\Delta(\mathbf{X}, q, j)_{2,F}\,\frac{\sigma_{j+1}}{\sigma_j} + \Delta^*(\mathbf{Y}_q, t, k)_{2,F}\,\frac{\sigma_{k+1}}{\sigma_k}\,. \tag{24}$$

By the hypothesis in Eq. (9) and Eq. (24) we see that $\|\sin\mathbf{\Theta}(\mathcal{K}_{q,t}^*,\tilde{\mathcal{S}})\|_2 \leq \frac{1}{\sqrt{2}}$ and then,

$$\mathbf{\Theta}(\mathcal{K}_{q,t}^*,\tilde{\mathcal{S}}) \leq \frac{\pi}{4}\,\mathbf{I}_h\,.$$

On the other hand, notice that $\mathbf{A}(\mathcal{K}_{q,t}^*) \subset \mathcal{K}_{q+t+1}$.

Step 2: applying the DIKM-I theory. We now argue as in the proof of [5, Theorem 2.3.]. Indeed, by [3, Lemma 8] we have that

$$\mathbf{A} - \hat{\mathbf{U}}_i\hat{\mathbf{U}}_i^*\mathbf{A} = \mathbf{A} - \mathbf{U}_K(\mathbf{U}_K^*\mathbf{A})_i \quad \text{for} \quad 1 \leq i \leq h\,,$$

where $(\mathbf{U}_K^*\mathbf{A})_i$ denotes a best rank-i approximation of $\mathbf{U}_K^*\mathbf{A}$. By the same result, we also get that $\mathbf{U}_K(\mathbf{U}_K^*\mathbf{A})_i$ is the best rank i approximation of $\mathbf{A}$ from $\mathcal{K}_{q,t}^*$ in the Frobenius norm, i.e.

$$\|\mathbf{A} - \mathbf{U}_K(\mathbf{U}_K^*\mathbf{A})_i\|_F = \min_{\mathrm{rank}(\mathbf{Y})\leq i} \|\mathbf{A} - \mathbf{U}_K\mathbf{Y}\|_F\,. \tag{25}$$

We now consider a SVD, $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^*$ such that the top h columns of $\mathbf{V}$ span the h-dimensional right dominant subspace $R(\mathbf{V}_h) = \mathcal{V}_h = \tilde{\mathcal{S}}$ (recall that this can always we done). We now set

$$\mathbf{A} = \mathbf{A}_i + \mathbf{A}_{i,\perp} \quad \text{where} \quad \mathbf{A}_i = \mathbf{U}_i\mathbf{\Sigma}_i\mathbf{V}_i^* \quad \text{and} \quad \mathbf{A}_{i,\perp} = \mathbf{A} - \mathbf{A}_i\,.$$

Then, by [5, Lemma 7.2] we get that

$$\|\mathbf{A} - \hat{\mathbf{U}}_i\hat{\mathbf{U}}_i^*\mathbf{A}\|_F^2 \leq \|\mathbf{A} - \mathbf{A}_i\|_F^2 + \|\mathbf{A}_i - \mathbf{U}_K\mathbf{U}_K^*\mathbf{A}_i\|_F^2\,. \tag{26}$$

Step 3: bounding the second term in Eq. (26). Since $\mathbf{\Theta}(R(\mathbf{V}_h),\mathcal{K}_{q,t}^*) \leq \frac{\pi}{4}\,\mathbf{I}_h$ we have that $\mathbf{V}_h^*(\mathcal{K}_{q,t}^*) = R(\mathbf{V}_h^*)$; then,

$$\mathrm{rank}(\mathbf{V}_i^*(\mathbf{A}^*\mathbf{A})\mathbf{X}) = \mathrm{rank}(\mathbf{\Sigma}_i^2\mathbf{V}_i^*\mathbf{X}) = \mathrm{rank}(\mathbf{V}_i^*\mathbf{X}) = i$$

and we see that $\mathbf{V}_i^*(\mathcal{K}_{q,t}^*) = R(\mathbf{V}_i^*)$. Thus, we can apply Lemma 6.6 in this context. Hence, we consider the principal vectors $\{\mathbf{w}_1,\ldots,\mathbf{w}_i\} \subset \mathcal{K}_{q,t}^*$ corresponding to the pair $(\mathcal{K}_{q,t}^*, R(\mathbf{V}_i))$. Moreover, we let $\mathbf{Q} \in \mathbb{K}^{n\times i}$ be an isometry with $R(\mathbf{Q}) = \mathrm{Span}\{\mathbf{w}_1,\ldots,\mathbf{w}_i\}$ so that $R(\mathbf{AQ}) \subset \mathbf{A}(\mathcal{K}_{q,t}^*) \subset \mathcal{K}_{q+t+1}$. The previous facts together with Lemma 6.6 show that

$$\begin{aligned}
\|\mathbf{A}_i - \mathbf{U}_K\mathbf{U}_K^*\mathbf{A}_i\|_F &\leq \|(\mathbf{I} - \mathbf{AQ}(\mathbf{AQ})^\dagger)\mathbf{A}_i\|_F = \|\mathbf{A}_i - \mathbf{AQ}(\mathbf{AQ})^\dagger\mathbf{A}_i\|_F \\
&\leq \|\mathbf{A} - \mathbf{A}_i\|_2\,\|\tan\mathbf{\Theta}(\mathcal{K}_{q,t}^*, R(\mathbf{V}_i))\|_F \\
&\leq \|\mathbf{A} - \mathbf{A}_i\|_2\,\|\tan\mathbf{\Theta}(\mathcal{K}_{q,t}^*, R(\mathbf{V}_h))\|_F\,.
\end{aligned}$$

Since $\mathbf{\Theta}(R(\mathbf{V}_h),\mathcal{K}_{q,t}^*) \leq \frac{\pi}{4}\,\mathbf{I}_h$ then

$$\begin{aligned}
\|\tan\mathbf{\Theta}(\mathcal{K}_{q,t}^*, R(\mathbf{V}_h))\|_F &\leq \sqrt{2}\,\|\sin\mathbf{\Theta}(\mathcal{K}_{q,t}^*, R(\mathbf{V}_h))\|_F \\
&\leq \sqrt{2}\,(4\,\Delta(\mathbf{X},q,j)_F\,\frac{\sigma_{j+1}}{\sigma_j} + \Delta^*(\mathbf{Y}_q,t,k)_F\,\frac{\sigma_{k+1}}{\sigma_k})\,,
\end{aligned}$$

where we have used Eq. (24). Therefore, the previous inequalities imply that

$$\|\mathbf{A} - \hat{\mathbf{U}}_i\hat{\mathbf{U}}_i^*\mathbf{A}\|_F \leq \|\mathbf{A} - \mathbf{A}_i\|_F + \delta_i \tag{27}$$

where δ_i is as defined in Theorem 2.11. This proves the upper bound in Eq. (10) for the Frobenius norm. In order to prove the bound for the spectral norm, recall that by [11, Theorem 3.4.] we get that Eq. (27) implies that

$$\|\mathbf{A} - \hat{\mathbf{U}}_i\hat{\mathbf{U}}_i^*\mathbf{A}\|_2 \leq \|\mathbf{A} - \mathbf{A}_i\|_2 + \delta_i\,,$$

since $\mathrm{rank}(\hat{\mathbf{U}}_i\hat{\mathbf{U}}_i^*\mathbf{A}) \leq i$. □

6 Appendix

In this section we include a number of technical results that are needed for the proofs of our main results. Most of these results are elementary and can be found in the literature; we include the versions that are well suited for our exposition together with their proofs, for the convenience of the reader.

Proposition 6.1. *Let* $\mathbf{V} \in \mathbb{K}^{n\times k}$ *be an isometry and let* $\mathcal{V}'$, $\mathcal{X} \subset \mathbb{K}^n$ *be subspaces such that* $\dim \mathcal{X} \geq \dim \mathcal{V}' = j$, $\mathcal{V}' \subset (\ker \mathbf{V}^*)^\perp = R(\mathbf{V})$ *and* $\mathbf{\Theta}(\mathcal{X}, \mathcal{V}') < \pi/2\, \mathbf{I}_j$. *Then* $\mathcal{W} = \mathbf{V}^*\mathcal{X} \subset \mathbb{K}^k$ *is such that* $\dim \mathcal{W} \geq j$ *and if we let* $\mathcal{H}' = \mathbf{V}^*\mathcal{V}'$ *then*

$$\mathbf{\Theta}(\mathcal{W}, \mathcal{H}') \leq \mathbf{\Theta}(\mathcal{X}, \mathcal{V}') \in \mathbb{R}^{j\times j} .$$

Proof. First notice that

$$\mathbf{P}_\mathcal{X}\mathbf{P}_{\mathcal{V}'}\mathbf{P}_\mathcal{X} \leq \mathbf{P}_\mathcal{X}\mathbf{V}\mathbf{V}^*\mathbf{P}_\mathcal{X} .$$

By hypothesis $\operatorname{rank}(\mathbf{P}_\mathcal{X}\mathbf{P}_{\mathcal{V}'}\mathbf{P}_\mathcal{X}) = j$ which shows that $\dim \mathcal{W} = \operatorname{rank}(\mathbf{V}^*\mathbf{P}_\mathcal{X}) \geq j$. On the other hand, since $\mathbf{V}$ is an isometry then

$$\mathbf{\Theta}(\mathcal{W}, \mathcal{H}') = \mathbf{\Theta}(\mathbf{V}\mathcal{W}, \mathbf{V}\mathcal{H}') = \mathbf{\Theta}(\mathbf{V}\mathbf{V}^*\mathcal{X}, \mathcal{V}') .$$

Consider $\mathbf{D} = \mathbf{V}\mathbf{V}^*\mathbf{P}_\mathcal{X}\mathbf{V}\mathbf{V}^*$; then $R(\mathbf{D}) = \mathbf{V}\mathbf{V}^*\mathcal{X}$, so $\dim R(\mathbf{D}) = \dim \mathcal{W} \geq j$. Moreover,

$$0 \leq \mathbf{D} \leq \mathbf{P}_{R(\mathbf{D})} \implies \mathbf{P}_{\mathcal{V}'}\mathbf{P}_\mathcal{X}\mathbf{P}_{\mathcal{V}'} = \mathbf{P}_{\mathcal{V}'}\mathbf{D}\mathbf{P}_{\mathcal{V}'} \leq \mathbf{P}_{\mathcal{V}'}\mathbf{P}_{R(\mathbf{D})}\mathbf{P}_{\mathcal{V}'} ,$$

where we used that $\mathbf{P}_{\mathcal{V}'}\mathbf{V}\mathbf{V}^* = \mathbf{P}_{\mathcal{V}'}$. Then, $\cos^2\mathbf{\Theta}(\mathcal{X}, \mathcal{V}') \leq \cos^2\mathbf{\Theta}(\mathbf{V}\mathbf{V}^*\mathcal{X}, \mathcal{V}') \in \mathbb{R}^{j\times j}$ and the result follows from the fact that $f(x) = \cos^2(x)$ is a decreasing function on $[0, \pi/2]$. □

Proposition 6.2. *Let* $\mathbf{B} \in \mathbb{K}^{k\times k}$ *be such that* $\mathbf{B} = \mathbf{B}^*$ *and let* $\mathcal{H}'$, $\mathcal{W}' \subset \mathbb{K}^k$ *be subspaces such that* $\mathbf{P}_{\mathcal{H}'}\mathbf{B} = \mathbf{B}\mathbf{P}_{\mathcal{H}'}$, $\dim \mathcal{H}' = \dim \mathcal{W}'$ *and* $\mathcal{H}'$, $\mathcal{W}' \subset \ker(\mathbf{B})^\perp$. *If we let* $\mathbf{B}\mathcal{W}' = \mathcal{T}'$,

$$\| \sin \mathbf{\Theta}(\mathcal{H}', \mathcal{T}')\|_{2,F} \leq \|\mathbf{B}(\mathbf{I} - \mathbf{P}_{\mathcal{H}'})\|_2 \,\|\mathbf{B}^\dagger\|_2 \| \sin \mathbf{\Theta}(\mathcal{H}', \mathcal{W}')\|_{2,F} .$$

Proof. Notice that $(\mathbf{B}\mathbf{P}_{\mathcal{H}'})^\dagger = \mathbf{P}_{\mathcal{H}'}\mathbf{B}^\dagger = \mathbf{B}^\dagger\mathbf{P}_{\mathcal{H}'}$. Then,

$$(\mathbf{I} - \mathbf{P}_{\mathcal{H}'})\mathbf{P}_{\mathcal{T}'} = (\mathbf{I} - \mathbf{P}_{\mathcal{H}'})(\mathbf{B}\mathbf{P}_{\mathcal{W}'})\,(\mathbf{B}\mathbf{P}_{\mathcal{W}'})^\dagger = (\mathbf{B}(\mathbf{I} - \mathbf{P}_{\mathcal{H}'}))\,(\mathbf{I} - \mathbf{P}_{\mathcal{H}'})\mathbf{P}_{\mathcal{W}'}\,(\mathbf{B}\mathbf{P}_{\mathcal{W}'})^\dagger .$$

Also, notice that $(\mathbf{B}\mathbf{P}_{\mathcal{W}'})^\dagger = \mathbf{P}_{\mathcal{W}'}\mathbf{B}^\dagger\mathbf{P}_{\mathcal{T}'}$; in particular, $\|(\mathbf{B}\mathbf{P}_{\mathcal{W}'})^\dagger\|_2 \leq \|\mathbf{B}^\dagger\|_2$. Finally, since $\dim \mathcal{T}' = \dim \mathcal{W}' = \dim \mathcal{H}'$ the previous facts imply that

$$\| \sin \mathbf{\Theta}(\mathcal{H}', \mathcal{T}')\|_{2,F} \leq \|\mathbf{B}(\mathbf{I} - \mathbf{P}_{\mathcal{H}'})\|_2 \,\|\mathbf{B}^\dagger\|_2 \| \sin \mathbf{\Theta}(\mathcal{H}', \mathcal{W}')\|_{2,F} .$$

□

Proposition 6.3. *Let* $\mathcal{T}'$, $\mathcal{T}''$ *and* $\mathcal{H}'$, $\mathcal{H}''$ *be pairs of mutually orthogonal subspaces in* $\mathbb{K}^k$, *such that* $\dim(\mathcal{H}') \leq \dim(\mathcal{T}')$ *and* $\dim(\mathcal{H}'') \leq \dim(\mathcal{T}'')$. *Consider the subspaces in* $\mathbb{K}^k$ *given by the (orthogonal) sums* $\mathcal{T} = \mathcal{T}' \oplus \mathcal{T}''$ *and* $\mathcal{H} = \mathcal{H}' \oplus \mathcal{H}''$, *so* $\dim(\mathcal{H}) \leq \dim(\mathcal{T})$. *In this case we have that*

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

Proof. As usual, we compute the sines of the principal angles in terms of singular values of products of projections: in this case, using that $\mathbf{P}_{\mathcal{H}} = \mathbf{P}_{\mathcal{H}'} + \mathbf{P}_{\mathcal{H}''}$ and $\mathbf{P}_{\mathcal{T}} = \mathbf{P}_{\mathcal{T}'} + \mathbf{P}_{\mathcal{T}''}$ we have that

$$\begin{aligned} \|\sin \mathbf{\Theta}(\mathcal{T},\mathcal{H})\|_{2,F} &= \|(\mathbf{I}-\mathbf{P}_{\mathcal{T}})\mathbf{P}_{\mathcal{H}}\|_{2,F} = \|(\mathbf{P}_{\mathcal{H}}-\mathbf{P}_{\mathcal{T}})\mathbf{P}_{\mathcal{H}}\|_{2,F} \\ &= \|(\mathbf{P}_{\mathcal{H}'}-\mathbf{P}_{\mathcal{T}'}+\mathbf{P}_{\mathcal{H}''}-\mathbf{P}_{\mathcal{T}''})(\mathbf{P}_{\mathcal{H}'}+\mathbf{P}_{\mathcal{H}''})\|_{2,F} \\ &\leq \|(\mathbf{I}-\mathbf{P}_{\mathcal{T}'})\mathbf{P}_{\mathcal{H}'}\|_{2,F} + \|(\mathbf{I}-\mathbf{P}_{\mathcal{T}''})\mathbf{P}_{\mathcal{H}''}\|_{2,F} + \\ &\quad \|\mathbf{P}_{\mathcal{T}'}\mathbf{P}_{\mathcal{H}''}\|_{2,F} + \|\mathbf{P}_{\mathcal{T}''}\mathbf{P}_{\mathcal{H}'}\|_{2,F} \,. \end{aligned}$$

Now, notice that $\mathbf{P}_{\mathcal{T}'} \leq \mathbf{I} - \mathbf{P}_{\mathcal{T}''}$ so

$$\|\mathbf{P}_{\mathcal{T}'}\mathbf{P}_{\mathcal{H}''}\|_{2,F} \leq \|(\mathbf{I}-\mathbf{P}_{\mathcal{T}''})\mathbf{P}_{\mathcal{H}''}\|_{2,F} = \|\sin \mathbf{\Theta}(\mathcal{T}'',\mathcal{H}'')\|_{2,F} \,.$$

Similarly, $\|\mathbf{P}_{\mathcal{T}''}\mathbf{P}_{\mathcal{H}'}\|_{2,F} \leq \|(\mathbf{I}-\mathbf{P}_{\mathcal{T}'})\mathbf{P}_{\mathcal{H}'}\|_{2,F} = \|\sin \mathbf{\Theta}(\mathcal{T}',\mathcal{H}')\|_{2,F}$. □

Proposition 6.4. *Let* $\mathbf{B} \in \mathbb{K}^{p\times q}$ *and let* $C \in \mathbb{K}^{q\times r}$ *with* $R(\mathbf{C}) = \mathcal{V} \subset \mathbb{K}^q$ *such that* $\mathcal{V} \subset \ker \mathbf{B}^{\perp}$*. Then*

$$(\mathbf{BC})^{\dagger} = \mathbf{C}^{\dagger}(\mathbf{BP}_{\mathcal{V}})^{\dagger} \,.$$

Proof. In this case $R(\mathbf{BC}) = \mathbf{B}\mathcal{V}$ and $\ker \mathbf{BC} = \ker \mathbf{C}$. Moreover,

$$\mathbf{BCC}^{\dagger}(\mathbf{BP}_{\mathcal{V}})^{\dagger} = \mathbf{BP}_{\mathcal{V}}(\mathbf{BP}_{\mathcal{V}})^{\dagger} = \mathbf{P}_{\mathbf{B}\mathcal{V}}$$

and

$$\mathbf{C}^{\dagger}(\mathbf{BP}_{\mathcal{V}})^{\dagger}\mathbf{BC}=\mathbf{C}^{\dagger}(\mathbf{BP}_{\mathcal{V}})^{\dagger}\mathbf{BP}_{\mathcal{V}}\mathbf{C}=\mathbf{C}^{\dagger}\mathbf{P}_{\ker(\mathbf{BP}_{\mathcal{V}})^{\perp}}\mathbf{C}=\mathbf{P}_{\ker \mathbf{C}^{\perp}} \,,$$

where we used that $\ker(\mathbf{BP}_{\mathcal{V}}) = \mathcal{V}^{\perp}$, since $\mathcal{V} \subset \ker \mathbf{B}^{\perp}$. □

Let $\mathbf{C} \in \mathbb{K}^{m\times c}$ have rank p. For $1 \leq i \leq p$ we define

$$\mathcal{P}^{\xi}_{\mathbf{C},\,i}(\mathbf{A}) = \mathbf{C} \cdot \operatorname{argmin}_{\operatorname{rank}(\mathbf{Y})\leq i}\|\mathbf{A}-\mathbf{CY}\|_{\xi} \quad \text{for} \quad \xi = 2, F \,.$$

Due to the optimality properties of the projection $\mathbf{CC}^{\dagger}$ (see [11]) we get that

$$\|\mathbf{A}-\mathbf{CC}^{\dagger}\mathbf{A}\|_{2,F} \leq \|\mathbf{A}-\mathcal{P}^{\xi}_{\mathbf{C},\,i}(\mathbf{A})\|_{2,F} \,. \tag{28}$$

The following result is [21, Lemma C.5] (see also [3]).

Lemma 6.5 ([21])**.** *Let* $\mathbf{A} \in \mathbb{K}^{m\times n}$ *and consider a decomposition* $\mathbf{A} = \mathbf{A}_1 + \mathbf{A}_2$*, with rank*$(\mathbf{A}_1) = i$*. Let* $\mathbf{V}_1 \in \mathbb{K}^{n\times i}$ *denote the top right singular vectors of* $\mathbf{A}_1$*. Let* $\mathbf{Z} \in \mathbb{K}^{n\times p}$ *such that rank*$(\mathbf{V}_1^*\mathbf{Z}) = i$ *and let* $\mathbf{C} = \mathbf{AZ}$*. Then rank*$(\mathbf{C}) \geq i$ *and*

$$\|\mathbf{A}_1 - \mathcal{P}^{\xi}_{\mathbf{C},\,i}(\mathbf{A}_1)\|_{2,F} \leq \|\mathbf{A}_2\mathbf{Z}(\mathbf{V}_1^*\mathbf{Z})^{\dagger}\|_{2,F} \,.$$

□

The following is a small variation of [21, Lemma C.1]

Lemma 6.6. *Let* $\mathbf{A} \in \mathbb{K}^{m\times n}$ *and consider the decomposition* $\mathbf{A} = \mathbf{A}_1 + \mathbf{A}_2$*, with* $\mathbf{A}_1\mathbf{A}_2^* = 0$ *and rank*$(\mathbf{A}_1) = i$*. Let* $\mathbf{V}_1 \in \mathbb{K}^{n\times i}$ *and* $\mathbf{V}_2 \in \mathbb{K}^{n\times(n-i)}$ *denote the top right singular vectors of* $\mathbf{A}_1$ *and* $\mathbf{A}_2$ *respectively. Let* $\mathcal{K}^* \subset \mathbb{K}^n$ *be a subspace such that* $\mathbf{V}_1^*(\mathcal{K}^*) = R(\mathbf{V}_1^*)$*. Let* $\{\mathbf{x}_1, \ldots, \mathbf{x}_i\} \subset \mathcal{K}^*$ *denote the principal vectors corresponding to the pair* $(\mathcal{K}^*, R(\mathbf{V}_1))$ *and let* $\mathbf{Q} \in \mathbb{K}^{n\times i}$ *be an isometry with* $R(\mathbf{Q}) =$ *Span*$(\{\mathbf{x}_1, \ldots, \mathbf{x}_i\}) \subset \mathcal{K}^*$*. Then,*

$$\|\mathbf{A}_1 - (\mathbf{AQ})(\mathbf{AQ})^{\dagger}\mathbf{A}_1\|_{2,F} \leq \|\mathbf{A}-\mathbf{A}_1\|_2 \; \|\tan \mathbf{\Theta}(\mathcal{K}^*, R(\mathbf{V}_1))\|_{2,F} \,.$$

Proof. Notice that by construction

$$\mathbf{\Theta}(R(\mathbf{Q}), R(\mathbf{V}_1)) = \mathbf{\Theta}(\mathcal{K}^*, R(\mathbf{V}_1)) < \frac{\pi}{2}\,\mathbf{I}_i\,.$$

Then, we get that $\operatorname{rank}(\mathbf{AQ}) = i$. Hence, we have that

$$\begin{aligned} \|\mathbf{A}_1 - (\mathbf{AQ})(\mathbf{AQ})^\dagger \mathbf{A}_1\|_{2,F} &\leq \|\mathbf{A}_1 - \mathcal{P}^{2,F}_{\mathbf{AQ},\,i}(\mathbf{A}_1)\|_{2,F} \leq \|\mathbf{A}_2\mathbf{Q}(\mathbf{V}_1^*\mathbf{Q})^\dagger\|_{2,F} \\ &\leq \|\mathbf{A}_2\|_2\,\|\mathbf{V}_2^*\mathbf{Q}(\mathbf{V}_1^*\mathbf{Q})^\dagger\|_{2,F} = \|\mathbf{A}_2\|_2\;\|\tan\mathbf{\Theta}(R(\mathbf{Q}), R(\mathbf{V}_1))\|_{2,F} \\ &= \|\mathbf{A} - \mathbf{A}_1\|_2\;\|\tan\mathbf{\Theta}(\mathcal{K}^*, R(\mathbf{V}_1))\|_{2,F}\,, \end{aligned}$$

where we have used Eq. (28), Lemma 6.5, that the isometry $\mathbf{V}_2$ satisfies that $\mathbf{A}_2 = \mathbf{A}_2\mathbf{V}_2\mathbf{V}_2^*$ and the identity

$$\|\mathbf{V}_2^*\mathbf{Q}(\mathbf{V}_1^*\mathbf{Q})^\dagger\|_{2,F} = \|\tan\mathbf{\Theta}(R(\mathbf{Q}), R(\mathbf{V}_1))\|_{2,F}$$

that holds by [5, Lemma 4.3], since $\operatorname{rank}(\mathbf{V}_1^*\mathbf{Q}) = i$. □

Acknowledgment

This research was partially supported by CONICET (PIP 2016 0525/ PIP 0152 CO), ANPCyT (2015 1505/ 2017 0883) and FCE-UNLP (11X829).

References

[1] F. Arrigo, M. Benzi, and C. Fenu, Computation of generalized matrix functions, SIAM J. Matrix Anal. Appl., 37 (2016), pp. 836-860.

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

[3] C. Boutsidis, P. Drineas, M. Magdon-Ismail, Near-optimal column-based matrix reconstruction. SIAM J. Comput. 43 (2014), no. 2, 687-717.

[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.M. 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] P. Drineas, R. Kannan, and M. W. Mahoney, Fast Monte Carlo algorithms for matrices. II. Computing a low-rank approximation to a matrix, SIAM J. Comput., 36 (2006), pp. 158-183.

[7] P. Drineas and M. W. Mahoney, RandNLA: Randomized Numerical Linear Algebra, Commun. ACM, 59 (2016), pp. 80-90.

[8] P. Drineas, M. W. Mahoney, and S. Muthukrishnan, Relative error CUR matrix decompositions, SIAM J. Matrix Anal. Appl., 30 (2008), pp. 844-881.

[9] P. Drineas, M. W. Mahoney, S. Muthukrishnan, and T. Sarlós, Faster least squares approximation, Numer. Math., 117 (2011), pp. 219-249.

[10] G. H. Golub and C. F. Van Loan, Matrix Computations, 4th ed., The Johns Hopkins University Press, Baltimore, 2013.

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

[12] N. Halko, P.-G. Martinsson, and J. Tropp, Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions. SIAM Review, 53(2):217-288, 2011.

[13] J. B. Hawkins and A. Ben-Israel, On generalized matrix functions, Linear Multilinear Algebra, 1 (1973), pp. 163-171.

[14] R. A. Horn and C. R. Johnson, Topics in Matrix Analysis, Cambridge University Press, Cambridge, UK, 1991.

[15] I. T. Jolliffe, Principal Component Analysis. Springer Verlag, New York, 1986.

[16] P. J. Maher, Some norm inequalities concerning generalized inverses, Linear Algebra Appl., 174 (1992), pp. 99-110.

[17] M. Mahoney and P. Drineas, Structural properties underlying high-quality randomized numerical linear algebra algorithms, in Handbook of Big Data, CRC Press, Boca Raton, FL, 2016, pp. 137-154.

[18] C. Musco and C. Musco, Randomized block Krylov methods for stronger and faster approximate singular value decomposition, in Advances in Neural Information Processing Systems 28, C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett, eds., Curran Associates, 2015, pp. 1396-1404.

[19] Y. Saad, Numerical Methods for Large Eigenvalue Problems: Revised Edition, Classics Appl. Math. 66, SIAM, Philadelphia, 2011.

[20] G. W. Stewart and J. Sun, Matrix Perturbation Theory, Academic Press, San Diego, 1990.

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

[22] D. P. Woodruff, Sketching as a tool for numerical linear algebra, Found. Trends Theor. Comput. Sci., 10 (2014), pp. 1-157.

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