Local Lidskii's theorems for unitarily invariant norms

Pedro G. Massey *, Noelia B. Rios * and Demetrio Stojanoff * †
Depto. de Matemática, FCE-UNLP and IAM-CONICET, Argentina

Abstract

Lidskii's additive inequalities (both for eigenvalues and singular values) can be interpreted as an explicit description of global minimizers of functions that are built on unitarily invariant norms, with domains consisting of certain orbits of matrices (under the action of the unitary group). In this paper, we show that Lidskii's inequalities actually describe all global minimizers of such functions and that local minimizers are also global minimizers. We use these results to obtain partial results related to local minimizers of generalized frame operator distances in the context of finite frame theory.

AMS subject classification: 42C15, 15A60.
Keywords: Lidskii's inequality, unitarily invariant norms, majorization, frame operator distance.

Contents

1 Introduction

Lidskii's additive inequalities [17] are ubiquitous in matrix analysis. They are part of the fundamental toolkit to deal with some of the most natural problems in this theory, such as matrix approximation problems (matrix nearness problems) and singular values/eigenvalues inequalities (see [3, 12, 13] and the references therein). Lidskii's inequalities are expressed in terms of an important pre-order between real vectors called majorization. Since majorization is intimately related to tracial inequalities involving convex functions, Lidskii's inequalities can be used to describe the structure of matrices that are optimal with respect to families of entropic-like functionals (see [18, 20, 22, 23]). Lidskii's inequalities also provide some simple relations between the spectra of the sum of selfadjoint matrices and its summands, related to the solution of Horn's conjecture [11]

*Partially supported by CONICET (PIP 0150/14), FONCyT (PICT 1506/15) and FCE-UNLP (11X681), Argentina.

†e-mail addresses: massey@mate.unlp.edu.ar, nbrios@mate.unlp.edu.ar, demetrio@mate.unlp.edu.ar

on the spectra of the sum of selfadjoint matrices, based on the work of A. Klyachko [14] and A. Knutson and T. Tao [15] (see [9] for a historical account and a comprehensive description of the solution of Horn's conjecture).

In the present paper, we consider local versions of Lidskii's inequalities with respect to unitarily invariant norms (u.i.n.). To be more precise, consider a strictly convex u.i.n., denoted by N, on $\mathcal{M}_d(\mathbb{C})$ - the algebra of complex $d \times d$ matrices - and fix a selfadjoint matrix $S \in \mathcal{M}_d(\mathbb{C})$. Fix $\mu \in \mathbb{R}^d$ and let $\mathcal{O}_\mu = \{U^* D_\mu U : U \in \mathcal{U}(d)\}$, where $\mathcal{U}(d)$ denotes the group of unitary matrices and $D_\mu \in \mathcal{M}_d(\mathbb{C})$ denotes the diagonal matrix with main diagonal μ. Then, we consider

$$\Phi : \mathcal{O}_\mu \to \mathbb{R}_{\geq 0} \quad \text{given by} \quad \Phi(G) = N(S - G)\,.$$

Using Lidskii's additive inequality for eigenvalues of selfadjoint matrices, we can construct $G^{\text{op}} \in \mathcal{O}_\mu$ such that $\Phi(G^{\text{op}}) \leq \Phi(G)$, for every $G \in \mathcal{O}_\mu$ (see Section 2 for details). That is, Lidskii's inequality allows us to construct (explicitly) global minimizers of Φ. It is natural to wonder about the structure of all possible minimizers of Φ in $\mathcal{O}_\mu$. Moreover, since $\mathcal{O}_\mu$ has a natural metric (induced by the spectral norm) then we can ask about the structure of local minimizers of Φ in $\mathcal{O}_\mu$. These local minimizers arise naturally when considering optimization of $\Phi_2(G) = \|S - G\|_2$, i.e. when N is the Frobenius norm. In this case, Φ_2^2 is a smooth function defined on a smooth manifold and thus we can apply (adapted) gradient descent algorithms to find minimizers of Φ_2; notice that local minimizers of Φ_2 are stability points of these algorithms and therefore their structure becomes part of the convergence analysis of these methods. Thus, our first main problem is to study the structure of global and local minimizers of Φ, for a general strictly convex u.i.n. N. We carry out a similar analysis for Lidskii's inequality for singular values. In both cases we show that local minimizers are indeed global minimizers and we compute their geometrical properties.

Finite frame theory is a well established and rapidly growing area of research (see [5]). It is well known by now that several fundamental results of finite frame theory are counter-parts of well known results in matrix analysis. For example, the so-called frame design problem with prescribed frame operator and norms - that has played a central role in finite frame theory - is equivalent to some formulations of the Schur-Horn theorem (see [19], [21], or the survey [4] and the reference therein). So it is no surprise that our results have implications in this area of research. Indeed, from the local version of Lidskii's theorem we derive some partial results related to the structure of local minimizers of the generalized frame operator distance (G-FOD) (see [2, 18, 24]).

The paper is organized as follows. In Section 2 we recall several results from matrix analysis that we use throughout the paper. In section 3 we state and prove our main results related to the local versions of Lidskii's theorem. Indeed, in Section 3.1 we obtain complete results showing that local minimizers of functions that are built on strictly convex u.i.n's (as above) are global minimizers. In Section 3.2 we consider the corresponding problem for Lidskii's singular value inequalities. In order to obtain these results, we consider some (differential) geometrical properties of some auxiliary smooth maps. In Section 4 we apply the results from the previous sections to the study of local minimizers of the G-FOD induced by strictly convex u.i.n's. We obtain some partial results regarding the general structure of these local minimizers and show that under some further hypothesis, they are global minimizers of the G-FOD.

2 Preliminaries

In this section we introduce the notations, terminology and results from matrix analysis that we will use throughout the paper (see the texts [3, 12, 13]).

Notation and terminology. We let $\mathcal{M}_{k,d}(\mathbb{C})$ be the space of complex $k \times d$ matrices and write $\mathcal{M}_{d,d}(\mathbb{C}) = \mathcal{M}_d(\mathbb{C})$ for the algebra of complex $d \times d$ matrices. We denote by $\mathcal{H}(d) \subset \mathcal{M}_d(\mathbb{C})$ the real

subspace of selfadjoint matrices and by $\mathcal{M}_d(\mathbb{C})^+ \subset \mathcal{H}(d)$ the cone of positive semidefinite matrices. We let $\mathcal{U}(d) \subset \mathcal{M}_d(\mathbb{C})$ denote the group of unitary matrices. For $d \in \mathbb{N}$, let $\mathbb{I}_d = \{1, \dots, d\}$. Given a vector $x \in \mathbb{C}^d$ we denote by D_x the diagonal matrix in $\mathcal{M}_d(\mathbb{C})$ whose main diagonal is x. Given $x = (x_i)_{i \in \mathbb{I}_d} \in \mathbb{R}^d$ we denote by $x^\downarrow = (x_i^\downarrow)_{i \in \mathbb{I}_d}$ the vector obtained by rearranging the entries of x in non-increasing order. We denote by $(\mathbb{R}^d)^\downarrow = \{x^\downarrow : \ x \in \mathbb{R}^d\}$ and $(\mathbb{R}^d_{\geq 0})^\downarrow = \{x^\downarrow : \ x \in \mathbb{R}^d_{\geq 0}\}$. Given a matrix $A \in \mathcal{H}(d)$ we denote by $\lambda(A) = \lambda(A)^\downarrow = (\lambda_i(A))_{i \in \mathbb{I}_d} \in (\mathbb{R}^d)^\downarrow$ the eigenvalues of A counting multiplicities and arranged in non-increasing order. For $B \in \mathcal{M}_d(\mathbb{C})$ we let $s(B) = \lambda(|B|)$ denote the singular values of B, i.e. the eigenvalues of $|B| = (B^*B)^{1/2} \in \mathcal{M}_d(\mathbb{C})^+$; we also let $\sigma(B) \subset \mathbb{C}$ denote the spectrum of B. If $x, \, y \in \mathbb{C}^d$ we denote by $x \otimes y \in \mathcal{M}_d(\mathbb{C})$ the rank-one matrix given by $(x \otimes y)\, z = \langle z \, , \, y \rangle \ x$, for $z \in \mathbb{C}^d$.

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

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

$$\sum_{i=1}^{j} x_i^\downarrow \leq \sum_{i=1}^{j} y_i^\downarrow \quad \text{for every} \quad 1 \leq j \leq \min\{k \, , \, d\} \, .$$

If $x \prec_w y$ and $\operatorname{tr} x = \sum_{i=1}^k x_i = \sum_{i=1}^d y_i = \operatorname{tr} y$, then x is *majorized* by y, and write $x \prec y$.

Remark 2.2. Given $x, \, y \in \mathbb{R}^d$ we write $x \leqslant y$ if $x_i \leq y_i$ for every $i \in \mathbb{I}_d$. It is a standard exercise to show that:

1. $x \leqslant y \implies x^\downarrow \leqslant y^\downarrow \implies x \prec_w y$.
2. $x \prec y \implies |x| \prec_w |y|$, where $|x| = (|x_i|)_{i \in \mathbb{I}_d} \in \mathbb{R}^d_{\geq 0}$.
3. $x \prec y, \ |x|^\downarrow = |y|^\downarrow \implies x^\downarrow = y^\downarrow$.
4. If $\operatorname{tr}(x) = \sum_{i \in \mathbb{I}_d} x_i = t$ then $\frac{t}{d} \, \mathbb{1}_d \prec x$. △

Although majorization is not a total order in $\mathbb{R}^d$, there are several fundamental inequalities in matrix theory that can be described in terms of this relation. As an example of this phenomenon we can consider Lidskii's (additive) inequality for eigenvalues of sums of hermitians (see [3, 12, 13]). In the following result we also include the characterization of the case of equality obtained in [22].

Theorem 2.3 (Lidskii's inequality)**.** Let $A, \, B \in \mathcal{H}(d)$ with eigenvalues $\lambda(A), \, \lambda(B) \in (\mathbb{R}^d)^\downarrow$ respectively. Then

1. $\lambda(A) - \lambda(B) \prec \lambda(A - B)$.
2. $(\lambda(A) - \lambda(B))^\downarrow = \lambda(A - B)$ if and only if there exists $\{v_i\}_{i \in \mathbb{I}_d}$ an orthonormal basis (ONB) of $\mathbb{C}^d$ such that

$$A = \sum_{i \in \mathbb{I}_d} \lambda_i(A) \ v_i \otimes v_i \quad \text{and} \quad B = \sum_{i \in \mathbb{I}_d} \lambda_i(B) \ v_i \otimes v_i \ . \tag{1}$$

   Notice that in this case, A and B commute. □

Recall that a norm $N(\cdot)$ in $\mathcal{M}_d(\mathbb{C})$ is unitarily invariant if

$$N(UAV) = N(A) \quad \text{for every} \quad A \in \mathcal{M}_d(\mathbb{C}) \quad \text{and} \quad U, \, V \in \mathcal{U}(d) \, .$$

Examples of unitarily invariant norms (u.i.n.) are the spectral norm $\| \cdot \|$ and the p-norms $\| \cdot \|_p$, for $p \geq 1$. It is well known that majorization relations between singular values of matrices are intimately related with inequalities with respect to u.i.n's. The following result summarizes these relations (see for example [3]):

Theorem 2.4. Let A, $B \in \mathcal{M}_d(\mathbb{C})$ be such that $s(A) \prec_w s(B)$. Then:

1. For every u.i.n. N in $\mathcal{M}_d(\mathbb{C})$ we have that $N(A) \leq N(B)$.
2. If we assume that there exists a strictly convex u.i.n. N in $\mathcal{M}_d(\mathbb{C})$ such that $N(A) = N(B)$ then we have that $s(A) = s(B)$.

□

3 Local Lidskii's theorems for unitarily invariant norms

Lidskii's additive inequalities (both for eigenvalues and singular values) can be interpreted as an explicit description of global minimizers of functions that are built on unitarily invariant norms and whose domains consist of certain orbits of matrices (under the action of the unitary group). In this section, we show that Lidskii's inequalities actually describe all global minimizers of such functions, and that local minimizers are also global minimizers. This last fact will play a central role in the next section, in which we state and study Strawn's generalized conjecture.

3.1 Selfadjoint matrices - eigenvalues

We begin with the following comments related to the classical Lidskii's inequality. Fix $S \in \mathcal{H}(d)$ and $\mu \in (\mathbb{R}^d)^{\downarrow}$, and consider $\mathcal{O}_\mu$ given by

$$\mathcal{O}_\mu = \{G \in \mathcal{H}(d) : \lambda(G) = \mu\} = \{U^* D_\mu U : \ U \in \mathcal{U}(d)\} \tag{2}$$

We consider the usual metric in $\mathcal{O}_\mu$ induced by the operator norm; hence $\mathcal{O}_\mu$ is a metric space.

For N a strictly convex u.i.n., let

$$\Phi = \Phi_{(N,S,\mu)} : \mathcal{O}_\mu \to \mathbb{R}_{\geq 0} \quad \text{ be given by } \quad \Phi(G) = N(S-G). \tag{3}$$

Using an orthonormal basis (ONB) of eigenvectors of S we can construct $G^{\text{op}} \in \mathcal{O}_\mu$ such that $\lambda(S - G^{\text{op}}) = (\lambda(S) - \mu)^{\downarrow}$. By Lidskii's inequality and Remark 2.2, we see that for every $G \in \mathcal{O}_\mu$ we have that

$$\lambda(S-G^{\text{op}}) \prec \lambda(S-G) \implies s(S-G^{\text{op}}) = |\lambda(S-G^{\text{op}})| \prec_w |\lambda(S-G)| = s(S-G)\,. \tag{4}$$

Hence, Theorem 2.4 implies that $\Phi(G^{\text{op}}) = N(S-G^{\text{op}}) \leq N(S-G) = \Phi(G)$, for $G \in \mathcal{O}_\mu$. Therefore, G^{op} is a global minimizer of Φ in $\mathcal{O}_\mu$. Conversely, let $G \in \mathcal{O}_\mu$ be a global minimizer of Φ in $\mathcal{O}_\mu$. The previous comments together with item 3. in Remark 2.2 show that

$$\lambda(S-G^{\text{op}}) \prec \lambda(S-G) \quad \text{ and } \quad N(S-G^{\text{op}}) = N(S-G) \implies \lambda(S-G^{\text{op}}) = \lambda(S-G) \tag{5}$$

where we have used the fact that N is strictly convex, the submajorization relation in Eq. (4) and Theorem 2.4. In turn, Eq. (5) together with Theorem 2.3 imply that there exists an ONB $\{v_i\}_{i\in\mathbb{I}_d}$ of $\mathbb{C}^d$ such

$$S = \sum_{i\in\mathbb{I}_d} \lambda_i \ v_i \otimes v_i \quad \text{ and } \quad G = \sum_{i\in\mathbb{I}_d} \mu_i \ v_i \otimes v_i\,,$$

where $(\lambda_i)_{i\in\mathbb{I}_d} = \lambda(S) \in (\mathbb{R}^d)^{\downarrow}$; that is, the global minimizer G is obtained from S as G^{op}.

It is then natural to ask about the structure of local minimizers G_0 of the map Φ in $\mathcal{O}_\mu$, which is our main problem in this section. As we will see, these local minimizers are actually global minimizers of Φ (see Theorem 3.5 below).

Definition 3.1. *Let $S, G_0 \in \mathcal{H}(d)$. We consider*

1. *The product manifold* $\mathcal{U}(d)\times\mathcal{U}(d)$ *endowed with the metric*

$$d((U_1,V_1),(U_2,V_2)) = \max\{\|I-U_1^*U_2\|,\, \|I-V_1^*V_2\|\}\,.$$

2. $\Gamma = \Gamma_{(S,G_0)} : \mathcal{U}(d)\times\mathcal{U}(d) \to \mathcal{H}(d)_\tau \stackrel{\text{def}}{=} \{M\in\mathcal{H}(d):\ \mathrm{tr}(M)=\tau\}$ *for* $\tau = \mathrm{tr}(S)-\mathrm{tr}(G_0)$*, given by*

$$\Gamma(U,V) = U^*S\,U - V^*G_0\,V \quad \textit{for} \quad U,\,V\in\mathcal{U}(d)\,.$$

3. *For a given u.i.n.* N *on* $\mathcal{M}_d(\mathbb{C})$*, we consider* $\Delta^N_{(S,G_0)} = \Delta : \mathcal{U}(d)\times\mathcal{U}(d)\to\mathbb{R}_{\geq 0}$*:*

$$\Delta(U,V) = N(\Gamma(U,V)) \quad \textit{for} \quad U,\,V\in\mathcal{U}(d)\,.$$

△

Our motivation for considering the previous notions comes from the following:

Lemma 3.2. *Let* $S\in\mathcal{H}(d)$*,* $\mu\in(\mathbb{R}^d)^\downarrow$*,* $G_0\in\mathcal{O}_\mu$ *and consider the notations from Definition 3.1. Given a u.i.n.* N *on* $\mathcal{M}_d(\mathbb{C})$*, the following conditions are equivalent:*

1. G_0 *is a local minimizer of* Φ *in* $\mathcal{O}_\mu$ *(defined in Eq.* (3)*);*
2. (I,I) *is a local minimizer of* Δ *on* $\mathcal{U}(d)\times\mathcal{U}(d)$*.*

Proof. 1. $\Longrightarrow$ 2. Consider $(U,W)\in\mathcal{U}(d)\times\mathcal{U}(d)$ such that

$$d((U,W),(I,I)) = \max\{\,\|I-U^*\|\,,\,\|I-W^*\|\,\} := \varepsilon\,.$$

Hence, if $Z = WU^*\in\mathcal{U}(d)$ then $U^*S\,U - W^*G_0\,W = U^*(S-Z^*G_0\,Z)\,U$. Notice that

$$\|Z-I\| = \|W\,(U^*-W^*)\| \leq \|U^*-I\| + \|I-W^*\| \leq 2\,\varepsilon \implies$$

$$\Delta(U,W) = N(U^*(S-Z^*G_0\,Z)\,U) = \Phi(Z^*G_0Z) \quad \text{with} \quad \|Z^*G_0Z-G_0\| \leq 4\,\varepsilon\|G_0\|\,.$$

2. $\Longrightarrow$ 1. This is a consequence of the fact that the map $\mathcal{U}(d)\ni Z\mapsto Z^*G_0\,Z\in\mathcal{O}_\mu$ is open (see, for example, [1, Thm. 4.1] or [7]). □

In what follows, given $\mathcal{S}\subset\mathcal{H}(d)$ we consider the commutant of $\mathcal{S}$, denoted $\mathcal{S}'$, that is the unital *-subalgebra of $\mathcal{M}_d(\mathbb{C})$ given by

$$\mathcal{S}' = \{\ C\in\mathcal{M}_d(\mathbb{C}):\ [C,D]=0\ \text{ for every }\ D\in\mathcal{S}\ \}\subset\mathcal{M}_d(\mathbb{C})\,,$$

where $[C,D] = CD-DC$ denotes the commutator of C and D.

Recall that $\mathcal{U}(d)$ has a natural smooth (differential) manifold structure. Hence, we can consider $\mathcal{U}(d)\times\mathcal{U}(d)$ as a smooth manifold, endowed with the product structure.

Lemma 3.3. *Consider the notations from Definition 3.1. Then*

$$\Gamma \quad \textit{is a submersion at} \quad (I,I) \quad\Longleftrightarrow\quad \{S,\,G_0\}' = \mathbb{C}\cdot I\ .$$

Proof. The (exponential) map $\mathcal{H}(d)\ni X\mapsto\exp(X)$ allows us to identify the tangent space $\mathcal{T}_I\mathcal{U}(d)$ with $i\cdot\mathcal{H}(d)$. Since we consider the product structure on $\mathcal{U}(d)\times\mathcal{U}(d)$ we conclude that the differential of Γ satisfies

$$D_{(I,I)}\Gamma(X,0) = [S,X] \quad \text{and} \quad D_{(I,I)}\Gamma(0,X) = [X,G_0] \quad \text{for} \quad X\in i\cdot\mathcal{H}(d)\,.$$

Therefore Γ is not a submersion at (I,I) if and only if there exists $0 \neq Y \in \mathcal{TH}(d)_\tau = \mathcal{H}(d)_0$ (i.e. $Y \in \mathcal{H}(d)$ such that $\operatorname{tr} Y = 0$) such that

$$\operatorname{tr}(Y\,[S,Z]) = \operatorname{tr}(Y\,[Z,G_0]) = 0 \quad \text{ for every } \quad Z \in i\cdot\mathcal{H}(d)\ . \tag{6}$$

Since $\operatorname{tr}(Y\,[S,Z]) = \operatorname{tr}([Y,S]\,Z)$ and similarly $\operatorname{tr}(Y\,[Z,G_0]) = \operatorname{tr}(Z\,[G_0,Y])$, we see that in this case

$$[Y,S] = 0 = [G_0,Y] \in i\cdot\mathcal{H}(d)\ .$$

Moreover, since $Y \neq 0$ and $\operatorname{tr} Y = 0$, then Y has some non-trivial spectral projection P which also satisfies that $[P,S] = [P,G_0] = 0$. Conversely, in case there exists a non-trivial projection P such that $[P,S] = [P,G_0] = 0$, we can construct $Y = \frac{P}{\operatorname{tr} P} - \frac{I-P}{\operatorname{tr}(I-P)}$ so that $\operatorname{tr} Y = 0$. Then $0 \neq Y \in \mathcal{TH}(d)_\tau$ and it satisfies Eq. (6), so that this matrix Y is orthogonal to the range of the operator $D_{(I,I)}\Gamma$. □

Proposition 3.4. *Consider the notations from Definition 3.1 and assume that N is a strictly convex u.i.n. If (I,I) is a local minimizer of Δ in $\mathcal{U}(d)\times\mathcal{U}(d)$ then $[S,G_0]=0$.*

Proof. Assume that $[S,G_0] \neq 0$. Then there exists a minimal projection P of the unital *-subalgebra $\mathcal{C} = \{S,\,G_0\}' \subseteq \mathcal{M}_d(\mathbb{C})$ such that $[P\,S, P\,G_0] \neq 0$. Indeed, $I \in \mathcal{C}$ is a projection such that $[I\,S, I\,G_0] \neq 0$. If I is not a minimal projection in $\mathcal{C}$ then there exists $P_1,\ P_2 \in \mathcal{C}$ non-zero projections such that $I = P_1 + P_2$; hence $[P_iS, P_iG_0] \neq 0$ for $i=1$ or $i=2$. If the corresponding P_i is not minimal in $\mathcal{C}$ we can repeat the previous argument (halving) applied to P_i. Since we deal with finite dimensional algebras, the previous procedure finds a minimal projection $P \in \mathcal{C}$ as above. By applying a convenient change of orthonormal basis we can assume that $R(P) = \operatorname{span}\{e_i : i \in \mathbb{I}_r\}$, where $r = \operatorname{rk}(P) > 1$. Since P reduces both S and G_0 we can consider $S_1 = S|_{R(P)} \in \mathcal{H}(r)$ and $G_1 = G_0|_{R(P)} \in \mathcal{H}(r)$. By minimality of P we conclude that $\{S_1,\,G_1\}' = \mathbb{C}I_r \subset \mathcal{M}_r(\mathbb{C})$. Using the case of equality of Lidskii's inequality (see Theorem 2.3), we conclude that

$$b := (\lambda(S_1) - \lambda(G_1))^\downarrow \prec a := \lambda(S_1 - G_1) \quad \text{ and } \quad a \neq b\,.$$

If we let $\sigma = \operatorname{tr}(S_1 - G_1)$ then, by Lemma 3.3 the map

$$\mathcal{U}(r)\times\mathcal{U}(r) \ni (U,V) \mapsto U^*S_1\,U - V^*G_1\,V \in \mathcal{H}(r)_\sigma$$

is a submersion at (I_r,I_r). In particular, for every open neighborhood $\mathcal{N}$ of (I_r,I_r) in $\mathcal{U}(r)\times\mathcal{U}(r)$ the set

$$\mathcal{M} := \{U^*S_1\,U - V^*G_1\,V :\ (U,V) \in \mathcal{N}\}$$

contains an open neighborhood of $S_1 - G_1$ in $\mathcal{H}(r)_\sigma$. Consider $\rho : [0,1] \to (\mathbb{R}^r_{\geq 0})^\downarrow$ given by $\rho(t) = (1-t)\,a + t\,b$ for $t \in [0,1]$. Notice that $\rho(t) \prec a$ and $\rho(t) \neq a$ for $t \in (0,1]$. If we let $S_1 - G_1 = W^*D_a\,W$ for $W \in \mathcal{U}(r)$ then the continuous curve $T(\cdot) : [0,1] \to \mathcal{H}(r)_\sigma$ given by $T(t) = W^*D_{\rho(t)}\,W$ for $t \in [0,1]$ satisfies that $T(0) = S_1 - G_1$, $\lambda(T(t)) \prec a$ and $\lambda(T(t)) \neq a$ for $t \in (0,1]$. Therefore, there exists $t_0 \in (0,1]$ such that $T(t) \in \mathcal{M}$ for $t \in [0,t_0]$ so, in particular, there exists $(U,V) \in \mathcal{N}$ such that

$$T(t_0) = U^*S_1\,U - V^*G_1\,V \implies \Delta(U\oplus P^\perp, V\oplus P^\perp) < \Delta(I_d,I_d)\ ,$$

because N is a strictly convex u.i.n., where $U\oplus P^\perp,\ V\oplus P^\perp \in \mathcal{U}(d)$ act as the identity on $R(P)^\perp \subset \mathbb{C}^d$. Since $\mathcal{N}$ was an arbitrary neighborhood of (I_r,I_r) we conclude that (I_d,I_d) is not a local minimizer of Δ in $\mathcal{U}(d)\times\mathcal{U}(d)$, which contradicts Lemma 3.2. □

Theorem 3.5 (Local Lidskii's theorem)**.** *Let $S \in \mathcal{H}(d)$ and $\mu = (\mu_i)_{i\in\mathbb{I}_d} \in (\mathbb{R}^d)^\downarrow$. Assume that N is a strictly convex u.i.n. and that $G_0 \in \mathcal{O}_\mu$ is a local minimizer of $\Phi = \Phi_{(N,\,S,\,\mu)}$ on $\mathcal{O}_\mu$. Then, there exists an ONB $\{v_i\}_{i\in\mathbb{I}_d}$ of $\mathbb{C}^d$ such that*

$$S = \sum_{i\in\mathbb{I}_d} \lambda_i\ v_i\otimes v_i \quad \text{ and } \quad G_0 = \sum_{i\in\mathbb{I}_d} \mu_i\ v_i\otimes v_i\,, \tag{7}$$

where $(\lambda_i)_{i\in\mathbb{I}_d} = \lambda(S) \in (\mathbb{R}^d)^{\downarrow}$. *In particular,* $\lambda(S - G_0) = (\lambda(S) - \lambda(G_0))^{\downarrow}$ *so* G_0 *is also a global minimizer of* Φ *on* $\mathcal{O}_\mu$.

Proof. By Lemma 3.2 and Proposition 3.4 we conclude that $[S, G_0] = 0$. Notice that in this case there exists $\mathcal{B} = \{v_i\}_{i\in\mathbb{I}_d}$ an ONB of $\mathbb{C}^d$ such that

$$S = \sum_{i\in\mathbb{I}_d} \lambda_i \; v_i \otimes v_i \;, \; G_0 = \sum_{i\in\mathbb{I}_d} \nu_i \; v_i \otimes v_i \;\; \text{with} \;\; \lambda = (\lambda_i)_{i\in\mathbb{I}_d} \in (\mathbb{R}^d_{\geq 0})^{\downarrow} \;,$$

for some $\nu_1, \dots, \nu_d \in \mathbb{R}$. We now show that under a suitable permutation of the elements of $\mathbb{I}_d$ we can obtain a representation as in Eq. (7) above. Indeed, assume that $j \in \mathbb{I}_{d-1}$ is such that $\nu_j < \nu_{j+1}$. If we assume that $\lambda_j > \lambda_{j+1}$ then consider the continuous curve of unitary operators $U(t) : [0, \pi/2) \to \mathcal{U}(d)$ given by

$$U(t) = \sum_{i\in\mathbb{I}_d\setminus\{j,\,j+1\}} v_i\otimes v_i + \cos(t)\,(v_j\otimes v_j + v_{j+1}\otimes v_{j+1}) + \sin(t)\,(v_j\otimes v_{j+1} - v_{j+1}\otimes v_j) \;, \;\; t\in[0,\pi/2)\,.$$

Notice that $U(0) = I_d$. We now define the continuous curve $G(t) = U(t)\, G_0\, U(t)^* \in \mathcal{O}_\mu$, for $t \in [0, \pi/2)$. Then $G(0) = G_0$ and we have that

$$S - G(t) = \sum_{i\in\mathbb{I}_d\setminus\{j,\,j+1\}} (\lambda_i - \nu_i)\; v_i \otimes v_i + \sum_{r,s=1}^{2} \gamma_{r,s}(t)\; v_{j+r} \otimes v_{j+s}\,, \tag{8}$$

where $M(t) = (\gamma_{r,s}(t))_{r,s=1}^2$ is determined by

$$M(t) = \begin{pmatrix} \lambda_j & 0 \\ 0 & \lambda_{j+1} \end{pmatrix} - V(t) \begin{pmatrix} \nu_j & 0 \\ 0 & \nu_{j+1} \end{pmatrix} V(t)^* \quad \text{and} \quad V(t) = \begin{pmatrix} \cos(t) & \sin(t) \\ -\sin(t) & \cos(t) \end{pmatrix} \;, \;\; t\in[0,\pi/2)\,.$$

Let us consider

$$R(t) = V^*(t) \begin{pmatrix} \lambda_j - \lambda_{j+1} & 0 \\ 0 & 0 \end{pmatrix} V(t) - \begin{pmatrix} \nu_j & 0 \\ 0 & \nu_{j+1} \end{pmatrix} \implies M(t) = V(t)\, R(t)\, V^*(t) + \lambda_{j+1}\, I_2\,. \tag{9}$$

We claim that $\lambda(R(t)) \prec \lambda(R(0))$ and $\lambda(R(t)) \neq \lambda(R(0))$ for $t \in (0, \pi/2)$ (i.e., the majorization relation is strict). Indeed, since $R(t)$ is a curve in $\mathcal{H}(2)$ such that $\mathrm{tr}(R(t))$ is constant, it is enough to show that the function $[0, \pi/2) \ni t \mapsto \mathrm{tr}(R(t)^2)$ is strictly decreasing in $[0, \pi/2)$. Using that $\lambda_j - \lambda_{j+1} > 0$ we have that

$$V^*(t) \begin{pmatrix} \lambda_j - \lambda_{j+1} & 0 \\ 0 & 0 \end{pmatrix} V(t) = g(t)\otimes g(t) \quad \text{where} \quad g(t) = (\lambda_j - \lambda_{j+1})^{1/2}(\cos(t), \sin(t)) \;, \;\; t\in[0,\pi/2)\,.$$

If $D \in \mathcal{M}_2(\mathbb{C})$ is the diagonal matrix with main diagonal $(\nu_j,\, \nu_{j+1})$ then $R(t) = g(t) \otimes g(t) - D$ so

$$\mathrm{tr}(R(t)^2) = \mathrm{tr}((g(t)\otimes g(t))^2) + \mathrm{tr}(D^2) - 2\;\mathrm{tr}(g(t)\otimes g(t)\; D) = c - 2\,\langle D\, g(t),\, g(t)\rangle$$

where $c = \|g(t)\|^4 + \nu_j^2 + \nu_{j+1}^2 = (\lambda_j - \lambda_{j+1})^2 + \nu_j^2 + \nu_{j+1}^2 \in \mathbb{R}$ is a constant and

$$\langle D\, g(t),\, g(t)\rangle = (\lambda_j - \lambda_{j+1})\; (\cos^2(t)\, \nu_j + \sin^2(t)\, \nu_{j+1})$$

is strictly increasing in $[0, \pi/2)$, since $\nu_j < \nu_{j+1}$. Thus, $\lambda(R(t)) \prec \lambda(R(0))$ and $\lambda(R(t)) \neq \lambda(R(0))$ for $t \in (0, \pi/2)$. Hence, by Eq. (9), we see that

$$\lambda(M(t)) = \lambda(R(t)) + \lambda_{j+1}\, \mathbb{1}_2 \implies \lambda(M(t)) \prec \lambda(M(0)) \;,\; \lambda(M(t)) \neq \lambda(M(0)) \;, \;\; t\in(0,\pi/2)\,.$$

Then, using Eq. (8) and Theorem 2.4, for $t \in (0, \pi/2)$

$$\lambda(S - G(t)) \prec \lambda(S - G_0) \;\; , \;\; \lambda(S - G(t)) \neq \lambda(S - G_0) \Rightarrow N(S - G(t)) < N(S - G_0).$$

This last inequality, which is a consequence of the assumption $\lambda_j < \lambda_{j+1}$, contradicts the local minimality of G_0 in $\mathcal{O}_\mu$. Hence, since $\lambda_j \leq \lambda_{j+1}$ we see that $\lambda_j = \lambda_{j+1}$; in this case, we can consider the basis $\mathcal{B}' = \{v_i'\}_{i\in\mathbb{I}_d}$ obtained by transposing the vectors v_j and v_{j+1} in the basis $\mathcal{B}$. In this case $S\; v_i' = \lambda_i\; v_i'$ for $i \in \mathbb{I}_d$, $G_0\; v_i = \nu_i\; v_i'$ for $i \in \mathbb{I}_d \setminus \{j,\, j+1\}$ and $G_0\; v_j' = \nu_{j+1}\, v_j'$, $G_0\; v_{j+1}' = \nu_j\, v_{j+1}'$. After performing this argument at most d times we get the desired ONB. □

3.2 Arbitrary matrices - singular values

In this section we obtain results related to a local Lidskii's theorem for arbitrary matrices with respect to singular values. As a consequence, we characterize the case of equality in the classical Lidskii's inequality for singular values.

Recall that if $A, B \in \mathcal{M}_d(\mathbb{C})$ then Lidskii's singular value inequality states that

$$|s(A) - s(B)| \prec_w s(A - B). \tag{10}$$

In what follows we fix $A \in \mathcal{M}_d(\mathbb{C})$, $s \in (\mathbb{R}^d_{\geq 0})^{\downarrow}$, $s \neq 0$, and N a strictly convex u.i.n. We consider the set of matrices whose vector of singular values is s, i.e.

$$\mathcal{V}_s := \{C \in \mathcal{M}_d(\mathbb{C}) : \, s(C) = s\}\,,$$

endowed with the usual metric, induced by the spectral norm. We further consider the function

$$\Psi_{(N,\,A,\,s)} = \Psi : \mathcal{V}_s \to \mathbb{R}_{\geq 0} \quad \text{ given by } \quad \Psi(C) = N(A - C).$$

With an argument similar to that in the beginning of Section 3.1, now based on the singular value decomposition (SVD) and Lidskii's inequality in Eq. (10), we can explicitly construct global minimizers of Ψ on $\mathcal{V}_s$. As before, we are interested in the structure of local minimizers of Ψ in $\mathcal{V}_s$.

We will describe the structure of local minimizers of Ψ in $\mathcal{V}_s$ and show that local minimizers are actually global minimizers. In order to do this we consider the following well known matrix construction: for $C \in \mathcal{M}_d(\mathbb{C})$, let $\widehat{C} \in \mathcal{H}(2d)$ be given by

$$\widehat{C} = \begin{pmatrix} 0 & C \\ C^* & 0 \end{pmatrix}.$$

Let $U, V \in \mathcal{U}(d)$ be such that $C = V^* D_{s(C)} U$, and define $W \in \mathcal{U}(2d)$ given by

$$W = \frac{1}{\sqrt{2}} \begin{pmatrix} V & U \\ -V & U \end{pmatrix}.$$

Then $\widehat{C} = W^* \left(D_{s(C)} \oplus -D_{s(C)}\right) W$, which implies that

$$\lambda(\widehat{C})_i = \begin{cases} s_i(C) & \text{if} \quad 1 \leq i \leq d \\ -s_{2d-i+1}(C) & \text{if} \quad d+1 \leq i \leq 2d\,. \end{cases} \tag{11}$$

Definition 3.6. *Let $A\,,\, B \in \mathcal{M}_d(\mathbb{C})$, let N be a u.i.n. on $\mathcal{M}_d(\mathbb{C})$ and $s \in (\mathbb{R}^d_{\geq 0})^{\downarrow}$, $s \neq 0$. We consider:*

1. *The real space $\mathcal{S} = \{\widehat{C} : \; C \in \mathcal{M}_d(\mathbb{C})\} \subset H(2d)$.*
2. *The map $\Pi_{(A\,,\,B)} = \Pi : \mathcal{U}(d)^4 \to \mathcal{S}$ given by*

$$\begin{aligned} \Pi(U_1\,,\,U_2\,,\,V_1\,,\,V_2) &= (U_1 \oplus V_1)^*\, \widehat{A}\, (U_1 \oplus V_1) - (U_2 \oplus V_2)^*\, \widehat{B}\, (U_2 \oplus V_2) \\ &= \widehat{U_1^*\, A\, V_1} - \widehat{U_2^*\, B\, V_2}\,. \end{aligned} \tag{12}$$

3. *The map $\Xi_{(A\,,\,B)} : \mathcal{U}(d)^4 \to \mathbb{R}_{\geq 0}$ given by*

$$\Xi(U_1\,,\,U_2\,,\,V_1\,,\,V_2) = N(U_1^*\, A\, V_1 - U_2^*\, B\, V_2)\,. \tag{13}$$

△

Lemma 3.7. *Let $A \in \mathcal{M}_d(\mathbb{C})$, $s \in (\mathbb{R}^d_{\geq 0})^{\downarrow}$, $s \neq 0$, and $B \in \mathcal{V}_s$. Given a u.i.n. N on $\mathcal{M}_d(\mathbb{C})$, the following conditions are equivalent:*

1. *B is a local minimizer of $\Psi_{(N\,,A\,,s)}$ in $\mathcal{V}_s$;*
2. *(I, I, I, I) is a local minimizer of $\Xi_{(A\,,B)}$ on $\mathcal{U}(d)^4$.*

Proof. An argument similar to that in the proof of Lemma 3.2 shows the equivalence of the items above. □

Next we develop some geometric properties of Π. As before, we consider $\mathcal{U}(d)^4$ as a smooth manifold, endowed with the product structure.

Lemma 3.8. *Let A, $B \in \mathcal{M}_d(\mathbb{C})$ and let Π be as Eq. (12). Then the following conditions are equivalent:*

1. *$\Pi_{(A\,,B)}$ is a submersion at (I, I, I, I);*
2. *Whenever $Z \in \mathcal{M}_d(\mathbb{C})$ is such that A^*Z, AZ^*, B^*Z, $BZ^* \in \mathcal{H}(d)$, then $Z = 0$.*

Proof. Notice that since Π is a smooth function, item 1. holds if and only if the differential map

$$D = D\Pi_{(I\,,I\,,I\,,I)} : (i \cdot \mathcal{H}(d))^4 \to \mathcal{S} \subset \mathcal{H}(2d) \quad \text{is surjective .}$$

We now check that D is *not* surjective if and only if there exists $Z \in \mathcal{M}_d(\mathbb{C})$, $Z \neq 0$, such that A^*Z, AZ^*, B^*Z, $BZ^* \in \mathcal{H}(d)$. Indeed, it is straightforward to compute

$$D(X_1\,,\,X_2\,,\,Y_1\,,\,Y_2) = -\widehat{X_1\,A} + \widehat{A\,Y_1} + \widehat{X_2\,B} - \widehat{B\,Y_2} \quad \text{for} \quad X_1\,,\,X_2\,,\,Y_1\,,\,Y_2 \in i \cdot \mathcal{H}(d)\,.$$

Hence, D is not surjective if and only if there exists $Z \in \mathcal{M}_d(\mathbb{C})$, $Z \neq 0$, such that

$$\widehat{Z} \;\perp\; -\widehat{X_1\,A} + \widehat{A\,Y_1} + \widehat{X_2\,B} - \widehat{B\,Y_2} \quad \text{for} \quad X_1\,,\,X_2\,,\,Y_1\,,\,Y_2 \in i \cdot \mathcal{H}(d)\,. \tag{14}$$

In this case (setting $X_2 = Y_2 = 0$) we have that

$$0 = \mathrm{tr}(\widehat{Z}\;(-\widehat{X_1\,A} + \widehat{A\,Y_1})) = 2\;\mathrm{Re}[\mathrm{tr}(Z^*(-X_1\,A + A\,Y_1))] \quad \text{for} \quad X_1\,,\,Y_1 \in i \cdot \mathcal{H}(d)\,. \tag{15}$$

Using that $\mathrm{Re}[\mathrm{tr}(C)] = \mathrm{tr}(\mathrm{Re}[C])$ and the tracial property, we see that Eq. (15) is equivalent to

$$0 = \mathrm{tr}(X_1\,(AZ^* - ZA^*)) + \mathrm{tr}(Y_1\,(A^*Z - Z^*A)) \quad \text{for} \quad X_1\,,\,Y_1 \in i \cdot \mathcal{H}(d)\,. \tag{16}$$

Since $(AZ^* - ZA^*)\,,\,(A^*Z - Z^*A) \in i \cdot \mathcal{H}(d)$, Eq. (16) holds if and only if

$$AZ^* - ZA^* = 0 \quad \text{and} \quad A^*Z - Z^*A = 0 \implies AZ^*\,,\,A^*Z \in \mathcal{H}(d)\,.$$

Similarly, by setting $X_1 = Y_1 = 0$ in Eq. (14) and arguing as before, we conclude that BZ^*, $B^*Z \in \mathcal{H}(d)$.

Conversely, assume that there exists $Z \in \mathcal{M}_d(\mathbb{C})$, $Z \neq 0$, such that A^*Z, AZ^*, B^*Z, $BZ^* \in \mathcal{H}(d)$. Then, arguing as before, it follows that Z verifies the perpendicularity condition in Eq. (14); thus, D is not surjective in this case. □

Proposition 3.9. *Fix $A \in \mathcal{M}_d(\mathbb{C})$, a strictly convex u.i.n. N on $\mathcal{M}_d(\mathbb{C})$ and $s \in (\mathbb{R}^d_{\geq 0})^{\downarrow}$, $s \neq 0$. If $B \in \mathcal{V}_s$ is a local minimizer of $\Psi = \Psi_{(N\,,A\,,s)}$, then $\Pi_{(A\,,B)}$ is not a submersion at $(I\,,\,I\,,\,I\,,\,I)$.*

Proof. Assume that $\Pi_{(A\,,\,B)}$ is a submersion at $(I\,,\,I\,,\,I\,,\,I)$. Assume further that any of the conditions $A^*B\,,\,A\,B^*\in\mathcal{H}(d)$ does not hold. In this case it is straightforward to check that $\widehat{A}$ and $\widehat{B}$ do not commute. In particular, by Theorem 2.3

$$b:=(\lambda(\widehat{A})-\lambda(\widehat{B}))^{\downarrow}\prec a:=\lambda(\widehat{A}-\widehat{B})\quad\text{ and }\quad a\neq b\,.$$

Now, by Eq. (11) we see that if we let

$$\tilde{b}:=(s(A)-s(B)\,,\,-[s(A)-s(B)])\;,\;\tilde{a}:=(s(A-B)\,,\,-s(A-B))\implies b=(\tilde{b})^{\downarrow}\;,\;a=(\tilde{a})^{\downarrow}\,.$$

Hence, if we let $\rho:[0,1]\to\mathbb{R}^{2d}$ be given by $\rho(t)=(1-t)\,\tilde{a}+t\,\tilde{b}$, for $t\in[0,1]$ then:

1. $\rho(t)\prec a$ and $\rho(t)^{\downarrow}\neq a$, for every $t\in(0,1]$;
2. $\rho(0)=\tilde{a}$;
3. For every $t\in[0,1]$ there exists $c_t\in\mathbb{R}^d$ such that $\rho(t)=(c_t\,,\,-c_t)$.

In order to see item 1. above, recall that

$$\rho(t)=(1-t)\,\tilde{a}+t\,\tilde{b}\prec(1-t)\,(\tilde{a})^{\downarrow}+t\,(\tilde{b})^{\downarrow}=(1-t)\,a+t\,b\prec a$$

and, $((1-t)\,a+t\,b)^{\downarrow}=(1-t)\,a+t\,b\neq a$ (since $a\neq b$), for $t\in(0,1]$. Consider a SVD for $A-B=V^*\,D_{s(A-B)}\,U$, for some $U\,,\,V\in\mathcal{U}(d)$, and define

$$W=\frac{1}{\sqrt{2}}\begin{pmatrix}V & U\\ -V & U\end{pmatrix}\in\mathcal{U}(2d)\,.$$

Then $\widehat{A-B}=W^*\,(D_{s(A-B)}\oplus -D_{s(A-B)})\,W=W^*\;D_{\tilde{a}}\;W$; Let us consider $T(t)=W^*\,D_{\rho(t)}\,W$ for $t\in[0,1]$. Then, using item 3. above, we see that $T(t)\in\mathcal{S}$ for $t\in[0,1]$. By the hypothesis on $\Pi_{(A\,,\,B)}=\Pi$, for every open neighborhood of $I\in\mathcal{N}\subset\mathcal{U}(d)$, the set

$$\mathcal{M}=\{\Pi(U_1\,,\,U_2\,,\,V_1\,,\,V_2):\;U_i\,,\,V_i\in\mathcal{N}\,,\;i=1,2\}$$

contains an open neighborhood of $\widehat{A-B}$ in $\mathcal{S}$. Since $T:[0,1]\to\mathcal{S}$ is a continuous curve such that $T(0)=\widehat{A-B}$, then there exists $t_0\in(0,1)$ such that $T(t)\in\mathcal{M}$, for $t\in[0,t_0]$. In particular, there exist $U_i\,,\,V_i\in\mathcal{N}$, for $i=1,2$ such that

$$T(t_0)=\widehat{U_1^*\,A\,V_1}-\widehat{U_2^*\,B\,V_2}\quad,\quad\lambda(T(t_0))=\rho(t)^{\downarrow}\prec a\quad,\quad\lambda(T(t_0))\neq a\,.$$

Hence, $s(U_1^*\,A\,V_1-U_2^*\,B\,V_2)\prec_w s(A-B)$ and $s(U_1^*\,A\,V_1-U_2^*\,B\,V_2)\neq s(A-B)$. Using that N is a strictly convex u.i.n. we conclude that

$$\Xi_{(A\,,\,B)}(U_1\,,\,U_2\,,\,V_1\,,\,V_2)=N(U_1^*\,A\,V_1-U_2^*\,B\,V_2)<N(A-B)=\Xi_{(A\,,\,B)}(I\,,\,I\,,\,I\,,\,I)\,.$$

Since $\mathcal{N}$ is an arbitrary neighborhood of I in $\mathcal{U}(d)$ we see that $(I\,,\,I\,,\,I\,,\,I)$ is not a local minimizer of $\Xi_{(A\,,\,B)}$, which contradicts Lemma 3.7.

The previous argument shows that $A^*B\,,\,AB^*\in\mathcal{H}(d)$. If we set $Z=B\in\mathcal{M}_d(\mathbb{C})$, we see that

$$Z\neq 0\quad\text{ and }\quad A^*Z\,,\,AZ^*\,,\,B^*Z\,,\,BZ^*\in\mathcal{H}(d)\,.$$

Now, Lemma 3.8 implies that Π is not a submersion at $(I\,,\,I\,,\,I\,,\,I)$, which contradicts our assumption on Π; this last fact proves the result. □

Remark 3.10. Let $A\,,\,B \in \mathcal{M}_d(\mathbb{C})$ be such that $A^*B,\, AB^* \in \mathcal{H}(d)$. Then, Eckart and Young [8] claimed that there exist matrices $U,\, V \in \mathcal{U}(d)$ such that

$$U^*A\,V = A \quad \text{and} \quad U^*B\,V = D_\beta \quad \text{with} \quad \beta \in \mathbb{R}^d\,.$$

Indeed, notice that the hypothesis also holds for X^*AY and X^*BY, for any $X\,,\,Y \in \mathcal{U}(d)$. Thus, by considering a SVD of A and the previous comment, we can assume that $A = \oplus_{i=1}^k \alpha_i\, I_i$ with $I_i \in \mathcal{M}_{d_i}(\mathbb{C})$ the identity matrix, $d_1 + \ldots + d_k = d$ and $\alpha_1 > \ldots > \alpha_k \geq 0$. Let $\mathbb{C}^d = \oplus_{i=1}^k \mathbb{C}^{d_i}$, and consider the block representation of B with respect to this decomposition, $B = (B_{ij})_{i,j=1}^k$. Under the previous assumption on A, we have that $A\,B, AB^* \in \mathcal{H}(d)$; then, $A\,B = (A\,B)^* = B^*\,A$ and $AB^* = (AB^*)^* = BA$. These equations imply that

$$B_{ji}^*\,\alpha_j = \alpha_i\,B_{ij} \quad \text{and} \quad \alpha_i\,B_{ji}^* = B_{ij}\,\alpha_j \quad \text{for} \quad 1 \leq i,j \leq k\,.$$

In particular, if $i \neq j$ and $\alpha_i \neq 0$ we get

$$\alpha_i\,B_{ij} = \alpha_j\,B_{ji}^* = \frac{\alpha_j^2}{\alpha_i} B_{ij} \implies B_{ij} = 0\,.$$

In case that $i \neq j$ and $\alpha_i = 0$ then $\alpha_j\,B_{ij} = 0 \implies B_{ij} = 0$ because $\alpha_j \neq \alpha_i = 0$. And if $\alpha_i \neq 0$ then

$$\alpha_i\,B_{ii}^* = \alpha_i\,B_{ii} \implies B_{ii} = B_{ii}^*\,.$$

Thus $B = \oplus_{i=1}^k B_{ii}$. Let note that if $\alpha_k = 0$, the block $B_{kk} \in \mathcal{M}_{d_k}(\mathbb{C})$ is arbitrary. Consider now the unitary matrices $U_i \in \mathcal{U}(d_i)$ such that $U^*B_{ii}U = D_{\gamma_i}$, with $\gamma_i \in \mathbb{R}^{d_i}$ for $\alpha_i \neq 0$ (that includes $1 \leq i \leq k-1$), and eventually (when $\alpha_k = 0$), a SVD $U_k^*B_{kk}V = D_{\gamma_k}$ for $U_k,\, V_k \in \mathcal{U}(d_k)$, with $\gamma_k \in \mathbb{R}^{d_k}_{\geq 0}$. Then, taking

$$U = \oplus_{i=1}^k U_i \quad \text{and} \quad V = \oplus_{i=1}^{k-1} U_i \oplus V_k,$$

and setting $\beta = (\gamma_1\,,\,\ldots\,,\,\gamma_k) \in \mathbb{R}^d$, we get

$$U^*AV = A \quad \text{and} \quad U^*BV = \oplus_{i=1}^k D_{\gamma_i} = D_\beta\,.$$

$\triangle$

Proposition 3.11. *Fix $A \in \mathcal{M}_d(\mathbb{C})$, a strictly convex u.i.n. N on $\mathcal{M}_d(\mathbb{C})$ and $s \in (\mathbb{R}^d_{\geq 0})^{\downarrow}$, $s \neq 0$. Let $B \in \mathcal{V}_s$ be a local minimizer of $\Psi = \Psi_{(N\,,A\,,s)}$. Then, $A^*B\,,\,A\,B^* \in \mathcal{H}(d)$.*

Proof. We argue by induction on the dimension $d \geq 1$. Indeed, in case $d = 1$ then the result follows from the fact that, given $a \in \mathbb{C}$, any local minimizer b of the function $f(c) = |a - c|$ for $c \in \{z \in \mathbb{C} :\ |z| = s > 0\}$ satisfies that $\bar{a} \cdot b \in \mathbb{R}$, and then also $a \cdot \bar{b} \in \mathbb{R}$.

We assume that the result holds for all dimension $\tilde{d}$ such that $1 \leq \tilde{d} \leq d-1$. Let $A\,,\,B \in \mathcal{M}_d(\mathbb{C})$ be such that B is a local minimizer of Ψ in $\mathcal{V}_s$. Notice that by Proposition 3.9, $\Pi_{(A\,,B)}$ is not a submersion at $(I\,,\,I\,,\,I\,,\,I)$. By Lemma 3.8, we conclude that there exists $Z \in \mathcal{M}_d(\mathbb{C})$, $Z \neq 0$, such that $A^*\,Z\,,\,A\,Z^*\,,\,B^*\,Z\,,\,B\,Z^* \in \mathcal{H}(d)$. Consider a SVD, $D_{s(Z)} = U^*\,Z\,V$, for $U,\, V \in \mathcal{U}(d)$. By replacing A and B by $U^*\,A\,V$ and $U^*\,B\,V$ we can further assume that $Z = D_{s(Z)}$, where $s(Z) = (s_i(Z))_{i \in \mathbb{I}_d} \in (\mathbb{R}^d_{\geq 0})^{\downarrow}$. We let:

1. $\sigma(Z) = \{\sigma_1 > \ldots > \sigma_k\}$ be the distinct eigenvalues of $Z = D_{s(Z)} \in \mathcal{M}_d(\mathbb{C})^+$.
2. $I_j = \{i \in \mathbb{I}_d :\ s_i(Z) = \sigma_j\}$ and $m_j = \#(I_j)$, for $j \in \mathbb{I}_k$.

Notice that since $Z \neq 0$ then $\sigma_1 > 0$. Using that $A^*Z\,,\ AZ^*\,,\ B^*Z\,,\ BZ^* \in \mathcal{H}(d)$ with $Z = \oplus_{j\in\mathbb{I}_k}\sigma_j\, I_j$ and Remark 3.10, we conclude that:

$$A = \oplus_{j\in\mathbb{I}_k}A_j \quad \text{ and } \quad B = \oplus_{j\in\mathbb{I}_k}B_j \quad \Longrightarrow \quad A - B = \oplus_{j\in\mathbb{I}_k}A_j - B_j\,, \tag{17}$$

where $I_j\,,\ A_j\,,\ B_j \in \mathcal{M}_{m_j}(\mathbb{C})$, for $j \in \mathbb{I}_k$; moreover, $A_j\,,\ B_j \in \mathcal{H}(m_j)$, whenever $\sigma_k \neq 0$, for $j \in \mathbb{I}_k$. Using the fact that B is a local minimizer of Ψ on $\mathcal{V}_s$ we see that

$$B_j \text{ is a local minimizer of } \Psi_{(N_j\,,\,A_j\,,\,s(B_j))}\ ,\ \text{ for } j \in \mathbb{I}_k\,,$$

where N_j is the strictly convex u.i.n. on $M_{m_j}(\mathbb{C})$ given by $N_j(C) = N(C \oplus 0_{d-m_j})$. In turn, this last fact shows that for each $j \in \mathbb{I}_k$ for which $\sigma_j \neq 0$ - which includes all $1 \leq j \leq \max\{k-1\,,\,1\}$ - B_j is a local minimizer of $\Phi_{(N_j\,,\,A_j\,,\,\lambda(B_j))}$; Theorem 3.5 shows that A_j and B_j commute, so $A_j^*B_j\,,\ A_j\,B_j^* \in \mathcal{H}(m_j)$, for $j \in \mathbb{I}_k$ such that $\sigma_j \neq 0$. Therefore, we consider two possible cases: on the one hand, if $\sigma_k \neq 0$ then the previous remarks show that

$$A^*\,B = \oplus_{j\in\mathbb{I}_k}A_j^*\,B_j \in \mathcal{H}(d)\,,$$

and similarly, $A\,B^* \in \mathcal{H}(d)$.

On the other hand, if $\sigma_k = 0$, notice that $m_k = \dim\,\ker Z < d$ (since $Z \neq 0$), and $B_k \in M_{m_k}(\mathbb{C})$ is a local minimizer of $\Psi_{(N_k\,,\,A_k\,,\,s(B_k))}$. In this case we can apply the inductive hypothesis and conclude that $A_k^*\,B_k\,,\ A_k\,B_k^* \in \mathcal{H}(m_k)$. Since we have already showed that $A_j^*\,B_j\,,\ A_j\,B_j^* \in \mathcal{H}(m_j)$, for $1 \leq j \leq k-1$, we now see that $A^*\,B\,,\ A\,B^* \in \mathcal{H}(d)$. □

Theorem 3.12. *Let $A \in \mathcal{M}_d(\mathbb{C})$, $s \in (\mathbb{R}^d_{\geq 0})^{\downarrow}$ and N be a strictly convex u.i.n. If B is a local minimizer of Ψ in $\mathcal{V}_s$ then A and B have a joint SVD i.e., there exist $U, V \in \mathcal{U}(d)$ such that*

$$A = U^*D_{s(A)}V \quad \textit{and} \quad B = U^*D_{s(B)}V.$$

In particular, $s(A-B) = |s(A) - s(B)|^{\downarrow}$ and B is a global minimizer of Ψ in $\mathcal{V}_s$.

Proof. Notice that if B is a local minimizer of Ψ in $\mathcal{V}_s$ and $X^*AY = D_{s(A)}$ is a SVD of A for some $X,\ Y \in \mathcal{U}(d)$, we can replace A by $D_{s(A)}$ and B by X^*BY to get

$$N(A-B) = N(D_{s(A)} - X^*BY).$$

Since $\mathcal{V}_s \ni C \mapsto X^*CY \in \mathcal{V}_s$ is a homeomorphism of $\mathcal{V}_s$ then we can assume, without loss of generality, that $A = D_{s(A)}$. By Proposition 3.11 we get that $A\,B,\ A\,B^* \in \mathcal{H}(d)$; then by [8] (see Remark 3.10) there exist matrices $U,\ V \in \mathcal{U}(d)$ such that

$$U^*A\,V = D_{s(A)}\,(= A) \quad \text{ and } \quad U^*B\,V = D_\beta \quad \text{ with } \quad \beta \in \mathbb{R}^d\,.$$

Suppose now that $\beta \notin \mathbb{R}^d_{\geq 0}$, so there exists $1 \leq \ell \leq d$ such that $\beta_\ell < 0$. Notice that the function $f(t) : [0,\pi] \to \mathbb{R}_{\geq 0}$ given by

$$f(t) = |s_\ell(A) - e^{it}\beta_\ell| \quad \text{ for } \quad t \in [0,\pi]$$

is strictly decreasing. Let $W(t) = (w_{jk})_{j,\,k\,\in\mathbb{I}_d} \in \mathcal{U}(d)$ be the diagonal matrix whose main diagonal is given by $w_{jj} = 1$ for all $j \neq \ell$, and $w_{\ell\ell} = e^{it}$ for $t \in [0,\pi]$; hence $W(0) = I$. Define

$$B(t) = U\,W(t)D_\beta V^* \quad \text{ for } \quad t \in [0,\pi]\,.$$

Then $B(t)$ is a continuous curve such that $B(0) = B$, $B(t) \in \mathcal{V}_s$ for $t \in [0,\pi]$, and

$$\Psi(B(t)) = N(U\,(D_{s(A)} - D_{\beta(t)})\,V^*) = N(D_{|\alpha-\beta(t)|})\ ,$$

where $\beta(t) = s(B(t))$ for $t \in [0,\pi]$. Hence, $\beta_j(t) = \beta_j$ for $j \neq \ell$ and $\beta_\ell(t) = e^{it}\,\beta_\ell$. Therefore, $|\alpha_j - \beta_j(t)| = \alpha_j - \beta_j$ is constant for $j \neq \ell$ and $|\alpha_\ell - \beta_\ell(t)| = f(t)$ for $t \in [0,\pi]$. Since f is strictly decreasing, we conclude that

$$|\alpha - \beta(t)| \prec_w |\alpha - \beta| \implies \Psi(B(t)) \text{ is strictly decreasing for } t \in [0,\pi]\,.$$

This last fact contradicts the assumption of B. Therefore $\beta \in \mathbb{R}^d_{\geq 0}$.

Suppose now that $\beta \neq s = s(B)$ i.e. $\beta \neq \beta^{\downarrow}$; since $A = D_{s(A)}$ with $s(A) = s(A)^{\downarrow}$ then, by Theorem 3.5, D_β is not a local minimizer of $\Phi = \Phi_{(N,\, D_{s(A)},\, s)}$ on $\mathcal{O}_s$. Then, there exists a continuous curve $\delta(t) : [0,1] \to \mathcal{U}(d)$ such that $\delta(0) = I$ and $h(t) = N(D_{s(A)} - \delta(t)^*\, D_\beta\, \delta(t)\,)$, is strictly decreasing in $[0,1]$. Therefore, if we let $\tilde{B}(t) = U\,\delta(t)^*\, D_\beta\, \delta(t)\, V^*$ for $t \in [0,1]$ then $\tilde{B}(0) = B$, $\tilde{B}(t) \in \mathcal{V}_s$ for $t \in [0,1]$, and the function $\Psi(\tilde{B}(t)) = h(t)$ is strictly decreasing in $[0,1]$. These facts contradict our assumption that B is a local minimizer of Ψ in $\mathcal{V}_s$. Hence, $U^*BV = D_s$ and then, $s(A-B) = |s(A) - s|^{\downarrow}$ which implies that B is a global minimizer of Ψ in $\mathcal{V}_s$. □

Corollary 3.13 (Equality in Lidskii's inequality for singular values)**.** *Let A, $B \in \mathcal{M}_d(\mathbb{C})$. Then $|s(A) - s(B)|^{\downarrow} = s(A-B)$ if and only if A and B have a joint SVD.*

Proof. In case $A\,,\, B \in \mathcal{M}_d(\mathbb{C})$ have a joint SVD, then it is straightforward to show that $|s(A) - s(B)|^{\downarrow} = s(A-B)$. Conversely, assume that $|s(A) - s(B)|^{\downarrow} = s(A-B)$ and choose your favorite strictly convex u.i.n. N on $\mathcal{M}_d(\mathbb{C})$. By the comments at the beginning of this section, we see that B is a global minimizer of $\Psi_{(N\,,\,A\,,\,s(B))} = \Psi$ on $\mathcal{V}_{s(B)}$. In particular, B is a local minimizer of Ψ in $\mathcal{V}_{s(B)}$; the result now follows from Theorem 3.12. □

4 Application: Generalized Strawn's conjecture

In this section we consider some problems within the theory of finite frames (see the texts [5, 6]. for general references on this topic). It is worth pointing out that our results can be also be described as the solution to certain matrix nearness problems, following the scheme of [10] (see Remark 4.3).

In what follows we adopt the following:

Notation and terminology: let $\mathcal{F} = \{f_i\}_{i\in\mathbb{I}_k}$ be a finite sequence in $\mathbb{C}^d$. Then,

1. $T_{\mathcal{F}} \in \mathcal{M}_{d,k}(\mathbb{C})$ denotes the synthesis operator of $\mathcal{F}$ given by $T_{\mathcal{F}} \cdot (\alpha_i)_{i\in\mathbb{I}_k} = \sum_{i\in\mathbb{I}_k} \alpha_i\, f_i$.
2. $T^*_{\mathcal{F}} \in \mathcal{M}_{k,d}(\mathbb{C})$ denotes the analysis operator of $\mathcal{F}$ and it is given by $T^*_{\mathcal{F}} \cdot f = (\langle f, f_i\rangle)_{i\in\mathbb{I}_k}$.
3. $S_{\mathcal{F}} \in \mathcal{M}_d(\mathbb{C})^+$ denotes the frame operator of $\mathcal{F}$ and it is given by $S_{\mathcal{F}} = T_{\mathcal{F}}\, T^*_{\mathcal{F}}$. Hence, $Sf = \sum_{i\in\mathbb{I}_k} \langle f, f_i\rangle f_i = \sum_{i\in\mathbb{I}_k} f_i \otimes f_i(f)$ for $f \in \mathbb{C}^d$.
4. We say that $\mathcal{F}$ is a frame for $\mathbb{C}^d$ if it spans $\mathbb{C}^d$; equivalently, $\mathcal{F}$ is a frame for $\mathbb{C}^d$ if $S_{\mathcal{F}}$ is a positive invertible operator acting on $\mathbb{C}^d$. △

4.1 Generalized frame operator distances

Let $S \in \mathcal{M}_d(\mathbb{C})^+$ and $\mathbf{a} = (a_i)_{i\in\mathbb{I}_k} \in (\mathbb{R}^k_{>0})^{\downarrow}$. In this case we consider

$$\mathbb{T}_d(\mathbf{a}) := \left\{ \mathcal{G} = \{g_i\}_{i\in\mathbb{I}_k} \in (\mathbb{C}^d)^k \ : \|g_i\|^2 = a_i\,, \text{ for every } i \in \mathbb{I}_k \right\}.$$

By definition, $\mathbb{T}_d(\mathbf{a})$ is the (Cartesian) product of spheres in $\mathbb{C}^d$; hence, we consider the product metric of the Euclidean metrics in each of these spheres, namely

$$d(\mathcal{G}\,,\,\mathcal{G}') = \max\{\|g_i - g'_i\| :\ i \in \mathbb{I}_k\} \quad \text{for} \quad \mathcal{G} = \{g_i\}_{i\in\mathbb{I}_k},\, \mathcal{G}' = \{g'_i\}_{i\in\mathbb{I}_k} \in \mathbb{T}_d(\mathbf{a})\,.$$

Notice that $\mathbb{T}_d(\mathbf{a})$ is a compact metric space with the product metric. Given a strictly convex u.i.n N on $\mathcal{M}_d(\mathbb{C})$, we can consider the generalized frame operator distance (G-FOD) in $\mathbb{T}_d(\mathbf{a})$ (see [18]) given by

$$\Theta_{(N,\,S,\,\mathbf{a})} = \Theta : \mathbb{T}_d(\mathbf{a}) \to \mathbb{R}_{\geq 0} \quad \text{given by} \quad \Theta(\mathcal{G}) = N(S - S_{\mathcal{G}})$$

where $S_{\mathcal{G}} = \sum_{i\in\mathbb{I}_k} g_i \otimes g_i$ denotes the frame operator of a family $\mathcal{G} \in \mathbb{T}_d(\mathbf{a})$. This notion is based on the frame operator distance (FOD) $\Theta_{(\|\cdot\|_2,\,S,\,\mathbf{a})}$ introduced by Strawn in [24], where $\|A\|_2^2 = \mathrm{tr}(A^*A)$ denotes the Frobenius norm, $A \in \mathcal{M}_d(\mathbb{C})$.

Based on his work and on numerical evidence, Strawn conjectured in [24] that local minimizers of $\Theta_{(\|\cdot\|_2,\,S,\,\mathbf{a})} : \mathbb{T}_d(\mathbf{a}) \to \mathbb{R}_{\geq 0}$ are global minimizers. In [18] we settled Strawn's conjecture in the affirmative; indeed, we obtained the following results related to the more general G-FOD induced by a strictly convex u.i.n.:

Theorem 4.1 (See[18]). *Let $S \in \mathcal{M}_d(\mathbb{C})^+$ and $\mathbf{a} = (a_i)_{i\in\mathbb{I}_k} \in (\mathbb{R}^k_{>0})^{\downarrow}$. Then, there exists $\nu^{\mathrm{op}} = \nu^{\mathrm{op}}(S\,,\,\mathbf{a}) \in (\mathbb{R}^d_{\geq 0})^{\downarrow}$ (that can be computed explicitly) such that:*

1. *There exists $\mathcal{G}^{\mathrm{op}} \in \mathbb{T}_d(\mathbf{a})$ such that $\lambda(S_{\mathcal{G}^{\mathrm{op}}}) = \nu$. In this case, if N is a u.i.n. in $\mathcal{M}_d(\mathbb{C})$ then*

$$\Theta_{(N,\,S,\,\mathbf{a})}(\mathcal{G}^{\mathrm{op}}) \leq \Theta_{(N,\,S,\,\mathbf{a})}(\mathcal{G}) \quad \text{for } \mathcal{G} \in \mathbb{T}_d(\mathbf{a})\,. \tag{18}$$

2. *If N is a strictly convex u.i.n. and $\mathcal{G}_0$ is a global minimizer of $\Theta_{(N,\,S,\,\mathbf{a})}$ on $\mathbb{T}_d(\mathbf{a})$ then $\lambda(S_{\mathcal{G}_0}) = \nu^{\mathrm{op}}$.*

3. *If $\mathcal{G}_0$ is a local minimizer of $\Theta_{(\|\cdot\|_2,\,S,\,\mathbf{a})}$ on $\mathbb{T}_d(\mathbf{a})$ then $\lambda(S_{\mathcal{G}_0}) = \nu^{\mathrm{op}}$; hence $\mathcal{G}_0$ is a global minimizer of $\Theta_{(\|\cdot\|_2,\,S,\,\mathbf{a})}$.* □

We point out that Theorem 4.1 is obtained in terms of a translation of G-FOD problems into frame completion problems with prescribed norms. Roughly speaking, given $S \in \mathcal{M}_d(\mathbb{C})^+$ and $\mathbf{a} = (a_i)_{i\in\mathbb{I}_k} \in (\mathbb{R}^k_{>0})^{\downarrow}$ as above, we can consider an auxiliary family $\mathcal{F}_0 = \{f_i\}_{i\in\mathbb{I}_d} \in \mathbb{C}^d$ such that $S_{\mathcal{F}_0} = \|S\|\,I - S \in \mathcal{M}_d(\mathbb{C})^+$, so that for each $\mathcal{G} \in \mathbb{T}_d(\mathbf{a})$ we get a representation of the operator difference

$$S - S_{\mathcal{G}} = \|S\|\,I - (S_{\mathcal{F}_0} + S_{\mathcal{G}})\,. \tag{19}$$

Notice that if we let $\mathcal{F} = (\mathcal{F}_0\,,\,\mathcal{G}) \in (\mathbb{C}^d)^{d+k}$ be the finite sequence obtained by juxtaposition of $\mathcal{F}_0$ and $\mathcal{G}$ then $S_{\mathcal{F}_0} + S_{\mathcal{G}} = S_{\mathcal{F}}$. In [18] any such $\mathcal{F}$ is called a completion of $\mathcal{F}_0$ by a family $\mathcal{G}$, with norms prescribed by the sequence $\mathbf{a}$. Eq. (19) can be used to show items 1. and 2. in Theorem 4.1 for a u.i.n. N. In order to get information about local minimizers of $\Theta_{(N,\,S,\,\mathbf{a})}$ from Eq. (19) we should assume further that N is the Frobenius norm. This obstruction to the general case of item 3. (for a strictly convex u.i.n. N) seems to be a limitation of the reduction methods from [18]. Hence, we state the following:

Conjecture 4.2. Given $S \in \mathcal{M}_d(\mathbb{C})^+$ and $\mathbf{a} = (a_i)_{i\in\mathbb{I}_k} \in (\mathbb{R}^k_{>0})^{\downarrow}$ and a strictly convex u.i.n. N in $\mathcal{M}_d(\mathbb{C})$, let $\Theta_{(N,\,S,\,\mathbf{a})} : \mathbb{T}_d(\mathbf{a}) \to \mathbb{R}_{\geq 0}$ be given by $\Theta_{(N,\,S,\,\mathbf{a})}(\mathcal{G}) = N(S - S_{\mathcal{G}})$. If G_0 is a local minimizer $\Theta_{(N,\,S,\,\mathbf{a})}$ in $\mathbb{T}_d(\mathbf{a})$ then:

1. $\lambda(S - S_{\mathcal{G}_0}) \prec \lambda(S - S_{\mathcal{G}})$, for every $\mathcal{G} \in \mathbb{T}_d(\mathbf{a})$;

2. In particular, $\mathcal{G}_0$ is a global minimizer of $\Theta_{(\tilde{N},\,S,\,\mathbf{a})}$, for every u.i.n. $\tilde{N}$ on $\mathcal{M}_d(\mathbb{C})$. △

We point out that item 2. in Conjecture 4.2 is a consequence of item 1. Nevertheless, item 2. is directly related with the possible applications of the solution of Conjecture 4.2 for the G-FOD problems.

In what follows, we will describe the first features of local minimizers of $\Theta_{(N,S,\mathbf{a})} : \mathbb{T}_d(\mathbf{a}) \to \mathbb{R}_{\geq 0}$, for an arbitrary strictly convex u.i.n. N in $\mathcal{M}_d(\mathbb{C})$. We will also show that Conjecture 4.2 holds under some further hypothesis on the spectral structure of local minimizers.

We end this section with the following remark, in which we show the connection between G-FOD problems and matrix nearness problems.

Remark 4.3. Let $S \in \mathcal{M}_d(\mathbb{C})^+$ and consider a strictly convex u.i.n. N in $\mathcal{M}_d(\mathbb{C})$. Let $\mu_j \in (\mathbb{R}^d)^{\downarrow}$, for $j \in \mathbb{I}_k$, and consider the orbits

$$\mathcal{O}_{\mu_j} = \{G \in \mathcal{H}(d) : \ \lambda(G) = \mu_k\} \ , \ j \in \mathbb{I}_k \, .$$

We can then consider the matrix nearness problem (as described in [10], see also [16])

$$\operatorname{argmin} \{N(S-H)) : \ H \in \mathcal{O}_{\mu_1} + \ldots + \mathcal{O}_{\mu_k}\} \, . \tag{20}$$

Let $\mathbf{a} = (a_i)_{i \in \mathbb{I}_k} \in (\mathbb{R}^k_{>0})^{\downarrow}$ and consider the particular case: $\mu_j = a_j \, e_1$, for $j \in \mathbb{I}_k$, where $\{e_i\}_{i=1}^d$ denotes the canonical basis of $\mathbb{C}^d$. Then $G \in \mathcal{O}_{\mu_j}$ if and only if $G = g \otimes g$ for some $g \in \mathbb{C}^d$ with $\|g\|^2 = a_j$, $j \in \mathbb{I}_k$. Hence, the matrix nearness problem in Eq. (20) coincides with the problem of computing global minimizers on $\Theta_{(N\,,S\,,\mathbf{a})}$ in $\mathbb{T}_d(\mathbf{a})$. Similarly, the study of local minimizers of the matrix nearness problem corresponds to the study of local minimizers of $\Theta_{(N\,,S\,,\mathbf{a})}$. It is worth pointing out that for the Frobenius norm, local minimizers of the matrix nearness problem arise naturally as stability points of (effective) gradient descent algorithms, as those considered in [16]. Hence, settling Conjecture 4.2 in the affirmative would be a relevant result from an applied point of view. △

4.2 Properties of local minimizers of the G-FOD on $\mathbb{T}_d(\mathbf{a})$

In this section we consider the following

Notation 4.4. Fix $S \in \mathcal{M}_d(\mathbb{C})^+$, $\mathbf{a} = (a_i)_{i \in \mathbb{I}_k} \in (\mathbb{R}^k_{>0})^{\downarrow}$ and a strictly convex u.i.n. N on $\mathcal{M}_d(\mathbb{C})$. We consider

1. $\Theta_{(N,S,\mathbf{a})} = \Theta : \mathbb{T}_d(\mathbf{a}) \to \mathbb{R}_{\geq 0}$ given by $\Theta(\mathcal{G}) = N(S - S_{\mathcal{G}})$.
2. A local minimizer $\mathcal{G}_0 = \{g_i\}_{i \in \mathbb{I}_k} \in \mathbb{T}_d(\mathbf{a})$ of $\Theta_{(N,S,\mathbf{a})}$, with frame operator $S_0 = S_{\mathcal{G}_0}$.
3. For $\mu \in (\mathbb{R}^d)^{\downarrow}$, the unitary orbit $\mathcal{O}_\mu$ given by

$$\mathcal{O}_\mu = \{G \in \mathcal{H}(d) : \lambda(G) = \mu\} = \{U^* D_\mu \, U : \ U \in \mathcal{U}(d)\} \, ,$$

   with the usual metric, induced by the operator norm;
4. The function $\Phi = \Phi_{(N,S,\mu)} : \mathcal{O}_\mu \to \mathbb{R}_{\geq 0}$ given by $\Phi(G) = N(S - G)$.

Theorem 4.5. *Consider Notation 4.4. Then,*

1. *$S - S_0$ and $g_j \otimes g_j$ commute, for $j \in \mathbb{I}_k$. Hence, g_j is an eigenvector of $S - S_0$, for $j \in \mathbb{I}_k$.*
2. *There exists $\{v_i\}_{i \in \mathbb{I}_d}$ an ONB of $\mathbb{C}^d$ such that*

$$S = \sum_{i \in \mathbb{I}_d} \lambda_i(S) \ v_i \otimes v_i \quad \text{and} \quad S_0 = \sum_{i \in \mathbb{I}_d} \lambda_i(S_0) \ v_i \otimes v_i \, .$$

   In particular, we have that $\lambda(S - S_0) = \big[\lambda(S) - \lambda(S_0)\big]^{\downarrow}$.

Proof. For $j \in \mathbb{I}_k$ define

$$S_{[j]} = S - \sum_{i \neq j} g_i \otimes g_i \ \in \mathcal{H}(d) \quad \text{and} \quad \mu_{[j]} = a_j e_1 \ \in \mathbb{R}^d_{\geq 0}\,.$$

Then, $\mathcal{O}_{\mu_{[j]}} = \left\{ g \otimes g : \|g\|^2 = a_j \right\}$ and it is straightforward to check that $g_j \otimes g_j$ is a local minimizer of $\Theta_{(N,\, S_{[j]},\, \mu_{[j]})}$ in $\mathcal{O}_{\mu_{[j]}}$. Thus, by Theorem 3.5, $g_j \otimes g_j$ commutes with $S_{[j]}$, for $j \in \mathbb{I}_k$. This last fact implies that $S - S_0$ and $g_j \otimes g_j$ commute, for $j \in \mathbb{I}_k$, which proves item 1.

Since $\mathcal{G}_0$ is a local minimizer of Θ in $\mathbb{T}_d(\mathbf{a})$, there exists $\varepsilon > 0$ such that

$$U \in B_{(I\,,\,\varepsilon)} \stackrel{\text{def}}{=} \{U \in \mathcal{U}(d) : \|I - U\| < \epsilon\} \implies \Phi_{(N,\,S,\,\mu)}(U\, S_0\, U^*) \geq \Phi_{(N,\,S,\,\mu)}(S_0)\,, \tag{21}$$

where we are using Notation 4.4, with $\mu = \lambda(S_0)$. Indeed, let $\varepsilon > 0$ be such that for $\mathcal{G}' \in \mathbb{T}_d(\mathbf{a})$ with $d(\mathcal{G}_0, \mathcal{G}') < \varepsilon$ we have that $\Theta(\mathcal{G}') \geq \Theta(\mathcal{G}_0)$. Notice that if $U \in B_{(I\,,\,\varepsilon)}$ then $U \cdot \mathcal{G}_0 = \{U\, g_i\}_{i \in \mathbb{I}_k} \in \mathbb{T}_d(\mathbf{a})$ is such that $d(\mathcal{G}_0\,,\, U \cdot \mathcal{G}_0) < \epsilon$. Therefore,

$$\Phi_{(N,\,S,\,\mu)}(U\, S_0\, U^*) = \Theta(U \cdot \mathcal{G}_0) \geq \Theta(\mathcal{G}_0) = \Phi_{(N,\,S,\,\mu)}(S_0)\,.$$

Now, the map $\pi : \mathcal{U}(d) \to \mathcal{O}_\mu$ given by $\pi(U) = U\,(S_0)\,U^*$ is open (see [1, Thm 4.1]), so that $\pi(B_{(I\,,\,\varepsilon)})$ is an open neighborhood of S_0 in $\mathcal{O}_\mu$, and S_0 is a local minimum for the map $\Phi_{(N,\,S,\,\mu)}$ on $\mathcal{O}_\mu$. Item 2 now follows from Theorem 3.5 and the fact that $\mu = \lambda(S_0) \in (\mathbb{R}^d)^{\downarrow}$. □

Corollary 4.6. *Consider Notation 4.4. Let $W = R(S_0) \subset \mathbb{C}^d$; then,*

1. *W reduces $S - S_0 \in \mathcal{H}(d)$; hence, $D := (S - S_0)|_W \in L(W)$ is a selfadjoint operator;*
2. *Let $\sigma(D) = \{c_1\,,\, \ldots\,,\, c_p\}$ be such that $c_1 < c_2 < \ldots < c_p$ and let*

$$J_j = \{\ell \in \mathbb{I}_k :\ D\, g_\ell = c_j\, g_\ell\} \quad \text{for} \quad j \in \mathbb{I}_p\,.$$

   Then $\mathbb{I}_k$ is the disjoint union of $\{J_j\}_{j \in \mathbb{I}_p}$;

3. *If we let $W_j = span\{g_\ell :\ \ell \in J_j\}$ then W_j reduces both S and S_0, for $j \in \mathbb{I}_p$. Moreover, $W = \oplus_{j \in \mathbb{I}_p} W_j$.*

Proof. Notice that $W = \text{span}\{g_i :\ i \in \mathbb{I}_k\}$; on the other hand, by Theorem 4.5, g_i is an eigenvector of $S - S_0$, for each $i \in \mathbb{I}_k$. These two facts show that W is an invariant subspace of $S - S_0$; since $S - S_0$ is selfadjoint, W reduces $S - S_0$. Thus, the restriction $D = (S - S_0)|_W \in L(W)$ is a well defined selfadjoint operator acting on W. The previous remarks also show that $\mathbb{I}_k$ is the disjoint union of $\{J_i\}_{i \in \mathbb{I}_p}$.

Let j, $\ell \in \mathbb{I}_p$ with $j \neq \ell$ and let $r \in J_j$ and $s \in J_\ell$. Then, $g_r \perp g_s$, since these vectors are eigenvectors of a selfadjoint operator, corresponding to different eigenvalues. Hence, $W_j \perp W_\ell$ and

$$S_0\, g_r = \sum_{u \in J_j} \langle g_r\,,\, g_u \rangle\ g_u \in W_j\,.$$

Thus, in particular, W_j reduces S_0; using that W_j also reduces $S - S_0$ we conclude that W_j reduces $S = (S - S_0) + S_0$, for $j \in \mathbb{I}_p$. On the other hand, since $W = \sum_{j \in \mathbb{I}_p} W_j$ then $W = \oplus_{j \in \mathbb{I}_p} W_j$. □

Theorem 4.7. *Consider Notation 4.4. Let $W = R(S_0)$ and let $\sigma((S - S_0)|_W) = \{c_1\,,\, \ldots\,,\, c_p\}$ as in Corollary 4.6. Let $j \in \mathbb{I}_p$ and assume that there exists $c \in \sigma(S - S_0)$ such that $c_j < c$. Then, the family $\{g_j\}_{j \in J_j}$ is linearly independent.*

Proof. Suppose that for some $j \in \mathbb{I}_p$ the family $\{g_i\}_{i\in J_j}$ is linearly dependent. Hence there exist coefficients $z_l \in \mathbb{C}$, $l \in J_j$ (not all zero) such that every $|z_l| \leq 1/2$ and

$$\sum_{l\in J_j} \overline{z_l}\; a_l^{1/2}\; g_l = 0 \; . \tag{22}$$

Let $I_j \subseteq J_j$ be given by $I_j = \{l \in J_j : \; z_l \neq 0\}$. Assume that there exists $c \in \sigma(S - S_0)$ such that $c_j < c$ and let $h \in \mathbb{C}^d$ be such that $\|h\| = 1$ and $(S - S_0)\, h = c\, h$. For $t \in (-1/2, 1/2)$ let $\mathcal{G}(t) = \{g_i(t)\}_{i\in\mathbb{I}_k}$ be given by

$$g_l(t) = \begin{cases} (1 - t^2\, |z_l|^2)^{1/2} g_l + t\, z_l\, a_l^{1/2}\, h & \text{if } \; l \in I_j \, ; \\ \qquad g_l & \text{if } \; l \in \mathbb{I}_k \setminus I_j \; . \end{cases}$$

Notice that $\mathcal{G}(t) \in \mathbb{T}_d(\mathbf{a})$ for $t \in (-1/2, 1/2)$. Let $\mathrm{Re}(A) = \frac{A+A^*}{2}$ denote the real part of $A \in \mathcal{M}_d(\mathbb{C})$. For $l \in I_j$ then

$$g_l(t) \otimes g_l(t) = (1 - t^2\, |z_l|^2)\; g_l \otimes g_l + t^2\, |z_l|^2\, a_l\; h \otimes h + 2\, (1 - t^2\, |z_l|^2)^{1/2}\, t\;\; \mathrm{Re}(h \otimes \overline{z_l}\, a_l^{1/2}\, g_l)$$

Notice that $\mathcal{G}(t)$ is a continuous curve in $\mathbb{T}_d(\mathbf{a})$ such that $\mathcal{G}(0) = \mathcal{G}_0$. Let $S(t)$ denote the frame operator of $\mathcal{G}(t) \in \mathbb{T}_d(\mathbf{a})$, so that $S(0) = S_0$, and let $T(t) = S - S(t)$ for $t \in (-1/2\, ,\, 1/2)$. Note that

$$T(t) = S - S_0 + t^2 \sum_{l\in I_j} |z_l|^2\, (g_l \otimes g_l - a_l\; h \otimes h) + R(t)$$

where $R(t) = -2 \sum\limits_{l\in I_j} (1 - t^2\, |z_l|^2)^{1/2}\, t\;\; \mathrm{Re}(h \otimes a_l^{1/2}\, \overline{z_l}\, g_l)$. Then $R(t)$ is a smooth function such that

$$R(0) = 0 \;\; , \;\; R'(0) = -\sum_{l\in I_j} \mathrm{Re}(h \otimes \overline{z_l}\, a_l^{1/2}\, g_l) = -\,\mathrm{Re}(h \otimes \sum_{l\in I_j} \overline{z_l}\, a_l^{1/2}\, g_l) \overset{(22)}{=} 0 \; ,$$

and such that $R''(0) = 0$. Therefore $\lim\limits_{t\to 0}\; t^{-2}\; R(t) = 0$. We now consider

$$V = \mathrm{span}\, \big(\{g_l : \; l \in I_j\} \cup \{h\} \big) = \mathrm{span}\, \big\{ g_l : \; l \in I_j \big\} \overset{\perp}{\oplus} \mathbb{C} \cdot h \; .$$

Then $\dim V = s + 1$, for $s = \dim \mathrm{span}\{g_l : \; l \in I_j\} \geq 1$. By construction, the subspace V reduces $S - S_0$ and $T(t)$ in such a way that $(S - S_0)|_{V^\perp} = T(t)|_{V^\perp}$, for $t \in (-1/2\, ,\, 1/2)$. On the other hand

$$T(t)|_V = (S - S_0)|_V + t^2 \sum_{l\in I_j} |z_l|^2\, (g_l \otimes g_l - a_l\; h \otimes h) + R(t) = A(t) + R(t) \in L(V) \; , \tag{23}$$

where we use the fact that the ranges of the selfadjoint operators in the second and third term in the formula above clearly lie in V. Then $\lambda\big((S - S_0)|_V \big) = \big(c\, ,\, c_j\, \mathbb{1}_s \big) \in (\mathbb{R}^{s+1}_{>0})^{\downarrow}$ and

$$\lambda\Big(\textstyle\sum_{l\in I_j} |z_l|^2 g_l \otimes g_l \Big) \;\; = (\gamma_1\, ,\, \ldots\, ,\, \gamma_s\, ,\, 0) \in (\mathbb{R}^{s+1}_{\geq 0})^{\downarrow} \quad \text{with} \quad \gamma_s > 0 \; ,$$

where we have used the definition of s and the fact that $|z_l| > 0$ for $l \in I_j$ (and the known fact that if $S\, ,\, T \in \mathcal{M}_d(\mathbb{C})^+ \implies R(S + T) = R(S) + R(T)$). Hence, for sufficiently small t, the spectrum of the operator $A(t) \in L(V)$ defined in Eq. (23) is

$$\lambda\big(A(t) \big) \;\; = \big(c - t^2 \textstyle\sum_{l\in I_j} a_l\, |z_l|^2\, ,\, c_j + t^2\, \gamma_1\, ,\, \ldots\, ,\, c_j + t^2\, \gamma_s \big) \in (\mathbb{R}^{s+1}_{\geq 0})^{\downarrow} \; ,$$

where we have used the fact that $\langle g_l\, ,\, h\rangle = 0$ for every $l \in I_j$. Let us now consider

$$\lambda\big(R(t) \big) = \big(\delta_1(t)\, ,\, \ldots\, ,\, \delta_{s+1}(t) \big) \in (\mathbb{R}^{s+1}_{\geq 0})^{\downarrow} \quad \text{for} \quad t \in \mathbb{R} \; .$$

Recall that in this case $\lim_{t\to 0} t^{-2}\delta_j(t) = 0$ for $1 \leq j \leq s+1$. Using Weyl's inequality on Eq. (23), we now see that

$$\lambda\big(\, T(t)|_V \,\big) \prec \lambda\big(\, A(t) \,\big) + \lambda\big(\, R(t) \,\big) \stackrel{\text{def}}{=} \rho(t) \in (\mathbb{R}^{s+1}_{\geq 0})^{\downarrow}\,. \tag{24}$$

We know that

$$\begin{aligned}\rho(t) &= \big(\, c - t^2 \textstyle\sum_{l\in I_j} a_l\,|z_l|^2 + \delta_1(t)\,,\, c_j + t^2\,\gamma_1 + \delta_2(t)\,,\,\ldots\,,\, c_j + t^2\,\gamma_s + \delta_{s+1}(t)\,\big)\\ &= \Big(\, c - t^2\,(\textstyle\sum_{l\in I_j} a_l\,|z_l|^2 + \frac{\delta_1(t)}{t^2})\,,\, c_j + t^2\,(\gamma_1 + \frac{\delta_2(t)}{t^2})\,,\,\ldots\,,\, c_j + t^2\,(\gamma_s + \frac{\delta_{s+1}(t)}{t^2})\,\Big)\,.\end{aligned}$$

Since by hypothesis $c_j < c$ then, the previous remarks show that there exists $\varepsilon > 0$ such that if $t \in (0,\varepsilon)$ then, for every $i \in \mathbb{I}_s$

$$c > c - t^2\,(\sum_{l\in I_j} a_l\,|z_l|^2 + \frac{\delta_1(t)}{t^2}) > c_j + t^2(\gamma_i + \frac{\delta_{i+1}(t)}{t^2}) > c_j\,.$$

The previous facts show that for $t \in (0,\varepsilon)$ then $\rho(t) \prec \lambda((S-S_0)|_V) = \big(\, c\,,\, c_j\,\mathbb{1}_s\,\big)$ strictly. Therefore,

$$\begin{aligned}\lambda(T(t)) &= \big(\lambda((S-S_0)|_{V^\perp})\,,\, T(t)|_V\,\big)^{\downarrow} \stackrel{(24)}{\prec} \big(\lambda((S-S_0)|_{V^\perp})\,,\, \rho(t)\,\big)\\ &\prec \big(\lambda((S-S_0)|_{V^\perp})\,,\, \lambda(S-S_0)|_V\,\big)^{\downarrow} = \lambda(S-S_0)\,,\end{aligned}$$

where the second majorization relation is strict (i.e. $(\lambda((S-S_0)|_{V^\perp})\,,\,\rho(t))^{\downarrow} \neq (\lambda((S-S_0)|_{V^\perp})\,,\,\lambda(S-S_0)|_V)^{\downarrow}$). Since N is strictly convex, for every $t \in (0,\varepsilon)$ we have that

$$\Theta(\mathcal{G}(t)) = N(T(t)) < N(S-S_0) = \Theta(\mathcal{G})\,.$$

This last fact contradicts the assumption that $\mathcal{G}_0$ is a local minimizer of Θ in $\mathbb{T}_d(\mathbf{a})$. □

4.3 Some special cases of Conjecture 4.2

Consider Notation 4.4 and assume that $k \geq d$; if we let $W = R(S_0) \subset \mathbb{C}^d$ then, as shown in Corollary 4.6, W reduces the self-adjoint operator $S - S_0 \in \mathcal{H}(d)$. In this section we show that in case W is an eigenspace of $S - S_0$ then Conjecture 4.2 holds for $\mathcal{G}_0$ i.e., $\mathcal{G}_0$ is a global minimizer of $\Theta_{(N\,,S\,,\mathbf{a})}$ in $\mathbb{T}_d(\mathbf{a})$. In order to tackle this particular case, we introduce the following

Remark 4.8 (A naive model)**.** Fix $S \in \mathcal{M}_d(\mathbb{C})^+$, $\mathbf{a} = (a_i)_{i\in\mathbb{I}_k} \in (\mathbb{R}^k_{>0})^{\downarrow}$ and a strictly convex u.i.n. N. We let $t = \mathrm{tr}(\mathbf{a}) = \sum_{i\in\mathbb{I}_k} a_i > 0$. If $\mathcal{G} \in \mathbb{T}_d(\mathbf{a})$ then it is clear that

$$S_{\mathcal{G}} \in \mathcal{M}_d(\mathbb{C})^+ \quad \text{ and } \quad \mathrm{tr}(S_{\mathcal{G}}) = \sum_{i\in\mathbb{I}_k} \|g_i\|^2 = \mathrm{tr}(\mathbf{a}) = t\,.$$

Hence, we consider

$$\mathcal{M}_d(\mathbb{C})^+_t = \{A:\ A \in \mathcal{M}_d(\mathbb{C})^+\,,\,\mathrm{tr}(A) = t\} \supset \{S_{\mathcal{G}}:\ \mathcal{G} \in \mathbb{T}_d(\mathbf{a})\}\,, \tag{25}$$

endowed with the metric induced by the operator norm. Moreover, we consider the map

$$\mathcal{D}_{(N,\,S,\,t)} = \mathcal{D} : \mathcal{M}_d(\mathbb{C})^+_t \to \mathbb{R}_{\geq 0} \quad \text{ given by } \quad \mathcal{D}(A) = N(S-A)\,. \tag{26}$$

By Eq. (25) we see that

$$\min\{\mathcal{D}(A):\ A \in \mathcal{M}_d(\mathbb{C})^+_t\} \leq \min\{\Theta(\mathcal{G}):\ \mathcal{G} \in \mathbb{T}_d(\mathbf{a})\}\,. \tag{27}$$

The inequality in Eq. (27) can be strict. Yet, we will show that under some additional hypothesis equality holds in Eq. (27). Moreover, since $\mathcal{M}_d(\mathbb{C})^+_t$ is a (larger but) simpler set, we are able to compute those $A \in \mathcal{M}_d(\mathbb{C})^+_t$ that attain the minimum in the left hand side of Eq. (27) (see Theorem 4.9 below); these facts together will allow us to prove Conjecture 4.2 in some special cases. □

Theorem 4.9. *Let $S \in \mathcal{M}_d(\mathbb{C})^+$, $\lambda(S) = (\lambda_i)_{i\in\mathbb{I}_d} \in (\mathbb{R}^d_{\geq 0})^{\downarrow}$, $t > 0$ and let N be a u.i.n. Consider $\{v_i\}_{i\in\mathbb{I}_d}$ an ONB of $\mathbb{C}^d$ such that $S\,v_i = \lambda_i\, v_i$, for $i \in \mathbb{I}_d$. Let $c \leq \lambda_1$ be uniquely determined by $\sum_{i\in\mathbb{I}_d}(\lambda_i - c)^+ = t$ and set*

$$A^{\rm op} = \sum_{i\in\mathbb{I}_d}(\lambda_i - c)^+\, v_i \otimes v_i \in \mathcal{M}_d(\mathbb{C})^+_t \quad \text{so that} \quad \lambda(S - A^{\rm op}) = (\min\{c\,,\,\lambda_i\})_{i\in\mathbb{I}_d} \in (\mathbb{R}^d)^{\downarrow}\,.$$

Then, $A^{\rm op}$ is a global minimizer of $\mathcal{D}$, defined as in Eq. (26).

Proof. By construction we see that $\lambda(S - A^{\rm op}) = (\min\{c\,,\,\lambda_i\})_{i\in\mathbb{I}_d}$. Let $A \in \mathcal{M}_d(\mathbb{C})^+_t$ be arbitrary; we consider the following cases:

In case $c \leq \lambda_d$ then we see that $\lambda(S - A^{\rm op}) = c\,\mathbb{1}_d$. Since $\mathrm{tr}(A) = t$, then $\mathrm{tr}(\lambda(S - A)) = \mathrm{tr}(S - A) = \mathrm{tr}(S) - t = \mathrm{tr}(\lambda(S - A^{\rm op}))$. Thus, in this case we have (see item 4. in Remark 2.2) that $\lambda(S - A^{\rm op}) = c\,\mathbb{1}_d \prec \lambda(S - A)$. Hence, we conclude that $\mathcal{D}(A^{\rm op}) = N(S - A^{\rm op}) \leq N(S - A) = \mathcal{D}(A)$.

In case $c > \lambda_d$, there exists $r \in \mathbb{I}_{d-1}$ such that $\lambda_r \geq c > \lambda_r + 1$. Then,

$$(\gamma_i)_{i\in\mathbb{I}_d} := \lambda(S - A^{\rm op}) = (c\,\mathbb{1}_r\,,\,\lambda_{r+1}\,,\,\ldots\,,\,\lambda_d) \in (\mathbb{R}^d)^{\downarrow}\,.$$

If we let $\lambda(A) = (\alpha_i)_{i\in\mathbb{I}_d} \in (\mathbb{R}^d_{\geq 0})^{\downarrow}$ then, by Lidskii's additive inequality, we get that

$$(\delta_i)_{i\in\mathbb{I}_d} := ((\lambda_i - \alpha_i)_{i\in\mathbb{I}_d})^{\downarrow} = (\lambda(S) - \lambda(A))^{\downarrow} \prec \lambda(S - A)\,. \tag{28}$$

We now show that $(\gamma_i)_{i\in\mathbb{I}_d} \prec (\delta_i)_{i\in\mathbb{I}_d}$; by construction $\mathrm{tr}((\gamma_i)_{i\in\mathbb{I}_d}) = \mathrm{tr}((\delta_i)_{i\in\mathbb{I}_d})$ that is

$$\mathrm{tr}(\gamma) = \sum_{j=1}^{d}\gamma_j = r\,c + \sum_{j=r+1}^{d}\lambda_j = \mathrm{tr}(\delta) = \sum_{j=1}^{d}\delta_j = \sum_{j=1}^{d}(\lambda_j - \alpha_j)\,. \tag{29}$$

Thus, in order to show that $(\gamma_i)_{i\in\mathbb{I}_d} \prec (\delta_i)_{i\in\mathbb{I}_d}$ we need to prove that $\sum_{j=k}^{d}\gamma_j \geq \sum_{j=k}^{d}\delta_j$, for every $k \in \mathbb{I}_d$, since the vectors are arranged in non-increasing order. Notice that $\lambda_i \geq \lambda_i - \alpha_i$, for every $i \in \mathbb{I}_d$; then, by Remark 2.2, we conclude that $\lambda_i \geq \delta_i$, for $i \in \mathbb{I}_d$. This guarantees that, for $r + 1 \leq k \leq d$,

$$\sum_{j=k}^{d}\gamma_j = \sum_{j=k}^{d}\lambda_j \geq \sum_{j=1}^{k}\delta_j. \tag{30}$$

We now define $\beta = \sum_{j=r+1}^{d}(\gamma_j - \delta_j)$ and notice that Eq. (30) shows that $\beta \geq 0$. By Eq. (29),

$$\sum_{j=1}^{r}\delta_j = r\,(c + \beta/r),$$

which implies that $(c + \beta/r)\mathbb{1}_r \prec (\delta_1, \cdots, \delta_r)$. Hence, if $1 \leq k \leq r$ then

$$\sum_{j=k}^{r}\delta_j \leq (r - k + 1)\,(c + \beta/r) \leq (r - k + 1)\,c + \beta\,. \tag{31}$$

Therefore, for $1 \leq k \leq r$,

$$\sum_{j=k}^{d}\gamma_j - \sum_{j=k}^{d}\delta_j = (r - k + 1)\,c + \beta - \sum_{j=k}^{r}\delta_j \overset{(31)}{\geq} 0\,. \tag{32}$$

Then, Eqs. (29), (31) and (32) show that $\gamma \prec \delta$. Finally, if N is a (strictly convex) u.i.n. then

$$N(S - A^{\rm op}) = N(D_\gamma) \leq N(D_\delta) \overset{(28)}{\leq} N(S - A)$$

so $A^{\rm op}$ is a global minimizer of $\mathcal{D}$ in $\mathcal{M}_d(\mathbb{C})^+_t$. □

The next result verifies Conjecture 4.2 under some additional assumptions on the spectral structure of local minimizers.

Theorem 4.10. *Let $S \in \mathcal{M}_d(\mathbb{C})^+$, $\mathbf{a} = (a_i)_{i\in\mathbb{I}_k} \in (\mathbb{R}^k_{>0})^{\downarrow}$, with $k \geq d$, and let N be a strictly convex u.i.n. in $\mathcal{M}_d(\mathbb{C})$. Let $\mathcal{G}_0 = \{g_i\}_{i\in\mathbb{I}_d}$ be a local minimizer of Θ in $\mathbb{T}_d(\mathbf{a})$ such that there exists $c_1 \in \mathbb{R}$ that satisfies $(S - S_{\mathcal{G}_0})g_i = c_1\, g_i$, for $i \in \mathbb{I}_k$. Then there exists an ONB $\{v_i\}_{i\in\mathbb{I}_d}$ of $\mathbb{C}^d$ such that*

$$S = \sum_{i\in\mathbb{I}_d} \lambda_i\ v_i \otimes v_i \quad \text{and} \quad S_{\mathcal{G}_0} = \sum_{i\in\mathbb{I}_d} (\lambda_i - c_1)^+\ v_i \otimes v_i\,, \tag{33}$$

where $(\lambda_i)_{i\in\mathbb{I}_d} = \lambda(S) \in (\mathbb{R}^d_{\geq 0})^{\downarrow}$. Moreover, $\lambda(S - S_{\mathcal{G}_0}) \prec \lambda(S - S_{\mathcal{G}})$ for $\mathcal{G} \in \mathbb{T}_d(\mathbf{a})$. In particular, $\mathcal{G}_0$ is a global minimizer of Θ in $\mathbb{T}_d(\mathbf{a})$.

Proof. Let $S_0 = S_{\mathcal{G}_0}$. By Theorem 4.5 there exists an ONB $\{v_i\}_{i\in\mathbb{I}_d}$ of $\mathbb{C}^d$ such that

$$S = \sum_{i\in\mathbb{I}_d} \lambda_i\ v_i \otimes v_i \quad \text{and} \quad S_0 = \sum_{i\in\mathbb{I}_d} \lambda_i(S_0)\ v_i \otimes v_i\,. \tag{34}$$

In particular, $\lambda(S - S_0) = (\lambda(S) - \lambda(S_0))^{\downarrow}$. Let $W = R(S_0) = \operatorname{span}\{g_i :\ i \in \mathbb{I}_k\}$, which reduces $S - S_0$ by Corollary 4.6. Then, by hypothesis we have that $\sigma((S - S_0)|_W) = \{c_1\}$. We consider the following two cases:

Assume that $W = \mathbb{C}^d$. In this case $\sigma(S - S_0) = \{c_1\}$ and therefore $\lambda(S - S_0) = c_1\, \mathbb{1}_d$. Thus, $\lambda_i - \lambda_i(S_0) = c_1$ which implies that $\lambda_i(S_0) = (\lambda_i - c_1)^+$, for $i \in \mathbb{I}_d$. Notice that for every $\mathcal{G} \in \mathbb{T}_d(\mathbf{a})$ we have that $\operatorname{tr}(S - S_{\mathcal{G}}) = \operatorname{tr}(S) - \operatorname{tr}(\mathbf{a})$; then we see that $c_1\, \mathbb{1}_d = \operatorname{tr}(\lambda(S - S_0)) = \operatorname{tr}(S - S_{\mathcal{G}})$ which shows (see item 4. in Remark 2.2) that $\lambda(S - S_0) \prec \lambda(S - S_{\mathcal{G}})$ for every $\mathcal{G} \in \mathbb{T}_d(\mathbf{a})$. This last fact implies that $\Theta(S_{\mathcal{G}_0}) = N(S - S_0) \leq N(S - S_{\mathcal{G}}) = \Theta(\mathcal{G})$, for every $\mathcal{G} \in \mathbb{T}_d(\mathbf{a})$. Thus, $\mathcal{G}_0$ is a global minimizer of Θ in $\mathbb{T}_d(\mathbf{a})$.

Assume now that $W \neq \mathbb{C}^d$. Hence, $d > \dim W = \operatorname{span}\{g_i :\ i \in \mathbb{I}_k\}$ which shows that $\mathcal{G}_0$ is a linearly dependent family, since $k \geq d$. Then, Theorem 4.7 implies that $c \leq c_1$ for every $c \in \sigma(S - S_0)$. Let $1 \leq r \leq d - 1$ be such that $\dim W = r$. Hence, $\lambda(S_0) = (\lambda_1(S_0)\,, \dots\,, \lambda_r(S_0)\,, 0_{d-r})$ and $W = \operatorname{span}\{v_i : 1 \leq i \leq r\}$. Therefore, using Eq. (34) and the previous facts we conclude that

$$S - S_0 = \sum_{i=1}^{r} (\lambda_i - \lambda_i(S_0))\ v_i \otimes v_i + \sum_{i=r+1}^{d} \lambda_i\ v_i \otimes v_i = c_1 \sum_{i=1}^{r} v_i \otimes v_i + \sum_{i=r+1}^{d} \lambda_i\ v_i \otimes v_i$$

Thus, $\sigma(S - S_0) \ni \lambda_i \leq c_1$ and $\lambda_i(S_0) = 0$, for $r + 1 \leq i \leq d$; hence, $\lambda_i(S_0) = (\lambda_i(S) - c_1)^+$, for $r + 1 \leq i \leq d$. Then, $\lambda_i(S_0) = (\lambda_i - c_1)^+$ for $i \in \mathbb{I}_d$ and therefore we obtain the representation of S_0 as in Eq. (33).

Notice that $c_1 = \lambda_1 - (\lambda_1 - c_1)^+$, since $W \neq \{0\}$. This shows that $c_1 \leq \lambda_1$. Moreover, if we let $\operatorname{tr}(\mathbf{a}) = t > 0$ then

$$\sum_{i\in\mathbb{I}_d} (\lambda_i - c_1)^+ = \operatorname{tr}(S_{\mathcal{G}_0}) = \operatorname{tr}(\mathbf{a}) = t\,.$$

Using Remark 4.8 and Theorem 4.9 we now see that for every $\mathcal{G} \in \mathbb{T}_d(\mathbf{a})$ we have that $\lambda(S - S_{\mathcal{G}_0}) \prec \lambda(S - S_{\mathcal{G}})$; in particular,

$$\Theta(\mathcal{G}) = N(S - S_{\mathcal{G}}) = \mathcal{D}(S_{\mathcal{G}}) \geq \mathcal{D}(S_0) = N(S - S_0) \quad \text{since} \quad S_{\mathcal{G}} \in \mathcal{M}_d(\mathbb{C})^+_t\,.$$

Thus, $\mathcal{G}_0$ is a global minimizer of Θ in $\mathbb{T}_d(\mathbf{a})$. □

Acknowledgment. We would like to thank Professor Eduardo Chiumiento for fruitful conversations related to the content of this work.

References

[1] E. Andruchow and D. Stojanoff, Geometry of Unitary Orbits, J. Operator Theory **26** (1991), 25-41.

[2] J. Antezana, E. Chiumiento, Approximation by partial isometries and symmetric approximation of finite frames, J. Fourier Anal Appl (2017). https://doi.org/10.1007/s00041-017-9547-5.

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

[4] M. Bownik, J. Jasper, Existence of frames with prescribed norms and frame operator. Excursions in harmonic analysis. Vol. 4, 103-117, Appl. Numer. Harmon. Anal., Birkhäuser/Springer, Cham, 2015.

[5] P. G. Casazza and G. Kutyniok eds., Finite Frames: Theory and Applications. Birkhauser, 2012. xii + 483 pp.

[6] O. Christensen, An introduction to frames and Riesz bases. Applied and Numerical Harmonic Analysis. Birkhäuser Boston, 2003. xxii+440 pp.

[7] D. Deckard and L. A. Fialkow, Characterization of Hilbert space operators with unitary cross sections, J. Operator Theory **2** (1979), 153-158.

[8] G: Eckart and G: Young, The approximation of one matrix by another of lower rank, Psychometrika **1** (1936), 211-218.

[9] W. Fulton, Eigenvalues, invariant factors, highest weights, and Schubert calculus, Bull. Amer. Math. Soc. (N.S.) 37 (2000), no. 3, 209-249.

[10] N.J. Higham, Matrix nearness problems and applications. Applications of matrix theory (Bradford, 1988), 1-27, Inst. Math. Appl. Conf. Ser. New Ser., 22, Oxford Univ. Press, New York, 1989.

[11] A. Horn, Eigenvalues of sums of Hermitian matrices, Pacific J. Math. 12 (1962), 225-241.

[12] R.A. Horn, C.R. Johnson, Matrix analysis. Second edition. Cambridge University Press, Cambridge, 2013.

[13] R.A. Horn, C.R. Johnson, Topics in matrix analysis. Corrected reprint of the 1991 original. Cambridge University Press, Cambridge, 1994.

[14] A. A. Klyachko, Stable bundles, representation theory and Hermitian operators, Selecta Math. 4 (1998), 419-445.

[15] A. Knutson and T. Tao, The honeycomb model of GLn(C) tensor products I: proof of the saturation conjecture, J. Amer. Math. Soc. 12 (1999), 1055-1090,

[16] C.K. Li, Y.T. Poon, T. Schulte-Herbrüggen, Least-squares approximation by elements from matrix orbits achieved by gradient flows on compact Lie groups. Math. Comp. 80 (2011), no. 275, 1601-1621.

[17] V.B. Lidskii, On the characteristic numbers of the sum and product of symmetric matrices. (Russian) Doklady Akad. Nauk SSSR (N.S.) 75, (1950). 769-772.

[18] P. Massey, N. B. Rios, D. Stojanoff, Frame completions with prescribed norms: local minimizers and applications. Adv. Comput. Math. 44 (2018), no. 1, 51–86.

[19] P. Massey, M.A. Ruiz, Tight frame completions with prescribed norms. Sampl. Theory Signal Image Process. 7 (2008), no. 1, 1-13.

[20] P. Massey, M. Ruiz; Minimization of convex functionals over frame operators. Adv. Comput. Math. 32 (2010), no. 2, 131-153.

[21] P. Massey, M. Ruiz, D. Stojanoff; Optimal dual frames and frame completions for majorization. Appl. Comput. Harmon. Anal. 34 (2013), 201-223.

[22] P.G. Massey, M.A. Ruiz, D. Stojanoff; Optimal frame completions. Advances in Computational Mathematics 40 (2014), 1011-1042.

[23] P. Massey, M. Ruiz, D. Stojanoff; Optimal frame completions with prescribed norms for majorization. J. Fourier Anal. Appl. 20 (2014), no. 5, 1111-1140.

[24] N. Strawn; Optimization over finite frame varieties and structured dictionary design, Appl. Comput. Harmon. Anal. 32 (2012) 413-434.