JOTA manuscript No.
(will be inserted by the editor)

Uniqueness in nuclear norm minimization

Flatness of the nuclear norm sphere and simultaneous polarization

Tim Hoheisel · Elliot Paquette

October 18, 2022

Abstract In this paper we establish necessary and sufficient conditions for the existence of line segments (or *flats*) in the sphere of the nuclear norm via the notion of *simultaneous polarization* and a refined expression for the subdifferential of the nuclear norm. This is then leveraged to provide (point-based) necessary and sufficient conditions for uniqueness of solutions for minimizing the nuclear norm over an affine subspace. We further establish an alternative set of sufficient conditions for uniqueness, based on the interplay of the subdifferential of the nuclear norm and the range of the problem-defining linear operator. Finally, we show how to transfer the uniqueness results for the original problem to a whole class of nuclear norm-regularized minimization problems with a strictly convex fidelity term.

Keywords Nuclear norm · singular value decomposition · polar decomposition · convex analysis · convex subdifferential · Fenchel conjugate · low rank minimization

Mathematics Subject Classification (2000) 15A1 · 47N10 · 65F22 · 90C25 · 90C27

Tim Hoheisel (corresponding author)
Department of Mathematics and Statistics, McGill University, 805 Sherbrooke St West, Montréal, Québec, Canada H3A 0B9
tim.hoheisel@mcgill.ca

Elliot Paquette
Department of Mathematics and Statistics, McGill University, 805 Sherbrooke St West, Montréal, Québec, Canada H3A 0B9
elliot.paquette@mcgill.ca

1 Introduction

One of the most ubiquitous paradigms for linear inverse problems in matrix space is *low rank approximation*, often cast in the form

$$\min_{X\in\mathbb{R}^{n\times p}} \operatorname{rank} X \text{ s.t. } \mathcal{A}(X)=b. \tag{1}$$

Here $\mathcal{A}:\mathbb{R}^{n\times p}\to\mathbb{E}$ is a linear map (into a Euclidean space $\mathbb{E}$) whose action is often simply a matrix multiplication $\mathcal{A}(X)=AX$ for some $A\in\mathbb{R}^{m\times n}$ or a selection operator which projects X onto the matrix composed of its entries from a prescribed index set $J\subset\{1,\dots,n\}\times\{1,\dots,p\}$. We direct the interested reader to Fazel's thesis [4], the important paper by Candès and Recht [2] as well as the survey article by Recht et al. [16] for applications, solution methods and pointers to the abundant literature for the low rank minimization problem (1) and the low rank minimization paradigm in general.

Due to the combinatorial nature of the rank function, problem (1) is, generally, NP-hard (as it contains cardinality minimization as a special case, which is NP-hard [5, 15]), and therefore many continuous relaxations for its numerical solution have been proposed. The predominant class of convex relaxations uses the *nuclear norm* (or *trace norm*) $\|\cdot\|_*$ as a convex approximation of the rank function. The justification for this stems from the fact that the nuclear norm is the *convex envelope* (i.e. the largest convex minorant) of the rank function when restricted to a spectral norm unit ball around the origin, a fact that was first established by Fazel in her thesis [4] (see also the approach by Hiriart-Urruty and Le [7]). On the other hand, the nuclear norm is simply the ℓ_1-norm of the vector of singular values, and the ℓ_1-norm is known to promote sparsity [3], hence the nuclear norm promotes low rank. Various nuclear norm-based approximations of problem (1) have been proposed, the most obvious one being

$$\min_{X\in\mathbb{R}^{n\times p}} \|X\|_* \text{ s.t. } \mathcal{A}(X)=b. \tag{2}$$

Existence of solutions for this problem[1] is readily established (whenever the problem is feasible) as the objective function is *coercive* (and the suitable continuity properties are satisfied). Given a solution $\bar{X}$ of (2), the goal of this paper is to establish conditions that guarantee that $\bar{X}$ is, in fact, the unique solution. This is inspired by the study by Zhang et al. [26] which establishes uniqueness results for ℓ_1-minimization problems[2]

We approach this task by combining tools from convex analysis and linear algebra. The natural interplay of these areas is most obvious in the study of *unitarily invariant norms* [9] which comes into play here since the nuclear norm (and its dual norm, the spectral norm) are unitarily invariant. This theory goes back to work of von Neumann [14], expanded on by various authors including Watson [22, 23], Zietak [24, 25] and de Sá [19, 20], and then vastly generalized beyond norms in Lewis' seminal work [10–12].

[1] Of course, we assume throughout that this problem is feasible.

[2] Nuclear norm minimization contains ℓ_1-minimization as a special case since $x\in\mathbb{R}^n$ can be identified with a diagonal matrix $\operatorname{diag}(x)$ for which $\|\operatorname{diag}(x)\|_*=\|x\|_1$.

1.1 Contributions

Our first main contribution, Theorem 3.1, provides a characterization of the existence of line segments (*flatness*) in the boundary of the nuclear norm ball, based on the notion of *simultaneous polarizability* (Definition 3.1). In Corollary 3.2 we give a reformulation of this characterization using the singular value decomposition of a point in the nuclear norm sphere, and this directly carries over to *necessary and sufficient* conditions for uniqueness (Corollary 4.1) for solutions of the nuclear norm minimization problem (2).

We then extend the study by Zhang et al. [26] to the nuclear norm setting, starting from the following observation of Gilbert [6] for any (proper) convex function f (see Proposition 4.1): $\bar{x}$ is the unique minimizer of f if 0 is in the interior of the subdifferential of f at $\bar{x}$. We make these conditions concrete for problem (2) in Proposition 4.2. We then bridge between these convex-analytic conditions and the linear-algebraic ones established earlier in Corollary 4.1 explicitly in Proposition 4.3, thus illuminating their connection. By means of a counterexample (Example 4.1) we show that the sufficient conditions (Assumption 1) are not necessary for uniqueness, which is in contrast to the (polyhedral convex) ℓ_1-case.

Through convex analysis (Proposition 4.4) we are able to transfer our findings for problem (2) to another class of nuclear norm minimization problems (see Corollary 4.4) including nuclear norm-regularized least-squares.

1.2 Roadmap

We present in Section 2 the necessary background from linear algebra and convex analysis, including a novel result on the convex geometry of the subdifferential of the nuclear norm. Section 3 is devoted to characterizing the existence of line segments in the nuclear norm sphere. We transfer these findings to nuclear norm minimization problems in Section 4. We close out with some final remarks in Section 5.

1.3 Notation

The vector $e_i \in \mathbb{R}^n$ is the i-th standard unit vector in $\mathbb{R}^n$. For a vector $x \in \mathbb{R}^n$, $\operatorname{diag}(x)$ will be a diagonal matrix with x on its diagonal, whose size will be clear from the context (and which may be rectangular). For a matrix X, we will generate the vector of its diagonal entries via $\mathrm{DIAG}(X)$. The space of $n \times n$ (real) symmetric matrices is denoted by $\mathbb{S}^n$, $\mathbb{S}^n_+$ is the positive semidefinite cone while $\mathbb{S}^n_{++}$ denotes the positive definite matrices in $\mathbb{S}^n$. For $S \in \mathbb{S}^n_+$, we denote its square root by $\sqrt{S}$. The set of $n \times n$ orthogonal matrices is denoted by $O(n)$. For a set C in a real vector space, we define $\mathbb{R}_+ C := \{tx \mid t \geq 0,\ x \in C\}$, the smallest cone that contains C. The line segment between two points x, y in a real vector space is denoted by $[x, y]$. The set of all linear maps between two Euclidean spaces V, W is denoted by $\mathcal{L}(V, W)$. For $\mathcal{A} \in \mathcal{L}(V, W)$, we write $\ker \mathcal{A}$ and $\operatorname{rge} \mathcal{A}$ for its *kernel* and *range*, respectively. Its adjoint map is denoted by $\mathcal{A}^*$.

2 Preliminaries

In what follows, $\mathbb{E}$ will be a finite-dimensional Euclidean space with inner product denoted by $\langle \cdot, \cdot \rangle$. The induced norm is denoted by $\| \cdot \|$, i.e. $\|x\| := \sqrt{\langle x, x \rangle}$ for all $x \in \mathbb{E}$. We equip $\mathbb{R}^{n \times p}$ with the (Frobenius) inner product

$$\langle X, Y \rangle := \operatorname{tr}(X^T Y) \quad \forall X, Y \in \mathbb{R}^{n \times p},$$

which induces the *Frobenius norm*

$$\|X\| := \sqrt{\langle X, X \rangle} = \sqrt{\sum_{i=1}^n \sum_{j=1}^p x_{ij}^2} \quad \forall X \in \mathbb{R}^{n \times p}.$$

For $X \in \mathbb{R}^{n \times p}$, its *operator norm* is given by

$$\|X\|_{op} := \max_{\|v\| \le 1} \|Xv\|,$$

while its *nuclear norm* is given by

$$\|X\|_* := \operatorname{tr}(\sqrt{X^T X}) = \operatorname{tr}(\sqrt{X X^T}) = \sum_{i=1}^n \sqrt{\lambda_i(X X^T)},$$

where $\lambda_i(X X^T)$ $(i = 1, \dots, n)$ are the (nonnegative) eigenvalues of $X X^T$. The following fact is frequently used in our study:

$$\|X\|_* = \operatorname{tr}(X) \quad \forall X \in \mathbb{S}^n_+. \tag{3}$$

The operator and the nuclear norm are *dual norms* in that

$$\|X\|_* = \max_{\|Y\|_{op} \le 1} \langle X, Y \rangle \quad \text{and} \quad \|Y\|_{op} = \max_{\|X\|_* \le 1} \langle X, Y \rangle .$$

The following simple estimate for the operator norm will be useful for our study.

Lemma 2.1 *For $A \in \mathbb{R}^{n \times p}$ the Euclidean norm of every column and row of A is bounded above by $\|A\|_{op}$. In particular, we have $a_{ij} \le \|A\|_{op}$ for all $i = 1, \dots, n,\ j = 1, \dots, p$.*

Proof Let a_j be the j-th column of A. Then

$$\|a_j\| = \|Ae_j\| \le \sup_{\|x\|=1} \|Ax\| = \|A\|_{op}.$$

Multiplying standard unit vectors e_j^T from the left, we get the analogous statement for rows. □

Lemma 2.2 *Let $n > p$, $X \in \mathbb{R}^{n\times p}$ and $Y \in \mathbb{R}^{n\times(n-p)}$. Then*

$$\|X\|_* \leq \|[X\ Y]\|_*,$$

where equality holds if and only if $Y = 0$.

Proof Recall that for $S \in \mathbb{S}^n_+$, we have $\lambda_i(\sqrt{S}) = \sqrt{\lambda_i(S)}$ for all $i = 1, \dots, n$. Now, observe that $A := XX^T + YY^T \succeq XX^T =: B$. Then

$$\lambda_i(A) \geq \lambda_i(B) \quad \forall i = 1, \dots, n,$$

see [9, Corollary 7.7.4(c)], which already proves the inequality. Now, using [9, Corollary 7.7.4(d)] and the above, we find

$$\begin{aligned} Y = 0 &\iff A = B \\ &\iff \mathrm{tr}\,(A) = \mathrm{tr}\,(B) \\ &\iff \lambda_i(A) = \lambda_i(B) \quad \forall i = 1, \dots, n \\ &\iff \lambda_i(\sqrt{A}) = \lambda_i(\sqrt{B}) \quad \forall i = 1, \dots, n \\ &\iff \|X\|_* = \|[XY]\|_*. \end{aligned}$$

This proves the result. □

We point out that the above result allows one to always embed problem (2) (defined by $\mathcal{A} \in \mathcal{L}(\mathbb{R}^{n\times p}, \mathbb{E})$ and $b \in \mathbb{E}$) in a rectangular matrix space $\mathbb{R}^{n\times p}$ (w.l.o.g.[3] $n \geq p$) into the square matrix space $\mathbb{R}^{n\times n}$. To this end, identify every element $\tilde{X} \in \mathbb{R}^{n\times n}$ with the block matrix $\tilde{X} = [X\ Y]$ for $X \in \mathbb{R}^{n\times p}, Y \in \mathbb{R}^{n\times(n-p)}$, define the linear operator $\tilde{\mathcal{A}} : \tilde{X} \to \mathcal{A}(X)$ and the right-hand side $\tilde{b} := b$. If we now consider the 'padded' problem

$$\min_{[X\ Y]\in\mathbb{R}^{n\times n}} \|[X\ Y]\|_* \text{ s.t. } \tilde{\mathcal{A}}([X\ Y]) = \tilde{b}, \tag{4}$$

it is an immediate consequence of Lemma 2.2 that $\bar{X}$ is a solution of (2) if and only if $[\bar{X}\ 0]$ is a solution of (4). In fact, we have

$$\operatorname*{argmin}_{[X\ Y]\in\mathbb{R}^{n\times n}} \left\{ \|[X\ Y]\|_* \;\middle|\; \tilde{\mathcal{A}}([X\ Y])\} = \tilde{b} \right\} = \left\{ [\bar{X}\ 0] \in \mathbb{R}^{n\times n} \;\middle|\; \bar{X} \text{ solves } (2) \right\};$$

in particular, uniqueness also carries over from one problem to the other.

[3] The nuclear norm of X and X^T are equal; the linear equation $\mathcal{A}(X) = b$ can always be rewritten in terms of X^T.

2.1 Singular value decomposition

For the facts and concepts presented in this paragraph we refer the reader to Horn and Johnson [9] for details. Throughout (w.l.o.g.) we assume that $n \geq p$. For $X \in \mathbb{R}^{n\times p}$, with rank $X = r$, there exist orthogonal matrices $U \in O(n)$ and $V \in O(p)$ (with columns $u_1, \dots, u_n$ and $v_1, \dots, v_p$, respectively) and unique real numbers

$$\sigma_1(X) \geq \sigma_2(X) \geq \sigma_r(X) > 0 = \sigma_{r+1} = \cdots = \sigma_n(X)$$

such that

$$X = U\mathrm{diag}(\sigma(X))V^T = \sum_{i=1}^{r} \sigma_i(X)u_i v_i^T.$$

This is called a *singular value decomposition* (SVD) of X. Note that the positive singular values of X are exactly the square roots of the nonzero eigenvalues of XX^T (or X^TX).
Through the singular value decomposition, we generate the map

$$\sigma : X \in \mathbb{R}^{n\times p} \to \sigma(X) \in \mathbb{R}^p.$$

We say that two matrices $X, Y \in \mathbb{R}^{n\times p}$ have a *simultaneous singular value decomposition* if there exist $(\bar{U}, \bar{V}) \in O(n) \times O(p)$ such that

$$X = \bar{U}\mathrm{diag}(\sigma(X))\bar{V}^T \quad \text{and} \quad Y = \bar{U}\mathrm{diag}(\sigma(Y))\bar{V}^T.$$

The next result, see e.g. [10,13,14], due to von Neumann, characterizes simultaneous singular value decompositions.

Theorem 2.1 (von Neumann) *For $X, Y \in \mathbb{R}^{n\times p}$ we have*

$$\langle X, Y\rangle \leq \langle \sigma(X), \sigma(Y)\rangle .$$

Equality holds if and only if X and Y have simultaneous singular value decompositions.

The nuclear and operator norm of X, respectively, can be expressed as the ℓ_1- and ℓ_∞-norm, respectively, of the vector of singular values of X, i.e.

$$\|X\|_* = \sum_{i=1}^{r} \sigma_i(X) \quad \text{and} \quad \|X\|_{op} = \sigma_1(X).$$

Moreover, the nuclear and the operator norm are *orthogonally invariant*, i.e. for all $X \in \mathbb{R}^{n\times p}$, we have

$$\|UXV\|_* = \|X\|_* \quad \text{and} \quad \|UXV\|_{op} = \|X\|_{op} \quad \forall (U, V) \in O(n) \times O(p). \tag{5}$$

There is an important extension of the above equation in the rectangular case. To formulate it, we recall the *Stiefel manifold* [21].

Definition 2.1 (Stiefel manifold) The *Stiefel manifold* $\mathcal{V}_{n,p}$ is the collection of matrices in $\mathbb{R}^{n\times p}$ with orthonormal columns, i.e.

$$\mathcal{V}_{n,p} := \left\{ U \in \mathbb{R}^{n\times p} \;\middle|\; U^T U = I_p \right\}.$$

The nuclear norm also has invariance on one side by multiplication by elements of the Stiefel manifold.

Lemma 2.3 *Let $X \in \mathbb{R}^{n\times p}$ and $U \in \mathcal{V}_{n,p}$. Then $\|XU^T\|_* = \|X\|_*$.*

Proof As U has orthonormal columns, we may extend it to an orthogonal matrix $[U\ W] \in O(n)$. Then

$$\|XU^T\|_* = \left\| [X\ 0] \cdot \begin{bmatrix} U^T \\ W^T \end{bmatrix} \right\|_* = \|[X\ 0]\|_* = \|X\|_*,$$

where the second identity uses the orthogonal invariance from (5) and the third is due to Lemma 2.2. □

2.2 Tools from convex analysis

For the facts and concepts presented in this paragraph we refer the uninitiated reader to the textbooks by Rockafellar [17], Hiriart-Urruty and Lemaréchal [8], Borwein and Lewis [1] or Rockafellar and Wets [18, Chapter 11].

A function $f : \mathbb{E} \to \mathbb{R} \cup \{+\infty\}$ is called proper if its domain, $\operatorname{dom} f := \{x \mid f(x) < +\infty\}$, is nonempty. We say that f is *convex* if its epigraph $\operatorname{epi} f := \{(x,\alpha) \in \mathbb{E}\times\mathbb{R} \mid f(x) \le \alpha\}$ is convex, and we say that it is *closed* if $\operatorname{epi} f$ is closed. Its conjugate $f^* : \mathbb{E} \to \mathbb{R}\cup\{+\infty\}$ is $f^*(y) := \sup_{x\in\operatorname{dom} f}\{\langle y,\, x\rangle - f(x)\}$. Its (convex) subdifferential at $\bar{x} \in \operatorname{dom} f$ is given by

$$\partial f(\bar{x}) := \{y \in \mathbb{E} \mid f(\bar{x}) + \langle y,\, x - \bar{x}\rangle \le f(x)\ \forall x \in \operatorname{dom} f\}.$$

An important (proper, convex) extended real-valued function is the *indicator function* of a (nonempty, convex) set $C \subset \mathbb{E}$ which is

$$\delta_C : \mathbb{E} \to \mathbb{R}\cup\{+\infty\}, \quad \delta_C(x) = \begin{cases} 0, & x \in C, \\ +\infty, & \text{else.} \end{cases}$$

Its subdifferential at $\bar{x} \in C$ is $\partial\delta_C(\bar{x}) = \{v \mid \langle v,\, x-\bar{x}\rangle \le 0\ \forall x \in C\}$. Its conjugate is the support function of C, i.e. $\delta_C^*(y) = \sup_{x\in C}\langle x,\, y\rangle =: \sigma_C(y)$. We point out that the support functions of the compact, convex, symmetric sets C that contain 0 in their interior are exactly the norms on $\mathbb{E}$ [17, Theorem 15.2] .

The subdifferential of the ℓ_1-norm $\|\cdot\|_1 : \mathbb{R}^n \to \mathbb{R}$, $\|x\|_1 = \sum_{i=1}|x_i|$, reads

$$\partial\|\cdot\|_1(x) = \left(\begin{cases} \operatorname{sgn}(x_i), & \text{if } x_i \ne 0, \\ [-1,1], & \text{if } x_i = 0 \end{cases} \right)_{i=1}^{n} = \{y \in \mathbb{B}_\infty \mid \langle x,\, y\rangle = \|x\|_1\}. \quad (6)$$

An important example to our study is the subdifferential of the nuclear norm.

Proposition 2.1 (Subdifferential of nuclear norm) *Let $\bar{X} \in \mathbb{R}^{n\times p}$ and let $(\bar{U}, \bar{V}) \in O(n) \times O(p)$ such that*

$$\bar{U}\mathrm{diag}(\sigma(\bar{X}))\bar{V}^T = \bar{X}.$$

The following hold:

(a) We have $Y \in \partial\|\cdot\|_*(\bar{X})$ if and only if $\bar{X}$ and Y have a simultaneous singular value decomposition and $\sigma(Y) \in \partial\|\cdot\|_1(\sigma(\bar{X}))$.
(b) It holds that

$$\partial\|\cdot\|_*(\bar{X}) = \left\{Y \;\middle|\; \langle \bar{X}, Y\rangle = \|\bar{X}\|_*,\ \|Y\|_{op} \le 1\right\}^4 \tag{7}$$

$$= \bar{U}\partial\|\cdot\|_*(\mathrm{diag}(\sigma(\bar{X}))\bar{V}^T. \tag{8}$$

Proof (a) See [10, Corollary 2.5].

(b) The expressions for the subdifferential can be found in [25], see, in particular, [25, Theorem 3.1] for the characterization in (8). □

For a convex set $C \subset \mathbb{E}$, its *affine hull*, denoted by $\mathrm{aff}\, C$, is the smallest affine set that contains C. In particular, $\mathrm{aff}\, C$ is a subspace if and only if it contains 0. The *subspace parallel to* C is defined to be the unique subspace parallel to $\mathrm{aff}\, C$ and given by $\mathrm{par}\, C := \mathrm{aff}\, C - \bar{x}$ for any $\bar{x} \in C$. Clearly, this entails that $\mathrm{ri}\, C = \mathrm{int}\, C$ if and only if the latter is nonempty, i.e. when $\mathrm{par}\, C = \mathbb{E}$.

The *relative interior* $\mathrm{ri}\, C$ of C is its interior in the relative topology with respect to its affine hull. The following characterization of relative interior points is useful to our study, see, e.g., [1, Exerc. 13, Ch. 1]:

$$x \in \mathrm{ri}\, C \quad\Longleftrightarrow\quad \mathbb{R}_+(C - x) = \mathrm{par}\, C. \tag{9}$$

For more details on the relative interior we refer the reader to Rockafellar [17, Chapter 6].

We will now exploit the representation in (8) to derive yet another representation of the subdifferential of the nuclear norm as well as its relative interior and parallel subspace. This is useful to our study but also of independent interest. We need the following lemma.

Lemma 2.4 *For $r \le p (\le n)$ set*

$$\mathcal{T} := \left\{B \in \mathbb{R}^{n\times p} \;\middle|\; \mathrm{DIAG}(B) \in \{1\}^r \times \mathbb{R}^{n-r},\ \|B\|_{op} \le 1\right\}.$$

Then

$$\mathcal{T} = \left\{\begin{pmatrix} I_r & 0 \\ 0 & R\end{pmatrix} \;\middle|\; R \in \mathbb{R}^{(n-r)\times(p-r)},\ \|R\|_{op} \le 1\right\}.$$

Proof Let $B \in \mathcal{T}$. Then, by Lemma 2.1 and the fact that $b_{ii} = 1$ for all $i = 1, \dots, r$, we find that $B = \begin{pmatrix} I_r & 0 \\ 0 & R\end{pmatrix}$ for some $R \in \mathbb{R}^{(n-r)\times(p-r)}$. Now observe that $\max\{1, \|R\|_{op}\} = \left\|\begin{pmatrix} I_r & 0 \\ 0 & R\end{pmatrix}\right\|_{op}$, which implies

$$\|R\|_{op} \le 1 \quad\Longleftrightarrow\quad \left\|\begin{pmatrix} I_r & 0 \\ 0 & R\end{pmatrix}\right\|_{op} \le 1.$$

This shows the desired equality. □

Proposition 2.2 (Convex geometry of $\partial\|\cdot\|_*(\bar{X})$) *Let $\bar{X} \in \mathbb{R}^{n\times p}$ with $r := \operatorname{rank} \bar{X}$ and let $(\bar{U}, \bar{V}) \in O(n) \times O(p)$ such that*

$$\bar{U}\operatorname{diag}(\sigma(\bar{X}))\bar{V}^T = \bar{X}.$$

Then the following hold:

(a) $\partial\|\cdot\|_*(\bar{X}) = \bar{U}\left\{\left(\begin{smallmatrix} I_r & 0 \\ 0 & R\end{smallmatrix}\right) \;\middle|\; R \in \mathbb{R}^{(n-r)\times(p-r)},\ \|R\|_{op} \le 1\right\}\bar{V}^T$;
(b) $\operatorname{ri}(\partial\|\cdot\|_*(\bar{X})) = \bar{U}\left\{\left(\begin{smallmatrix} I_r & 0 \\ 0 & R\end{smallmatrix}\right) \;\middle|\; R \in \mathbb{R}^{(n-r)\times(p-r)},\ \|R\|_{op} < 1\right\}\bar{V}^T$;
(c) $\operatorname{par}(\partial\|\cdot\|_*(\bar{X})) = \bar{U}\left\{\left(\begin{smallmatrix} 0 & 0 \\ 0 & R\end{smallmatrix}\right) \;\middle|\; R \in \mathbb{R}^{(n-r)\times(p-r)}\right\}\bar{V}^T$.

Proof (a) From (8), we find that $\partial\|\cdot\|_*(\bar{X}) = \bar{U}\partial\|\cdot\|_*(\operatorname{diag}(\sigma(\bar{X}))\bar{V}^T$. In turn, by (7), we find that

$$\partial\|\cdot\|_*(\operatorname{diag}(\sigma(\bar{X})) = \left\{B \in \mathbb{R}^{n\times p} \;\middle|\; \left\langle B, \operatorname{diag}(\sigma(\bar{X})\right\rangle = \|\operatorname{diag}(\sigma(\bar{X}))\|_*,\ \|B\|_{op} \le 1\right\}.$$

Now, observe that, by taking adjoints, $\left\langle B, \operatorname{diag}(\sigma(\bar{X})\right\rangle = \langle \operatorname{DIAG}(B), \sigma(\bar{x})\rangle$ and also $\|\operatorname{diag}(\sigma(\bar{X}))\|_* = \|\sigma(\bar{X})\|_1$. Hence

$$\partial\|\cdot\|_*(\operatorname{diag}(\sigma(\bar{X})) = \left\{B \in \mathbb{R}^{n\times p} \;\middle|\; \operatorname{DIAG}(B) \in \{1\}^r \times \mathbb{R}^{n-r},\ \|B\|_{op} \le 1\right\},$$

and thus Lemma 2.4 gives the desired result.
(b) Define the affine map $F : \mathbb{R}^{(n-r)\times(p-r)} \to \mathbb{R}^{n\times n}$ by $F(R) = \bar{U}\left(\begin{smallmatrix} I_r & 0 \\ 0 & R\end{smallmatrix}\right)\bar{V}^T$. Then, in view of (a), we find that $\partial\|\cdot\|_*(\bar{X}) = F(\mathbb{B}_{op})$. Therefore the desired formula follows from [17, Theorem 6.6].
(c) Follows immediately from (a) or (b). □

In the setting of Proposition 2.2, it follows immediately from part (c) that

$$\operatorname{par}(\partial\|\cdot\|_*(\bar{X})) = \operatorname{span}\left\{u_i v_j^T \mid i = r+1, \dots, n,\ j = r+1, \dots, p\right\}, \quad (10)$$

where u_i $(i = 1, \dots, n)$ and v_j $(j = 1, \dots, p)$ are the columns of $\bar{U}$ and $\bar{V}$, respectively.

3 Flatness of the nuclear norm and simultaneous polarizability

As before, we assume (w.l.o.g.) that $n \ge p$. In this section we present our main results on the geometry of the nuclear norm sphere, specifically a characterization of the *flats*[5]. We then leverage this to characterize the uniqueness of certain nuclear norm optimization problems. The next definition is central to this analysis.

Definition 3.1 (Polarizability) Let $X, \hat{X} \in \mathbb{R}^{n\times p}$.

(a) We say that $U \in \mathbb{R}^{n\times p}$ *polarizes*[6] X if $U \in \mathcal{V}_{n,p}$ and $XU^T \in \mathbb{S}^n_+$.

[5] Flats, in the context of Riemannian geometry, are (uncurved) Euclidean submanifolds.
[6] Sometimes this is also called the 'angular' part of the polar decomposition.

(b) We say that X and $\hat{X}$ are *simultaneously polarizable* if there exists a matrix $U \in \mathcal{V}_{n,p}$ that polarizes both X and $\hat{X}$.

A polarization in the sense of Definition 3.1 (a) always exists as the following result shows, which is based on *polar decomposition.* Note that conventionally, for the case of the rectangular polar decomposition, the polarizing matrix U usually appears on the large side of the matrix, see Horn and Johnson [9, Theorem 7.3.1]. In Definition 3.1, we have placed it on the short side, but we observe that by padding, it is possible to conclude the existence of the short-side polarization as well.

Proposition 3.1 (Existence of polarization) *Let $X \in \mathbb{R}^{n\times p}$. Then there exists $U \in \mathcal{V}_{n,p}$ that polarizes X.*

Proof Consider the augmented matrix $[X\ 0] \in \mathbb{R}^{n\times n}$. By polar decomposition, see, e.g., [9, Theorem 7.3.1], there exists $Q \in O(n)$ and $S \in \mathbb{S}^n_+$ such that $[X\ 0] = SQ$. Now, partition $Q = [U\ W]$ according to $[X\ 0]$. Then

$$XU^T = [X\ 0] \cdot \begin{bmatrix} U^T \\ W^T \end{bmatrix} = S \succeq 0,$$

and, by construction, U has orthonormal columns. □

Polarizability can be expressed in terms of the subdifferential of the nuclear norm.

Lemma 3.1 *Let $X \in \mathbb{R}^{n\times p}$ and let $U \in \mathcal{V}_{n,p}$. The following are equivalent:*

(i) $U \in \partial\|\cdot\|_*(X)$;
(ii) $\langle U, X\rangle = \|X\|_*$;
(iii) *U polarizes X, i.e.* $XU^T \in \mathbb{S}^n_+$.

Proof (i)⇒(ii): By the subdifferential representation of $\|\cdot\|_*$ in (7).
(ii)⇒(iii): Observe that $U^TU = I_p$. In particular, $\sigma(U) = [1,\dots,1]^T \in \mathbb{R}^p$. By assumption, we hence have $\langle U, X\rangle = \|X\|_* = \langle \sigma(U), \sigma(X)\rangle$. By (von Neumann's) Theorem 2.1, we thus find $\bar{U} \in O(n), \bar{V} \in O(p)$ such that $X = \bar{U}\mathrm{diag}(\sigma(X))\bar{V}^T$ and $U = \bar{U}\begin{pmatrix} I_p \\ 0 \end{pmatrix}\bar{V}^T$. Consequently

$$XU^T = \bar{U}\begin{pmatrix} \mathrm{diag}(\sigma(\bar{X})) \\ 0 \end{pmatrix}\bar{V}^T\bar{V}\begin{pmatrix} I_p \\ 0 \end{pmatrix}\bar{U}^T = \bar{U}\begin{pmatrix} \mathrm{diag}(\sigma(\bar{X})) & 0 \\ 0 & 0 \end{pmatrix}\bar{U}^T \succeq 0.$$

(iii)⇒(i): We have

$$\mathrm{tr}\,(XU^T) = \|XU^T\|_* = \|X\|_*,$$

where the first identity employs the assumption that $XU^T \in \mathbb{S}^n_+$ combined with (3) and the second one is due to Lemma 2.3. Since $\|U\|_{op} = 1$ (as $U^TU = I_p$), the desired statement follows from Proposition 2.1(b). □

Combining Lemma 3.1 with Proposition 3.1, we find that $\mathcal{V}_{n,p} \cap \partial\|\cdot\|(X) \neq \emptyset$ for any $X \in \mathbb{R}^{n\times p}$.

We now present our first main result which characterizes the existence of (proper) line segments in the nuclear norm sphere.

Theorem 3.1 (Flats in the nuclear norm sphere) *Let $\bar{X}, \hat{X} \in \mathbb{R}^{n\times p}$ and define*

$$X(t) := \bar{X} + t(\hat{X} - \bar{X}) \quad \forall t \in [0,1].$$

Then the following are equivalent:

(i) $\|X(t)\|_* = \|\bar{X}\|_*$ *for all* $t \in [0,1]$.
(ii) $\hat{X}$ *and* $\bar{X}$ *are simultaneously polarizable and* $\|\bar{X}\|_* = \|\hat{X}\|_*$.

Proof Note that there is nothing to prove if $\bar{X} = \hat{X}$. So we assume the contrary from now on.
(i)⇒(ii): By assumption, the convex function $f : \mathbb{R} \to \mathbb{R}$, $f(t) = \|X(t))\|_*$, is constant on $[0,1]$. Hence, by the (subdifferential) chain rule [17, Theorem 23.8], we have

$$\{0\} = \{f'(t)\} = \left\{\left\langle \hat{X} - \bar{X}, Y\right\rangle \mid Y \in \partial\|\cdot\|_*(X(t))\right\} \quad \forall t \in (0,1),$$

i.e.

$$\left\langle \hat{X} - \bar{X}, Y\right\rangle = 0 \quad \forall Y \in \partial\|\cdot\|_*(X(t)),\ t \in (0,1). \tag{11}$$

Now, for any $t \in (0,1)$ and any $Y \in \partial\|\cdot\|_*(X(t))$, we have

$$\left\langle t\hat{X} + (1-t)\bar{X}, Y\right\rangle = \langle X(t), Y\rangle = \|X(t)\|_* = \|\bar{X}\|_*. \tag{12}$$

Multiplying (11) by $-t$ and adding to (12) then yields

$$\left\langle \bar{X}, Y\right\rangle = \|\bar{X}\|_* \quad \forall Y \in \partial\|\cdot\|_*(X(t)),\ t \in (0,1),$$

hence

$$\partial\|\cdot\|_*(X(t)) \subset \partial\|\cdot\|_*(\bar{X}) \quad \forall t \in (0,1). \tag{13}$$

Similarly, multiplying (11) by $(1-t)$ and adding to (12) ultimately yields

$$\partial\|\cdot\|_*(X(t)) \subset \partial\|\cdot\|_*(\hat{X}) \quad \forall t \in (0,1). \tag{14}$$

Combining (13) and (14) we thus find

$$\partial\|\cdot\|_*(X(t)) \subset \partial\|\cdot\|_*(\hat{X}) \cap \partial\|\cdot\|_*(\bar{X}) \quad \forall t \in (0,1). \tag{15}$$

Now, for $t \in (0,1)$, set $X_t := X(t)$. Choose $U_t \in \mathcal{V}_{n,p}$ that polarizes X_t by means of Proposition 3.1. Then, by Lemma 3.1, we have $U_t \in \partial\|\cdot\|_*(X_t)$, and consequently, by (15), we find $U_t \in \partial\|\cdot\|_*(\hat{X}) \cap \partial\|\cdot\|_*(\bar{X})$. Therefore, we find

$$\left\langle \bar{X}, U_t\right\rangle = \|\bar{X}\|_* \text{ and } \left\langle \hat{X}, U_t\right\rangle = \|\hat{X}\|_*.$$

By Lemma 3.1 we thus infer that U_t polarizes both $\bar{X}$ and $\hat{X}$.
(ii) ⇒ (i): Let $U \in \mathbb{R}^{n\times p}$ polarize $\bar{X}$ and $\hat{X}$. Consequently, U polarizes $X(t)$, and hence, by Lemma 3.1, $U \in \partial\|\cdot\|_*(X(t))$ for all $t \in [0,1]$. Therefore

$$\|X(t)\|_* = \operatorname{tr}(X(t)U^T) = t \cdot \operatorname{tr}(\bar{X}U^T) + (1-t)\cdot \operatorname{tr}(\hat{X}U^T) = \|\bar{X}\|_*.$$

Here, the last identity uses that $\operatorname{tr}(\bar{X}U^T) = \|\bar{X}\|_* = \|\hat{X}\|_* = \operatorname{tr}(\hat{X}U^T)$ as U polarizes both $\bar{X}$ and $\hat{X}$ which have the same nuclear norm (by assumption). □

An immediate consequence is the following corollary.

Corollary 3.1 *For $\bar{X}, \hat{X} \in \mathbb{R}^{n\times p}$ and $\mathcal{A} \in \mathcal{L}(\mathbb{R}^{n\times p}, \mathbb{E})$ the following are equivalent:*

(i) $\|\cdot\|_*$ *and* $\mathcal{A}$ *are constant on the line segment* $[\bar{X}, \hat{X}]$.
(ii) $\hat{X} - \bar{X} \in \ker \mathcal{A}$, $\|\bar{X}\|_* = \|\hat{X}\|_*$ *and* $\hat{X}$ *and* $\bar{X}$ *are simultaneously polarizable.*

The previous result, while geometrically elegant, is potentially difficult to evaluate. By working with the singular value decomposition of the base point $\bar{X}$, one can further specify exactly the set of directions which should not be contained in the kernel of the linear operator $\mathcal{A}$. To state this result, we use the following notation for some $r \in \{1, \dots, n\}$:

$$\mathcal{F}^{n,r} := \left\{ \begin{pmatrix} A & 0 \\ 0 & 0 \end{pmatrix} \in \mathbb{S}^n_+ \;\middle|\; A \in \mathbb{S}^r_{++} \right\} \subset \mathbb{S}^n_+.$$

In the sequel, we call a line segment *proper* if it contains more than a single point.

Corollary 3.2 *Let $\bar{X} \in \mathbb{R}^{n\times p}$ and let $r := \operatorname{rank} \bar{X}$. Let there be posed a singular value decomposition $\bar{X} = \bar{U}\operatorname{diag}(\sigma(\bar{X}))\bar{V}^T$ and let $\mathcal{A} \in \mathcal{L}(\mathbb{R}^{n\times p}, \mathbb{E})$. Set*[7]

$$W(\bar{X}) := \left\{ \bar{U}M \begin{pmatrix} I_r & 0 \\ 0 & R \end{pmatrix} \bar{V}^T \;\middle|\; \begin{array}{l} M \in \mathbb{S}^n_+ - \mathcal{F}^{n,r},\ \operatorname{tr}(M) = 0, \\ R \in \mathcal{V}_{n-r,p-r},\ M \begin{pmatrix} I_r & 0 \\ 0 & RR^T \end{pmatrix} = M \end{array} \right\}.$$

The following are equivalent:

(i) $\mathcal{X} := \left\{ X \in \mathbb{R}^{n\times n} \;\middle|\; \mathcal{A}(X) = \mathcal{A}(\bar{X}),\ \|X\|_* = \|\bar{X}\|_* \right\}$ *does not contain a proper line segment including* $\bar{X}$.
(ii) $\ker \mathcal{A} \cap W(\bar{X}) = \{0\}$.

Proof (ii)⇒(i): Assume (i) does not hold, i.e. there is $\hat{X} \neq \bar{X}$ such that $[\bar{X}, \hat{X}] \subset \mathcal{X}$. By Corollary 3.1, we find that $\bar{X}$ and $\hat{X}$ are simultaneously polarizable, i.e. there exists $U \in \mathcal{V}_{n,p}$ such that $\bar{X}U^T \in \mathbb{S}^n_+$ and $\hat{X}U^T \in \mathbb{S}^n_+$. In particular, by Lemma 3.1, $U \in \partial\|\cdot\|_*(\bar{X})$, hence, by Proposition 2.2, $U = \bar{U}\begin{pmatrix} I_r & 0 \\ 0 & R \end{pmatrix}\bar{V}^T$ for some $R \in \mathcal{V}_{n-r,p-r}$. The latter comes from the fact that $U \in \mathcal{V}_{n,p}$. Moreover, since $\hat{X}U^T \in \mathbb{S}^n_+$, we find that

$$\bar{U}^T\hat{X}\bar{V}\begin{pmatrix} I_r & 0 \\ 0 & R^T \end{pmatrix} = \bar{U}^T(\hat{X}U^T)\bar{U} \in \mathbb{S}^n_+. \tag{16}$$

On the other hand, we also have

$$\bar{U}^T\bar{X}\bar{V}\begin{pmatrix} I_r & 0 \\ 0 & R^T \end{pmatrix} = \operatorname{diag}(\sigma(\bar{X})) \in \mathcal{F}^{n,r}.$$

Combining this with (16), we find that

$$M := \bar{U}^T(\hat{X} - \bar{X})\bar{V}\begin{pmatrix} I_r & 0 \\ 0 & R^T \end{pmatrix} \in \mathbb{S}^n_+ - \mathcal{F}^{n,r},$$

[7] See Section 3.1 for a discussion of $W(\bar{X})$.

and, trivially, $M\begin{pmatrix} I_r & 0 \\ 0 & RR^T \end{pmatrix} = M$. Moreover

$$\begin{aligned}
\operatorname{tr}(M) &= \operatorname{tr}\left(\bar{U}^T \hat{X} \bar{V}\begin{pmatrix} I_r & 0 \\ 0 & R^T \end{pmatrix}\right) - \operatorname{tr}\left(\bar{U}^T \bar{X} \bar{V}\begin{pmatrix} I_r & 0 \\ 0 & R^T \end{pmatrix}\right) \\
&= \|\bar{U}^T \hat{X} \bar{V}\begin{pmatrix} I_r & 0 \\ 0 & R^T \end{pmatrix}\|_* - \|\bar{U}^T \bar{X} \bar{V}\begin{pmatrix} I_r & 0 \\ 0 & R^T \end{pmatrix}\|_* \\
&= \|\hat{X}\|_* - \|\bar{X}\|_* \\
&= 0,
\end{aligned}$$

where the second identity uses the positive semidefiniteness of the matrices in question (combined with (3)), and the third one uses orthogonal invariance and Lemma 2.3. Since also $\mathcal{A}(\hat{X}) = \mathcal{A}(\bar{X})$, we consequently have

$$0 \neq \hat{X} - \bar{X} = \bar{U} M \begin{pmatrix} I_r & 0 \\ 0 & R \end{pmatrix} \bar{V}^T \in W(\bar{X}) \cap \ker \mathcal{A}.$$

(i)⇒(ii): Assume (ii) does not hold, i.e. there exists $M \in (\mathbb{S}^n_+ - \mathcal{F}^{n,r}) \setminus \{0\}$ with $\operatorname{tr}(M) = 0$, $R \in \mathcal{V}_{n-r,p-r}$ and $M\begin{pmatrix} I_r & 0 \\ 0 & RR^T \end{pmatrix} = M$ such that $Y := \bar{U} M \begin{pmatrix} I_r & 0 \\ 0 & R \end{pmatrix} \bar{V}^T \in \ker \mathcal{A}$. Define $U := \bar{U}\begin{pmatrix} I_r & 0 \\ 0 & R \end{pmatrix} \bar{V}^T$. Since $\|R\|_{op} = 1$, in view of Proposition 2.2, we have $U \in \partial \| \cdot \|_*(\bar{X}) \cap \mathcal{V}_{n,p}$. Now, for $\varepsilon > 0$ set $X(\varepsilon) := \bar{X} + \varepsilon Y$. Then

$$\begin{aligned}
\bar{U}^T (X(\varepsilon) U^T) \bar{U} &= \bar{U}^T \bar{X} U^T \bar{U} + \varepsilon \bar{U}^T Y U^T \bar{U} \\
&= \operatorname{diag}(\sigma(\bar{X})) + \varepsilon M \begin{pmatrix} I_r & 0 \\ 0 & RR^T \end{pmatrix} \\
&= \operatorname{diag}(\sigma(\bar{X})) + \varepsilon M,
\end{aligned}$$

as $M\begin{pmatrix} I_r & 0 \\ 0 & RR^T \end{pmatrix} = M$ by assumption. Now, recall that $M \in \mathbb{S}^n_+ - \mathcal{F}^{n,r}$, and $\operatorname{diag}(\sigma(\bar{X})) \in \mathcal{F}^{n,r}$, and hence we can find $\hat{\varepsilon} > 0$, sufficiently small, such that $\operatorname{diag}(\sigma(\bar{X})) + \varepsilon M \in \mathbb{S}^n_+$ for all $\varepsilon \in [0, \hat{\varepsilon}]$. Consequently, for $\hat{X} := X(\hat{\varepsilon})$, we have $\hat{X} U^T \in \mathbb{S}^n_+$ and $U \in \mathcal{V}_{n,p}$, i.e. U polarizes $\hat{X}$ (and $\bar{X}$). In addition, we find that

$$\|\hat{X}\|_* = \operatorname{tr}(\hat{X} U^T) = \operatorname{tr}(\bar{U}^T (\hat{X} U^T) \bar{U}) = \operatorname{tr}(\operatorname{diag}(\sigma(\bar{X}))) + \hat{\varepsilon} \cdot \operatorname{tr}(M) = \|\bar{X}\|_*,$$

as $\operatorname{tr}(M) = 0$. Since we have $\mathcal{A}(\hat{X}) = \mathcal{A}(\bar{X})$ as well, Corollary 3.1 now gives the desired conclusion. □

3.1 The set $W(\bar{X})$ in Corollary 3.2

Some comments on the contents of Corollary 3.2 are in order, given its complexity. We note that that the set $W(\bar{X})$ is a cone, owing to the set $\mathbb{S}^n_+ - \mathcal{F}^{n,r}$ being a cone. While the cone is not generally centrally symmetric (meaning there are $A \in W(\bar{X})$ so that $-A \notin W(\bar{X})$), the condition (ii) is equivalently formulated with the image of $W(\bar{X})$ under the reflection $x \mapsto -x$. The symmetrization of $W(\bar{X})$ has the interpretation as the subset of the tangent space at $\bar{X}$ (in $\mathbb{R}^{n \times p}$) in which the nuclear norm changes linearly. The cone $W(\bar{X})$

is *not* generally convex, save for the case that $\bar{X}$ is full rank; in that case, the set $W(\bar{X})$ simplifies to

$$W(\bar{X}) := \left\{\bar{U}\ M[I_p\ 0]^T\ \bar{V}^T \ \middle|\ M \in \mathbb{S}^n,\ \mathrm{tr}\,(M) = 0\right\}, \tag{17}$$

which is a vector space.

In Section 4, the condition $\ker \mathcal{A} \cap W(\bar{X}) = \{0\}$ will be used to characterize the fact that $\bar{X}$ is the unique solution of (2). One way to produce a more user-friendly sufficient condition is to check this intersection condition holds with $W(\bar{X})$ replaced by a more easily described superset. So, we evaluate in the next proposition $\operatorname{span} W(\bar{X})$, which we note, in the case that $\overline{X}$ is full rank, is simply $W(\bar{X})$ as given in (17).

Proposition 3.2 (Span of $\boldsymbol{W(\bar{X})}$) *Let $\bar{X} \in \mathbb{R}^{n\times p}$ and let $r := \operatorname{rank} \bar{X}$. Let there be posed a singular value decomposition $\bar{X} = \bar{U}\operatorname{diag}(\sigma(\bar{X}))\bar{V}^T$ and let $\mathcal{A} \in \mathcal{L}(\mathbb{R}^{n\times p}, \mathbb{E})$. Let $W(\bar{X})$ be as in Corollary 3.2. Then*

$$\operatorname{span} W(\bar{X}) = \left\{ \bar{U} \left(\begin{smallmatrix} A & B \\ C & D \end{smallmatrix}\right) \bar{V}^T \ \middle|\ \begin{array}{l} A \in \mathbb{S}^r, B \in \mathbb{R}^{r\times(p-r)}, \\ C \in \mathbb{R}^{(n-r)\times r}, D \in \mathbb{R}^{(n-r)\times(p-r)} \end{array} \right\}.$$

Proof As observed in (17), there is nothing to show when $p = r$ and so we assume $p > r$. Let W_4 be the right-hand side of the displayed equation. The containment of $\operatorname{span} W(\bar{X}) \subset W_4$ is immediate from the containment $W(\bar{X}) \subset W_4$ and the fact that the latter is a subspace. For the reverse, we argue by construction of a flag $W_1 \subset W_2 \subset W_3 \subset W_4$ each of which we show is in $\operatorname{span} W(\bar{X})$. Set

$$\begin{aligned} W_1 &:= \left\{\bar{U} \left(\begin{smallmatrix} 0 & 0 \\ 0 & D \end{smallmatrix}\right) \bar{V}^T \ \middle|\ D \in \mathbb{R}^{(n-r)\times(p-r)}\right\}, \\ W_2 &:= \left\{\bar{U} \left(\begin{smallmatrix} A & 0 \\ 0 & D \end{smallmatrix}\right) \bar{V}^T \ \middle|\ A \in \mathbb{S}^r, D \in \mathbb{R}^{(n-r)\times(p-r)}\right\}, \\ W_3 &:= \left\{\bar{U} \left(\begin{smallmatrix} A & 0 \\ C & D \end{smallmatrix}\right) \bar{V}^T \ \middle|\ A \in \mathbb{S}^r, C \in \mathbb{R}^{(n-r)\times r}, D \in \mathbb{R}^{(n-r)\times(p-r)}\right\}. \end{aligned} \tag{18}$$

The matrices $\bar{U}$ and $\bar{V}$ play no role in the proof, so we drop them for simplicity.

$W_1 \subset \operatorname{span} W(\bar{X})$: For any element of W_1, let it be given as in (18) for some $D \in \mathbb{R}^{(n-r)\times(p-r)}$, and let $R \in \mathcal{V}_{n-r,p-r}$ be a polarizing matrix such that $P = DR^T \in \mathbb{S}^{n-r}_+$. Now set

$$M := \left(\begin{smallmatrix} aI_r & 0 \\ 0 & P \end{smallmatrix}\right) \in \mathbb{S}^n - \mathcal{F}^{n,r} \quad \text{where} \quad a := -\frac{\mathrm{tr}\,(P)}{r}.$$

Then

$$\tfrac{1}{2}M \left(\begin{smallmatrix} I_r & 0 \\ 0 & R \end{smallmatrix}\right) - \tfrac{1}{2}M \left(\begin{smallmatrix} I_r & 0 \\ 0 & -R \end{smallmatrix}\right) = M \left(\begin{smallmatrix} 0 & 0 \\ 0 & R \end{smallmatrix}\right) = \left(\begin{smallmatrix} 0 & 0 \\ 0 & D \end{smallmatrix}\right). \tag{19}$$

Since $PRR^T = P$, hence $M \begin{pmatrix} I_r & 0 \\ 0 & RR^T \end{pmatrix} = M$, and $\mathrm{tr}\,(M) = 0$, we conclude from (19) that $W_1 \subset \operatorname{span} W(\bar{X})$.

$\underline{W_2 \subset \operatorname{span} W(\bar{X})}$: For arbitrary $A \in \mathbb{S}^r$, let $P_1, P_2 \in \mathbb{S}^r_{++}$ be such that $A = P_1 - P_2$. Now set

$$M_i := \begin{pmatrix} -P_i & 0 & 0 \\ 0 & a_i I_{p-r} & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad \text{where} \quad a_i = \frac{\operatorname{tr}(P_i)}{p-r} \quad (i = 1, 2).$$

Then $\operatorname{tr}(M_i) = 0$ $(i = 1, 2)$, and taking $R := [I_{p-r}\ 0]^T \in \mathcal{V}_{n-r,p-r}$, we have

$$\left(\begin{smallmatrix} A & 0 \\ 0 & * \end{smallmatrix}\right) = -M_1 \left(\begin{smallmatrix} I_r & 0 \\ 0 & R \end{smallmatrix}\right) + M_2 \left(\begin{smallmatrix} I_r & 0 \\ 0 & R \end{smallmatrix}\right) \in \operatorname{span} W(\bar{X}),$$

where the $(*)$ represents an $(n-r) \times (p-r)$ matrix of no importance. Therefore, since $W_1 \subset \operatorname{span} W(\bar{X})$, $W_2 \subset \operatorname{span} W(\bar{X})$.

$\underline{W_3 \subset \operatorname{span} W(\bar{X})}$: Take an arbitary rank 1 matrix $C \in \mathbb{R}^{(n-r)\times r}$. For any $d > 0$, set $a := \frac{d^2}{r} \operatorname{tr}(CC^T) > 0$ and define

$$M := \begin{pmatrix} -aI_r & C^T \\ C & d^2 CC^T \end{pmatrix},$$

which we observe always has trace 0. Then taking $b := a - \frac{1}{d^2}$, we have, for all d sufficiently large, that

$$M = \begin{pmatrix} \frac{1}{d} I_r & 0 \\ dC & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{d} I_r & dC^T \\ 0 & 0 \end{pmatrix} - \left(\begin{smallmatrix} bI_r & 0 \\ 0 & 0 \end{smallmatrix}\right) \in \mathbb{S}^n_+ - \mathcal{F}^{n,r}.$$

Since C is rank 1, we can find an element $R \in \mathcal{V}_{n-r,p-r}$ so that $C^T RR^T = C^T$, for example by taking the first column of R to be in the column space of C. Then $M \begin{pmatrix} I_r & 0 \\ 0 & RR^T \end{pmatrix} = M$, hence

$$\left(\begin{smallmatrix} * & 0 \\ C & 0 \end{smallmatrix}\right) = M \left(\begin{smallmatrix} I_r & 0 \\ 0 & 0 \end{smallmatrix}\right) = \tfrac{1}{2} M \left(\begin{smallmatrix} I_r & 0 \\ 0 & R \end{smallmatrix}\right) + \tfrac{1}{2} M \left(\begin{smallmatrix} I_r & 0 \\ 0 & -R \end{smallmatrix}\right) \in \operatorname{span} W(\bar{X}).$$

Hence since $W_2 \in \operatorname{span} W(\bar{X})$, $\left(\begin{smallmatrix} 0 & 0 \\ C & 0 \end{smallmatrix}\right) \in \operatorname{span} W(\bar{X})$. As there is a basis of rank-1 matrices of $\mathbb{R}^{(n-r)\times r}$ (say those with a single non-zero entry which is 1), we conclude that $W_3 \subset \operatorname{span} W(\bar{X})$.

$\underline{W_4 \subset \operatorname{span} W(\bar{X})}$: Take an arbitary matrix $B \in \mathbb{R}^{r\times(p-r)}$. Then for all c sufficiently large,

$$M = \begin{pmatrix} -\frac{p-r}{r} cI_r & B & 0 \\ B^T & cI_{p-r} & 0 \\ 0 & 0 & 0 \end{pmatrix} \in \{M \in \mathbb{S}^n_+ - \mathcal{F}^{n,r} \mid \operatorname{tr}(M) = 0\}.$$

Taking $R = [I_{p-r}\ 0]^T$, we find

$$M \left(\begin{smallmatrix} I_r & 0 \\ 0 & R \end{smallmatrix}\right) = \begin{pmatrix} -\frac{p-r}{r} cI_r & B \\ B^T & cI_{p-r} \\ 0 & 0 \end{pmatrix} \in W(\bar{X}),$$

and hence we conclude that, since $W_3 \in \operatorname{span} W(\bar{X})$, so is W_4. This concludes the proof. $\square$

4 Unique solutions in nuclear norm minimization

Throughout this section, let $(\mathbb{E}, \langle\cdot,\cdot\rangle)$ be a (finite-dimensional) Euclidean space and let $\mathcal{A} \in \mathcal{L}(\mathbb{R}^{n\times p}, \mathbb{E})$. Our study above immediately yields uniqueness results for the nuclear norm minimization

$$\min_{X\in\mathbb{R}^{n\times p}} \|X\|_* \text{ s.t. } \mathcal{A}(X) = b. \tag{20}$$

Corollary 4.1 *Let $\bar{X} \in \mathbb{R}^{n\times p}$ be a solution of* (20) *with* $\operatorname{rank}\bar{X} = r$*, and let $W(\bar{X})$ be defined as in Corollary 3.2. Then the following are equivalent:*

(i) *$\bar{X}$ is the unique solution of* (20).
(ii) $\ker\mathcal{A} \cap W(\bar{X}) = \{0\}$.

Proof Observe that the solution set $\mathcal{X}$ of (20) is convex and can be written as $\mathcal{X} = \left\{X \in \mathbb{R}^{n\times n} \;\middle|\; \mathcal{A}(X) = \mathcal{A}(\bar{X}),\ \|X\|_* = \|\bar{X}\|_*\right\}$. Now, by convexity, $\mathcal{X}$ does not contain a proper line segment including $\bar{X}$ if and only if $\bar{X}$ is the unique solution of (20). Thus Corollary 3.2 gives the desired statement. □

We further produce a sufficient condition by considering the span of $W(\bar{X})$, which is given in Proposition 3.2.

Corollary 4.2 *Let $\bar{X} \in \mathbb{R}^{n\times p}$ be a solution of* (20) *with* $\operatorname{rank}\bar{X} = r$*, and let $W(\bar{X})$ be defined as in Corollary 3.2. Then if* $\ker\mathcal{A} \cap \operatorname{span} W(\bar{X}) = \{0\}$*, $\bar{X}$ is the unique solution of* (20).

4.1 Sufficient conditions through convex analysis

The following result is a generic convex analysis result which, given a solution, provides a sufficient condition for uniqueness of solutions to a(ny) convex optimization problem. It was established in [6] that in the polyhedral convex case it is also necessary which was then exploited to establish uniqueness of solutions for ℓ_1-minimization problems.

Proposition 4.1 *Let $f : \mathbb{E} \to \mathbb{R}\cup\{+\infty\}$ be proper, convex and assume that $0 \in \operatorname{int}\partial f(\bar{x})$. Then* $\operatorname{argmin} f = \{\bar{x}\}$.

Proof Let $x \in \mathbb{E}$. By assumption, there exists $\varepsilon > 0$ such that $\varepsilon(x-\bar{x}) \in \partial f(\bar{x})$. Consequently

$$f(x) \ge f(\bar{x}) + \langle\varepsilon(x-\bar{x}),\, x-\bar{x}\rangle = f(\bar{x}) + \varepsilon\|x-\bar{x}\|^2 > f(\bar{x}).$$

□

We will, of course, apply this to the objective function $f = \|\cdot\|_* + \delta_{\{0\}}(\mathcal{A}(\cdot)-b)$ of (20). It turns out (in Proposition 4.2) that the following conditions at some (feasible) point $\bar{X}$ are equivalent to having $0 \in \operatorname{int}(\partial f(\bar{X}))$.

Assumption 1 *For $\bar{X} \in \mathbb{R}^{n\times p}$ such that $\mathcal{A}(\bar{X}) = b$ it holds that:*

(i) $\mathrm{ri}\,(\partial\|\cdot\|_*(\bar{X})) \cap \mathrm{rge}\,\mathcal{A}^* \neq \emptyset$*;*
(ii) $\mathrm{par}\,\big(\partial\|\cdot\|_*(\bar{X})\big) + \mathrm{rge}\,\mathcal{A}^* = \mathbb{R}^{n\times p}$*.*

The reader can make these conditions even more tangible by inserting the respective expressions for the relative interior and parallel subspace of the $\partial\|\cdot\|_*(\bar{X})$ provided in Proposition 2.2 (and (10)).

We now provide the advertized characterization.

Proposition 4.2 *Let $\mathcal{A} \in \mathcal{L}(\mathbb{R}^{n\times p}, \mathbb{E})$, $b \in \mathbb{E}$ and define the (closed) proper, convex function $f : \mathbb{R}^{n\times p} \to \mathbb{R} \cup \{+\infty\}$ by $f(X) = \|X\|_* + \delta_{\{0\}}(\mathcal{A}(X) - b)$. For $\bar{X} \in \mathbb{R}^{n\times p}$ such that $\mathcal{A}(\bar{X}) = b$, the following are equivalent:*

(I) $0 \in \mathrm{int}\,\partial f(\bar{X})$.
(II) Assumption 1 holds at $\bar{X}$.

Proof Observe that $\partial(\delta_{\{0\}}((\cdot) - b) \circ \mathcal{A})(\bar{X}) = \mathcal{A}^*\partial\delta_{\{0\}}(0) = \mathcal{A}^*\mathbb{E} = \mathrm{rge}\,\mathcal{A}^*$, by the chain rule [17, Theorem 23.9], and consequently

$$\partial f(\bar{X}) = \partial\|\cdot\|_*(\bar{X}) + \mathrm{rge}\,\mathcal{A}^*,$$

by the sum rule [17, Theorem 23.8]. Thus, we observe that $0 \in \mathrm{int}\,\partial f(\bar{X})$ is equivalent to the following conditions holding simultaneously:

(1) $0 \in \mathrm{ri}\,(\partial\|\cdot\|_*(\bar{X}) + \mathrm{rge}\,\mathcal{A}^*)$;
(2) $\mathrm{aff}\,(\partial\|\cdot\|_*(\bar{X}) + \mathrm{rge}\,\mathcal{A}^*) = \mathbb{R}^{n\times p}$.

Clearly, condition (1) is equivalent to (i). Now, observe that, for any $y \in \partial\|\cdot\|_*(\bar{X}) \cap \mathrm{rge}\,\mathcal{A}^*$, we have

$$\begin{aligned}\mathrm{aff}\,(\partial\|\cdot\|_*(\bar{X}) + \mathrm{rge}\,\mathcal{A}^*) &= \mathrm{aff}\,(\partial\|\cdot\|_*(\bar{X})) + \mathrm{rge}\,\mathcal{A}^* \\ &= \mathrm{aff}\,(\partial\|\cdot\|_*(\bar{X}) - y) + y + \mathrm{rge}\,\mathcal{A}^* \\ &= \mathrm{par}\,\partial\|\cdot\|_*(\bar{X}) + \mathrm{rge}\,\mathcal{A}^*.\end{aligned}$$

This shows that (2) is equivalent to (ii), and therefore concludes the proof. □

Corollary 4.3 *Let $\bar{X}$ be a solution of* (20) *such that Assumption 1 holds at $\bar{X}$. Then $\bar{X}$ is the unique solution of* (20)

Proof Combine Proposition 4.1 and Proposition 4.2. □

4.2 Connecting the convex analytic and geometric conditions

Combining Corollary 4.3 and Corollary 3.2, it follows readily that Assumption 1 at $\bar{X}$ implies that $W(\bar{X}) \cap \ker\mathcal{A} = \{0\}$. On the other hand, this argument is not very illuminating when trying to understand the exact interplay of these two types of conditions. Moreover, it is not clear whether Assumption 1 might also be necessary for uniqueness of solutions (as it is for its ℓ_1-analog). We shed some light on these issues now and start with an auxiliary result.

Lemma 4.1 *Let $\bar{X}$ satisfy Assumption 1. Then*

$$\mathbb{R}^{n\times p} = \operatorname{rge}\mathcal{A}^* + \mathbb{R}_+\partial\|\cdot\|_*(\bar{X}).$$

Proof Set $\mathcal{S} := \partial\|\cdot\|_*(\bar{X})$, and let $Y \in \operatorname{rge}\mathcal{A}^* \cap \operatorname{ri}\mathcal{S}$ which exists by Assumption 1 (i). Then

$$\begin{aligned}
\mathbb{R}^{n\times p} &= \operatorname{rge}\mathcal{A}^* + \operatorname{par}(\partial\|\cdot\|_*(\bar{X}))\\
&= \operatorname{rge}\mathcal{A}^* + \mathbb{R}_+(\mathcal{S} - Y)\\
&= \mathbb{R}_+(\operatorname{rge}\mathcal{A}^* + \mathcal{S} - Y)\\
&= \mathbb{R}_+(\operatorname{rge}\mathcal{A}^* + \mathcal{S})\\
&= \operatorname{rge}\mathcal{A}^* + \mathbb{R}_+\mathcal{S}.
\end{aligned}$$

The second identity uses (9) and that $y \in \operatorname{ri}\mathcal{S}$, and the last to last one uses that $Y \in \operatorname{rge}\mathcal{A}^*$. □

As alluded to above, the following result is clear from our previous analysis. We give an explicit proof in the hopes of consolidating the different flavors of the conditions in Assumption 1 and Corollary 3.2, respectively.

Proposition 4.3 *Let $\bar{X} \in \mathbb{R}^{n\times p}$ with $r := \operatorname{rank}\bar{X}$ and singular value decomposition $\bar{X} = \bar{U}\operatorname{diag}(\sigma(\bar{X}))\bar{V}^T$. If Assumption 1 holds, then $\ker\mathcal{A} \cap W(\bar{X}) = \{0\}$.*

Proof Let $X \in \ker\mathcal{A} \cap W(\bar{X})$. Then, by definition of $W(\bar{X})$, there exists $M = \begin{pmatrix} A-D & B\\ B^T & C\end{pmatrix}$ with $A \in \mathbb{S}^r_+, C \in \mathbb{S}^{n-r}_+$, $D \in \mathbb{S}^r_{++}$, $\operatorname{tr}(A) + \operatorname{tr}(C) = \operatorname{tr}(D)$, and $R \in \mathcal{V}_{n-r,p-r}$ such that

$$X = \bar{U}M\begin{pmatrix} I_r & 0\\ 0 & R\end{pmatrix}\bar{V}^T.$$

On the other hand, by Lemma 4.1 and Proposition 2.2 we find $Z \in \operatorname{rge}\mathcal{A}^*$, $F \in \mathbb{B}_{op}$ and $t \geq 0$ such that

$$X = Z + t\cdot\bar{U}\begin{pmatrix} I_r & 0\\ 0 & F\end{pmatrix}\bar{V}^T.$$

Consequently, we have

$$\begin{aligned}
\|X\|^2 &= \left\langle \bar{U}M\begin{pmatrix} I_r & 0\\ 0 & R\end{pmatrix}\bar{V}^T,\, Z + t\cdot\bar{U}\begin{pmatrix} I_r & 0\\ 0 & F\end{pmatrix}\bar{V}^T\right\rangle\\
&= t\cdot\left\langle \bar{U}M\begin{pmatrix} I_r & 0\\ 0 & R\end{pmatrix}\bar{V}^T,\, \bar{U}\begin{pmatrix} I_r & 0\\ 0 & F\end{pmatrix}\bar{V}^T\right\rangle\\
&= t\cdot\left\langle M\begin{pmatrix} I_r & 0\\ 0 & R\end{pmatrix},\, \begin{pmatrix} I_r & 0\\ 0 & F\end{pmatrix}\right\rangle\\
&= t\cdot\operatorname{tr}\left(\begin{pmatrix} I_r & 0\\ 0 & RF^T\end{pmatrix}\cdot\begin{pmatrix} A-D & B\\ B^T & C\end{pmatrix}\right)\\
&= t\cdot\operatorname{tr}\left(\begin{pmatrix} A-D & B\\ RF^TB^T & RF^TC\end{pmatrix}\right)\\
&= t\cdot\left(\operatorname{tr}(A) - \operatorname{tr}(D) + \operatorname{tr}(RF^TC)\right)\\
&= t\cdot\left(\operatorname{tr}(RF^TC) - \operatorname{tr}(C)\right)\\
&\leq t\cdot\left(\|RF^T\|_{op}\cdot\|C\|_* - \|C\|_*\right)\\
&\leq 0.
\end{aligned}$$

Here, the second identity takes into account that $Z \in \operatorname{rge} \mathcal{A}^*$ while $\bar{U} M \left(\begin{smallmatrix} I_r & 0 \\ 0 & R \end{smallmatrix}\right) \bar{V}^T = X \in \ker \mathcal{A}$. The seventh (last) equality uses the fact that $\operatorname{tr}(A) + \operatorname{tr}(C) = \operatorname{tr}(D)$. The first inequality uses the fact that C is positive semidefinite as well as the 'Hölder inequality' for the operator and nuclear norm. The last inequality is due to the fact that $\|R\|_{op} = 1$, $\|F\|_{op} \le 1$ and the submulitiplicativity of the operator norm.

All in all, we find that $X = 0$ which proves the desired result. □

4.3 Discussion of the uniqueness conditions

We have established two sufficient conditions for uniqueness, Corollary 4.3 and Corollary 4.2, as well as one characterization in Corollary 4.1. We conclude this section by showing, in Example 4.1, that neither of these sufficient conditions imply one another, and we further show that neither is necessary. We summarize our findings in Figure 1 below.

Example 4.1 In this example we consider the nuclear norm minimization problem

$$\min_{X \in \mathbb{R}^{2\times 2}} \|X\|_* \text{ s.t. } \mathcal{A}(X) = b \tag{21}$$

for different choices of $\mathbb{E}$, $\mathcal{A} \in \mathcal{L}(\mathbb{R}^{2\times 2}, \mathbb{E})$ and $b \in \mathbb{E}$. In what follows, $\mathbb{A}^2$ denotes the 2×2 skew symmetric matrices.

(a) Consider (21) with

$$\mathbb{E} := \mathbb{R}^{2\times 2} \times \mathbb{R}^{2\times 2}, \quad b := [\left(\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix}\right), \left(\begin{smallmatrix} 0 & 0 \\ 0 & 0 \end{smallmatrix}\right)], \quad \mathcal{A}(X) = [\left(\begin{smallmatrix} 1 & 1 \\ 0 & 0 \end{smallmatrix}\right) X,\ P_{\mathbb{A}^2}(X)],$$

where $P_{\mathbb{A}^2}(X) := \frac{1}{2}(X - X^T)$ is the projection onto $\mathbb{A}^2$. We find that

$$\ker \mathcal{A} = \mathbb{R}\{\left(\begin{smallmatrix} 1 & -1 \\ -1 & 1 \end{smallmatrix}\right)\} \text{ and } \operatorname{rge} \mathcal{A}^* = \{\left(\begin{smallmatrix} t & s \\ t & s \end{smallmatrix}\right) \mid t, s \in \mathbb{R}\} + \mathbb{A}^2.$$

Now, set $\bar{X} := \left(\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix}\right)$. Then $\mathcal{A}(\bar{X}) = b$, i.e. $\bar{X}$ is feasible for (21). Moreover, by Proposition 2.2, observe that

$$\partial\|\cdot\|_*(\bar{X}) = \left\{\left(\begin{smallmatrix} 1 & 0 \\ 0 & \beta \end{smallmatrix}\right) \mid \beta \in [-1, 1]\right\} \text{ and } \operatorname{par}(\partial\|\cdot\|_*(\bar{X})) = \left\{\left(\begin{smallmatrix} 0 & 0 \\ 0 & \beta \end{smallmatrix}\right) \mid \beta \in \mathbb{R}\right\}.$$

Moreover, we observe that

$$\begin{aligned} 0 \in \partial\|\cdot\|_*(\bar{X}) + \operatorname{rge} \mathcal{A}^* &\iff \partial\|\cdot\|_*(\bar{X}) \cap \operatorname{rge} \mathcal{A}^* \ne \emptyset \\ &\iff \exists \beta \in [-1, 1], t, s, q \in \mathbb{R} : \left(\begin{smallmatrix} 1 & 0 \\ 0 & \beta \end{smallmatrix}\right) = \left(\begin{smallmatrix} t & s \\ t & s \end{smallmatrix}\right) + \left(\begin{smallmatrix} 0 & q \\ -q & 0 \end{smallmatrix}\right). \end{aligned}$$

The latter system has only one solution: $t = q = 1, s = \beta = -1$. In particular, we see that $\bar{X}$ is a minimizer of (21) and that

$$\operatorname{ri}(\partial\|\cdot\|_*(\bar{X})) \cap \operatorname{rge} \mathcal{A}^* = \emptyset.$$

This shows that Assumption 1 (part (i)) is violated at $\bar{X}$. In addition, in view of Proposition 3.2, we find that $\operatorname{span} W(\bar{X}) = \mathbb{R}^{2\times 2}$, and consequently $\operatorname{span} W(\bar{X}) \cap \ker \mathcal{A} \supsetneq \{0\}$.

On the other hand, we have

$$W(\bar X)=\left\{\left(\begin{smallmatrix} a-d & b\\ b & c\end{smallmatrix}\right)\left(\begin{smallmatrix} 1 & 0\\ 0 & R\end{smallmatrix}\right)\;\middle|\; R^2=1, \left(\begin{smallmatrix} a & b\\ b & c\end{smallmatrix}\right)\in\mathbb{S}^2_+, \operatorname{tr}\left(\begin{smallmatrix} a-d & b\\ b & c\end{smallmatrix}\right)=0, d>0\right\},$$

and thus

$$\begin{aligned} X\in W(\bar X)\cap\ker\mathcal{A} &\Longrightarrow X=\left(\begin{smallmatrix} x & -x\\ -x & x\end{smallmatrix}\right)=\left(\begin{smallmatrix} a-d & \pm b\\ b & \pm c\end{smallmatrix}\right), a-d+c=0.\\ &\Longrightarrow X=0.\end{aligned}$$

Therefore, by Corollary 3.2, $\bar X$ is the unique solution of (21).

(b) Consider (21) with

$$\mathbb{E}:=\mathbb{R}^{2\times 2}\times\mathbb{R}^{2\times 2},\quad b:=[\left(\begin{smallmatrix} 1 & 0\\ 1 & 0\end{smallmatrix}\right), \left(\begin{smallmatrix} 0 & 0\\ 0 & 0\end{smallmatrix}\right)]\in\mathbb{E},\quad \mathcal{A}(X)=[\left(\begin{smallmatrix} 1 & 0\\ 1 & 0\end{smallmatrix}\right)X,\ P_{\mathbb{A}^2}(X)].$$

Then

$$\ker\mathcal{A}=\mathbb{R}\{\left(\begin{smallmatrix} 0 & 0\\ 0 & 1\end{smallmatrix}\right)\}\ \text{ and }\ \operatorname{rge}\mathcal{A}^*=\{\left(\begin{smallmatrix} a & b\\ c & 0\end{smallmatrix}\right)\mid a,b,c\in\mathbb{R}\}.$$

Set $\bar X:=\left(\begin{smallmatrix} 1 & 0\\ 0 & 0\end{smallmatrix}\right)$. Then $\bar X$ is feasible. In particular, it is clear that the feasible set of (21) is

$$\left(\begin{smallmatrix} 1 & 0\\ 0 & 0\end{smallmatrix}\right)+\ker\mathcal{A}^*=\{\left(\begin{smallmatrix} 1 & 0\\ 0 & t\end{smallmatrix}\right)\mid t\in\mathbb{R}\},$$

and consequently $\bar X$ solves (21). Moreover, with the subdifferential-related expressions given in (a), we find that

$$\operatorname{ri}\partial\|\cdot\|_*(\bar X)\cap\operatorname{rge}\mathcal{A}^*=\{\bar X\}\neq\emptyset\ \text{ and }\ \operatorname{rge}\mathcal{A}^*+\operatorname{par}\partial\|\cdot\|_*(\bar X)=\mathbb{R}^{2\times 2},$$

i.e. Assumption 1 is satisfied. On the other hand, as in part (a), we find that $\operatorname{span}W(\bar X)=\mathbb{R}^{2\times 2}$, and thus $\operatorname{span}W(\bar X)\cap\ker\mathcal{A}=\mathbb{R}\{\left(\begin{smallmatrix} 0 & 0\\ 0 & 1\end{smallmatrix}\right)\}\supsetneq\{0\}$.

(c) Consider (21) with

$$\mathbb{E}:=\mathbb{R}^{2\times 2},\quad b:=I,\quad \mathcal{A}(X):=P_{\mathbb{S}^2}(X)=\frac{1}{2}(X+X^T).$$

We have

$$\ker\mathcal{A}=\mathbb{A}^2\ \text{ and }\ \operatorname{rge}\mathcal{A}^*=\mathbb{S}^2.$$

In particular, the feasible set is $I+\mathbb{A}^2$, and consequently the (unique) solution is $\bar X=I$ with $\partial\|\cdot\|_*(\bar X)=\{I\}$ and thus

$$\operatorname{ri}\partial\|\cdot\|_*(\bar X)=\{I\}\ \text{ and }\ \operatorname{par}\partial\|\cdot\|_*(\bar X)=\{0\}.$$

Consequently, we have

$$\operatorname{par}\partial\|\cdot\|_*(\bar X)+\operatorname{rge}\mathcal{A}^*=\mathbb{S}^2\subsetneq\mathbb{R}^{2\times 2},$$

hence Assumption 1 (part (ii)) is violated. On the other hand, by Proposition 3.2, we find that $\operatorname{span}W(\bar X)=\mathbb{S}^2$, and, consequently, $\ker\mathcal{A}\cap W(\bar X)=\{0\}$.

◇

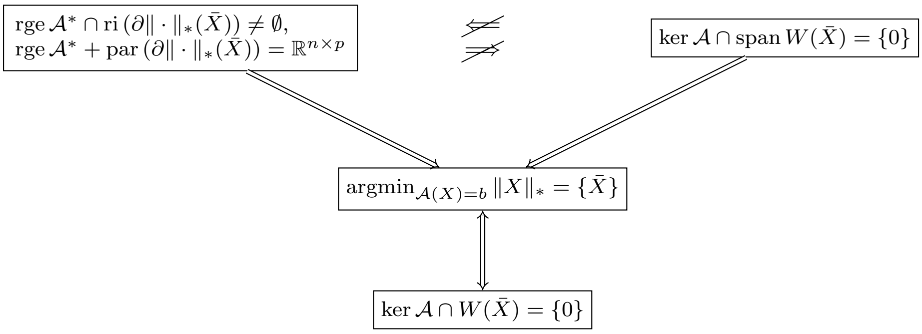

Fig. 1: Uniqueness conditions in nuclear norm minimization

4.4 Other nuclear norm minimization problems

The following general result affords us to carry over uniqueness results from above to other nuclear norm minimization problems involving a linear operator.

Proposition 4.4 *Let $\mathcal{A} \in \mathcal{L}(\mathbb{E}_1, \mathbb{E}_2)$, $g : \mathbb{E} \to \mathbb{R}$ strictly convex, and $h : \mathbb{E}_1 \to \mathbb{R} \cup \{+\infty\}$ closed, proper convex. Then $\mathcal{A}$ and h are constant on the solution set $\mathcal{X}^* := \mathrm{argmin}_{x\in\mathbb{E}_1}\{g(\mathcal{A}(x)) + h(x)\}$.*

Proof Assume there were two solutions $\hat{x}$ and $\bar{x}$ (attaining the optimal value f^*) with $\mathcal{A}(\bar{x}) \neq A(\hat{x})$. Then

$$g\left(\mathcal{A}\left(\frac{\hat{x}+\bar{x}}{2}\right)\right) < \frac{1}{2}g(\mathcal{A}(\hat{x})) + \frac{1}{2}g(\mathcal{A}(\bar{x})) \quad \text{and} \quad h\left(\frac{\hat{x}+\bar{x}}{2}\right) \leq \frac{1}{2}h(\hat{x}) + \frac{1}{2}h(\bar{x}),$$

by strict convexity of g and convexity of h, respectively. Consequently

$$g\left(\mathcal{A}\left(\frac{\hat{x}+\bar{x}}{2}\right)\right) + h\left(\frac{\hat{x}+\bar{x}}{2}\right) < \frac{1}{2}(g(\mathcal{A}(\hat{x})) + h(\hat{x})) + \frac{1}{2}(g(\mathcal{A}(\bar{x})) + h(\bar{x})) = f^*,$$

which yields a contradiction. □

Corollary 4.4 *Let $\mathcal{A} \in \mathcal{L}(\mathbb{R}^{n\times p}, \mathbb{E})$, $b \in \mathbb{E}$, $\lambda > 0$, $f : \mathbb{E} \to \mathbb{R}$ strictly convex, and let $\bar{X}$ be a solution of*

$$\min_{X\in\mathbb{R}^{n\times n}} f(\mathcal{A}(X) - b) + \lambda\|X\|_*. \tag{22}$$

Then $\bar{X}$ is the unique solution if and only if

$$\left\{X \in \mathbb{R}^{n\times p} \;\middle|\; \mathcal{A}(X) = \mathcal{A}(\bar{X}),\ \|X\|_* = \|\bar{X}\|_*\right\} = \{\bar{X}\}.$$

(all of which is the case if and only if $W(\bar{X}) \cap \ker\mathcal{A} = \{0\}$).

Proof Let $\mathcal{X} = \mathrm{argmin}_{\mathbb{R}^{n\times p}}\{f(\mathcal{A}(\cdot) - b) + \lambda\|\cdot\|_*\}$ be the solution set of (22). Applying Proposition 4.4 to $g := f((\cdot) - b)$ and $h := \lambda\|\cdot\|_*$ yields that, in fact, $\mathcal{X} = \left\{X \in \mathbb{R}^{n\times p} \;\middle|\; \mathcal{A}(X) = \mathcal{A}(\bar{X}),\ \|X\|_* = \|\bar{X}\|_*\right\}$. Therefore, the claim follows. □

5 Final remarks

In this paper, starting from a study of line segments in the nuclear norm sphere, we established necessary and sufficient conditions for uniqueness of solutions for minimizing the nuclear norm over an affine subspace. The central linear-algebraic notion in this regard is *simultaneous polarizability*, which formalizes the idea of rotating two (square) matrices in the same fashion to render them positive semidefinite. We then gave another set of sufficient conditions based on the convex geometry of the subdifferential (of the nuclear norm) and its interplay with (the range of) the ambient linear operator. A duality-based argument enabled us to transfer these findings to a whole class of nuclear norm-regularized optimization problems with strictly convex fidelity term.

As a topic of future research, we intend to build on this analysis to study stability of nuclear norm(-regularized) optimization problems in terms of the right-hand side b and the regularization parameter λ. In particular, we would like to study Lipschitz properties of the solution function

$$(b,\lambda) \mapsto \operatorname*{argmin}_{X\in\mathbb{R}^{n\times p}} \left\{ \frac{1}{2}\|\mathcal{A}(X)-b\|^2 + \lambda\|X\|_* \right\}.$$

This study will rely on a suitable representation of the graph of the subdifferential of the nuclear norm.

Acknowledgments The authors would like to thank two anonymous referees for their extremely valuable comments which substantially improved the presentation of the material.

References

1. J.M. Borwein and A.S. Lewis: *Convex Analysis and Nonlinear Optimization. Theory and Examples.* CMS Books in Mathematics, Springer-Verlag, New York, 2000.
2. E.J. Candès, and B. Recht: *Exact matrix completion via convex optimization.* Foundations of Computational Mathematics 9, 2009, pp. 717–772.
3. E.J. Candès, and T. Tao: *Decoding by linear programming.* IEEE Transactions on Information Theory 51(12), 2005, pp. 4203–4215.
4. M. Fazel: *Matrix Rank Minimization with Applications.* Ph.D. thesis, Stanford University, Stanford, CA, 2002.
5. S. Foucart and H. Rauhut: *A Mathematical Introduction to Compressive Sensing.* Birkhäuser, Series on Applied and Numerical Harmonic Analysis, Springer, New York, Heidelberg, Dordrecht London, 2013.
6. J.C. Gilbert: *On the solution uniqueness characterization in the L1 norm and polyhderal gauge recovery.* Journal of Optimization Theory and Applications 172, 2017, pp. 70–101.
7. J.-B. Hiriart-Urruty and H.Y Le: *A variational approach of the rank function.* TOP 21, 2013, pp. 207–240.
8. J.-B. Hiriart-Urrruty and C. Lemaréchal: *Fundamentals of Convex Analysis.* Grundlehren Tex Editions, Springer, Berlin, Heidelberg, 2001.
9. R. Horn and C. R. Johnson: *Matrix Analysis.* Cambridge University Press, Cambridge, 2nd Edition, 2013.
10. A.S. Lewis: *The convex analysis of unitarily invariant matrix functions.* Journal of Convex Analysis 2(1,2), 1995, pp. 173–183.

11. A.S. Lewis: *The convex analysis of hermitian matrices.* SIAM Journal on Optimization 6(1), 1995, pp. 165–177.
12. A.S. Lewis and H.S. Sendov: *Nonsmooth Analysis of Singular Values. Part I: Theory* Set-Valued Analysis 13, 2005, pp. 213–241.
13. L. Mirsky: *A trace inequality of John von Neumann.* Monatshefte für Mathematik 79, 1975, pp. 303–306.
14. J. von Neumann: *Some matrix inequalities and metrization of matric-space.* Tomsk University Review 1, 1937, pp. 286–300. In: Collected Works Vol. IV, Pergamon, Oxford, 1962, pp. 205–218.
15. B. K. Natarajan: *Sparse approximate solutions to linear systems.* SIAM Journal on Computing 24, 1995, pp. 227–234.
16. B. Recht, M. Fazel, and P.A. Parrilo: *Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization.* SIAM Review 52(3), 2010, pp. 471–501.
17. R.T. Rockafellar: *Convex Analysis.* Princeton Mathematical Series, No. 28. Princeton University Press, Princeton, N.J. 1970.
18. R.T. Rockafellar and R.J.-B. Wets: *Variational Analysis.* Grundlehren der Mathematischen Wissenschaften, Vol. 317, Springer-Verlag, Berlin, 1998.
19. E.M. de Sá: *Exposed faces and duality for symmetric and unitarily invariant norms.* Linear Algebra and its Applications 197-198,1994, pp. 429–450.
20. E.M. de Sá: *Faces of the unit ball of a unitarily invariant norm.* Linear Algebra and its Applications 197-198, 1994, pp. 451–493.
21. E. Stiefel: *Richtungsfelder und Fernparallelismus in n-dimensionalen Mannigfaltigkeiten.* Commentarii Mathematici Helvetici 8(4), 1935–1936, pp. 305–353.
22. G.A. Watson: *Characterization of the subdifferential of some matrix norms.* Linear Algebra and its Applications 170, 1992, pp. 33–45.
23. G.A. Watson: *On matrix approximation problems with Ky Fan k norms.* Numerical Algorithms 5, 1993, pp. 263–272.
24. K. Zietak: *On the characterization of the extremal points of the unit sphere of matrices.* Linear Algebra and its Applications 106, 1988, pp. 57–75.
25. K. Zietak: *Subdifferentials, faces, and dual matrices.* Linear Algebra and its Applications 185, 1993, pp. 125–141.
26. H. Zhang, W. Yin, and L. Cheng: *Necessary and sufficient conditions of solution uniqueness in 1-norm minimization.* Journal of Optimization Theory and Applications 164, 2015, pp. 109–122.