NUMERICAL FUNCTIONAL ANALYSIS AND OPTIMIZATION
2021, VOL. 42, NO. 6, 679–695
https://doi.org/10.1080/01630563.2021.1922438

Taylor & Francis
Taylor & Francis Group

OPEN ACCESS Check for updates

Duality Mapping for Schatten Matrix Norms

Shayan Aziznejad and Michael Unser

École polytechnique fédérale de Lausanne, Lausanne, Switzerland

ABSTRACT

In this paper, we fully characterize the duality mapping over the space of matrices that are equipped with Schatten norms. Our approach is based on the analysis of the saturation of the Hölder inequality for Schatten norms. We prove in our main result that, for $p \in (1, \infty)$, the duality mapping over the space of real-valued matrices with Schatten-p norm is a continuous and single-valued function and provide an explicit form for its computation. For the special case $p = 1$, the mapping is set-valued; by adding a rank constraint, we show that it can be reduced to a Borel-measurable single-valued function for which we also provide a closed-form expression.

ARTICLE HISTORY
Received 18 January 2021
Revised 22 April 2021
Accepted 22 April 2021

KEYWORDS
Banach spaces; duality mapping; Hölder inequality; Schatten norm; singular value decomposition

1. Introduction

In linear algebra and matrix analysis, Schatten norms are a family of spectral matrix norms that are defined via the singular-value decomposition [1]. They have appeared in many applications such as image reconstruction [2, 3], image denoising [4], and tensor decomposition [5], to name a few.

Generally, the Schatten-p norm of a matrix is the ℓ_p norm of its singular values [6]. The family contains some well-known matrix norms: The Frobenius and the spectral (operator) norms are special cases in the family, with $p = 2$ and $p = \infty$, respectively. The case $p = 1$ (trace or nuclear norm) is of particular interest for applications as it can be used to recover low-rank matrices [7]. This is the current paradigm in matrix completion, where the goal is to recover an unknown matrix given some of its entries [8]. Prominent examples of applications that can be reduced to low-rank matrix-recovery problems are phase retrieval [9], sensor-array processing [10], system identification [11], and index coding [12, 13].

CONTACT Shayan Aziznejad shayan.aziznejad@epfl.ch École polytechnique fédérale de Lausanne, Lausanne, Switzerland.

© 2021 The Author(s). Published with license by Taylor and Francis Group, LLC
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (http://creativecommons.org/licenses/by-nc-nd/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

In addition to their many applications in data science, Schatten norms have been extensively studied from a theoretical point of view. Various inequalities concerning Schatten norms have been proven [14–22]; sharp bounds for commutators in Schatten spaces have been given [23, 24]; moreover, facial structure [25], Fréchet differentiablity [26], and various other aspects [27, 28] have been studied already.

Our objective in this paper is to investigate the duality mapping in spaces of matrices that are equipped with Schatten norms. The duality mapping is a powerful tool to understand the topological structure of Banach spaces [29, 30]. It has been used to derive powerful characterizations of the solution of variational problems in function spaces [31, 32] and also to determine generalized linear inverse operators [33]. Here, we prove that the duality mapping over Schatten-p spaces with $p \in (1, +\infty)$ is a single-valued and continuous function which, in fact, highlights the strict convexity of these spaces. Although the provided characterization is intuitive, we could not find it in the literature and this is, to the best of our knowledge, the first work which provides a direct way of computing this mapping in this case. For the special case $p = 1$, the mapping is set-valued. However, we prove that, by adding a rank constraint, it reduces to a single-valued Borel-measurable function. In both cases, we also derive closed-form expressions that allow one to compute them explicitly.

The paper is organized as follows: In Section 2, we present relevant mathematical tools and concepts that are used in this paper. We study the duality mapping of Schatten spaces and propose our main result in Section 3. We provide further discussions regarding the introduced mappings in Section 4.

2. Preliminaries

2.1. Dual norms, Hölder inequality, and duality mapping

Let V be a finite-dimensional vector space that is equipped with an inner-product $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$ and let $||\cdot||_X : V \to \mathbb{R}_{\geq 0}$ be an arbitrary norm on V. We then denote by X the space V equipped with $||\cdot||_X$. Clearly, X is a Banach space, because all finite-dimensional normed spaces are complete. The dual norm of X, denoted by $||\cdot||_{X'} : V \to \mathbb{R}_{\geq 0}$, is defined as

$$||\mathbf{v}||_{X'} = \sup_{\mathbf{u} \in V \setminus \{\mathbf{0}\}} \frac{\langle \mathbf{v}, \mathbf{u} \rangle}{||\mathbf{u}||_X}, \tag{1}$$

for any $\mathbf{v} \in V$. Following this definition, one would directly obtain the generic duality bound

$$\langle \mathbf{v}, \mathbf{u} \rangle \leq ||\mathbf{v}||_{X'} ||\mathbf{u}||_X, \tag{2}$$

for any $\mathbf{v}, \mathbf{u} \in V$. Saturation of Inequality (2) is the key concept of dual conjugates that is formulated in the following definition.

Definition 1. Let V be a finite-dimensional vector space and let $(||\cdot||_X, ||\cdot||_{X'})$ be a pair of dual norms that are defined over V. The pair $(\mathbf{u}, \mathbf{v}) \in V \times V$ is said to be a (X, X')-conjugate, if

- $\langle \mathbf{v}, \mathbf{u} \rangle = ||\mathbf{v}||_{X'} ||\mathbf{u}||_X$,
- $||\mathbf{v}||_{X'} = ||\mathbf{u}||_X$.

For any $\mathbf{u} \in V$, the set of all elements $\mathbf{v} \in V$ such that $(\mathbf{u}, \mathbf{v})$ forms an (X, X')-conjugate is denoted by $\mathcal{J}_X(\mathbf{u}) \subseteq V$. We refer to the set-valued mapping $\mathcal{J}_X : V \to 2^V$ as the duality mapping. If, for all $\mathbf{u} \in V$, the set $\mathcal{J}_X(\mathbf{u})$ is a singleton, then we indicate the duality mapping for the X-norm via the single-valued function $\mathrm{J}_X : V \to V$ with $\mathcal{J}_X(\mathbf{u}) = \{\mathrm{J}_X(\mathbf{u})\}$.

It is worth mentioning that, for any $\mathbf{u} \in V$, the set $\mathcal{J}_X(\mathbf{u})$ is nonempty. In fact, the closed ball $B = \{\mathbf{v} \in V : ||\mathbf{v}||_{X'} = ||\mathbf{u}||_X\}$ is compact and, hence, the function $\mathbf{v} \mapsto \langle \mathbf{v}, \mathbf{u} \rangle$ attains its maximum value at some $\mathbf{v}^* \in B$. Now, following Definition 1, one readily verifies that $(\mathbf{u}, \mathbf{v}^*)$ is an (X, X')-conjugate.

We conclude this part by providing a classical and illustrative example. Let $V = \mathbb{R}^n$ for some $n \in \mathbb{N}$. For any $p \in [1, +\infty]$, the ℓ_p-norm of a vector $\mathbf{u} = (u_i) \in \mathbb{R}^n$ is defined as

$$||\mathbf{u}||_p = \begin{cases} \left(\sum_{i=1}^{n} |u_i|^p \right)^{\frac{1}{p}}, & p < +\infty \\ \max_i |u_i|, & p = +\infty. \end{cases} \tag{3}$$

It is widely known that the dual norm of ℓ_p is the ℓ_q-norm, where (p, q) are Hölder conjugates (*i.e.*, $1/p + 1/q = 1$) [34]. This stems from the Hölder inequality which states that

$$\langle \mathbf{v}, \mathbf{u} \rangle \leq ||\mathbf{u}||_p ||\mathbf{v}||_q, \tag{4}$$

for all $\mathbf{u} = (u_i), \mathbf{v} = (v_i) \in \mathbb{R}^n$. In the sequel, we exclude the trivial cases $\mathbf{u} = \mathbf{0}$ and $\mathbf{v} = \mathbf{0}$ to avoid unnecessary complexities in our statements.

When $1 < p < +\infty$, Inequality (4) is saturated if and only if $u_i v_i \geq 0$ for $i = 1, \ldots, n$ and there exists a constant $c > 0$ such that $|\mathbf{u}|^p = c|\mathbf{v}|^q$, where $|\mathbf{u}|^p = (|u_i|^p)$. This ensures that the duality mapping is single-valued and also yields the map

$$\mathrm{J}_p(\mathbf{u}) = \mathrm{sign}(\mathbf{u}) \frac{|\mathbf{u}|^{p-1}}{||\mathbf{u}||_p^{p-2}}. \tag{5}$$

For $p = 1$, one can verify that the equality happens if and only if, for any index $i = 1, \ldots, n$ with $u_i \neq 0$, one has that

$$v_i = \text{sign}(u_i)||\mathbf{v}||_\infty. \tag{6}$$

In other words, the vector $\mathbf{v}$ should attain its extreme values at places where $\mathbf{u}$ has nonzero values, with the sign being determined by the corresponding element in $\mathbf{u}$.

Due to (6), the set $\mathcal{J}_1(\mathbf{u})$ is not necessarily a singleton. However, if we add an additional sparsity constraint, then the mapping becomes single-valued. This leads us to introduce the new notion of *sparse duality mapping* in Definition 2.

Definition 2. Let V be a finite-dimensional vector space and let $s_0 : V \to \mathbb{N}$ be an integer-valued function that acts as a sparsity measure. Assuming a pair $(||\cdot||_X, ||\cdot||_{X'})$ of dual norms over V, we call the pair $(\mathbf{u}, \mathbf{v}) \in V \times V$ a sparse (X, X')-conjugate if

- $(\mathbf{u}, \mathbf{v})$ forms an (X, X')-conjugate pair. In other words, $\mathbf{v} \in \mathcal{J}(\mathbf{u})$.
- The quantity $s_0(\mathbf{v})$ attains its minimal value over the set $\mathcal{J}(\mathbf{u})$.

We denote the set of sparse conjugates of $\mathbf{u}$ by $\mathcal{J}_{X,s_0}(\mathbf{u})$. Whenever $\mathcal{J}_{X,s_0}(\mathbf{u})$ is a singleton for any $\mathbf{u} \in V$, we refer to the single-valued function $\mathrm{J}_{X,s_0} : V \to V$ with $\mathcal{J}_{X,s_0}(\mathbf{u}) = \{\mathrm{J}_{X,s_0}(\mathbf{u})\}$ as the sparse duality mapping.

Following Definition 2, if we use the ℓ_0-norm as the sparsity measure, that is $s_0(\mathbf{u}) = ||\mathbf{u}||_0 = \text{Card}(\{i : u_i \neq 0\})$[1], then we have the sparse duality mapping

$$\mathrm{J}_{1,0} : \mathbb{R}^n \to \mathbb{R}^n : \mathbf{u} = (u_i) \mapsto \mathbf{v} = (v_i) = \mathrm{J}_{1,0}(\mathbf{u}),$$
$$v_i = \begin{cases} \text{sign}(u_i)||\mathbf{u}||_1, & u_i \neq 0 \\ 0, & u_i = 0. \end{cases} \tag{7}$$

Finally, we mention that, for $p = +\infty$, the reduced set $\mathcal{J}_{\infty,0}$ is not single-valued. Indeed, let us define $I_{\max}(\mathbf{u}) = \{i : |u_i| = ||\mathbf{u}||_\infty\} \subseteq \{1, ..., n\}$. We readily deduce from (6) that $\mathbf{v} = (v_1, ..., v_n) \in \mathcal{J}_\infty(\mathbf{u})$ if and only if $v_i = 0$ whenever $i \notin I_{\max}(\mathbf{u})$ and $\text{sign}(v_i) = \text{sign}(u_i)$ for $i \in I_{\max}(\mathbf{u})$ with $\sum_{i \in I_{\max}(\mathbf{u})} |v_i| = ||\mathbf{u}||_\infty$. This shows that $\mathcal{J}_\infty(\mathbf{u})$ is a convex set with $\mathcal{J}_{\infty,0}(\mathbf{u})$ being its extreme points, where $\mathcal{J}_{\infty,0}(\mathbf{u}) = \{u_i \mathbf{e}_i : i \in I_{\max}(\mathbf{u})\}$.

2.2. Schatten p-norm

It is widely known that any matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ can be decomposed as

$$\mathbf{A} = \mathbf{U}\mathbf{S}\mathbf{V}^T, \tag{8}$$

[1]Although this functional does not satisfy the homogeneity property of a norm, it has been widely referred to as the ℓ_0-norm.

where $\mathbf{U} \in \mathbb{R}^{m\times m}$ and $\mathbf{V} \in \mathbb{R}^{n\times n}$ are orthogonal matrices and $\mathbf{S}$ is an m by n rectangular diagonal matrix with nonnegative real entries $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_{\min(m,n)} \geq 0$ sorted in descending order [35]. In the literature, (8) is known as the singular-value decomposition (SVD) and the entries σ_i are the singular values of $\mathbf{A}$. In general, the SVD of a matrix $\mathbf{A}$ is not unique. However, the diagonal matrix $\mathbf{S}$ and, consequently, its entries, are fully determined from $\mathbf{A}$. In other words, the values of σ_i are invariant to a specific choice of decomposition. This is why one can refer to the diagonal entries of $\mathbf{S}$ as the "singular values" of $\mathbf{A}$.

When $\mathbf{A}$ is not full rank, one can obtain a reduced version of (8). Indeed, if we denote the rank of $\mathbf{A}$ by r, then we have that

$$\mathbf{A} = \mathbf{U}_r\mathbf{S}_r\mathbf{V}_r^T, \tag{9}$$

where $\mathbf{U}_r \in \mathbb{R}^{m\times r}$ and $\mathbf{V}_r \in \mathbb{R}^{n\times r}$ are (sub)-orthogonal matrices such that $\mathbf{U}_r^T\mathbf{U}_r = \mathbf{V}_r^T\mathbf{V}_r = \mathbf{I}_r$ and $\mathbf{S}_r = \mathrm{diag}(\boldsymbol{\sigma})$ is a diagonal matrix that contains positive singular values $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_r) \in \mathbb{R}^r$ of $\mathbf{A}$.

For any $p \in [1, +\infty]$, the Schatten-p norm of $\mathbf{A}$ is defined as

$$||\mathbf{A}||_{S_p} = \begin{cases} \left(\sum_{i=1}^{r} \sigma_i^p\right)^{\frac{1}{p}}, & p < +\infty \\ \sigma_1, & p = +\infty. \end{cases} \tag{10}$$

We remark that (10) defines a family of quasi norms for $p \in (0, 1)$. In the extreme case $p = 0$, the Schatten-0 norm actually coincides with the rank of the matrix, *i.e.* $||\mathbf{A}||_{S_0} = \mathrm{rank}(\mathbf{A})$. The Schatten quasi norms have also been studied in the literature from both theoretical and practical point of views (see, [36–39], and references therein).

3. Duality mapping in Schatten spaces

For any $p \in [1, \infty]$, the dual of the Schatten-p norm is the Schatten-q norm, where $q \in [1, \infty]$ is such that $\frac{1}{p} + \frac{1}{q} = 1$ [1]. This is due to the generalized version of Hölder's inequality for Schatten norms, as stated in Proposition 1. While this is a known result (see, for example, [40]), it is also the basis for the present work, which is the reason why we provide a proof in A.

Proposition 1. *For any pair* $(p, q) \in [1, +\infty]^2$ *of Hölder conjugates with* $\frac{1}{p} + \frac{1}{q} = 1$ *and any pair of matrices* $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{m\times n}$, *we have that*

$$\langle \mathbf{A}, \mathbf{B} \rangle = \mathrm{Tr}(\mathbf{A}^T\mathbf{B}) \leq ||\mathbf{A}||_{S_p}||\mathbf{B}||_{S_q}. \tag{11}$$

In Proposition 2, we investigate the case where the Hölder inequality is saturated, in the sense that

$$\mathrm{Tr}(\mathbf{A}^T\mathbf{B}) = ||\mathbf{A}||_{S_p}||\mathbf{B}||_{S_q}. \tag{12}$$

This saturation is central to our work, as it is tightly linked to the notion of duality mapping.

Proposition 2. *Let (p, q) be a pair of Hölder conjugates and let* $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{m\times n}$ *be a pair of nonzero matrices with reduced SVDs of the form*

$$\mathbf{A} = \mathbf{U}_r\mathrm{diag}(\boldsymbol{\sigma})\mathbf{V}_r^T, \quad \mathbf{B} = \tilde{\mathbf{U}}_{\tilde{r}}\mathrm{diag}(\tilde{\boldsymbol{\sigma}})\tilde{\mathbf{V}}\tilde{r}^T. \tag{13}$$

- *If* $p \in (1, \infty)$, *then the Hölder inequality is saturated if and only if we have that*

$$\mathbf{B} = c\mathbf{U}_r\mathrm{diag}(\mathrm{J}_p(\boldsymbol{\sigma}))\mathbf{V}_r^T \tag{14}$$

  or, equivalently,

$$\mathbf{A} = c^{-1}\tilde{\mathbf{U}}_{\tilde{r}}\mathrm{diag}(\mathrm{J}_q(\tilde{\boldsymbol{\sigma}}))\tilde{\mathbf{V}}\tilde{r}^T, \tag{15}$$

  where $c = \frac{||\mathbf{B}||_{S_q}}{||\mathbf{A}||_{S_p}}$ *and* $\mathrm{J}_p(\cdot)$ *and* $\mathrm{J}_q(\cdot)$ *are the duality mappings for the* ℓ_p *and* ℓ_q *norms, respectively (see (5)).*
- *If* $p = 1$, *then a necessary condition for the saturation of the Hölder inequality is that*

$$\mathrm{rank}(\mathbf{A}) \le r_1 \le \mathrm{rank}(\mathbf{B}), \tag{16}$$

  where $r_1 = \mathrm{Card}(\{i : \tilde{\sigma}_i = \tilde{\sigma}_1\})$ *is the multiplicity of the first singular value of* $\mathbf{B}$. *Moreover, if we denote the first* r_1 *singular vectors of* $\mathbf{B}$ *in* (13) *by* $\tilde{\mathbf{U}}_1 \in \mathbb{R}^{m\times r_1}$ *and* $\tilde{\mathbf{V}}_1 \in \mathbb{R}^{n\times r_1}$, *then the Hölder inequality is saturated if and only if there exists a symmetric matrix* $\mathbf{X} \in \mathbb{R}^{r_1\times r_1}$ *such that*

$$\mathbf{A} = \tilde{\mathbf{U}}_1\mathbf{X}\tilde{\mathbf{V}}_1^T. \tag{17}$$

  Finally in the rank-equality case $\mathrm{rank}(\mathbf{A}) = \mathrm{rank}(\mathbf{B})$, *we have saturation if and only if*

$$\mathbf{B} = c\mathbf{U}_r\mathbf{V}_r^T, \tag{18}$$

  where $c = ||\mathbf{B}||_{S_\infty}$ *and the matrices* $\boldsymbol{U}_r$ *and* $\boldsymbol{V}_r$ *are defined in* (13).

Remark 1. Note that even though the reduced SVD is not unique (i.e. there are multiple choices for the sub-orthogonal matrices in (13)), the parametric forms given in Proposition 2 do not depend on a specific decomposition and the results are invariant to any arbitrary choice of these reduced SVDs, primarily due to the "only if" parts of the statements.

The proof of Proposition 2 can be found in B. We observe that, in the case $p \in (1, \infty)$, the saturation of Hölder inequality provides a very tight link between the two matrices: If we know one of them, then the other lies in a one-dimensional ray that is parameterized by the constant $c > 0$. However, in the special case $p = 1$, the identification is not as simple. There again, for a given matrix $\mathbf{B}$, one can fully characterize the set of admissible matrices $\mathbf{A}$. However, for the reverse direction, an additional rank-equality constraint is essential to reduce the set of admissible matrices $\mathbf{B}$ to just one ray.

Inspired from Proposition 2, we now propose our main result in Theorem 1, where we explicitly characterize the duality mapping for the Schatten p-norms. The proof of Theorem 1 can be found in C.

Theorem 1. *Let $p, q \in [1, +\infty]$ be a pair of Hölder conjugates with $\frac{1}{p} + \frac{1}{q} = 1$ and $\mathbf{A} \in \mathbb{R}^{m \times n}$ a matrix whose reduced SVD is specified in (9).*

- *If $1 < p < +\infty$, then the single-valued duality mapping $\mathrm{J}_{S_p} : \mathbb{R}^{m \times n} \to \mathbb{R}^{m \times n}$ is well-defined and can be expressed as*

$$\mathrm{J}_{S_p} : \mathbf{A} = \mathbf{U}_r \mathrm{diag}(\boldsymbol{\sigma}) \mathbf{V}_r^T \mapsto \mathbf{A}^* = \mathbf{U}_r \mathrm{diag}(\mathrm{J}_p(\boldsymbol{\sigma})) \mathbf{V}_r^T. \tag{19}$$

- *If $p = 1$ and if we consider the rank function as the sparsity measure in Definition 2, then the sparse duality mapping $\mathrm{J}_{S_1, \mathrm{rank}} : \mathbb{R}^{m \times n} \to \mathbb{R}^{m \times n}$ is well-defined (singleton) and is given as*

$$\mathrm{J}_{S_1, \mathrm{rank}} : \mathbf{A} = \mathbf{U}_r \mathrm{diag}(\boldsymbol{\sigma}) \mathbf{V}_r^T \mapsto \mathbf{A}^* = ||\boldsymbol{\sigma}||_1 \mathbf{U}_r \mathbf{V}_r^T. \tag{20}$$

- *If $p = +\infty$, then the set-valued mapping $\mathcal{J}_{S_\infty}(\cdot)$ can be described as*

$$\mathcal{J}_{S_\infty}(\mathbf{A}) = \left\{ \sigma_1 \mathbf{U}_1 \mathbf{X} \mathbf{V}_1^T : \mathbf{X} \in \mathbb{R}^{r_1 \times r_1} \text{ is symmetric and } ||\mathbf{X}||_{S_1} = 1 \right\}, \tag{21}$$

where r_1 denotes the multiplicity of the first singular value σ_1 of $\mathbf{A}$ and $\mathbf{U}_1, \mathbf{V}_1$ are singular vectors that correspond to σ_1 in (9). Finally, the set of sparse dual conjugates is the collection of rank-1 elements of $\mathcal{J}_{S_\infty}(\mathbf{A})$ which can be characterized as

$$\mathcal{J}_{S_\infty, \mathrm{rank}}(\mathbf{A}) = \{ \sigma_1 \mathbf{U}_1 \mathbf{p} \mathbf{p}^T \mathbf{V}_1^T : \mathbf{p} \in \mathbb{R}^{r_1}, ||\mathbf{p}||_2 = 1 \}. \tag{22}$$

4. Discussion

Theorem 1 provides an interesting characterization of the duality mapping in three scenarios: The first case is $1 < p < +\infty$ which is the most straightforward one. Theorem 1 tells us that the mapping is single-valued and also gives a formula to compute the dual conjugate $\mathbf{A}^*$ of any matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$.

We use this result to deduce the continuity of the duality mapping as well as the strict convexity of the Schatten space in this case (see Corollary 1). In the second case, with $p = 1$, the mapping is not single-valued. However, there is a unique element in the set of dual conjugates with the minimal rank (that is equal to the rank of $\mathbf{A}$) and, hence, we can construct a single-valued sparse duality mapping. Finally, we showed in the third case, characterized by $p = +\infty$, that neither the set of dual conjugates nor the ones with the minimal rank are unique.

In Corollary 1, we highlight some consequences of Theorem 1 concerning the strict convexity of Schatten spaces and the continuity of the duality mapping.

Corollary 1. *The Banach space of m by n matrices equipped with the Schatten-p norm is strictly convex, if and only if $p \in (1, +\infty)$. In this case, the function $\mathrm{J}_{S_p} : \mathbb{R}^{m\times n} \to \mathbb{R}^{m\times n}$ is continuous.*

Proof. For $p \in (1, +\infty)$, we know from Theorem 1 that the duality mapping J_{S_p} is bijective. Moreover, it is known that all finite-dimensional Banach spaces are reflexive. Now, following [41], we deduce the strict convexity of the space of m by n matrices with Schatten-p norm.

For $p = 1$ and $p = +\infty$, we can readily verify that

$$\left\|\alpha\begin{pmatrix}1 & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & 0\end{pmatrix} + (1-\alpha)\begin{pmatrix}0 & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & 1\end{pmatrix}\right\|_{S_1} = \left\|\begin{pmatrix}\alpha & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & (1-\alpha)\end{pmatrix}\right\|_{S_1} = 1,$$

$$\left\|\alpha\begin{pmatrix}1 & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & 0\end{pmatrix} + (1-\alpha)\begin{pmatrix}1 & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & 1\end{pmatrix}\right\|_{S_\infty} = \left\|\begin{pmatrix}1 & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & (1-\alpha)\end{pmatrix}\right\|_{S_\infty} = 1,$$

for all $\alpha \in (0, 1)$, which shows that the Schatten space is not strictly convex for $p = 1, +\infty$.

Finally, the Schatten-p norm is known to be Fréchet differentiable for $p \in (1, +\infty)$ [26]. Moreover, the duality mapping of any Banach space with Fréchet-differentiable norms is guaranteed to be continuous [42, 43]. Combining the two statements, we deduce the continuity of the duality mapping in this case. □

By contrast, the sparse duality mapping $\mathrm{J}_{S_1,\mathrm{rank}}(\cdot)$ is not continuous. This is best explained by providing a counterexample. Specifically, let us consider the sequence of 2 by 2 matrices

$$\mathbf{S}_k = \begin{pmatrix}1 & 0\\ 0 & \frac{1}{k}\end{pmatrix}, \quad k \in \mathbb{N}.$$

It is clear that $\mathbf{S}_k \to \mathbf{S}_\infty = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$. However, we have that

$$\forall k \in \mathbb{N} : \mathrm{J}_{S_1,\mathrm{rank}}(\mathbf{S}_k) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \text{while} \quad \mathrm{J}_{S_1,\mathrm{rank}}(\mathbf{S}_\infty) = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix},$$

which shows the discontinuity of $\mathrm{J}_{S_1,\mathrm{rank}}$ in the space of 2 by 2 matrices. This can be generalized to space of matrices with arbitrary dimensions $m, n \in \mathbb{N}$.

Although $\mathrm{J}_{S_1,\mathrm{rank}}$ is not continuous, we now show that it is Borel-measurable and, hence, that it can be approximated with arbitrary precision by a continuous mapping due to Lusin's theorem [34].

Proposition 3. *For any $m, n \in \mathbb{N}$, the sparse duality mapping $\mathrm{J}_{S_1,\mathrm{rank}}$ is a Borel-measurable matrix-valued function over the space of m by n matrices.*

Before going into the proof of Proposition 3, we present a preliminary result.

Lemma 1. *The set $\mathcal{R}_r \subseteq \mathbb{R}^{m\times n}$ of m by n matrices of rank r is Borel-measurable.*

Proof. First note that

$$\mathcal{R}_1 = \{\mathbf{u}\mathbf{v}^T : \mathbf{u} \in \mathbb{R}^m, \mathbf{v} \in \mathbb{R}^n\}.$$

The set $\mathcal{R}_1$ is the image of the continuous mapping $\mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}^{m\times n} : (\mathbf{u}, \mathbf{v}) \mapsto \mathbf{u}\mathbf{v}^T$ and, hence, is Borel-measurable.

Now, denote by $\mathcal{R}_{\le r} \subseteq \mathbb{R}^{m\times n}$, the set of matrices with rank no more than r. Using the identity

$$\mathcal{R}_{\le r} = \mathcal{R}_1 + \cdots + \mathcal{R}_1, \quad (\text{r times}),$$

we deduce that $\mathcal{R}_{\le r}$ and, consequently, $\mathcal{R}_r = \mathcal{R}_{\le r} \backslash \mathcal{R}_{\le (r-1)}$ are also Borel-measurable sets. □

Proof of Proposition 3. Consider a Borel-measurable set $\mathcal{B} \subseteq \mathbb{R}^{m\times n}$. We show that $\mathcal{B}_{\mathrm{inv}} = \mathrm{J}^{-1}_{S_1,\mathrm{rank}}(\mathcal{B})$ is also Borel-measurable. By defining $\mathcal{B}_{\mathrm{inv},r} = \mathcal{B}_{\mathrm{inv}} \cap \mathcal{R}_r$, we can partition $\mathcal{B}_{\mathrm{inv}}$ as

$$\mathcal{B}_{\mathrm{inv}} = \bigcup_{r=1}^{\min(m,n)} \mathcal{B}_{\mathrm{inv},r}.$$

Hence, it is sufficient to show that each partition $\mathcal{B}_{\mathrm{inv},r}$ is Borel-measurable.

Define the set $\mathcal{P}_r \subseteq \mathcal{R}_r^2$ as

$$\mathcal{P}_r = \{(\mathbf{A}, \mathbf{B}) \in \mathcal{R}_r \times \mathcal{B} : \mathrm{Tr}(\mathbf{A}^T\mathbf{B}) = ||\mathbf{A}||_{S_1}||\mathbf{B}||_{S_\infty}, \quad ||\mathbf{A}||_{S_1} = ||\mathbf{B}||_{S_\infty}\}.$$

The set $\mathcal{P}_r$ introduces a relation over $\mathcal{R}_r$ whose domain is $\mathcal{B}_{\text{inv},r}$. In other words, we have that

$$\mathcal{B}_{\text{inv},r} = \{\mathbf{A} \in \mathcal{R}_r : \exists \mathbf{B} \in \mathcal{B}, (\mathbf{A}, \mathbf{B}) \in \mathcal{P}_r\}.$$

Since the trace and norm are continuous (and, consequently, Borel-measurable) functions and $\mathcal{R}_r \times \mathcal{B}$ is a Borel-measurable set (using Lemma 1), we deduce that the relation induced from $\mathcal{P}_r$ is Borel-measurable as well. Finally, we use [44, Proposition 2.1] to show that its domain is Borel-measurable. □

5. Conclusion

In this paper, we studied the duality mapping in finite-dimensional Schatten spaces. Based on a careful investigation of the cases where the Hölder inequality saturates, we provided an explicit form for this mapping when $p \in (1, +\infty)$. Furthermore, by adding a rank constraint, we proved that the mapping becomes single-valued for the special case $p = 1$. As for $p = +\infty$, we showed that the mapping yields a convex set whose elements are explicitly characterized. Finally, we discussed our theorem and studied the continuity of the introduced mappings as well as the strict convexity of the Schatten spaces. A possible future direction of research is to extend the results of this paper to infinite-dimensional Schatten spaces and even, in full generality, to linear operators over Hilbert spaces.

Funding

This work was supported in part by the European Research Council (H2020-ERC Project GlobalBioIm) under Grant 692726 and in part by the Swiss National Science Foundation, Grant 200020_184646/1.

ORCID

Shayan Aziznejad http://orcid.org/0000-0002-6871-9703

References

[1] Bhatia, R. (1997). *Matrix Analysis*, Vol. 169. New York: Springer-Verlag.

[2] Lefkimmiatis, S., Unser, M. (2013). Poisson image reconstruction with Hessian Schatten-norm regularization. *IEEE Trans. Image Process.* 22(11):4314–4327. DOI: 10.1109/TIP.2013.2271852.

[3] Lefkimmiatis, S., Ward, J., Unser, M. (2013). Hessian Schatten-norm regularization for linear inverse problems. *IEEE Trans. Image Process.* 22(5):1873–1888. DOI: 10.1109/TIP.2013.2237919.

[4] Xie, Y., Gu, S., Liu, Y., Zuo, W., Zhang, W., Zhang, L. (2016). Weighted Schatten *p*-norm minimization for image denoising and background subtraction. *IEEE Trans. Image Process*. 25(10):4842–4857. DOI: 10.1109/TIP.2016.2599290.
[5] Gao, S., Fan, Q. (2020). Robust Schatten-*p* norm based approach for tensor completion. *J. Sci. Comput*. 82(1):1–23.
[6] Horn, R. A., Johnson, C. R. (2012). *Matrix Analysis*. Cambridge University Press.
[7] Davenport, M. A., Romberg, J. (2016). An overview of low-rank matrix recovery from incomplete observations. *IEEE J. Sel. Top. Signal Process*. 10(4):608–622. DOI: 10.1109/JSTSP.2016.2539100.
[8] Candès, E. J., Recht, B. (2009). Exact matrix completion via convex optimization. *Found. Comput. Math*. 9(6):717–772. DOI: 10.1007/s10208-009-9045-5.
[9] Candès, E. J., Eldar, Y. C., Strohmer, T., Voroninski, V. (2015). Phase retrieval via matrix completion. *SIAM Rev*. 57(2):225–251. DOI: 10.1137/151005099.
[10] Davies, M. E., Eldar, Y. C. (2012). Rank awareness in joint sparse recovery. *IEEE Trans. Inform. Theory* 58(2):1135–1146. DOI: 10.1109/TIT.2011.2173722.
[11] Fazel, M., Pong, T. K., Sun, D., Tseng, P. (2013). Hankel matrix rank minimization with applications to system identification and realization. *SIAM J. Matrix Anal. Appl*. 34(3):946–977. DOI: 10.1137/110853996.
[12] Asadi, E., Aziznejad, S., Amerimehr, M. H., Amini, A. (2017). A fast matrix completion method for index coding. In: *Proceedings of the Twenty-Fifth European Signal Processing Conference (EUSIPCO'17)*. Kos Island, Greece: IEEE, pp. 2606–2610.
[13] Esfahanizadeh, H., Lahouti, F., Hassibi, B. (2014). A matrix completion approach to linear index coding problem. In: *Proceedings of the Information Theory Workshop (ITW 2014)*. Hobart, Australia: IEEE, pp. 531–535.
[14] Kittaneh, F. (1985). Inequalities for the Schatten *p*-norm. *Glasgow Math. J*. 26(2): 141–143. DOI: 10.1017/S0017089500005905.
[15] Kittaneh, F. (1987). Inequalities for the Schatten *p*-norm II. *Glasgow Math. J*. 29(1): 99–104. DOI: 10.1017/S0017089500006716.
[16] Kittaneh, F. (1986). Inequalities for the Schatten *p*-norm III. *Communmath. Phys*. 104(2):307–310. DOI: 10.1007/BF01211597.
[17] Kittaneh, F. (1986). Inequalities for the Schatten *p*-norm IV. *Communmath. Phys*. 106(4):581–585. DOI: 10.1007/BF01463397.
[18] Kittaneh, F., Kosaki, H. (1987). Inequalities for the Schatten *p*-norm V. *Publ. Res. Inst. Math. Sci*. 23(2):433–443. DOI: 10.2977/prims/1195176547.
[19] Bourin, J.-C. (2006). Matrix versions of some classical inequalities. *Linear Algebra Appl*. 416(2–3):890–907. DOI: 10.1016/j.laa.2006.01.002.
[20] Hirzallah, O., Kittaneh, F., Moslehian, M. (2010). Schatten *p*-norm inequalities related to a characterization of inner product spaces. *Math. Inequal. Appl*. 13(2): 235–241. DOI: 10.7153/mia-13-19.
[21] Moslehian, M. S., Tominaga, M., Saito, K.-S. (2011). Schatten *p*-norm inequalities related to an extended operator parallelogram law. *Linear Algebra Appl*. 435(4): 823–829. DOI: 10.1016/j.laa.2011.01.046.
[22] Conde, C., Moslehian, M. S. (2016). Norm inequalities related to *p*-Schatten class. *Linear Algebra Appl*. 498:441–449. DOI: 10.1016/j.laa.2015.11.031.
[23] Wenzel, D., Audenaert, K. M. (2010). Impressions of convexity: An illustration for commutator bounds. *Linear Algebra Appl*. 433(11–12):1726–1759. DOI: 10.1016/j.laa.2010.06.039.
[24] Cheng, C.-M., Lei, C. (2015). On Schatten *p*-norms of commutators. *Linear Algebra Appl*. 484:409–434. DOI: 10.1016/j.laa.2015.07.009.

[25] So, W. (1990). Facial structures of Schatten p-norms. *Linear Multilinear Algebra*. 27(3):207–212. DOI: 10.1080/03081089008818012.
[26] Potapov, D., Sukochev, F. (2014). Fréchet differentiability of S_p norms. *Adv. Math*. 262:436–475.
[27] Kittaneh, F. (1989). On the continuity of the absolute value map in the Schatten classes. *Linear Algebra Appl*. 118:61–68. DOI: 10.1016/0024-3795(89)90571-5.
[28] Bhatia, R., Kittaneh, F. (2000). Cartesian decompositions and Schatten norms. *Linear Algebra Appl*. 318(1-3):109–116. DOI: 10.1016/S0024-3795(00)00206-8.
[29] Beurling, A., Livingston, A. (1962). A theorem on duality mappings in Banach spaces. *Ark. Mat*. 4(5):405–411. DOI: 10.1007/BF02591622.
[30] Cioranescu, I. (1990). *Geometry of Banach Spaces. Duality Mappings and Nonlinear Problems*, Vol. 62, Netherlands: Springer.
[31] de Boor, C. (1976). On "best" interpolation. *J. Approximation Theory*. 16(1):28–42. DOI: 10.1016/0021-9045(76)90093-9.
[32] Unser, M. (2020). *A Unifying Representer Theorem for Inverse Problems and Machine Learning*. Foundations of Computational Mathematics.
[33] Liu, P., Wang, Y.-W. (2007). The best generalized inverse of the linear operator in normed linear space. *Linear Algebra Appl*. 420(1):9–19. DOI: 10.1016/j.laa.2006.04.024.
[34] Rudin, W. (1991). *Functional Analysis. International Series in Pure and Applied Mathematics*. New York: McGraw-Hill, Inc.
[35] Johnson, C. R., Horn, R. A. (1985). *Matrix Analysis*. Cambridge University Press,
[36] Nie, F., Huang, H., Ding, C. (2012). Low-rank matrix recovery via efficient Schatten p-norm minimization. In: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 26.
[37] Shang, F., Liu, Y., Cheng, J. (2016). Scalable algorithms for tractable Schatten quasi-norm minimization. In: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 30.
[38] Shang, F., Liu, Y., Shang, F., Liu, H., Kong, L., Jiao, L. (2020). A unified scalable equivalent formulation for schatten quasi-norms. *Mathematics*. 8(8):1325. DOI: 10.3390/math8081325.
[39] Giampouras, P., Vidal, R., Rontogiannis, A., Haeffele, B. (2020). A novel variational form of the Schatten-p quasi-norm. arXiv preprint arXiv:2010.13927.
[40] Lefkimmiatis, S., Roussos, A., Maragos, P., Unser, M. (2015). Structure tensor total variation. *SIAM J. Imaging Sci*. 8(2):1090–1122. DOI: 10.1137/14098154X.
[41] Petryshyn, W. (1970). A characterization of strict convexity of Banach spaces and other uses of duality mappings. *J. Funct. Anal*. 6(2):282–291. DOI: 10.1016/0022-1236(70)90061-3.
[42] Giles, J., Gregory, D., Sims, B. (1978). Geometrical implications of upper semi-continuity of the duality mapping on a Banach space. *Pacific J. Math*. 79(1):99–109. DOI: 10.2140/pjm.1978.79.99.
[43] Contreras, M. D., Payá, R. (1994). On upper semicontinuity of duality mappings. *Proc. Am. Math. Soc*. 121(2):451–459. DOI: 10.1090/S0002-9939-1994-1215199-4.
[44] Himmelberg, C. J., Parthasarathy, T. (1975). Measurable relations. *Fund. Math*. 87(1):53–72. DOI: 10.4064/fm-87-1-53-72.
[45] Cvetkovski, Z. (2012). *Inequalities: Theorems, Techniques and Selected Problems*. Berlin Heidelberg: Springer-Verlag.

A. Proof of Proposition 1

Proof. Let us recall the reduced SVD of the matrix $\mathbf{A}$ as

$$\mathbf{A} = \mathbf{U}_r \mathbf{S}_r \mathbf{V}_r^T, \tag{23}$$

where $r = \text{rank}(\mathbf{A}), \mathbf{U}_r = [\mathbf{u}_1 \cdots \mathbf{u}_r] \in \mathbb{R}^{m\times r}, \mathbf{V}_r = [\mathbf{v}_1 \cdots \mathbf{v}_r] \in \mathbb{R}^{n\times r}$, and $\mathbf{S} = \text{diag}(\sigma_1, ..., \sigma_r)$. Similarly, for the matrix $\mathbf{B}$, we have that

$$\mathbf{B} = \tilde{\mathbf{U}}_{\tilde{r}} \tilde{\mathbf{S}}_{\tilde{r}} \tilde{\mathbf{V}} \tilde{r}^T, \tag{24}$$

where $\tilde{r} = \text{rank}(\mathbf{A}), \tilde{\mathbf{U}}_{\tilde{r}} = [\tilde{\mathbf{u}}_1 \cdots \tilde{\mathbf{u}}_{\tilde{r}}] \in \mathbb{R}^{m\times \tilde{r}}, \tilde{\mathbf{V}}_{\tilde{r}} = [\tilde{\mathbf{v}}_1 \cdots \tilde{\mathbf{v}}_r] \in \mathbb{R}^{n\times \tilde{r}}$, and $\tilde{\mathbf{S}} = \text{diag}(\tilde{\sigma}_1, ..., \tilde{\sigma}_{\tilde{r}})$. A direct computation then reveals that

$$\text{Tr}(\mathbf{A}^T\mathbf{B}) = \sum_{i=1}^{r} \sum_{j=1\tilde{r}} \sigma_i \tilde{\sigma}_j \mathbf{u}_i^T \tilde{\mathbf{u}}_j \mathbf{v}_i^T \tilde{\mathbf{v}}_j. \tag{25}$$

By using the weighted Hölder inequality for vectors [45], we obtain for $p \neq 1$ that

$$\sum_{i=1}^{r} \sum_{j=1\tilde{r}} \sigma_i \tilde{\sigma}_j \mathbf{u}_i^T \tilde{\mathbf{u}}_j \mathbf{v}_i^T \tilde{\mathbf{v}}_j \leq \left(\sum_{i=1}^{r} \sigma_i^p \sum_{j=1\tilde{r}} |\mathbf{u}_i^T \tilde{\mathbf{u}}_j \mathbf{v}_i^T \tilde{\mathbf{v}}_j| \right)^{\frac{1}{p}} \left(\sum_{j=1\tilde{r}} \sigma_j^p \sum_{i=1}^{r} |\mathbf{u}_i^T \tilde{\mathbf{u}}_j \mathbf{v}_i^T \tilde{\mathbf{v}}_j| \right)^{\frac{1}{q}} \tag{26}$$

and for $p = 1$ that

$$\sum_{i=1}^{r} \sum_{j=1\tilde{r}} \sigma_i \tilde{\sigma}_j \mathbf{u}_i^T \tilde{\mathbf{u}}_j \mathbf{v}_i^T \tilde{\mathbf{v}}_j \leq \left(\sum_{i=1}^{r} \sigma_i \sum_{j=1\tilde{r}} |\mathbf{u}_i^T \tilde{\mathbf{u}}_j \mathbf{v}_i^T \tilde{\mathbf{v}}_j| \right) ||\tilde{\boldsymbol{\sigma}}||_\infty. \tag{27}$$

Finally, by invoking Cauchy-Schwartz and the orthonormality of the matrices $\mathbf{U}_r, \mathbf{V}_r, \tilde{\mathbf{U}}_{\tilde{r}}, \tilde{\mathbf{V}}_{\tilde{r}}$, we deduce for $i = 1, ..., r$ that

$$\sum_{j=1\tilde{r}} |\mathbf{u}_i^T \tilde{\mathbf{u}}_j \mathbf{v}_i^T \tilde{\mathbf{v}}_j| \leq \left(\sum_{j=1\tilde{r}} (\mathbf{u}_i^T \tilde{\mathbf{u}}_j)^2 \right)^{\frac{1}{2}} \left(\sum_{j=1\tilde{r}} (\mathbf{v}_i^T \tilde{\mathbf{v}}_j)^2 \right)^{\frac{1}{2}} \leq ||\mathbf{u}_i||_2 ||\mathbf{v}_i||_2 = 1, \tag{28}$$

For $j = 1, ..., \tilde{r}$, we deduce that

$$\sum_{i=1}^{r} |\mathbf{u}_i^T \tilde{\mathbf{u}}_j \mathbf{v}_i^T \tilde{\mathbf{v}}_j| \leq \left(\sum_{i=1}^{r} (\mathbf{u}_i^T \tilde{\mathbf{u}}_j)^2 \right)^{\frac{1}{2}} \left(\sum_{i=1}^{r} (\mathbf{v}_i^T \tilde{\mathbf{v}}_j)^2 \right)^{\frac{1}{2}} \leq ||\tilde{\mathbf{u}}_i||_2 ||\tilde{\mathbf{v}}_i||_2 = 1. \tag{29}$$

The combination of these inequalities completes the proof. □

B. Proof of Proposition 2

Proof. We separate the two cases and analyze each one independently.

Case 1: $1<p<+\infty$. We prove (14) and deduce (15) by symmetry. Following the proof of Proposition 2 and considering the reduced SVD of the matrices $\mathbf{A}$ and $\mathbf{B}$ given in (23) and (24), we immediately see that the inequalities (26), (28), and (29) should all be saturated. The equality condition of the weighted Hölder implies the existence of a positive constant $\alpha>0$ such that, for all $(i,j) \in \{1, ..., r\} \times \{1, ..., \tilde{r}\}$, we have one of the following

conditions:

$$\mathbf{u}_i^T \tilde{\mathbf{u}}_j \mathbf{v}_i^T \tilde{\mathbf{v}}_j = 0, \quad \text{or} \tag{30}$$

$$\mathbf{u}_i^T \tilde{\mathbf{u}}_j \mathbf{v}_i^T \tilde{\mathbf{v}}_j > 0 \quad \text{and} \quad \tilde{\sigma}_j^q = \alpha \sigma_i^p. \tag{31}$$

Moreover, the saturation of (28) implies that

$$\mathbf{u}_i \in \text{Range}(\tilde{\mathbf{U}}_{\tilde{r}}), \quad \mathbf{v}_i \in \text{Range}(\tilde{\mathbf{V}}_{\tilde{r}}) \qquad \forall i = 1, \ldots, r \tag{32}$$

and also that there exists a positive constant $\beta_i > 0$ (positivity follows from (31) and (30)) such that

$$\mathbf{u}_i^T \tilde{\mathbf{u}}_j = \beta_i \mathbf{v}_i^T \tilde{\mathbf{v}}_j, \quad \forall j = 1, \ldots, \tilde{r}. \tag{33}$$

However, from the normality of $\mathbf{u}_i$ and (32), we have that

$$1 = ||\mathbf{u}_i||_2^2 = \sum_{j=1\tilde{r}} |\mathbf{u}_i^T \tilde{\mathbf{u}}_j|^2 = \beta_i^2 \sum_{j=1\tilde{r}} |\mathbf{v}_i^T \tilde{\mathbf{v}}_j|^2 = \beta_i^2 ||\mathbf{v}||_2^2 = \beta_i^2 \tag{34}$$

which, together with the positivity of β_i, leads to the conclusion that $\beta_i = 1$ for $i = 1, \ldots, r$. Using this, we rewrite (33) in matrix form as

$$\mathbf{U}_r^T \tilde{\mathbf{U}}_{\tilde{r}} = \mathbf{V}_r^T \tilde{\mathbf{V}}_{\tilde{r}}. \tag{35}$$

Similarly, the saturation of (29) implies that

$$\tilde{\mathbf{u}}_j \in \text{Range}(\mathbf{U}_r), \quad \tilde{\mathbf{v}}_i \in \text{Range}(\mathbf{V}_r), \tag{36}$$

for all $j = 1, \ldots, \tilde{r}$. Putting together (32) and (36), we deduce that $r = \tilde{r}$ and

$$\text{Range}(\mathbf{U}_r) = \text{Range}(\tilde{\mathbf{U}}_{\tilde{r}}), \quad \text{Range}(\mathbf{V}_r) = \text{Range}(\tilde{\mathbf{V}}_{\tilde{r}}). \tag{37}$$

This implies the existence of two orthogonal matrices $\mathbf{P}, \mathbf{Q} \in \mathbb{R}^{r \times r}$ such that

$$\tilde{\mathbf{U}}_{\tilde{r}} = \mathbf{U}_r \mathbf{P}, \quad \tilde{\mathbf{V}}_{\tilde{r}} = \mathbf{V}_r \mathbf{Q}. \tag{38}$$

However, replacing (38) in (35), we conclude that

$$\mathbf{P} = \mathbf{U}_r^T \mathbf{U}_r \mathbf{P} = \mathbf{U}_r^T \tilde{\mathbf{U}}_{\tilde{r}} = \mathbf{V}_r^T \tilde{\mathbf{V}}_{\tilde{r}} = \mathbf{V}_r^T \mathbf{V}_r \mathbf{Q} = \mathbf{Q}. \tag{39}$$

This implies that the matrix $\mathbf{B}$ can be represented as

$$\mathbf{B} = \mathbf{U}_r \mathbf{P} \tilde{\mathbf{S}}_{\tilde{r}} \mathbf{P}^T \mathbf{V}_r^T = \mathbf{U}_r \mathbf{S}_0 \mathbf{V}_r^T, \tag{40}$$

where $\mathbf{S}_0 = \mathbf{P} \tilde{\mathbf{S}}_{\tilde{r}} \mathbf{P}^T$. We now show that $\mathbf{S}_0$ is a diagonal matrix. Indeed, by denoting the (i, j)-th entry of $\mathbf{P}$ as $p_{i,j}$ such that $\mathbf{P} = [\mathbf{p}_1 \cdots \mathbf{p}_r] = [p_{i,j}]$, we rewrite (30) and (31) as

$$p_{i,j} = 0, \quad \text{or} \tag{41}$$

$$p_{i,j} > 0 \quad \text{and} \quad \tilde{\sigma}_j^q = \alpha \sigma_i^p, \tag{42}$$

for all $(i, j) \in \{1, \ldots, r\}^2$. Moreover, by expanding the (i, j)-th entry of the matrix $\mathbf{S}_0$, we have that

$$[\mathbf{S}_0]_{i,j} = \left[\mathbf{P} \tilde{\mathbf{S}}_{\tilde{r}} \mathbf{P}^T\right]_{i,j} = \sum_{k=1}^{r} p_{i,k} \tilde{\sigma}_k p_{j,k} = \sum_{k=1}^{r} p_{i,k} \sigma_i^{\frac{p}{q}} \alpha^{\frac{1}{q}} p_{j,k}$$

$$= \sigma_i^{\frac{p}{q}} \alpha^{\frac{1}{q}} \mathbf{p}_i^T \mathbf{p}_j = \left[\mathrm{J}_p(\boldsymbol{\sigma})\right]_i c_{\mathbf{B}} \delta[i-j],$$

where $\delta[\cdot]$ denotes the Kronecker delta and $c_{\mathbf{B}} = \alpha^{\frac{1}{q}} > 0$ is a positive constant. Finally, we obtain the announced expression in (14) by replacing the above characterization of $\mathbf{S}_0$ in (40).

For the converse, we note that, if the matrix $\mathbf{B}$ is in the form of (14), then we have that

$$\begin{aligned}\mathrm{Tr}(\mathbf{A}^T\mathbf{B}) &= \mathrm{Tr}\left(\mathbf{U}_r \mathrm{diag}(\boldsymbol{\sigma})\mathbf{V}_r^T\left(\mathbf{U}_r \mathrm{diag}(\mathrm{J}_p(\boldsymbol{\sigma}))\mathbf{V}_r^T\right)^T\right)\\ &= c_{\mathbf{B}}\mathrm{Tr}\left(\mathrm{diag}(\boldsymbol{\sigma})\mathbf{V}_r^T\mathbf{V}_r \mathrm{diag}(\mathrm{J}_p(\boldsymbol{\sigma}))\mathbf{U}_r^T\mathbf{U}_r\right)\\ &= c_{\mathbf{B}}\boldsymbol{\sigma}^T\mathrm{J}_p(\boldsymbol{\sigma})\\ &= c_{\mathbf{B}}||\boldsymbol{\sigma}||_p||\mathrm{J}_p(\boldsymbol{\sigma})||_q = ||\mathbf{A}||_{S_p}||\mathbf{B}||_{S_q},\end{aligned}$$

which shows that the equality is indeed saturated in this case.

Case 2: $p = 1$. In this case, the saturation of the weighted Hölder inequality implies that, for all $(i,j) \in \{1,\ldots,r\} \times \{1,\ldots,\tilde{r}\}$, we have that

$$\mathbf{u}_i^T\tilde{\mathbf{u}}_j\mathbf{v}_i^T\tilde{\mathbf{v}}_j = 0, \quad \text{or} \tag{43}$$

$$\mathbf{u}_i^T\tilde{\mathbf{u}}_j\mathbf{v}_i^T\tilde{\mathbf{v}}_j > 0 \quad \text{and} \quad \tilde{\sigma}_j = \tilde{\sigma}_1. \tag{44}$$

For equality, we also need to have the saturation of (28), which we showed to be equivalent to (32) and (35). From (32), we deduce the existence of matrices $\mathbf{P}_1, \mathbf{P}_2 \in \mathbb{R}^{\tilde{r}\times r}$ such that

$$\mathbf{U}_r = \tilde{\mathbf{U}}_{\tilde{r}}\mathbf{P}_1, \qquad \mathbf{V}_r = \tilde{\mathbf{V}}_{\tilde{r}}\mathbf{P}_2. \tag{45}$$

The replacement of these in (35) implies that

$$\mathbf{P}_1^T = \mathbf{P}_1^T\tilde{\mathbf{U}}\tilde{r}^T\tilde{\mathbf{U}}_{\tilde{r}} = \mathbf{U}_r^T\tilde{\mathbf{U}}_{\tilde{r}} = \mathbf{V}_r^T\tilde{\mathbf{V}}_{\tilde{r}} = \mathbf{P}_2^T\tilde{\mathbf{V}}\tilde{r}^T\tilde{\mathbf{V}}_{\tilde{r}} = \mathbf{P}_2^T, \tag{46}$$

and, hence, that $\mathbf{P}_1 = \mathbf{P}_2 = [p_{i,j}] \in \mathbb{R}^{\tilde{r}\times r}$. Now, one can rewrite the conditions (43) and (44) and deduce that, for any $j = 1,\ldots,\tilde{r}$, we have that

$$p_{j,i} = 0, \quad \forall i = 1,\ldots,r \quad \text{or} \quad \tilde{\sigma}_j = \tilde{\sigma}_1. \tag{47}$$

From Conditions (47) and following the definition of r_1 (the multiplicity of the largest singular value), we deduce that

$$\mathbf{P}_1 = \begin{bmatrix}\mathbf{P}\\ \mathbf{0}_{r_{\mathrm{res}}\times r}\end{bmatrix}, \tag{48}$$

where $\mathbf{P} \in \mathbb{R}^{r_1\times r}$ and $r_{\mathrm{res}} = (\tilde{r} - r_1)$. Using this form and the definition of $\tilde{\mathbf{U}}_1$ and $\tilde{\mathbf{V}}_1$ (given in the statement of the proposition), we rewrite (45) as

$$\mathbf{U}_r = \tilde{\mathbf{U}}_1\mathbf{P}, \qquad \mathbf{V}_r = \tilde{\mathbf{V}}_1\mathbf{P}. \tag{49}$$

Therefore,

$$\mathbf{I}_r = \mathbf{U}_r^T\mathbf{U}_r = \mathbf{P}^T\tilde{\mathbf{U}}_1^T\tilde{\mathbf{U}}_1\mathbf{P} = \mathbf{P}^T\mathbf{P}. \tag{50}$$

Hence, $\mathbf{P}$ is a sub-orthogonal matrix and

$$\mathrm{rank}(\mathbf{B}) = \tilde{r} \geq r_1 \geq \mathrm{rank}(\mathbf{P}) \geq r = \mathrm{rank}(\mathbf{A}).$$

The replacement of (49) in the reduced SVD of $\mathbf{A}$ yields the announced expression with $\mathbf{X} = \mathbf{PSP}^T$.

Based on the definitions of r_1, $\tilde{\mathbf{U}}_1$, and $\tilde{\mathbf{V}}_1$, we note that one can rewrite the reduced SVD of $\mathbf{B}$ as

$$\mathbf{B} = \tilde{\sigma}_1 \tilde{\mathbf{U}}_1 \tilde{\mathbf{V}}_1^T + \tilde{\mathbf{U}}_{\text{res}} \tilde{\mathbf{S}}_{\text{res}} \tilde{\mathbf{V}}_{\text{res}}^T, \tag{51}$$

where $\tilde{\mathbf{U}}_{\text{res}} \in \mathbb{R}^{m \times r_{\text{res}}}$, $\tilde{\mathbf{S}}_{\text{res}} \in \mathbb{R}^{r_{\text{res}} \times r_{\text{res}}}$, and $\tilde{\mathbf{V}}_{\text{res}} \in \mathbb{R}^{n \times r_{\text{res}}}$ are the remaining singular values and vectors such that

$$\tilde{\mathbf{U}} = \begin{bmatrix} \tilde{\mathbf{U}}_1 & \tilde{\mathbf{U}}_{\text{res}} \end{bmatrix}, \quad \tilde{\mathbf{V}} = \begin{bmatrix} \tilde{\mathbf{V}}_1 & \tilde{\mathbf{V}}_{\text{res}} \end{bmatrix}, \quad \tilde{\mathbf{S}} = \begin{bmatrix} \tilde{\sigma}_1 \mathbf{I}_{r_1} & \mathbf{0} \\ \mathbf{0} & \tilde{\mathbf{S}}_{\text{res}} \end{bmatrix}.$$

Now, if $\mathbf{A}$ admits the form (17) and if we consider the SVD of $\mathbf{X} = \mathbf{PSP}^T$ (the assumption that $\mathbf{X}$ is symmetric ensures that is has an orthogonal eigen-decomposition), then

$$\begin{aligned} \text{Tr}(\mathbf{A}^T\mathbf{B}) &= \text{Tr}(\tilde{\mathbf{V}}_1 \mathbf{PSP}^T \tilde{\mathbf{U}}_1^T (\tilde{\sigma}_1 \tilde{\mathbf{U}}_1 \tilde{\mathbf{V}}_1^T + \tilde{\mathbf{U}}_{\text{res}} \tilde{\mathbf{S}}_{\text{res}} \tilde{\mathbf{V}}_{\text{res}}^T)) \\ &= \tilde{\sigma}_1 \text{Tr}(\tilde{\mathbf{V}}_1 \mathbf{PSP}^T \tilde{\mathbf{U}}_1^T \tilde{\mathbf{U}}_1 \tilde{\mathbf{V}}_1^T) + \text{Tr}(\tilde{\mathbf{V}}_1 \mathbf{PSP}^T \tilde{\mathbf{U}}_1^T \tilde{\mathbf{U}}_{\text{res}} \tilde{\mathbf{S}}_{\text{res}} \tilde{\mathbf{V}}_{\text{res}}^T) \\ &= \tilde{\sigma}_1 \text{Tr}(\tilde{\mathbf{V}}_1 \mathbf{PSP}^T \mathbf{I}_{r_1} \tilde{\mathbf{V}}_1^T) + \text{Tr}(\tilde{\mathbf{V}}_1 \mathbf{PSP}^T \mathbf{0}_{r_1 \times r_{\text{res}}} \tilde{\mathbf{S}}_{\text{res}} \tilde{\mathbf{V}}_{\text{res}}^T) \\ &= \tilde{\sigma}_1 \text{Tr}(\mathbf{SP}^T \tilde{\mathbf{V}}_1^T \tilde{\mathbf{V}}_1 \mathbf{P}) + 0 \\ &= \tilde{\sigma}_1 \text{Tr}(\mathbf{SP}^T \mathbf{P}) \\ &= \tilde{\sigma}_1 \text{Tr}(\mathbf{S}) = ||\mathbf{B}||_{S_\infty} ||\mathbf{A}||_{S_1}, \end{aligned}$$

which establishes the sufficiency in this case.

Finally, assuming that $r = r_1 = \tilde{r}$, we deduce that $\mathbf{P} \in \mathbb{R}^{r \times r}$ is an orthogonal matrix and, hence, that $\mathbf{P}^{-1} = \mathbf{P}^T$. Now, using (49) and the rank assumption, we can simplify the expansion (51) as

$$\mathbf{B} = \tilde{\sigma}_1 \tilde{\mathbf{U}}_1 \tilde{\mathbf{V}}_1^T = \tilde{\sigma}_1 \mathbf{U}_r \mathbf{P}^T \left(\mathbf{V}_r \mathbf{P}^T\right)^T = \tilde{\sigma}_1 \mathbf{U}_r \mathbf{P}^T \mathbf{P} \mathbf{V}_r^T = \tilde{\sigma}_1 \mathbf{U}_r \mathbf{V}_r^T. \tag{52}$$

C. Proof of Theorem 1

Proof. **Case I:** $1 < p < +\infty$. Assume that $(\mathbf{A}, \mathbf{B})$ forms an (S_p, S_q)-conjugate pair. Hence, we have that $\langle \mathbf{A}, \mathbf{B} \rangle = ||\mathbf{A}||_{S_p} ||\mathbf{B}||_{S_q}$ which, together with Proposition 2, implies that $\mathbf{B}$ admits the form

$$\mathbf{B} = \frac{||\mathbf{B}||_{S_q}}{||\mathbf{A}||_{S_p}} \mathbf{U}_r \text{diag}(\mathrm{J}_p(\boldsymbol{\sigma})) \mathbf{V}_r^T = \mathbf{U}_r \text{diag}(\mathrm{J}_p(\boldsymbol{\sigma})) \mathbf{V}_r^T.$$

Case II: $p = 1$. Similarly to the previous case, consider $\mathbf{A} \in \mathbb{R}^{m \times n}$ and $\mathbf{B} \in \mathcal{J}_{S_1, \text{rank}}(\mathbf{A})$. We have that

$$\text{Tr}(\mathbf{A}^T\mathbf{B}) = ||\mathbf{A}||_{S_1} ||\mathbf{B}||_{S_\infty} \tag{53}$$

$$||\mathbf{A}||_{S_1} = ||\mathbf{B}||_{S_\infty} \tag{54}$$

$$\text{rank}(\mathbf{B}) \leq \text{rank}(\mathbf{C}), \quad \forall \mathbf{C} \in \mathcal{J}_{S_1}(\mathbf{A}). \tag{55}$$

From (53) and using Proposition 2, we deduce that $\text{rank}(\mathbf{B}) \geq \text{rank}(\mathbf{A})$ which, together with (55), implies that $\mathbf{B}$ should be equal to

$$\mathbf{B} = ||\mathbf{B}||_{S_\infty} \mathbf{U}_r \mathbf{V}_r^T = ||\boldsymbol{\sigma}||_1 \mathbf{U}_r \mathbf{V}_r^T,$$

where the last equality is obtained using (54).

Case III: $p = +\infty$. Following Proposition 2, any matrix $\mathbf{B} \in \mathcal{J}_{S_\infty}(\mathbf{A})$ can be expressed as

$$\mathbf{B} = \mathbf{U}_1 \tilde{\mathbf{X}} \mathbf{V}_1^T,$$

where $\tilde{\mathbf{X}} \in \mathbb{R}^{r_1 \times r_1}$ is a symmetric matrix. By defining $\mathbf{X} = \sigma_1^{-1}\tilde{\mathbf{X}}$, one readily verifies that $\mathbf{B} = \sigma_1 \mathbf{U}_1 \mathbf{X} \mathbf{V}_1^T$. By recalling the normalization constraint $||\mathbf{A}||_{S_\infty} = ||\mathbf{B}||_{S_1}$, we therefore obtain that

$$\sigma_1 = ||\mathbf{A}||_{S_\infty} = ||\mathbf{B}||_{S_1} = \sigma_1 ||\mathbf{X}||_{S_1},$$

which implies that $||\mathbf{X}||_{S_1} = 1$. To show that $\mathcal{J}_{S_\infty}(\mathbf{A})$ is convex, consider two symmetric matrices $\mathbf{X}_0$ and $\mathbf{X}_1$ in the unit ball of Schatten-1 norm and define

$$\mathbf{B}_\alpha = \sigma_1 \mathbf{U}_1 \mathbf{X}_\alpha \mathbf{V}_1^T, \quad \mathbf{X}_\alpha = \alpha \mathbf{X}_0 + (1-\alpha)\mathbf{X}_1$$

for $\alpha \in [0,1]$. On one hand, from the linearity of traces, we have that

$$\mathrm{Tr}(\mathbf{A}^T \mathbf{B}_\alpha) = \mathrm{Tr}\big(\mathbf{A}^T(\alpha \mathbf{B}_1 + (1-\alpha)\mathbf{B}_0)\big) = \alpha \mathrm{Tr}(\mathbf{A}^T \mathbf{B}_1) + (1-\alpha)\mathrm{Tr}(\mathbf{A}^T \mathbf{B}_0).$$

On the other hand, from the definition of $\mathbf{X}_0$ and $\mathbf{X}_1$, we deduce that $\mathbf{B}_0, \mathbf{B}_1 \in \mathcal{J}_{S_\infty}(\mathbf{A})$. Hence,

$$\mathrm{Tr}(\mathbf{A}^T \mathbf{B}_\alpha) = \alpha ||\mathbf{A}||_{S_\infty}^2 + (1-\alpha)||\mathbf{A}||_{S_\infty}^2 = ||\mathbf{A}||_{S_\infty}^2.$$

However, from the Hölder inequality and the convexity of norms, we have that

$$\mathrm{Tr}(\mathbf{A}^T \mathbf{B}_\alpha) \leq ||\mathbf{A}||_{S_\infty} ||\mathbf{B}_\alpha||_{S_1} \leq ||\mathbf{A}||_{S_\infty} \big(\alpha ||\mathbf{B}_1||_{S_1} + (1-\alpha)||\mathbf{B}_0||_{S_1}\big) = ||\mathbf{A}||_{S_\infty}.$$

This implies that the Hölder inequality is saturated and also that $||\mathbf{B}_\alpha||_{S_1} = ||\mathbf{A}||_{S_\infty}$ which, altogether, implies that $\mathbf{B}_\alpha \in \mathcal{J}_{S_\infty}(\mathbf{A})$ for all $\alpha \in [0,1]$.

Finally, we observe that the set $\mathcal{J}_{S_\infty}(\mathbf{A})$ contains all matrices of the form $\mathbf{B} = \mathbf{U}_1 \mathbf{p}\mathbf{p}^T \mathbf{V}_1^T$ for any vector $\mathbf{p} \in \mathbb{R}^{r_1}$ with $||\mathbf{p}||_2 = 1$. These are indeed all the rank-1 elements of $\mathcal{J}_{S_\infty}(\mathbf{A})$ which, due to the Definition 2, forms the set of sparse dual conjugates. □