Convex Analysis in Spectral Decomposition Systems*

Hòa T. Bùi[1], Minh N. Bùi[2], and Christian Clason[3]

[1]Curtin University
Centre for Optimisation and Decision Science
Kent Street, Bentley, Western Australia 6102, Australia
hoa.bui@curtin.edu.au

[2]University of Graz
Department of Mathematics and Scientific Computing, NAWI Graz
Heinrichstraße 36, 8010 Graz, Austria
minh.bui@uni-graz.at

[3]University of Graz
Department of Mathematics and Scientific Computing, NAWI Graz
Heinrichstraße 36, 8010 Graz, Austria
c.clason@uni-graz.at

Abstract. This work is concerned with convex analysis of so-called spectral functions of matrices that only depend on eigenvalues of the matrix. An abstract framework of spectral decomposition systems is proposed that covers a wide range of previously studied settings, including eigenvalue decomposition of Hermitian matrices and singular value decomposition of rectangular matrices and allows deriving new results in more general settings such as Euclidean Jordan algebras. The main results characterize convexity, lower semicontinuity, Fenchel conjugates, convex subdifferentials, and Bregman proximity operators of spectral functions in terms of the reduced functions. As a byproduct, a generalization of the Ky Fan majorization theorem is obtained.

Keywords. Bregman proximity operator; conjugation; convexity; Ky Fan majorization theorem; spectral decomposition system; spectral function; convex subdifferential;

*Corresponding author: M. N. Bùi. The work of H. T. Bùi was supported by the Australian Research Council through the Centre for Transforming Maintenance through Data Science grant number IC180100030. This research was funded in whole or in part by the Austrian Science Fund (FWF) 10.55776/F100800.

1. Introduction

Many practically relevant optimization problems involve minimizing over spaces or cones of matrices instead of vectors; by way of example we only mention non-negative matrix factorization [28], matrix completion [10], low-rank approximation [21, 39], or operator learning [32]. Particularly – but not only – in the last example, one is actually interested in optimizing over (finite-dimensional) linear operators and not their particular matrix representations. This implies that the functions to be minimized should be invariant under basis changes. It is a well-known fact from linear algebra that, under appropriate assumptions, such functions are fully characterized by their dependence on the eigenvalues (or singular values) of its argument; the most well-known example is probably the nuclear norm of a matrix. Correspondingly, examples of such *spectral functions* are ubiquitous in applications areas as diverse as robust matrix estimation [8], signal processing [11], conic programming [16], semi-definite programming [29], nonlinear elasticity [48], and brain network analysis [56].

Of particular interest in this context are nonsmooth convex problems, e.g., nuclear norm minimization or problems with rank constraints. Here the main question is to relate the convexity of the spectral function to that of the invariant function of the eigenvalues (or singular values), and to characterize the fundamental objects of convex analysis such as subdifferentials, Fenchel conjugates, and proximal operators of the former in terms of the latter. The central challenge in this is the fact that the invariant function only depends on the *set* of eigenvalues but not their ordering.

Correspondingly, this problem has received significant attention the last years; see, e.g., [6, 7, 42] and the literature cited therein. In particular, [33] studied absolutely symmetric functions of singular values of rectangular matrices, [34] studied spectral functions on the space of Hermitian matrices, and [35] studied spectral functions in normal decomposition systems (which encompass in particular the spaces of rectangular and Hermitian matrices; see also [36] for applications to Cartan subspaces). In a related direction, convexity and conjugation of a certain type of invariant functions of signed singular values of a matrix was addressed in [18], while [1] investigated the convexity, conjugate, and subdifferentiability of spectral functions in the context of Euclidean Jordan algebras; see also [49] for a special case. Regarding convex geometry, [30] treated operations (such as closure and convex hull) and properties (such as closedness, compactness, convexity) of spectral sets in Euclidean Jordan algebras. Finally, [8] was concerned with Bregman proximity operators of lower semicontinuous convex spectral functions in the context of symmetric matrices.

However, each of these works treated a specific setting in isolation. The aim of this work is therefore to develop a general framework that covers all these settings and – more importantly – allows deriving results more easily for settings and objects not covered so far. In a nutshell, we introduce a *spectral decomposition system* consisting of

(i) a family of *spectral decompositions* that generalize constructing a matrix with given eigenvalues (e.g., via a basis of eigenvectors);

(ii) a *spectral mapping* that generalizes computing the eigenvalues from a given matrix;

(iii) an *ordering mapping* that generalizes sorting eigenvalues in decreasing order;

that satisfy some natural compatibility conditions such as a generalization of von Neumann's trace inequality; see Definition 2.1 for a precise definition. We will show that this definition covers all previously considered settings in uniform generality (applying, for instance, in each case to matrices over the real, complex, or quaternion fields) and give a full characterization of the convexity of spectral functions in this framework. Along the way, we establish a generalization of the well-known Ky Fan's majorization theorem (Theorem 3.9). Our main technical contribution is then to establish a general “reduced minimization principle” (Theorem 5.1), which in particular allows us to compute Fenchel conjugates, subdifferentials, and Fréchet derivatives of convex spectral functions in spectral decomposition systems. We then further apply these results to compute Bregman proximity operators, which generalize the classical proximity (or proximal point) operators that are the basic building blocks of modern first-order nonsmooth optimization algorithms [6, 7, 15] and which to the best of our knowledge are new even in some of the classical settings studied in the literature so far.

This work is structured as follows. In the next Section 2, we give a precise definition of spectral decomposition systems and show how it covers previously studied settings. Section 3 collects a number of technical results on spectral mappings and spectral ordering mappings that are crucial for the following analysis, including the above-mentioned generalization of the Ky Fan majorization theorem (Theorem 3.9). In Section 4, we then define the class of spectral functions studied in this work and characterize their fundamental properties such as lower semicontinuity (Proposition 4.5) and convexity (Theorem 4.6). We also derive fundamental properties of spectral sets. Section 5 is devoted to the reduced minimization principle (Theorem 5.1) and its application to Fenchel conjugates and convex subdifferentials of spectral functions (Corollary 5.3 and Proposition 5.5, respectively). Finally, we study Bregman proximity operators of (not necessarily convex) spectral functions in Section 6.

2. Spectral decomposition systems

To motivate our definition, recall from the introduction that various important optimization problems involve a function Φ acting on a Euclidean (i.e., finite-dimensional Hilbert) space $\mathfrak{H}$ whose values depend only on a “spectral mapping” $\gamma\colon \mathfrak{H} \to \mathcal{X}$ on some “smaller” Euclidean space $\mathcal{X}$, in the sense that $\Phi(X) = \Phi(Y)$ for $X, Y \in \mathfrak{H}$ whenever $\gamma(X) = \gamma(Y)$. For example, in the context of symmetric matrices, $\mathfrak{H} = \mathbb{S}^N$ is the space of $N \times N$ real symmetric matrices, $\mathcal{X} = \mathbb{R}^N$, and γ is the mapping that outputs the eigenvalues of a matrix $X \in \mathfrak{H}$ in decreasing order. Under suitable assumptions, such a function Φ can be “decomposed” as $\Phi = \varphi \circ \gamma$ for some function $\varphi\colon \mathcal{X} \to [-\infty, +\infty]$. A closer inspection of the literature cited in the introduction reveals three additional important ingredients to such decomposition results:

(i) A family $(\Lambda_{a})_{a\in\mathcal{A}}$ of linear isometries from $\mathcal{X}$ to $\mathfrak{H}$ that allows for a “spectral decomposition” of elements of $\mathfrak{H}$ through the spectral mapping γ in the sense that every $X \in \mathfrak{H}$ can be written as $X = \Lambda_a \gamma(X)$ for some $a \in \mathcal{A}$. For instance, in the motivating example $\mathfrak{H} = \mathbb{S}^N$ and $\mathcal{X} = \mathbb{R}^N$, this is the eigenvalue decomposition of matrices in $\mathfrak{H}$ given by

$$\Lambda_U\colon \mathcal{X} \to \mathfrak{H}\colon x \mapsto U(\operatorname{Diag} x)U^{\mathsf{T}}, \tag{2.1}$$

where U is an orthogonal matrix and $\operatorname{Diag} x$ is the diagonal matrix with elements given by those of the vector x.

(ii) A von Neumann-type inequality relating the scalar product of elements in $\mathfrak{H}$ to that of their spectra in $\mathcal{X}$.

It turns out that these must satisfy some natural invariance conditions with respect to the action of a certain group S on $\mathcal{X}$ (of permutations, in the example of symmetric matrices) in order to allow the desired decomposition. Making these precise requires introducing some notation that will be used throughout this work. First, the scalar product and the associated norm of a Euclidean (i.e., finite-dimensional Hilbert) space are denoted by $\langle \cdot \mid \cdot \rangle$ and $\|\cdot\|$, respectively. Given a Euclidean space $\mathcal{H}$ and a group G acting on $\mathcal{H}$, we will denote the composition of two elements $\mathsf{g} \in \mathsf{G}$ and $\mathsf{h} \in \mathsf{G}$ by gh and use the symbol $\cdot$ to denote the group action. Furthermore:

- The *orbit* of an element $x \in \mathcal{H}$ is defined by $\mathsf{G} \cdot x = \{\mathsf{g} \cdot x \mid \mathsf{g} \in \mathsf{G}\}$.
- Given a set $\mathcal{U}$, a mapping $\Phi\colon \mathcal{H} \to \mathcal{U}$ is said to be G-*invariant* if $(\forall x \in \mathcal{H})(\forall \mathsf{g} \in \mathsf{G})\ \Phi(\mathsf{g} \cdot x) = \Phi(x)$.
- A subset C of $\mathcal{H}$ is said to be G-*invariant* if its indicator function

$$\iota_C\colon \mathcal{H} \to \left]-\infty, +\infty\right] : x \mapsto \begin{cases} 0 & \text{if } x \in C, \\ +\infty & \text{otherwise}, \end{cases} \tag{2.2}$$

  is G-invariant or, equivalently, $(\forall x \in C)(\forall \mathsf{g} \in \mathsf{G})\ \mathsf{g} \cdot x \in C$.
- We say that G *acts on* $\mathcal{H}$ *by linear isometries* if, for every $\mathsf{g} \in \mathsf{G}$, the mapping $\mathcal{H} \to \mathcal{H} : x \mapsto \mathsf{g} \cdot x$ is a linear isometry.

We are now ready to introduce the abstract framework we will work in.

Definition 2.1 (spectral decomposition system). Let $\mathfrak{H}$ and $\mathcal{X}$ be Euclidean spaces, let S be a group which acts on $\mathcal{X}$ by linear isometries, let $\gamma\colon \mathfrak{H} \to \mathcal{X}$, and let $(\Lambda_{a})_{a \in \mathscr{A}}$ be a family of linear operators from $\mathcal{X}$ to $\mathfrak{H}$. We say that the tuple $\mathfrak{S} = (\mathcal{X}, \mathsf{S}, \gamma, (\Lambda_{a})_{a \in \mathscr{A}})$ is a *spectral decomposition system* for $\mathfrak{H}$ if the following are satisfied:

[A] For every $a \in \mathscr{A}$, Λ_{a} is an isometry.

[B] There exists an S-invariant mapping $\tau\colon \mathcal{X} \to \mathcal{X}$ such that

$$\begin{cases} (\forall x \in \mathcal{X}) & \tau(x) \in \mathsf{S} \cdot x \\ (\forall a \in \mathscr{A}) & \gamma \circ \Lambda_{a} = \tau. \end{cases} \tag{2.3}$$

[C] $(\forall X \in \mathfrak{H})(\exists\, a \in \mathscr{A})\ X = \Lambda_{a}\gamma(X)$.

[D] $(\forall X \in \mathfrak{H})(\forall Y \in \mathfrak{H})\ \langle X \mid Y \rangle \leqslant \langle \gamma(X) \mid \gamma(Y) \rangle$.

Here:

- The mapping γ is called the *spectral mapping* of the system $\mathfrak{S}$.
- The mapping τ in property [B] is called the *spectral-induced ordering mapping* of the system $\mathfrak{S}$. (The motivation for this term will become apparent from the concrete scenarios described in Examples 2.5, 2.6, 2.7, and 2.8.)
- We set

$$(\forall X \in \mathfrak{H}) \quad \mathscr{A}_X = \{a \in \mathscr{A} \mid X = \Lambda_{a}\gamma(X)\}. \tag{2.4}$$

  Note that property [C] ensures that the sets $(\mathscr{A}_X)_{X \in \mathfrak{H}}$ are nonempty.
- Given $X \in \mathfrak{H}$, the vector $\gamma(X)$ is called the *spectrum* of X with respect to $\mathfrak{S}$ and, for every $a \in \mathscr{A}_X$, the identity

$$X = \Lambda_{a}\gamma(X) \tag{2.5}$$

  is called a *spectral decomposition* of X with respect to $\mathfrak{S}$.

The remainder of this section is concerned with examples that illustrate the breadth of the proposed abstract definition of a spectral decomposition system through a broad range of concrete settings found in the literature. To this end, we first collect some further notation that will be used in these examples, also in the following section.

Notation 2.2. Let M and N be strictly positive integers.

- $\mathbb{K}$ denotes one of the following: the field $\mathbb{R}$ of real numbers, the field $\mathbb{C}$ of complex numbers, or the skew-field $\mathbb{H}$ of Hamiltonian quaternions (we refer the reader to [47] for background on quaternions).
- The canonical involution on $\mathbb{K}$ is denoted by $\xi \mapsto \overline{\xi}$, which fixes only the elements of $\mathbb{R}$, and the real part of $\xi \in \mathbb{K}$ is $\operatorname{Re}\xi = (\xi + \overline{\xi})/2$.
- $\mathbb{K}^{M\times N}$ denotes the real vector space of $M \times N$ matrices with entries in $\mathbb{K}$.
- Given a matrix $X = [\xi_{ij}]_{\substack{1\leqslant i\leqslant M\\1\leqslant j\leqslant N}} \in \mathbb{K}^{M\times N}$, its conjugate transpose is $X^* = [\overline{\xi_{ji}}]_{\substack{1\leqslant j\leqslant N\\1\leqslant i\leqslant M}} \in \mathbb{K}^{N\times M}$.
- The trace of a matrix $X \in \mathbb{K}^{N\times N}$ is denoted by $\operatorname{tra} X$.
- $\mathsf{H}^N(\mathbb{K})$ denotes the vector subspace of $\mathbb{K}^{N\times N}$ which consists of Hermitian matrices, that is,

$$\mathsf{H}^N(\mathbb{K}) = \big\{X \in \mathbb{K}^{N\times N} \mid X = X^*\big\}. \tag{2.6}$$

- $\mathsf{U}^N(\mathbb{K}) = \big\{U \in \mathbb{K}^{N\times N} \mid U^*U = \mathrm{Id}\big\}$ is the group of unitary matrices.
- $\mathsf{SO}^N = \big\{U \in \mathsf{U}^N(\mathbb{R}) \mid \det U = 1\big\}$ is the special orthogonal group.
- $\mathsf{P}_{\pm}^N$ denotes the multiplicative group of all matrices in $\mathbb{R}^{N\times N}$ with entries in $\{-1, 0, 1\}$ which contain exactly one nonzero entry in every row and every column. (The elements of $\mathsf{P}_{\pm}^N$ are called *signed permutation matrices*.)
- P^N denotes the subgroup of $\mathsf{P}_{\pm}^N$ which consists of matrices with entries in $\{0, 1\}$. (The elements of P^N are called *permutation matrices*.)
- For every $x = (\xi_i)_{1\leqslant i\leqslant N} \in \mathbb{R}^N$, $x^{\downarrow}$ denotes the rearrangement vector of x with entries listed in decreasing order, and $|x|^{\downarrow}$ denotes the rearrangement vector of $(|\xi_i|)_{1\leqslant i\leqslant N}$ with entries listed in decreasing order.

We begin with a simple example.

Example 2.3. Let $2 \leqslant N \in \mathbb{N}$, let the (multiplicative) group $\{-1, 1\}$ act on $\mathbb{R}$ via multiplication, and set

$$e = (1, 0, \ldots, 0) \in \mathbb{R}^N \quad \text{and} \quad \big(\forall U \in \mathsf{U}^N(\mathbb{R})\big)\ \Lambda_U : \mathbb{R} \to \mathbb{R}^N : \xi \mapsto \xi U e. \tag{2.7}$$

Then $(\mathbb{R}, \{-1, 1\}, \|\cdot\|_2, (\Lambda_U)_{U\in\mathsf{U}^N(\mathbb{R})})$ is a spectral decomposition system for $\mathbb{R}^N$.

Proof. One can verify that property [B] in Definition 2.1 is satisfied with $\tau = |\cdot|$, and recognize at once that property [D] in Definition 2.1 is precisely the Cauchy–Schwarz inequality. It therefore remains to verify property [C] in Definition 2.1. To this end, let $x \in \mathbb{R}^N \smallsetminus \{0\}$ and set

$$y = \frac{1}{\|x\|_2}x, \quad u = \frac{1}{\|e - y\|_2}(e - y), \quad \text{and} \quad U = \mathrm{Id} - 2uu^{\mathsf{T}}. \tag{2.8}$$

Then, a straightforward computation shows that $U \in \mathsf{U}^N(\mathbb{R})$ and $x = \Lambda_U(\|x\|_2)$. ☐

Our first nontrivial example is the framework of [35], which subsumes, in particular, the semisimple real Lie algebra framework of [36, 51, 54].

Example 2.4 (normal decomposition system). Let $(\mathfrak{H}, \mathsf{G}, \gamma)$ be a *normal decomposition system* in the sense of [35, Definition 2.1], that is, $\mathfrak{H}$ is a Euclidean space, G is a group which acts on $\mathfrak{H}$ by linear isometries, and $\gamma\colon \mathfrak{H} \to \mathfrak{H}$ is a G-invariant mapping such that

$$\begin{cases} (\forall X \in \mathfrak{H})(\exists\, \mathsf{g} \in \mathsf{G}) \quad X = \mathsf{g} \cdot \gamma(X) \\ (\forall X \in \mathfrak{H})(\forall Y \in \mathfrak{H}) \quad \langle X \mid Y \rangle \leqslant \langle \gamma(X) \mid \gamma(Y) \rangle. \end{cases} \tag{2.9}$$

Additionally, let $\mathcal{X}$ be a vector subspace of $\mathfrak{H}$ which contains the range of γ, define a subgroup S of G by

$$\mathsf{S} = \big\{ \mathsf{s} \in \mathsf{G} \mid \mathsf{s} \cdot \mathcal{X} = \mathcal{X} \big\}, \tag{2.10}$$

and set $(\forall \mathsf{g} \in \mathsf{G})$ $\Lambda_{\mathsf{g}}\colon \mathcal{X} \to \mathfrak{H}\colon X \mapsto \mathsf{g} \cdot X$. Suppose that

$$(\forall X \in \mathcal{X})(\exists\, \mathsf{s} \in \mathsf{S}) \quad X = \mathsf{s} \cdot \gamma(X). \tag{2.11}$$

Then $\mathfrak{S} = (\mathcal{X}, \mathsf{S}, \gamma, (\Lambda_{\mathsf{g}})_{\mathsf{g} \in \mathsf{G}})$ is a spectral decomposition system for $\mathfrak{H}$.

Proof. In fact, property [B] in Definition 2.1 is satisfied with $\tau = \gamma|_{\mathcal{X}}$. □

The next example concerns the Euclidean Jordan algebra framework of [1, 30, 38, 50], which captures in particular the space $\mathsf{H}^N(\mathbb{K})$ of Hermitian matrices (see Example 2.6) that arises in applications such as robust matrix estimation [8], semi-definite programming [29], and brain network analysis [56]. As shown in [43], in general Euclidean Jordan algebras cannot be embedded into a normal decomposition system of Example 2.4.

Example 2.5 (Euclidean Jordan algebra). Let $\mathfrak{H}$ be a *Euclidean Jordan algebra* (also known as *formally real Jordan algebra*), that is, $\mathfrak{H}$ is a finite-dimensional real vector space which is endowed with a bilinear form

$$\mathfrak{H} \times \mathfrak{H} \to \mathfrak{H}\colon (X, Y) \mapsto X \circledast Y \tag{2.12}$$

such that the following are satisfied:

[A] $(\forall X \in \mathfrak{H})(\forall Y \in \mathfrak{H})$ $X \circledast Y = Y \circledast X$ and $X \circledast ((X \circledast X) \circledast Y) = (X \circledast X) \circledast (X \circledast Y)$.

[B] There exists a scalar product $(\cdot \mid \cdot)$ on $\mathfrak{H}$ such that $(\forall X \in \mathfrak{H})(\forall Y \in \mathfrak{H})(\forall Z \in \mathfrak{H})$ $(X \circledast Y \mid Z) = (X \mid Y \circledast Z)$.

(We refer to [22] for background on and concrete examples of Euclidean Jordan algebras.) We equip $\mathfrak{H}$ with the scalar product

$$(\forall X \in \mathfrak{H})(\forall Y \in \mathfrak{H}) \quad \langle X \mid Y \rangle = \operatorname{Tra}(X \circledast Y), \tag{2.13}$$

where $\operatorname{Tra} X$ is the trace in $\mathfrak{H}$ of an element $X \in \mathfrak{H}$; see [22, Section II.2]. In addition, we denote by E the identity element of $\mathfrak{H}$ and by N the rank of $\mathfrak{H}$. Additionally, let P^N act on $\mathbb{R}^N$ via matrix-vector multiplication. Following [1], we define a *Jordan frame* of $\mathfrak{H}$ as a family $(A_i)_{1 \leqslant i \leqslant N}$ in $\mathfrak{H}^N \smallsetminus \{0\}$ such that

$$\begin{cases} (\forall i \in \{1, \ldots, N\})(\forall j \in \{1, \ldots, N\}) \quad A_i \circledast A_j = \begin{cases} A_i & \text{if } i = j, \\ 0 & \text{if } i \neq, \end{cases} \\ \displaystyle\sum_{i=1}^N A_i = E. \end{cases} \tag{2.14}$$

(This is equivalent to the standard definition, e.g., on [22, p. 44], of a complete system of orthogonal primitive idempotents satisfying (2.14) since such a system necessarily is of size N and vice versa.)

According to the spectral decomposition theorem for Euclidean Jordan algebras [22, Theorem III.1.2], for every $X \in \mathfrak{H}$, there exist a unique vector $(\lambda_1(X), \ldots, \lambda_N(X)) \in \mathbb{R}^N$, the entries of which are called the *eigenvalues* of X, and a Jordan frame $(A_i)_{1\leqslant i\leqslant N}$ such that

$$\lambda_1(X) \geqslant \cdots \geqslant \lambda_N(X) \quad \text{and} \quad X = \sum_{i=1}^{N} \lambda_i(X) A_i; \tag{2.15}$$

this decomposition process thus defines a mapping

$$\lambda\colon \mathfrak{H} \to \mathbb{R}^N \colon X \mapsto \big(\lambda_1(X), \ldots, \lambda_N(X)\big). \tag{2.16}$$

Further, let $\mathscr{A}$ be the set of Jordan frames of $\mathfrak{H}$ and define

$$\big(\forall \mathscr{a} = (A_i)_{1\leqslant i\leqslant N} \in \mathscr{A}\big) \quad \Lambda_{\mathscr{a}} \colon \mathbb{R}^N \to \mathfrak{H} \colon x = (\xi_i)_{1\leqslant i\leqslant N} \mapsto \sum_{i=1}^{N} \xi_i A_i. \tag{2.17}$$

Then $\mathfrak{S} = (\mathbb{R}^N, \mathsf{P}^N, \lambda, (\Lambda_{\mathscr{a}})_{\mathscr{a}\in\mathscr{A}})$ is a spectral decomposition system for $\mathfrak{H}$.

Proof. If A is an element of a Jordan frame of $\mathfrak{H}$, then it follows from the spectral decomposition theorem that $\lambda(A) = (1, 0, \ldots, 0)$ and, in turn, from (2.13), (2.14), and [22, Theorem III.1.2] that

$$\|A\|^2 = \operatorname{Tra}(A \circledast A) = \operatorname{Tra} A = \sum_{i=1}^{N} \lambda_i(A) = 1. \tag{2.18}$$

Therefore, for every $\mathscr{a} = (A_i)_{1\leqslant i\leqslant N} \in \mathscr{A}$, inasmuch as $(A_i)_{1\leqslant i\leqslant N}$ is an orthonormal family thanks to (2.14) and (2.13), we derive from (2.17) that

$$\big(\forall x = (\xi_i)_{1\leqslant i\leqslant N} \in \mathbb{R}^N\big) \quad \|\Lambda_{\mathscr{a}} x\|^2 = \sum_{i=1}^{N} \xi_i^2 = \|x\|_2^2. \tag{2.19}$$

This implies that $\Lambda_{\mathscr{a}}$ is a linear isometry. Next, define $\tau\colon \mathbb{R}^N \to \mathbb{R}^N \colon x \mapsto x^{\downarrow}$. It is clear that τ is P^N-invariant and $(\forall x \in \mathbb{R}^N)$ $\tau(x) \in \mathsf{P}^N \cdot x$. At the same time, invoking the spectral decomposition theorem once more, we get

$$\big(\forall \mathscr{a} = (A_i)_{1\leqslant i\leqslant N} \in \mathscr{A}\big)\big(\forall x = (\xi_i)_{1\leqslant i\leqslant N} \in \mathbb{R}^N\big) \quad \lambda(\Lambda_{\mathscr{a}} x) = \lambda\left(\sum_{i=1}^{N} \xi_i A_i\right) = x^{\downarrow} = \tau(x) \tag{2.20}$$

and

$$(\forall X \in \mathfrak{H})\big(\exists\, \mathscr{a} = (A_i)_{1\leqslant i\leqslant N} \in \mathscr{A}\big) \quad X = \sum_{i=1}^{N} \lambda_i(X) A_i = \Lambda_{\mathscr{a}} \lambda(X), \tag{2.21}$$

where the last equality follows from (2.16) and (2.17). Furthermore, we deduce from (2.13) and [1, Theorem 23] that

$$(\forall X \in \mathfrak{H})(\forall Y \in \mathfrak{H}) \quad \langle X \mid Y \rangle \leqslant \sum_{i=1}^{N} \lambda_i(X)\lambda_i(Y) = \langle \lambda(X) \mid \lambda(Y) \rangle. \tag{2.22}$$

Altogether, $\mathfrak{S}$ is a spectral decomposition system for $\mathfrak{H}$. ☐

If in Example 2.5 $\mathfrak{H}$ is the Euclidean Jordan algebra of Hermitian matrices (see [22, Section V.2]), we obtain the following.

Example 2.6 (eigenvalue decomposition). Let $2 \leqslant N \in \mathbb{N}$. We equip $\mathsf{H}^N(\mathbb{K})$ with the scalar product

$$\langle \cdot \mid \cdot \rangle \colon (X, Y) \mapsto \operatorname{Re}\operatorname{tra}(XY) \tag{2.23}$$

and let P^N act on $\mathbb{R}^N$ via matrix-vector multiplication. For every $X \in \mathsf{H}^N(\mathbb{K})$, we denote by $\lambda(X) = (\lambda_1(X), \ldots, \lambda_N(X))$ the vector of the N (not necessarily distinct) eigenvalues of X listed in decreasing order; see [47, Theorem 5.3.6(c)] for the quaternion case. Additionally, set

$$\big(\forall U \in \mathsf{U}^N(\mathbb{K})\big) \quad \Lambda_U \colon \mathbb{R}^N \to \mathsf{H}^N(\mathbb{K}) \colon x \mapsto U(\operatorname{Diag} x)U^*. \tag{2.24}$$

Then $\mathfrak{S} = (\mathbb{R}^N, \mathsf{P}^N, \lambda, (\Lambda_U)_{U \in \mathsf{U}^N(\mathbb{K})})$ is a spectral decomposition system for $\mathsf{H}^N(\mathbb{K})$.

We now turn to a setting which has been extensively used in robust principal component analysis and signal processing; see for instance [10, 11, 12, 31].

Example 2.7 (singular value decomposition). Let M and N be strictly positive integers and set $m = \min\{M, N\}$. Let $\mathfrak{H}$ be the Euclidean space obtained by equipping $\mathbb{K}^{M\times N}$ with the scalar product

$$(X, Y) \mapsto \operatorname{Re}\operatorname{tra}(X^*Y), \tag{2.25}$$

and let $\mathsf{P}^m_\pm$ act on $\mathbb{R}^m$ via matrix-vector multiplication. Given a matrix $X \in \mathfrak{H}$, the vector in $\mathbb{R}^m_+$ formed by the m (not necessarily distinct) singular values of X, with the convention that they are listed in decreasing order, is denoted by $(\sigma_1(X), \ldots, \sigma_m(X))$; see [47, Proposition 3.2.5(f)] for singular value decomposition of matrices in $\mathbb{H}^{M\times N}$. This thus defines a mapping

$$\sigma \colon \mathfrak{H} \to \mathbb{R}^m \colon X \mapsto \big(\sigma_1(X), \ldots, \sigma_m(X)\big). \tag{2.26}$$

Further, set $\mathcal{A} = \mathsf{U}^M(\mathbb{K}) \times \mathsf{U}^N(\mathbb{K})$ and

$$\big(\forall \mathscr{a} = (U, V) \in \mathcal{A}\big) \quad \Lambda_{\mathscr{a}} \colon \mathbb{R}^m \to \mathfrak{H} \colon x \mapsto U(\operatorname{Diag} x)V^*, \tag{2.27}$$

where the operator $\operatorname{Diag} \colon \mathbb{R}^m \to \mathfrak{H}$ maps a vector $(\xi_i)_{1\leqslant i\leqslant m}$ to the diagonal matrix in $\mathfrak{H}$ of which the diagonal entries are $\xi_1, \ldots, \xi_m$. Then $\mathfrak{S} = (\mathbb{R}^m, \mathsf{P}^m_\pm, \sigma, (\Lambda_{\mathscr{a}})_{\mathscr{a}\in\mathcal{A}})$ is a spectral decomposition system for $\mathfrak{H}$.

Proof. We derive from (2.25), [47, properties (c) and (d), p. 30], and [22, Proposition V.2.1(i)] that

$$\begin{aligned}
\big(\forall \mathscr{a} = (U, V) \in \mathcal{A}\big)(\forall x \in \mathbb{R}^m) \quad \|\Lambda_{\mathscr{a}} x\|^2 &= \operatorname{Re}\operatorname{tra}\Big(V\big(\operatorname{Diag} x\big)^*U^*U\big(\operatorname{Diag} x\big)V^*\Big)\\
&= \operatorname{Re}\operatorname{tra}\Big(V^*V\big(\operatorname{Diag} x\big)^*\big(\operatorname{Diag} x\big)\Big)\\
&= \operatorname{Re}\operatorname{tra}\Big(\big(\operatorname{Diag} x\big)^*\big(\operatorname{Diag} x\big)\Big)\\
&= \|x\|_2^2.
\end{aligned} \tag{2.28}$$

This confirms that the linear operators $(\Lambda_{\mathscr{a}})_{\mathscr{a}\in\mathcal{A}}$ are isometries. To verify property [B] in Definition 2.1, set $\tau \colon \mathbb{R}^m \to \mathbb{R}^m \colon x \mapsto |x|^\downarrow$. It is evident that τ is $\mathsf{P}^m_\pm$-invariant and $(\forall x \in \mathbb{R}^m)$ $\tau(x) \in \mathsf{P}^m_\pm \cdot x$. In addition, it follows from (2.27) and the uniqueness of singular values (see [47, Proposition 3.2.5(f)] for the quaternion case) that

$$\big(\forall \mathscr{a} = (U, V) \in \mathcal{A}\big)(\forall x \in \mathbb{R}^m) \quad \sigma(\Lambda_{\mathscr{a}} x) = \sigma\big(\operatorname{Diag} x\big) = |x|^\downarrow = \tau(x). \tag{2.29}$$

Finally, observe that, in the present setting, property [C] in Definition 2.1 is precisely the singular value decomposition of matrices in $\mathfrak{H}$, and property [D] in Definition 2.1 is precisely the von Neumann trace inequality; see [27, Theorem 7.4.1.1] for the real and complex cases, and [11, Lemma 3] for the quaternion case. □

We next consider a setting that arises in the study of isotropic stored energy functions in nonlinear elasticity [2, 17, 48], as well as the study of existence of a matrix with prescribed singular values and main diagonal elements [53].

Example 2.8 (signed singular value decomposition). Let $2 \leqslant N \in \mathbb{N}$ and let $\mathfrak{H}$ be the Euclidean space obtained by equipping $\mathbb{R}^{N\times N}$ with the scalar product

$$(X, Y) \mapsto \operatorname{tra}(X^{\mathsf{T}} Y), \tag{2.30}$$

let S be the subgroup of $\mathsf{P}^N_{\pm}$ which consists of all matrices with an even number of entries equal to -1, and let S act on $\mathbb{R}^N$ via matrix-vector multiplication. As in Example 2.7, $\sigma(X) = (\sigma_1(X), \ldots, \sigma_N(X))$ designates the vector of the N singular values of a matrix $X \in \mathfrak{H}$, with the convention that $\sigma_1(X) \geqslant \cdots \geqslant \sigma_N(X)$. Define a mapping

$$\gamma\colon \mathfrak{H} \to \mathbb{R}^N : X \mapsto \big(\gamma_1(X), \ldots, \gamma_N(X)\big) \tag{2.31}$$

by

$$(\forall X \in \mathfrak{H})\big(\forall i \in \{1, \ldots, N\}\big) \quad \gamma_i(X) = \begin{cases} \sigma_i(X) & \text{if } 1 \leqslant i \leqslant N-1, \\ \sigma_N(X)\operatorname{sign}(\det X) & \text{if } i = N. \end{cases} \tag{2.32}$$

Finally, set $\mathcal{A} = \mathrm{SO}^N \times \mathrm{SO}^N$ and

$$\big(\forall a = (U, V) \in \mathcal{A}\big) \quad \Lambda_a : \mathbb{R}^N \to \mathfrak{H} : x \mapsto U(\operatorname{Diag} x)V^{\mathsf{T}}. \tag{2.33}$$

Then $\mathfrak{S} = (\mathbb{R}^N, \mathsf{S}, \gamma, (\Lambda_a)_{a\in\mathcal{A}})$ is a spectral decomposition system for $\mathfrak{H}$.

Proof. It is clear that the operators $(\Lambda_a)_{a\in\mathcal{A}}$ are linear isometries. Next, to verify property [B] in Definition 2.1, we define a mapping $\tau\colon \mathbb{R}^N \to \mathbb{R}^N$ as follows: for every $x = (\xi_i)_{1\leqslant i\leqslant N} \in \mathbb{R}^N$, let $(\mu_i)_{1\leqslant i\leqslant N} = |x|^{\downarrow}$ and set $\tau(x) = (\nu_i)_{1\leqslant i\leqslant N}$, where

$$\big(\forall i \in \{1, \ldots, N\}\big) \quad \nu_i = \begin{cases} \mu_i & \text{if } 1 \leqslant i \leqslant N-1, \\ \mu_N \operatorname{sign}(\xi_1 \cdots \xi_N) & \text{if } i = N. \end{cases} \tag{2.34}$$

One can check that τ is S-invariant and $(\forall x \in \mathbb{R}^N)\ \tau(x) \in \mathsf{S}\cdot x$. The condition

$$(\forall a \in \mathcal{A}) \quad \gamma \circ \Lambda_a = \tau \tag{2.35}$$

in property [B] in Definition 2.1 follows from the orthogonal invariance of σ and the construction of γ. Moreover, in the present setting, property [C] in Definition 2.1 reads

$$\big(\forall X \in \mathbb{R}^{N\times N}\big)\big(\exists\, U \in \mathrm{SO}^N\big)\big(\exists\, V \in \mathrm{SO}^N\big) \quad X = U\big(\operatorname{Diag}\gamma(X)\big)V^{\mathsf{T}}, \tag{2.36}$$

which is an easy consequence of the singular value decomposition theorem. Finally, property [D] in Definition 2.1 follows from [48, Lemma 1.3]. □

Our last example of this section describes a construction of spectral decomposition systems from existing ones.

Example 2.9. Let $\mathfrak{H}$ and $\mathcal{U}$ be Euclidean spaces, let $\mathfrak{S} = (\mathcal{X}, \mathsf{S}, \gamma, (\Lambda_{a})_{a\in\mathscr{A}})$ be a spectral decomposition system for $\mathfrak{H}$, let S act on $\mathcal{X} \oplus \mathcal{U}$ via $(\mathsf{s}, (x, u)) \mapsto (\mathsf{s} \cdot x, u)$, and define

$$\begin{cases} \boldsymbol{\gamma}\colon \mathfrak{H} \oplus \mathcal{U} \to \mathcal{X} \oplus \mathcal{U}\colon (X, u) \mapsto \big(\gamma(X), u\big) \\ (\forall a \in \mathscr{A}) \quad \boldsymbol{\Lambda}_{a}\colon \mathcal{X} \oplus \mathcal{U} \to \mathfrak{H} \oplus \mathcal{U}\colon (x, u) \mapsto (\Lambda_{a} x, u). \end{cases} \tag{2.37}$$

Then $(\mathcal{X} \oplus \mathcal{U}, \mathsf{S}, \boldsymbol{\gamma}, (\boldsymbol{\Lambda}_{a})_{a\in\mathscr{A}})$ is a spectral decomposition system for $\mathfrak{H} \oplus \mathcal{U}$.

3. Properties of spectral and spectral-induced ordering mappings

We present in this section properties of spectral and spectral-induced ordering mappings that will assist us in our analysis. We shall operate under the umbrella of the following assumption.

Assumption 3.1. $\mathfrak{H}$ is a Euclidean space, $\mathfrak{S} = (\mathcal{X}, \mathsf{S}, \gamma, (\Lambda_{a})_{a\in\mathscr{A}})$ is a spectral decomposition system for $\mathfrak{H}$, and $\tau\colon \mathcal{X} \to \mathcal{X}$ is the spectral-induced ordering mapping of the system $\mathfrak{S}$, that is, τ is S-invariant and satisfies

$$(\forall x \in \mathcal{X}) \quad \tau(x) \in \mathsf{S} \cdot x \tag{3.1}$$

and

$$(\forall a \in \mathscr{A}) \quad \gamma \circ \Lambda_{a} = \tau. \tag{3.2}$$

The following simple observation will be employed frequently.

Lemma 3.2. *Let $\mathcal{H}$ be a Euclidean space, let G be a group which acts on $\mathcal{H}$ by linear isometries, and let $\mathsf{g} \in \mathsf{G}$. Then the adjoint of the operator $L_{\mathsf{g}}\colon \mathcal{H} \to \mathcal{H}\colon x \mapsto \mathsf{g} \cdot x$ is given by $\mathcal{H} \to \mathcal{H}\colon x \mapsto \mathsf{g}^{-1} \cdot x$.*

Proof. Since L_{g} is a linear isometry, we have $L_{\mathsf{g}}^{*} \circ L_{\mathsf{g}} = \mathrm{Id}$. Hence

$$(\forall x \in \mathcal{H}) \quad \mathsf{g}^{-1} \cdot x = L_{\mathsf{g}}^{*}\big(L_{\mathsf{g}}(\mathsf{g}^{-1} \cdot x)\big) = L_{\mathsf{g}}^{*}\big(\mathsf{g} \cdot (\mathsf{g}^{-1} \cdot x)\big) = L_{\mathsf{g}}^{*} x, \tag{3.3}$$

as claimed. ☐

We first establish basic properties of spectral-induced ordering mappings.

Proposition 3.3. *Suppose that Assumption 3.1 is in force, and let x and y be in $\mathcal{X}$. Then the following hold:*

(i) *τ is idempotent, that is, $\tau \circ \tau = \tau$.*
(ii) *$(\forall a \in \mathscr{A})$ $\|\Lambda_{a} x\| = \|x\| = \|\tau(x)\|$.*
(iii) *$\langle x \mid y\rangle \leqslant \langle \tau(x) \mid \tau(y)\rangle$.*
(iv) *$\langle \tau(x) \mid \tau(y)\rangle = \max_{\mathsf{s}\in\mathsf{S}} \langle \mathsf{s} \cdot x \mid y\rangle$.*
(v) *$\|\tau(x) - \tau(y)\| \leqslant \|x - y\|$.*
(vi) *$\operatorname{range} \tau$ is a closed convex cone in $\mathcal{X}$.*
(vii) *The orbit $\mathsf{S} \cdot x$ is compact.*

Proof. (i): A consequence of (3.1) and the S-invariance of τ.

(ii): Elementary.

(iii): Take $\mathscr{a} \in \mathscr{A}$ and note that $\Lambda_{\mathscr{a}}^* \circ \Lambda_{\mathscr{a}} = \mathrm{Id}_{\mathcal{X}}$. We derive from property [D] in Definition 2.1 and (3.2) that $\langle x \mid y \rangle = \langle \Lambda_{\mathscr{a}} x \mid \Lambda_{\mathscr{a}} y \rangle \leqslant \langle \gamma(\Lambda_{\mathscr{a}} x) \mid \gamma(\Lambda_{\mathscr{a}} y) \rangle = \langle \tau(x) \mid \tau(y) \rangle$.

(iv): Use (iii), the S-invariance of τ, (3.1), and Lemma 3.2.

(v): Combine (iii) and (ii).

(vi): We infer from (3.1) and (iii) that $(\mathcal{X}, \mathrm{S}, \tau)$ is a normal decomposition system in the sense of [35, Definition 2.1], and then conclude via [35, Theorem 2.4] that range τ is a closed convex cone in $\mathcal{X}$.

(vii): Observe that the orbit $\mathrm{S}{\cdot}x$ is bounded because $(\forall \mathsf{s} \in \mathrm{S})\ \|\mathsf{s} \cdot x\| = \|x\|$. Hence, it remains to show that $\mathrm{S}{\cdot}x$ is closed. To do so, let $(\mathsf{s}_n)_{n\in\mathbb{N}}$ be a sequence in S such that the sequence $(\mathsf{s}_n{\cdot}x)_{n\in\mathbb{N}}$ converges to some $z \in \mathcal{X}$. Using the S-invariance of τ and (v), we deduce that $\|\tau(x) - \tau(z)\| = \|\tau(\mathsf{s}_n \cdot x) - \tau(z)\| \leqslant \|\mathsf{s}_n \cdot x - z\| \to 0$ and, thus, that $\tau(x) = \tau(z)$. Therefore, (3.1) yields $z \in \mathrm{S} \cdot \tau(z) = \mathrm{S} \cdot \tau(x) = \mathrm{S} \cdot x$. ☐

Proposition 3.3 (iii) allows us to provide another example of a spectral decomposition system.

Example 3.4. Suppose that Assumption 3.1 is in force and define

$$(\forall \mathsf{s} \in \mathrm{S}) \quad \pi_{\mathsf{s}} \colon \mathcal{X} \to \mathcal{X} \colon x \mapsto \mathsf{s} \cdot x. \tag{3.4}$$

Then $(\mathcal{X}, \mathrm{S}, \tau, (\pi_{\mathsf{s}})_{\mathsf{s}\in\mathrm{S}})$ is a spectral decomposition system for $\mathcal{X}$.

Our next result provides basic properties of spectral mappings.

Proposition 3.5. *Suppose that Assumption 3.1 is in force. Then the following hold:*

(i) $\tau \circ \gamma = \gamma$ *and* $(\forall \mathscr{a} \in \mathscr{A})\ \gamma \circ \Lambda_{\mathscr{a}} \circ \gamma = \gamma$.

(ii) range γ = range τ *is a closed convex cone in* $\mathcal{X}$.

(iii) $(\forall X \in \mathfrak{H})(\forall \mathscr{a} \in \mathscr{A})\ \|X\| = \|\Lambda_{\mathscr{a}} \gamma(X)\| = \|\gamma(X)\|$.

(iv) γ *is nonexpansive, that is,* $(\forall X \in \mathfrak{H})(\forall Y \in \mathfrak{H})\ \|\gamma(X) - \gamma(Y)\| \leqslant \|X - Y\|$.

(v) $(\forall X \in \mathfrak{H})(\forall \alpha \in \mathbb{R}_+)\ \gamma(\alpha X) = \alpha\gamma(X)$.

(vi) $(\forall X \in \mathfrak{H})\ \gamma(-X) \in -\mathrm{S} \cdot \gamma(X)$.

Proof. (i): In view of (3.2), it is enough to show that $\tau \circ \gamma = \gamma$. Take $X \in \mathfrak{H}$ and $\mathscr{a} \in \mathscr{A}_X$. By (3.2), $\tau = \gamma \circ \Lambda_{\mathscr{a}}$. However, by the very definition of $\mathscr{A}_X$, $X = \Lambda_{\mathscr{a}}\gamma(X)$. Hence

$$\tau\big(\gamma(X)\big) = (\gamma \circ \Lambda_{\mathscr{a}})\big(\gamma(X)\big) = \gamma\big(\Lambda_{\mathscr{a}}\gamma(X)\big) = \gamma(X). \tag{3.5}$$

(ii): It follows from (i) that range $\gamma \subset$ range τ, while (3.2) yields range $\tau \subset$ range γ. Thus range γ = range τ, which is a closed convex cone in $\mathcal{X}$ thanks to Proposition 3.3 (vi).

(iii): This follows from the fact that the operators $(\Lambda_{\mathscr{a}})_{\mathscr{a}\in\mathscr{A}}$ are linear isometries, and from the spectral decomposition of elements in $\mathfrak{H}$; see property [C] in Definition 2.1.

(iv): Combine property [D] in Definition 2.1 and (iii).

(v): For every $X \in \mathfrak{H}$ and every $\alpha \in \mathbb{R}_+$, we derive from (iii) and property [D] in Definition 2.1 that

$$\begin{aligned}\|\gamma(\alpha X) - \alpha\gamma(X)\|^2 &= \|\gamma(\alpha X)\|^2 - 2\alpha\langle \gamma(\alpha X) \mid \gamma(X) \rangle + \alpha^2\|\gamma(X)\|^2 \\ &\leqslant \|\alpha X\|^2 - 2\alpha\langle \alpha X \mid X \rangle + \alpha^2\|X\|^2 \\ &= 0,\end{aligned} \tag{3.6}$$

which leads to $\gamma(\alpha X) = \alpha\gamma(X)$.

(vi): For every $X \in \mathfrak{H}$ and every $\mathscr{a} \in \mathscr{A}_X$, using the linearity of $\Lambda_{\mathscr{a}}$, together with (3.2) and (3.1), we get

$$\gamma(-X) = \gamma\big(-\Lambda_{\mathscr{a}}\gamma(X)\big) = (\gamma \circ \Lambda_{\mathscr{a}})\big(-\gamma(X)\big) = \tau\big(-\gamma(X)\big) \in -\mathrm{S} \cdot \gamma(X), \tag{3.7}$$

which concludes the proof. ☐

We now establish necessary and sufficient conditions for equality in property [D] in Definition 2.1.

Proposition 3.6. *Suppose that Assumption 3.1 is in force, let $\mathscr{D}$ be a nonempty subset of $\mathfrak{H}$, and set*

$$\mathscr{K} = \left\{ \sum_{i=1}^{m} \alpha_i X_i \;\middle|\; m \in \mathbb{N} \smallsetminus \{0\},\ (X_i)_{1 \leqslant i \leqslant m} \in \mathscr{D}^m,\ \textit{and } (\alpha_i)_{1 \leqslant i \leqslant m} \in \mathbb{R}_+^m \right\}. \tag{3.8}$$

Then the following are equivalent:

(i) $(\forall X \in \mathscr{D})(\forall Y \in \mathscr{D})\ \langle X \mid Y \rangle = \langle \gamma(X) \mid \gamma(Y) \rangle$.

(ii) $(\forall X \in \mathscr{D})(\forall Y \in \mathscr{D})\ \|X - Y\| = \|\gamma(X) - \gamma(Y)\|$.

(iii) *There exists* $\mathscr{a} \in \mathscr{A}$ *such that* $(\forall X \in \mathscr{D})\ X = \Lambda_{\mathscr{a}}\gamma(X)$.

(iv) *There exists* $\mathscr{a} \in \mathscr{A}$ *such that* $(\forall X \in \mathscr{K})\ X = \Lambda_{\mathscr{a}}\gamma(X)$.

Moreover, if one of the statements (i)–(iv) holds, then

$$\big(\forall m \in \mathbb{N} \smallsetminus \{0\}\big)\big(\forall (X_i)_{1 \leqslant i \leqslant m} \in \mathscr{D}^m\big)\big(\forall (\alpha_i)_{1 \leqslant i \leqslant m} \in \mathbb{R}_+^m\big) \quad \gamma\left(\sum_{i=1}^{m} \alpha_i X_i\right) = \sum_{i=1}^{m} \alpha_i \gamma(X_i). \tag{3.9}$$

Proof. (i) ⇔ (ii): This follows from Proposition 3.5 (iii).

(i) ⇒ (iii): We employ the techniques of the proof of [35, Theorem 2.2]. Take $Z \in \operatorname{ri}(\operatorname{conv}\mathscr{D})$ and $\mathscr{a} \in \mathscr{A}_Z$. Assume that there exists $X \in \mathscr{D}$ for which $X \neq \Lambda_{\mathscr{a}}\gamma(X)$. We get from Proposition 3.5 (iii) that

$$\begin{aligned} 2\langle X \mid \Lambda_{\mathscr{a}}\gamma(X) \rangle &= \|X\|^2 + \|\Lambda_{\mathscr{a}}\gamma(X)\|^2 - \|X - \Lambda_{\mathscr{a}}\gamma(X)\|^2 \\ &= 2\|X\|^2 - \|X - \Lambda_{\mathscr{a}}\gamma(X)\|^2 \\ &< 2\|X\|^2. \end{aligned} \tag{3.10}$$

On the other hand, since $Z \in \operatorname{ri}(\operatorname{conv}\mathscr{D})$ and $X \in \mathscr{D}$, it results from [45, Theorem 6.4] that there exist families $(\alpha_i)_{0 \leqslant i \leqslant n}$ in $]0,1[$ and $(X_i)_{1 \leqslant i \leqslant n}$ in $\mathscr{D}$ such that $\sum_{i=0}^{n} \alpha_i = 1$ and $Z = \alpha_0 X + \sum_{i=1}^{n} \alpha_i X_i$; additionally, by assumption,

$$\big(\forall i \in \{1, \ldots, n\}\big) \quad \langle \gamma(X_i) \mid \gamma(X) \rangle = \langle X_i \mid X \rangle. \tag{3.11}$$

Therefore, we derive from property [D] in Definition 2.1 and Proposition 3.5 (i) that

$$\begin{aligned}
\langle \gamma(Z) \mid \gamma(X) \rangle &= \langle \Lambda_a \gamma(Z) \mid \Lambda_a \gamma(X) \rangle \\
&= \langle Z \mid \Lambda_a \gamma(X) \rangle \\
&= \alpha_0 \langle X \mid \Lambda_a \gamma(X) \rangle + \sum_{i=1}^{n} \alpha_i \langle X_i \mid \Lambda_a \gamma(X) \rangle \\
&< \alpha_0 \|X\|^2 + \sum_{i=1}^{n} \alpha_i \big\langle \gamma(X_i) \,\big|\, \gamma\big(\Lambda_a \gamma(X)\big) \big\rangle \\
&= \alpha_0 \|X\|^2 + \sum_{i=1}^{n} \alpha_i \langle \gamma(X_i) \mid \gamma(X) \rangle \\
&= \alpha_0 \|X\|^2 + \sum_{i=1}^{n} \alpha_i \langle X_i \mid X \rangle \\
&= \langle Z \mid X \rangle \\
&\leqslant \langle \gamma(Z) \mid \gamma(X) \rangle,
\end{aligned} \tag{3.12}$$

which is impossible.

(iii) ⟹ (i): Use the fact that $\Lambda_a^* \circ \Lambda_a = \mathrm{Id}_{\mathcal{X}}$.

(iii) ⟹ (iv): Let $X \in \mathcal{K}$, say $X = \sum_{i=1}^{m} \alpha_i X_i$, where $(X_i)_{1 \leqslant i \leqslant m} \in \mathcal{D}^m$ and $(\alpha_i)_{1 \leqslant i \leqslant m} \in \mathbb{R}_+^m$. Then $(\forall i \in \{1, \ldots, m\})\ X_i = \Lambda_a \gamma(X_i)$. Thus, because Λ_a is an isometry,

$$\big(\forall i \in \{1, \ldots, m\}\big)\big(\forall j \in \{1, \ldots, m\}\big) \quad \|X_i - X_j\| = \|\Lambda_a \gamma(X_i) - \Lambda_a \gamma(X_j)\| = \|\gamma(X_i) - \gamma(X_j)\|. \tag{3.13}$$

In turn, upon setting $\alpha = \sum_{i=1}^{m} \alpha_i$, we derive from [5, Lemma 2.1(i)] and items (iii) and (iv) in Proposition 3.5 that

$$\begin{aligned}
\left\| \gamma(X) - \sum_{i=1}^{m} \alpha_i \gamma(X_i) \right\|^2 &= (1-\alpha)\|\gamma(X)\|^2 + \sum_{i=1}^{m} \alpha_i \|\gamma(X) - \gamma(X_i)\|^2 + (\alpha - 1) \sum_{i=1}^{m} \alpha_i \|\gamma(X_i)\|^2 \\
&\quad - \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \alpha_i \alpha_j \|\gamma(X_i) - \gamma(X_j)\|^2 \\
&\leqslant (1-\alpha)\|X\|^2 + \sum_{i=1}^{m} \alpha_i \|X - X_i\|^2 + (\alpha - 1) \sum_{i=1}^{m} \alpha_i \|X_i\|^2 \\
&\quad - \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \alpha_i \alpha_j \|X_i - X_j\|^2 \\
&= \left\| X - \sum_{i=1}^{m} \alpha_i X_i \right\|^2 \\
&= 0.
\end{aligned} \tag{3.14}$$

Hence $\gamma(X) = \sum_{i=1}^{m} \alpha_i \gamma(X_i)$, which verifies (3.9). Consequently, by linearity of Λ_a,

$$\Lambda_a \gamma(X) = \Lambda_a \left(\sum_{i=1}^{m} \alpha_i \gamma(X_i) \right) = \sum_{i=1}^{m} \alpha_i \Lambda_a \gamma(X_i) = \sum_{i=1}^{m} \alpha_i X_i = X. \tag{3.15}$$

(iv) ⟹ (iii): Clear. □

Remark 3.7. Some comments are in order.

(i) The equivalence (i) ⇔ (iii) in Proposition 3.6 corresponds to [35, Theorem 2.2] in the context of normal decomposition systems of Example 2.4, and to [24, Theorem 4.1] in the context of Euclidean Jordan algebras of Example 2.5.

(ii) In the settings of Examples 2.6, 2.7, and 2.8, where $\mathbb{K}$ is taken to be $\mathbb{R}$ or $\mathbb{C}$, and with the choice $\mathscr{D} = \{X, Y\}$, item (iii) in Proposition 3.6 reduces to the classical condition for equality in von Neumann-type trace inequalities; see [37, 48, 52, 55].

Central to our approach will be Theorem 3.9, which compares the spectrum of a sum and the sum of the spectra. It can be regarded as a generalization of Ky Fan's classical majorization theorem on the eigenvalues of a sum of Hermitian matrices; see Remark 3.10 for further discussion. Proving Theorem 3.9 necessitates the following technical result.

Proposition 3.8. *Suppose that Assumption 3.1 is in force, and let x and y be in $\mathcal{X}$. Then the following are equivalent:*

(i) $y \in \operatorname{conv}(\mathsf{S} \cdot x)$.

(ii) $(\forall z \in \mathcal{X})\ \langle y \mid z\rangle \leqslant \langle \tau(x) \mid \tau(z)\rangle$.

Moreover, if x and y lie in range γ*, then each of the statements (i) and (ii) is equivalent to*

(iii) $(\forall z \in \operatorname{range}\gamma)\ \langle y - x \mid z\rangle \leqslant 0$.

Proof. It results from Proposition 3.3 (vii) and [46, Corollary 2.30] that $\operatorname{conv}(\mathsf{S} \cdot x)$ is compact. In turn, we infer from [6, Proposition 11.1(iii)] and Proposition 3.3 (iv) that

$$(\forall z \in \mathcal{X}) \quad \sigma_{\operatorname{conv}(\mathsf{S}\cdot x)}(z) = \sigma_{\mathsf{S}\cdot x}(z) = \max_{\mathsf{s}\in\mathsf{S}} \langle \mathsf{s} \cdot x \mid z\rangle = \langle \tau(x) \mid \tau(z)\rangle. \tag{3.16}$$

Therefore, the equivalence (i) ⇔ (ii) follows from [6, Proposition 7.11].

To proceed further, let us assume that $\{x, y\} \subset \operatorname{range}\gamma$. Note that, for every $z \in \operatorname{range}\gamma$, thanks to Proposition 3.5 (ii) and Proposition 3.3 (i), we get $\tau(z) = z$; in particular, $\tau(x) = x$ and $\tau(y) = y$.

(ii) ⇒ (iii): Clear.

(iii) ⇒ (ii): For every $z \in \mathcal{X}$, because $\tau(z) \in \operatorname{range}\gamma$ by Proposition 3.5 (ii), we get from Proposition 3.3 (iii) that $\langle y \mid z\rangle \leqslant \langle y \mid \tau(z)\rangle \leqslant \langle x \mid \tau(z)\rangle = \langle \tau(x) \mid \tau(z)\rangle$. ☐

Theorem 3.9. *Suppose that Assumption 3.1 is in force, let $(X_i)_{1\leqslant i\leqslant m}$ be a family in $\mathfrak{H}$, and let $(\alpha_i)_{1\leqslant i\leqslant m}$ be a family in $\mathbb{R}_+$. Then*

$$\gamma\left(\sum_{i=1}^{m} \alpha_i X_i\right) \in \operatorname{conv}\left(\mathsf{S} \cdot \sum_{i=1}^{m} \alpha_i \gamma(X_i)\right). \tag{3.17}$$

Proof. Set $X = \sum_{i=1}^{m} \alpha_i X_i$ and let $\mathscr{a} \in \mathscr{A}_X$. We derive from property [D] in Definition 2.1 and Proposition 3.5 (i) that

$$(\forall Y \in \mathfrak{H}) \quad \langle \gamma(X) \mid \gamma(Y) \rangle = \langle \Lambda_a \gamma(X) \mid \Lambda_a \gamma(Y) \rangle$$
$$= \langle X \mid \Lambda_a \gamma(Y) \rangle$$
$$= \sum_{i=1}^{m} \alpha_i \langle X_i \mid \Lambda_a \gamma(Y) \rangle$$
$$\leqslant \sum_{i=1}^{m} \alpha_i \big\langle \gamma(X_i) \,\big|\, \gamma\big(\Lambda_a \gamma(Y)\big) \big\rangle$$
$$= \left\langle \sum_{i=1}^{m} \alpha_i \gamma(X_i) \,\middle|\, \gamma(Y) \right\rangle. \tag{3.18}$$

Therefore, inasmuch as $\sum_{i=1}^{m} \alpha_i \gamma(X_i) \in \operatorname{range} \gamma$ thanks to Proposition 3.5 (ii), the conclusion follows from the equivalence (i) ⇔ (iii) in Proposition 3.8 applied to $x = \sum_{i=1}^{m} \alpha_i \gamma(X_i)$ and $y = \gamma(X)$. □

Remark 3.10. Here are some noteworthy instances of Theorem 3.9.

(i) We infer from [36, Propositions 3.3 and 3.4] that the semisimple real Lie algebra setting of [51, Theorem 2] is a normal decomposition system (see Example 2.4), and we therefore recover [51, Theorem 2].

(ii) Specializing Theorem 3.9 to Example 2.5 yields

$$(\forall X_1 \in \mathfrak{H})(\forall X_2 \in \mathfrak{H}) \quad \lambda(X_1 + X_2) \in \operatorname{conv}\Big(\mathsf{P}^N \cdot \big(\lambda(X_1) + \lambda(X_2)\big)\Big), \tag{3.19}$$

which improves [30, Proposition 8]. In particular, by choosing $\mathfrak{H} = \mathsf{H}^N(\mathbb{K})$ (see Example 2.6), we obtain

$$\big(\forall X_1 \in \mathsf{H}^N(\mathbb{K})\big)\big(\forall X_2 \in \mathsf{H}^N(\mathbb{K})\big) \quad \lambda(X_1 + X_2) \in \operatorname{conv}\Big(\mathsf{P}^N \cdot \big(\lambda(X_1) + \lambda(X_2)\big)\Big), \tag{3.20}$$

which is Ky Fan's majorization theorem for the eigenvalues of a sum of Hermitian matrices in the real and complex cases [40, Theorem 9.G.1].

(iii) In the setting of Example 2.7, we recover from Theorem 3.9 the inclusion

$$\big(\forall X_1 \in \mathbb{C}^{M \times N}\big)\big(\forall X_2 \in \mathbb{C}^{M \times N}\big) \quad \sigma(X_1 + X_2) \in \operatorname{conv}\Big(\mathsf{P}_{\pm}^m \cdot \big(\sigma(X_1) + \sigma(X_2)\big)\Big), \tag{3.21}$$

which is Ky Fan's weak majorization theorem for the singular values of a sum of rectangular matrices [40, 9.G.1.d].

4. Spectral functions and their basic properties

Consider the setting of Assumption 3.1. Of prime interest in applications in the context of Examples 2.5, 2.6, 2.7, and 2.8 is the class of functions $\Phi\colon \mathfrak{H} \to \left]-\infty, +\infty\right]$ whose values are "spectrally invariant", i.e., for which $\Phi(X) = \Phi(Y)$ whenever $\gamma(X) = \gamma(Y)$. We follow customary convention and call such functions *spectral functions*. For instance, in the setting of Hermitian matrices (Example 2.6 with $\mathbb{K} = \mathbb{C}$), these are functions that are unitarily invariant; see, e.g., [34, p. 164]. The following result shows that, under Assumption 3.1, spectral functions are precisely those of the form $\varphi \circ \gamma$ where $\varphi\colon \mathcal{X} \to [-\infty, +\infty]$ is S-invariant.

Proposition 4.1. *Suppose that Assumption 3.1 is in force and let $\Phi\colon \mathfrak{H} \to [-\infty,+\infty]$. Then the following are equivalent:*

(i) *Φ is a* spectral function *in the sense of*

$$(\forall X \in \mathfrak{H})(\forall Y \in \mathfrak{H}) \quad \gamma(X) = \gamma(Y) \quad \Rightarrow \quad \Phi(X) = \Phi(Y). \tag{4.1}$$

(ii) $(\forall \mathscr{a} \in \mathscr{A})(\forall \mathscr{b} \in \mathscr{A})\ \Phi \circ \Lambda_{\mathscr{a}} = \Phi \circ \Lambda_{\mathscr{b}}$.

(iii) *There exists an* S*-invariant function $\varphi\colon \mathcal{X} \to [-\infty,+\infty]$ such that $\Phi = \varphi \circ \gamma$.*

Moreover, if (iii) holds, then $(\forall \mathscr{a} \in \mathscr{A})\ \varphi = \Phi \circ \Lambda_{\mathscr{a}}$.

Proof. (i)⇒(ii): Let $\mathscr{a}$ and $\mathscr{b}$ be in $\mathscr{A}$. For every $x \in \mathcal{X}$, since (3.2) yields $\gamma(\Lambda_{\mathscr{a}}x) = \tau(x) = \gamma(\Lambda_{\mathscr{b}}x)$, it results from (4.1) that $\Phi(\Lambda_{\mathscr{a}}x) = \Phi(\Lambda_{\mathscr{b}}x)$.

(ii)⇒(iii): Fix $\mathscr{b} \in \mathscr{A}$ and set $\varphi = \Phi \circ \Lambda_{\mathscr{b}}$. For every $X \in \mathfrak{H}$ and every $\mathscr{a} \in \mathscr{A}_X$, we have

$$\varphi\big(\gamma(X)\big) = (\Phi \circ \Lambda_{\mathscr{b}})\big(\gamma(X)\big) = (\Phi \circ \Lambda_{\mathscr{a}})\big(\gamma(X)\big) = \Phi\big(\Lambda_{\mathscr{a}}\gamma(X)\big) = \Phi(X), \tag{4.2}$$

which confirms that $\varphi \circ \gamma = \Phi$. In turn, appealing to (3.2), we obtain $\varphi = \Phi \circ \Lambda_{\mathscr{b}} = \varphi \circ (\gamma \circ \Lambda_{\mathscr{b}}) = \varphi \circ \tau$. Hence, the S-invariance of φ follows from that of τ.

(iii)⇒(i): Clear.

Finally, if (iii) holds, then we derive from (3.2), (3.1), and the S-invariance of φ that $(\forall \mathscr{a} \in \mathscr{A})$ $\Phi \circ \Lambda_{\mathscr{a}} = \varphi \circ (\gamma \circ \Lambda_{\mathscr{a}}) = \varphi \circ \tau = \varphi$. □

In the light of Proposition 4.1, the following notion is well defined.

Definition 4.2. Suppose that Assumption 3.1 is in force and let $\Phi\colon \mathfrak{H} \to [-\infty,+\infty]$ be a spectral function. The unique S-invariant function $\varphi\colon \mathcal{X} \to [-\infty,+\infty]$ such that $\Phi = \varphi \circ \gamma$ is called the *invariant function associated with Φ*.

Considering the class of spectral indicator functions leads to the notion of a *spectral set*.

Corollary 4.3. *Suppose that Assumption 3.1 is in force and let $\mathfrak{D}$ be a subset of $\mathfrak{H}$. Then the following are equivalent:*

(i) *$\mathfrak{D}$ is a* spectral set *in the sense that*

$$(\forall X \in \mathfrak{H})(\forall Y \in \mathfrak{H}) \quad \begin{cases} \gamma(X) = \gamma(Y) \\ X \in \mathfrak{D} \end{cases} \quad \Rightarrow \quad Y \in \mathfrak{D}. \tag{4.3}$$

(ii) *There exists an* S*-invariant subset D of $\mathcal{X}$ such that $\mathfrak{D} = \gamma^{-1}(D)$.*

Moreover, if (ii) holds, then $(\forall \mathscr{a} \in \mathscr{A})\ D = \Lambda_{\mathscr{a}}^{-1}(\mathfrak{D})$.

Proof. Apply Proposition 4.1 to $\Phi = \iota_{\mathfrak{D}}$. □

Let $\mathcal{H}$ be a Euclidean space and let $f\colon \mathcal{H} \to [-\infty,+\infty]$. The lower semicontinuous envelope of f is defined by

$$\overline{f} = \sup\big\{g\colon \mathcal{H} \to [-\infty,+\infty] \mid g \text{ is lower semicontinuous and } g \leqslant f\big\} \tag{4.4}$$

and, on account of [6, Lemma 1.32(iv)] and [46, Lemma 1.7], we have

$$(\forall x \in \mathcal{H}) \quad \overline{f}(x) = \min\Big\{\xi \in [-\infty, +\infty] \;\Big|\; \big(\exists\, (x_n)_{n\in\mathbb{N}} \in \mathcal{H}^{\mathbb{N}}\big)\ x_n \to x \text{ and } f(x_n) \to \xi\Big\}; \tag{4.5}$$

in particular, for every subset C of $\mathcal{H}$, $\overline{\iota_C} = \iota_{\overline{C}}$. Next, the function f is said to be convex if

$$(\forall x \in \operatorname{dom} f)(\forall y \in \operatorname{dom} f)(\forall \alpha \in \,]0,1[) \quad f\big(\alpha x + (1-\alpha)y\big) \leqslant \alpha f(x) + (1-\alpha) f(y), \tag{4.6}$$

where $\operatorname{dom} f = \{x \in \mathcal{H} \mid f(x) < +\infty\}$ is the domain of f, and f is said to be proper if $-\infty \notin f(\mathcal{H})$ and $\operatorname{dom} f \neq \varnothing$. Finally, the set of all proper lower semicontinuous convex functions from $\mathcal{H}$ to $]-\infty, +\infty]$ is denoted by $\Gamma_0(\mathcal{H})$.

Below is a characterization of invariant functions that will be utilized repeatedly.

Proposition 4.4. *Suppose that Assumption 3.1 is in force and let $\varphi\colon \mathcal{X} \to [-\infty, +\infty]$. Then the following hold:*

(i) *φ is S-invariant $\Leftrightarrow$ $\varphi \circ \tau = \varphi$ $\Leftrightarrow$ $(\forall \mathscr{a} \in \mathscr{A})$ $\varphi \circ \gamma \circ \Lambda_{\mathscr{a}} = \varphi$.*

(ii) *Suppose that φ is S-invariant. Then $\operatorname{int}\operatorname{dom}(\varphi \circ \gamma) = \gamma^{-1}(\operatorname{int}\operatorname{dom}\varphi)$.*

Proof. (i): Use (3.1), the S-invariance of τ, and (3.2).

(ii): Suppose that $X \in \operatorname{int}\operatorname{dom}(\varphi \circ \gamma)$ and let $\varepsilon \in \,]0, +\infty[$ be such that the closed ball $B(X;\varepsilon)$ is contained in $\operatorname{dom}(\varphi \circ \gamma)$. We must show that $\gamma(X) \in \operatorname{int}\operatorname{dom}\varphi$. To do so, let $y \in B(\gamma(X);\varepsilon)$ and $\mathscr{a} \in \mathscr{A}_X$. Then $\|\Lambda_{\mathscr{a}} y - X\| = \|\Lambda_{\mathscr{a}} y - \Lambda_{\mathscr{a}}\gamma(X)\| = \|y - \gamma(X)\| \leqslant \varepsilon$, which implies that $\Lambda_{\mathscr{a}} y \in B(X;\varepsilon) \subset \operatorname{dom}(\varphi \circ \gamma)$. Hence, we get from (i) that $\varphi(y) = (\varphi \circ \gamma)(\Lambda_{\mathscr{a}} y) < +\infty$. Therefore $B(\gamma(X);\varepsilon) \subset \operatorname{dom}\varphi$ or, equivalently, $\gamma(X) \in \operatorname{int}\operatorname{dom}\varphi$. The reverse inclusion is proved similarly. ☐

Next, we study lower semicontinuity and convexity of spectral functions in terms of the associated invariant functions.

Proposition 4.5. *Suppose that Assumption 3.1 is in force and let $\varphi\colon \mathcal{X} \to [-\infty, +\infty]$ be S-invariant. Then the following hold:*

(i) *$\overline{\varphi \circ \gamma} = \overline{\varphi} \circ \gamma$.*

(ii) *Let $X \in \mathfrak{H}$. Then $\varphi \circ \gamma$ is lower semicontinuous at X if and only if φ is lower semicontinuous at $\gamma(X)$.*

(iii) *$\varphi \circ \gamma$ is lower semicontinuous if and only if φ is lower semicontinuous.*

Proof. (i): Take $X \in \mathfrak{H}$ and let $(X_n)_{n\in\mathbb{N}}$ be a sequence in $\mathfrak{H}$ such that $X_n \to X$ and $\overline{\varphi \circ \gamma}(X) = \lim(\varphi \circ \gamma)(X_n)$. Proposition 3.5 (iv) yields $\gamma(X_n) \to \gamma(X)$ and, thus, (4.5) implies that

$$\overline{\varphi \circ \gamma}(X) = \lim \varphi\big(\gamma(X_n)\big) \geqslant \overline{\varphi}\big(\gamma(X)\big). \tag{4.7}$$

Now let $(x_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{X}$ such that $x_n \to \gamma(X)$ and $\overline{\varphi}(\gamma(X)) = \lim \varphi(x_n)$, and take $\mathscr{a} \in \mathscr{A}_X$. Since $\|\Lambda_{\mathscr{a}} x_n - X\| = \|\Lambda_{\mathscr{a}} x_n - \Lambda_{\mathscr{a}}\gamma(X)\| = \|x_n - \gamma(X)\| \to 0$, we deduce from Proposition 4.4 (i) and (4.5) that

$$\overline{\varphi}\big(\gamma(X)\big) = \lim \varphi(x_n) = \lim(\varphi \circ \gamma)(\Lambda_{\mathscr{a}} x_n) \geqslant \overline{\varphi \circ \gamma}(X). \tag{4.8}$$

Consequently, combining (4.7) and (4.8) yields the conclusion.

(ii): We derive from [6, Lemma 1.32(v)] and (i) that

$$\begin{aligned}\varphi\circ\gamma \text{ is lower semicontinuous at } X &\Leftrightarrow \overline{\varphi\circ\gamma}(X)=(\varphi\circ\gamma)(X)\\ &\Leftrightarrow \overline{\varphi}\big(\gamma(X)\big)=\varphi\big(\gamma(X)\big)\\ &\Leftrightarrow \varphi \text{ is lower semicontinuous at } \gamma(X),\end{aligned}\tag{4.9}$$

as announced.

(iii): Suppose that $\varphi\circ\gamma$ is lower semicontinuous. For every sequence $(x_n)_{n\in\mathbb{N}}$ in $\mathcal{X}$ which converges to some $x\in\mathcal{X}$ and every $a\in\mathsf{A}$, because $\|\Lambda_a x_n-\Lambda_a x\|=\|x_n-x\|\to 0$, we derive from Proposition 4.4 (i) that $\underline{\lim}\,\varphi(x_n)=\underline{\lim}(\varphi\circ\gamma)(\Lambda_a x_n)\geqslant(\varphi\circ\gamma)(\Lambda_a x)=\varphi(x)$. As a result, φ is lower semicontinuous. The reverse implication is a consequence of (ii). ☐

The following result characterizing convexity of spectral functions is the main result of this section.

Theorem 4.6. *Suppose that Assumption 3.1 is in force and let $\varphi\colon\mathcal{X}\to\,]-\infty,+\infty]$ be* S*-invariant. Then the following hold:*

(i) *$\varphi\circ\gamma$ is convex if and only if φ is convex.*

(ii) *Suppose that φ is proper. Then $\varphi\circ\gamma\in\Gamma_0(\mathfrak{H})$ if and only if $\varphi\in\Gamma_0(\mathcal{X})$.*

Proof. (i): If $\varphi\circ\gamma$ is convex, then, upon taking $a\in\mathsf{A}$, we infer from Proposition 4.4 (i) that $\varphi=(\varphi\circ\gamma)\circ\Lambda_a$ is convex as the composition of a convex function and a linear operator. Conversely, suppose that φ is convex, and that X and Y lie in $\operatorname{dom}(\varphi\circ\gamma)$, and take $\alpha\in\,]0,1[$. On account of Theorem 3.9, we obtain finite families $(\mathsf{s}_i)_{i\in I}$ in S and $(\alpha_i)_{i\in I}$ in $]0,1]$ such that

$$\sum_{i\in I}\alpha_i=1\quad\text{and}\quad\gamma\big(\alpha X+(1-\alpha)Y\big)=\sum_{i\in I}\alpha_i\Big(\mathsf{s}_i\cdot\big(\alpha\gamma(X)+(1-\alpha)\gamma(Y)\big)\Big).\tag{4.10}$$

Therefore, using the convexity and the S-invariance of φ, we get

$$\begin{aligned}\varphi\Big(\gamma\big(\alpha X+(1-\alpha)Y\big)\Big)&\leqslant\sum_{i\in I}\alpha_i\varphi\Big(\mathsf{s}_i\cdot\big(\alpha\gamma(X)+(1-\alpha)\gamma(Y)\big)\Big)\\ &=\sum_{i\in I}\alpha_i\varphi\big(\alpha\gamma(X)+(1-\alpha)\gamma(Y)\big)\\ &=\varphi\big(\alpha\gamma(X)+(1-\alpha)\gamma(Y)\big)\\ &\leqslant\alpha\varphi\big(\gamma(X)\big)+(1-\alpha)\varphi\big(\gamma(Y)\big),\end{aligned}\tag{4.11}$$

which verifies that $\varphi\circ\gamma$ is convex.

(ii): Proposition 4.4 (i) ensures that $\varphi\circ\gamma$ is proper. Hence, we obtain the conclusion by combining (i) and Proposition 4.5 (iii). ☐

Remark 4.7. Theorem 4.6 encompasses various results on the convexity of spectral functions in the context of Examples 2.4, 2.5, 2.6, 2.7, and 2.8; see, e.g., [1, Theorem 41], [18, Theorem 4.5], [19, Theorem, p. 276], [30, Theorem 6], [35, Theorem 4.3], [48, Theorem 2.2], [49, Theorems 3.1 and 3.2], and [54, Théorème 1.2].

From Theorem 4.6, we can immediately obtain characterizations of various useful properties of sets defined via spectra.

Corollary 4.8. *Suppose that Assumption 3.1 is in force and let D be an* S*-invariant subset of $\mathcal{X}$. Then the following hold:*

(i) $\operatorname{int}\gamma^{-1}(D) = \gamma^{-1}(\operatorname{int} D)$.

(ii) $\overline{\gamma^{-1}(D)} = \gamma^{-1}(\overline{D})$.

(iii) $\gamma^{-1}(D)$ *is closed if and only if* D *is closed.*

(iv) $\gamma^{-1}(D)$ *is convex if and only if* D *is convex.*

Proof. Apply respectively Proposition 4.4 (ii), Proposition 4.5 (i), Proposition 4.5 (iii), and Theorem 4.6 (i) to ι_D. ☐

We close this section with formulas for the convex hull and the set of extreme points of spectral sets in terms of that of the associated invariant sets, together with several illustrating examples. Here, given a subset C of a Euclidean space $\mathcal{H}$, we denote by $\operatorname{conv} C$ its convex hull and, if C is convex, by $\operatorname{ext} C$ its set of extreme points.

Proposition 4.9. *Suppose that Assumption 3.1 is in force and let D be a nonempty* S*-invariant subset of $\mathcal{X}$. Then the following hold:*

(i) $\operatorname{conv}\gamma^{-1}(D) = \gamma^{-1}(\operatorname{conv} D)$.

(ii) *Suppose that D is convex. Then* $\operatorname{ext}\gamma^{-1}(D) = \gamma^{-1}(\operatorname{ext} D)$.

Proof. (i): Let $X \in \mathfrak{H}$. First, suppose that $X \in \operatorname{conv}\gamma^{-1}(D)$, and let $(X_i)_{i\in I}$ and $(\alpha_i)_{i\in I}$ be finite families in $\operatorname{conv}\gamma^{-1}(D)$ and $]0,1]$, respectively, such that

$$\sum_{i\in I}\alpha_i = 1 \quad\text{and}\quad X = \sum_{i\in I}\alpha_i X_i. \tag{4.12}$$

Note that

$$(\forall i \in I)\quad \gamma(X_i) \in D. \tag{4.13}$$

In turn, appealing to Theorem 3.9, we obtain finite families $(\mathsf{s}_j)_{j\in J}$ in S and $(\beta_j)_{j\in J}$ in $]0,1]$ such that

$$\sum_{j\in J}\beta_j = 1 \quad\text{and}\quad \gamma(X) = \sum_{j\in J}\beta_j\left(\mathsf{s}_j\cdot\left(\sum_{i\in I}\alpha_i\gamma(X_i)\right)\right). \tag{4.14}$$

On the one hand, we have

$$\sum_{(i,j)\in I\times J}\alpha_i\beta_j = 1 \quad\text{and}\quad \gamma(X) = \sum_{(i,j)\in I\times J}\alpha_i\beta_j\big(\mathsf{s}_j\cdot\gamma(X_i)\big). \tag{4.15}$$

On the other hand, using the S-invariance of D and (4.13), we get $(\forall i \in I)(\forall j \in J)\ \mathsf{s}_j\cdot\gamma(X_i) \in D$. Therefore $\gamma(X) \in \operatorname{conv} D$ or, equivalently, $X \in \gamma^{-1}(\operatorname{conv} D)$. Conversely, suppose that $X \in \gamma^{-1}(\operatorname{conv} D)$, that is, there exist finite families $(y_k)_{k\in K}$ in D and $(\delta_k)_{k\in K}$ in $]0,1]$ such that

$$\sum_{k\in K}\delta_k = 1 \quad\text{and}\quad \gamma(X) = \sum_{k\in K}\delta_k y_k. \tag{4.16}$$

Take $a \in \mathcal{A}_X$ and set $(\forall k \in K)\ Y_k = \Lambda_a y_k$. For every $k \in K$, it follows (3.2) and (3.1) that $\gamma(Y_k) = (\gamma\circ\Lambda_a)(y_k) = \tau(y_k) \in \mathsf{S}\cdot y_k \subset D$, which confirms that $Y_k \in \gamma^{-1}(D)$. However, the linearity of Λ_a gives

$$X = \Lambda_a\gamma(X) = \sum_{k\in K}\delta_k\Lambda_a y_k = \sum_{k\in K}\delta_k Y_k. \tag{4.17}$$

As a result, $X \in \operatorname{conv} \gamma^{-1}(D)$.

(ii): Corollary 4.8 (iv) asserts that $\gamma^{-1}(D)$ is convex. Now take $X \in \mathfrak{H}$ and $\mathcal{a} \in \mathcal{A}_X$. First, suppose that $X \in \operatorname{ext} \gamma^{-1}(D)$, and let y and z be in D and $\alpha \in]0,1[$ be such that $\gamma(X) = \alpha y + (1-\alpha)z$. We must show that $y = z$. Towards this end, set $Y = \Lambda_{\mathcal{a}} y$ and $Z = \Lambda_{\mathcal{a}} z$. We deduce from (3.2) and (3.1) that $\gamma(Y) = (\gamma \circ \Lambda_{\mathcal{a}})(y) = \tau(y) \in \mathsf{S} \cdot y \subset D$ or, equivalently, $Y \in \gamma^{-1}(D)$. Likewise $Z \in \gamma^{-1}(D)$. At the same time, by linearity of $\Lambda_{\mathcal{a}}$,

$$X = \Lambda_{\mathcal{a}} \gamma(X) = \alpha \Lambda_{\mathcal{a}} y + (1-\alpha)\Lambda_{\mathcal{a}} z = \alpha Y + (1-\alpha)Z. \tag{4.18}$$

Therefore, because X is an extreme point of $\gamma^{-1}(D)$, we must have $Y = Z$, which leads to $\|y - z\| = \|\Lambda_{\mathcal{a}} y - \Lambda_{\mathcal{a}} z\| = \|Y - Z\| = 0$. Conversely, assume that $\gamma(X) \in \operatorname{ext} D$, and let $U \in \gamma^{-1}(D)$, $V \in \gamma^{-1}(D)$, and $\beta \in]0,1[$ be such that $X = \beta U + (1-\beta)V$. We infer from Theorem 3.9 that there exist finite families $(\mathsf{s}_i)_{i\in I}$ in S and $(\alpha_i)_{i\in I}$ in $]0,1]$ such that

$$\sum_{i\in I} \alpha_i = 1 \quad \text{and} \quad \gamma(X) = \sum_{i\in I} \alpha_i \Big(\mathsf{s}_i \cdot \big(\beta\gamma(U) + (1-\beta)\gamma(V)\big)\Big). \tag{4.19}$$

By linearity,

$$\gamma(X) = \sum_{i\in I} \alpha_i \beta \big(\mathsf{s}_i \cdot \gamma(U)\big) + \sum_{i\in I} \alpha_i (1-\beta)\big(\mathsf{s}_i \cdot \gamma(V)\big). \tag{4.20}$$

On the other hand, for every $i \in I$, since $\{\gamma(U), \gamma(V)\} \subset D$ and D is S-invariant, we get $\mathsf{s}_i \cdot \gamma(U) \in D$ and $\mathsf{s}_i \cdot \gamma(V) \in D$. Thus, inasmuch as $\gamma(X)$ is an extreme point of D, it follows that

$$(\forall i \in I) \quad \gamma(X) = \mathsf{s}_i \cdot \gamma(U) = \mathsf{s}_i \cdot \gamma(V). \tag{4.21}$$

In turn, by Proposition 3.5 (iii) and the assumption that S acts on $\mathcal{X}$ by linear isometries, $\|X\| = \|U\| = \|V\|$. Therefore, we derive from [6, Corollary 2.15] that

$$\begin{aligned} \beta(1-\beta)\|U - V\|^2 &= \beta\|U\|^2 + (1-\beta)\|V\|^2 - \|\beta U + (1-\beta)V\|^2 \\ &= \beta\|X\|^2 + (1-\beta)\|X\|^2 - \|X\|^2 \\ &= 0, \end{aligned} \tag{4.22}$$

which completes the proof. ☐

As a consequence, we can describe explicitly the convex hull of a set of elements with prescribed spectrum.

Example 4.10. Suppose that Assumption 3.1 is in force and let $x \in \operatorname{range} \gamma$. Then

$$\operatorname{conv}\big\{X \in \mathfrak{H} \mid \gamma(X) = x\big\} = \big\{X \in \mathfrak{H} \mid \gamma(X) \in \operatorname{conv}(\mathsf{S} \cdot x)\big\}. \tag{4.23}$$

Proof. Denote by D the S-invariant set $\mathsf{S} \cdot x$. Then, it results from Proposition 3.5 (i) and Proposition 3.3 (i) that

$$\begin{aligned} (\forall X \in \mathfrak{H}) \quad \gamma(X) \in D &\Leftrightarrow (\exists\, \mathsf{s} \in \mathsf{S})\ \gamma(X) = \mathsf{s} \cdot x. \\ &\Rightarrow \gamma(X) = \tau\big(\gamma(X)\big) = \tau(\mathsf{s} \cdot x) = \tau(x) = x, \end{aligned} \tag{4.24}$$

which confirms that $\big\{X \in \mathfrak{H} \mid \gamma(X) = x\big\} = \big\{X \in \mathfrak{H} \mid \gamma(X) \in D\big\} = \gamma^{-1}(D)$. Thus, we derive from Proposition 4.9 (i) that

$$\operatorname{conv}\big\{X \in \mathfrak{H} \mid \gamma(X) = x\big\} = \operatorname{conv} \gamma^{-1}(D) = \gamma^{-1}(\operatorname{conv} D) = \big\{X \in \mathfrak{H} \mid \gamma(X) \in \operatorname{conv}(\mathsf{S} \cdot x)\big\}, \tag{4.25}$$

as desired. ☐

Example 4.11 (eigenvalue decomposition). In the context of Example 2.6, (4.23) reads

$$\operatorname{conv}\big\{X \in \mathsf{H}^N(\mathbb{K}) \mid \gamma(X) = x\big\} = \big\{X \in \mathsf{H}^N(\mathbb{K}) \mid \gamma(X) \in \operatorname{conv}(\mathsf{P}^N \cdot x)\big\}, \tag{4.26}$$

where $x = (\xi_i)_{1\leqslant i\leqslant N} \in \mathbb{R}^N$ satisfies $\xi_1 \geqslant \cdots \geqslant \xi_N$. When $\mathbb{K}$ is $\mathbb{R}$ or $\mathbb{C}$, we recover several existing results on the convex hull of a set of Hermitian matrices with prescribed spectrum, e.g., [26, Proposition 2.1], [44, Corollary 9.3], and [53, Theorem 12].

Specializing Example 4.10 to the context of Example 2.8 and choosing $x = (1, \ldots, 1)$, we recover an expression for the convex hull of the special orthogonal group which first appeared in [53, Corollary 10].

Example 4.12 (signed singular value decomposition). Consider the setting of Example 2.8 and set $x = (1, \ldots, 1) \in \mathbb{R}^N$. Then

$$\operatorname{conv} \mathsf{SO}^N = \big\{Y \in \mathbb{R}^{N\times N} \mid \big(\sigma_1(Y), \ldots, \sigma_{N-1}(Y), \sigma_N(Y) \operatorname{sign}(\det Y)\big) \in \operatorname{conv}(\mathsf{S} \cdot x)\big\}. \tag{4.27}$$

The last illustration of Proposition 4.9 is a formula for the set of extreme points of the unit ball of a unitarily invariant norm on $\mathbb{K}^{M\times N}$. It covers, in particular, [57, Theorem 5.1] in the real and complex cases.

Example 4.13 (singular value decomposition). Consider the setting of Example 2.7. Let φ be a norm on $\mathbb{R}^m$ which is $\mathsf{P}^m_\pm$-invariant, and set $\Phi = \varphi \circ \sigma$. Then the following hold:

(i) Φ is a unitarily invariant norm on $\mathbb{K}^{M\times N}$, that is, Φ is a norm on $\mathbb{K}^{M\times N}$ which satisfies

$$\big(\forall X \in \mathbb{K}^{M\times N}\big)\big(\forall U \in \mathsf{U}^M(\mathbb{K})\big)\big(\forall V \in \mathsf{U}^N(\mathbb{K})\big) \quad \Phi(UXV^*) = \Phi(X). \tag{4.28}$$

(ii) Denote by $\mathscr{B}$ and B the unit balls of Φ and φ, respectively. Then $\operatorname{ext}\mathscr{B} = \sigma^{-1}(\operatorname{ext} B)$.

Proof. (i): An easy consequence of Proposition 4.1 and the singular value decomposition theorem.
(ii): Note that $\mathscr{B} = \sigma^{-1}(B)$ and apply Proposition 4.9 (ii). ☐

5. Conjugation and subdifferentiability of spectral functions

Two fundamental convex analytical objects attached to a proper function $f : \mathcal{H} \to \,]-\infty, +\infty]$ defined on a Euclidean space $\mathcal{H}$ are its conjugate

$$f^* : \mathcal{H} \to [-\infty, +\infty] : y \mapsto \sup_{x\in\mathcal{H}} \big(\langle x \mid y\rangle - f(x)\big), \tag{5.1}$$

and its subdifferential

$$\partial f : \mathcal{H} \to 2^{\mathcal{H}} : x \mapsto \big\{y \in \mathcal{H} \mid (\forall z \in \mathcal{H})\ \langle z - x \mid y\rangle + f(x) \leqslant f(z)\big\}, \tag{5.2}$$

which are related by the Fenchel–Young equality

$$(\forall x \in \mathcal{H})(\forall y \in \mathcal{H}) \quad y \in \partial f(x) \;\Leftrightarrow\; f(x) + f^*(y) = \langle x \mid y\rangle \;\Leftrightarrow\; x \in \operatorname{Argmin}\big(f - \langle \cdot \mid y\rangle\big), \tag{5.3}$$

where $\operatorname{Argmin} g$ denotes the set of minimizers of a function $g : \mathcal{H} \to \,]-\infty, +\infty]$, that is,

$$\operatorname{Argmin} g = \begin{cases} \big\{x \in \mathcal{H} \mid g(x) = \inf g(\mathcal{H})\big\} & \text{if } \inf g(\mathcal{H}) < +\infty, \\ \varnothing & \text{otherwise.} \end{cases} \tag{5.4}$$

In this section, we derive formulas for evaluating the conjugate and subdifferential of a spectral function in terms of that of the associated invariant function. Towards this end, we first establish a *reduced minimization* principle, which will play a fundamental role in our analysis. In essence, one can replace an optimality condition involving spectral functions by one that involves the associated invariant functions and the spectral mapping.

Theorem 5.1. *Suppose that Assumption 3.1 is in force, let $\varphi\colon \mathcal{X} \to \left]-\infty,+\infty\right]$ be proper and* S*-invariant, and let $Y \in \mathfrak{H}$. Set*

$$\mathcal{S} = \operatorname{Argmin}\big(\varphi\circ\gamma - \langle\cdot\mid Y\rangle\big) \quad \textit{and} \quad S = \operatorname{Argmin}\big(\varphi - \langle\cdot\mid\gamma(Y)\rangle\big). \tag{5.5}$$

Then the following hold:

(i) $\inf_{X\in\mathfrak{H}}\big(\varphi(\gamma(X)) - \langle X\mid Y\rangle\big) = \inf_{x\in\mathcal{X}}\big(\varphi(x) - \langle x\mid\gamma(Y)\rangle\big)$.

(ii) *Let $X \in \mathfrak{H}$. Then*

$$X\in\mathcal{S} \quad\Leftrightarrow\quad \begin{cases}\gamma(X)\in S\\ (\exists\,\mathscr{a}\in\mathscr{A})\ X = \Lambda_{\mathscr{a}}\gamma(X)\ \textit{and}\ Y = \Lambda_{\mathscr{a}}\gamma(Y).\end{cases} \tag{5.6}$$

(iii) *Let $x\in\mathcal{X}$ and $\mathscr{b}\in\mathscr{A}_Y$. Then $\Lambda_{\mathscr{b}}x\in\mathcal{S} \Leftrightarrow x\in S$.*

(iv) $\mathcal{S} = \big\{\Lambda_{\mathscr{b}}x \mid x\in S\ \textit{and}\ \mathscr{b}\in\mathscr{A}_Y\big\}$.

(v) *$\mathcal{S}$ is convex if and only if S is convex.*

(vi) *$\mathcal{S}$ is a singleton if and only if S is a singleton.*

Proof. We denote by μ and ν the infima on the left-hand and right-hand sides of (i), respectively. Using the S-invariance of φ and Proposition 4.4 (i), we get

$$\begin{aligned}(\forall x\in\mathcal{X})(\forall\mathscr{b}\in\mathscr{A}_Y)\quad \varphi(x) - \langle x\mid\gamma(Y)\rangle &= \varphi\big(\gamma(\Lambda_{\mathscr{b}}x)\big) - \langle\Lambda_{\mathscr{b}}x\mid\Lambda_{\mathscr{b}}\gamma(Y)\rangle\\ &= \varphi\big(\gamma(\Lambda_{\mathscr{b}}x)\big) - \langle\Lambda_{\mathscr{b}}x\mid Y\rangle \qquad (5.7)\\ &\geqslant \mu. \qquad (5.8)\end{aligned}$$

At the same time, since φ is proper, Proposition 4.4 (i) ensures that $\varphi\circ\gamma$ is likewise.

(i): It results from property [D] in Definition 2.1 that

$$(\forall X\in\mathfrak{H})\quad \varphi\big(\gamma(X)\big) - \langle X\mid Y\rangle \geqslant \varphi\big(\gamma(X)\big) - \langle\gamma(X)\mid\gamma(Y)\rangle \geqslant \nu. \tag{5.9}$$

Therefore $\mu\geqslant\nu$. On the other hand, taking the infimum over $x\in\mathcal{X}$ in (5.8) yields $\nu\geqslant\mu$. Altogether $\mu=\nu$.

(ii): First, assume that $X\in\mathcal{S}$. Then $\varphi(\gamma(X))\in\mathbb{R}$. In turn, since (i) and (5.9) entail that

$$\nu = \mu = \varphi\big(\gamma(X)\big) - \langle X\mid Y\rangle \geqslant \varphi\big(\gamma(X)\big) - \langle\gamma(X)\mid\gamma(Y)\rangle \geqslant \nu. \tag{5.10}$$

we get

$$\varphi\big(\gamma(X)\big) - \langle\gamma(X)\mid\gamma(Y)\rangle = \nu = \inf_{x\in\mathcal{X}}\big(\varphi(x) - \langle x\mid\gamma(Y)\rangle\big) \tag{5.11}$$

and

$$\langle X\mid Y\rangle = \langle\gamma(X)\mid\gamma(Y)\rangle. \tag{5.12}$$

Hence $\gamma(X) \in S$ by (5.11), while (5.12) and Proposition 3.6 force $\mathscr{A}_X \cap \mathscr{A}_Y \neq \varnothing$. Conversely, assume that $\gamma(X) \in S$ and that there exists $\mathscr{a} \in \mathscr{A}$ for which $X = \Lambda_{\mathscr{a}}\gamma(X)$ and $Y = \Lambda_{\mathscr{a}}\gamma(Y)$. Then $\langle X \mid Y\rangle = \langle \gamma(X) \mid \gamma(Y)\rangle$ due to the identity $\Lambda_{\mathscr{a}}^* \circ \Lambda_{\mathscr{a}} = \mathrm{Id}_{\mathcal{X}}$. Thus, we derive from (i) that

$$\varphi\big(\gamma(X)\big) - \langle X \mid Y\rangle = \varphi\big(\gamma(X)\big) - \langle \gamma(X) \mid \gamma(Y)\rangle = \nu = \mu \tag{5.13}$$

and, therefore, that $X \in \mathcal{S}$.

(iii): This follows from (5.7) and (i).

(iv): Combine (ii) and (iii).

(v): Let $\alpha \in\,]0,1[$ and $\mathscr{b} \in \mathscr{A}_Y$. Suppose first that $\mathcal{S}$ is convex and that x_0 and x_1 lie in S, and set $x = (1-\alpha)x_0 + \alpha x_1$. Item (iv) implies that $\Lambda_{\mathscr{b}}x_0$ and $\Lambda_{\mathscr{b}}x_1$ lie in the convex set $\mathcal{S}$, which leads to $\Lambda_{\mathscr{b}}x = (1-\alpha)\Lambda_{\mathscr{b}}x_0 + \alpha\Lambda_{\mathscr{b}}x_1 \in \mathcal{S}$. Hence, (iii) forces $x \in S$. Conversely, suppose that S is convex and that X_0 and X_1 lie in $\mathcal{S}$, and set $X = (1-\alpha)X_0 + \alpha X_1$. We infer from (ii) and Proposition 3.6 that

$$\big(\forall i \in \{0,1\}\big) \quad \gamma(X_i) \in S \quad \text{and} \quad \langle X_i \mid Y\rangle = \langle \gamma(X_i) \mid \gamma(Y)\rangle. \tag{5.14}$$

At the same time, Theorem 3.9 ensures the existence of finite families $(\mathsf{s}_j)_{j\in J}$ in S and $(\beta_j)_{j\in J}$ in $]0,1]$ such that

$$\sum_{j\in J}\beta_j = 1 \quad \text{and} \quad \gamma(X) = \sum_{j\in J}\beta_j\Big(\mathsf{s}_j \cdot \big((1-\alpha)\gamma(X_0) + \alpha\gamma(X_1)\big)\Big). \tag{5.15}$$

In turn, it results from Proposition 3.3 (iv), Proposition 3.5 (i), and (5.14) that

$$\big(\forall i \in \{0,1\}\big)(\forall j \in J) \quad \langle \mathsf{s}_j \cdot \gamma(X_i) \mid \gamma(Y)\rangle \leqslant \langle \gamma(X_i) \mid \gamma(Y)\rangle = \langle X_i \mid Y\rangle. \tag{5.16}$$

Combining this with property [D] in Definition 2.1 and (5.15), we obtain

$$\begin{aligned}\langle X \mid Y\rangle &\leqslant \langle \gamma(X) \mid \gamma(Y)\rangle \\ &= \sum_{j\in J}\beta_j\Big((1-\alpha)\langle \mathsf{s}_j \cdot \gamma(X_0) \mid \gamma(Y)\rangle + \alpha\langle \mathsf{s}_j \cdot \gamma(X_1) \mid \gamma(Y)\rangle\Big) \\ &\leqslant \sum_{j\in J}\beta_j\Big((1-\alpha)\langle X_0 \mid Y\rangle + \alpha\langle X_1 \mid Y\rangle\Big) \\ &= \langle X \mid Y\rangle.\end{aligned} \tag{5.17}$$

Hence

$$\langle X \mid Y\rangle = \langle \gamma(X) \mid \gamma(Y)\rangle, \tag{5.18}$$

which entails that $\mathscr{A}_X \cap \mathscr{A}_Y \neq \varnothing$. Moreover, we deduce from (5.16) that

$$\big(\forall i \in \{0,1\}\big)(\forall j \in J) \quad \langle \mathsf{s}_j \cdot \gamma(X_i) \mid \gamma(Y)\rangle = \langle \gamma(X_i) \mid \gamma(Y)\rangle. \tag{5.19}$$

In the light of (ii), it remains to show that $\gamma(X) \in S$. For every $i \in \{0,1\}$ and every $j \in J$, using the S-invariance of φ, (5.19), and (5.14), we get

$$\varphi\big(\mathsf{s}_j \cdot \gamma(X_i)\big) - \langle \mathsf{s}_j \cdot \gamma(X_i) \mid \gamma(Y)\rangle = \varphi\big(\gamma(X_i)\big) - \langle \gamma(X_i) \mid \gamma(Y)\rangle = \min_{z\in\mathcal{X}}\big(\varphi(z) - \langle z \mid \gamma(Y)\rangle\big), \tag{5.20}$$

which verifies that $\mathsf{s}_j \cdot \gamma(X_i) \in S$. Consequently, appealing to (5.15) and the convexity of S, we obtain $\gamma(X) \in S$.

(vi): Suppose that $\mathcal{S}$ is a singleton, say $\mathcal{S} = \{X\}$. Then, on the one hand, it results from (v) that S is convex. On the other hand, for every $x \in S$ and every $\mathscr{b} \in \mathscr{A}_Y$, we deduce from (iv) that $\Lambda_{\mathscr{b}} x = X$ and, thus, since $\Lambda_{\mathscr{b}}$ is a linear isometry, that $\|x\| = \|\Lambda_{\mathscr{b}} x\| = \|X\|$. Altogether, [6, Proposition 3.7] implies that S is a singleton. Conversely, assume that S is a singleton, say $S = \{x\}$. By (v), $\mathcal{S}$ is convex. At the same time, for every $X \in \mathcal{S}$, (iv) entails that there exists $\mathscr{b} \in \mathscr{A}_Y$ such that $X = \Lambda_{\mathscr{b}} x$, from which we obtain $\|X\| = \|x\|$. Therefore, invoking [6, Proposition 3.7] once more, we conclude that $\mathcal{S}$ is a singleton. ☐

Remark 5.2. Here are several observations on Theorem 5.1.

(i) Theorem 5.1 (iv) subsumes, in particular, [8, Theorem 1].

(ii) We derive from the first inequality in (5.9), Proposition 3.5 (i), and Theorem 5.1 (i) that

$$\mu \geqslant \inf_{\substack{x \in \mathcal{X} \\ \tau(x) = x}} \big(\varphi(x) - \langle x \mid \gamma(Y) \rangle\big) \geqslant \nu = \mu. \tag{5.21}$$

Hence

$$\inf_{X \in \mathfrak{H}} \big(\varphi(\gamma(X)) - \langle X \mid Y \rangle\big) = \inf_{x \in \mathcal{X}} \big(\varphi(x) - \langle x \mid \gamma(Y) \rangle\big) = \inf_{\substack{x \in \mathcal{X} \\ \tau(x) = x}} \big(\varphi(x) - \langle x \mid \gamma(Y) \rangle\big). \tag{5.22}$$

Additionally, arguing as in the proof of Theorem 5.1 (ii), we obtain

$$(\forall X \in \mathfrak{H}) \quad X \in \operatorname{Argmin}\big(\varphi \circ \gamma - \langle \cdot \mid Y \rangle\big) \quad \Leftrightarrow \quad \begin{cases} \gamma(X) \in \underset{\substack{x \in \mathcal{X} \\ \tau(x) = x}}{\operatorname{Argmin}} \big(\varphi(x) - \langle x \mid \gamma(Y) \rangle\big) \\ (\exists\, \mathscr{a} \in \mathscr{A})\ X = \Lambda_{\mathscr{a}} \gamma(X) \text{ and } Y = \Lambda_{\mathscr{a}} \gamma(Y). \end{cases} \tag{5.23}$$

Let us illustrate these identities via a problem in eigenvalue optimization [23]. Consider the setting of Example 2.6 with $\mathbb{K} = \mathbb{R}$, let $A \in \mathbb{R}^{M \times N}$, let $b \in \mathbb{R}^M$, and set

$$\begin{cases} \mathscr{D} = \big\{X \in \mathsf{H}^N(\mathbb{R}) \mid A\lambda(X) - b \in \left]-\infty, 0\right]^M\big\}, \\ D = \big\{x \in \mathbb{R}^N \mid A(x^{\downarrow}) - b \in \left]-\infty, 0\right]^M\big\}, \\ C = \big\{x \in \mathbb{R}^N \mid x^{\downarrow} = x \text{ and } Ax - b \in \left]-\infty, 0\right]^M\big\}. \end{cases} \tag{5.24}$$

Then C is convex, D is P^N-invariant, and $\iota_{\mathscr{D}} = \iota_D \circ \lambda$. By noting that, in the current situation, $\tau \colon x \mapsto x^{\downarrow}$, we apply the above identities to the P^N-invariant function $\varphi = \iota_D + \|\cdot\|_2^2/2$ to get

$$\begin{aligned} \big(\forall Y \in \mathsf{H}^N(\mathbb{R})\big) \quad \inf_{X \in \mathscr{D}} \frac{1}{2}\|X - Y\|^2 &= \frac{1}{2}\|Y\|^2 + \inf_{X \in \mathsf{H}^N(\mathbb{R})} \left(\frac{1}{2}\|X\|^2 + \iota_{\mathscr{D}}(X) - \langle X \mid Y \rangle\right) \\ &= \frac{1}{2}\|\lambda(Y)\|_2^2 + \inf_{\substack{x \in \mathbb{R}^N \\ x^{\downarrow} = x}} \left(\frac{1}{2}\|x\|_2^2 + \iota_D(x) - \langle x \mid \lambda(Y) \rangle\right) \\ &= \inf_{\substack{x \in D \\ x^{\downarrow} = x}} \frac{1}{2}\|x - \lambda(Y)\|_2^2 \\ &= \inf_{x \in C} \frac{1}{2}\|x - \lambda(Y)\|_2^2, \end{aligned} \tag{5.25}$$

which implies [23, Theorem 6].

Theorem 5.1 (i) yields at once the following calculus rule for evaluating the conjugate of a spectral function.

Corollary 5.3. *Suppose that Assumption 3.1 is in force and let* $\varphi\colon \mathcal{X} \to \left]-\infty,+\infty\right]$ *be* S*-invariant. Then* $(\varphi\circ\gamma)^* = \varphi^*\circ\gamma$.

The following fact will assist us in studying subdifferentiability of spectral functions.

Lemma 5.4. *Let* $\mathcal{H}$ *be a Euclidean space, let* G *be a group which acts on* $\mathcal{H}$ *by linear isometries, and let* $f\colon \mathcal{H} \to [-\infty,+\infty]$ *be* G*-invariant. Then the following hold:*

(i) f^* *is* G*-invariant.*

(ii) *Suppose that* f *is proper, and let* $x \in \mathcal{H}$ *and* $\mathrm{g} \in \mathrm{G}$*. Then* $\partial f(\mathrm{g}\cdot x) = \mathrm{g}\cdot\partial f(x)$.

Proof. (i): In fact, for every $y \in \mathcal{H}$ and every $\mathrm{g} \in \mathrm{G}$, using Lemma 3.2 and the fact that the mapping $x \mapsto \mathrm{g}^{-1}\cdot x$ is a bijection, we obtain

$$\begin{aligned} f^*(\mathrm{g}\cdot y) &= \sup_{x\in\mathcal{H}}\big(\langle x \mid \mathrm{g}\cdot y\rangle - f(x)\big)\\ &= \sup_{x\in\mathcal{H}}\big(\langle \mathrm{g}^{-1}\cdot x \mid y\rangle - f(\mathrm{g}^{-1}\cdot x)\big)\\ &= \sup_{x\in\mathcal{H}}\big(\langle x \mid y\rangle - f(x)\big)\\ &= f^*(y), \end{aligned} \tag{5.26}$$

as desired.

(ii): It results from (5.3), (i), and Lemma 3.2 that

$$\begin{aligned} (\forall y \in \mathcal{H})\quad y \in \partial f(\mathrm{g}\cdot x) &\Leftrightarrow f(\mathrm{g}\cdot x) + f^*(y) = \langle \mathrm{g}\cdot x \mid y\rangle\\ &\Leftrightarrow f(x) + f^*(\mathrm{g}^{-1}\cdot y) = \langle x \mid \mathrm{g}^{-1}\cdot y\rangle\\ &\Leftrightarrow \mathrm{g}^{-1}\cdot y \in \partial f(x)\\ &\Leftrightarrow y \in \mathrm{g}\cdot\partial f(x), \end{aligned} \tag{5.27}$$

which completes the proof. ☐

Next, we examine subdifferentiability of spectral functions.

Proposition 5.5. *Suppose that Assumption 3.1 is in force, let* $\varphi\colon \mathcal{X} \to \left]-\infty,+\infty\right]$ *be proper and* S*-invariant, and let* $X \in \mathfrak{H}$*. Then the following hold:*

(i) *For every* $Y \in \mathfrak{H}$, $Y \in \partial(\varphi\circ\gamma)(X)$ *if and only if* $\gamma(Y) \in \partial\varphi(\gamma(X))$ *and there exists* $a \in \mathcal{A}$ *such that* $X = \Lambda_a\gamma(X)$ *and* $Y = \Lambda_a\gamma(Y)$.

(ii) $\partial(\varphi\circ\gamma)(X) = \big\{\Lambda_a y \mid y \in \partial\varphi(\gamma(X)) \text{ and } a \in \mathcal{A}_X\big\}$.

(iii) $\partial(\varphi\circ\gamma)(X)$ *is a singleton if and only if* $\partial\varphi(\gamma(X))$ *is a singleton.*

Proof. (i): In view of (5.3), this follows from Theorem 5.1 (ii).

(ii): The inclusion $\partial(\varphi\circ\gamma)(X) \subset \big\{\Lambda_a y \mid y \in \partial\varphi(\gamma(X)) \text{ and } a \in \mathcal{A}_X\big\}$ follows from (i). To establish the reverse inclusion, suppose that $y \in \partial\varphi(\gamma(X))$ and let $a \in \mathcal{A}_X$. Lemma 5.4 (i) asserts that φ^* is

S-invariant and, in turn, Proposition 4.4 (i) implies that $\varphi^* \circ \gamma \circ \Lambda_a = \varphi^*$. Therefore, we derive from Corollary 5.3 and (5.3) that

$$\begin{aligned}(\varphi \circ \gamma)(X) + (\varphi \circ \gamma)^*(\Lambda_a y) &= \varphi\big(\gamma(X)\big) + \varphi^*\big(\gamma(\Lambda_a y)\big)\\ &= \varphi\big(\gamma(X)\big) + \varphi^*(y)\\ &= \langle \gamma(X) \mid y\rangle\\ &= \langle \Lambda_a \gamma(X) \mid \Lambda_a y\rangle\\ &= \langle X \mid \Lambda_a y\rangle. \end{aligned} \tag{5.28}$$

Consequently, invoking (5.3) once more, we infer that $\Lambda_a y \in \partial(\varphi \circ \gamma)(X)$.

(iii): Note that both of the sets $\partial(\varphi\circ\gamma)(X)$ and $\partial\varphi(\gamma(X))$ are convex due to [6, Proposition 16.4(iii)]. We can now use (ii) and proceed analogously to the proof of Theorem 5.1 (vi) to arrive at the conclusion. □

Remark 5.6. Let us make a connection between Proposition 5.5 and existing works.

(i) The realization of Proposition 5.5 (i) and (ii) in normal decomposition systems (Example 2.4) is established in [35, Theorem 4.5]; see [33, 34] for further special cases in the context of matrices.

(ii) The special case of Proposition 5.5 (i) in the context of Euclidean Jordan algebras (Example 2.5) is established in [1, Corollary 31].

Instantiations of Corollary 5.3, Proposition 5.5, and Corollary 5.8 in the frameworks described in Examples 2.6, 2.7, and 2.8 can be found in, e.g., [18] and the references therein.

We deduce at once from Proposition 5.5 (i) the so-called commutation principle [25, Theorem 1.3].

Example 5.7 (normal decomposition system)**.** Consider the normal decomposition system setting of Example 2.4. Let $\varphi\colon \mathcal{X} \to\ ]-\infty, +\infty]$ be proper, convex, and S-invariant, let $\Psi\colon \mathfrak{H} \to \mathbb{R}$ be Fréchet differentiable, and let $X \in \mathfrak{H}$ be such that $\gamma(X) \in \operatorname{dom}\varphi$. Suppose that X is a local minimizer of $\varphi \circ \gamma + \Psi$. Then

$$\gamma\big(-\nabla\Psi(X)\big) \in \partial\varphi\big(\gamma(X)\big) \quad \text{and} \quad \big[\,(\exists\, \mathrm{g} \in \mathrm{G})\ X = \mathrm{g}\cdot\gamma(X) \text{ and } \nabla\Psi(X) = -\mathrm{g}\cdot\gamma\big(-\nabla\Psi(X)\big)\,\big]. \tag{5.29}$$

Proof. Theorem 4.6 asserts that $\varphi \circ \gamma$ is convex. Therefore, since $X \in \operatorname{dom}(\varphi \circ \gamma + \Psi)$, we infer from [41, Propositions 1.114 and 1.107(i), and Theorem 1.93] that $0 \in \partial(\varphi\circ\gamma)(X) + \nabla\Psi(X)$. The conclusion thus follows from Proposition 5.5 (i). □

We end this section with Fréchet differentiability property of spectral functions.

Corollary 5.8. *Suppose that Assumption 3.1 is in force and let $\varphi\colon \mathcal{X} \to\ ]-\infty,+\infty]$ be proper, convex, and S-invariant, and let $X \in \mathfrak{H}$. Then $\varphi\circ\gamma$ is differentiable at X if and only if φ is differentiable at $\gamma(X)$, in which case*

$$\gamma\big(\nabla(\varphi\circ\gamma)(X)\big) = \nabla\varphi\big(\gamma(X)\big) \quad \textit{and} \quad (\forall a \in \mathcal{A}_X)\ \nabla(\varphi\circ\gamma)(X) = \Lambda_a\big(\nabla\varphi\big(\gamma(X)\big)\big). \tag{5.30}$$

Proof. Theorem 4.6 (i) ensures that $\varphi \circ \gamma$ is convex. Thus, according to [45, Theorem 25.1] and Proposition 5.5 (iii),

$$\begin{aligned}\varphi\circ\gamma \text{ is differentiable at } X &\Leftrightarrow \partial(\varphi\circ\gamma)(X) \text{ is a singleton}\\ &\Leftrightarrow \partial\varphi\big(\gamma(X)\big) \text{ is a singleton}\\ &\Leftrightarrow \varphi \text{ is differentiable at } \gamma(X). \end{aligned}\tag{5.31}$$

Thus, if φ is differentiable at $\gamma(X)$, then by the above equivalences and [45, Theorem 25.1], we get $\partial(\varphi\circ\gamma)(X) = \{\nabla(\varphi\circ\gamma)(X)\}$ and $\partial\varphi(\gamma(X)) = \{\nabla\varphi(\gamma(X))\}$, from which and items (i) and (ii) in Proposition 5.5 we obtain (5.30). □

6. Bregman proximity operators of spectral functions

In this last section, we characterize the Bregman proximity (or proximal point) operator of spectral functions in the setting of general spectral decomposition systems. These operators generalize the classical proximal point operators that are the basic building blocks of modern first-order nonsmooth optimization algorithms [6, 7, 15].

Let $\mathcal{H}$ be a Euclidean space. A function $g \in \Gamma_0(\mathcal{H})$ is said to be:

- essentially smooth if ∂g is at most single-valued;
- essentially strictly convex if it is strictly convex on every convex subset of $\operatorname{dom} \partial g = \{x \in \mathcal{H} \mid \partial g(x) \neq \varnothing\}$;
- a Legendre function if it is both essentially smooth and essentially strictly convex;

see [45, Section 26]. If $g \in \Gamma_0(\mathcal{H})$ is a Legendre function, then

$$D_g \colon \mathcal{H} \times \mathcal{H} \to [0, +\infty] \\ (y, x) \mapsto \begin{cases} g(y) - g(x) - \langle y - x \mid \nabla g(x) \rangle & \text{if } x \in \operatorname{int} \operatorname{dom} g, \\ +\infty, & \text{otherwise} \end{cases} \tag{6.1}$$

is called the *Bregman distance* associated with g. Now let $f \colon \mathcal{H} \to \,]-\infty, +\infty]$ and let $g \in \Gamma_0(\mathcal{H})$ be a Legendre function such that $(\operatorname{dom} f) \cap (\operatorname{dom} g) \neq \varnothing$. The *Bregman envelope* of f with respect to g is defined by

$$\operatorname{env}_f^g \colon \mathcal{H} \to [-\infty, +\infty] \colon x \mapsto \inf_{y \in \mathcal{H}} \big(f(y) + D_g(y, x) \big), \tag{6.2}$$

and the *Bregman proximity operator* (or, in the terminology of [4], D-prox operator) of f with respect to g is defined by

$$\operatorname{Prox}_f^g \colon \mathcal{H} \to 2^{\mathcal{H}} \\ x \mapsto \operatorname{Argmin}\big(f + D_g(\cdot, x)\big) = \begin{cases} \operatorname{Argmin}\big(f + g - \langle \cdot \mid \nabla g(x) \rangle\big) & \text{if } x \in \operatorname{int} \operatorname{dom} g, \\ \varnothing & \text{otherwise.} \end{cases} \tag{6.3}$$

In particular, given a subset C of $\mathcal{H}$ such that $C \cap \operatorname{dom} g \neq \varnothing$, we define the Bregman distance to C with respect to g as

$$\operatorname{dist}_C^g = \operatorname{env}_{\iota_C}^g, \tag{6.4}$$

and the Bregman projector onto C with respect to g as

$$\operatorname{Proj}_C^g = \operatorname{Prox}_{\iota_C}^g. \tag{6.5}$$

Finally, when $g = \|\cdot\|^2/2$, we shall omit the superscript and simply write env_f, Prox_f, dist_C, and Proj_C.

We can now establish, under the umbrella of Assumption 3.1, relationship between the Bregman proximity operators $\operatorname{Prox}_{\varphi\circ\gamma}^{\psi\circ\gamma}$ and $\operatorname{Prox}_{\varphi}^{\psi}$, where $\varphi \colon \mathcal{X} \to \,]-\infty, +\infty]$ is S-invariant and $\psi \circ \gamma$ is a Legendre spectral function, without any additional assumptions on φ and ψ. This necessitates the following technical result.

Proposition 6.1. *Suppose that Assumption 3.1 is in force and let $\varphi \in \Gamma_0(\mathcal{X})$ be S-invariant. Then the following hold:*

(i) $\varphi \circ \gamma$ *is essentially smooth if and only if* φ *is essentially smooth.*

(ii) $\varphi \circ \gamma$ *is essentially strictly convex if and only if* φ *is essentially strictly convex.*

(iii) $\varphi \circ \gamma$ *is a Legendre function if and only if* φ *is a Legendre function.*

Proof. Theorem 4.6 (ii) states that $\varphi \circ \gamma \in \Gamma_0(\mathfrak{H})$.

(i): Suppose that $\varphi \circ \gamma$ is essentially smooth, and take $x \in \operatorname{dom} \partial\varphi$ and $\mathscr{a} \in \mathscr{A}$. By (3.1), there exists $\mathsf{s} \in \mathsf{S}$ such that

$$\gamma(\Lambda_{\mathscr{a}} x) = \tau(x) = \mathsf{s} \cdot x, \tag{6.6}$$

where the first equality follows from (3.2). In turn, because $\partial\varphi(x)$ is nonempty, Lemma 5.4 (ii) ensures that $\partial\varphi(\gamma(\Lambda_{\mathscr{a}} x))$ is likewise, and it thus results from Proposition 5.5 (ii) that $\partial(\varphi \circ \gamma)(\Lambda_{\mathscr{a}} x) \neq \varnothing$. Hence, the essential smoothness of $\varphi \circ \gamma$ forces $\partial(\varphi \circ \gamma)(\Lambda_{\mathscr{a}} x)$ to be a singleton and, therefore, Proposition 5.5 (iii) entails that $\partial\varphi(\gamma(\Lambda_{\mathscr{a}} x))$ is likewise. Combining with (6.6), we infer from Lemma 5.4 (ii) that $\partial\varphi(x)$ must be a singleton. As a result, φ is essentially smooth. Conversely, suppose that φ is essentially smooth and let $X \in \operatorname{dom} \partial(\varphi \circ \gamma)$. Then, Proposition 5.5 (i) implies that $\gamma(X) \in \operatorname{dom} \partial\varphi$ and φ is therefore differentiable at $\gamma(X)$. Consequently, we deduce from Corollary 5.8 that $\varphi \circ \gamma$ is differentiable at X and conclude that $\varphi \circ \gamma$ is essentially smooth.

(ii): We deduce from Lemma 5.4 (i) that φ^* is an S-invariant function in $\Gamma_0(\mathcal{X})$. Thus, on account of [45, Theorem 26.3], Corollary 5.3, and (i), we obtain

$$\begin{aligned}
\varphi \circ \gamma \text{ is essentially strictly convex} &\Leftrightarrow \varphi^* \circ \gamma = (\varphi \circ \gamma)^* \text{ is essentially smooth} \\
&\Leftrightarrow \varphi^* \text{ is essentially smooth} \\
&\Leftrightarrow \varphi \text{ is essentially strictly convex,}
\end{aligned} \tag{6.7}$$

as claimed.

(iii): Combine (i) and (ii). ☐

The main results of this section are laid out in the following theorem.

Theorem 6.2. *Suppose that Assumption 3.1 is in force. Let* $\varphi\colon \mathcal{X} \to \left]-\infty, +\infty\right]$ *be a proper* S*-invariant function, let* $\psi \in \Gamma_0(\mathcal{X})$ *be an* S*-invariant Legendre function such that* $(\operatorname{dom}\varphi) \cap (\operatorname{dom}\psi) \neq \varnothing$*, and let* $X \in \mathfrak{H}$*. Then the following hold:*

(i) $\operatorname{env}_{\varphi\circ\gamma}^{\psi\circ\gamma} X = (\operatorname{env}_{\varphi}^{\psi})(\gamma(X))$.

(ii) *Let* $Z \in \mathfrak{H}$*. Then*

$$Z \in \operatorname{Prox}_{\varphi\circ\gamma}^{\psi\circ\gamma} X \quad \Leftrightarrow \quad \begin{cases} \gamma(Z) \in \operatorname{Prox}_{\varphi}^{\psi} \gamma(X) \\ (\exists\, \mathscr{a} \in \mathscr{A})\ X = \Lambda_{\mathscr{a}}\gamma(X) \ \text{and}\ Z = \Lambda_{\mathscr{a}}\gamma(Z). \end{cases} \tag{6.8}$$

(iii) *Let* $z \in \mathcal{X}$ *and* $\mathscr{a} \in \mathscr{A}_X$*. Then* $\Lambda_{\mathscr{a}} z \in \operatorname{Prox}_{\varphi\circ\gamma}^{\psi\circ\gamma} X \Leftrightarrow z \in \operatorname{Prox}_{\varphi}^{\psi} \gamma(X)$.

(iv) $\operatorname{Prox}_{\varphi\circ\gamma}^{\psi\circ\gamma} X = \big\{\Lambda_{\mathscr{a}} z \mid z \in \operatorname{Prox}_{\varphi}^{\psi} \gamma(X) \ \text{and}\ \mathscr{a} \in \mathscr{A}_X\big\}$.

(v) $\operatorname{Prox}_{\varphi\circ\gamma}^{\psi\circ\gamma} X$ *is convex if and only if* $\operatorname{Prox}_{\varphi}^{\psi} \gamma(X)$ *is convex.*

(vi) $\operatorname{Prox}_{\varphi\circ\gamma}^{\psi\circ\gamma} X$ *is a singleton if and only if* $\operatorname{Prox}_{\varphi}^{\psi} \gamma(X)$ *is a singleton.*

Proof. Proposition 6.1 (iii) asserts that $\psi \circ \gamma$ is a Legendre function, while Proposition 4.4 (ii) states that $\operatorname{int}\operatorname{dom}(\psi \circ \gamma) = \gamma^{-1}(\operatorname{int}\operatorname{dom}\psi)$. We assume henceforth that

$$\gamma(X) \in \operatorname{int}\operatorname{dom}\psi \tag{6.9}$$

since otherwise the assertions are clear. Then ψ is differentiable at $\gamma(X)$ and $\psi \circ \gamma$ is differentiable at X. In turn, upon setting

$$Y = \nabla(\psi \circ \gamma)(X) \quad \text{and} \quad y = \nabla\psi\big(\gamma(X)\big), \tag{6.10}$$

we get from Corollary 5.8 that

$$\gamma(Y) = y \quad \text{and} \quad (\forall \mathscr{a} \in \mathscr{A}_X)\ Y = \Lambda_{\mathscr{a}} y = \Lambda_{\mathscr{a}} \gamma(Y). \tag{6.11}$$

Therefore, by Proposition 3.6,

$$\langle X \mid Y \rangle = \langle \gamma(X) \mid \gamma(Y) \rangle = \langle \gamma(X) \mid y \rangle. \tag{6.12}$$

We claim that

$$\mathscr{A}_X = \mathscr{A}_Y. \tag{6.13}$$

Indeed, it results from (6.11) that $\mathscr{A}_X \subset \mathscr{A}_Y$. Now take $\mathscr{b} \in \mathscr{A}_Y$ and set $W = \Lambda_{\mathscr{b}} \gamma(X)$. Proposition 3.5 (i) implies that $\gamma(W) = \gamma(X)$ and, therefore, that

$$\Lambda_{\mathscr{b}} \gamma(W) = \Lambda_{\mathscr{b}} \gamma(X) = W \tag{6.14}$$

Thus $\mathscr{b} \in \mathscr{A}_W$. Additionally, we have

$$\partial\psi\big(\gamma(W)\big) = \partial\psi\big(\gamma(X)\big) = \big\{\nabla\psi\big(\gamma(X)\big)\big\}, \tag{6.15}$$

where the last equality follows from the differentiability of ψ at $\gamma(X)$. Hence, on account of Proposition 5.5 (ii), $\Lambda_{\mathscr{b}}(\nabla\psi(\gamma(X))) \in \partial(\psi \circ \gamma)(W)$ and, because $\psi \circ \gamma$ is a Legendre function, we must have $\nabla(\psi \circ \gamma)(W) = \Lambda_{\mathscr{b}}(\nabla\psi(\gamma(X)))$. In turn, since $\mathscr{b} \in \mathscr{A}_Y$, we derive from (6.10) and (6.11) that

$$\nabla(\psi \circ \gamma)(W) = \Lambda_{\mathscr{b}} \gamma(Y) = Y = \nabla(\psi \circ \gamma)(X). \tag{6.16}$$

Therefore, since $\psi \circ \gamma$ is a Legendre function, we infer from [45, Theorem 26.5] and (6.14) that $X = W = \Lambda_{\mathscr{b}} \gamma(X)$ and, therefore, that $\mathscr{b} \in \mathscr{A}_X$.

(i): We derive from (6.1), (6.10), Theorem 5.1 (i) applied to the function $\varphi + \psi$, (6.11), and (6.12) that

$$\begin{aligned}
\operatorname{env}_{\varphi\circ\gamma}^{\psi\circ\gamma} X &= \inf_{Z \in \mathfrak{H}} \Big(\varphi\big(\gamma(Z)\big) + D_{\psi\circ\gamma}(Z, X)\Big) \\
&= \inf_{Z \in \mathfrak{H}} \Big(\varphi\big(\gamma(Z)\big) + \psi\big(\gamma(Z)\big) - \psi\big(\gamma(X)\big) - \langle Z \mid Y \rangle + \langle X \mid Y \rangle\Big) && (6.17) \\
&= \inf_{z \in \mathcal{X}} \Big(\varphi(z) + \psi(z) - \psi\big(\gamma(X)\big) - \langle z \mid y \rangle + \langle \gamma(X) \mid y \rangle\Big) && (6.18) \\
&= \big(\operatorname{env}_{\varphi}^{\psi}\big)\big(\gamma(X)\big), && (6.19)
\end{aligned}$$

as desired.

(ii)–(vi): On account of (6.3) and (6.10), we have

$$\operatorname{Prox}_{\varphi\circ\gamma}^{\psi\circ\gamma} X = \operatorname{Argmin}\big(\varphi\circ\gamma + \psi\circ\gamma - \langle\cdot \mid Y\rangle\big) \tag{6.20}$$

and, by using (6.11),

$$\operatorname{Prox}_{\varphi}^{\psi} \gamma(X) = \operatorname{Argmin}\big(\varphi + \psi - \langle\cdot \mid y\rangle\big) = \operatorname{Argmin}\big(\varphi + \psi - \langle\cdot \mid \gamma(Y)\rangle\big). \tag{6.21}$$

In the light of (6.13), the assertions respectively follow from items (ii)–(vi) in Theorem 5.1 applied to the function $\varphi + \psi$. ☐

Remark 6.3. Let us relate Theorem 6.2 to existing works.

(i) Consider the special case of Theorem 6.2 where $\varphi \in \Gamma_0(\mathcal{X})$ and $\psi = \|\cdot\|^2/2$.
- In the context of Example 2.3, Theorem 6.2 (iv) reduces to [7, Theorem 6.18].
- In the context of Examples 2.6 and 2.7 where $\mathbb{K}$ is $\mathbb{R}$ or $\mathbb{C}$, we recover from Theorem 6.2 (iv) well-known expressions for the proximity operator of a lower semicontinuous convex spectral function; see, for instance, [6, Corollary 24.65 and Proposition 24.68] and [7, Theorems 7.18 and 7.29].
- In the context of Example 2.7 where $\mathbb{K} = \mathbb{H}$, concrete instantiations of Theorem 6.2 (iv) which arise in machine learning applications can be found in [11, 12].
- Theorem 6.2 (iv) appears to be new in the context of Examples 2.4, 2.5, and 2.8.

(ii) In the context of Example 2.6 where $\mathbb{K} = \mathbb{R}$, [8, Corollary 1] follows from items (iv) and (vi) in Theorem 6.2 with $\varphi \in \Gamma_0(\mathcal{X})$, and [20, Theorem 4.1] follows from items (i) and (iv) in Theorem 6.2 with $\psi = \|\cdot\|^2/2$.

(iii) Let us point out that Theorem 6.2 (iv) fully describes the set-valued operator $\operatorname{Prox}_{\varphi\circ\gamma}^{\psi\circ\gamma}$ in terms of $\operatorname{Prox}_{\varphi}^{\psi}$ and the spectral mapping γ. To the best of our knowledge, this result is new, even in the Hermitian matrix setting of Example 2.6.

(iv) Concrete examples of proximity operators of functions on matrix spaces can be found in [3, 6, 7, 8, 13].

Theorem 6.2 yields the following descriptions of the Bregman distance to a spectral set, as well as the associated Bregman projector, in terms of those of the associated invariant set.

Corollary 6.4. *Suppose that Assumption 3.1 is in force. Let D be a nonempty* S*-invariant subset of $\mathcal{X}$, let $\psi \in \Gamma_0(\mathcal{X})$ be an* S*-invariant Legendre function such that $D \cap \operatorname{dom}\psi \neq \varnothing$, and let $X \in \mathfrak{H}$. Then the following hold:*

(i) $\operatorname{dist}_{\gamma^{-1}(D)}^{\psi\circ\gamma}(X) = \operatorname{dist}_D^{\psi}(\gamma(X))$.

(ii) *For every $Z \in \mathfrak{H}$,*

$$Z \in \operatorname{Proj}_{\varphi\circ\gamma}^{\psi\circ\gamma} X \quad\Leftrightarrow\quad \begin{cases} \gamma(Z) \in \operatorname{Proj}_{\varphi}^{\psi} \gamma(X) \\ (\exists\, a \in \mathcal{A})\ X = \Lambda_a \gamma(X) \text{ and } Z = \Lambda_a \gamma(Z). \end{cases} \tag{6.22}$$

(iii) $\operatorname{Proj}_{\gamma^{-1}(D)}^{\psi\circ\gamma} X = \big\{\Lambda_a z \mid z \in \operatorname{Proj}_D^{\psi} \gamma(X) \text{ and } a \in \mathcal{A}_X\big\}$.

(iv) $\operatorname{Proj}_{\gamma^{-1}(D)}^{\psi\circ\gamma} X$ *is a singleton if and only if* $\operatorname{Proj}_D^{\psi} \gamma(X)$ *is a singleton.*

Proof. Apply respectively items (i), (ii), (iv), and (vi) in Theorem 6.2 to the proper S-invariant function ι_D, and observe that $\iota_{\gamma^{-1}(D)} = \iota_D \circ \gamma$. ☐

Example 6.5 (Euclidean Jordan algebra). Consider the setting of Example 2.5. For every $X \in \mathfrak{H}$, we define the *rank of* X, in symbol $\operatorname{Rank} X$, to be the number of nonzero entries of the vector $\lambda(X)$ of eigenvalues of X. Now let $r \in \{1, \dots, N-1\}$, set

$$\mathscr{D} = \big\{X \in \mathfrak{H} \mid \lambda(X) \in \mathbb{R}^N_+ \ \text{and} \ \operatorname{Rank} X \leqslant r\big\}, \tag{6.23}$$

let $X \in \mathfrak{H}$, and recall that (see (2.4) and (2.17)) $\mathscr{A}_X$ is the set of Jordan frames $(A_i)_{1\leqslant i\leqslant N}$ in $\mathfrak{H}$ for which $X = \sum_{i=1}^N \lambda_i(X)A_i$. Furthermore, denote by P^N_X the set of $N \times N$ permutation matrices that fix $\lambda(X)$, that is,

$$\mathsf{P}^N_X = \big\{P \in \mathsf{P}^N \mid \lambda(X) = P\lambda(X)\big\}. \tag{6.24}$$

Then exactly one of the following holds:

(i) $(\exists\, i \in \{1, \dots, N\})\ \lambda_i(X) \geqslant 0$: Set

$$q = \max\big\{i \in \{1, \dots, N\} \mid \lambda_i(X) \geqslant 0\big\} \quad \text{and} \quad \overline{x} = \big(\lambda_1(X), \dots, \lambda_{\min\{q,r\}}(X), 0, \dots, 0\big). \tag{6.25}$$

Then

$$\operatorname{Proj}_{\mathscr{D}} X = \left\{\sum_{i=1}^N \zeta_i A_i \;\middle|\; (A_i)_{1\leqslant i\leqslant N} \in \mathscr{A}_X,\ P \in \mathsf{P}^N_X,\ (\zeta_i)_{1\leqslant i\leqslant N} = P\overline{x}\right\}. \tag{6.26}$$

(ii) $(\forall i \in \{1, \dots, N\})\ \lambda_i(X) < 0$: Then $\operatorname{Proj}_{\mathscr{D}} X = \{0\}$.

Proof. Define

$$D = \big\{y \in \mathbb{R}^N_+ \mid \|y\|_0 \leqslant r\big\}, \tag{6.27}$$

where $\|y\|_0$ denotes the number of nonzero entries of a vector $y \in \mathbb{R}^N$. Then D is P^N-invariant and $\mathscr{D} = \lambda^{-1}(D)$. Thus, using (2.17), we deduce from Corollary 6.4 (iii) applied to the Legendre function $\psi = \|\cdot\|^2/2$ that

$$\operatorname{Proj}_{\mathscr{D}} X = \left\{\sum_{i=1}^N \zeta_i A_i \;\middle|\; (\zeta_i)_{1\leqslant i\leqslant N} \in \operatorname{Proj}_D \lambda(X) \ \text{and} \ (A_i)_{1\leqslant i\leqslant N} \in \mathscr{A}_X\right\}, \tag{6.28}$$

and it therefore suffices to determine $\operatorname{Proj}_D \lambda(X)$. Towards this goal, take $z \in D$ and write (see Notation 2.2)

$$z^{\downarrow} = (\pi_i)_{1\leqslant i\leqslant N}. \tag{6.29}$$

Since $\pi_1 \geqslant \cdots \geqslant \pi_N \geqslant 0$ and $\|z^{\downarrow}\|_0 = \|z\|_0 \leqslant r$, we must have

$$\pi_1 \geqslant \cdots \geqslant \pi_r \geqslant 0 = \pi_{r+1} = \cdots = \pi_N. \tag{6.30}$$

(i): Let us consider two cases.
(a) $q < r$: The very definition of q yields

$$\big(\forall i \in \{q+1, \dots, r\}\big) \quad \lambda_i(X) < 0. \tag{6.31}$$

Hence, we derive from the Hardy–Littlewood–Pólya rearrangement inequality that

$$\begin{aligned}\|z-\lambda(X)\|^2 &\geqslant \left\|z^{\downarrow}-\lambda(X)\right\|^2\\ &= \sum_{i=1}^{q}\big(\pi_i-\lambda_i(X)\big)^2+\sum_{i=q+1}^{r}\big(\pi_i-\lambda_i(X)\big)^2+\sum_{i=r+1}^{N}\lambda_i(X)^2\\ &\geqslant \sum_{i=q+1}^{r}\big(\pi_i^2-\underbrace{2\pi_i\lambda_i(X)}_{\leqslant 0}+\lambda_i(X)^2\big)+\sum_{i=r+1}^{N}\lambda_i(X)^2\\ &\geqslant \sum_{i=q+1}^{N}\lambda_i(X)^2. \end{aligned} \tag{6.32}$$

In turn, appealing to the condition for equality in the rearrangement inequality, we obtain

$$\begin{aligned}\|z-\lambda(X)\|^2=\sum_{i=q+1}^{N}\lambda_i(X)^2 &\Leftrightarrow \begin{cases}\|z-\lambda(X)\|=\left\|z^{\downarrow}-\lambda(X)\right\|\\ \sum_{i=1}^{q}\big(\pi_i-\lambda_i(X)\big)^2=\sum_{i=q+1}^{r}\pi_i^2=\sum_{i=q+1}^{r}\pi_i\lambda_i(X)=0\end{cases}\\ &\Leftrightarrow \begin{cases}\big(\exists\, P\in \mathsf{P}_X^N\big)\ z=Pz^{\downarrow}\\ \big(\forall i\in\{1,\ldots,q\}\big)\ \pi_i=\lambda_i(X)\\ \big(\forall i\in\{q+1,\ldots,r\}\big)\ \pi_i=0\end{cases}\\ &\Leftrightarrow \big(\exists\, P\in \mathsf{P}_X^N\big)\ z=P\big(\lambda_1(X),\ldots,\lambda_q(X),0,\ldots,0\big)=P\overline{x}, \end{aligned} \tag{6.33}$$

which entails that $\operatorname{Proj}_D\lambda(X)=\big\{P\overline{x}\mid P\in\mathsf{P}_X^N\big\}$.

(b) $q\geqslant r$: Argue as in case (a).

(ii): Upon setting $(\zeta_i)_{1\leqslant i\leqslant N}=z$, we deduce that

$$\|z-\lambda(X)\|^2=\sum_{i=1}^{N}\big(\zeta_i-\lambda_i(X)\big)^2=\sum_{i=1}^{N}\big(\zeta_i^2-\underbrace{2\zeta_i\lambda_i(X)}_{\leqslant 0}+\lambda_i(X)^2\big)\geqslant\sum_{i=1}^{N}\lambda_i(X)^2 \tag{6.34}$$

and, in turn, that

$$\|z-\lambda(X)\|^2=\sum_{i=1}^{N}\lambda_i(X)^2 \quad\Leftrightarrow\quad \sum_{i=1}^{N}\zeta_i^2=\sum_{i=1}^{N}\zeta_i\lambda_i(X)=0 \quad\Leftrightarrow\quad \big(\forall i\in\{1,\ldots,N\}\big)\ \zeta_i=0. \tag{6.35}$$

Thus $\operatorname{Proj}_D\lambda(X)=\{0\}$. ☐

Remark 6.6. Here are several comments on Example 6.5.

(i) Set

$$\mathscr{R}=\big\{X\in\mathfrak{H}\mid \operatorname{Rank}X\leqslant r\big\}\quad\text{and}\quad \mathscr{C}=\big\{X\in\mathfrak{H}\mid \lambda(X)\in\mathbb{R}_+^N\big\}. \tag{6.36}$$

Along the same lines of the proof of Example 6.5, one can show that

$$\operatorname{Proj}_{\mathscr{D}}X=\operatorname{Proj}_{\mathscr{R}}\big(\operatorname{Proj}_{\mathscr{C}}X\big). \tag{6.37}$$

In general, nevertheless, given nonempty closed subsets R and C of a Euclidean space,

$$\operatorname{Proj}_{R\cap C}\neq\operatorname{Proj}_R\circ\operatorname{Proj}_C. \tag{6.38}$$

(ii) Choosing $\mathfrak{H}$ to be the Euclidean Jordan algebra of Hermitian matrices (see Example 2.6), we obtain an expression for the projector onto the set of positive semidefinite matrices of rank at most r, that is, the set

$$\{X \in \mathrm{H}^N(\mathbb{K}) \mid \lambda(X) \in \mathbb{R}^N_+ \text{ and } \operatorname{Rank} X \leqslant r\}, \tag{6.39}$$

which is of interest in low-rank factor analysis [9]. Such expression is known in the real and complex cases (see, e.g., [14, Lemma 6.1]) but appears to be new in the quaternion case.

7. Conclusion

We have introduced an abstract framework of spectral decomposition systems that covers a wide range of related settings such as eigenvalue or singular value decompositions of real, complex, and quaternion matrices, or Euclidean Jordan algebras. In this framework, we have derived results on convex analytical properties and objects related to spectral functions in spectral decomposition systems that unify corresponding results in individual settings studied in the literature as well as new ones. Along the way, we have obtained a generalization of the Ky Fan majorization theorem and an abstract reduced minimization principle that can be used to derive novel results on spectral functions. We have also studied Bregman proximity operators of spectral functions, which for non-convex functions are new even in the case of Hermitian matrices.

Future work will be concerned with applying these representations for proximal point and splitting algorithms [6, 7, 15] and with extending our results to objects of (nonconvex) variational analysis such as the Clarke, Fréchet, and limiting subdifferentials [46, 15] of spectral functions in spectral decomposition systems.

References

[1] M. Baes, Convexity and differentiability properties of spectral functions and spectral mappings on Euclidean Jordan algebras, *Linear Algebra Appl.* 422 (2007), 664–700, doi:10.1016/j.laa.2006.11.025.

[2] J. M. Ball, Differentiability properties of symmetric and isotropic functions, *Duke Math. J.* 51 (1984), 699–728, doi:10.1215/s0012-7094-84-05134-2.

[3] H. H. Bauschke and J. M. Borwein, Legendre functions and the method of random Bregman projections, *J. Convex Anal.* 4 (1997), 27–67, https://www.heldermann-verlag.de/jca/jca04/jca04002.pdf.

[4] H. H. Bauschke, J. M. Borwein, and P. L. Combettes, Bregman monotone optimization algorithms, *SIAM J. Control Optim.* 42 (2003), 596–636, doi:10.1137/s0363012902407120.

[5] H. H. Bauschke, M. N. Bùi, and X. Wang, On sums and convex combinations of projectors onto convex sets, *J. Approx. Theory* 242 (2019), 31–57, doi:10.1016/j.jat.2019.02.001.

[6] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, Springer, Cham, Switzerland, 2nd edition, 2017, doi:10.1007/978-3-319-48311-5.

[7] A. Beck, *First-Order Methods in Optimization*, SIAM, Philadelphia, PA, 2017, doi:10.1137/1.9781611974997.

[8] A. Benfenati, E. Chouzenoux, and J. C. Pesquet, Proximal approaches for matrix optimization problems: Application to robust precision matrix estimation, *Signal Process.* 169 (2020), 107417, doi:10.1016/j.sigpro.2019.107417.

[9] D. Bertsimas, M. S. Copenhaver, and R. Mazumder, Certifiably optimal low rank factor analysis, *J. Machine Learn. Res.* 18 (2017), 1–53, http://jmlr.org/papers/v18/15-613.html.

[10] J. F. Cai, E. J. Candès, and Z. Shen, A singular value thresholding algorithm for matrix completion, *SIAM J. Optim.* 20 (2010), 1956–1982, doi:10.1137/080738970.

[11] T. S. T. Chan and Y. H. Yang, Complex and quaternionic principal component pursuit and its application to audio separation, *IEEE Signal Process. Lett.* 23 (2016), 287–291, doi:10.1109/lsp.2016.2514845.

[12] Y. Chen, X. Xiao, and Y. Zhou, Low-rank quaternion approximation for color image processing, *IEEE. Trans. Image Process.* 29 (2019), 1426–1439, doi:10.1109/tip.2019.2941319.

[13] G. Chierchia, E. Chouzenoux, P. L. Combettes, and J. C. Pesquet, The Proximity Operator Repository, 2016, https://proximity-operator.net/.

[14] K. L. Clarkson and D. P. Woodruff, Low-rank PSD approximation in input-sparsity time, in *Proceedings of the 2017 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, P. N. Klein (ed.), SIAM, Philadelphia, PA, 2017, 2061–2072, doi:10.1137/1.9781611974782.134.

[15] C. Clason and T. Valkonen, Introduction to Nonsmooth Analysis and Optimization, 2024, arXiv:2001.00216.

[16] C. Coey, L. Kapelevich, and J. P. Vielma, Conic optimization with spectral functions on Euclidean Jordan algebras, *Math. Oper. Res.* 48 (2023), 1906–1933, doi:10.1287/moor.2022.1324.

[17] B. Dacorogna, *Direct Methods in the Calculus of Variations*, Springer, New York, 2nd edition, 2008, doi:10.1007/978-0-387-55249-1.

[18] B. Dacorogna and P. Maréchal, Convex $\mathrm{SO}(N) \times \mathrm{SO}(n)$-invariant functions and refinements of von Neumann's inequality, *Ann. Fac. Sci. Toulouse Math. (6)* 16 (2007), 71–89, doi:10.5802/afst.1139.

[19] C. Davis, All convex invariant functions of Hermitian matrices, *Arch. Math.* 8 (1957), 276–278, doi:10.1007/bf01898787.

[20] D. Drusvyatskiy and C. Paquette, Variational analysis of spectral functions simplified, *J. Convex Anal.* 25 (2018), 119–134, https://www.heldermann.de/JCA/JCA25/JCA251/jca25008.htm.

[21] C. Eckart and G. Young, The approximation of one matrix by another of lower rank, *Psychometrika* 1 (1936), 211–218, doi:10.1007/bf02288367.

[22] J. Faraut and A. Korányi, *Analysis on Symmetric Cones*, Oxford University Press, Oxford, 1994, doi:10.1093/oso/9780198534778.001.0001.

[23] C. Garner, G. Lerman, and S. Zhang, Spectrally constrained optimization, *J. Sci. Comput.* 100 (2024), 29 pp., doi:10.1007/s10915-024-02636-9.

[24] M. S. Gowda, Simultaneous spectral decomposition in Euclidean Jordan algebras and related systems, *Linear Multilinear Algebra* 70 (2022), 6535–6547, doi:10.1080/03081087.2021.1960259.

[25] M. S. Gowda and J. Jeong, Commutation principles in Euclidean Jordan algebras and normal decomposition systems, *SIAM J. Optim.* 27 (2017), 1390–1402, doi:10.1137/16m1071006.

[26] J. B. Hiriart-Urruty and D. Ye, Sensitivity analysis of all eigenvalues of a symmetric matrix, *Numer. Math.* 70 (1995), 45–72, doi:10.1007/s002110050109.

[27] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, New York, 2nd edition, 2013, doi:10.1017/cbo9780511810817.

[28] P. O. Hoyer, Non-negative matrix factorization with sparseness constraints, *J. Mach. Learn. Res.* 5 (2003/04), 1457–1469, https://www.jmlr.org/papers/volume5/hoyer04a/hoyer04a.pdf.

[29] H. Hu, R. Sotirov, and H. Wolkowicz, Facial reduction for symmetry reduced semidefinite and doubly nonnegative programs, *Math. Program.* A200 (2023), 475–529, doi:10.1007/s10107-022-01890-9.

[30] J. Jeong and M. S. Gowda, Spectral sets and functions on Euclidean Jordan algebras, *Linear Algebra Appl.* 518 (2017), 31–56, doi:10.1016/j.laa.2016.12.020.

[31] Z. Jia, M. K. Ng, and G. J. Song, Robust quaternion matrix completion with applications to image inpainting, *Numer. Linear Algebra Appl.* 26 (2019), e2245, doi:10.1002/nla.2245.

[32] N. Kovachki, Z. Li, B. Liu, K. Azizzadenesheli, K. Bhattacharya, A. Stuart, and A. Anandkumar, Neural operator: learning maps between function spaces with applications to PDEs, *J. Mach. Learn. Res.* 24 (2023), 1–97, https://www.jmlr.org/papers/volume24/21-1524/21-1524.pdf.

[33] A. S. Lewis, The convex analysis of unitarily invariant matrix functions, *J. Convex Anal.* 2 (1995), 173–183, https://www.heldermann-verlag.de/jca/jca02/jca02012.pdf.

[34] A. S. Lewis, Convex analysis on the Hermitian matrices, *SIAM J. Optim.* 6 (1996), 164–177, doi:10.1137/0806009.

[35] A. S. Lewis, Group invariance and convex matrix analysis, *SIAM J. Matrix Anal. Appl.* 17 (1996), 927–949, doi:10.1137/s0895479895283173.

[36] A. S. Lewis, Convex analysis on Cartan subspaces, *Nonlinear Anal.* 42 (2000), 813–820, doi:10.1016/s0362-546x(99)00126-1.

[37] A. S. Lewis and H. S. Sendov, Nonsmooth analysis of singular values. Part I: Theory, *Set-Valued Var. Anal.* 13 (2005), 213–241, doi:10.1007/s11228-004-7197-7.

[38] B. F. Lourenço and A. Takeda, Generalized subdifferentials of spectral functions over Euclidean Jordan algebras, *SIAM J. Optim.* 30 (2020), 3387–3414, doi:10.1137/19m1245001.

[39] I. Markovsky, Structured low-rank approximation and its applications, *Automatica J. IFAC* 44 (2008), 891–909, doi:10.1016/j.automatica.2007.09.011.

[40] A. W. Marshall, I. Olkin, and B. C. Arnold, *Inequalities: Theory of Majorization and Its Applications*, Springer, New York, 2nd edition, 2011, doi:10.1007/978-0-387-68276-1.

[41] B. S. Mordukhovich, *Variational Analysis and Generalized Differentiation I*, Springer, Heidelberg, 2006, doi:10.1007/3-540-31247-1.

[42] B. S. Mordukhovich and N. M. Nam, *Convex Analysis and Beyond. Volume I: Basic Theory*, Springer, Cham, Switzerland, 2022, doi:10.1007/978-3-030-94785-9.

[43] M. Orlitzky, Proscribed normal decompositions of Euclidean Jordan algebras, *J. Convex Anal.* 29 (2022), 755–766, https://www.heldermann.de/JCA/JCA29/JCA293/jca29043.htm.

[44] T. Pennanen, Graph-convex mappings and K-convex functions, *J. Convex Anal.* 6 (1999), 235–266, https://www.heldermann.de/JCA/JCA06/JCA062/jca06015.htm.

[45] R. T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, NJ, 1970, doi:10.1515/9781400873173.

[46] R. T. Rockafellar and R. J. B. Wets, *Variational Analysis*, Springer, Heidelberg, 2009, doi:10.1007/978-3-642-02431-3.

[47] L. Rodman, *Topics in Quaternion Linear Algebra*, Princeton University Press, Princeton, NJ, 2014, doi:10.1515/9781400852741.

[48] P. Rosakis, Characterization of convex isotropic functions, *J. Elasticity* 49 (1998), 257–267, doi:10.1023/a:1007468902439.

[49] H. S. Sendov, Nonsmooth analysis of Lorentz invariant functions, *SIAM J. Optim.* 18 (2007), 1106–1127, doi:10.1137/060658370.

[50] D. Sun and J. Sun, Löwner's operator and spectral functions in Euclidean Jordan algebras, *Math. Oper. Res.* 33 (2008), 421–445, doi:10.1287/moor.1070.0300.

[51] T. Y. Tam, A unified extension of two results of Ky Fan on the sum of matrices, *Proc. Amer. Math. Soc.* 126 (1998), 2607–2614, doi:10.1090/s0002-9939-98-04770-4.

[52] C. M. Theobald, An inequality for the trace of the product of two symmetric matrices, *Math. Proc. Cambridge Philos. Soc.* 77 (1975), 265–267, doi:10.1017/s0305004100051070.

[53] R. C. Thompson, Singular values, diagonal elements, and convexity, *SIAM J. Appl. Math.* 32 (1977), 39–63, doi:10.1137/0132003.

[54] F. Vincent, Une note sur les fonctions invariantes, *Ann. Fac. Sci. Toulouse (6)* 6 (1997), 357–363, http://www.numdam.org/item/?id=AFST_1997_6_6_2_357_0.

[55] J. von Neumann, Some matrix-inequalities and metrization of matrix-space, *Tomsk Univ. Rev.* 1 (1937), 286–300.

[56] S. Yang, Z. Lu, X. Shen, P. Wonka, and J. Ye, Fused multiple graphical lasso, *SIAM J. Optim.* 25 (2015), 916–943, doi:10.1137/130936397.

[57] K. Ziętak, On the characterization of the extremal points of the unit sphere of matrices, *Linear Algebra Appl.* 106 (1988), 57–75, doi:10.1016/0024-3795(88)90023-7.