Clarke Generalized Jacobian of the Projection onto the Cone of Positive Semidefinite Matrices — Source link

Jérôme Malick, Hristo S. Sendov

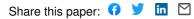**Institutions:** University of Guelph

Related papers:

- A nonsmooth version of Newton's method

- Optimization and nonsmooth analysis

- Semismooth Matrix-Valued Functions

- The Strong Second-Order Sufficient Condition and Constraint Nondegeneracy in Nonlinear Semidefinite Programming and Their Implications

- Nonlinear Programming

Clarke Generalized Jacobian of the Projection onto the Cone of Positive Semidefinite Matrices

Jérôme Malick, Hristo S. Sendov

HAL Id: inria-00070517
https://hal.inria.fr/inria-00070517v2

Submitted on 25 Mar 2013

Clarke Generalized Jacobian of the Projection onto the Cone of Positive Semidefinite Matrices

Jérôme Malick[*]   Hristo S. Sendov[†]

August 10, 2005

Abstract

This paper studies the differentiability properties of the projection onto the cone of positive semidefinite matrices. In particular, the expression of the Clarke generalized Jacobian of the projection at any symmetric matrix is given.

Keywords: Positive semidefinite matrices, projection, spectral function, differentiability, Clarke generalized Jacobian, eigenvalue optimization, perturbation theory, nonsmooth analysis.

Mathematics Subject Classification (2000): primary: 49R50, 49J52, secondary: 26E25, 47A75.

1 Introduction

1.1 Motivations

Let M^n be the Euclidean space of $n \times n$ real matrices equipped with the standard inner product $\langle A, B \rangle = \text{trace}\,(A^\top B) = \sum_{ij} A_{ij} B_{ij}$. Denote by S^n

---

[*]INRIA, Montbonnot, F-38334 Saint Ismier, France. Email: `jerome.malick@inria.fr`

[†]Department of Mathematics and Statistics, University of Guelph, Guelph, Ontario, Canada N1G 2W1. Email: `hssendov@uoguelph.ca`. Research supported by NSERC.

1

the subspace of symmetric matrices and by S^n_+ the cone of symmetric positive semi-definite matrices. Denote the projection from S^n onto S^n_+ by $\mathrm{P}_{S^n_+}$. Since S^n_+ is a closed and convex set, this projection is well defined and globally Lipschitz on S^n, see [7, Proposition A.3.1.3]. By the Rademacher Theorem, [3, Chapter 3], $\mathrm{P}_{S^n_+}$ is differentiable almost everywhere (in Lebesgue sense) on S^n. Natural questions are therefore arising:

(i) Where is $\mathrm{P}_{S^n_+}$ differentiable?

(ii) What is the Jacobian, $\nabla \mathrm{P}_{S^n_+}(X)$, of $\mathrm{P}_{S^n_+}$ at a point $X \in S^n$ where $\mathrm{P}_{S^n_+}$ is differentiable?

(iii) What is the Clarke generalized Jacobian, $\partial_C \mathrm{P}_{S^n_+}(X)$, of $\mathrm{P}_{S^n_+}$ at a point $X \in S^n$ where $\mathrm{P}_{S^n_+}$ is not differentiable?

These are the questions that we answer is this paper. In particular, using differentiability properties of *spectral functions*, our development culminates in Theorem 3.7 giving a complete description of the Clarke generalized Jacobian.

In the general context of the projection onto a closed convex set, Question (iii) above was first formulated in [5].

1.2   Basic notation

By $\mathbb{R}^n_\downarrow$ we denote the set of vectors from $\mathbb{R}^n$ with coordinates ordered in non increasing order. The function $\lambda : S^n \to \mathbb{R}^n_\downarrow$ maps a symmetric matrix to the vector of its eigenvalues, ordered non increasingly:

$$\lambda_1(X) \geq \lambda_2(X) \geq \cdots \geq \lambda_n(X).$$

For any vector x in $\mathbb{R}^n$, by $\mathrm{Diag}\, x$ we denote the diagonal matrix with vector x on the main diagonal, and by $\mathrm{diag} : M^n \to \mathbb{R}^n$ we denote its dual operator: $\mathrm{diag}\,(X) = (X^{11}, \ldots, X^{nn})$.

By p_X we denote the number of the positive eigenvalues of $X \in S^n$, by q_X the number of negative ones, and by n_X the multiplicity of 0 as an eigenvalue. Clearly

(1) $$p_X + q_X = n - n_X = \text{rank of } X.$$

We denote by $O(n)$ the compact group of $n \times n$ orthogonal matrices and by P^n the finite subgroup of permutations matrices. We denote the set of

all orthogonal matrices that give the ordered spectral decomposition of X by $O(n)^X$, that is,

$$(2) \qquad O(n)^X = \{U \in O(n) \mid X = U(\mathrm{Diag}\,\lambda(X))U^\top\}.$$

For instance, observe that $O(n)^X = O(n)$ if $X = 0$.

Spectral functions are functions on S^n invariant under conjugation by elements from $O(n)$: $F(X) = F(U^\top X U)$. Restricting a spectral function F to the subspace of diagonal matrices shows that

$$\forall\, X \in S^n, \qquad F(X) = f(\lambda(X)),$$

where $f(x) = F(\mathrm{Diag}\,x)$ is a symmetric function on $\mathbb{R}^n$, that is, $f(x) = f(Px)$ for any $P \in P^n$. Thus, spectral and symmetric functions are in one-to-one correspondence. Many properties of F are inherited from its corresponding f and vice versa. For example, F is convex on S^n if, and only if, f is convex on $\mathbb{R}^n$, see [2]. Moreover, F is differentiable or twice differentiable at $X \in S^n$ if, and only if, f is such at $\lambda(X)$, see Theorem 1.5 below.

1.3   Tensor notation

To conveniently handle the formulae for the first and second derivatives of spectral functions $f \circ \lambda$, we use some of the notation introduced in [11].

A k-tensor on $\mathbb{R}^n$ is a multi-linear, real valued function on $\mathbb{R}^n \times \cdots \times \mathbb{R}^n$ (k-times). The set of all k-tensors on $\mathbb{R}^n$ will be denoted by $T^{k,n}$. By $\mathbb{N}_n$ we will denote the set $\{1, 2, \ldots, n\}$. For a fixed basis in $\mathbb{R}^n$, by $T^{i_1 \ldots i_k}$ we denote the $(i_1, \ldots, i_k)$-th entry of $T \in T^{k,n}$. Vectors are viewed as 1-tensors and matrices as 2-tensors, in the natural way. In this paper we work with tensors of dimension at most four.

By $\{H_{pq} : 1 \le p, q \le n\}$ we denote the standard basis in the space M^n. That is, the matrices H_{pq} are such that $(H_{pq})^{ij}$ is one if $(i, j) = (p, q)$, and zero otherwise.

Definition 1.1 (σ-Hadamard product) For each permutation σ on $\mathbb{N}_2$, we define σ-*Hadamard* product, $H_1 \circ_\sigma H_2$, between the two matrices H_1 and H_2 to be a 2-tensor on $\mathbb{R}^n$ as follows.

$$(3) \qquad (H_1 \circ_\sigma H_2)^{i_1 i_2} = H_1^{i_1 i_{\sigma^{-1}(1)}} H_2^{i_2 i_{\sigma^{-1}(2)}}.$$

In other words, when $\sigma = (1)(2)$ we have $H_1 \circ_{(1)(2)} H_2 = \mathrm{diag}\,(H_1)\mathrm{diag}\,(H_2)^\top$ and when $\sigma = (12)$ we have $H_1 \circ_{(12)} H_2 = H_1 \circ H_2^\top$, where $A \circ B = [A^{ij}B^{ij}]$ is the usual Hadamard product between $A = [A^{ij}]$ and $B = [B^{ij}]$. Analogously when σ is a permutation on $\mathbb{N}_1$, that is $\sigma = (1)$, we define $\circ_{(1)} H = \mathrm{diag}\,H$.

Let T be an arbitrary matrix in M^n and let σ be a permutation on $\mathbb{N}_2$. We define $\mathrm{Diag}^\sigma T$ to be a 4-tensor on $\mathbb{R}^n$ in the following way

$$(4) \qquad (\mathrm{Diag}^\sigma T)^{i_1 i_2}_{j_1 j_2} = \begin{cases} T^{i_1 i_2}, & \text{if } i_s = j_{\sigma(s)} \text{ for } s = 1, 2, \\ 0, & \text{otherwise.} \end{cases}$$

Similarly, when $T \in \mathbb{R}^n$ and $\sigma = (1)$ we define $\mathrm{Diag}^{(1)} T = \mathrm{Diag}\,T$.

Notice that any 4-tensor, T, on $\mathbb{R}^n$ can naturally be viewed as a 2-tensor on M^n, and vice versa, for example

$$T[H_1, H_2] = \sum_{p_1, q_1 = 1}^n \sum_{p_2, q_2 = 1}^n T^{p_1 p_2}_{q_1 q_2} H_1^{p_1 q_1} H_2^{p_2 q_2}.$$

Define inner product between two tensors in $T^{k,n}$ in the usual way:

$$\langle T_1, T_2 \rangle = \sum_{p_1, \ldots, p_k = 1}^n T_1^{p_1 \ldots p_k} T_2^{p_1 \ldots p_k}.$$

Definition 1.2 (Conjugation) We define an action of the orthogonal group $O(n)$ on the space of all k-tensors on $\mathbb{R}^n$, that will be denoted by $UTU^\top$ or by $U \cdot T$:

$$(5) \qquad (UTU^\top)^{i_1 \ldots i_k} = \sum_{p_1, \ldots, p_k = 1}^n \left(T^{p_1 \ldots p_k} U^{i_1 p_1} \cdots U^{i_k p_k} \right).$$

It is easy to see that this action is, as expected, norm preserving and associative: $V(UTU^\top)V^\top = (VU)T(VU)^\top$ for all $U, V \in O(n)$. The Diag^σ operator, the σ-Hadamard product, and conjugation by an orthogonal matrix are connected in the following way (see [11, Theorem 4.4] for a generalization).

Lemma 1.3 *For any vector $T \in \mathbb{R}^n$, any matrix H in M^n and any orthogonal matrix $U \in O(n)$, we have:*

$$\langle T, \circ_{(1)}(U^\top H U) \rangle = \left(U(\mathrm{Diag}^{(1)} T)U^\top \right)[H].$$

4

For any matrix $T \in M^n$, any matrices H_1 and H_2 in M^n, any orthogonal matrix $U \in O(n)$ and any permutation σ on $\mathbb{N}_2$, we have:

$$\langle T, (U^\top H_1 U) \circ_\sigma (U^\top H_2 U) \rangle = \left(U (\mathrm{Diag}^\sigma T) U^\top \right)[H_1, H_2].$$

We conclude this section with an easy lemma about conjugation. In this lemma and throughout the paper, $\mathrm{conv}\, S$ denotes the convex hull of the set S (a subset of a real vector space).

Lemma 1.4 *Let S be a subset of the vector space of k-tensors on $\mathbb{R}^n$ and G be a subset of the group $O(n)$. Then*

(i) $G \cdot (\mathrm{conv}\, S) \subseteq \mathrm{conv}\,(G \cdot S)$,

(ii) $\mathrm{conv}\,(G \cdot (\mathrm{conv}\, S)) \subseteq \mathrm{conv}\,(G \cdot S)$.

Proof. Let $g \in G$, observe on (5) that the function $T \mapsto g \cdot T$ is linear. Therefore
$$g \cdot \mathrm{conv}\, S = \mathrm{conv}\, g \cdot S \subset \mathrm{conv}\, G \cdot S.$$

This yields the first inclusion. Since the right-hand side is convex, the second inclusion follows. ∎

1.4 Gradient and Hessian of spectral functions

Theorem 1.5 (Derivatives of spectral functions) a) *Let $X \in S^n$ and let U be any element in $O(n)^X$. A symmetric function $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable at $\lambda(X)$ if, and only if, $f \circ \lambda$ is differentiable at X. In that case, the gradient of $f \circ \lambda$ is given by the following well-defined (independent of the choice of U) formula*

$$(6) \qquad \nabla(f \circ \lambda)(X) = U \left(\mathrm{Diag}^{(1)} \nabla f(\lambda(X)) \right) U^\top,$$

b) *A symmetric function $f : \mathbb{R}^n \to \mathbb{R}$ is twice differentiable at $\lambda(X)$ if, and only if, $f \circ \lambda$ is twice differentiable at X. In that case, the Hessian of $f \circ \lambda$ is given by the following well-defined formula*

$$(7) \quad \nabla^2(f \circ \lambda)(X) = U \left(\mathrm{Diag}^{(1)(2)} \nabla^2 f(\lambda(X)) + \mathrm{Diag}^{(12)} \mathcal{A}(\lambda(X)) \right) U^\top,$$

where $\mathcal{A} : \mathbb{R}^n \to S^n$ is the map defined by

$$(8) \qquad \mathcal{A}^{ij}(x) = \begin{cases} 0, & if \ i = j \\ f_{ii}''(x) - f_{ij}''(x), & if \ i \neq j \ but \ x_i = x_j \\ \dfrac{f_i'(x) - f_j'(x)}{x_i - x_j}, & otherwise. \end{cases}$$

Proof. The differentiability result is established in [8, Theorem 1.1] and the twice differentiability in [9, Theorem 3.3]. The particular expression for the Hessian is given in [11, Example 4.9]. ■

Notice that the entries of $\nabla^2 f(x)$ and $\mathcal{A}(x)$ do not depend on the eigenvalues but only on the partial derivatives of f up to order two. Thus, Formulae (6) and (7) depend on the eigenvalues only through the orthogonal matrix U and the composition with $\lambda(X)$. One should keep in mind that, in the formula for the Hessian, U conjugates a four dimensional tensor on $\mathbb{R}^n$.

The proof of the next corollary is an interesting application of the differentiability properties of spectral functions.

Corollary 1.6 a) *The set of all singular matrices in S^n has Lebesgue measure zero.* **b)** *The set of all matrices in S^n with repeated eigenvalues has Lebesgue measure zero.*

Proof. a) Consider the symmetric convex function $f(x) = \sum_{i=1}^n |x_i|$ defined for all $x \in \mathbb{R}^n$. As we mentioned at the end of Section 1.2, the spectral function $f \circ \lambda$ is convex. Moreover, by Theorem 1.5, $f \circ \lambda$ is differentiable at X if, and only if, f is differentiable at $\lambda(X)$. Observe now that f is differentiable at $x \in \mathbb{R}^n$ if, and only if, all x_i are nonzero. Thus the convex function $f \circ \lambda$ is differentiable at $X \in S^n$ if, and only if, X is nonsingular. Since a convex function on S^n (or $\mathbb{R}^n$) is differentiable almost everywhere [10, 6] we are done.

b) An analogous argument for the function $f(x) = \sum_{i<j} |x_i - x_j|$. ■

2 Differentiability of the projection

The main idea in this section is the realization that the projection onto the cone of positive semi-definite matrices is the gradient mapping of a spectral

function (see Subsection 2.2). Then to investigate the smoothness of the projection, we employ Formula (7) for the Hessian of spectral functions.

2.1 Explicit expression of the projection

It is well-known, see [4] for example, that the projection operator onto the cone of positive semi-definite matrices allows a closed form representation. Indeed, projecting X amounts to replacing the negative eigenvalues by 0 in the spectral decomposition of X. Recall that the polar cone $(S_+^n)^\circ = \{B \in S^n, \ \langle A, B \rangle \leq 0\}$ is actually $S_-^n = -S_+^n$, the cone of negative semi-definite matrices.

Theorem 2.1 (Expression for $\mathrm{P}_{S_+^n}$) *Let $X \in S^n$ and let U be any element in $O(n)^X$. Then, the projection of X onto S_+^n is given by*

$$(9) \qquad \mathrm{P}_{S_+^n}(X) = U\big(\mathrm{Diag}\,(\max\{\lambda_1(X), 0\}, \ldots, \max\{\lambda_n(X), 0\})\big)U^\top.$$

The equivalent result for the projection onto S_-^n replaces the positive eigenvalues with 0.

It is worth mentioning that this result is also a straightforward application of Moreau's theorem (see [7, Theorem 3.2.5 and Exercise A.15]) which in our case can be stated as follows.

Theorem 2.2 (J.-J. Moreau) *For any three matrices X, X_1, and X_2 in S^n the properties below are equivalent:*

(i) $X = X_1 + X_2$ with $X_1 \in S_+^n$, $X_2 \in S_-^n$, and $\langle X_1, X_2 \rangle = 0$;

(ii) $X_1 = \mathrm{P}_{S_+^n}(X)$ and $X_2 = \mathrm{P}_{S_-^n}(X)$.

In particular, $\mathrm{P}_{S_+^n} + \mathrm{P}_{S_-^n} = I_n$.

2.2 A primitive function of the projection

Denote by $I_{S_-^n}$ the indicator function on S_-^n. Let $\Delta : S^n \to \mathbb{R}$ be the *Moreau regularization* of $I_{S_-^n}$ defined by

$$(10) \qquad \Delta(X) = \min_{Y \in S^n} \left\{\frac{1}{2}\|X - Y\|^2 + I_{S_-^n}(Y)\right\}.$$

In Lemma 2.3, we observe that $\Delta(X)$ is a spectral function and in Lemma 2.4 we show that $\mathrm{P}_{S_+^n}$ is its gradient.

Lemma 2.3 *The function Δ is the spectral function $\Delta = \delta \circ \lambda$, where*

$$(11) \qquad\qquad \delta(x_1, \ldots, x_n) = \frac{1}{2} \sum_{i=1}^n \max\{0, x_i\}^2 .$$

Proof.   For any $X \in S^n$, write

$$
\begin{aligned}
\Delta(X) &= \tfrac{1}{2}\|X - \mathrm{P}_{S^n_-}(X)\|^2 && \text{[by definition of $\mathrm{P}_{S^n_-}(X)$]}\\
&= \tfrac{1}{2}\|\mathrm{P}_{S^n_+}(X)\|^2 && \text{[by Theorem 2.2]}\\
&= \tfrac{1}{2}\sum_{i=1}^n \max\{0, \lambda_i(X)\}^2 = \delta\big(\lambda(X)\big). && \text{[by Theorem 2.1]}
\end{aligned}
$$

This ends the proof. ∎

Lemma 2.4 *The function Δ defined by (10) is convex and differentiable. Moreover, its gradient at $X \in S^n$ is $\mathrm{P}_{S^n_+}(X)$.*

Proof.   The function δ defined by (11) is obviously convex on $\mathbb{R}^n$. Then [8, Corollary 2.7] yields the convexity of Δ on S^n. The differentiability of Δ and the formula for the gradient follow from Theorem 1.5. ∎

Remark 2.5 This result is actually independent of spectral functions properties: it holds if S^n_+ is replaced by any closed, convex cone. See more about the properties of the Moreau regularization in [6, Ch. XV]. ∎

2.3   Differentiability of the projection

Studying differentiability of $\mathrm{P}_{S^n_+}$ amounts to studying second-order differentiability of $\Delta = \delta \circ \lambda$. The point is that $\delta \circ \lambda$ is a spectral function and this makes things easier: we have a characterization of twice differentiable spectral functions and an easy-to-use formula for their Hessian.

   We begin by recalling the nature of objects that come into play. When it exists, the derivative of $\mathrm{P}_{S^n_+}$ at X, $\nabla \mathrm{P}_{S^n_+}(X)$ is a linear function from S^n to S^n. The second derivative of $\delta \circ \lambda$ at X, $\nabla^2(\delta \circ \lambda)(X)$, is a 2-tensor on S^n. Notice that these derivatives are connected by the following equation

$$(12) \qquad \forall\, H_1, H_2 \in S^n, \qquad \nabla^2(\delta \circ \lambda)(X)[H_1, H_2] = \langle \nabla \mathrm{P}_{S^n_+}(X)[H_1], H_2 \rangle.$$

In the rest of the paper, we identify these objects via this identity. We now introduce a map that plays an important role in the description of the gradient and the Clarke generalized Jacobian of $\mathrm{P}_{S^n_+}$.

Definition 2.6 (The map $\mathcal{B}(x)$) For any $x \in \mathbb{R}^n_\downarrow$ with non-zeros entries (p positive, q negative, where $p + q = n$), we define the matrix $\mathcal{B}(x) \in S^n$ as follows:

- If $i \leq p$, $j \leq p$, then $\mathcal{B}^{ij}(x) = 1$;

- If $i > p$, $j > p$, then $\mathcal{B}^{ij}(x) = 0$;

- If $i \leq p$, $j > p$, then $\mathcal{B}^{ij}(x) = \dfrac{x_i}{x_i - x_j}$;

- If $i > p$, $j \leq p$, then $\mathcal{B}^{ij}(x) = \dfrac{x_j}{x_j - x_i}$.

Theorem 2.7 *For any $X \in S^n$ the following are equivalent:*

(i) $\mathrm{P}_{S^n_+}$ *is differentiable at X,*

(ii) $\Delta = \delta \circ \lambda$ *is twice differentiable at X,*

(iii) X *is non-singular.*

In that case, the functions $\delta \circ \lambda$ and $\mathrm{P}_{S^n_+}$ are actually C^∞ in a neighbourhood of X, and

(13)
$$\nabla \mathrm{P}_{S^n_+}(X) = U\big(\mathrm{Diag}^{(12)}\,\mathcal{B}(\lambda(X))\big)U^\top,$$

for any $U \in O(n)^X$.

Proof. By Lemma 2.4, the differentiability of $\mathrm{P}_{S^n_+}$ at $X \in S^n$ is equivalent to the twice differentiability of Δ at X. By (11), the function δ is twice differentiable at $x = (x_i)_i$ if, and only if, each x_i is nonzero. Theorem 1.5 implies that Δ is twice differentiable at X if, and only if, the function δ is twice differentiable at $\lambda(X)$, and that happens if, and only if, 0 is not an eigenvalue of X.

Since $\delta(x)$ is C^∞ at a vector x with nonzero coordinates, we obtain by [15, Section 3] that $\delta \circ \lambda$ and $\mathrm{P}_{S^n_+}$ are C^∞ at a non-singular matrix X.

9

We now compute the Hessian of $\delta \circ \lambda$ at a non-singular $X \in S^n$. For a vector $x \in \mathbb{R}^n_\downarrow$ with nonzero coordinates (the first p strictly positive), we have:

$$(14) \qquad \nabla^2 \delta(x) = \left(\begin{array}{c|c} I_p & 0 \\ \hline 0 & 0 \end{array} \right),$$

and it is easy to see that

$$(15) \qquad \mathrm{Diag}^{(1)(2)} \nabla^2 \delta(x) = \mathrm{Diag}^{(12)} \nabla^2 \delta(x).$$

Now, using (8), we compute the matrix $\mathcal{A}(x)$:

- If $i = j$, then $\mathcal{A}^{ij}(x) = 0$.

- If $i \leq p, j \leq p$, with $i \neq j$, then $\mathcal{A}^{ij}(x) = 1$.

- If $i > p, j > p$, then $\mathcal{A}^{ij}(x) = 0$.

- If $i \leq p, j > p$, then $\mathcal{A}^{ij}(x) = \dfrac{x_i}{x_i - x_j}$.

- If $i > p, j \leq p$, then $\mathcal{A}^{ij}(x) = \dfrac{x_j}{x_j - x_i}$.

Combining that with Equations (14) and (15), and letting $x = \lambda(X)$, particularizes (7) into (13), and completes the proof. ■

At the very beginning of the paper, we stated that $\mathrm{P}_{S^n_+}$ is differentiable almost everywhere. This property can be checked using Theorem 2.7 and the fact that the set of singular matrices has a (Lebesgue) measure zero, see Corollary 1.6.

2.4 Study of $\mathcal{B}$

In this subsection we present several technical properties of $\mathcal{B}(x)$ and define two subsets of M^n that will be useful in the sequel. We start with the following observation.

Lemma 2.8 *Let $x \in \mathbb{R}^n_\downarrow$ be with non-zeros entries: p positive and q negative $(p + q = n)$. Then the entries of the matrix $\mathcal{B}(x)$ are in the closed interval $[0, 1]$ and form non increasing sequences in each row (considered from left to right) and each column (from top to bottom).*

10

Proof. Write $x = (\lambda_1, \ldots, \lambda_p, \mu_{p+1}, \ldots, \mu_n)$ with

$$\lambda_1 > \lambda_2 > \cdots > \lambda_p > 0 > \mu_{p+1} > \mu_{p+2} > \cdots > \mu_n.$$

Using the expressions for the entries of $\mathcal{B}(x)$ in Definition 2.6, one can see that the only case to study is $i \leq p$ and $j > p$. In that case we have

$$\mathcal{B}^{ij}(x) = \frac{\lambda_i}{\lambda_i - \mu_j} = \frac{\lambda_i/\mu_j}{\lambda_i/\mu_j - 1}.$$

Notice that the ratio λ_i/μ_j is negative and

$$\frac{\lambda_i}{\mu_l} \leq \frac{\lambda_j}{\mu_l} \qquad \text{whenever } i < j \leq p, \ l \in \{p+1, \ldots, n\}$$

$$\frac{\lambda_l}{\mu_i} \leq \frac{\lambda_l}{\mu_j} \qquad \text{whenever } p < i < j, \ l \in \{1, \ldots, p\}.$$

Since the function $y \mapsto \frac{y}{y-1}$ is decreasing on the interval $(-\infty, 0)$, with values in $(0, 1)$, the proof is complete. ∎

This property leads us to define the following sets:

- By $\mathcal{D}_{\{01\}}(m)$ we denote the finite set of $m \times m$ symmetric matrices with entries from the set $\{0, 1\}$ such that the entries in each row (considered from left to right) or column (from top to bottom) form a non increasing sequence.

- By $\mathcal{D}_{[01]}(m)$ we denote the convex set of $m \times m$ symmetric matrices with entries from the interval $[0, 1]$, with the same property of non increasing rows and columns.

For example,

$$\mathcal{D}_{\{01\}}(2) = \left\{ \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \right\}, \text{ and}$$

(16)

$$\mathcal{D}_{[01]}(2) = \left\{ \begin{pmatrix} a & b \\ b & c \end{pmatrix} \mid a, b, c \in [0, 1] \text{ and } a \geq b \geq c \right\}.$$

The next corollary is immediate.

11

Corollary 2.9 *For any $x \in \mathbb{R}^n_\downarrow$, the matrix $\mathcal{B}(x)$ is in $\mathcal{D}_{[01]}(n)$. Moreover, any $m \times m$ submatrix of a matrix in $\mathcal{D}_{[01]}(n)$ is an element of $\mathcal{D}_{[01]}(m)$, $m \leq n$.*

Trivially, $\mathcal{D}_{\{01\}}(m) \subset \mathcal{D}_{[01]}(m)$, and the following result gives the precise relationship between $\mathcal{D}_{\{01\}}(m)$ and $\mathcal{D}_{[01]}(m)$.

Lemma 2.10 *For any $m \in \mathbb{N}$, we have*

$$\mathcal{D}_{[01]}(m) = \mathrm{conv}\left(\mathcal{D}_{\{01\}}(m)\right).$$

The elements of $\mathcal{D}_{\{01\}}(m)$ are the extreme points of $\mathcal{D}_{[01]}(m)$.

Proof. Let $\{E_{ij} \,|\, 1 \leq i \leq j \leq m\}$ be the standard basis for the set of $m \times m$ symmetric matrices. That is, if $i \neq j$, then E_{ij} has one in positions (i, j) and (j, i) and zero everywhere else; and if $i = j$, then E_{ii} has one in position (i, i) and zero everywhere else.

Take any matrix $Q \in \mathcal{D}_{[01]}(m)$. We will express it as a convex combination of matrices from $\mathcal{D}_{\{01\}}(m)$. If Q is the all-zero matrix, we are done. Otherwise, define the matrix $B \in \mathcal{D}_{\{01\}}(m)$ as follows

$$B^{ij} = \begin{cases} 1, & \text{if } Q^{ij} \neq 0 \\ 0, & \text{otherwise.} \end{cases}$$

Next, we find basic matrices $E_{i_1 j_1}, \ldots, E_{i_s j_s}$ such that

- $B = \sum_{k=1}^{s} E_{i_k j_k}$;

- For every $l \in \{1, \ldots, s\}$, the sum of the first l of these basic matrices is in $\mathcal{D}_{\{01\}}(m)$ (this implies that $(i_1, j_1) = (1, 1)$);

- $Q^{i_r j_r} \leq Q^{i_{r-1} j_{r-1}}$, for all $r \in \{2, \ldots, s\}$.

Now define

$$B_l^0 = \sum_{k=1}^{l} E_{i_k j_k}, \qquad \text{for all } l = 1, 2, .., s.$$

We will express matrix Q as a convex combination of the matrices $\{B_l^0\}_{l=1}^s$ in s steps. Three points about the notation should be immediately clear: $B_s^0 = B$, $Q^{i_r j_r} > 0$, and $Q \circ E_{i_r j_r} = Q^{i_r j_r} E_{i_r j_r}$ for all $r \in \{1, 2, \ldots, s\}$.

12

In the first step, we let $\alpha_1 = Q^{i_1 j_1} \in [0, 1]$ and define

$$B_l^1 = \alpha_1 B_l^0, \qquad \text{for all } l = 1, \ldots, s.$$

The fact that the zero matrix is in $\mathcal{D}_{\{01\}}(m)$ implies that these matrices are in conv $(\mathcal{D}_{\{01\}}(m))$. Notice that we have

$$B_l^1 \circ E_{i_1 j_1} = Q \circ E_{i_1 j_1} \quad \text{for } l = 1, 2, \ldots, s$$
$$B_l^1 \circ E_{i_q j_q} = Q^{i_1 j_1} E_{i_q j_q} \quad \text{for } 1 \leq q \leq l \leq s,$$
$$B_l^1 \circ E_{i_q j_q} = 0 \qquad \text{for } 1 \leq l < q \leq s$$

If $s = 1$, we are done, thus suppose that $s \geq 2$.

Suppose that on the $(r-1)$-st step we defined the matrices

$$\{B_l^{r-1} : l = r - 1, r, \ldots, s\}$$

with the following four properties:

(i) $B_l^{r-1} \in$ conv $(\mathcal{D}_{\{01\}}(m))$ for all $l = r - 1, r, \ldots, s$;

(ii) $B_l^{r-1} \circ E_{i_q j_q} = Q \circ E_{i_q j_q}$ for all l and q such that $q \leq r - 1 \leq l \leq s$;

(iii) $B_l^{r-1} \circ E_{i_q j_q} = Q^{i_{r-1} j_{r-1}} E_{i_q j_q}$ for all l and q such that $r - 1 \leq q \leq l \leq s$;

(iv) $B_l^{r-1} \circ E_{i_q j_q} = 0$ for all l and q such that $r - 1 \leq l < q \leq s$.

On the r-th step, since $0 < Q^{i_r j_r} \leq Q^{i_{r-1} j_{r-1}}$, set $\alpha_r = Q^{i_r j_r}/Q^{i_{r-1} j_{r-1}} \in [0, 1]$ and form the convex combinations:

$$B_l^r = \alpha_r B_l^{r-1} + (1 - \alpha_r) B_{r-1}^{r-1}, \text{ for all } l = r, r + 1, \ldots, s.$$

We now check that $\{B_l^r : l = r, r + 1, \ldots, s\}$ satisfy the four properties in the indiction hypothesis.

(i) The fact that $B_l^r \in$ conv $(\mathcal{D}_{\{01\}}(m))$ for all $l = r, r + 1, \ldots, s$ is clear from the definition.

(ii) For fixed indexes l and q such that $q < r \leq l \leq s$ (the case $q = r$ is covered in Property (iii) below), we have:

$$\begin{aligned} B_l^r \circ E_{i_q j_q} &= \left(\alpha_r B_l^{r-1} + (1 - \alpha_r) B_{r-1}^{r-1} \right) \circ E_{i_q j_q} && \text{[by the definition of } B_l^r] \\ &= \alpha_r Q \circ E_{i_q j_q} + (1 - \alpha_r) Q \circ E_{i_q j_q} && \text{[by Property (ii)]} \\ &= Q \circ E_{i_q j_q}. \end{aligned}$$

13

(iii) For fixed indexes l and q such that $r \leq q \leq l \leq s$, we have:

$$
\begin{aligned}
B_l^r \circ E_{i_q j_q} &= \left(\alpha_r B_l^{r-1} + (1 - \alpha_r) B_{r-1}^{r-1}\right) \circ E_{i_q j_q} && \text{[by the definition of } B_l^r] \\
&= \alpha_r Q^{i_{r-1} j_{r-1}} E_{i_q j_q} && \text{[by Property (iii)]} \\
&\quad + (1 - \alpha_r) B_{r-1}^{r-1} \circ E_{i_q j_q} \\
&= \alpha_r Q^{i_{r-1} j_{r-1}} E_{i_q j_q} && \text{[by Property (iv)]} \\
&= Q^{i_r j_r} E_{i_q j_q}. && \text{[by the definition of } \alpha_r]
\end{aligned}
$$

(iv) For fixed indexes l and q such that $r \leq l < q \leq s$, by Property (iv), we have:

$$
B_l^r \circ E_{i_q j_q} = \left(\alpha_r B_l^{r-1} + (1 - \alpha_r) B_{r-1}^{r-1}\right) \circ E_{i_q j_q} = 0.
$$

With this the induction is complete. After s inductive steps we define the matrix $B_s^s \in \operatorname{conv}\left(\mathcal{D}_{\{01\}}(m)\right)$ with the property that

$$
B_s^s \circ E_{i_q j_q} = Q \circ E_{i_q j_q}, \quad \text{for all } 1 \leq q \leq s.
$$

Summing over all q's we get that $B_s^s = B_s^s \circ B_s^0 = Q \circ B_s^0 = Q$.

The second claim in the lemma is clear. ∎

Note 2.11 Denoting by $\mathcal{I}(n)$ the set of all $n \times n$ symmetric matrices with entries from the interval $[0, 1]$, it is easy to see, using Example (16), that, in general, the following inclusion is strict

$$
\operatorname{conv}\left(P^n \cdot \mathcal{D}_{\{01\}}(n)\right) \subsetneq \mathcal{I}(n). \qquad ∎
$$

Lemma 2.12 a) *For any $A \in \mathcal{D}_{\{01\}}(m)$, there is a sequence of vectors $\{x^k\}_{k=1}^{\infty}$ in $\mathbb{R}_{\downarrow}^m$, converging to 0, with each x^k having distinct, non-zero coordinates, and such that*

$$
\lim_{k \to \infty} \mathcal{B}(x^k) = A.
$$

b) *If $\{x^k\}_{k=1}^{\infty}$ is any sequence of vectors in $\mathbb{R}_{\downarrow}^m$ with non-zero entries, for which the sequence $\{\mathcal{B}(x^k)\}_{k=1}^{\infty}$ converges, then the number of positive (and therefore negative) coordinates of x^k is constant, for all large enough k.*

Proof. **a)** Suppose that matrix A has r_1 ones in the first row, r_2 in the second and so on. It is clear that $n \geq r_1 \geq r_2 \geq \cdots \geq r_n \geq 0$ and that

14

vector $(r_1, r_2, \ldots, r_n)$ uniquely determines A. Define $p = \max\{t \mid r_t \geq t\}$. Informally, this is the size of the largest principle submatrix, with every entry equal to one, that we can find in the upper-left corner of A. For each $j \in \{p + 1, \ldots, n\}$ define the sequence

$$x_j^k = -\frac{1}{k^{n+p-j+1}}, \qquad k = 1, 2, \ldots,$$

and for each $i \in \{1, \ldots, p\}$, define the sequence

$$x_i^k = \frac{1}{k^{n+p-r_i+0.5}}, \qquad k = 1, 2, \ldots.$$

Then, for every k the symmetric matrix $\mathcal{B}(x^k)$ is given by:

- If $i \leq p$, $j \leq p$, then $\mathcal{B}^{ij}(x^k) = 1$;

- If $i > p$, $j > p$, then $\mathcal{B}^{ij}(x^k) = 0$;

- If $i \leq p$, $j > p$, then $\mathcal{B}^{ij}(x^k) = \dfrac{1}{1 + k^{j-r_i-0.5}}$.

It is easy to see now, that $\mathcal{B}(x^k)$ converges to A.

b) Let p_k be the number of positive entries in the vector x^k. Assume the sequence $\{p_k\}$ does not become constant from some point on. Since p_k is an integer between 1 and n, there are two integers s and t ($1 \leq s < t \leq n$) that appear infinitely many times in the sequence. Without loss of generality, we may assume that the constant subsequence with elements equal to s (resp. t) is p_{2k} (resp. p_{2k+1}). The fact that

$$\mathcal{B}^{tt}(x^{2k}) = 0 \qquad \text{and} \qquad \mathcal{B}^{tt}(x^{2k+1}) = 1,$$

contradicts the convergence of $\{\mathcal{B}(x^k)\}$. ∎

3   Clarke generalized Jacobian

In this section, we compute the Clarke generalized Jacobian [1, Section 2.6] of the projection onto S_+^n, defined by

$$(17) \quad \partial_C \mathrm{P}_{S_+^n}(X) = \mathrm{conv}\Big\{ \lim_{k \to \infty} \nabla \mathrm{P}_{S_+^n}(X_k) \mid X_k \to X, \nabla \mathrm{P}_{S_+^n}(X_k) \text{ exists}, X_k \notin N \Big\},$$

where $N \subset S^n$ has Lebesgue measure zero. The fact, that the definition of the Clarke generalized Jacobian does not depend on the choice of the null set N, is part of [13, Theorem 4].

By Corollary 1.6, the set of matrices in S^n with multiple eigenvalues has (Lebesgue) measure zero. Thus, we may assume without loss of generality, that the set N in (17) contains the set of all symmetric matrices with repeated eigenvalues. Also, by Theorem 2.7, for the gradient $\nabla \mathrm{P}_{S^n_+}(X_k)$ to exist, the matrix X_k must be non-singular. These two observations are behind point (ii) in the definition of $\mathcal{C}(X)$ below.

Remark 3.1 (X non-singular) Notice that the generalized Jacobian is already known at non-singular matrices. Let X be a non-singular matrix in S^n. Theorem 2.7 states that $\delta \circ \lambda$ and $\mathrm{P}_{S^n_+}$ are C^∞ at X and therefore

$$(18) \qquad \partial_C \mathrm{P}_{S^n_+}(X) = \{\nabla \mathrm{P}_{S^n_+}(X)\} = \{O(n)^X \cdot \left(\mathrm{Diag}^{(12)} \mathcal{B}(\lambda(X)))\right)\},$$

where the second equality follows from Equation (13). In Theorem 3.7, we generalize this formula for an arbitrary matrix X. ∎

3.1 Simpler representation of the generalized Jacobian

In this subsection we will derive more convenient representation of the right-hand side of Equation (17). For any X in S^n, we denote by $\mathcal{C}(X)$ the set of sequences in S^n with the following three properties:

(i) $\{X_k\}_{k=0}^\infty$ converges to X;

(ii) X_k is non-singular and has distinct eigenvalues for every k;

(iii) The limit $\lim_{k \to \infty} \mathcal{B}(\lambda(X_k))$ exists.

We introduce the set of all possible such limits

$$(19) \qquad \mathcal{L}(X) = \left\{ \lim_{k \to \infty} \mathcal{B}(\lambda(X_k)) \,|\, \{X_k\}_{k=0}^\infty \in \mathcal{C}(X) \right\}.$$

Lemma 3.2 *Let $X \in S^n$. Then*

$$(20) \qquad \partial_C \mathrm{P}_{S^n_+}(X) = \mathrm{conv} \left\{ O(n)^X \cdot (\mathrm{Diag}^{(12)} \mathcal{L}(X)) \right\}.$$

16

Proof. Denote the right-hand side of Equation (17) by $\mathcal{E}$. The inclusion $\supseteq$ is easy. Indeed, take $U \in O(n)^X$, $B \in \mathcal{L}(X)$, and let X_k be a sequence in $\mathcal{C}(X)$, such that $\mathcal{B}(X_k)$ tends to B. Then, for every $k \in \mathbb{N}$ set

$$Y_k = U(\mathrm{Diag}\, \lambda(X_k))U^\top,$$

and note, using (13), that

$$\nabla \mathrm{P}_{S_+^n}(Y_k) = U(\mathrm{Diag}^{(12)}\mathcal{B}(\lambda(X_k)))U^\top \longrightarrow U(\mathrm{Diag}^{(12)} B)U^\top.$$

Hence $U(\mathrm{Diag}^{(12)} B)U^\top \in \mathcal{E}$ and the inclusion is proved.

To prove $\subseteq$, take $M \in \mathcal{E}$ and a sequence $\{X_k\}$ of non-singular matrices with distinct eigenvalues converging to X, such that

$$(21) \qquad \lim_{k \to \infty} \nabla \mathrm{P}_{S_+^n}(X_k) = M.$$

Let $X_k = U_k(\mathrm{Diag}\, \lambda(X_k))U_k^\top$ be the ordered spectral decompositions of the X_k. By Theorem 2.7, Limit (21) becomes

$$(22) \qquad \lim_{k \to \infty} U_k(\mathrm{Diag}^{(12)}\mathcal{B}(\lambda(X_k)))U_k^\top = M.$$

Now, we extract subsequences from X_k:

- Since the orthogonal group is compact, there is a subsequence, (denote it again by $\{X_k\}$) for which $\{U_k\}$ converges to some $U \in O(n)$. Moreover, by continuity of the eigenvalues on S^n (see e.g. [14, Section 2.3.1]), we have that $X = U(\mathrm{Diag}\, \lambda(X))U^\top$.

- By Lemma 2.8, the sequence $\{\mathcal{B}(\lambda(X_k))\}$ is bounded and therefore we may take a subsequence (denote it again by $\{X_k\}$), such that $\mathcal{B}(\lambda(X_k))$ converges to some $B \in \mathcal{D}_{[01]}(n)$.

Passing to the limit in (22) we obtain $M = U(\mathrm{Diag}^{(12)} B)U^\top$, and the second inclusion is proved. ∎

3.2 Properties of the set $\mathcal{L}(X)$

Let X be in S^n and recall Identity (1): $p_X + q_X = n - n_X$. Let also

$$(23) \qquad \lambda(X) = (\lambda_1(X), \ldots, \lambda_{p_X}(X), 0, \ldots, 0, \mu_{p_X+n_X+1}(X), \ldots, \mu_n(X)),$$

17

where there are n_X zeros in between. We define two subsets of S^n, denoted by $\mathcal{D}_{\{01\}}(X)$, $\mathcal{D}_{[01]}(X)$ respectively, of all matrices of the form:

(24)
$$
\begin{pmatrix}
1_{p_X \times p_X} & 1_{p_X \times n_X} & \Lambda \\
1_{n_X \times p_X} & \mathcal{S}(n_X) & 0_{n_X \times q_X} \\
\Lambda^\top & 0_{q_X \times n_X} & 0_{q_X \times q_X}
\end{pmatrix}.
$$

where $\mathcal{S}(n_X)$ is equal to $\mathcal{D}_{\{01\}}(n_X)$, $\mathcal{D}_{[01]}(n_X)$ respectively, and the $p_X \times q_X$ matrix Λ has entries

$$
\Lambda^{ij} = \frac{\lambda_i}{\lambda_i - \mu_{p_X + n_X + j}}, \qquad \text{for all } \ 1 \le i \le p_X, \ 1 \le j \le q_X.
$$

Example 3.3 • If X is the zero matrix in M^n, we have

$$
\mathcal{D}_{\{01\}}(X) = \mathcal{D}_{\{01\}}(n) \qquad \text{and} \qquad \mathcal{D}_{[01]}(X) = \mathcal{D}_{[01]}(n).
$$

• If X is non-singular, observe from Definition 2.6 that we have

$$
\mathcal{D}_{\{01\}}(X) = \mathcal{D}_{[01]}(X) = \{\mathcal{B}(\lambda(X))\}. \qquad \blacksquare
$$

The goal in this subsection is to prove the inclusions

$$
\mathcal{D}_{\{01\}}(X) \subseteq \mathcal{L}(X) \subseteq \mathcal{D}_{[01]}(X),
$$

where the set $\mathcal{L}(X)$ is defined by Equation (19). The next two propositions show the first and the second inclusion, respectively.

Proposition 3.4 *Fix a matrix X in S^n. For any $A \in \mathcal{D}_{\{01\}}(X)$, there exists a sequence $\{X_k\}$ in $\mathcal{C}(X)$, such that $\lim\limits_{k \to \infty} \mathcal{B}(\lambda(X_k)) = A$.*

Proof. Let $X = U\big(\mathrm{Diag}\,\lambda(X)\big)U^\top$ be an ordered spectral decomposition of X, where vector $\lambda(X)$ is described by (23). Since $A \in \mathcal{D}_{\{01\}}(X)$, A has a block structure as described by (24). Let $B \in \mathcal{D}_{\{01\}}(n_X)$ be the central

18

block in this representation. Applying Lemma 2.12 with $m = n_X$ and B gives a sequence $\{x^k\} \subset \mathbb{R}^{n_X}_\downarrow$ converging to $0 \in \mathbb{R}^{n_X}$ with each vector x^k having distinct, non-zeros entries, such that $\lim_{k \to \infty} \mathcal{B}(x^k) = B$ and such that the number of positive entries of x^k is constant for every k. Define

$$d_k = \left(\lambda_1(X) + \frac{p_X}{k}, \ldots, \lambda_{p_X}(X) + \frac{1}{k}, x^k, \mu_{p_X + n_X + 1}(X) + \frac{q_X}{k}, \ldots, \mu_n(X) + \frac{1}{k}\right),$$

and $X_k = U(\mathrm{Diag}\, d_k)U^\top$. Then, for k large enough, X_k has distinct eigenvalues and the sequence $\{X_k\}$ converges to X. Furthermore, for all large enough k we have $\lambda(X_k) = d_k \in \mathbb{R}^n_\downarrow$. Finally, using Definition 2.6, it is not difficult to see that $\lim_{k \to \infty} \mathcal{B}(\lambda(X_k)) = A$. Clearly, $\{X_k\}$ is in $\mathcal{C}(X)$. ∎

Proposition 3.5 *Let $X \in S^n$ and let $\{X_k\} \in \mathcal{C}(X)$ be such that $\{\mathcal{B}(\lambda(X_k))\}$ converges to B, then B lies in $\mathcal{D}_{[01]}(X)$.*

Proof. By the second part of Lemma 2.12, we may suppose without loss of generality that $p_{X_k} = p$. By continuity of eigenvalues on S^n (see e.g. [14, 2.3.1]) and the fact that $\{X_k\}$ converges to X we get $p_X \le p \le p_X + n_X$. Let

$$\lambda(X_k) = (\lambda_1^k, \ldots, \lambda_p^k, \mu_{p+1}^k, \ldots, \mu_n^k), \text{ and}$$
$$\lambda(X) = (\lambda_1, \ldots, \lambda_{p_X}, 0, ..., 0, \mu_{p_X + n_X + 1}, \ldots, \mu_n^k)$$

We compute $B = \lim_{k \to \infty} \mathcal{B}(\lambda(X_k))$ by considering several cases.
- If $i \le p_X$ and $j \le p$, then $\mathcal{B}^{ij}(\lambda(X_k)) = 1$ for all k implies that $B^{ij} = 1$;
- If $i \le p_X$ and $p < j \le p_X + n_X$, then $\mathcal{B}^{ij}(\lambda(X_k)) = \frac{\lambda_i^k}{\lambda_i^k - \mu_j^k}$ implies $B^{ij} = 1$;
- If $i \le p_X$ and $p_X + n_X < j$, then $\mathcal{B}^{ij}(\lambda(X_k)) = \frac{\lambda_i^k}{\lambda_i^k - \mu_j^k}$ implies $B^{ij} = \frac{\lambda_i}{\lambda_i - \mu_j}$;
- The principle submatrix of $\mathcal{B}(\lambda(X_k))$ formed by the entries with indexes $i, j \in \{p_X + 1, ..., p_X + n_X\}$ is, by Lemma 2.8, in the set $\mathcal{D}_{[01]}(n_X)$. Since $\mathcal{D}_{[01]}(n_X)$ is closed, the same submatrix of B will also be in $\mathcal{D}_{[01]}(n_X)$.
- If $p_X < i \le p$ and $p_X + n_X < j$, then $\mathcal{B}^{ij}(\lambda(X_k)) = \frac{\lambda_i^k}{\lambda_i^k - \mu_j^k}$ implies $B^{ij} = 0$;
- If $p < i$ and $p_X + n_X < j$, then $\mathcal{B}^{ij}(\lambda(X_k)) = 0$, and then $B^{ij} = 0$.

Since $\mathcal{B}$ is a symmetric matrix, we have its complete description, which shows that $\mathcal{B} \in \mathcal{D}_{[01]}(X)$. ∎

We summarize the results in this subsection in the following proposition.

Proposition 3.6 *For any matrix $X \in S^n$ we have the inclusions*

$$\mathcal{D}_{\{01\}}(X) \subseteq \mathcal{L}(X) \subseteq \mathcal{D}_{[01]}(X).$$

If, in particular, X is non-singular, the inclusions hold with equalities

$$\mathcal{D}_{\{01\}}(X) = \mathcal{L}(X) = \mathcal{D}_{[01]}(X) = \{\mathcal{B}(\lambda(X))\}.$$

3.3   A formula for the Clarke generalized Jacobian

We are now in a position to compute the Clarke generalized Jacobian of $\mathrm{P}_{S_+^n}$ at an arbitrary symmetric matrix X.

Theorem 3.7 (Clarke generalized Jacobian) *Let X be any $n \times n$ symmetric matrix. The Clarke generalized Jacobian of $\mathrm{P}_{S_+^n}$, the projection map onto the cone of positive semi-definite matrices, at X is*

(25) $$\partial_C \mathrm{P}_{S_+^n}(X) = \mathrm{conv}\left(O(n)^X \cdot (\mathrm{Diag}^{(12)} \mathcal{D}_{\{01\}}(X))\right),$$

where $\mathcal{D}_{\{01\}}(X)$ is defined by (24).

Proof.   By Proposition 3.6, we have the inclusions

$$\mathcal{D}_{\{01\}}(X) \subseteq \mathcal{L}(X) \subseteq \mathcal{D}_{[01]}(X).$$

Applying the $\mathrm{Diag}^{(12)}$ operator and conjugating, we get

$$O(n)^X \cdot (\mathrm{Diag}^{(12)} \mathcal{D}_{\{01\}}(X)) \subseteq O(n)^X \cdot (\mathrm{Diag}^{(12)} \mathcal{L}(X))$$
$$\subseteq O(n)^X \cdot (\mathrm{Diag}^{(12)} \mathcal{D}_{[01]}(X)).$$

Taking the convex hull throughout, we get

$$\mathrm{conv}\left(O(n)^X \cdot \left(\mathrm{Diag}^{(12)} \mathcal{D}_{\{01\}}(X)\right)\right)$$

$$\begin{aligned}
&\subseteq \partial_C \mathrm{P}_{S_+^n}(X) && \text{[by Lemma 3.2]} \\
&\subseteq \mathrm{conv}\left(O(n)^X \cdot \left(\mathrm{Diag}^{(12)} \mathcal{D}_{[01]}(X)\right)\right) && \\
&= \mathrm{conv}\left(O(n)^X \cdot \left(\mathrm{Diag}^{(12)} \mathrm{conv}\, \mathcal{D}_{\{01\}}(X)\right)\right) && \text{[by Lemma 2.10]} \\
&= \mathrm{conv}\left(O(n)^X \cdot \left(\mathrm{conv}\, \mathrm{Diag}^{(12)} \mathcal{D}_{\{01\}}(X)\right)\right) && \text{[since $\mathrm{Diag}^{(12)}$ is linear]} \\
&\subseteq \mathrm{conv}\left(O(n)^X \cdot \left(\mathrm{Diag}^{(12)} \mathcal{D}_{\{01\}}(X)\right)\right). && \text{[by Lemma 1.4(ii)]}
\end{aligned}$$

Thus, we must have equalities throughout and the result follows. ■

Example 3.8 ($X = 0$) In particular, when $X = 0$, we obtain the formula

$$\partial_C \mathrm{P}_{S^n_+}(0) = \mathrm{conv}\left(O(n) \cdot (\mathrm{Diag}^{(12)} \mathcal{D}_{\{01\}}(n))\right). \qquad \blacksquare$$

Example 3.9 (X non-singular) When X is a non-singular matrix in S^n Formula (25) reduces to Formula (18) via Proposition 3.6. $\qquad \blacksquare$

Remark 3.10 (What about $\mathrm{P}_{S^n_-}$?) We can adapt this study to obtain similar results about $\mathrm{P}_{S^n_-}$ – the projection onto the cone of negative definite matrices. This projection is also differentiable at any non-singular matrix X and its gradient at X is expressed via

$$\nabla \mathrm{P}_{S^n_+}(X) = U\left(\mathrm{Diag}^{(12)} \mathcal{B}'(\lambda(X))\right) U^\top,$$

with $U \in O(n)^X$ and $\mathcal{B}'(x) \in S^n$ defined for any $x \in \mathbb{R}^n_\downarrow$ with non-zeros entries (p positive, q negative, and $p + q = n$) by

- If $i \leq p$, $j \leq p$, then $\mathcal{B}'^{ij}(x) = 0$;

- If $i > p$, $j > p$, then $\mathcal{B}'^{ij}(x) = 1$;

- If $i \leq p$, $j > p$, then $\mathcal{B}'^{ij}(x) = \dfrac{x_j}{x_j - x_i}$;

- If $i > p$, $j \leq p$, then $\mathcal{B}'^{ij}(x) = \dfrac{x_i}{x_i - x_j}$.

Further results for $\mathrm{P}_{S^n_-}$ follow, in an analogous way as those for $\mathrm{P}_{S^n_+}$. Notice that we also can retrieve the differentiability properties of $\mathrm{P}_{S^n_-}$ from those of $\mathrm{P}_{S^n_+}$ via Theorem 2.2. $\qquad \blacksquare$

4 Interchanging the conjugation and the convex hull

Looking at Formula (25), one is tempted to exchange the operations of taking the convex hull and taking the orbits under the action of $O(n)^X$. In this section we address this issue. We show that if the multiplicity of the zero eigenvalue is zero or one we can exchange conv with $O(n)^X$, but if the multiplicity is at least two we, in general, cannot exchange these operations.

4.1   Block diagonal and block constant matrices

Let X be a symmetric matrix with eigenvalues

$$\lambda_1(X) = \cdots = \lambda_{k_1}(X) > \lambda_{k_1+1}(X) = \cdots = \lambda_{k_2}(X) > \lambda_{k_2+1}(X) \cdots \lambda_{k_r}(X),$$

where $k_r = n$. The indices of the equal eigenvalues partition the set $\{1, ..., n\}$ in to r blocks $I_1 = \{1, 2, ..., k_1\}, ..., I_r = \{k_{r-1} + 1, ..., k_r\}$.

A matrix M is said to be *block constant* if $M^{i_1 i_2} = M^{j_1 j_2}$ whenever i_1, j_1 are in the same block and i_2, j_2 are in the same, possibly different, block. We say that a matrix is *block diagonal* when it has r square blocks on the diagonal with sizes equal to the sizes of the sets $I_1, ..., I_r$, and zeros elsewhere. We denote by $O(X, n)$ the group of all block diagonal orthogonal matrices. We need the following two lemmas.

Lemma 4.1 *For any block constant matrix M and any block diagonal orthogonal matrix U we have*

$$U\left(\mathrm{Diag}^{(12)} M\right)U^T = \mathrm{Diag}^{(12)} M.$$

Proof.   This lemma is a particular case of Lemma 5.3 in [12]. It follows easily from the definitions.   ∎

Lemma 4.2 *Let U be a arbitrary matrix of $O(n)^X$. We have the representation $O(n)^X = U\, O(X, n)$.*

Proof.   Observe that if $V \in O(n)^X$, then $U^\top V$ is a block diagonal, orthogonal matrix. Thus, we have $U^\top O(n)^X \subset O(X, n)$ or, equivalently, $O(n)^X \subset U\, O(X, n)$. The converse inclusion follows from definitions.   ∎

Thus, with $U \in O(n)^X$, we can write

$$\partial_C \mathrm{P}_{S_+^n}(X) = \mathrm{conv}\left(O(n)^X \cdot (\mathrm{Diag}^{(12)} \mathcal{D}_{\{01\}}(X))\right)$$
$$= U \cdot \mathrm{conv}\left(O(X, n) \cdot (\mathrm{Diag}^{(12)} \mathcal{D}_{\{01\}}(X))\right).$$

We see that conv and $O(n)^X$ can be interchanged, if and only if conv and $O(X, n)$ can be interchanged.

22

4.2 Different multiplicities of the zero eigenvalue

Recall that p_X, q_X are the number of positive and negative eigenvalues respectively and that n_X is the multiplicity of zero eigenvalue. Notice that

- if $n_X = 0$, the set $\mathcal{D}_{\{01\}}(X)$ has one element (namely $\mathcal{B}(\lambda(X))$),

- if $n_X = 1$, the set $\mathcal{D}_{\{01\}}(X)$ has two elements.

Observe also that, in these two cases, every matrix of $\mathcal{D}_{\{01\}}(X)$ is block constant. Using Lemma 4.1 the following result then becomes clear.

Theorem 4.3 *Let X be any $n \times n$ symmetric matrix having eigenvalue zero with multiplicity zero or one. The Clarke generalized Jacobian of $\mathrm{P}_{S_+^n}$, the projection map onto the cone of positive semi-definite matrices, at X is*

$$(26) \qquad \partial_C \mathrm{P}_{S_+^n}(X) = U\big(\mathrm{Diag}^{(12)} \mathcal{D}_{[01]}(X)\big) U^T,$$

where $\mathcal{D}_{[01]}(X)$ is defined by (24) and $U \in O(n)^X$.

When $n_X \geq 2$, in general, we cannot interchange the operations conv and $O(n)^X$ in Formula (25). We do have that

$$O(n)^X \cdot \big(\mathrm{Diag}^{(12)} \mathcal{D}_{[01]}(X)\big) = O(n)^X \cdot \mathrm{conv}\big(\mathrm{Diag}^{(12)} \mathcal{D}_{\{01\}}(X)\big)$$
$$\subsetneq \mathrm{conv}\big(O(n)^X \cdot \mathrm{Diag}^{(12)} \mathcal{D}_{\{01\}}(X)\big),$$

where the first equality follows from Lemma 2.10 and the inclusion from Lemma 1.4(i). We show that the inclusion is strict in general, by showing that the set $O(n)^X \cdot (\mathrm{Diag}^{(12)} \mathcal{D}_{[01]}(X))$ is not convex when X is the 2×2 zero matrix.

Proposition 4.4 *If X is the 2×2 zero matrix then the set*

$$(27) \qquad O(n)^X \cdot (\mathrm{Diag}^{(12)} \mathcal{D}_{[01]}(X))$$

is not convex.

Proof. Define

$$T(a, b, c) = \begin{pmatrix} a & b \\ b & c \end{pmatrix},$$

23

as well as

$$U(\phi) = \begin{pmatrix} \cos(\phi) & -\sin(\phi) \\ \sin(\phi) & \cos(\phi) \end{pmatrix} \quad \text{and} \quad V(\phi) = \begin{pmatrix} \cos(\phi) & \sin(\phi) \\ \sin(\phi) & -\cos(\phi) \end{pmatrix}.$$

Then we have

$$O(2)^X = O(2) = \{U(\phi) \,|\, 0 \le \phi < 2\pi\} \cup \{V(\phi) \,|\, 0 \le \phi < 2\pi\},$$
$$\mathcal{D}_{[01]}(X) = \{T(a,b,c) \,|\, a,b,c \in [0,1] \text{ and } a \ge b \ge c\}.$$

For given parameters ϕ, a, b and c, direct computation shows

$$U(\phi) \cdot (\mathrm{Diag}^{(12)} T(a,b,c)) = V(\phi) \cdot (\mathrm{Diag}^{(12)} T(a,b,c)),$$

implying that it is enough to work only with the matrices $U(\phi)$. Consider two particular entries in this 4-tensor:

$$f(\phi, a, b, c) = \left(U(\phi) \cdot (\mathrm{Diag}^{(12)} T(a,b,c))\right)^{12}_{11},$$
$$g(\phi, a, b, c) = \left(U(\phi) \cdot (\mathrm{Diag}^{(12)} T(a,b,c))\right)^{22}_{11}.$$

Using Equations (4), (5) we compute f and g explicitly:

$$f(\phi, a, b, c) = \cos^3(\phi) \sin(\phi)(a - b) + \cos(\phi) \sin^3(\phi)(b - c),$$
$$g(\phi, a, b, c) = \cos^2(\phi) \sin^2(\phi)(a - b) - \cos^2(\phi) \sin^2(\phi)(b - c).$$

Consider the following two points

$$M_1 = (f(5\pi/6, 1, 0, 0), g(5\pi/6, 1, 0, 0)) = (-3\sqrt{3}/16, 3/16)$$
$$M_2 = (f(4\pi/6, 1, 1, 0), g(4\pi/6, 1, 1, 0)) = (-3\sqrt{3}/16, -3/16).$$

To show that (27) is not convex, it is sufficient to show that the middle, $(-3\sqrt{3}/16, 0)$, of the segment $[M_1, M_2]$ is not in the set

$$\{(f(\phi, a, b, c), g(\phi, a, b, c)) \,|\, 0 \le \phi < 2\pi, \, 0 \le a, b, c \le 1\}.$$

Suppose on the opposite that there are values of ϕ, a, b and c for which $(f(\phi, a, b, c), g(\phi, a, b, c)) = (-3\sqrt{3}/16, 0)$. The fact $f(\phi, a, b, c) \ne 0$ implies that $\cos(\phi) \ne 0 \ne \sin(\phi)$. Then, $g(\phi, a, b, c) = 0$ implies that

(28) $$a - b = b - c,$$

24

and substituting that into $f(\phi, a, b, c) = -\frac{3\sqrt{3}}{16}$ we get

$$\cos(\phi)\sin(\phi)(a - b) = -\frac{3\sqrt{3}}{16}.$$

The last equation is equivalent to

$$\sin(2\phi)(a - b) = -\frac{3\sqrt{3}}{8}.$$

Notice that since $a, b, c \in [0, 1]$, Equation (28) also implies that $|a - b| \leq 1/2$. This gives us the contradiction $|\sin(2\phi)| \geq 3\sqrt{3}/4 > 1$. Thus the middle of $U(5\pi/6) \cdot \mathrm{Diag}^{(12)} T(1, 0, 0)$ and $U(4\pi/6) \cdot \mathrm{Diag}^{(12)} T(1, 1, 0)$ does not lie in $O(n)^X \cdot (\mathrm{Diag}^{(12)} \mathcal{D}_{[01]}(X))$. ■

Acknowledgements

References

[1] F.H. CLARKE. *Optimization and Nonsmooth Analysis*. Wiley, New York, 1983.

[2] C. DAVIS. All convex invariant functions of hermitian matrices. *Archiv der Mathematik*, 8:276–278, 1957.

[3] L.C. EVANS and R.F. GARIEPY. *Measure Theory and Tine Properties of Functions*. CRC Press, Inc., Studies in Advanced Mathematics, 1992.

[4] N. HIGHAM. Computing a nearest symmetric positive semidefinite matrix. *Linear Algebra and its Applications*, 103:103–118, 1988.

[5] J.-B. HIRIART-URRUTY. At what points is the projection mapping differentiable?. *American Mathematical Monthly*, 456–458, 1982.

[6] J.-B. HIRIART-URRUTY and C. LEMARÉCHAL. *Convex Analysis and Minimization Algorithms I and II.* Grundlehren der Mathematischen Wissenschaften. Springer-Verlag, Berlin, 1993.

[7] J.-B. HIRIART-URRUTY and C. LEMARÉCHAL. *Fundamentals of Convex Analysis.* Springer-Verlag, Heidelberg, 2001.

[8] A.S. LEWIS. Derivatives of spectral functions. *Mathematics of Operations Research*, 21:576–588, 1996.

[9] A.S. LEWIS and H.S. SENDOV. Twice differentiable spectral functions. *SIAM Journal of Matrix Analysis and Applications*, 23(2):368–386, 2001.

[10] R.T. ROCKAFELLAR. *Convex Analysis.* Princeton University Press, Princeton, N.J., 1970.

[11] H.S. SENDOV. Generalized Hadamard product and the derivatives of spectral functions. *Research Report, Department of Mathematics and Statistics, University of Guelph, Ontario*, 2004. Mathematical Series 2004-308.

[12] H.S. SENDOV. Interactions Between the Generalized Hadamard Product and the Eigenvalues of Symmetric Matrices. *Research Report, Department of Mathematics and Statistics, University of Guelph, Ontario*, 2004. Mathematical Series 2004-309.

[13] J. WARGA. Fat homeomorphisms and unbounded derivate containers. *Journal of Mathematical Analysis and Applications*, 81:545–560, 1981.

[14] H.WOLKOWICZ, R. SAIGAL and L. VANDERBERGHE. *Handbook of Semidefinite-Programming.* Kluwer, 2000.

[15] N-K. TSING, M.K.H FAN, E.I. VERRIEST. On analycity of functions involving eigenvalues. *Linear Algebra and its Applications*, 207:159-180, 1994.