A Generalization of S-Divergence to Symmetric Cone Programming via Euclidean Jordan Algebra

by

Zhuochen (Jaden) Wang

A thesis submitted to the

College of Engineering and Applied Sciences of the

University of Colorado in partial fulfillment

of the Engineering Honors Program requirements for

Bachelor of Science

Department of Applied Mathematics

2022

Committee Members:	*Affiliations:*
Advisor: Dr. Stephen Becker	Applied Mathematics
Dr. Richard Green	Mathematics
Dr. Jonathan Belcher	Applied Mathematics

Wang, Zhuochen (Jaden) (Applied Mathematics)

A Generalization of S-Divergence to Symmetric Cone Programming via Euclidean Jordan Algebra

Thesis directed by Dr. Stephen Becker

Abstract

Symmetric cone programming encompasses a vast majority of tractable convex optimization problems, including linear programming, semidefinite programming, and second-order cone programming. It turns out that we can generalize many results from semidefinite matrices to symmetric cones under the abstract framework of Euclidean Jordan algebra. In particular, S-divergence was previously proposed as a numerical alternative to Riemannian distance for the Hermitian positive definite cone. The goal of this thesis is to generalize S-divergence to symmetric cones and prove that its nice properties in the matrix case are preserved. Specifically, we wish to show that S-divergence induces a metric on the cone and is geodesically-convex. After an extensive exposition of necessary background, we successfully proved most of our claims, with only one conjecture remaining to be proven.

Dedication

To my parents, who gave me everything so I can pursue my passion in mathematics.

Acknowledgements

I would like to express my deepest gratitude toward my advisor Stephen Becker. Two and half years ago, I cold-walked into his office and he has been generously sharing his expertise in optimization and great life advice with me ever since. The seed for this thesis was planted when Stephen forwarded me a Riemannian optimization paper, making me realize the importance of geometry. Stephen further watered the seed by introducing me to the work of Sra and encouraging me to explore topics that give me joy. The seed finally sprouted when Stephen found Permenter's paper, and soon after I knew that I found my thesis topic. It was his pivotal guidance that brings this thesis to blossom. Besides the research skills I have learned from Stephen, I have gained so much independence. I realize that being able to take ownership of my research progress yet feel 100% supported by Stephen is the perfect combination to make math research exceptionally rewarding. I am very grateful to have this experience as an undergraduate.

I am also deeply indebted to Dr. Green. Frankly, I took algebra as an applied math major because some schools require it for graduate admission. I was blown away by Dr. Green's masterful and engaging lectures; the course quickly became a highlight of my undergraduate experience despite the pandemic. By serendipity, Dr. Green taught the graduate version as well, where he challenged and engaged me to the fullest. The rigorous algebra background, the passion for abstract reasoning, and the appreciation for elegance I gained due to Dr. Green are invaluable, and they are all reflected in this thesis. What made this thesis particularly fun to write is that it deepens my understanding of concepts I learned in algebra. For the same reason, I would also like to thank Prof. Mayr for teaching me half of the background in this thesis!

Despite only knowing him for a few months, I am beyond grateful for Jon Belcher's involvement in this thesis. He helped me catch multiple mistakes and clarify challenging concepts. His encouragement and validation was also exactly what I needed during this stressful time. I hope that this thesis will be useful for Jon if he chooses to pursue this topic further. For similar reasons, I also want to thank Jacob Gaiter and Theodore Gonzales for helping me refine some details in this thesis.

Moreover, I cannot thank Prof. Muddappa Gowda and Prof. Jiyuan Tao enough for their generous expert guidance on some of the resources and techniques I needed for this generalization. I learned many important insights from them that greatly accelerated my progress. Not to mention that this thesis would not have been possible without their (and their students') seminal contributions to EJA and optimization.

Finally, I would like to thank Dr. Frank Permenter for acquainting me with EJA that sparked the central theme of this thesis. I was completely unaware of it until he pointed out to me, and what an exhilarating revelation it was!

Contents

Chapter 1

Introduction

Symmetric cone programming (SCP) is a central topic in the study of optimization and has plethora of real-world applications. It encompasses a vast majority of the most important classes of optimization problems including linear programming (LP), semidefinite programming (SDP), and second-order cone programming (SOCP). In the 90's, a breakthrough in SCP emerged from the discovery of a highly efficient method called the primal-dual interior-point method (IPM), which was mostly refined by Nesterov and Todd [NT97, NT98]. The method exploits self-concordant barrier functions extensively studied by Nesterov and Nemirovskii [NN94] and remains the state-of-art for SCP as of today.

However, although Nesterov and Todd provided a way to understand primal-dual IPM through self-concordant barrier functions, this perspective is largely attributed to the brilliant numerical insights of Nesterov and Todd. It is not easy to motivate self-concordant barrier functions from an abstract framework. Specifically, the definitions of self-concordance and the specific barrier functions used for common SCP problems such as semidefinite or second-order programming appear to be somewhat *ad hoc*. Although self-concordance is sufficient for the convergence analysis of primal-dual IPM, the abstract reason for its effectiveness remained elusive until another paper by Nesterov and Todd [NT02]. It is known that the Hessian of the barrier function induces a Riemannian metric tensor on the symmetric cone, and this metric tensor endows the cone with a Riemannian manifold structure. This is called a *Hessian manifold*. In this paper, Nesterov and Todd showed that the path tracked by the primal-dual IPM, namely the central path, is close to following a geodesic on the cone as a Hessian manifold. However, understanding SCP through Riemannian geometry alone is also unsatisfying. In order to see why, we first review three of the most common cones in SCP:

- Nonnegative Orthant $\mathbb{R}^n_+$. This cone is associated with the conic constraints in linear programming (LP) and is defined as

  $$\mathbb{R}^n_+ := \{(x_1, \ldots, x_n)^T \in \mathbb{R}^n : x_i \geq 0 \ \forall\ 1 \leq i \leq n\}.$$

  The barrier function is $B(x) = -\sum_{i=1}^n \log(x_i)$. This barrier function induces a Riemannian manifold. The geodesic between x and y is

  $$c(t) = x^{1-t} \circ y^t,$$

  where $\circ$ denotes the element-wise (Hadamard) multiplication.

- Positive semidefinite matrix cone $\mathbb{S}^n_+$. This cone is associated with the conic constraints in semidefinite programming (SDP) and is defined exactly as the name suggests. The barrier function is $B(X) = -\log\det X$, where log is the usual logarithmic function on the real numbers. The geodesic between X and Y is

  $$c(t) = X^{\frac{1}{2}} \left(X^{-\frac{1}{2}} Y X^{-\frac{1}{2}} \right)^t X^{\frac{1}{2}}.$$

  We see that for $n = 1$, $\mathbb{S}^1_+$ coincides with $\mathbb{R}_+$ and so do their barrier functions and geodesics.

- Second-order cone $\mathbb{L}^{n+1}$. This cone is associated with the conic constraints in second-order cone programming (SOCP) and is defined as

  $$\mathbb{L}^{n+1} := \{(x_0, x_1, \ldots, x_n)^T \in \mathbb{R}^{n+1} : x_1^2 + \cdots + x_n^2 \leq x_0^2\}.$$

  Denote $\overline{x} := (x_1 \ldots, x_n)^T \in \mathbb{R}^n$ and let $\|\cdot\|$ represent the Euclidean norm. The barrier function is $B(x) = -\log(x_0^2 - \|\overline{x}\|^2)$. The square root of elements in $\mathbb{L}^{n+1}$ is well-defined. So given $x, y \in \mathbb{L}^{n+1}$, let $w := x^{1/2}$, and define

$$M := \begin{pmatrix} \|w\|^2 & 2w_0\overline{w}^T \\ 2w_0\overline{w} & (w_0^2 - \|\overline{w}\|^2)I_n + 2\overline{w}\overline{w}^T \end{pmatrix}.$$

This matrix is in fact invertible so we can define $z := M^{-1}y$. The geodesic between x and y is

$$c(t) = \frac{1}{2}M\left((z_0 + \|\overline{z}\|)^t \begin{pmatrix} 1 \\ \frac{\overline{z}}{\|\overline{z}\|} \end{pmatrix} + (z_0 - \|\overline{z}\|)^t \begin{pmatrix} 1 \\ -\frac{\overline{z}}{\|\overline{z}\|} \end{pmatrix}\right).$$

It is not hard to see how agonizing it would be to prove results on these cones individually. Despite the increasingly unfriendly geodesics, we see that the barrier functions for these three cones look very similar. This raises the question: is there a way to systematically derive all the results (*e.g.*, the geodesics) for these cones using a unifying theory without dealing with the messy details?

It turns out the answer is yes for any cone belonging to the family of *symmetric cones*, and the unifying theory involves Euclidean Jordan algebra (EJA). In particular, EJA completely describes the family of symmetric cones, including the three canonical examples above. This connection was first discovered by Güler [Gül96]. Since then, EJA has experienced a small renaissance in the conic optimization literature. The advantage of studying symmetric cones under the EJA framework is apparent: working with specific instance of symmetric cones is cumbersome and often does not offer deep insight into the structural properties. Adopting an EJA framework allows us to study the structure of all symmetric cones at once. The disadvantage is that despite its increasing popularity, EJA has a higher barrier of entry for many optimizers as it was previously a niche topic in algebra. The existing limited expositions on EJA tend to be fairly theoretical and not written with a non-algebraist audience in mind. Moreover, some tools readily available in linear algebra and operator theory have not been generalized to EJA, so a straightforward result in linear algebra or operator theory might be quite tricky to prove in EJA. This is presumably why

Nesterov and Todd were well-aware of EJA but understandably decided not to adopt this framework for their analysis, as they discussed in the introduction of [NT98] and [NT02]. Some progress on the accessibility of EJA has been made since then, and hopefully this thesis can also contribute to making EJA accessible to a wider audience.

Combining both insights that primal-dual IPM is close to following the geodesics and that EJA can describe the geodesics of all symmetric cones elegantly, Permenter proposed two SCP algorithms that follow the Riemannian geodesics directly, exploiting the Hessian manifold structure [Per20]. The geodesic IPM has comparable convergence results with the primal-dual IPM. In particular, Permenter suggested that the Nesterov–Todd direction from the primal-dual IPM in the best case is a linear approximation of the geodesic update. Thus, it would not be surprising if geodesic IPM has superior performance. However, Permenter did not compare the two methods numerically in the paper. Moreover, the Riemannian distance used in the geodesic IPM can be expensive to compute as discussed in [CSBP11]. In the same paper, a numerical alternative to the Riemannian distance, the S-divergence, was proposed. Sra further investigated the properties of S-divergence in [Sra15] and showed that the metric derived from S-divergence enjoys similar desirable properties as the Riemannian distance, including being geodesically-convex (g-convex). We conjecture that the results in this paper generalize to any symmetric cone. If this conjecture holds, then this metric induced by the S-divergence might be a more efficient metric choice for algorithms that rely on Riemannian distance and further improve their performance.

This thesis is organized as follows: In Chapter 2, we introduce key concepts from conic programming and EJA. All results and examples in this chapter already exist in the literature, so only the exposition is original. In Chapter 3, we prove that S-divergence induces a valid metric in Theorem 3.1.15 and prove that S-divergence is g-convex on any symmetric cone with certain caveats Theorem 3.2.10, providing a partial generalization of major results from Sra's paper. As far as we know, these are novel generalizations. In Chapter 4, we discuss the implications of our results and propose some interesting future directions.

Chapter 2

Background

We shall take the liberty to assume that the reader is familiar with the basics of convexity, ring theory, and vector space theory. Due to the limited time constraint, we cannot cover all background needed so we encourage the reader to look up unfamiliar convexity concepts from Boyd and Vandenberghe [BV04] and algebra concepts from Dummit and Foote [DF91]. The geometric concepts presented in this thesis are meant to be intuitive but loose on the rigor.

2.1 Conic programming

The importance of optimization is self-evident. Who does not want to find the optimal solutions to their quantifiable problems? We especially care about convex optimization because convexity offers lots of desirable properties. In particular, all local minima of a convex function are global minima, enabling us to find global solutions using local information such as the derivatives.

Definition 2.1.1. A function $f : S \subseteq \mathbb{R}^n \to \mathbb{R}$ is **convex** if S is a convex set and if for all $x, y \in S$, and $t \in [0, 1]$, we have

$$f(tx + (1-t)y) \leq tf(x) + (1-t)f(y).$$

This definition can be simplified if f is continuous:

Lemma 2.1.2. *A continuous function $f : S \subseteq \mathbb{R}^n \to \mathbb{R}$ is convex if and only if f is **mid-point convex**:*

$$f\left(\frac{x+y}{2}\right) \leq \frac{1}{2}f(x) + \frac{1}{2}f(y).$$

With possibly additional assumptions such as continuity, differentiability, or the Lipschitz continuity of the derivative of some order, we can employ a variety of algorithms and tools to find a global minimum of a convex function efficiently, often with convergence guarantees. For example, first-order Taylor approximation of a differentiable function leads us to the method of gradient descent. Similarly, second-order Taylor approximation of twice-differentiable functions leads to Newton's method.

In a convex optimization problem, we often have convex constraints on the feasible solutions. The standard form of a convex optimization problem looks like

$$\begin{aligned} \min \quad & f_0(x) \\ \text{subject to} \quad & f_i(x) \leq 0, i = 1, \ldots, m \\ & Ax = b, \end{aligned}$$

where $f_i : \mathbb{R}^n \to \mathbb{R}$ are convex functions and f_0 is called the **objective function**. A quick intuition for this formulation is that a company has the objective to maximize its profits (represented by $-f_0$) but has various budget constraints (represented by the inequalities). We can convert this real world problem into the standard form by a change of variable through the affine constraint. Since the composition of an affine function with a convex function is still convex, the nice behaviors of convex functions are preserved.

Moreover, we can replace the inequality constraints with generalized inequality constraints. In particular, some nice cones that occur in many real world optimization problems induce a partial order on $\mathbb{R}^n$.

Definition 2.1.3. A subset $\mathcal{K}$ of a real vector space is a **cone** if for every $x \in \mathcal{K}$ and $t \geq 0$, we have $tx \in \mathcal{K}$. A cone $\mathcal{K}$ is a **proper cone** if

(i) $\mathcal{K}$ is convex, *i.e.* $sx + ty \in \mathcal{K}$ for any $x, y \in \mathcal{K}$ and $s, t \geq 0$;

(ii) $\mathcal{K}$ is closed, *i.e.* $\overline{\mathcal{K}} = \mathcal{K}$;

(iii) $\mathcal{K}$ is solid, *i.e.* $\mathring{\mathcal{K}} \neq \emptyset$;

(iv) $\mathcal{K}$ is pointed, *i.e.* if both x and $-x$ are in K, then $x = 0$.

A proper cone is defined this way so that it induces a partial order on $\mathbb{R}^n$ defined as

$$x \preceq_{\mathcal{K}} y \iff y - x \in \mathcal{K}.$$

We can check that this is indeed a partial order: $x - x = 0 \in \mathcal{K}$ so $x \preceq_{\mathcal{K}} x$. If $y - x \in \mathcal{K}$ and $x - y = -(y - x) \in \mathcal{K}$, then since $\mathcal{K}$ is pointed, $y - x = 0$ so $x = y$. Finally, if $y - x \in \mathcal{K}$ and $z - y \in \mathcal{K}$, then $z - y + y - x = z - x \in \mathcal{K}$ by convexity, so $x \preceq_{\mathcal{K}} z$. Strict inequality is defined similarly:

$$x \prec_{\mathcal{K}} y \iff y - x \in \mathring{\mathcal{K}}.$$

It is also straightforward to check that the direct product of proper cones is still a proper cone.

Given a proper cone $\mathcal{K}$, a variable $x \in \mathbb{R}^n$, fixed $c \in \mathbb{R}^n$ and $b \in \mathbb{R}^m$, and $A \in \mathbb{R}^{m \times n}$, a **conic program** has the following standard form:

$$\begin{aligned} \min \quad & \langle c, x \rangle \\ \text{subject to} \quad & x \succeq_{\mathcal{K}} 0 \\ & Ax = b, \end{aligned}$$

where $\langle \cdot, \cdot \rangle$ denotes the usual dot product.

From this formulation, we see that in classical conic programming, we solve a linear objective function together with an affine constraint and a proper cone constraint. That is, the feasible set must intersect with $\mathcal{K}$. This inspired the *interior-point methods* (IPM) where solutions are forced to stay inside $\mathcal{K}$ by a **barrier function**. The intuition is that if we can find a function with the proper cone as the domain and the function value goes to infinity as the input approaches the boundary, we can simply add this barrier function to the objective function to ensure that the solution stays inside the proper cone. Then for each iteration, we can solve the system of equations given by the modified KKT conditions using Newton's method. By gradually reducing the weight of the barrier, we can eventually reach a solution while staying inside the cone. For a thorough treatment of IPM, see the book by Nesterov and Nemirovskii [NN94] and the book by Renegar [Ren01].

However, the original IPM using Newton's method relied on the assumption that the Hessian of the objective is Lipschitz continuous. Not only is this a strong assumption, it also suffers from the fact that an affine change of variables might alter the Lipschitz constant, making convergence analysis challenging. Moreover, this is an unnatural assumption for Newton's method because Newton's method itself is *affine-invariant*. That is, an affine change of variables does not affect the solution of Newton's method at each update step. We should therefore expect our assumption to be affine-invariant as well. A condition called *self-concordance* was discovered by Nesterov and Nemirovskii [NN94] that exactly fits our need.

Definition 2.1.4. A thrice differentiable convex function $f : \mathbb{R} \to \mathbb{R}$ is **self-concordant** if

$$|f'''(x)| \leq 2f''(x)^{3/2}.$$

A function $g : \mathbb{R}^n \to \mathbb{R}$ is self-concordant if it is self-concordant along every line in its domain.

Combining the self-concordant assumption with Newton's method, Nesterov and Todd introduced the *primal-dual IPM* [NT97, NT98], an affine-invariant and highly practical algorithm whose convergence does not depend on the condition number of the Hessians.

Even more surprisingly, Theorem 2.5.1 from [NN94] shows that every closed convex domain of $\mathbb{R}^n$ admits a self-concordant barrier function referred as the *universal barrier*.

Although the proof is not constructive, so we might not be able to find such barrier explicitly for arbitrary closed convex sets, this existence result still implies that theoretically the primal-dual IPM can be used to solve an arbitrary convex optimization problem that satisfies certain assumptions (see Theorem 3.2.1 of [NN94]).

Hence as long as we can assume the objective function to be compatible with the notion of self-concordance (Proposition 3.2.1 in [NN94]) and can find a self-concordant barrier function, we can expect the primal-dual IPM to work very well. Due to this reason and the fact that self-concordant barrier functions of many important cones are known, the primal-dual IPM proposed by Nesterov and Todd soon became a gold-standard method for solving conic programs.

However, some important questions remained. First, why does the primal-dual IPM work so well? Second, is the primal-dual IPM the best we can achieve? Nesterov and Todd addressed both of these questions in another paper [NT02]. In particular, by treating the cone as a *Hessian manifold*, they found that the update steps of the primal-dual IPM follow a path close to the geodesic on the cone. A natural follow-up question is, why don't we design an algorithm that tracks the geodesics directly? In the same paper, Nesterov and Todd indeed presented the geodesics of some common cones to facilitate such pursuit, but the derivation was computation-heavy and intuition-driven on a case-by-case basis. This would make studying the theoretical properties of any such algorithm highly challenging. Luckily, Euclidean Jordan algebra (EJA) comes to the rescue.

2.2 Euclidean Jordan algebra

Now we present the EJA results necessary to generalize S-divergence to symmetric cones. Therefore, the content of this chapter is not meant to be exhaustive or completely self-contained, but to help the reader acquire just enough understanding of EJA to follow the next chapter. Some important EJA topics such as the Peirce decomposition and the canonical trace product are skipped due to their limited relevance for our goal. Similarly, many highly technical proofs whose techniques do not immediately help us in the generalization are omitted, so we leave references for the reader. Throughout this chapter, we illustrate various definitions with three canonical examples of symmetric cones or their ambient EJAs: the nonnegative orthant, the positive semidefinite matrices, and the Lorentz/second-order cone. They respectively are the cones used by linear programming (LP), semidefinite programming (SDP), and second-order cone programming (SOCP).

We note that most of the concepts and examples in the rest of this chapter can be found in the following resources: the canonical textbook on symmetric cones by Faraut and Koranyi for advanced users [FK94], a new EJA textbook draft by Michael Orlitzky for optimization researchers [Orl21], Michel Baes's PhD thesis [Bae06], and Manuel Vieira's PhD thesis [Vie07]. In particular, we recommend Michael Orlitzky's book to fill in any gap in the prerequisite knowledge.

The EJAs are a special case of a more general family named Jordan algebras. Note that the terminology *algebra* here is different than the canonical definition, since we do not require multiplication to be associative. Our precise definition is:

Definition 2.2.1. An **algebra** $(M, R, \circ)$ consists of

(i) An R-module $(M, +)$ where R is a commutative ring;

(ii) A binary operation $\circ : M \times M \to M$ which we call "multiplication".

Moreover, multiplication is bilinear with respect to addition and scalar multiplication in M. That is, $\forall\ x, y, z \in M, \alpha \in R$, we have

$$(x + y) \circ z = x \circ z + y \circ z,$$

$$x \circ (y + z) = x \circ y + x \circ z,$$

$$(\alpha x) \circ y = x \circ (\alpha y) = \alpha(x \circ y).$$

Notice that an algebra does not have to contain the multiplicative identity 1_M. An algebra with 1_M is called a **unital algebra**. The **subalgebra generated by $\boldsymbol{x}$**, denoted $R(x)$, in a unital algebra is the smallest unital subalgebra that contains x.

Definition 2.2.2. An algebra $(V, F, \circ)$ is a **Jordan algebra** if F is a field not of characteristic 2 and if multiplication $\circ$ satisfies the following two conditions:

(i) Commutativity: $\forall\, x, y \in V : x \circ y = y \circ x$.

(ii) The Jordan identity: $\forall\, x, y \in V : x \circ ((x \circ x) \circ y) = (x \circ x) \circ (x \circ y)$.

We shall write $x^2 = x \circ x$ for convenience. Since V is an algebra, multiplication is bilinear. Hence left multiplication by any $x \in V$ is linear and therefore a vector space endomorphism of V. We use L_x to denote the endomorphism that represents left multiplication by x. Then the Jordan identity is equivalent to $L_x L_{x^2} = L_{x^2} L_x$. In such case, we say that x and x^2 **operator-commute** because their left multiplication operators commute.

Moreover, L_x is also linear in the subscript. That is, $L_{x+\alpha y}(z) = (x + \alpha y) \circ z = x \circ z + \alpha y \circ z = L_x(z) + \alpha L_y(z)$ so $L_{x+\alpha y} = L_x + \alpha L_y$.

The reader might be appalled by the lack of associativity of multiplication. After all, the notation x^3 would not even well-defined if we don't have $x \circ (x \circ x) = (x \circ x) \circ x$. Thankfully, the Jordan identity is designed this way so that we can at least have well-defined powers:

Theorem 2.2.3. *If V is a Jordan algebra, then V is a power-associative algebra. That is, for any $x \in V$,*

$$x^m \circ x^n = x^{m+n}.$$

Moreover, x^m and x^n operator-commute. That is,

$$L_{x^m} L_{x^n} = L_{x^n} L_{x^m}.$$

The proof of this theorem relies on induction and some *polarization identities* for Jordan algebras derived from the Jordan identity. The proof is not particularly enlightening and we will not use these polarization identities in our proofs, so we refer the reader to [Orl21]. The theorem itself is however very important. It is not hard to see that power-associativity is the bare minimum we need to define minimal and characteristic polynomials on V, which in turn allow us to generalize the almighty spectral theorem from linear algebra.

Note that power-associativity does not apply to the subscripts of multiplication operators. That is, in general $L_{x^m} L_{x^n} \neq L_{x^{m+n}}$. This fact makes working directly with these multiplication operators in EJA challenging. Thus, many straightforward results from linear algebra do not generalize trivially.

Definition 2.2.4. A **Euclidean Jordan algebra** is a triple $(V, \circ, \langle \cdot, \cdot \rangle)$ consisting of a finite-dimensional Jordan algebra $(V, \mathbb{R}, \circ)$ and an inner product that satisfies

$$\forall\ x, y, z \in V : \langle x \circ y, z \rangle = \langle y, x \circ z \rangle,$$

and a multiplicative identity 1_V such that

$$\forall\ x \in V : 1_V \circ x = x = x \circ 1_V.$$

The condition on the inner product is equivalent to the condition that for any $x \in V$, the linear operator L_x is self-adjoint, *i.e.* $L_x^* = L_x$.

Being finite dimensional and a Hilbert space, EJAs behave a lot more nicely than generic Jordan algebras. Many of the definitions and theorems we present below apply to general Jordan algebras, so we will specify in the assumptions. Thankfully, all examples we care about in optimization are EJAs, and we introduce some common examples:

Example 2.2.5 (Hadamard EJA)**.** We can endow the vector space $\mathbb{R}^n$ over $\mathbb{R}$ with a EJA structure. Given $x, y \in \mathbb{R}^n$, define multiplication $\circ$ as entrywise (Hadamard) multiplication:

$$x \circ y := \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \circ \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_1 y_1 \\ \vdots \\ x_n y_n \end{pmatrix}.$$

Let $\langle \cdot, \cdot \rangle$ be the usual inner product on $\mathbb{R}^n$:

$$\langle x, y \rangle := \sum_{i=1}^{n} x_i y_i.$$

We can easily check that this yields a EJA structure with the identity $1_V = (1, 1, \ldots, 1)^T$.

Example 2.2.6 (real symmetric EJA)**.** Since matrix multiplication is not commutative, we cannot obtain a EJA structure on the symmetric matrices $\mathbb{S}^n$ through it. Instead we shall use the symmetrized multiplication:

$$X \circ Y := \frac{XY + YX}{2}.$$

This way, $\circ$ is clearly commutative and the product is always in $\mathbb{S}^n$, unlike matrix multiplication. The inner product is the usual trace inner product:

$$\langle X, Y \rangle := \operatorname{tr}(X^T Y) = \operatorname{tr}(XY).$$

The identity is still the identity matrix I_n.

Notice that since $X \circ X = X^2$ which coincides with matrix multiplication, familiar results from linear algebra still apply to powers of X.

Example 2.2.7 (Jordan spin EJA)**.** We can endow $\mathbb{R}^{n+1}$ over $\mathbb{R}$ with a very different EJA structure. If we write $x = (x_0, \overline{x})^T \in \mathbb{R}^{n+1}$ where $\overline{x} := (x_1 \ldots, x_n)^T \in \mathbb{R}^n$, define

$$x \circ y := \begin{pmatrix} x_0 \\ \overline{x} \end{pmatrix} \circ \begin{pmatrix} y_0 \\ \overline{y} \end{pmatrix} = \begin{pmatrix} \langle x, y \rangle \\ y_0 \overline{x} + x_0 \overline{y} \end{pmatrix},$$

where $\langle \cdot, \cdot \rangle$ is the usual dot product. The identity $1_V = (1, 0, \ldots, 0)^T$. Checking that this indeed defines an EJA is a tedious computational exercise.

Example 2.2.8 (direct product EJA)**.** For any two EJAs $(V, \circ_V, \langle \cdot, \cdot \rangle_V)$ and $(W, \circ_W, \langle \cdot, \cdot \rangle_W)$, we can define multiplication of any $(v_1, w_1), (v_2, w_2)$ in the direct product $V \times W$ the natural way:

$$(v_1, w_1) \circ (v_2, w_2) = (v_1 \circ_V v_2, w_1 \circ_W w_2).$$

Similarly,

$$\langle (v_1, w_1), (v_2, w_2) \rangle = \langle v_1, v_2 \rangle_V + \langle w_1, w_2 \rangle_W .$$

The identity $1_{V\times W}$ is given by $(1_V, 1_W)$.

From this point we will refer to a EJA by its vector space when the multiplication and the inner product is clear from context.

Since $\circ$ is not associative, it is often clumsy to work with multiplication directly. The next definition allows us to work with a special operator instead that often simplifies the computation.

Definition 2.2.9. Let V be a Jordan algebra. The **quadratic representation** $P_x : V \to V$ of $x \in V$ is defined as

$$P_x = 2L_x^2 - L_{x^2}.$$

We know that $P_x \in \mathrm{End}(V)$ since it is a linear combination of compositions of two endomorphisms. However, unlike L_x, P_x is not linear in the subscript. That is, in general P_{x+y} is different from $P_x + P_y$.

Example 2.2.10 (Hadamard EJA)**.** This is a special case when the quadratic representation coincides with the multiplication operator:

$$\begin{aligned} P_x(y) &= 2x \circ (x \circ y) - x^2 \circ y \\ &= \begin{pmatrix} 2x_1(x_1y_1) - x_1^2 y_1 \\ \vdots \\ 2x_n(x_ny_n) - x_n^2 y_n \end{pmatrix} \\ &= L_x(y). \end{aligned}$$

Example 2.2.11 (real symmetric EJA)**.** Given $X, Y \in V$, we have

$$\begin{aligned}
P_X(Y) &= 2L_X^2(Y) - L_{X^2}(Y) \\
&= 2X \circ (X \circ Y) - X^2 \circ Y \\
&= 2X \circ \frac{XY + YX}{2} - \frac{X^2Y + YX^2}{2} \\
&= \frac{X^2Y + XYX + XYX + YX^2 - X^2Y - YX^2}{2} \\
&= XYX = XYX^T.
\end{aligned}$$

Hence we see that the quadratic representation plays the role of matrix congruence operator in the real symmetric EJA. This operator has the representation $X \otimes X$, where $\otimes$ denotes the Kronecker product.

Example 2.2.12 (Jordan spin EJA)**.** Given $x, y \in V$, we unapologetically skip the computation and directly present

$$P_x = 2xx^T - (x_0^2 - \|\overline{x}\|^2) \begin{pmatrix} 1 & 0 \\ 0 & -I_n \end{pmatrix}.$$

It is not immediately obvious why the quadratic representation P_x is in some sense the more natural operator to work with than the multiplication operator L_x. The advantage of the quadratic representation will be much more evident when we discuss symmetric cones.

Below we present some basic properties of the quadratic representation that we use all the time in Chapter 3. An element x is said to be **invertible** if there exists an element $y \in F(x)$ s.t. $x \circ y = 1_V$.

Proposition 2.2.13. *Let V be a EJA, $\alpha \in \mathbb{R}$, and $x, y \in V$. Then we have*

(1) $P_{\alpha x} = \alpha^2 P_x$.

(2) $P_x(1_V) = x^2$.

(3) More generally, $P_x(x^t) = x^{t+2}$ *for* $t \in \mathbb{N}$*. If* x *is invertible, then this holds for* $t \in \mathbb{Q}$.

For a differentiable function $f : V \to \mathbb{R}$, we can define the **gradient** ∇f the usual way since we have an inner product. Below we present some results about inversion and a proof can be found in Section 2.3 of [Vie07].

Proposition 2.2.14. *Let V be a EJA and $x \in V$. Then x is invertible if and only if P_x is invertible. For an invertible x, we have the following results:*

(1) $(P_x)^{-1} = P_{x^{-1}}$;

(2) $\nabla x^{-1} = -P_x^{-1}$;

(3) $(P_x(y))^{-1} = P_x^{-1}(y^{-1})$;

(4) the fundamental identity: $P_{P_x y} = P_x P_y P_x$.

The reason *(4)* is given such an impressive name is that we can define a Jordan algebra using this identity in lieu of the Jordan identity.

We now present a classification result to conclude this section.

Definition 2.2.15. An EJA is **simple** if its only algebra ideals are itself and the trivial algebra ideal $\{0\}$.

Theorem 2.2.16 (complete classification of EJAs)**.** *Every finite-dimensional EJA can be decomposed into a direct sum of a finite number of simple EJAs in a unique way up to indexing. Every simple EJA is isomorphic to one of the following EJAs:*

(1) a bilinear form EJA on $\mathbb{R}^{n+1}$ over $\mathbb{R}$ (generalized spin algebra EJA where we can change the dot product to any valid inner product),

(2) a real symmetric EJA on $\mathbb{R}^{n\times n}$ over $\mathbb{R}$,

(3) a complex Hermitian EJA on $\mathbb{C}^{n\times n}$ over $\mathbb{R}$,

(4) a quaternion Hermitian EJA on $\mathbb{H}^{n\times n}$ over $\mathbb{R}$,

(5) an octonion Hermitian EJA on $\mathbb{O}^{3\times 3}$ over $\mathbb{R}$ (the Albert algbera).

A proof can be found in Proposition III 4.4 and Chapter V of [FK94].

Example 2.2.17 (Hadamard EJA)**.** It might be surprising that the Hadamard EJA isn't a simple EJA even though it looks "simple". It is in fact a direct product of 1×1 real symmetric EJAs, which are just $\mathbb{R}$ with the usual algebraic structure.

Remark 2.2.18. EJAs were originally developed by Pascual Jordan, not to be mistaken with Camille Jordan of the Jordan canonical form and the Jordan curve theorem fame, to model quantum mechanics so that operations on Hermitian matrices representing observables always return an observable. It was the complete classification of EJAs proved by Jordan, Wigner, and von Neumann that doomed its chance to fulfill its purpose, since the largest dimension of any potential quantum model candidate from simple EJAs is 27 (the Albert algebra), too small for what it was designed to do. This classification is however very helpful for optimization researchers to know exactly what types of problems EJA can help to solve. That is, it is *the* theoretical framework to use for studying symmetric cones. Before we diving into that topic, let us understand the concept of determinant on EJA first. This requires the spectral theorem.

2.3 Spectral decomposition

For this section, we shall assume that F is an infinite field. This allows us to obtain an isomorphism between polynomial functions and polynomials:

Lemma 2.3.1. *If $(V, R, \circ)$ is a nontrivial, power-associative, and unital algebra over an infinite integral domain R and if the polynomial functions $p = q$ on V, then $p(\Lambda) = q(\Lambda)$ in $R[\Lambda]$. Hence $p(\Lambda) \mapsto p$ is a ring isomorphism.*

This is Corollary 6 from [Orl21] and implies that we can switch back and forth between a polynomial function and a polynomial.

Throughout this thesis, we shall use Det to denote the usual matrix determinant, and det to denote the determinant of EJA. We reserve captial greek letters Λ and Ξ to denote indeterminates of a polynomial, and capital English letters to denote matrices. Given a polynomial $p(\Lambda)$, we use the notational shortcut $p(x)$ to denote $\phi_x(p(\Lambda))$ where ϕ_x is the evaluation homomorphism at x. When we treat x as a variable, we refer to $p(x)$ as a *polynomial function* of x. To make it even more confusing, we shall call elements of the field of fractions of a polynomial integral domain *rational functions* per convention even though they are not functions. Finally, for an element x in V and a basis B of V, we denote the vector of coefficients of x under the basis B as $(x)_B = (x_1, \ldots, x_n) \in F^n$.

One of the most celebrated triumphs in the studying of EJA is the successful generalization of the spectral theorem from linear algebra. The finite dimension and power-associativity of EJA are the key ingredients that make it work.

Consider the polynomial algebra $F[\Lambda]$ over a field F with indeterminate Λ. Define $I_x := \{p(\Lambda) \in F[\Lambda] : p(x) = 0\}$. Recall that an algebra ideal is a ring ideal that is closed under scalar multiplication. Since $I_x = \ker \phi_x$ where ϕ_x is the evaluation algebra homomorphism at x, I_x is an algebra ideal. Since F is a field, $F[\Lambda]$ is a principal ideal domain (PID), and I_x is generated by some polynomial $m_x(\Lambda) \in F[\Lambda]$ and we can assume this generator to be monic. It is unique because if there is another monic generator $m'_x(\Lambda)$, then $m_x(\Lambda) - m'_x(\Lambda) \in I_x$. But since $m_x(\Lambda)$ and $m'_x(\Lambda)$ have the same degree and are monic, their difference must have degree strictly less than $\deg m_x(\Lambda)$, contradicting the minimality of the degree of $m_x(\Lambda)$ in I_x. Thus we call $m_x(\Lambda)$ the **minimal polynomial** of x.

By the first isomorphism theorem we obtain $\operatorname{im} \phi_x \cong F[\Lambda]/\langle m_x(\Lambda)\rangle$. Notice that $\operatorname{im} \phi_x$ is just the set of all possible linear combinations of powers of x, which is exactly $F(x)$, the unital subalgebra generated by x. By power-associativity of V, we conclude that $F(x)$ is an associative algebra.

Definition 2.3.2. Let V be a finite-dimensional power-associative unital F-algebra and

$x \in V$. Let $F(x)$ denote the unital subalgebra generated by x. Then the **degree** of x is:

$$\deg(x) := \dim(F(x)) = \deg m_x(\Lambda).$$

The **rank** of V is defined as

$$\operatorname{rank} V = \max\{\deg(x) : x \in V\}.$$

An element x is **regular** if $\deg(x) = \operatorname{rank} V$.

We note that this definition of the minimal polynomial mirrors the minimal polynomial of a vector space endomorphism. However, we cannot easily generalizes the definition of the characteristic polynomial from vector spaces since we have yet to define the determinant on EJA. Without the *a priori* determinant, we would need to work a little harder. The roadmap is that we first define the characteristic polynomial of a regular element to be its minimal polynomial. Then we use the fact that regular elements are dense in EJA and continuity of polynomial functions to extend the characteristic polynomial of regular elements to all elements. This is a sophisticated proof that invokes techniques from algebraic geometry. Since for the purpose of this thesis we only need the characteristic polynomials of the regular elements, we shall only cover the proof for regular elements and refer the reader to [Orl21] for a complete proof.

Lemma 2.3.3 (Gauss). *Let R be a unique factorization domain (UFD) with field of fractions F and let $p(\Xi) \in R[\Xi]$. If $p(\Xi)$ is reducible in $F[\Xi]$ then $p(\Xi)$ is reducible in $R[\Xi]$.*

Proposition 2.3.4. *Let V be an n-dimensional power-associative unital F-algebra with rank r. Then for any basis B, there exists unique polynomials $a_1(\Xi_1, \ldots, \Xi_n), a_2(\Xi_1, \ldots, \Xi_n), \ldots, a_r(\Xi_1, \ldots, \Xi_n) \in F[\Xi_1, \ldots, \Xi_n]$ such that each $a_i(x) := a_i((x)_B) = a_i(x_1, \ldots, x_n) \in F$ and the minimal polynomial of every regular element $x \in V$ is given by*

$$m_x(\Lambda) = \Lambda^r - a_1(x)\Lambda^{r-1} + a_2(x)\Lambda^{r-2} + \cdots + (-1)^r a_r(x)\Lambda^0.$$

Proof. Note that this proof frequently uses Lemma 2.3.1: since $F[\Xi_1, \ldots, \Xi_n]$ is an infinite unique factorization domain (UFD), there exists a natural isomorphism between polynomials in $F[\Xi_1, \ldots, \Xi_n]$ and polynomial functions with variables $x_1, \ldots, x_n \in F$.

Fix a regular element $y \in V$. Then $\{1_V, y, y^2, \ldots, y^{r-1}\}$ is a basis of the subalgebra $F(y)$ and therefore linearly independent. Since V is finite-dimensional, by the Building-Up Lemma there exists a basis $B = \{1_V, y, y^2, \ldots, y^{r-1}, b_1, \ldots, b_{n-r}\}$ of V.

For any $x \in V$ (treating as a variable), define

$$q(x) := \mathrm{Det}((1_V)_B, (x)_B, \cdots, (x^{r-1})_B, (b_1)_B, \ldots, (b_{n-r})_B).$$

Denote this matrix inside Det as M. Since we can express the coefficents of $(x^k)_B$ as a linear combination of $x_1, \ldots, x_n$, $q(x)$ is in fact a polynomial function with variables $x_1, \ldots, x_n$. Hence $q(\Xi_1, \ldots, \Xi_n)$ is a polynomial in $F[\Xi_1, \ldots, \Xi_n]$. Notice $q(y) = 1_F$ since it is just the determinant of the identity matrix, so $q(\Xi_1, \ldots, \Xi_n)$ must not be the zero polynomial. This means that we can put it in the denominator of rational functions. If $q(x) \neq 0 \in F$, then M has full rank so $1_V, x, \ldots, x^{r-1}$ must be linearly independent, implying that x has degree r and is a regular element. So by definition of the degree of x, the minimal polynomial of x has degree r. That is, we have

$$m_x(\Lambda) = \Lambda^r - a_1(x)\Lambda^{r-1} + \cdots + (-1)^r a_r(x)\Lambda^0.$$

Since $m_x(x) = 0$, we have

$$x^r = a_1(x)x^{r-1} + \cdots + (-1)^{r-1} a_r(x)1_V.$$

Now, by expressing x^k in the basis B and treating $a_j(x)$ as unknowns in F, we obtain a system of equations:

$$\begin{pmatrix} (1_V)_B & (x)_B & \cdots & (x^{r-1})_B \end{pmatrix} \begin{pmatrix} a_1(x) \\ a_2(x) \\ \vdots \\ a_r(x) \end{pmatrix} = (x^r)_B.$$

We can solve each $a_j(x)$ using Cramer's rule:

$$a_j(x) = (-1)^{j-1}\frac{\mathrm{Det}((1_V)_B, \ldots, (x^{j-1})_B, (x^r)_B, (x^{j+1})_B, \ldots, (x^{r-1})_B, (b_1)_B, \ldots, (b_{n-r})_B)}{q(x)}.$$

Since $q(\Xi_1, \ldots, \Xi_n) \neq 0$, each $a_j(\Xi_1, \ldots, \Xi_n)$ associated with the coefficients $a_j(x)$ of $m_x(\Lambda)$ is a rational function in the field of fractions of $F[\Xi_1, \ldots, \Xi_n]$. Let $c_{L_x}(\Lambda) = \mathrm{Det}(\Lambda I - L_x)$ be the characteristic polynomial of the vector space endomorphism L_x. By the Cayley–Hamilton Theorem, $c_{L_x}(L_x) = 0$. Recall that $L_x 1_V = x$. Hence by linearity of $\circ$, we can obtain a polynomial function of x via a polynomial function of L_x acting on 1_V as left multiplication. Thus, $c_{L_x}(x) = c_{L_x}(L_x)1_V = 0 \cdot 1_V = 0$. So $c_{L_x}(\Lambda) \in I_x$ and $m_x(\Lambda)$ divides $c_{L_x}(\Lambda)$. Since L_x has entries in F, $c_{L_x}(\Lambda)$ has coefficients in $F \subset F[\Lambda]$. Since $F[\Xi_1, \ldots, \Xi_n]$ is a UFD, by Gauss's Lemma $m_x(\Lambda)$ as a factor of $c_{L_x}(\Lambda)$ also has coefficients in $F[\Xi_1, \ldots, \Xi_n]$ as well. That is, we can choose $a_j(\Xi_1, \ldots, \Xi_n)$ to be polynomials. The uniqueness of each $a_j(\Xi_1, \ldots, \Xi_n)$ follows from the uniqueness of $m_x(\Lambda)$. $\square$

For regular elements, we simply define characteristic polynomials to be their minimal polynomials. Finally, by the believable magic of density of regular elements and continuity of polynomial functions, we can extend characteristic polynomials of regular elements to all elements of V, completing the construction of characteristic polynomials in EJA.

Definition 2.3.5. Let V be a finite-dimensional power-associative unital F-algebra with rank r. Suppose the characteristic polynomial of $x \in V$ is

$$c_x(\Lambda) = \Lambda^r - a_1(x)\Lambda^{r-1} + a_2(x)\Lambda^{r-2} + \cdots + (-1)^r a_r(x)\Lambda^0.$$

The **determinant** and the **trace** of x are defined respectively as

$$\det(x) := a_r(x), \qquad \mathrm{tr}(x) := a_1(x).$$

Let $\widetilde{L}_x$ denote L_x with domain restricted to the subalgebra $F(x)$. If x is regular, we know that under the canonical basis $\{1_V, x, x^2, \ldots, x^{r-1}\}$ of $F(x)$, multiplication by x can be represented by the following matrix:

$$\widetilde{L}_x = \begin{pmatrix} 0 & 0 & \cdots & 0 & (-1)^{r-1}a_r(x) \\ 1 & 0 & \cdots & 0 & (-1)^{r-2}a_{r-1}(x) \\ 0 & 1 & \cdots & 0 & (-1)^{r-3}a_{r-2}(x) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & a_1(x) \end{pmatrix}.$$

This is the companion matrix of $c_x(\Lambda)$, so the characteristic polynomial of $\widetilde{L}_x$ is just $c_x(\Lambda)$. It is easy to see that due to the abundance of zeros, $\mathrm{Det}(\widetilde{L}_x) = a_r(x) =: \det x$. Together with the identity that $\widetilde{L}_{(\Lambda 1_V - x)} = \Lambda I - \widetilde{L}_x$ by linearity of multiplication, we obtain

$$c_x(\Lambda) = c_{\widetilde{L}_x}(\Lambda) := \mathrm{Det}(\Lambda I - \widetilde{L}_x) = \mathrm{Det}(\widetilde{L}_{(\Lambda 1_V - x)}) = \det(\Lambda 1_V - x).$$

So we recover the definition of characteristic polynomial for matrices in EJA:

Proposition 2.3.6. *Let V be a finite-dimensional power-associative unital F-algebra with rank r. Then the characteristic polynomial of any element $x \in V$ can be expressed as:*

$$c_x(\Lambda) = \det(\Lambda 1_V - x).$$

Example 2.3.7 (Hadamard EJA)**.** Given $x = (x_1, \ldots, x_n)^T \in V$, we have

$$\det x = \prod_{i=1}^{n} x_i,$$
$$c_x(\Lambda) = \prod_{i=1}^{n} (\Lambda - x_i).$$

Hence the rank of V is n.

Example 2.3.8 (real symmetric EJA)**.** We have

$$\det X = \mathrm{Det}\, X,$$
$$c_X(\Lambda) = \det(\Lambda I - X) = \mathrm{Det}(\Lambda I - X).$$

This coincides with the usual determinant definition for matrices, as we would expect from power-associativity and uniqueness of minimal polynomials.

Example 2.3.9 (Jordan spin EJA)**.** We have

$$\begin{aligned} \det x &= x_0^2 - \|\overline{x}\|^2, \\ c_x(\Lambda) &= \det(\Lambda 1_V - x) \\ &= \Lambda^2 - 2x_0\Lambda + x_0^2 - \|\overline{x}\|^2. \end{aligned}$$

This reveals that $\operatorname{rank} V = 2$.

Example 2.3.10 (direct product EJA)**.** Let $V = V_1 \times V_2$ be the direct product EJA. Then given $(x_1, x_2) \in V$, we have

$$\det(x_1, x_2) = \det x_1 \det x_2.$$

Proposition 2.3.11. *For all $v \in V$, and $x, y \in F(v)$, we have*

$$\det(x \circ y) = \det x \det y.$$

Moreover, $\det(1_V)$ *equals 1.*

Note that this is only true in the subalgebra $F(v)$ and not true for elements in V in general!

We again have the familiar result from linear algebra:

Proposition 2.3.12. *Let V be a power-associative unital F-algebra with rank r. An element $x \in V$ is invertible if and only if $\det x \neq 0$. An invertible x has inverse*

$$x^{-1} = \frac{Q(x)}{\det x},$$

where $Q[\Lambda]$ is a polynomial in $F[\Lambda]$ of degree $r - 1$.

Proof. ($\implies$) : Suppose x is invertible, then $x^{-1} \in F(x)$ and $\det x \det(x^{-1}) = \det(x \circ x^{-1}) = \det(1_V) = 1$. So $\det x$ is not 0.

($\Longleftarrow$) : suppose $\det x \neq 0$, then since $c_x(x) = 0$, we have

$$\begin{aligned}
0 &= x^r - a_1(x)x^{r-1} + a_2(x)x^{r-2} + \cdots + (-1)^r \det(x) 1_V \\
1_V &= (-1)^{r-1} \frac{x^r - a_1(x)x^{r-1} + \cdots + (-1)^{r-1} a_{r-1}(x)x}{\det x} \\
1_V &= x \circ \left((-1)^{r-1} \frac{x^{r-1} - a_1(x)x^{r-2} + \cdots + (-1)^{r-1} a_{r-1}(x)}{\det x} \right) \\
&=: x \circ x^{-1}.
\end{aligned}$$

The last calculation also proves the second statement of the proposition. $\square$

The next proposition is perhaps the most useful identity for this thesis.

Proposition 2.3.13. *Let V be a simple EJA with dimension n and rank r, and let $x, y \in V$. We have*

$$\det(P_x(y)) = \det(x)^2 \det(y).$$

Proof. By the fundamental identity, we have

$$\operatorname{Det} P_{P_x y} = \operatorname{Det}(P_x P_y P_x) = \operatorname{Det}(P_x)^2 \operatorname{Det}(P_y).$$

Then by Proposition III 4.2 of [FK94], we have $\operatorname{Det} P_x = (\det x)^{\frac{2n}{r}}$, so

$$\begin{aligned}
(\det(P_x(y)))^{\frac{2n}{r}} &= (\det(x)^2 \det(y))^{\frac{2n}{r}} \\
\det(P_x(y)) &= \det(x)^2 \det(y).
\end{aligned}$$

$\square$

With the characteristic polynomial defined, we can proceed to define another familiar concept:

Definition 2.3.14. The **eigenvalues** of x are the roots of its characteristic polynomial $c_x(\Lambda)$.

The next proposition easily follows from this definition by writing $c_x(\Lambda)$ as a product of roots:

Proposition 2.3.15. *Let V be a power-associative unital F-algebra with rank r and $x \in V$. Then*

$$\det(x) = \prod_{i=1}^{r} \lambda_i \quad \textit{and} \quad \operatorname{tr}(x) = \sum_{i=1}^{r} \lambda_i,$$

where λ_i are the eigenvalues of x.

This result allows us to mostly forget about the characteristic polynomials and greatly simplifies working with the determinant. Linearity of trace is self-evident.

Definition 2.3.16. Let V be a EJA. An **idempotent** $c \in V$ is an element satisfying $c^2 = c$. A **primitive idempotent** $c \in V$ is a non-zero idempotent such that there exist no nonzero idempotents $c_1, c_2 \in V$ satisfying $c = c_1 + c_2$.

Definition 2.3.17. A set $\{c_1, \ldots, c_k\} \subseteq V$ is a **complete system of orthogonal idempotents** if

(i) Each c_i is an idempotent.

(ii) If $i \neq j$, then $c_i \circ c_j = 0$.

(iii) $\sum_{i=1}^{k} c_i = 1_V$.

A complete system of orthogonal *primitive* idempotents is called a **Jordan frame**.

As the name suggests, $c_i \circ c_j = 0$ is in fact equivalent to $\langle c_i, c_j \rangle = 0$. One direction is straightforward: suppose $c \circ d = 0$, then by the property of the EJA inner proudct, $\langle c, d \rangle = \langle c^2, d \rangle = \langle c, c \circ d \rangle = \langle c, 0 \rangle = 0$. The other direction requires the canonical trace product, so we leave the details to [Orl21].

Here we have more operator-commuting results.

Proposition 2.3.18. *Let V be an EJA. If c and d are idempotents, then they operator-commute.*

Corollary 2.3.19. *Let V be an EJA. If x and y are linear combinations of a set of orthogonal idempotents, then x and y operator-commute.*

We are ready to present the spectral theorem. There are in fact two versions of it. The first one is referred as the *unique EJA spectral theorem* by Orlitzsky and only concerns with the spectral decomposition of x within the unital subalgebra generated by x. We shall only present the second version because it is much more general and therefore useful, at the expense of losing a bit of uniqueness:

Theorem 2.3.20 (full EJA spectral theorem)**.** *Suppose that V is an EJA with rank r. Then for any $x \in V$, there exists a Jordan frame $\{c_1, \ldots, c_r\}$ and unique real numbers $\lambda_1 \geq \ldots \geq \lambda_r$ such that*

$$x = \sum_{i=1}^{r} \lambda_i c_i.$$

The λ_i are the eigenvalues of x. The decomposition is unique in the sense that if there exists another Jordan frame $\{d_1, \ldots, d_r\}$ such that $x = \sum_{i=1}^{r} \lambda_i d_i$, then for every eigenvalue t of x, we have

$$\sum_{\lambda_i = t} c_i = \sum_{\lambda_i = t} d_i.$$

Notice that if the eigenvalues of x are all distinct, then the Jordan frame of x is unique. Usually we do not need a unique Jordan frame. Existence is often enough, as is in the case of the next lemma:

Lemma 2.3.21. *Let V be a EJA of rank r. They $x, y \in V$ operator-commute if and only if x and y have full spectral decompositions with respect to a common Jordan frame. That is, $L_x L_y = L_y L_x$ if and only if there exists a Jordan frame $\{c_1, \ldots, c_r\}$ and sets of real numbers*

$\{\lambda_1, \ldots, \lambda_r\}$ and $\{\mu_1, \ldots, \mu_r\}$ such that

$$x = \sum_{i=1}^{r} \lambda_i c_i \text{ and } y = \sum_{i=1}^{r} \mu_i c_i.$$

The backward direction follows from Corollary 2.3.19. The forward direction is highly technical so we delegate the proof to [Orl21].

Example 2.3.22 (Hadamard EJA)**.** We can check that the standard basis coincides with the Jordan frame of any element $x \in V$. The eigenvalues are just $\{x_1, \ldots, x_n\}$, the entries of x.

Example 2.3.23 (real symmetric EJA)**.** Perhaps not surprisingly, the EJA spectral decomposition of the symmetric matrices is equivalent to the linear algebra spectral decomposition, although they might be presented differently. Recall that for any symmetric matrix $X \in V$, the linear algebra spectral decomposition yields a diagonal matrix Λ with eigenvalues on the diagonal and an orthogonal matrix Q with normalized eigenvectors as columns such that $X = Q\Lambda Q^{-1}$. We can rewrite this as a sum:

$$X = \sum_{i=1}^{n} \lambda_i q_i q_i^T,$$

which is in the form of EJA spectral decomposition. Notice that $q_i q_i^T$ is projection onto the invariant subspace spanned by the eigenvector q_i. Hence $q_i q_i^T$ is an idempotent. It is straightforward to check that $\{q_1 q_1^T, \ldots, q_n q_n^T\}$ forms a Jordan frame. So we see that the concept of eigenvectors from linear algebra and Jordan frame are not equivalent but still highly connected through invariant subspaces. In fact, in operator theory, the EJA version of the spectral theorem is used all the time on matrices.

Example 2.3.24 (Jordan spin EJA)**.** Given $x \in V$, since we observe earlier that $\operatorname{rank} V = 2$, then whenever $\overline{x} \neq 0$, x has the following spectral decomposition: $x = \lambda_1(x)c_1(x) + \lambda_2(x)c_2(x)$, where

$$\lambda_1(x) = x_0 + \|\overline{x}\| \quad \text{and} \quad \lambda_2(x) = x_0 - \|\overline{x}\|$$

are the solutions to the characteristic polynomial $c_x(\Lambda) = \Lambda^2 - 2x_0\Lambda + (x_0^2 - \|\overline{x}\|)$, and

$$c_1(x) = \frac{1}{2}\begin{pmatrix} 1 \\ \\ \frac{\overline{x}}{\|\overline{x}\|} \end{pmatrix} \quad \text{and} \quad c_2(x) = \frac{1}{2}\begin{pmatrix} 1 \\ \\ -\frac{\overline{x}}{\|\overline{x}\|} \end{pmatrix}.$$

If $\overline{x} = 0$, then the spectral decomposition of x is just x itself.

In the Lorentz cone $\mathbb{L}^3 \subseteq \mathbb{R}^3$, we can visualize $c_1(x)$ and $c_2(x)$ as two opposing vectors in the cone with height 1 whose projections to the xy plane coincide with $\pm\overline{x}$ (see page 19 of [AG03]).

Given $S \subseteq \mathbb{R}$, denote V_S as the set of all elements in V with eigenvalues completely in S. Thanks to the spectral decomposition, we can now extend any function $f : S \to \mathbb{R}$ to a **spectral function** $F : V_S \to V$ by applying f to the eigenvalues of any element: if $x = \lambda_1 c_1 + \cdots + \lambda_r c_r$, then

$$F(x) = \sum_{i=1}^{r} f(\lambda_i)c_i.$$

This allows us to define the power function on all elements in V with nonnegative eigenvalues:

$$x^t := \sum_{i=1}^{r} \lambda_i^t c_r$$

for any $t \in \mathbb{R}$. When $t = -1$, we see that x^{-1} from this definition indeed coincides with the inverse of x.

2.4 Symmetric cones

With the determinant of EJA defined, we are now ready to understand symmetric cones and their natural barrier functions.

Definition 2.4.1. Let $\mathcal{K}$ be a cone in an ambient vector space V. Define the **automorphism group** of $\mathcal{K}$ to be

$$\mathrm{Aut}(\mathcal{K}) := \{g \in GL(V) : g\mathcal{K} = \mathcal{K}\},$$

where $GL(V)$ is the general linear group on V.

Although [FK94] only define the automorphism group on the interior, we choose this equivalent definition (see Page 4 of [FK94]) to align with the common definition in the literature. Under either definition, the automorphism group is a closed subset of $GL(V)$ and therefore a Lie group , but this fact is not important for our purpose.

Recall that a group action of G on a nonempty set A is **transitive** if for any $a, b \in A$, there exists $g \in G$ such that $g.b = a$. We are ready to define the protagonist of this thesis:

Definition 2.4.2. A cone $\mathcal{K}$ in a finite-dimensional real inner product space is a **symmetric cone** if it is self-dual and homogeneous. That is, $\mathcal{K} = \mathcal{K}^* := \{y \in V : \langle y, x\rangle \geq 0 \ \forall\ x \in \mathcal{K}\}$ and $\mathrm{Aut}(\mathcal{K})$ acts transitively on $\mathring{\mathcal{K}}$.

Note that since the dual cone is always closed and convex, a symmetric cone must be closed and convex as well. It turns out that a symmetric cone is in fact proper. Moreover, *self-scaled cones* described in [NT97] coincide with symmetric cones. But we shall soon see that symmetric cones have more than one alter ego.

Definition 2.4.3. Let V be a EJA. Define the **cone of squares** of V as

$$\mathcal{K}(V) = \{x^2 : x \in V\}.$$

We simply refer to $\mathcal{K}(V)$ as $\mathcal{K}$ when the ambient EJA V is clear from context.

Note that the three examples $\mathbb{R}^n_+, \mathbb{S}^n_+$ and $\mathbb{L}^{n+1}$ from Chapter 1 are respectively the cone of squares of Hadamard EJA, real symmetric EJA, and Jordan spin EJA. We have finally unified them using EJA!

Denote the set of invertible elements of an EJA V by $\mathcal{I}$. It turns out that the interior of the cone of squares coincides with

$$\mathring{\mathcal{K}} = \{x^2 : x \in \mathcal{I}\}.$$

Proposition III 2.2 of [FK94] states an important result:

Proposition 2.4.4. *If $x \in \mathcal{I}$, then $P_x \in \mathrm{Aut}(\mathcal{K})$.*

It follows that P_x acts transitively on $\mathring{\mathcal{K}}$. This proposition has an even more important consequence that we will use frequently:

Corollary 2.4.5. *For any $a \in \mathcal{I}$, P_a is an order-isomorphism on $\mathcal{K}$.*

Proof. Let $x, y \in \mathcal{K}$ satisfy $x \preceq_{\mathcal{K}} y$. Since $y - x \in \mathcal{K}$ and $P_a \in \mathrm{Aut}(\mathcal{K})$, $P_a(y - x) = P_a(y) - P_b(y) \in \mathcal{K}$. That is, $P_a(x) \preceq_{\mathcal{K}} P_a(y)$. Strict inequality follows similarly. □

We now present a marvelous bridge between geometry and algebra.

Theorem 2.4.6. *A cone is symmetric if and only if it is the cone of squares of some EJA.*

We refer the reader to Chapter III of [FK94] for a proof. It is much easier to see that any symmetric cone is pointed and solid by showing that any cone of squares is. Combining this theorem with the complete classification theorem, it is not hard to see that if we define an **irreducible symmetric cone** to be the cone of squares arise from a simple EJA, then we only have five classes of irreducible symmetric cones corresponding to five classes of simple EJAs, and

Proposition 2.4.7. *Any symmetric cone can be decomposed into a direct sum of a finite number of irreducible symmetric cones in a unique way up to indexing.*

The converse is also true: any finite direct product of symmetric cones is a symmetric cone, since we can write every element as a tuple of squares so it is a cone of squares under elementwise multiplication.

Definition 2.4.8. Let V be an EJA and $\mathcal{K}$ be its symmetric cone. Then an endomorphism $A : V \to V$ is **positive definite**, denoted $A > 0$, if for all nonzero $x \in V$, we have

$$\langle Ax, x\rangle > 0.$$

Proposition 2.4.9. *Let $\mathcal{K}$ be a symmetric cone and $x \in \mathcal{K}$. Then the following statements hold:*

(i) The quadratic representation P_x is positive semidefinite and is positive definite if $x \in \mathring{\mathcal{K}}$.

(ii) $P_x^{1/2} = P_{x^{1/2}}$. More generally, $P_x^t = P_{x^t}$ whenever x^t is defined.

See Section 2.5 of [Vie07] and Lemma 2.5 of [TWK21] for proofs.

Definition 2.4.10. Let V be an EJA and $\mathcal{K}$ be its symmetric cone. The **natural barrier function** $B : \mathring{\mathcal{K}} \to \mathbb{R}$ is defined as

$$B(x) = -\log\det(x),$$

where $\log : \mathbb{R}_{++} \to \mathbb{R}$ is the usual natural logarithmic function.

Since any element $x \in \mathring{\mathcal{K}}$ has strictly positive eigenvalues, $\det(x) > 0$ so the composition is well-defined. As x approaches $\partial\mathcal{K}$, we see that at least one eigenvalue approaches 0 so $\det(x) \to 0$ and $B(x) \to \infty$. This is exactly the behavior we expect from a barrier function. Moreover, $B(x)$ is self-concordant. We shall not prove it here but the sketch is that log is self-concordant and we can show that $B(x)$ is self-concordant on every geodesic in V by replacing the determinant with the eigenvalues. See Example 9.5 of [BV04] for the special case of $\mathbb{S}_+^n$.

This is how we obtain the barrier functions of $\mathbb{R}_+^n, \mathbb{S}_+^n$, and $\mathbb{L}^{n+1}$ in Chapter 1. Now we see that instead of working with the barrier functions individually, we can simply prove results and design algorithms using this simple expression for all symmetric cones at once.

Example 2.4.11 (product cone)**.** Suppose $\mathcal{K}_1$ and $\mathcal{K}_2$ are symmetric cones. Then the natural barrier function of the product cone $\mathcal{K}_1 \times \mathcal{K}_2$ is simply the sum of their individual barrier functions:

$$B(x, y) = -(\log\det x + \log\det y) = -\log(\det x \det y).$$

See Theorem 2.1 of [NT97] for details.

Proposition 2.4.12. *Let V be an EJA and $\mathcal{K}$ be its symmetric cone. Let B be the natural barrier function of $\mathcal{K}$. Then in the usual sense of the gradient and the Hessian in the Hilbert space V, for all $x \in \mathring{\mathcal{K}}$, we have*

(i) $\nabla B(x) = -x^{-1}$,

(ii) $\nabla^2 B(x) = P_x^{-1}$.

See Section 2.6 of [Vie07] for a proof.

Corollary 2.4.13. *The natural barrier function $B(x)$ is strictly convex.*

Proof. For any nonzero $x \in \mathring{\mathcal{K}}$, we have $x^{-1} \succ_{\mathcal{K}} 0$ so $P_{x^{-1}} = P_x^{-1} > 0$ by Proposition 2.4.9 (ii). Since $\mathring{\mathcal{K}}$ is convex, by the second-order conditions of convexity B is a strictly convex function. □

Finally, we introduce some definitions that will be useful in the next chapter.

Definition 2.4.14. Let V be a simple EJA of rank r and $\mathcal{K}$ be its symmetric cone. For any given $f : S \subseteq \mathbb{R} \to \mathbb{R}$, f induces a spectral function $F : V_S \to V$, where V_S is the set of elements of V with eigenvalues completely contained in S. Then we say

(a) f is **SC-monotone of order $\boldsymbol{r}$** if for any $x, y \in V_S$,

$$x \preceq_{\mathcal{K}} y \implies F(x) \preceq_{\mathcal{K}} F(y);$$

(b) f is **SC-concave of order $\boldsymbol{r}$** if for any $x, y \in V_S$,

$$\forall\ s \in [0,1], \qquad F(sx + (1-s)y) \preceq_{\mathcal{K}} sF(x) + (1-s)F(y).$$

We say f is **SC-monotone** or **SC-concave** if it is respectively SC-monotone or SC-concave of all orders.

These definitions generalize the concepts of operator monotone and operator concave from operator theory.

2.5 Geodesic convexity

So far we have introduced some of the most powerful Jordan algebraic tools to study symmetric cones. Recall that we are interested in understanding the geometry (*e.g.*, the geodesics) of symmetric cones, but it is still unclear how the Jordan algebraic tools we presented above help us achieve that. Below we illustrate the connection between symmetric cones as Riemannian manifolds and as cone of squares of EJAs. Since Riemannian geometry is out of the scope of this thesis, we only aim to provide the big picture and delegate the complete geometric details to [Vis18] and Chapter 10 of [Bou20]. For symmetric cones, it suffices to understand their geodesic segments intuitively as the "shortest path between two points" or "generalized line segments on the cone".

First, we briefly justify why the interior of a symmetric cone $\mathring{\mathcal{K}}$ is a smooth manifold. Since we have a finite-dimensional vector space, $V \cong \mathbb{R}^n$ for some n and this linear isomorphism yields a global chart. Hence V is a smooth manifold. Since $\mathring{\mathcal{K}}$ is the preimage of the open set $\mathbb{R}_{++}$ under the continuous function det, it is open in V and therefore is a smooth open submanifold.

The following theorem allow us to unify the geometric and algebraic structures of symmetric cones:

Theorem 2.5.1. *Let $\mathcal{K}$ be a symmetric cone with ambient EJA V. Then the Hessian $P_x^{-1}$of the natural barrier function $B(x) = -\log\det(x)$ of $\mathcal{K}$ induces a Riemannian metric tensor g, where $g_x(u, v) = \langle P_x^{-1}u, P_x^{-1}v\rangle$. Furthermore, $(\mathring{\mathcal{K}}, g)$ is a complete, connected Riemannian manifold.*

Proof. Since $\langle\cdot,\cdot\rangle$ is defined to be symmetric bilinear and P_x^{-1} is positive definite by Proposition 2.4.9 (i), g_x is a valid metric tensor. Since P_x^{-1} is a composition of smooth functions in x, g_x varies smoothly with $x \in \mathring{\mathcal{K}}$. So $(\mathring{\mathcal{K}}, g)$ is a Riemannian manifold. Since $\mathring{\mathcal{K}}$ endowed with the Riemannian distance is a complete metric space [NT02], by Hopf–Rinow Theorem in Chapter 10 of [Bou20], $\mathring{\mathcal{K}}$ is geodesically-complete. Finally, $\mathring{\mathcal{K}}$ is convex and therefore connected. □

Then by Theorem 10.8 of [Bou20], every two points in $\mathring{\mathcal{K}}$ have a minimizing geodesic segment connecting them. This geodesic segment is in fact unique:

Proposition 2.5.2. *Let $\mathcal{K}$ be a symmetric cone. The unique geodesic segment $c : [0,1] \to \mathring{\mathcal{K}}$ connecting $x, y \in \mathring{\mathcal{K}}$ can be expressed as:*

$$c(t) = P_{x^{1/2}} \left(P_{x^{-1/2}}(y)\right)^t \;\; \forall \; t \in [0,1].$$

We also use the notation $x\#_t y := c(t)$ to make the end points more explicit.

See Proposition 2.6 of [Lim01] for a proof. In addition, for an abstract treatment of geodesic midpoint in the C^*-algebra setting, see [LL07a, LL07b].

This is exactly how we find the geodesics of $\mathbb{R}^n_+, \mathbb{S}^n_+$ and $\mathbb{L}^{n+1}$ in Chapter 1. Compared to the geodesic of the second-order cone, this abstract formula is a godsend.

Corollary 2.5.3. *If $x \in \mathring{\mathcal{K}}$, then we have*

$$x^t = 1\#_t x.$$

Definition 2.5.4. The **geometric mean** of x and y, denoted by $x\#y$, is defined as the geodesic midpoint of x and y. That is, $x\#y := c\left(\frac{1}{2}\right) = y\#x$.

Finally, we define g-convexity.

Definition 2.5.5. A subset S of a Riemannian manifold $\mathcal{M}$ is **geodesically convex** (g-convex) if, for every $x, y \in S$, there exists a geodesic segment $c : [0,1] \to \mathcal{M}$ such that $c(0) = x, c(1) = y$ and $c([0,1]) \subseteq S$.

Definition 2.5.6. Let $\mathcal{M}$ be a Riemannian manifold and $S \subseteq \mathcal{M}$. Then a function $f : S \to \mathbb{R}$ is **geodesically convex** (g-convex) if S is g-convex and $f \circ c : [0,1] \to \mathbb{R}$ is convex for each geodesic segment $c : [0,1] \to \mathcal{M}$ whose image is in S (with $c(0) \neq c(1)$). That is, f is g-convex if for all $x, y \in S$ and all geodesics c connecting x to y,

$$f(c(t)) \leq (1-t)f(x) + tf(y) \;\forall\; t \in [0,1].$$

It is not hard to see that g-convexity is a natural generalization of convexity and therefore inherits many benefits of convexity, including the fact that all local minima are global minima. Algorithms designed for optimizing convex problems can also be generalized to g-convex problems.

Corollary 2.5.7. *The interior of any symmetric cone, $\mathring{\mathcal{K}}$, is g-convex.*

Proof. By Theorem 10.8 of [Bou20], any complete and connected Riemannian manifold is g-convex. □

Now we have the background knowledge we need in order to generalize Sra's results.

Chapter 3

S-divergence

Throughout this chapter, we shall assume that V is a simple EJA with rank r and $\mathcal{K}$ is the corresponding irreducible symmetric cone. This allows us to use Proposition 2.3.13. We discuss the non-simple case in the next chapter.

Let f be a continuously differentiable, strictly convex function with domain in a Hilbert space. The **Bregman divergence** of f is defined as:

$$D_f(x, y) := f(x) - f(y) - \langle \nabla f(y), (x - y) \rangle.$$

Intuitively, the Bregman divergence measures how far the convex function deviates from a local linear approximation at y between an initial point y and a final point x. This is not a metric as it satisfies neither symmetry nor the triangle inequality. It is positive definite because any local linear approximation of a strictly convex function lies strictly below the function except at $x = y$. So the positive definiteness still provide some "distance-like" information.

The **Jenson–Shannon divergence** of f is the symmetrized version of Bregman divergence:

$$S_f(x, y) := \frac{1}{2} \left(D_f \left(x, \frac{x+y}{2} \right) + D_f \left(y, \frac{x+y}{2} \right) \right).$$

We define S-divergence on the symmetric cone as the Jenson-Shannon divergence of

$f = -\log\det$, its natural barrier function. We provide the derivation below:

$$\begin{aligned}
\delta_S^2(x,y) := \frac{1}{2}\Bigg(&-\log\det x + \log\det\left(\frac{x+y}{2}\right) + \left\langle \left(\frac{x+y}{2}\right)^{-1}, \left(x - \frac{x+y}{2}\right)\right\rangle \\
&-\log\det y + \log\det\left(\frac{x+y}{2}\right) + \left\langle \left(\frac{x+y}{2}\right)^{-1}, \left(y - \frac{x+y}{2}\right)\right\rangle\Bigg) \\
= \log\det\left(\frac{x+y}{2}\right) &- \frac{1}{2}\log\det x - \frac{1}{2}\log\det y + \left\langle \left(\frac{x+y}{2}\right)^{-1}, \left(\frac{x+y}{2} - \frac{x+y}{2}\right)\right\rangle \\
= \log\det\left(\frac{x+y}{2}\right) &- \frac{1}{2}(\log\det x + \log\det y).
\end{aligned}$$

We rewrite δ_S^2 in a more revealing form:

Definition 3.0.1. The **S-divergence** $\delta_S^2 : \mathring{\mathcal{K}} \times \mathring{\mathcal{K}} \to \mathbb{R}$ of the natural barrier function $f = -\log\det$ is defined as

$$\delta_S^2(x,y) := \log\frac{\det\left(\frac{x+y}{2}\right)}{\sqrt{\det x \det y}}.$$

That is, δ_S^2 is the log of the determinant of the arithmetic mean of x and y over the geometric mean of their determinants. In fact, the denominator equals the determinant of their geometric mean, which we shall prove next. For the rest of this thesis, we shall drop the subscript S for convenience.

3.1 δ is a metric

The goal of this section is to prove that $\delta : \mathring{\mathcal{K}} \times \mathring{\mathcal{K}} \to \mathbb{R}$ is a valid metric. This generalization is novel, and the proof ideas largely mirror that of Sra in the case of Hermitian positive definite matrices [Sra15].

The following notation will appear frequently in this section: let V be a simple EJA with rank r and let $x \in V$. Then $\lambda(x) = \lambda^{\downarrow}(x)$ denotes the vector of unique eigenvalues of x in descending order, *i.e.*, $\lambda_1 \geq \ldots \geq \lambda_r$, which is the default order given by the spectral decomposition. Let $\lambda_i(x)$ denote the ith entry of $\lambda(x)$. Let $\lambda^{\uparrow}(x)$ denote the vector of eigenvalues in ascending order. Let $C = \{c_1, \ldots, c_r\}$ be any Jordan frame in V. Define

$\lambda_C(x) := \lambda_1 c_1 + \cdots + \lambda_r c_r$. Here we simply replace the Jordan frame of x from its full EJA spectral decomposition with C. We see that $\det x = \det(\lambda_C(x)) = \det(\lambda_C^{\uparrow}(x)) = \prod_{i=1}^r \lambda_i$.

First, we show that δ is non-negative. This requires a few steps.

Proposition 3.1.1. *For any $a, b \in \mathring{\mathcal{K}}$, we have*

$$\det(a\#b) = \sqrt{\det a \det b}.$$

Proof. Notice that for any $x \in \mathring{\mathcal{K}}$ and $t \in \mathbb{R}$, since $x^t := \sum_{i=1}^r \lambda_i^t c_i$, we have $\det(x^t) = \prod_{i=1}^r \lambda_i(x)^t = \left(\prod_{i=1}^r \lambda_i(x)\right)^t = (\det x)^t$. Let $t = \frac{1}{2}$, and the rest is straightforward computation using Proposition 2.3.13:

$$\begin{aligned}
\det(a\#b) &= \det\left(P_{a^{1/2}}\left(P_{a^{-1/2}}(b)\right)^{1/2}\right) \\
&= \det a \det\left(\left(P_{a^{-1/2}}(b)\right)^{1/2}\right) \\
&= \det a \left(\det\left(P_{a^{-1/2}}(b)\right)\right)^{1/2} \\
&= \det a \sqrt{\det\left(a^{-1}\right)\det b} \\
&= \sqrt{\det a \det b}.
\end{aligned}$$

□

This shows that the geometric mean of the determinants of a, b is indeed the determinant of the geometric means of a, b. Hence the S-divergence encodes information about how much the arithmetic mean of two elements deviates from their geometric mean.

Proposition 3.1.2. *The determinant* $\det : \mathring{\mathcal{K}} \to \mathbb{R}$ *is monotone with respect to the partial order induced by $\mathcal{K}$.*

Proof. Given $x \preceq_{\mathcal{K}} y \in \mathring{\mathcal{K}}$, suppose x, y operator-commute, then they share a common Jordan frame C so we have $\lambda(x) \leq \lambda(y)$ entrywise. Therefore, we have $\det x \leq \det y$. If x, y are arbitrary, then there exists a $d \in \mathring{\mathcal{K}}$ such that $P_d(x) = a$ and $P_d(y) = b \in \mathring{\mathcal{K}}$ where

a, b operator-commute. Since P_d is an order-isomorphism, we have $a \preceq_{\mathcal{K}} b$ and therefore $\det(a) \leq \det(b)$. Thus we have

$$\det(P_d(x)) \leq \det(P_d(y))$$
$$\det(d)^2 \det(x) \leq \det(d)^2 \det(y)$$
$$\det x \leq \det y,$$

since $\det d \neq 0$. Hence det is monotone. $\square$

Corollary 3.1.3. *The function* $\log\det : \mathring{\mathcal{K}} \to \mathbb{R}$ *is monotone.*

Proof. We know log is monotone increasing so the composition is also monotone. $\square$

Corollary 3.1.4. *For any* $x, y \in \mathring{\mathcal{K}}$, *we have* $\delta(x, y) \geq 0$.

Proof. By of Theorem 2.8 of [Lim00], $x\#y \preceq_{\mathcal{K}} \frac{x+y}{2}$. Then monotonicity of det yields

$$\frac{\det\left(\frac{x+y}{2}\right)}{\sqrt{\det x \det y}} = \frac{\det\left(\frac{x+y}{2}\right)}{\det(x\#y)}$$
$$\geq 1.$$

Thus δ^2 as the composition with the logarithmic function is always non-negative, and so is δ. $\square$

This allows us to treat squaring as a monotone function and prove results about δ using δ^2. Next, we tackle the most difficult part the proof, the triangle inequality. It requires a series of technical results:

Lemma 3.1.5. *Let* $a \in \mathring{\mathcal{K}}$ *and* C *be any Jordan frame of* V. *Then we have*

$$\delta(1_V, a) = \delta(1_V, \lambda_C(a)).$$

Proof. There is nothing mysterious here because δ only depends on the eigenvalues so it is invariant under replacing the Jordan frame. Recall that all elements operator-commute with

1_V and share a Jordan frame. Given $a \in \mathring{\mathcal{K}}$ so its eigenvalues are all strictly positive, we have the spectral decompositions $1_V = \sum_{i=1}^r e_i$ and $a = \sum_{i=1}^r \lambda_i e_i$ with respect to a common Jordan frame $\{e_1, \ldots, e_r\}$. Then we have

$$\begin{aligned}
\delta^2(1_V, a) &= \log \frac{\det\left(\frac{1_V + a}{2}\right)}{\sqrt{\det(1_V)\det(a)}} \\
&= \log \frac{\det\left(\frac{\sum_{i=1}^r (1+\lambda_i)e_i}{2}\right)}{\sqrt{\det(1_V)\det(a)}} \\
&= \log \frac{\prod_{i=1}^r \left(\frac{1+\lambda_i}{2}\right)}{\sqrt{\det(1_V)\det(a)}} \\
&= \log \frac{\det\left(\frac{\sum_{i=1}^r (1+\lambda_i)c_i}{2}\right)}{\sqrt{\det(1_V)\det(\lambda_C(a))}} \\
&= \delta^2(1_V, \lambda_C(a)).
\end{aligned}$$

□

Proposition 3.1.6. *Let $x, y, d \in \mathring{\mathcal{K}}$, then we have*

$$\delta(P_d(x), P_d(y)) = \delta(x, y).$$

Proof. We compute

$$\begin{aligned}
\delta^2(P_d(x), P_d(y)) &= \log \frac{\det\left(\frac{P_d(x)+P_d(y)}{2}\right)}{\sqrt{\det(P_d(x))\det(P_d(y))}} \\
&= \log \frac{\det\left(P_d\left(\frac{x+y}{2}\right)\right)}{\sqrt{\det(P_d(x))\det(P_d(y))}} && \text{linearity of quadratic rep} \\
&= \log \frac{\det^2(d)\det\left(\frac{x+y}{2}\right)}{\sqrt{\det^2(d)\det(x)\det^2(d)\det(y)}} && \text{Proposition 2.3.13} \\
&= \log \frac{\det\left(\frac{x+y}{2}\right)}{\sqrt{\det(x)\det(y)}} \\
&= \delta^2(x, y).
\end{aligned}$$

□

We also need some inequality results. The next lemma is Corollary 3.5 of [Sra15].

Lemma 3.1.7. *Define δ_s to be the scalar version of δ_S. That is, $\delta_s(x,y) := \sqrt{\log((x+y)/(2\sqrt{xy}))}$. Let $x, y, z \in \mathbb{R}^n_{++}$ and let $p \geq 1$ be an integer. Then we have*

$$\left(\sum_{i=1}^{n} \delta_s^p(x_i, y_i)\right)^{1/p} \leq \left(\sum_{i=1}^{n} \delta_s^p(x_i, z_i)\right)^{1/p} + \left(\sum_{i=1}^{n} \delta_s^p(y_i, z_i)\right)^{1/p}.$$

Proposition 3.1.8. *Suppose $x, y, z \in \mathring{\mathcal{K}}$ share the same Jordan frame $\{c_1, \ldots, c_r\}$. Then we have*

$$\delta(x,y) \leq \delta(x,z) + \delta(y,z).$$

Proof. By the full EJA spectral theorem, since x, y, z share the same Jordan frame, we have the following decompositions:

$$\begin{aligned} x &= \lambda_1 c_1 + \cdots + \lambda_r c_r, \\ y &= \mu_1 c_1 + \cdots + \mu_r c_r, \\ \text{and} \quad z &= \nu_1 c_1 + \cdots + \nu_r c_r. \end{aligned}$$

It follows that

$$\begin{aligned} \delta^2(x,y) &= \log \frac{\left(\frac{\lambda_1+\mu_1}{2}\right)\cdots\left(\frac{\lambda_n+\mu_n}{2}\right)}{\sqrt{\lambda_1\cdots\lambda_n\mu_1\cdots\mu_n}} \\ &= \sum_{i=1}^{n} \log \frac{\lambda_i+\mu_i}{2\sqrt{\lambda_i\mu_i}} \\ &= \sum_{i=1}^{n} \delta_s^2(\lambda_i, \mu_i), \end{aligned}$$

and likewise for the other two terms. Setting $p = 2$, the desired inequality follows from the previous lemma. □

The next theorem is Corollary 3.6.1 from Baes's PhD thesis [Bae06]. It is a generalization of Lidskii's inequalities from operator theory:

Theorem 3.1.9 (Generalized Lidskii's inequalities)**.** *Fix a positive integer $k \leq r$ and let $1 \leq i_1 \leq \ldots \leq i_k \leq r$. Then for every $u, v \in V$, we have*

$$\sum_{j=1}^{k} \lambda_{i_j}(u) + \sum_{j=1}^{k} \lambda_j(v-u) \geq \sum_{j=1}^{k} \lambda_{i_j}(v) \geq \sum_{j=1}^{k} \lambda_{i_j}(u) + \sum_{j=1}^{k} \lambda_{r-j+1}(v-u).$$

This allows us to generalize Fiedler's inequalities for positive definite matrices to symmetric cones:

Theorem 3.1.10 (Generalized Fiedler's inequalities)**.** *Let V be an EJA with rank r and let $x, y \in \mathring{\mathcal{K}}$. Then we have*

$$\prod_{i=1}^{r}(\lambda_i(x)+\lambda_i(y)) \le \det(x+y) \le \prod_{i=1}^{r}\left(\lambda_i(x)+\lambda_i^{\uparrow}(y)\right).$$

We need to introduce some technical definitions from [Bha13] for the proof:

Definition 3.1.11. Let $x, y \in \mathbb{R}^n$. We say that x is **majorized** by y, denoted by $x \prec y$, if

$$\sum_{j=1}^{k} x_j^{\downarrow} \le \sum_{j=1}^{k} y_j^{\downarrow}, \qquad \forall\, 1 \le k \le n,$$

and

$$\sum_{j=1}^{n} x_j^{\downarrow} = \sum_{j=1}^{n} y_j^{\downarrow}.$$

Definition 3.1.12. A function $f : \mathbb{R}^n \to \mathbb{R}$ is called **Schur-concave** if

$$x \succ y \implies f(x) \le f(y).$$

Proof of generalized Fiedler's inequalities. Take $x = u$, and $y = v - u$ in Lidskii's inequalities. Since the inequalities hold for every $k \le r$, letting $i_j = j$ for each k yields a majorization result:

$$\lambda(x)+\lambda(y) \succ \lambda(x+y) \succ \lambda(x)+\lambda^{\uparrow}(y),$$

where equality when $k = r$ holds because $\mathrm{tr}(x)+\mathrm{tr}(y) = \mathrm{tr}(x+y)$.

Since all eigenvalues are positive, and the elementary symmetric polynomial function $s_r : \mathbb{R}^r \to \mathbb{R}, (x_1, \ldots, x_r) \mapsto \prod_{i=1}^{r} x_i$ is Schur-concave on $\mathbb{R}^r_+$ (Example II 3.16 of [Bha13]), applying s_r to the majorization above yields:

$$\prod_{i=1}^{r}(\lambda_i(x)+\lambda_i(y)) \le \prod_{i=1}^{r}\lambda_i(x+y) = \det(x+y) \le \prod_{i=1}^{r}\left(\lambda_i(x)+\lambda_i^{\uparrow}(y)\right).$$

□

The above generalization of Fiedler's inequalities is routine but novel as far as we know.

Corollary 3.1.13. *Let $x, y \in \mathring{\mathcal{K}}$ and $C = \{c_1, \dots, c_r\}$ be a Jordan frame of x. Then we have*

$$\delta\left(\lambda_C(x), \lambda_C(y)\right) \leq \delta(x, y) \leq \delta(\lambda_C(x), \lambda_C^{\uparrow}(y)).$$

Proof. Dividing Fiedler's inequalities by $\sqrt{\det x \det y}$ which must be a positive real number, we obtain

$$\frac{\prod_{i=1}^{r}\left(\lambda_i\left(\frac{x}{2}\right)+\lambda_i\left(\frac{y}{2}\right)\right)}{\sqrt{\det x \det y}} \leq \frac{\det\left(\frac{x+y}{2}\right)}{\sqrt{\det x \det y}} \leq \frac{\prod_{i=1}^{r}\left(\lambda_i\left(\frac{x}{2}\right)+\lambda_i^{\uparrow}\left(\frac{y}{2}\right)\right)}{\sqrt{\det x \det y}}$$

$$\frac{\det\left(\frac{\lambda_C(x)+\lambda_C(y)}{2}\right)}{\sqrt{\det \lambda_C(x) \det \lambda_C(y)}} \leq \delta^2(x, y) \leq \frac{\det\left(\frac{\lambda_C(x)+\lambda_C^{\uparrow}(y)}{2}\right)}{\sqrt{\det \lambda_C(x) \det \lambda_C^{\uparrow}(y)}}$$

$$\delta^2\left(\lambda_C(x), \lambda_C(y)\right) \leq \delta^2(x, y) \leq \delta^2(\lambda_C(x), \lambda_C^{\uparrow}(y))$$

$$\delta\left(\lambda_C(x), \lambda_C(y)\right) \leq \delta(x, y) \leq \delta(\lambda_C(x), \lambda_C^{\uparrow}(y)).$$

□

The next lemma is Lemma 1 from [GT11]. It allows us to generalize statements that are true for elements that operator-commute to arbitrary elements in $\mathring{\mathcal{K}}$. It is a standard trick in operator theory to prove statements for the easier case of operator-commuting elements first and apply this lemma to prove the general case. It is instrumental for several proofs throughout this chapter.

Lemma 3.1.14. *If $x, y \in \mathcal{K}$, then there exists $d \in \mathring{\mathcal{K}}$ such that $P_d(x) = a$ and $P_d(y) = b$ where $a, b \in \mathcal{K}$ operator-commute. If $x \in \mathring{\mathcal{K}}$, then we can take $a = 1_V$.*

Note that if $y \in \mathring{\mathcal{K}}$ above, then by the order-isomorphism of P_d, we have $b \in \mathring{\mathcal{K}}$ as well. Now we have all the results needed to prove the main theorem of this section:

Theorem 3.1.15. *The square root of S-divergence, δ, is a metric on $\mathring{\mathcal{K}}$.*

Proof. Symmetry is true by construction of δ. Corollary 3.1.4 shows δ is nonnegative. Since $-\log\det$ is strictly convex, the Bregman divergence induced by $-\log\det$ is positive definite. It follows that the symmetrized version, δ^2, is also positive definite. Hence δ is positive definite.

It remains to prove the triangle inequality. Given $x, y, z \in \mathring{\mathcal{K}}$, by Lemma 3.1.14 there exists $d \in \mathring{\mathcal{K}}$ such that $1_V = P_d(x)$ and $a = P_d(y)$. Let $z' := P_d(z)$ and C be the common Jordan frame shared by 1_V and a.

Since $1_V, a$, and $\lambda_C(z')$ share a common Jordan frame C, Corollary 3.1.13 holds:

$$
\begin{aligned}
\delta(x,y) &= \delta(P_d(x), P_d(y)) \\
&= \delta(1_V, a) \\
&\leq \delta(1_V, \lambda_C(z')) + \delta(a, \lambda_C(z')) && \text{Corollary 3.1.13} \\
&\leq \delta(1_V, z') + \delta(a, z') && \text{left Fiedler inequality} \\
&= \delta(P_d(x), P_d(z)) + \delta(P_d(y), P_d(z)) \\
&= \delta(x,z) + \delta(y,z).
\end{aligned}
$$

Hence δ is indeed a metric on $\mathring{\mathcal{K}}$. $\square$

3.2 The geometric mean and g-convexity of δ^2

If we wish to use δ as a numerical alternative to the Riemannian distance, then it would be important for δ to enjoy similar properties. The most important property for the purpose of optimization is that δ^2 needs to be g-convex. That is what we aim to prove in this section.

However, Sra's proofs used many operator theory results that are not readily available for symmetric cones. Below we present a few results that help us overcome this challenge.

Definition 3.2.1. For a Jordan algebra V and $x, y \in V$, the **polarization of quadratic**

representation is defined as

$$P_{(x,y)} := \frac{1}{2}\left(P_{x+y} - P_x - P_y\right).$$

As a side note, this polarization plays an important role in the study of quadratic Jordan algebras.

Proposition 3.2.2. *Let $x, y \in \mathring{\mathcal{K}}$, then the polarization of the quadratic representation $P_{(x,y)}$ is positive definite.*

We are grateful for the help we received from Dr. Muddappa Gowda for this proof.

Proof. Let us first consider the easier case where x, y operator-commute. Then by Lemma 2.3.21 there exists a common Jordan frame $\{c_1, \ldots, c_r\}$ of x and y such that $x = \sum_{i=1}^r \lambda_i c_i$ and $y = \sum_{i=1}^r \mu_i c_i$, where the eigenvalues are all positive. It follows that $x + y = \sum_{i=1}^r (\lambda_i + \mu_i) c_i$. Then for any nonzero $z \in V$, according to Proposition 1 from [GT11] we have the following three equations:

$$\begin{aligned}
\langle P_{x+y}(z), z\rangle &= \sum_{i \leq j} (\lambda_i + \mu_i)(\lambda_j + \mu_j)\langle z_{ij}, z_{ij}\rangle, \\
\langle P_x(z), z\rangle &= \sum_{i \leq j} \lambda_i \lambda_j \langle z_{ij}, z_{ij}\rangle, \\
\langle P_y(z), z\rangle &= \sum_{i \leq j} \mu_i \mu_j \langle z_{ij}, z_{ij}\rangle.
\end{aligned}$$

By the bilinearity of the inner product, we subtract the first equation by the other two and obtain

$$\begin{aligned}
\langle (P_{x+y} - P_x - P_y)(z), z\rangle &= \sum_{i \leq j} (\lambda_i\lambda_j + \lambda_i\mu_i + \lambda_j\mu_i + \mu_i\mu_j - \lambda_i\lambda_j - \mu_i\mu_j)\langle z_{ij}, z_{ij}\rangle \\
&= \sum_{i \leq j} (\lambda_i\mu_j + \lambda_j\mu_i)\langle z_{ij}, z_{ij}\rangle \\
&> 0,
\end{aligned}$$

since each $\lambda_i\mu_j + \lambda_j\mu_i > 0$ and there exists a $\langle z_{ij}, z_{ij}\rangle > 0$ because $z \neq 0$. That is, we have $P_{(x,y)} > 0$.

Now suppose that $x, y \in \mathring{\mathcal{K}}$ are arbitrary. Then by Lemma 3.1.14, there exists $d \in \mathring{\mathcal{K}}$ such that $P_d(x) = a$ and $P_d(y) = b$ and $a, b \in \mathring{\mathcal{K}}$ operator-commute. Thus by the previous case, $P_{(a,b)} > 0$ holds. That is, we have

$$\begin{aligned} P_{P_d(x+y)} - P_{P_d(x)} - P_{P_d(y)} &> 0 \\ P_d P_{x+y} P_d - P_d P_x P_d - P_d P_y P_d &> 0 && \text{fundamental identity} \\ P_d(P_{x+y} - P_x - P_y)P_d &> 0 \\ P_{P_d}(P_{(x,y)}) &> 0. \end{aligned}$$

Since $d \in \mathring{\mathcal{K}}$, we have $P_d > 0$, so P_{P_d} is an order-isomorphism and thus $P_{(x,y)} > 0$, completing the proof. □

The next proposition is Proposition 2.4 of [Lim00].

Proposition 3.2.3. *If $a, b \in \mathring{\mathcal{K}}$, then the Riccati equation*

$$P_x(a^{-1}) = b$$

has a unique solution $x = a\#b \in \mathring{\mathcal{K}}$.

In Sra's proof of the upcoming theorem, a parallel sum result $(A^{-1} + B^{-1})^{-1} = A(A + B)^{-1}B$ for invertible matrices A, B was used. This does not generalize nicely to symmetric cones due to the lack of associativity of multiplication. The next proposition from Proposition 3.8 of McCrimmon [McC78] offers a workaround for this problem.

Proposition 3.2.4 (Hua's identity)**.** *Let V be a Jordan algebra. If $a, b, a - b \in V$ are invertible, then we have*

$$a^{-1} = (a - b)^{-1} + P_{a^{-1}}(a^{-1} - b^{-1})^{-1}.$$

The proof idea of the next theorem is heavily inspired by that of Sra, but generalizing the details to symmetric cones is not trivial and the proof is novel.

Theorem 3.2.5. *Let $a, b \in \mathring{\mathcal{K}}$, then we have*

$$a\#b = \arg\min_{x \in \mathring{\mathcal{K}}} h(x),$$

where

$$h(x) = \delta^2(x, a) + \delta^2(x, b).$$

Moreover, $a\#b$ is equidistant from a and b. That is,

$$\delta(a\#b, a) = \delta(a\#b, b).$$

Proof. We need to establish an identity first. Let $y := x + a$, then we have $x = y - a$. Since x, a, and $x + a$ are elements of $\mathring{\mathcal{K}}$ and are therefore invertible, we can apply Hua's identity to obtain

$$\begin{aligned}(x+a)^{-1} = y^{-1} &= (y-a)^{-1} + P_{y^{-1}}(y^{-1} - a^{-1})^{-1} \\ &= x^{-1} + P_{(x+a)^{-1}}\left((x+a)^{-1} - a^{-1}\right)^{-1}.\end{aligned}$$

Hence, using the fact that a and a^{-1} operator-commute, we have

$$
\begin{aligned}
\left(x^{-1}-(x+a)^{-1}\right)^{-1} &= \left(x^{-1}-\left(x^{-1}+P_{(x+a)^{-1}}\left((x+a)^{-1}-a^{-1}\right)^{-1}\right)\right)^{-1} \\
&= -\left(P_{(x+a)^{-1}}\left((x+a)^{-1}-a^{-1}\right)^{-1}\right)^{-1} \\
&= -P_{x+a}\left((x+a)^{-1}-a^{-1}\right) && \text{Proposition 2.2.14} \\
&= P_{x+a}(a^{-1})-(x+a) \\
&= 2(x+a)\circ\left((x+a)\circ a^{-1}\right)-(x+a)^2\circ a^{-1}-(x+a) \\
&= 2(x+a)\circ(x\circ a^{-1}+1_V)-(x^2+2a\circ x+a^2)\circ a^{-1}-(x+a) \\
&= 2x\circ(x\circ a^{-1})+2a\circ(a^{-1}\circ x)+2x+2a-x^2\circ a^{-1} \\
&\quad -2a^{-1}\circ(a\circ x)-a-x-a \\
&= (2L_x^2(a^{-1})-L_{x^2}(a^{-1}))+2(L_aL_{a^{-1}}(x)-L_{a^{-1}}L_a(x))+x \\
&= P_x(a^{-1})+x.
\end{aligned}
$$

Now we begin proof proper. To find the minimizer, we can find all the critical points of h by setting the gradient to zero. Recall that $\nabla-\log\det(x)=-x^{-1}$, so we obtain

$$
\begin{aligned}
\nabla h(x)=\frac{1}{2}\left(\frac{x+a}{2}\right)^{-1}+\frac{1}{2}\left(\frac{x+b}{2}\right)^{-1}-x^{-1} &= 0 \\
x^{-1}-(x+a)^{-1} &= (x+b)^{-1} \\
\left(x^{-1}-(x+a)^{-1}\right)^{-1} &= x+b \\
P_x(a^{-1})+x &= x+b \\
P_x(a^{-1}) &= b.
\end{aligned}
$$

By Proposition 3.2.3, we know that this is the Riccati equation and has the geometric mean $x_0 := a\#b$ as the unique solution. Since $x_0^{-1}=(x_0+a)^{-1}+(x_0+b)^{-1}$ holds, we have

$$
\begin{aligned}
\nabla^2 h(x_0) &= P_{x_0^{-1}}-(P_{(x_0+a)^{-1}}+P_{(x_0+b)^{-1}}) \\
&= P_{(x_0+a)^{-1}+(x_0+b)^{-1}}-(P_{(x_0+a)^{-1}}+P_{(x_0+b)^{-1}}) \\
&> 0.
\end{aligned}
$$

Hence $a\#b$ is a local minimum of h. Since it is the unique solution of $\nabla h(x) = 0$, there is no other local minima. It remains to check the boundary. But as x approaches $\partial\mathcal{K}$, $\det x \to 0$. Thus in the expression of $\delta^2(x,a)$, the denominator approaches 0 but the numerator is lower-bounded by the positive constant $\det\left(\frac{a}{2}\right)$ by Corollary 2 of [GT11]. This forces $h(x)$ to approach ∞. Therefore, $a\#b$ must be the unique global minimum of h.

Since $b\#a = a\#b$ is the midpoint of the geodesic, we have $P_{a\#b}(a^{-1}) = b$ and $P_{a\#b}(b^{-1}) = P_{b\#a}(b^{-1}) = a$. Then by Proposition 3.1.6, we have

$$\begin{aligned}\delta(a^{-1}, b^{-1}) &= \delta\left(P_{a\#b}(a^{-1}), P_{a\#b}(b^{-1})\right)\\ &= \delta(b,a)\\ &= \delta(a,b).\end{aligned}$$

It follows that

$$\begin{aligned}\delta(a, a\#b) &= \delta\left(a^{-1}, (a\#b)^{-1}\right)\\ &= \delta\left(P_{a\#b}(a^{-1}), P_{a\#b}(a\#b)^{-1})\right)\\ &= \delta(b, a\#b).\end{aligned}$$

□

We collect a result from the above proof as a proposition:

Proposition 3.2.6. *For $a, b \in \mathring{\mathcal{K}}$, we have*

$$\delta(a^{-1}, b^{-1}) = \delta(a,b).$$

Now we state a conjecture that has already been proven true for four out of five simple EJA classes [CCP16], with the Albert algebra being the only remaining case to prove. We shall discuss this conjecture further in the next chapter.

Conjecture 3.2.7. *For $t \in [0,1]$, the power map $\phi : [0,\infty) \to [0,\infty), a \mapsto a^t$ is SC-concave. That is, for any irreducible symmetric cone $\mathcal{K}$ and $x, y \in \mathring{\mathcal{K}}$, we have*

$$\forall\ s \in [0,1], \quad (1-s)x^t + sy^t \leq \left((1-s)x + sy\right)^t.$$

We present the proven part of this conjecture as a lemma:

Lemma 3.2.8. *Suppose V is a simple EJA not isomorphic to the Albert algebra, then for $t \in [0,1]$, the power map $\phi : [0,\infty) \to [0,\infty), a \mapsto a^t$ is SC-concave.*

From this point, all EJAs are assumed to be simple EJAs not isomorphic to the Albert algebra. Suppose the conjecture is true, then the following results immediately generalize to all simple EJAs.

Proposition 3.2.9. *Let V be a simple EJA not isomorphic to the Albert algebra. Let $x, y \in \mathring{\mathcal{K}}$, $t \in [0,1]$. Then $\delta^2(x^t, y^t) \leq t\delta^2(x,y)$.*

Proof. By the lemma above, we have

$$\frac{1}{2}(x^t + y^t) \leq \left(\frac{x+y}{2}\right)^t.$$

Then by monotonicity of $\log\det$, we have

$$\log\det\left(\frac{1}{2}(x^t + y^t)\right) \leq \log\det\left(\frac{x+y}{2}\right)^t$$

$$\log\det\left(\frac{1}{2}(x^t + y^t)\right) - \frac{1}{2}\log(\det(x^t)\det(y^t)) \leq \log\det\left(\frac{x+y}{2}\right)^t - \frac{1}{2}\log(\det(x)^t\det(y)^t)$$

$$\delta^2(x^t, y^t) \leq t\delta^2(x,y).$$

□

Theorem 3.2.10. *Let V be a simple EJA not isomorphic to the Albert algebra. The S-divergence $\delta^2(x,y)$ is jointly g-convex for $x, y \in \mathring{\mathcal{K}}$.*

Proof. A function $f : \mathring{\mathcal{K}} \times \mathring{\mathcal{K}} \to \mathbb{R}$ is jointly g-convex if and only if the composition $f \circ c$ is convex for $c : [0,1] \to \mathring{\mathcal{K}} \times \mathring{\mathcal{K}}$. It suffices to show midpoint convexity for $f \circ c$ by Lemma 2.1.2. Suppose $(x_1, y_1), (x_2, y_2)$ are two arbitrary points in $\mathring{\mathcal{K}} \times \mathring{\mathcal{K}}$, let c be the geodesic of the product cone such that $c(0) = (x_1, y_1)$ and $c(1) = (x_2, y_2)$. Then $\delta^2 \circ c\left(\frac{1}{2}\right)$ is equal to $\delta^2(x_1 \# x_2, y_1 \# y_2)$. Thus midpoint convexity requires us to show:

$$\delta^2(x_1 \# x_2, y_1 \# y_2) \leq \frac{1}{2}\delta^2(x_1, y_1) + \frac{1}{2}\delta^2(x_2, y_2).$$

First, we show that the result holds when the starting point of the geodesic is on the diagonal. Let $u, v_1, v_2 \in \mathring{\mathcal{K}}$ and let $c(0) = (u, u)$. Since $u \in \mathring{\mathcal{K}}$, define $d := u^{-1/2}$ so we have $P_d(u) = 1_V$, $P_d(v_1) = a_1$, and $P_d(v_2) = a_2$ for some $a_1, a_2 \in \mathring{\mathcal{K}}$. Then for any $v \in \mathring{\mathcal{K}}$, we have

$$\begin{aligned}
P_d(u\#v) &= P_{u^{-1/2}}\left(P_{u^{1/2}}\left(P_{u^{-1/2}}(v)\right)^{1/2}\right) \\
&= \left(P_{u^{1/2}}^{-1}P_{u^{1/2}}\right)\left(P_{u^{-1/2}}(v)\right)^{1/2} \\
&= \left(P_d(v)\right)^{1/2} \\
&= 1_V\#P_d(v). && \text{Corollary 2.5.3}
\end{aligned}$$

Therefore, we obtain

$$\begin{aligned}
\delta^2(u\#v_1, u\#v_2) &= \delta^2(P_d(u\#v_1), P_d(u\#v_2)) \\
&= \delta^2(1_V\#P_d(v_1), 1_V\#P_d(v_2)) \\
&= \delta^2(1_V\#a_1, 1_V\#a_2) \\
&= \delta^2\left(a_1^{1/2}, a_2^{1/2}\right) \\
&\leq \frac{1}{2}\delta^2(a_1, a_2) \\
&= \frac{1}{2}\delta^2(P_d(v_1), P_d(v_2)) \\
&= \frac{1}{2}\delta^2(v_1, v_2).
\end{aligned}$$

The triangle inequality yields

$$\begin{aligned}
\delta^2(x_1\#x_2, y_1\#y_2) &\leq \delta^2(y_1\#y_2, x_1\#y_2) + \delta^2(x_1\#x_2, x_1\#y_2) \\
&\leq \frac{1}{2}\delta^2(x_1, y_1) + \frac{1}{2}\delta^2(x_2, y_2).
\end{aligned}$$

$\square$

Chapter 4

Discussion

4.1 Conjecture

We believe Conjecture 3.2.7 is likely true and thus worth proving for three reasons.

First, Corollary 3.1 of [CCP16] implies that the fractional power function ϕ is SC-concave for four out of five cone of squares of simple EJAs, with only the Albert algebra left out. For optimization purposes, one can argue that this is sufficient since we are likely to never encounter objects isomorphic to $\mathbb{O}^{3\times 3}$ in optimization. From a theoretical perspective, this is unsatisfying.

Second, from an abstract perspective we are able to show that the conjecture is true for operator-commuting elements.

Lemma 4.1.1. *For $t \in [0,1]$, the power map $\phi : [0,\infty) \to [0,\infty), a \mapsto a^t$ is SC-concave. That is, for any irreducible symmetric cone $\mathcal{K}$ and operator-commuting elements $x, y \in \mathring{\mathcal{K}}$, we have*

$$\forall\ s \in [0,1], \quad (1-s)x^t + sy^t \leq ((1-s)x + sy)^t\,.$$

Proof. Suppose $x, y \in \mathring{\mathcal{K}}$ operator-commute. Recall from the equivalence of norm inequalities that for $a, b \in \mathbb{R}$, $|a| + |b| \leq \sqrt{2}\sqrt{a^2 + b^2}$. Since $\mathcal{K}$ is the cone of squares of V, every element in $\mathring{\mathcal{K}}$ can be written as a square of an invertible element in V. That is, there exists $z, w \in V$ such that $x = z^2$ and $y = w^2$. Since x, y share the same Jordan frame $\{c_1, \ldots, c_r\}$, z, w share the same Jordan frame. We obtain the necessary inequality via the following computation:

$$\begin{aligned}
\delta^2(z,w) &= \log \frac{\det\left(\frac{z+w}{2}\right)}{\sqrt{\det z \det w}} \\
&= \log \frac{\det\left(\sum_{i=1}^r \frac{\lambda_i+\mu_i}{2} c_i\right)}{\sqrt{\det z \det w}} \\
&= \log \frac{\prod_{i=1}^r \frac{\lambda_i+\mu_i}{2}}{\sqrt{\det z \det w}} \\
&\leq \log \frac{\prod_{i=1}^r \sqrt{2}/2\sqrt{\lambda_i^2+\mu_i^2}}{\sqrt{\det z \det w}} \\
&= \log \frac{\prod_{i=1}^r \sqrt{\frac{\lambda_i^2+\mu_i^2}{2}}}{\sqrt{\det z \det w}} \\
&= \frac{1}{2} \log \frac{\prod_{i=1}^r \left(\frac{\lambda_i^2+\mu_i^2}{2}\right)}{\sqrt{\det(z^2)\det(w^2)}} \\
&= \frac{1}{2} \log \frac{\det\left(\sum_{i=1}^r \frac{z^2+w^2}{2} c_i\right)}{\sqrt{\det(z^2)\det(w^2)}} \\
&= \frac{1}{2}\delta^2(z^2,w^2).
\end{aligned}$$

A substitution yields $\delta^2(x^{1/2},y^{1/2}) \leq \frac{1}{2}\delta^2(x,y)$. Midpoint concavity and continuity of δ^2 implies concavity.

□

However, we have trouble generalizing this result to non-operator-commuting elements. Here we describe our failed attempts so others can save some time.

Suppose $x, y \in \mathring{\mathcal{K}}$ are arbitrary, there exist invertible z, w such that $x = z^2$ and $y = w^2$. There exists $d \in \mathring{\mathcal{K}}$ such that $P_d(z) = 1_V$ and $P_d(w) = a \in \mathring{\mathcal{K}}$. Then we have

$$\begin{aligned}
\delta^2(z,w) &= \delta^2(P_d(z), P_d(w)) \\
&= \delta^2(1_V, a) \\
&\leq \frac{1}{2}\delta^2(1_V^2, a^2) \\
&= \frac{1}{2}\delta^2((P_d(z))^2, (P_d(w))^2).
\end{aligned}$$

Our goal is to get to $\frac{1}{2}\delta^2(z^2, w^2)$. However, since in general $(P_d(z))^2 \neq P_d(z^2)$, we do not have the result immediately. But notice that we do not need equality; we simply need the determinants of their arithmetic and geometric means to give the same ratio. Since squaring a positive element does not change its Jordan frame nor its spectral ordering, it suffices to show that they have the same eigenvalues. That is, we wish to show that they are similar: $(P_d(z))^2 \sim P_d(z^2)$. We know from Proposition 3.2.3 of [Vie07] that $P_d(z^2) \sim P_z(d^2)$. However, in the matrix case, clearly we have $(P_D(Z))^2 = DZD^2ZD \not\sim ZD^2Z = P_Z(D^2) \sim P_D(Z^2)$, since $Z = 1, D = 2$ is a one-dimensional counterexample. Therefore, we cannot proceed to the final step using this approach.

Third, we already know that ϕ is SC-monotone from Corollary 9 of [Lim01]. In operator theory, knowing that ϕ is operator monotone is enough to prove it is also operator concave. The argument goes as follows:

Definition 4.1.2. A function f is **matrix monotone of order n** if it is monotone with respect to the partial order on $n \times n$ Hermitian matrices. We say f is **operator monotone** if f is matrix monotone of order n for all n.

Definition 4.1.3. A function f is **matrix concave of order n** if for all $n \times n$ Hermitian matrices A and B and for all real numbers $t \in [0, 1]$,

$$f((1-t)A + tB) \geq (1-t)f(A) + tf(B).$$

We say f is **operator concave** if f is matrix concave of order n for all n.

The following three results are from Bhatia [Bha13]. Next is Theorem 2.5:

Theorem 4.1.4. *Let f be a continuous function mapping $[0, \infty)$ into itself. Then f is operator monotone if and only if it is operator concave.*

Proposition 4.1.5. *The function $f(X) = X^t$ is operator monotone on $[0, \infty)$ for $t \in [0, 1]$.*

Corollary 4.1.6. *The power function $f(X) = X^t$ is operator concave for $t \in [0, 1]$.*

It would not be surprising if we can generalize this line of argument to symmetric cones. Specifically, generalizing Theorem 4.1.4 is sufficient to prove the conjecture. However, this would be completely out of the scope of this thesis. We encourage operator theory enthusiasts to look into this generalization.

4.2 Product cones

In Chapter 3, we restricted our analysis to irreducible symmetric cones only. An immediate question is whether our results apply to arbitrary symmetric cones using Proposition 2.4.7.

First, we consider the possibility of generalizing our results to a metric defined on a product cone. For simplicity, we present the case when the cone is the product of two irreducible symmetric cones, *i.e.* $\mathcal{K} = \mathcal{K}_1 \times \mathcal{K}_2$, and let δ_1 and δ_2 be their δ metrics respectively. We see that $\mathring{\mathcal{K}} = \mathring{\mathcal{K}}_1 \times \mathring{\mathcal{K}}_2$. Using $B(x, y) = -(\log\det x + \log\det y)$ from Example 2.4.11 as the seed function for the Jensen-Shannon divergence, the S-divergence becomes $\Delta : \mathring{\mathcal{K}} \times \mathring{\mathcal{K}} \to \mathbb{R}$:

$$\Delta((x_1, x_2), (y_1, y_2)) := \sqrt{\delta_1^2(x_1, y_1) + \delta_2^2(x_2, y_2)}.$$

Theorem 4.2.1. *The square root of the S-divergence, Δ, is a metric on $\mathring{\mathcal{K}}$.*

Proof. By triangle inequality, we have

$$\begin{aligned}\delta_1^2(x_1, y_1) &\le (\delta_1(x_1, z_1) + \delta_1(y_1, z_1))^2 \\ &\le \delta_1^2(x_1, z_1) + \delta_1^2(y_1, z_1) + 2\delta_1(x_1, z_1)\delta_1(y_1, z_1).\end{aligned}$$

Similarly, we have

$$\delta_2^2(x_2, y_2) \le \delta_2^2(x_2, z_2) + \delta_2^2(y_2, z_2) + 2\delta_2(x_2, z_2)\delta_2(y_2, z_2).$$

By the Cauchy–Schwarz inequality, we obtain

$$
\begin{aligned}
\delta_1(x_1,z_1)\delta_1(y_1,z_1)+\delta_2(x_2,z_2)\delta_2(y_2,z_2) &= \begin{pmatrix}\delta_1(x_1,z_1)\\ \delta_2(x_1,z_1)\end{pmatrix}^T \begin{pmatrix}\delta_2(y_2,z_2)\\ \delta_2(y_2,z_2)\end{pmatrix}\\
&\leq \sqrt{\delta_1^2(x_1,z_1)+\delta_2^2(x_2,z_2)}+\sqrt{\delta_1^2(y_1,z_1)+\delta_2^2(y_2,z_2)}.
\end{aligned}
$$

Taken together, this implies that

$$
\begin{aligned}
\Delta^2((x_1,x_2),(y_1,y_2)) &= \delta_1^2(x_1,y_1)+\delta_2^2(x_2,y_2)\\
&= \delta_1^2(x_1,z_1)+\delta_1^2(y_1,z_1)+2\delta_1(x_1,z_1)\delta_1(y_1,z_1)\\
&\quad+\delta_2^2(x_2,z_2)+\delta_2^2(y_2,z_2)+2\delta_2(x_2,z_2)\delta_2(y_2,z_2)\\
&\leq \delta_1^2(x_1,z_1)+\delta_1^2(y_1,z_1)+2\sqrt{\delta_1^2(x_1,z_1)+\delta_2^2(x_2,z_2)}\\
&\quad+\delta_2^2(x_2,z_2)+\delta_2^2(y_2,z_2)+2\sqrt{\delta_1^2(y_1,z_1)+\delta_2^2(y_2,z_2)}\\
&= \delta_1^2(x_1,z_1)+\delta_2^2(x_2,z_2)+\delta_1^2(y_1,z_1)+\delta_2^2(y_2,z_2)\\
&\quad+2\sqrt{\delta_1^2(x_1,z_1)+\delta_2^2(x_2,z_2)}\sqrt{\delta_1^2(y_1,z_1)+\delta_2^2(y_2,z_2)}\\
&= (\Delta((x_1,x_2),(z_1,z_2))+\Delta((y_1,y_2),(z_1,z_2)))^2.
\end{aligned}
$$

Since non-negativity, positive definiteness, and symmetry are inherited from δ_1 and δ_2, we conclude that Δ is a metric on the product. □

Theorem 4.2.2. *The S-divergence Δ^2 is g-convex.*

Proof. This result follows from g-convexity of δ_1^2 and δ_2^2. We compute

$$
\begin{aligned}
\Delta^2((x_1,x_1')\#(x_2,x_2'),(y_1,y_1')\#(y_2,y_2')) &= \Delta^2((x_1\#x_2,x_1'\#x_2'),(y_1\#y_2,y_1'\#y_2'))\\
&= \delta^2(x_1\#x_2,y_1\#y_2)+\delta^2(x_1'\#x_2',y_1'\#y_2')\\
&\leq \frac{1}{2}\delta^2(x_1,y_1)+\frac{1}{2}\delta^2(x_2,y_2)+\frac{1}{2}\delta^2(x_1',y_1')+\frac{1}{2}\delta^2(x_2',y_2')\\
&= \frac{1}{2}(\delta^2(x_1,y_1)+\delta^2(x_1',y_1'))+\frac{1}{2}(\delta^2(x_2,y_2)+\delta^2(x_2',y_2'))\\
&= \frac{1}{2}\Delta^2((x_1,x_1'),(y_1,y_1'))+\frac{1}{2}\Delta^2((x_2,x_2'),(y_2,y_2')).
\end{aligned}
$$

□

Notice that this proof would not generalize trivially if we have a product of three or more cones. In that case, we can define a product metric two-at-a-time recursively to preserve g-convexity. However, the complexity might erase any benefit from using a metric on the product.

In fact, notice that we do not need a product metric to perform optimization on the product. Given a symmetric cone, it suffices to decompose it into its irreducible components and perform optimization on each individually. Since its components do not interact with each other, this approach is in fact optimal and allows parallelization. Therefore, we conclude that even though a metric on any symmetric cone can be found, our results in Chapter 3 are sufficient for optimization on all symmetric cones.

4.3 Embedding EJAs into real symmetric EJA

It is known that we can embed four out of five simple EJAs into the real symmetric EJA. For example, the Hadamard EJA can be embedded as diagonal matrices, the Jordan spin EJA can be embedded as arrow-shaped matrices, and the hermitian complex and quaternion EJAs can be embedded as bigger real symmetric matrices. The results of [CCP16] were exactly accomplished by embedding these four simple EJAs into symmetric matrices. Since we are unlikely to encounter the Albert algebra in optimization, it is fair to ask whether it is worth the trouble to adopt an EJA framework when we can generalize results to common symmetric cones via embedding.

Our answer is an unequivocal "yes". First, it is not clear that the embedding preserves the geometric structure of the cone. For instance, in the second-order cone $\mathcal{K} \subset \mathbb{R}^{n+1}$, if we set

$$\operatorname{Arw}(x) := \begin{pmatrix} x_0 & \overline{x}^T \\ \overline{x} & x_0 I_n \end{pmatrix},$$

which looks like an arrow when the block matrices are expanded. The multiplication of the second-order cone can be rewritten in terms of symmetric matrices multiplication as

$$x \circ y = \operatorname{Arw}(x) \operatorname{Arw}(y) 1_V.$$

It turns out that the geodesic between x and y indeed coincides with the geodesic between $\operatorname{Arw}(x)$ and $\operatorname{Arw}(y)$ after right multiplying 1_V. But this requires proof, and we might have to prove properties one-by-one just for the second-order cone, and we would need to repeat for other irreducible symmetric cones. Such proofs using embeddings are typically tedious and unenlightening because they provide little insight on the structural properties of the cone itself (see [CCP16] for a taste). If we need to spend lots of effort proving something, we might as well use the natural abstract framework to prove results for every symmetric cone at once while gaining insights on their structure. After all, if we are content with understanding everything by just piggybacking on existing results ill-fitted for the task, we would be still stuck with the Plotemaic model of the universe, which would obscure most insights about gravity and the nature of our universe.

Second, theory allows us to connect the dots and generalize. If we wish to study important cones that are not necessarily symmetric, such as the power cones or the exponential cones, we can potentially study them by investigating how we can relax symmetric cone theoretical framework to describe these non-symmetric cones. By doing so we could gain lots of insights that we otherwise would not gain just by studying these messy cones individually. Embeddings have their limitations (as is in the case of Albert algebra), and it is unclear whether we can embed all important objects we care about in optimization into well-understood objects like matrices. If embedding ever fails and we do not have enough theory to understand the object on its own, then we would be forced to develop the theory anyway. We do not claim that adopting an abstract framework is essential for applications, but in the long run it should accelerate research progress in applications and is therefore time well-spent.

Third, in practice such as the implementation of primal-dual IPM, we never embed other cones into matrices because it would be a massive waste of computational resources by artificially inflating the dimension of the problem; we instead optimize natively on these cones by using the their idiosyncratic barrier functions. Therefore, embeddings are unnatural even from a practitioner's perspective and can make theory more confusing as opposed to more accessible.

Finally, beyond insights there is also elegance and beauty in this unifying theory. It is worth pursuing for the aesthetics alone.

4.4 Future directions

We outline some potentially fruitful endeavors that we could not pursue due to time constraints.

4.4.1 Replacing Riemannian distance in SCP solvers

For optimization purpose, results from Chapter 3 are sufficient for justifying the use of the metric δ derived from S-divergence as a numerical proxy for the Riemannian distance ρ. An immediate next step is to replace the Riemannian distance used in any SCP solver with this metric and compare the performance. We illustrate their computational difference below.

The Riemannian distance has the following form on a symmetric cone:

$$\rho(x, y) = \left\| \log(P_{y^{-1/2}}(x)) \right\|,$$

where log is the spectral function induced by the usual log function, and $\|\cdot\|$ is induced by the inner product from the EJA. In the case of positive definite cone, this requires matrix multiplication and eigenvalue decomposition. However, to compute the determinant of a positive definite matrix X, we can cheaply perform Cholesky decomposition $X = LL^T$, multiply the diagonal entries of L, and square it since $\operatorname{Det} X = \operatorname{Det}(L) \operatorname{Det}(L^T) = \operatorname{Det}^2(L)$. Depending on the dimension of the matrices involved, this could save a substantial amount of computation. Similarly in the case of second-order cone, recall from Example 2.2.12 that computing the quadratic representation in the Riemannian distance already involves computing the determinant $\det x = x_0^2 - \|\overline{x}\|^2$, so it is clearly more expensive than computing the determinant alone.

In particular, the geodesic IPM proposed by Permenter [Per20] uses Riemannian distance to evaluate convergence. Therefore, if Riemannian distance computation happens to be a bottleneck in Permenter's algorithms, then replacing it with δ should allow us to see significant performance gain. However, it may very well be the case that the Newton's direction is much more expensive to compute than the Riemannian distance. In that case, δ still has the theoretical advantage of being much easier to bound in convergence analysis. Indeed, Permenter used a similar divergence called the *Jeffrey divergence* to bound the Riemannian distance for this reason.

It is possible that as more optimization researchers adopt an EJA framenwork to study SCP in the future, we might have more SCP solvers that substantially rely on the Riemannian distance. Specifically, Then δ, as a computational and theoretically superior choice over the Riemannian distance, can make these solvers more competitive against the primal-dual IPM.

4.4.2 Generalizing operator theory to symmetric cones

From a theoretic perspective, given the abundant parallelism among existing results, we believe that there are a lot more results from operator theory that could be generalize to symmetric cones but no one has done so due to its relatively niche status. There are potentially low-hanging fruit suitable for undergraduate or graduate students interested in operator theory to attempt. The optimization community will be very grateful.

4.4.3 Generalizing linear algebra to EJA

The Gershgorin Circle Theorem and related results have been generalized to EJA by Moldovan's PhD thesis [Mol09]. Many applications of the Gershgorin Theorem, such as diagonally-dominant-sum-of-squares (DSOS) optimization [AM19], can potentially be generalized as well.

In addition, the Jordan canonical form is a powerful tool from linear algebra and can be generalized to EJA or in fact any power-associative algebra. Section VIII.3 of [FK94] has a concise treatment of it, but there might be a lot more results than what the book offers. Moreover, who would not want to see both Camille Jordan and Pascual Jordan honored in the same phrase?

Finally, we are aware that this thesis barely scratches the surface of EJAs and symmetric cones. Therefore, in addition to [FK94] for symmetric cones, we highlight a book, *A Taste of Jordan Algebras* by Kevin McCrimmon [McC04], for readers who wish to understand EJAs from a highly abstract treatment of general Jordan algebras.

Bibliography

[AG03] F. Alizadeh and D. Goldfarb. Second-order cone programming. *Mathematical Programming, Series B*, 95(1), 2003.

[AM19] Amir Ali Ahmadi and Anirudha Majumdar. DSOS and SDSOS optimization: More tractable alternatives to sum of squares and semidefinite optimization. *SIAM Journal on Applied Algebra and Geometry*, 3(2), 2019.

[Bae06] Michel Baes. Spectral functions and smoothing techniques on jordan algebras. *Catholic University of Louvain, PhD Thesis*, 2006.

[Bha13] Rajendra Bhatia. *Matrix analysis*, volume 169. Springer Science & Business Media, 2013.

[Bou20] Nicolas Boumal. *An introduction to optimization on smooth manifolds*, volume 3. Available online, May, 2020.

[BV04] Stephen Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

[CCP16] Yu Lin Chang, Jein Shan Chen, and Shaohua Pan. Symmetric cone monotone functions and symmetric cone convex functions. *Journal of Nonlinear and Convex Analysis*, 17(3), 2016.

[CSBP11] Anoop Cherian, Suvrit Sra, Arindam Banerjee, and Nikolaos Papanikolopoulos. Efficient similarity search for covariance matrices via the Jensen-Bregman LogDet Divergence. In *Proceedings of the IEEE International Conference on Computer Vision*, 2011.

[DF91] David S Dummit and Richard M Foote. *Abstract algebra*, volume 1999. Prentice Hall Englewood Cliffs, NJ, 1991.

[FK94] Jacques Faraut and Adam Koranyi. Analysis on symmetric cones. *Oxford mathematical monographs*, 1994.

[GT11] M. Seetharama Gowda and Jiyuan Tao. Some inequalities involving determinants, eigenvalues, and Schur complements in Euclidean Jordan algebras. *Positivity*, 15(3), 2011.

[Gül96] Osman Güler. Barrier functions in interior point methods. *Mathematics of Operations Research*, 21(4), 1996.

[Lim00] Yongdo Lim. Geometric means on symmetric cones. *Archiv der Mathematik*, 75(1):39–45, 2000.

[Lim01] Yongdo Lim. Applications of geometric means on symmetric cones. *Mathematische Annalen*, 319(3), 2001.

[LL07a] Jimmie Lawson and Yongdo Lim. Metric convexity of symmetric cones. *Osaka Journal of Mathematics*, 44(4):795–816, 2007.

[LL07b] Jimmie Lawson and Yongdo Lim. Symmetric spaces with convex metrics. 2007.

[McC78] Kevin McCrimmon. Jordan algebras and their applications. *Bulletin of the American Mathematical Society*, 84(4), 1978.

[McC04] Kevin McCrimmon. *A taste of Jordan algebras*. Springer, 2004.

[Mol09] Melania M Moldovan. *A Geršgorin type theorem, spectral inequalities, and simultaneous stability in Euclidean Jordan algebras*. University of Maryland, Baltimore County, PhD Thesis, 2009.

[NN94] Yurii Nesterov and Arkadii Nemirovskii. *Interior-Point Polynomial Algorithms in Convex Programming*. 1994.

[NT97] Yu E. Nesterov and M. J. Todd. Self-scaled barriers and interior-point methods for convex programming. *Mathematics of Operations Research*, 22(1), 1997.

[NT98] Yu E. Nesterov and M. J. Todd. Primal-dual interior-point methods for self-scaled cones. *SIAM Journal on Optimization*, 8(2), 1998.

[NT02] Yu E. Nesterov and M. J. Todd. On the Riemannian Geometry Defined by Self-Concordant Barriers and Interior-Point Methods. *Foundations of Computational Mathematics*, 2(4), 2002.

[Orl21] Michael Orlitzky. *Euclidean Jordan algebras for optimization*. Available online, March, 2021.

[Per20] Frank Permenter. A geodesic interior-point method for linear optimization over symmetric cones. *arXiv preprint*, 2020.

[Ren01] James Renegar. *A Mathematical View of Interior-Point Methods in Convex Optimization*. SIAM, 2001.

[Sra15] Suvrit Sra. Positive definite matrices and the S-divergence. *Proceedings of the American Mathematical Society*, 144(7), 2015.

[TWK21] Jiyuan Tao, GQ Wang, and L Kong. The Araki–Lieb–Thirring inequality and the Golden–Thompson inequality in Euclidean Jordan algebras. *Linear and Multilinear Algebra*, pages 1–16, 2021.

[Vie07] Manuel VC Vieira. Jordan algebraic approach to symmetric optimization. *Technische Universiteit Delft, PhD Thesis*, 2007.

[Vis18] Nisheeth K Vishnoi. Geodesic convex optimization: Differentiation on manifolds, geodesics, and convexity. *arXiv preprint*, 2018.