Jordan Algebraic Approach to Symmetric Optimization

Manuel V. C. Vieira

Jordan Algebraic Approach to Symmetric Optimization

PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus Prof. dr. ir. J.T. Fokkema,
voorzitter van het College voor Promoties,
in het openbaar te verdedigen op
dinsdag 13 november 2007 om 12.30 uur
door

Manuel Valdemar Cabral VIEIRA

Grau de mestre em Investigação Operacional
Universidade de Lisboa

geboren te Vila Nova de Foz Côa, Portugal.

Dit proefschrift is goedgekeurd door de promotor:

Prof. dr. ir. C. Roos

Samenstelling promotiecommissie:

Rector Magnificus	voorzitter
Prof. dr. ir. C. Roos	Technische Universiteit Delft, promotor
Prof. dr. F. Glineur	Université Catholique de Louvain
Prof. dr. A. Yoshise	Tsukuba University, Japan
Prof. dr. J.M.A.M. van Neerven	Technische Universiteit Delft
Prof. dr. Y. Bai	Shanghai University, China
Dr. J. Brinkhuis	Erasmus Universiteit Rotterdam
Dr. M. Laurent	Centrum voor Wiskunde en Informatica
Prof. dr. ir. C. Vuik	Technische Universiteit Delft, reservelid

Financial support was provided by the Faculty of Science and Technology of the New University of Lisbon and by the Portuguese Foundation for Science and Technology.

Dit proefschrift kwam tot stand onder auspicien van:

THOMAS STIELTJES INSTITUTE
FOR MATHEMATICS

Copyright © 2007 by M. V. C. Vieira

All rights reserved. No part of the material protected by this copyright notice may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying, recording or by any information storage and retrieval system, without the prior permission of the author.

ISBN 978-90-6464-189-3

Author email: mvcv@fct.unl.nl

To my wife

Acknowledgements

The time had run. If sometimes it seemed impossible to accomplish, now the result comes out and I am completing my PhD.

I would like first to thank my supervisor, Kees Roos, for his availability, support and enriching comments during the writing of the thesis. I am thankful for his extreme patience while reading my thesis.

I would like to thank my sponsors, the Faculty of Science and Technology of the New University of Lisbon and to the Portuguese Foundation for Science and Technology. Without them, this life experience of studying in the Netherlands would not be possible. I also would like to thank my host institution the Technical University of Delft, where I always felt welcome.

Thanks Etienne de Klerk, who friendly hosted me in my first visit to Delft.

I would like to thank François Glineur for his careful reading of my thesis and useful remarks to improve my thesis.

I also would like to thank all my colleagues from the optimization group. I thank my colleagues Ivan and Hossein for their help to handle some affairs with the university, after I left the Netherlands.

Now, I want to thank my family for their support that is crucial for pursuing the goal of doing a PhD. Thanks to all my friends. But most of all I want to apologize to my family and friends of being apart so long. I am sorry.

Of course, I specially thank my wife Élia, for her great support and love, specially during the hard moments.

Delft, November 2007

Manuel Vieira

Contents

List of notations

Sets

$\mathbb{R}$ - the field of real numbers;

$\mathbb{C}$ - the field of complex numbers;

$\mathbb{H}$ - the set of quaternions;

$\mathbb{O}$ - the set of Octonions;

$\mathcal{K}^*$ - the dual of the cone $\mathcal{K}$;

$\mathbf{S}^n$ - the vector space of real symmetric matrices;

$\mathbf{S}^n_+$ - the cone of positive semidefinite matrices;

$\mathbb{R}[x]$ - the subalgebra generated by e and x (see page 8);

$\mathrm{GL}(V)$ - the set of invertible linear mappings from V into itself;

End(V) - the set of endomorphisms of V;

$\mathrm{Aut}(\mathcal{K})$ - the automorphism group, $\{g \in \mathrm{GL}(V) : g(\mathcal{K}) = \mathcal{K}\}$ (see page 38);

$\mathrm{OAut}(\mathcal{K})$ - the set of orthogonal automorphism that leave $\mathcal{K}$ invariant (see page 38);

$\mathcal{K}(V)$ - the cone of squares, $\{x^2 : x \in V\}$ (see page 28);

$\mathcal{I}$ - the set of invertible elements of V (see page 13);

U^0 - the interior of the set U;

$\mathrm{cl}(U)$ - the closure of the set U;

$\mathrm{Aut}(V)$ - the set of automorphisms of V (see page 38);

V_{ij} - the Peirce spaces (see page 43).

Operators

$\circ$ - Jordan product (see page 13);

$L(x)$ - Operator of multiplication by x (see page 9);

$P(x)$ - Quadratic operator (see page 16);

$\langle\cdot,\cdot\rangle$ - inner product;

$\oplus$ - direct sum;

$P_{ii} := P(c_i)$ - the orthogonal projection onto V_{ii};

$P_{ij} := 4L(c_i)L(c_j)$ - the orthogonal projection onto V_{ij};

$[S,T] := ST - TS$ - the commutator of the endomorphisms S and T;

$x \# y$ - geometric mean of x and y (see page 80);

$D_x f(x)$ - the first derivative of f at x (see page 18);

$D_x^u f(x)$ - the derivative of f in the direction u at x (see page 19);

$\nabla f(x)$ - the gradient of f (see page 19);

$D_x^2 f(x)$ the second derivative of f at x (see page 32).

Functions

$\lambda_i(x)$ - an eigenvalue of x;

$\lambda_{\min}(x)$ - the smallest eigenvalue of x;

$\lambda_{\max}(x)$ - the largest eigenvalue of x;

$\det(x)$ - the determinant of x, $\prod_{i=1}^{r} \lambda_i(x)$;

$\operatorname{tr}(x)$ - the trace of x, $\sum_{i=1}^{r} \lambda_i(x)$;

$\psi(t)$ - kernel function (see page 71);

$\Psi(x)$ - barrier function (see page 73);

$B(x)$ - the natural barrier function (see page 32).

Special elements

e - the identity element;

x^{-1} - the inverse element of x;

$x^{1/2}$ - the square root of x.

Chapter 1

Introduction

To introduce the topic of this thesis we have to go back to the first paper on *interior-point methods* by Karmarkar [27] in 1984. He introduced a *polynomial-time projective algorithm* for *linear optimization* (LO), the first polynomial-time algorithm performing well in practice. Since then LO revived as an active area of research. Today the resulting interior-point methods are among the most effective methods for solving LO problems. Many researchers have proposed and analyzed various interior-point methods for LO and a large amount of results have been reported. For a survey we refer to recent books on the subject [45, 51, 52]. An interesting fact is that almost all known polynomial-time variants of interior-point methods use the so-called *central path* [52] as a guideline to the optimal set and some variant of Newton's method to follow the central path approximately. Therefore, analyzing the behavior of Newton's method has been a crucial issue in the theoretical investigation of interior-point methods. It is generally agreed that *primal-dual path-following methods* are the most efficient methods from the computational point of view (see e.g., Andersen et al. [3]). These methods use the Newton direction as a search direction; this direction is closely related to the well-known primal-dual logarithmic barrier function.

The interior-point methods developed for LO could be naturally extended to obtain polynomial-time methods for conic optimization. In conic optimization, a linear function is minimized over the intersection of an affine space and a closed convex cone. The foundation for solving these problems by interior-point methods was laid by Nesterov and Nemirovskii [37]. These authors considered primal (and dual) interior point methods based on so-called self-concordant barrier functions. Later, Nesterov and Todd [38, 39] introduced symmetric primal-dual interior-point methods on a special class of cones called self-scaled cones, which allowed a symmetric treatment of the primal and the dual problem. Conic optimization includes solving problems such as linear optimization, semi-definite optimization and second order cone optimization problems (see e.g. [2, 12]).

During the last two decades interior-point methods have proved to be a powerful tool to solve convex optimization problems (see for example [37]), provided that we have a self-concordant computationally tractable barrier function for the underlying cone. Until recently all the barrier functions considered were so-called logarithmic barrier functions. However, there is a gap between the practical behavior of the algorithms and the theoretical performance results, where the practical behavior is better than the worst-case complexity analysis. This is especially true for the so-called *large-update methods*. If n denotes the number of variables in the problem, then the theoretical complexity analysis of large-update methods yielded an $O(n \log(n/\epsilon))$ iteration bound, where ϵ represents the desired accuracy of the solution. In practice, however, large-update methods are much more efficient than the so-called small-update methods for which the theoretical iteration bound is only $O(\sqrt{n} \log(n/\epsilon))$. So the current theoretical bounds differ by a factor $\sqrt{n}$, in favor of the small-update methods. This gap is significant.

1.1 Kernel functions

Recently, the gap could be narrowed by deviating from the usual approach. Peng et al. [40–43] replaced the primal-dual logarithmic barrier by a so-called self-regular barrier function, which is determined by a simple univariate self-regular function, called its kernel function. The search direction was modified accordingly, and a large-update method was obtained for which the theoretical iteration bound is $O(\sqrt{n} \log n \log(\frac{n}{\epsilon}))$. Thus the gap between the theoretical iteration bounds for small- and large-update methods has been narrowed. They naturally extended their work to semi-definite optimization and second-order cone optimization.

Later, a new class of barrier functions was introduced whose members are not necessarily self-regular [9]. Some new analytic tools were developed for the analysis of interior-point methods based on such kernel functions. As a result the analysis is much simpler that in [40–43], whereas the iteration bounds are at least as good. In addition, the analysis also applies to some (self-regular) functions, see Bai et al. [9].

This simpler analysis in [9] motivated us to extend the analysis to *symmetric optimization* problems, which is the aim of this thesis.

1.2 Symmetric optimization

As we mentioned before, several interior-point methods for LO were extended to semi-definite optimization and second-order cone optimization. In fact, these optimization problems can be defined as minimizing a linear function over the intersection of an affine space and a closed convex cone. If the cone is the linear cone, the second order cone or the cone of real semi-definite positive symmetric matrices, then we have respectively a linear optimization problem, a second-order cone optimization problem or a semi-definite optimization problem. These three cones are the most relevant for the optimization field, and they were classified as belonging to the set of *self-dual* and *homogenous cones* (Section 2.5), also called *symmetric cones*. Thus, many authors developed interior-point methods for symmetric cones (e.g. [16–19, 46, 47]) by generalizing existing interior-point methods for LO.

1.3 Why Jordan algebras?

Jordan algebras were created to illuminate a particular aspect of physics: the quantum mechanical observables. However, Jordan algebras illuminated connections with many other areas of mathematics. A surprising observation was their relation to symmetric cones. This relation is as follows: any symmetric cone, can be realized as a *cone of squares* of some *Euclidean Jordan algebra*. It turns out that Euclidean Jordan algebras provided the tools to treat optimization problems involving symmetric cones. In short, Jordan algebras provide us with a simple structure to analyze, at once, all symmetric optimization problems.

1.4 Jordan algebras and optimization

The first work connecting Jordan algebras and optimization is due to Güler [22]. He observed that the family of the self-scaled cones ([38]) is identical to the set of symmetric cones for which there exists a complete classification theory. It is worth mentioning that Nesterov and Todd [38], who provided a theoretical foundation for the study of interior-point methods for symmetric optimization problems, did not use a Jordan algebraic approach.

Faybusovich analyzed several interior-point methods for symmetric optimization using the Jordan algebra framework: the primal-dual interior-point method analyzed by Alizadeh for semi-definite optimization [1] and a short-step path-following algorithm [17, 18]. With Arana [19], he derived complexity estimates for a long-step primal-dual interior-point algorithm. Later on, he described a primal-dual potential-reduction algorithm and its complexity estimates [16]. The complexity estimates were always obtained in terms of the rank of the Euclidean Jordan algebra.

Schmieta and Alizadeh [46] presented a general framework in which the analysis of interior-point algorithms for semi-definite optimization can be extended verbatim to optimization problems over symmetric cones derivable from associative algebras. In particular, their analysis is extendable to the cone of positive semi-definite Hermitian matrices with complex and quaternion entries, and to the second-order cone. They dealt with the case of the second-order cone by embedding its associated Jordan algebra into the Clifford algebra. Later, Schmieta and Alizadeh [47] showed that the so-called commutative class of primal-dual interior-point algorithms which were designed by Monteiro and Zhang for semi-definite optimization [35] extends word-by-word to all symmetric cones. They also proved polynomial-time worst-case bounds for variants of the short-, semi-long, and long-step path-following algorithms using the Nesterov-Todd, XS, or SX directions.

Rangarajan [44] established polynomial-time convergence of infeasible interior-point methods for symmetric optimization.

The well known software SeDuMi [49] was a major contribution of Jos Sturm. This software solves symmetric optimization problems. SeDuMi also solves optimization problems with complex variables. With Luo and Zhang he also analyzed the so-called self-dual embedding for symmetric optimization [32].

Other contributions to the application of Jordan algebraic techniques in optimization were given by Baes, Hauser and Lim [4, 23, 24, 31].

1.5 Subject of the thesis

In this thesis we deal with the symmetric optimization problem. As may be clear now, the object of symmetric optimization is to minimize a linear function over the intersection of an affine subspace and a symmetric cone. Symmetric optimization problems offer a unified framework for linear optimization, second-order cone optimization, semi-definite optimization and combinations of these.

1.5.1 Our approach to symmetric optimization

As said before, our aim is to generalize the kernel function based approach of Peng et al. [40–43] and Bai et al. [9] for LO to symmetric optimization. We thus obtain an interior-point method for solving symmetric optimization problems. It relies on a real univariate function, called kernel function, which generates an associated barrier function. The way to generate the barrier function is via the eigenvalues of its argument, for which the theory of Euclidean Jordan algebras provides the necessary framework. The barrier function is used to define the search direction and to define a proximity measure of the current iterate to the central path.

Many things are generalizable word-by-word but others do not. In fact, we have encountered some difficulties: while the analysis of the method in LO deals with the coordinates of the vectors, in symmetric optimization we have to deal with the eigenvalues (or spectral decomposition) and even sometimes with the so-called Peirce decomposition in a Euclidean Jordan algebra (see Section 2.9). This fact produced the main differences between LO and symmetric optimization. Particularly, we had to establish some similarity properties, which replaced some equalities in LO. This is important because if two elements are similar the value of the barrier function is the same for both elements.

By definition a kernel function is e-convex. A main issue in the kernel function based approach is to prove that the associated barrier function is also e-convex, because only then we are able to perform the analysis of the algorithm. The proof of e-convexity of the barrier function turned out to be the most demanding part of our work.

During our work we developed formulas for some derivatives of eigenvalues. They are interesting in themselves, despite the fact that we did not use them (see Section 3.4).

1.5.2 Contents of the thesis

We next summarize the contents of each chapter, giving a global picture of this thesis.

Chapter 2: We give a not so short but an as simple as possible introduction to the theory of Euclidean Jordan algebras and symmetric cones, also explaining the connection between Euclidean Jordan algebras and symmetric cones. This introduction is sufficient for the purpose of this thesis and does not require a priori knowledge of Jordan algebras, especially since we have included a large number of proofs. The results presented are not new, but we found it useful to give our own exposition.

Chapter 3: We present some properties of eigenvalues, especially similarity relations and inequalities. We also give formulas for the derivatives of eigenvalues and recall derivatives formulas for separable spectral functions, in terms of Euclidean Jordan algebras.

Chapter 4: We recall the notion of a kernel function and discuss how a kernel function can be used to define a barrier function for symmetric cones. Moreover, we achieve a crucial inequality for the barrier function based on the so-called exponential convexity of the kernel function.

Chapter 5: We first define the symmetric optimization problem and the new search direction based on kernel functions, using the framework of Euclidean Jordan algebras. Moreover, we prove that we can adapt the algorithm for LO presented in [9] to general symmetric optimization problems, using properties that we have developed before.

Chapter 6: This chapter contains some concluding remarks and suggestions for further research.

Appendix: The appendix consists of four sections. One of these contains a quick introduction to quaternions and octonions, including some results. We also prove three properties concerning matrices of quaternions. The remaining sections contains some notions from topology, some matrix properties and a few technical properties that are relevant for the thesis.

Chapter 2

Euclidean Jordan algebras and symmetric cones

This chapter offers an introduction to the theory of Euclidean Jordan algebras as needed for the optimization techniques that we present later. In order to make the thesis as self-supporting as possible, we decided to write down the proofs of quite a large number of properties. We omit proofs in cases that would require the introduction of concepts that are far behind the purpose of this work. The approach we present closely follows Faraut and Korányi [15].

2.1 Power associative algebras

It turns out later on that Jordan algebras are power associative. Therefore, to start with we introduce the notion of a power associative algebra and some of its properties.

Let V be a finite-dimensional vector space over $\mathbb{R}$. A map $h : V \times V \mapsto V$ is called bilinear if:

(i) $h(\alpha u + \beta v, w) = \alpha h(u, w) + \beta h(v, w)$ for all $u, v \in V$ and $\alpha, \beta \in \mathbb{R}$.

(ii) $h(w, \alpha u + \beta v) = \alpha h(w, u) + \beta h(w, v)$ for all $u, v \in V$ and $\alpha, \beta \in \mathbb{R}$.

If there exists a bilinear map $(x, y) \to x \circ y$ from $V \times V$ into V then $(V, \circ)$ is called an *algebra* over $\mathbb{R}$ (also called an $\mathbb{R}$-algebra). We call "$\circ$" the product of $(V, \circ)$. If the product "$\circ$" is *associative*, that is, for all $x, y, z \in V$,

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

then $(V, \circ)$ is called an *associative algebra*. Moreover, an $\mathbb{R}$-algebra $(V, \circ)$ is *commutative* if we have, for all $x, y \in V$,

$$x \circ y = y \circ x.$$

If for some $e \in V$,

$$x \circ e = e \circ x = x$$

for every $x \in V$ then e is an identity element of V. V can have at most one identity element, because if e_1 and e_2 are identity elements then $e_1 = e_1 \circ e_2 = e_2 \circ e_1 = e_2$. So, if it exists, the identity element e is unique.

In an algebra $(V, \circ)$ with identity element e, we recursively define *powers* of elements as follows:

$$x^0 := e, \; x^n := x \circ x^{n-1} \text{ for } n \in \mathbb{N}.$$

Definition 2.1.1. An $\mathbb{R}$-algebra $(V, \circ)$ is *power associative* if it has an identity element and for any $x \in V$ and nonnegative integers p, q, one has: $x^p \circ x^q = x^{p+q}$. □

From now on, we assume that $(V, \circ)$ is a power associative $\mathbb{R}$-algebra with identity element e. We let $\mathbb{R}[X]$ denote the algebra over $\mathbb{R}$ of all polynomials in the variable X with coefficients in $\mathbb{R}$ and with the usual product of polynomials.

For an element x in V we define

$$\mathbb{R}[x] := \{p(x) : p \in \mathbb{R}[X]\}.$$

Let U be a subspace of V. We say that U is a subalgebra of V if $x \circ y \in U$ for all $x, y \in U$. We say that the subalgebra of V generated by x and e consists of all linear combinations, with coefficients in $\mathbb{R}$, of e and powers of x. Obviously, the algebra $(V, \circ)$ is power associative if for each $x \in V$, the subalgebra generated by x and e, i.e. $\mathbb{R}[x]$, is associative.

Since V is a finite-dimensional vector space, for each $x \in V$, there exists a positive integer k such that the set $\{e, x, x^2, \ldots, x^k\}$ is linearly dependent. This implies the existence of a polynomial $p \neq 0$ such that $p(x) = 0$. Recall that a monic polynomial is a polynomial with the leading coefficient equal to 1. We define the *minimal polynomial* of $x \in V$ as the monic polynomial $p \in \mathbb{R}[X]$ of minimal degree such that $p(x) = 0$. The minimal polynomial is unique, because if p_1 and p_2 are two distinct minimal polynomials of x, then we have $p_1(x) - p_2(x) = 0$. Since p_1 and p_2 are monic polynomials, the degree of $p_1 - p_2$ is less than the degree of p_1, which contradicts the minimality of p_1.

We define the *degree* of x, denoted as degree(x), as the degree of the minimal polynomial of x. Obviously degree$(x) \leq \dim(V)$, where $\dim(V)$ denotes the dimension of the vector space V over $\mathbb{R}$.

We define the *rank* of V as

$$\text{rank}(V) := \max\{\text{degree}(x) : x \in V\}.$$

An element $x \in V$ is called *regular* if degree$(x) = \text{rank}(V)$.

The next proposition uses the notion of dense and open set. The meaning of these well known concepts are presented in Appendix A.

Proposition 2.1.2 (Proposition II.2.1 in [15]). *Let $(V, \circ)$ be a power associative $\mathbb{R}$-algebra with rank r. The set of regular elements is open and dense in V. There exist polynomials $a_1, a_2, \ldots, a_r \in \mathbb{R}[X]$, $i = 1 \ldots, r$ such that $a_i(x) \in \mathbb{R}$ and the minimal polynomial of every regular element $x \in V$ in the variable λ is given by*

$$f(\lambda; x) = \lambda^r - a_1(x)\lambda^{r-1} + a_2(x)\lambda^{r-2} + \cdots + (-1)^r a_r(x).$$

The polynomials $a_1, \ldots, a_r$ are unique and a_j is homogeneous of degree j, for $1 \leq j \leq r$.

The polynomial $f(\lambda; x)$ in the above proposition is called the *characteristic polynomial* of x. Furthermore, the proposition immediately implies that $f(x; x) = 0$. Its proof is beyond the scope of our work. But it is not hard to get the functions $a_j(x)$. Let y be a regular element. Then the elements $e, y, y^2, \ldots, y^{r-1} \in V$ are independent and there exist elements $b_1, \ldots, b_{n-r} \in V$ such that

$$B = \{e, y, y^2, \ldots, y^{r-1}, b_1, \ldots, b_{n-r}\}$$

is a basis of V. Note that we want to find polynomials a_j such that

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

This equation can be thought of a system of n linear equations in r unknowns $a_j(x)$, with respect to the basis B. By Cramer's rule we get

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

where Det denotes the usual matrix determinant. It can be proved that a_j, $j = 1, \ldots, r$ are indeed polynomials, see [15].

Since the regular elements are dense in V we can extend by continuity the polynomials $a_i(x)$ to all elements of V and consequently the characteristic polynomial. Moreover, the minimal polynomial is equal to the characteristic polynomial for regular elements (as stated in Proposition 2.1.2), but it divides the characteristic polynomial of non-regular elements.

For an element $x \in V$, let $L(x) : V \to V$ be the linear map defined by

$$L(x)y := x \circ y, \quad \forall y \in V. \tag{2.1}$$

By its definition, $L(x)$ is linear in x and $L(e)$ is the identity operator which we denote by I.

Let $x \in V$ be a regular element and $L_0(x)$ be the restriction of $L(x)$ to $\mathbb{R}[x]$. The set $B := \{e, x, \ldots, x^{r-1}\}$ is a basis of $\mathbb{R}[x]$ since x is regular. Then it is clear that the matrix of $L_0(x)$ with respect to the basis B is given by the $r \times r$ matrix

$$L_0(x) = \begin{bmatrix} 0 & 0 & \ldots & 0 & (-1)^{r-1}a_r(x) \\ 1 & 0 & \ldots & 0 & (-1)^{r-2}a_{r-1}(x) \\ 0 & 1 & \ldots & 0 & (-1)^{r-3}a_{r-2}(x) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & 1 & a_1(x) \end{bmatrix}. \tag{2.2}$$

To build the matrix we used that

$$L_0(x)x^{r-1} = x^r = a_1(x)x^{r-1} - a_2(x)x^{r-2} + \cdots + (-1)^{r-1}a_r(x)e,$$

by Proposition 2.1.2. Therefore, with Tr and Det denoting the usual trace and determinant of endomorphisms, we have

$$\begin{aligned} \mathrm{Tr} L_0(x) &= a_1(x), \\ \mathrm{Det} L_0(x) &= a_r(x). \end{aligned}$$

By the density of the regular elements in V these equalities extends to non-regular elements.

Definition 2.1.3. The coefficient $a_1(x)$ is called the *trace* of x, denoted as $\mathrm{tr}\,(x)$. The coefficient $a_r(x)$ is called the *determinant* of x, denoted as $\det(x)$. The roots of the characteristic polynomial are called the *eigenvalues* of x.

According to this definition we have

$$\begin{aligned} \mathrm{tr}\,(x) &= \mathrm{Tr} L_0(x) \\ \det x &= \mathrm{Det} L_0(x). \end{aligned}$$

We define the polynomial $F(\lambda; L_0(x)) := \mathrm{Det}(\lambda I - L_0(x))$, for $x \in V$ regular, in the algebra of matrices. The large amount of zeros in the matrix of $L_0(x)$, as given by 2.2, makes it easy to compute the determinant of $\lambda I - L_0(x)$ which turns out to be exactly the characteristic polynomial of x. Since

$$\mathrm{Det}(\lambda I - L_0(x)) = \mathrm{Det}(L_0(\lambda e - x)) = \det(x - \lambda e),$$

where the last equality is due to Definition 2.1.3. It follows that the characteristic polynomial of x is given by

$$f(\lambda; x) = \det(\lambda e - x).$$

Hence it follows that the eigenvalues of x are precisely the eigenvalues of $L_0(x)$.

Example 2.1.4. Let e be the identity element of V. Since $L_0(e)$ is the identity map, its matrix is the identity matrix. So we have $\mathrm{tr}\,(e) = r$ and $\det(e) = 1$. Clearly its minimal polynomial is $\lambda - 1$ and its characteristic polynomial is given by

$$f(\lambda; e) = \det(\lambda e - e) = \det((\lambda - 1)e) = a_r((\lambda - 1)e).$$

Since a_r is homogenous of degree r, it follows that

$$f(\lambda; e) = (\lambda - 1)^r a_r(e) = (\lambda - 1)^r \det(e) = (\lambda - 1)^r.$$

Note that $\lambda - 1$ divides $(\lambda - 1)^r$. □

Since the trace is a homogeneous polynomial of degree 1, we have

$$\mathrm{tr}\,(x+y) = \mathrm{tr}\,(x) + \mathrm{tr}\,(y)\,.$$

In general it is not true that $\det(x \circ y) = \det(x)\det(y)$. This is illustrated by a simple example.

Example 2.1.5. Let $\mathbf{S}^n$ be the Euclidean space of $n \times n$ real symmetric matrices. Let the binary operation "$\circ$" defined by

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

in $\mathbf{S}^n$, where XY denotes the usual matrix product. Since the usual product of matrices is associative, $(\mathbf{S}^n, \circ)$ is a power associative $\mathbb{R}$-algebra. Since for $X \in \mathbf{S}^n$ the characteristic polynomial is $\mathrm{Det}(\lambda I - X)$, the determinant of X, $\det(X)$, is the usual determinant of a matrix. Now, let

$$X := \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} \qquad \text{and} \qquad Y := \begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix}.$$

It easily follows that

$$\mathrm{Det}(X \circ Y) = 4 \neq 5 = \mathrm{Det}(X)\mathrm{Det}(Y).$$

Hence, in general, $\det(x \circ y) \neq \det(x)\det(y)$. □

After this example, the following result is interesting.

Proposition 2.1.6 (Proposition II.2.2 in [15]). *For all u in V, and x, y in $\mathbb{R}[u]$ we have*

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

Proof. We assume first that u is regular. For x in $\mathbb{R}[u]$ let $L_0(x)$ denote the restriction of $L(x)$ to $\mathbb{R}[u]$. The algebra $\mathbb{R}[u]$ is associative. Therefore, for any x and y in $\mathbb{R}[u]$,

$$L_0(x \circ y) = L_0(x)L_0(y),$$

and hence we have

$$\det(x \circ y) = \mathrm{Det}L_0(x \circ y) = \mathrm{Det}L_0(x)\mathrm{Det}L_0(y) = \det(x)\det(y),$$

i.e., for any polynomials p and q in $\mathbb{R}[u]$ we have

$$\det(p(u) \circ q(u)) = \det p(u) \det q(u).$$

Finally, the set of regular elements in V is dense in V, and by continuity in u it follows that the last equality holds for any u in V. ■

An element x is said to be an *invertible element* if there exists an element y in $\mathbb{R}[x]$ such that $x \circ y = e$. Since $\mathbb{R}[x]$ is associative, y is unique. It is called the *inverse* of x and is denoted by $y := x^{-1}$. If $y \in V$ and $x \circ y = e$, it is not necessarily true that any y is the inverse of x. Before giving an example we first deal with the following result.

Proposition 2.1.7 (Proposition II.2.3 in [15]). *If $L(x)$ is invertible, then x is invertible and $x^{-1} = L(x)^{-1}e$.*

Proof. If $L(x)$ is invertible, the restriction $L_0(x)$ of $L(x)$ to $\mathbb{R}[x]$ is one to one and onto. Hence $y = L(x)^{-1}e$ belongs to $\mathbb{R}[x]$ and $x \circ y = x \circ L(x)^{-1}e = L(x)L(x)^{-1}e = e$. ■

The following example also shows that the converse of Proposition 2.1.7 is not true.

Example 2.1.8. Let V be the vector space of 2×2 symmetric matrices and "$\circ$" the product defined in Example 2.1.5. Now consider

$$X := \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \; Y := \begin{bmatrix} 1 & a \\ a & -1 \end{bmatrix}, \; Z := \begin{bmatrix} 0 & a \\ a & 0 \end{bmatrix}, \; a \in \mathbb{R}.$$

Then X is invertible and $X^{-1} = X$. We have $X \circ Y = I$, but Y does not belong to $\mathbb{R}[X]$ for $a \neq 0$. Also, $L(X)$ is not invertible since $L(X)Z = 0$. □

Proposition 2.1.9 (Proposition II.2.4 in [15]). *An element x is invertible if and only if $\det x \neq 0$, and then*

$$x^{-1} = \frac{q(x)}{\det x}, \tag{2.3}$$

where

$$q(x) := (-1)^{r-1}\left(x^{r-1} - a_1(x)x^{r-2} + \cdots + (-1)^{r-1}a_{r-1}(x)e\right).$$

Proof. We have

$$\begin{aligned} x \circ q(x) &= x \circ ((-1)^{r-1}(x^{r-1} - a_1(x)x^{r-2} + \cdots + (-1)^{r-1}a_{r-1}(x)e) \\ &= (-1)^{r-1}(x^r - a_1(x)x^{r-1} + \cdots + (-1)^{r-1}a_{r-1}(x)x) \\ &\quad [\text{ by Proposition 2.1.2}] \\ &= (-1)^{r-1}((-1)(-1)^r a_r(x)e) \\ &= (-1)^{2r}a_r(x)e = \det(x)e. \end{aligned}$$

Therefore, if $\det(x) \neq 0$, then x is invertible and x^{-1} is given by (2.3). Conversely, if x is invertible, there exists a polynomial $p \in \mathbb{R}[x]$ such that $x \circ p(x) = e$, and, by Proposition 2.1.6

$$\det(x)\det(p(x)) = 1,$$

therefore $\det(x) \neq 0$. ■

The above proposition establishes that the set of the invertible elements is given by

$$\mathcal{I} := \{x \in V : \det x \neq 0\}. \tag{2.4}$$

By Proposition 2.1.7 we have

$$\mathcal{L} := \{x \in V : L(x) \text{ is invertible } \}$$

is a subset of $\mathcal{I}$. Moreover, we have the following result.

Proposition 2.1.10. *The set* $\mathcal{L} = \{x \in V : \text{Det}L(x) \neq 0\}$ *is dense in the set of the invertible elements. In other words,* $\text{cl}(\mathcal{L}) = \mathcal{I}$.

Proof. Since $\mathcal{L}$ is a subset of $\mathcal{I}$, it is enough to prove that $\mathcal{I}$ is a subset of the closure of $\mathcal{L}$. Let $y \in \mathcal{I}$. If $\text{Det}L(y) \neq 0$ then $y \in \mathcal{L}$. Suppose now that $\text{Det}L(y) = 0$. Then y belongs to $\text{cl}(\mathcal{L})$ if for small enough $\epsilon > 0$, we have $y - \epsilon e \in \mathcal{L}$. Let

$$p(\epsilon) := \text{Det}(L(y - \epsilon e)) = \text{Det}(L(y) - \epsilon I),$$

where I denotes the identity operator. Since the roots of p are the eigenvalues of $L(y)$ it follows that there exists α^* such that

$$\alpha^* := \min\{|\alpha| : p(\alpha) = 0 \text{ and } \alpha \neq 0\}.$$

Hence, for all ϵ such that $0 < \epsilon < \alpha^*$ we have $p(\epsilon) \neq 0$. Thus $\text{Det}(L(y - \epsilon e)) \neq 0$ which is equivalent to $y - \epsilon e \in \mathcal{L}$ and this implies that $y \in \text{cl}(\mathcal{L})$. The result follows. ■

2.2 Jordan algebras

Definition 2.2.1. Let $(V, \circ)$ be a finite-dimensional $\mathbb{R}$-algebra. Then $(V, \circ)$ is a *Jordan* $\mathbb{R}$*-algebra* if

(J1) $x \circ y = y \circ x \qquad \forall x, y \in V$,

(J2) $x \circ (x^2 \circ y) = x^2 \circ (x \circ y) \qquad \forall x, y \in V$. □

Using definition from (2.1) of $L(x)$ it is clear that (J1) and (J2) are equivalent to

(J1*) $L(x)y = L(y)x \quad \forall x, y \in V$,

(J2*) $L(x)L(x^2) = L(x^2)L(x) \quad \forall x \in V$.

The property (J2*) means that the operators $L(x)$ and $L(x^2)$ commute. The notation $[S, T] := ST - TS$ is called the *commutator* of S and T, for any two endomorphisms of the vector space V. Hence the property (J2*) can also be written as

$$[L(x), L(x^2)] = 0, \forall x \in V. \tag{2.5}$$

We sometimes abbreviate the notation $(V, \circ)$ to V when there is no possible confusion on the product.

Proposition 2.2.2. *Let $(V, \circ)$ be a Jordan $\mathbb{R}$-algebra. Then the following identities hold for all $x, y \in V$:*

(i) $[L(y), L(x^2)] + 2[L(x), L(x \circ y)] = 0,$

(ii) $[L(x), L(y \circ z)] + [L(y), L(z \circ x)] + [L(z), L(x \circ y)] = 0,$

(iii) $L(x^2 \circ y) - L(x^2)L(y) = 2(L(x \circ y) - L(x)L(y))L(x).$

Proof. By (2.5) we have for each $t \in \mathbb{R}$:

$$\begin{aligned} [L(x+ty), L((x+ty)^2)] &= [L(x) + tL(y), L(x^2) + 2tL(x \circ y) + t^2 L(y^2)] \\ &= [L(x), L(x^2)] + 2t[L(x), L(x \circ y)] + t[L(y), L(x^2)] \\ &\quad +2t^2[L(y), L(x \circ y)] + t^2[L(x), L(y^2)] + t^3[L(y), L(y^2)] \\ &= t\left(2[L(x), L(x \circ y)] + [L(y), L(x^2)]\right) \\ &\quad +t^2\left(2[L(y), L(x \circ y)] + [L(x), L(y^2)]\right) = 0. \end{aligned}$$

So we have for all $t \in \mathbb{R}$,

$$t\left(2[L(x), L(x \circ y)] + [L(y), L(x^2)]\right) + t^2\left(2[L(y), L(x \circ y)] + [L(x), L(y^2)]\right) = 0,$$

therefore

$$2[L(x), L(x \circ y)] + [L(y), L(x^2)] = 0$$

and (i) follows. If we replace x by $x + tz$ and y by and $y + tz$, with $t \in \mathbb{R}$, in (i) and perform as in the previous item, one obtains (ii). Applying both sides of (i) to an element $z \in V$, the resulting identity can be rewritten as

$$L(y)(x^2 \circ z) - L(x^2)(y \circ z) = 2L(x \circ y)(x \circ z) - 2L(x)((x \circ y) \circ z),$$

which is valid for all $x, y, z \in V$. Using (J1*), this is the same as

$$L(x^2 \circ z)y - L(x^2)L(z)y = 2L(x \circ z)L(x)y - 2L(x)L(z)L(x)y,$$

and this means that

$$L(x^2 \circ z) - L(x^2)L(z) = 2L(x \circ z)L(x) - 2L(x)L(z)L(x),$$

which is just the identity (iii), when z is replaced by y. ∎

Jordan algebras are not necessarily associative, but they are power associative, as becomes clear in the next result.

Proposition 2.2.3 (Proposition II.1.2 in [15])**.**

(i) Let V be a Jordan $\mathbb{R}$-algebra. Then, for any x in V and any positive integers p and q:

$$[L(x^p), L(x^q)] = 0.$$

(ii) Any Jordan algebra is power associative.

Proof. *(i)* Let End(V) be the set of endomorphisms of V. We use the identity *(iii)* of Proposition 2.2.2 with $y := x^{n-1}$ and $n \in \mathbb{N}$, which gives that

$$L(x^{n+1}) = L(x^2)L(x^{n-1}) + 2L(x^n)L(x) - 2L(x)L(x^{n-1})L(x), \ n \in \mathbb{N}.$$

This implies, by induction to n, that, for every n, $L(x^n)$ belongs to the subalgebra of End(V) generated by $L(x)$ and $L(x^2)$, which is commutative by (J2*). In other words, it follows that $L(x^n)$ is a polynomial in $L(x)$ and $L(x^2)$ and (i) follows.

(ii) We need to show that $x^p \circ x^q = x^{p+q}$ for all positive integers p and q. We start with $q = 2$ and use induction on p. By definition it is obviously true if $p = 1$. Using (J2) we may write

$$x^2 \circ x^{p+1} = x^2 \circ (x \circ x^p) = x \circ (x^2 \circ x^p) = x \circ x^{2+p} = x^{3+p}.$$

Now using induction on q we obtain

$$x^{p+q+1} = x^{p+q} \circ x = (x^p \circ x^q) \circ x = L(x)L(x^p)x^q,$$

and using *(i)*:

$$x^{p+q+1} = L(x^p)L(x)x^q = x^p \circ x^{q+1}.$$

The proposition is proved. ■

Corollary 2.2.4. *For all u in V, and x, y in $\mathbb{R}[u]$, we have*

$$L(x)L(y) = L(y)L(x).$$

Proof. The proof immediately follows from Proposition 2.2.3, since x and y are polynomials in u. ■

The following three examples illustrate some of the properties presented before.

Example 2.2.5. Defining for all $x, y \in \mathbb{R}^n$ the operation

$$x \circ y := (x_1y_1; \dots ; x_ny_n),$$

one easily verifies that $(\mathbb{R}^n, \circ)$ is a Jordan $\mathbb{R}$-algebra. The characteristic polynomial of $x \in \mathbb{R}^n$ is

$$f(\lambda; x) = (\lambda - x_1)(\lambda - x_2) \dots (\lambda - x_n),$$

and rank$(\mathbb{R}^n) = n$. Consequently, the trace of x is the sum of all components of the vector x and the determinant is their product. The identity element is $e = (1; \dots ; 1)$ and the inverse element of x is

$$x^{-1} = (x_1^{-1}; x_2^{-1}; \dots ; x_n^{-1}),$$

if it exists. □

Example 2.2.6. We denote $(x_0; x_1; \ldots; x_n) \in \mathbb{R}^{n+1}$ as $x = (x_0; \bar{x})$ with $\bar{x} := (x_1; \ldots; x_n)$ and define the product as

$$x \circ y := (x^T y; x_0 \bar{y} + y_0 \bar{x}).$$

Therefore $(\mathbb{R}^{n+1}, \circ)$ is a Jordan $\mathbb{R}$-algebra. We can see easily that the identity element is $(1; 0; \ldots; 0)$. Suppose that x and e are linearly independent. Now we want to show that $\text{rank}(\mathbb{R}^{n+1}) = 2$. So we need to verify whether there exist $\alpha, \beta \in \mathbb{R}$ such that

$$x^2 = \alpha e + \beta x.$$

Since $x^2 = (x^T x; 2x_0 \bar{x})$, we may rewrite this equation as a system of equations,

$$x^T x = \alpha + \beta x_0 \tag{2.6}$$

$$2x_0 x_i = \beta x_i \quad i = 1, \ldots, n. \tag{2.7}$$

Since $x \neq e$ there exists some i such that $x_i \neq 0$. Therefore, we get $\beta = 2x_0$ and consequently $\alpha = \|x\|^2 - 2x_0^2 = \|\bar{x}\|^2 - x_0^2$. Hence, the characteristic polynomial of x is

$$\lambda^2 - 2x_0 \lambda + x_0^2 - \|\bar{x}\|^2$$

and $\text{rank}(\mathbb{R}^{n+1}) = 2$. Obviously, the trace of x is $2x_0$ and the determinant is $x_0^2 - \|\bar{x}\|^2$. Moreover, we can obtain the inverse element using Proposition 2.1.9:

$$x^{-1} = \frac{1}{\det(x)}(-1)(x - a_1(x)e) = \frac{1}{\det(x)}(-1)(x - 2x_0 e) = \frac{1}{x_0^2 - \|\bar{x}\|^2}(x_0; -\bar{x}),$$

if it exists, i.e., if $x_0^2 - \|\bar{x}\|^2 \neq 0$. □

Example 2.2.7. Let $\mathbf{S}^n$ be the Euclidean space of real symmetric matrices. With "$\circ$" defined as in Example 2.1.5, $(\mathbf{S}^n, \circ)$ is a Jordan $\mathbb{R}$-algebra. Remark that the Jordan algebra is commutative but not associative, contrary to the usual matrix product which is associative but not commutative. Since $X \circ X = XX$, we can easily conclude that the powers of X are equal wether we consider the Jordan algebra $(\mathbf{S}^n, \circ)$ or the algebra of symmetric matrices with the usual product. Thus, the characteristic polynomial of X is $\det(\lambda I - X)$ and $\text{rank}(\mathbf{S}^n) = n$. Consequently, the trace and the determinant are the usual ones, the identity and the inverse elements are the identity matrix and the inverse matrix. □

2.3 Quadratic representation

In this section we give the notion of the *quadratic representation* of a Jordan algebra. Let $(V, \circ)$ be a finite-dimensional Jordan algebra over $\mathbb{R}$, with the identity element e. Given $x \in V$, we define:

$$P(x) := 2L(x)^2 - L(x^2).$$

It will turn out to be much easier to work with the operator $P(x)$ rather than with $L(x)$. The endomorphisms $L(x)$ and $P(x)$ commute, because we may write

$$\begin{aligned} P(x)L(x) &= 2L(x)^2L(x) - L(x^2)L(x) \\ &= 2L(x)L(x)^2 - L(x)L(x^2), \end{aligned}$$

where we used Proposition 2.2.3. Therefore,

$$P(x)L(x) = L(x)P(x). \tag{2.8}$$

Example 2.3.1. Let $(\mathbb{R}^n, \circ)$ be the Jordan algebra with $x \circ y := (x_1y_1; \dots; x_ny_n)$, as defined in Example 2.2.5. Then $L(x) = \mathrm{Diag}(x)$ and $P(x) = \mathrm{Diag}(x^2)$, where $\mathrm{Diag}(x)$ denotes a diagonal matrix whose entries are the x_i's in their natural order. □

Example 2.3.2. Let $(\mathbb{R}^{n+1}, \circ)$ be the Jordan algebra defined in Example 2.2.6. Consider the canonical basis in $\mathbb{R}^{n+1}$, it is quite standard to obtain the matrix of $L(x)$

$$L(x) = \begin{bmatrix} x_0 & \bar{x}^T \\ \bar{x} & x_0 I \end{bmatrix},$$

where we identify $L(x)$ with its matrix. By definition, we easily get

$$P(x) = \begin{bmatrix} x^T x & 2x_0\bar{x}^T \\ 2x_0\bar{x} & \det(x)I + 2\bar{x}\bar{x}^T \end{bmatrix},$$

after some elementary calculations. □

Example 2.3.3. Let $(\mathbf{S}^n, \circ)$ be the Jordan algebra defined in Example 2.2.7. Consider the matrix X in the vector space of $n \times n$ matrices, denoted as $\mathbf{M}_n(\mathbb{R})$. If we choose the basis $\{B_{11}, B_{12}, \dots, B_{1n}, \dots, B_{nn}\}$ and B_{ij} $i, j = 1, \dots, n$ are matrices such that in row i and column j its entry is equal to 1 and the others entries are equal to 0, after tedious calculations we get

$$L(X) = \frac{1}{2}(X \otimes I + I \otimes X),$$

where $\otimes$ denotes the Kronecker product. It easily follows that

$$P(X) = X \otimes X,$$

by using the properties of the Kronecker product. By definition, we obtain $P(X)Y = XYX$. □

Proposition 2.3.4 (Proposition II.3.1 in [15]). *An element $x \in V$ is invertible if and only if the linear operator $P(x) : V \mapsto V$ is invertible. In this case:*

$$P(x)x^{-1} = x, \tag{2.9}$$

$$P(x)^{-1} = P(x^{-1}). \tag{2.10}$$

Proof. If $P(x)$ is invertible, then the restriction of $P(x)$ to $\mathbb{R}[x]$ is a bijection and $y = P(x)^{-1}x$ belongs to $\mathbb{R}[x]$. Since

$$P(x)e = 2L(x)^2e - L(x^2)e = 2L(x)(x \circ e) - x^2 \circ e = x^2$$

and $P(x)$ and $L(x)$ commute we have

$$(P(x)^{-1}x) \circ x = L(x)P(x)^{-1}x = P(x)^{-1}L(x)x = P(x)^{-1}x^2 = e,$$

proving (2.9). Suppose that x is invertible. Since $\mathbb{R}[x]$ is associative we have,

$$P(x)x^{-1} = 2L(x)^2x^{-1} - L(x^2)x^{-1} = 2L(x)(x \circ x^{-1}) - x^2 \circ x^{-1} = x.$$

Since x^{-1} is a polynomial in x, by Proposition 2.2.3 $L(x)$ and $L(x^{-1})$ commute. Hence if we replace y by x^{-1} (respectively $y = x^{-2}$) in Proposition 2.2.2-(*iii*) we obtain

$$P(x)L(x^{-1}) = L(x)$$

and

$$2L(x)L(x^{-1}) - P(x)L(x^{-2}) = I,$$

respectively, where I represents the identity operator. If we substitute the first equation in the second one we conclude

$$P(x)P(x^{-1}) = I,$$

proving (2.10). ■

Corollary 2.3.5 (Corollary II.3.2 in [15]). *The set $\mathcal{I}$ of invertible elements is given by*

$$\mathcal{I} = \{x : \mathrm{Det}P(x) \neq 0\}.$$

Proof. By Propositions 2.1.9 and 2.3.4 the result follows. ■

In the remaining section we prove some more properties on the quadratic representation. For their proof we need of the well known concepts as derivative and gradient.

For now, we assume that V is endowed with an inner product denoted as $\langle \cdot, \cdot \rangle$ and with the induced norm, $\| \cdot \|$.

Definition 2.3.6. We say a function h is $o(u)$, denoted as $h = o(u)$, if

$$\lim_{u \to 0} \frac{\|h(u)\|}{\|u\|} = 0.$$

Definition 2.3.7. Let $f : U \mapsto V$ and U an open subset of V. The function f is *differentiable* at $x \in U$ if there exists a linear map $g(x) : V \mapsto V$ such that

$$f(x+u) - f(x) - g(x)u = o(u), \qquad \forall u \in V. \tag{2.11}$$

We denote $g(x)$ by $D_x f(x)$ or $f'(x)$. The function f is differentiable on U if f is differentiable at all points in U. In addition, if the map $x \mapsto g(x)$ is continuous at each $x \in U$ then f is said to be *continuously differentiable*.

If x is a point of the interior of the domain of f and $u \in V$, we say that the function f is differentiable in the direction u at the point x if the limit

$$D_x^u f(x) := \lim_{t \to 0} \frac{f(x+tu) - f(x)}{t}$$

exists.

If f is differentiable at x , then $D_x^u f(x)$ exists and is equal to $D_x f(x)u$. The linear map $D_x f(x)$ is called the *derivative* of f at x. It is often called the *gradient* of f and denoted by $\nabla f(x)$.

Let $v : V \mapsto V$ and $w : V \mapsto V$ be differentiable functions. The bilinearity of the operation "$\circ$" and the definition of D_x^u leads to

$$D_x^u (v \circ w) = v \circ D_x^u w + D_x^u v \circ w = L(v) D_x^u w + L(w) D_x^u v.$$

This is obtained as easily as for the product of univariate real functions. Consequently

$$\nabla(v \circ w) = L(v)\nabla w + L(w)\nabla v.$$

The definition of the powers of $x \in V$ implies

$$\nabla x^{k+1} = L(x)\nabla x^k + L(x^k).$$

Especially we have $\nabla x = I$ and $\nabla x^2 = 2L(x)$.

Proposition 2.3.8. *One has:*

(i) The gradient of the map $x \mapsto x^{-1}$is $-P(x)^{-1}$, i.e.,

$$\nabla x^{-1} = -P(x)^{-1}.$$

(ii) If $x, y \in V$ are both invertible, then $P(x)y$ is also invertible and moreover:

$$(P(x)y)^{-1} = P(x^{-1})y^{-1}.$$

(iii) For any $x, y \in V$:

$$P(P(y)x) = P(y)P(x)P(y).$$

Proof. By differentiating the relations

$$x^{-1} \circ x = e \text{ and } x^2 \circ x^{-1} = x$$

in the u direction, we obtain

$$L(x)D_x^u(x^{-1}) + L(x^{-1})u = 0 \tag{2.12}$$

and

$$2L(x)u \circ x^{-1} + L(x^2)D_x^u(x^{-1}) = u, \tag{2.13}$$

respectively. Multiplying equation (2.12) by $2L(x)$ and subtracting equation (2.13) yields (recalling that $L(x)$ and $L(x^{-1})$ commute)

$$P(x)D_x^u(x^{-1}) = -u$$

and (i) follows. The proof of second and third item will be done at the same time. Since, by the proof of Proposition 2.3.4

$$L(x^{-1})P(x) = P(x)L(x^{-1}) = L(x),$$

we have

$$x^{-1} \circ P(x)y = L(x)^{-1}P(x)y = L(x)y = x \circ y. \tag{2.14}$$

Since, $P(x)y = 2x \circ (x \circ y) - x^2 \circ y$ we get

$$\begin{aligned} D_x^u\,(P(x)y) &= 2u \circ (x \circ y) + 2x \circ (u \circ y) - 2(x \circ u) \circ y \\ &= 2(L(u)L(y)x - L(y)L(u)x + L(u \circ y)x) \\ &= 2Q(u, y)x, \end{aligned}$$

where we define $Q(u, y) := L(u)L(y) - L(y)L(u) + L(u \circ y)$. Regarding the left- and right-hand sides of (2.14) as functions of x and applying D_x^u we obtain

$$(-P(x^{-1})u) \circ (P(x)y) + 2x^{-1} \circ Q(u, y)x = u \circ y.$$

Setting $u = y^{-1}$, and since $L(y)$ and $L(y^{-1})$ commute, it follows that

$$(-P(x^{-1})y^{-1}) \circ (P(x)y) + 2x^{-1} \circ x = e,$$

or

$$(P(x^{-1})y^{-1}) \circ (P(x)y) = e.$$

If $L(P(x)y)$ is invertible, then $P(x)y$ is invertible and

$$(P(x)y)^{-1} = P(x^{-1})y^{-1}. \tag{2.15}$$

Since, by Proposition 2.1.10, the set $\{(x, y) : \det L(P(x)y) \neq 0\}$ is dense in $\{(x, y) : \det(P(x)y) \neq 0\}$, the identity (2.15) remains valid for $x, y \in V$ such that $\det(P(x)y) \neq 0$. Now, applying (i), i.e., use (i) to get the gradients of both sides of equation (2.15), we obtain

$$-P(P(y)x)^{-1}P(y) = -P(y)^{-1}P(x)^{-1},$$

which can be written as follows:

$$P(y)P(x)P(y) = P(P(y)x). \tag{2.16}$$

Using equation (2.16) and Proposition 2.3.4 it follows that

$$\{(x, y) : \det(P(x)y) \neq 0\} = \{(x, y) : \det(x)\det(y) \neq 0\}.$$

Thus, we conclude that the formula (2.15) is still valid for x and y invertible and the third item is proved for x and y invertible. Since both sides of (2.16) are polynomials in x and y, the result holds in general by continuity. Everything is proved. ■

The identity

$$P(P(y)x) = P(y)P(x)P(y) \tag{2.17}$$

is known as the *fundamental formula*.

Corollary 2.3.9. *Let $x \in V$ and k a positive integer. One has*

$$P(x^k) = P(x)^k.$$

Proof. If $k = 1$ then it is trivially true. We proceed with induction to k. Suppose that $P(x^k) = P(x)^k$. By the fundamental formula we have

$$P(P(x^k)x) = P(x^k)P(x)P(x^k).$$

By induction hypothesis we have

$$P(x^k)P(x)P(x^k) = P(x)^k P(x)P(x)^k = P(x)^{2k+1}.$$

Since $P(x^k)x = x^{2k+1}$ we obtain

$$P(x^{2k+1}) = P(x)^{2k+1}.$$

On the other hand, using the same arguments, we have

$$P(P(x^k)e) = P(x^k)P(e)P(x^k)$$

which, by the fundamental formula and induction hypothesis, implies

$$P(x^{2k}) = P(x)^{2k}.$$

It follows that we have proved the property for all positive integers. ■

2.4 Euclidean Jordan algebras

A Jordan algebra V over $\mathbb{R}$ is said to be *Euclidean* if there exists an inner product which is associative. In other words, if there exists an inner product, denoted by $\langle u, v\rangle$, such that

$$\langle x \circ u, v\rangle = \langle x, u \circ v\rangle$$

for all x, u, v, in V.

We will assume that $(V, \circ, \langle\cdot,\cdot\rangle)$ is an finite-dimensional Euclidean Jordan algebra over $\mathbb{R}$ with rank r.

An element $c \in V$ is said to be *idempotent* if $c^2 = c$. Two idempotents c and d are said to be *orthogonal* if $c \circ d = 0$. Since then

$$\langle c, d\rangle = \langle c^2, d\rangle = \langle c, c \circ d\rangle = \langle c, 0\rangle = 0,$$

orthogonal idempotents are orthogonal with respect to the inner product. One says that $c_1, \ldots, c_k$, with $k \leq r$, is a complete system of orthogonal idempotents if

$$\begin{aligned} &c_i^2 = c_i, \quad \forall i \\ &c_i c_j = 0 \text{ if } i \neq j, \\ &c_1 + c_2 + \cdots + c_k = e. \end{aligned}$$

Theorem 2.4.1 (Theorem III.1.1 in [15]). **(Spectral theorem, type I.)** *For x in V there exist unique real numbers $\lambda_1, \ldots, \lambda_k$, all distinct, and a unique complete system of orthogonal idempotents $c_1, \ldots, c_k$ such that*

$$x = \lambda_1 c_1 + \cdots + \lambda_k c_k,$$

with $k = \mathrm{degree}(x)$. For each $j = 1, \ldots, k$, we have $c_j \in \mathbb{R}[x]$. The numbers λ_j are the eigenvalues of x and $\sum_{j=1}^k \lambda_j c_j$ is called the spectral decomposition *of x.*

Sketch of the proof. Essentially, the proof of the theorem uses the spectral decomposition of an endomorphism; for details, we refer to [36]. For y in $\mathbb{R}[x]$, let $M(y)$ be the restriction of $L(y)$ to $\mathbb{R}[x]$. Then $M(x)$ is a symmetric endomorphism of the Euclidean space $\mathbb{R}[x]$. Therefore, applying the spectral decomposition to $M(x)$, there exist non-zero orthogonal projections $P_1, \ldots, P_k$ in $\mathbb{R}[x]$ such that $P_1 + \cdots + P_k = I$, and real numbers λ_i such that

$$M(x) = \lambda_1 P_1 + \cdots + \lambda_k P_k.$$

Furthermore, there exist polynomials p_j such that $P_j = p_j(M(x))$. We set $c_j := p_j(x)$. Using the associativity of the algebra $\mathbb{R}[x]$ we obtain

$$M(c_j) = M(p_j(x)) = p_j(M(x)) = P_j,$$

where the second equality follows from the fact that $M(u \circ v) = M(u)M(v)$ for $u, v \in \mathbb{R}[x]$ (see Corollary 2.2.4). Similarly,

$$\begin{aligned} &M(c_i \circ c_j) = M(c_i)M(c_j) = P_i P_j, \\ &M(\sum c_j) = \sum P_j = I, \\ &M(\sum \lambda_j c_j) = \sum \lambda_j P_j = M(x). \end{aligned}$$

Since M is injective (if $M(y) = 0$ then $y = M(y)e = 0$) it follows that

$$\begin{aligned} &c_i^2 = c_i, \ c_i \circ c_j = 0 \ if \ i \neq j \\ &\sum c_j = e, \ \sum \lambda_j c_j = x. \end{aligned}$$

To prove uniqueness, suppose $x = \sum \alpha_j d_j$ where α_j's are all distinct and $d_1, \ldots, d_k$ are complete system of orthogonal idempotents. We have $p(x) = \sum p(\alpha_j) d_j$ for every polynomial p. Defining, for fixed j,

$$p^{(j)}(X) = \prod_{i \neq j} (X - \alpha_i),$$

it follows that

$$p^{(j)}(x) = \prod_{i \neq j} (\alpha_j - \alpha_i) d_j,$$

or equivalently

$$d_j = \frac{p^{(j)}(x)}{\prod_{i \neq j} (\alpha_j - \alpha_i)},$$

since the α_j's are all different. Therefore d_j belongs to $\mathbb{R}[x]$. Since $d_1, \dots, d_j$ is complete system of idempotents and M is invertible, the $M(d_j)$'s are mutually orthogonal projections, and so the α_j's are necessarily just eigenvalues of $M(x)$. Each d_j is then the orthogonal projection, $d_j = M(d_j)e$, of e onto the eigenspace of $M(x)$ corresponding to α_j. This makes everything unique. ∎

We say that an idempotent is *primitive* if it is non-zero and cannot be written as the sum of two (necessarily orthogonal) non-zero idempotents. We say that $c_1, \dots, c_m$ is a complete system of orthogonal primitive idempotents, or a *Jordan frame*, if each c_j is a primitive idempotent and if

$$c_i \circ c_j = 0, i \neq j$$
$$\sum_{j=1}^{m} c_j = e.$$

Every Jordan frame has $r = \text{rank}(V)$, as we prove in the following result.

Theorem 2.4.2 (Theorem III.1.2 in [15]). **Spectral theorem, type II.** *Suppose V has rank r. Then for every x in V there exists a Jordan frame $c_1, \dots, c_r$ and real numbers $\lambda_1, \dots, \lambda_r$ such that*

$$x = \sum_{j=1}^{r} \lambda_j c_j.$$

The numbers λ_j (with their multiplicities) are uniquely determined by x. Furthermore,

$$\det(x) = \prod_{j=1}^{r} \lambda_j, \quad \text{tr}\,(x) = \sum_{j=1}^{n} \lambda_j.$$

More generally,

$$a_k(x) = \sum_{1 \leq i_1 \leq \cdots \leq i_k \leq r} \lambda_{i_1} \dots \lambda_{i_k},$$

where a_k $(1 \leq k \leq r)$ is the polynomial defined in Proposition 2.1.2.

Proof. If $x = \sum_{i=1}^{k} \lambda_i c_i$ is the spectral decomposition given by Theorem 2.4.1, then, clearly, $p(x) = \sum_{i=1}^{k} p(\lambda_i) c_i$ for every polynomial p. Therefore, the minimal polynomial of x is

$$f(X, x) = \prod_{i=1}^{k} (X - \lambda_i).$$

From this we see that $k \le r$, and $k = r$ exactly if x is regular. In the latter case each c_i is primitive, otherwise it would be possible to have a Jordan frame of more than r elements, and one could construct elements in V whose minimal polynomial had degree higher than r, contradicting the definition of rank.

The formulas for $\det(x)$, $\operatorname{tr}(x)$ and more generally $a_k(x)$ are now obvious when x is regular. When x is not regular, since the set of regular elements are dense in V, it is still the limit of a sequence $\left(x^{(n)}\right)_{n\in\mathbb{N}}$ of regular elements. Then $x^{(n)} = \sum_{i=1}^{r} \lambda_i^{(n)} c_i^{(n)}$ by the above, where $c_1^{(n)}, c_2^{(n)}, \ldots, c_r^{(n)}$ is a Jordan frame for every $n \in \mathbb{N}$. Since the converging sequence is a compact set in V there exists a subsequence such that there exist limits $\lambda_i = \lim_k \lambda_i^{(n_k)}$ and $c_i = \lim_k c_i^{(n_k)}$. Our statements follow from this. The uniqueness is immediate from Theorem 2.4.1. ■

Whenever we refer to spectral decomposition of an element we mean the spectral decomposition of type II given by Theorem 2.4.2, unless stated otherwise.

The decomposition of type I can be obtained from the decomposition of type II as follows. Let $x = \sum_{i=1}^{r} \lambda_i c_i$ be the spectral decomposition of type II of x. Let us define the integers $s, q_1, \ldots, q_s$ such that $q_s := r$ and

$$\lambda_1(x) = \cdots = \lambda_{q_1}(x) > \lambda_{q_1+1}(x) = \cdots = \lambda_{q_2}(x) > \cdots > \lambda_{q_s}(x).$$

Define $J_i := \{q_{i-1}+1, \ldots, q_i\}$ (with $q_0 = 0$) and put $e_i := \sum_{j\in J_i} c_j$. Then $e_1, \ldots, e_s$ is a complete system of idempotents and

$$x = \sum_{i=1}^{s} \lambda_{q_i}(x) e_i \tag{2.18}$$

is the spectral decomposition of type I.

The spectral decomposition of type II includes all the eigenvalues of x including multiplicities, while the type I decomposition includes only the different eigenvalues. A direct consequence of Theorems 2.4.1 and 2.4.2 is that eigenvalues of elements in a Euclidean Jordan algebra are always real. Since $e = c_1 + \cdots + c_r$, by Theorem 2.4.1 the eigenvalues of e are all equal to one, and it immediately follows that $\operatorname{tr}(e) = r$ and $\det(e) = 1$. This agrees with Example 2.1.4.

We give the following examples of spectral decompositions.

Example 2.4.3. The Jordan algebra $(\mathbb{R}^n, \circ)$ defined in Example 2.2.5 is Euclidean with the associative inner product defined as $\langle x, y\rangle := \operatorname{tr}(x \circ y) = x^T y$, $x, y \in \mathbb{R}^n$. The canonical basis in $\mathbb{R}^n$ is a Jordan frame, and we can write any $x \in \mathbb{R}^n$ as $(x_1; x_2; \ldots; x_n) = x_1 c_1 + \cdots + x_n c_n$. In this case, the Jordan frame is unique. We just need to observe that $x_i^2 = x_i$ is equivalent to $x_i = 0$ or $x_i = 1$, for $i = 1 \ldots n$. □

Example 2.4.4. Consider the Jordan algebra $(\mathbb{R}^{n+1}, \circ)$ defined in Example 2.2.6. We have that $(\mathbb{R}^{n+1}, \circ)$ is Euclidean with the associative inner product defined as

$\langle x, y \rangle := \mathrm{tr}\,(x \circ y) = 2x^T y$, $x, y \in \mathbb{R}^{n+1}$. Moreover, any $x \in \mathbb{R}^{n+1}$, with $\bar{x} \neq 0$, has the spectral decomposition

$$x = \lambda_1(x)c_1(x) + \lambda_2(x)c_2(x),$$

where

$$\lambda(x) = \begin{bmatrix} \lambda_1(x) \\ \lambda_2(x) \end{bmatrix} = \begin{bmatrix} x_0 - \|\bar{x}\| \\ x_0 + \|\bar{x}\| \end{bmatrix}$$

and

$$c_1(x) = \frac{1}{2} \begin{bmatrix} 1 \\ -\frac{\bar{x}}{\|\bar{x}\|} \end{bmatrix} \qquad c_2(x) = \frac{1}{2} \begin{bmatrix} 1 \\ \frac{\bar{x}}{\|\bar{x}\|} \end{bmatrix},$$

where $\|\bar{x}\|$ is the Euclidean norm in $\mathbb{R}^n$. When $\bar{x} = 0$ the spectral decomposition is trivial. □

Example 2.4.5. Let $(\mathbf{S}^n, \circ)$ be the Jordan algebra defined in Example 2.2.7. With the associative inner product defined as $\langle X, Y \rangle = \mathrm{tr}\,(X \circ Y) = \mathrm{tr}\,(XY)$ for all $X, Y \in \mathbf{S}^n$, $(\mathbf{S}^n, \circ)$ is a Euclidean Jordan algebra. Any $X \in \mathbf{S}^n$ has spectral decomposition:

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

where $\lambda_i,\ i = 1, \ldots, n$, are the eigenvalues of X and $q_i, i = 1, ..., n$ are unitary eigenvectors of X. Since X is symmetric, all eigenvalues are real numbers. The symmetric matrix $q_i q_i^T$ is an idempotent, since

$$q_i q_i^T \circ q_i q_i^T = q_i q_i^T q_i q_i^T = q_i q_i^T.$$

Moreover, it is easy to see that $\{q_1 q_1^T, \ldots, q_n q_n^T\}$ is a Jordan frame. □

We can extend the definition of any real valued univariate and continuous function $f(.)$ to elements of a Euclidean Jordan algebra, using the eigenvalues. This can be done as follows. Starting from the (unique) spectral decomposition of type I, $x = \sum_{i=1}^{s} \lambda_{q_i}(x) e_i$ (cf. (2.18)), we (uniquely) define the function $F : V \mapsto \mathbb{R}$:

$$F(x) := \sum_{i=1}^{s} f(\lambda_{q_i}(x)) e_i. \tag{2.19}$$

Since

$$\sum_{i=1}^{s} f(\lambda_{q_i}(x)) e_i = \sum_{i=1}^{r} f(\lambda_i(x)) c_i,$$

we can also use a spectral decomposition of type II to obtain the value of F at x.

Of particular interest are the following examples:

1. The square root: $x^{\frac{1}{2}} := \lambda_1^{\frac{1}{2}} c_1 + \cdots + \lambda_r^{\frac{1}{2}} c_r$, whenever all $\lambda_i \geq 0$, and undefined otherwise;

2. The inverse: $x^{-1} := \lambda_1^{-1} c_1 + \cdots + \lambda_r^{-1} c_r$ whenever all $\lambda_i \neq 0$ and undefined otherwise.

Note that x^{-1} given as a linear combination of $c_1, \ldots, c_r$ also belongs to $\mathbb{R}[x]$ because $c_i \in \mathbb{R}[x]$ for every $i = 1, \ldots, r$. Moreover $(\lambda_1^{-1} c_1 + \cdots + \lambda_r^{-1} c_r) \circ x = e$. Thus, the expression of x^{-1} using a spectral decomposition of x is the same as the algebraic inverse element considered before.

In Chapter 4 we will deal with more examples that are extensively used in this thesis.

Proposition 2.4.6 (Proposition III.1.3 in [15])**.** *Let c be an idempotent in a Jordan algebra. The only possible eigenvalues of $L(c)$ are $0, \frac{1}{2}$, and 1.*

Proof. Using identity (iii) of Proposition 2.2.2, with $x = y$, we obtain

$$L(x^3) = 3L(x^2)L(x) - 2L(x)^3,$$

and for $x = c$:

$$2L(c)^3 - 3L(c)^2 + L(c) = 0.$$

Therefore, an eigenvalue λ of $L(c)$ is a solution of

$$2\lambda^3 - 3\lambda^2 + \lambda = 0,$$

whose roots are 0, $\frac{1}{2}$ and 1. ■

Proposition 2.4.7 (Proposition III.1.5 in [15])**.** *Let V be a Jordan algebra over $\mathbb{R}$. The following properties are equivalent.*

(i) V is Euclidean.

(ii) The symmetric bilinear form $\operatorname{tr}(x \circ y)$ is positive definite.

Sketch of the proof. $(i) \Rightarrow (ii)$. By Theorem 2.4.2, any x in V has a spectral decomposition $x = \sum_{i=1}^{r} \lambda_j c_j$, where the λ_i's are real numbers, and $x^2 = \sum_{i=1}^{r} \lambda_j^2 c_j$. If $x \neq 0$, then

$$\operatorname{tr}\left(x^2\right) = \sum \lambda_i^2 > 0.$$

$(ii) \Rightarrow (i)$. Let $\langle x, y \rangle := \operatorname{tr}(x \circ y)$. Then $\langle x, y \rangle$ is a positive definite symmetric bilinear form. It remains to show that $\langle\ ,\ \rangle$ is associative. The proof is beyond the scope of the thesis. We refer to [15] or [4]. ■

Proposition 2.4.7 also means that $\operatorname{tr}(x \circ y)$ is an inner product. In the sequel, $\langle x, y \rangle$ will always denote this inner product. So

$$\langle x, y \rangle := \operatorname{tr}(x \circ y).$$

The norm induced by this inner product is given by

$$\|x\| := \sqrt{\langle x, x\rangle}.$$

The norm of $x \in V$ can be obtained by using its eigenvalues. So, let $x = \sum_{i=1}^{r} \lambda_i^2(x) c_i$ a spectral decomposition of x. It is straightforward to see that

$$x^2 = \sum_{i=1}^{r} \lambda_i^2(x) c_i.$$

Thus,

$$\|x\| := \sqrt{\langle x, x\rangle} = \sqrt{\operatorname{tr}(x^2)} = \sqrt{\sum_{i=1}^{r} \lambda_i^2(x)}.$$

The operator $L(x)$ and $P(x)$ are self-adjoint with respect to this inner product. To explain this we need some definitions. If g is a linear map from V to V, then we call the linear map g^* the *adjoint* of g, with respect to the inner product $\langle\ ,\ \rangle$, if

$$\langle gx, y\rangle = \langle x, g^* y\rangle, \text{ for all } x, y \in V.$$

If $g = g^*$ then g is said to be *self-adjoint*. Now we can easily prove that $L(x)$ is self-adjoint:

$$\langle L(x)y, z\rangle = \langle x \circ y, z\rangle = \langle y \circ x, z\rangle = \langle y, x \circ z\rangle = \langle y, L(x)z\rangle, \quad \forall x, y, z \in V,$$

where the third equality follows from the associativity of the inner product. As we now show this implies that the quadratic representation $P(x)$ is also self-adjoint. For $x, y, x \in V$ we have

$$\begin{aligned}
\langle P(x)y, z\rangle &= \langle 2L^2(x)y - L(x^2)y, z\rangle \\
&= \langle 2L^2(x)y, z\rangle - \langle L(x^2)y, z\rangle \\
&= \langle y, 2L^2(x)z\rangle - \langle y, L(x^2)z\rangle \\
&= \langle y, P(x)z\rangle,
\end{aligned}$$

where the equalities follows from the definition of the quadratic representation and since $L(x)$ is self-adjoint.

2.5 Symmetric cones

In this section the relevance of the theory of Euclidean Jordan algebras to Symmetric Optimization will be illuminated. It may be useful to recall the definition of a convex cone. Let V be a finite-dimensional real Euclidean space.

Definition 2.5.1. A non-empty subset $\mathcal{K}$ of V is a *cone* if $x \in \mathcal{K}$ and $\lambda \geq 0$ imply that $\lambda x \in \mathcal{K}$. □

A subset S of V is said to be *convex* if $x, y \in S$ and $0 \leq \lambda \leq 1$ imply that $\lambda x + (1-\lambda)y \in S$. It is clear that $\mathcal{K} \subset V$ is a *convex cone* if and only if $\mathcal{K}$ is a cone and a convex set. One easily verifies that a cone $\mathcal{K}$ is convex if and only if $x, y \in \mathcal{K}$ imply that $x + y \in \mathcal{K}$. The *dual cone* of $\mathcal{K}$ is defined as

$$\mathcal{K}^* := \{y \in V : \langle x, y\rangle \geq 0, \forall x \in \mathcal{K}\}.$$

It is easy to see that $\mathcal{K}^*$ is closed convex cone. If $\mathcal{K} = \mathcal{K}^*$ we say that $\mathcal{K}$ is *self-dual* . The cone $\mathcal{K}$ is said to be *pointed* if $\mathcal{K} \cap -\mathcal{K} = \{0\}$. In what follows, $\mathcal{K}$ will denote a non-empty pointed convex cone.

For a convex cone $\mathcal{K}$ we define the *automorphism group*, $\mathrm{Aut}(\mathcal{K}) \subset \mathrm{GL}(V)$ of $\mathcal{K}$ by

$$\mathrm{Aut}(\mathcal{K}) := \{g \in \mathrm{GL}(V) : g(\mathcal{K}) = \mathcal{K}\}, \tag{2.20}$$

where $\mathrm{GL}(V)$ is the set of invertible linear maps g from V into itself. The convex cone $\mathcal{K}$ is said to be *homogeneous* if $\mathrm{Aut}(\mathcal{K})$ acts transitively on $\mathcal{K}^0$, i.e., for all $x, y \in \mathcal{K}^0$ there exists $g \in \mathrm{Aut}(\mathcal{K})$ such that $gx = y$.

Definition 2.5.2. The convex cone $\mathcal{K}$ is said to be *symmetric* if it is homogeneous and self-dual. □

Let V be a Euclidean Jordan $\mathbb{R}$-algebra and let $\mathcal{K}(V)$ be the set of all squares, i.e,

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

If $x^2 \in \mathcal{K}(V)$ and $\alpha \geq 0$ then $\alpha x^2 = (\sqrt{\alpha}x)^2 \in \mathcal{K}(V)$. It follows that $\mathcal{K}(V)$ is a cone. We call $\mathcal{K}(V)$ the *cone of squares* of the Euclidean Jordan algebra V. To prove convexity we use the dual cone, which is given by

$$\mathcal{K}^*(V) = \{y \in V : \langle y, x^2\rangle \geq 0 \ \forall x \in V\}.$$

Since

$$\langle y, x^2\rangle = \langle y \circ x, x\rangle = \langle L(y)x, x\rangle,$$

we have

$$\mathcal{K}^*(V) = \{y \in V : L(y) \text{ is positive semidefinite}\}.$$

Proposition 2.5.3. *The cone $\mathcal{K}(V)$ is self-dual.*

Proof. Let $x^2 \in \mathcal{K}(V)$ with $x \in V$ and $x = \sum_{i=1}^r \lambda_i c_i$ its spectral decomposition. Then

$$x^2 = \sum_{i=1}^r \lambda_i^2 c_i,$$

and

$$L(x^2) = \sum_{i=1}^r \lambda_i^2 L(c_i).$$

By Proposition 2.4.6, the operators $L(c_i)$ are positive semidefinite, therefore $L(x^2)$ is positive semidefinite and hence $x^2 \in \mathcal{K}^*(V)$. Conversely, let x belong to $\mathcal{K}^*(V)$, then $L(x)$ is positive semidefinite and particularly

$$\langle L(x)c_i, c_i \rangle \geq 0.$$

Since the idempotents $c_1, \ldots, c_r$ are orthogonal we have

$$\langle L(x)c_i, c_i \rangle = \lambda_i \langle c_i, c_i \rangle = \lambda_i \|c_i\|^2 \geq 0, \quad \forall i.$$

Hence, the eigenvalues are non-negative, then we can write $x = y^2$ with

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

Therefore, it follows that $\mathcal{K}^*(V) = \mathcal{K}(V)$. ■

Corollary 2.5.4. *$\mathcal{K}(V)$ is a closed convex cone.*

Proof. Since $\mathcal{K}^*(V)$ is a closed convex cone, it immediately follows from Proposition 2.5.3 that $\mathcal{K}(V)$ is closed and convex. ■

In the sequel, unless stated otherwise, $\mathcal{K}$ will always denote the cone $\mathcal{K}(V)$.

Proposition 2.5.5. *As before (see (2.4)), we denote $\mathcal{I}$ the set of invertible elements in V. Then*

$$\mathcal{K}^0 = \{x^2 : x \in \mathcal{I}\},$$

where $\mathcal{K}^0$ denotes the interior of $\mathcal{K}$.

Proof. We will prove that

$$\Omega := \{x^2 : x \in \mathcal{I}\}$$

is equal to

$$\Gamma := \{x \in V : L(x) \text{ is positive definite}\}.$$

Let $y = w^2 \in \Omega$ with $w \in \mathcal{I}$. Then $L(y)$ is positive definite if

$$0 < \langle z, L(y)z \rangle = \langle z^2, w^2 \rangle \text{ for all } z \in V \setminus \{0\}.$$

If $w = \sum_{i=1}^r \lambda_i c_i$ then $w^2 = \sum_{i=1}^r \lambda_i^2 c_i$. Hence

$$\langle zL(w^2), z \rangle = \langle \sum_{i=1}^{r} \lambda_i^2 L(c_i) z, z \rangle.$$

We claim that for each $z \in V \setminus \{0\}$ there is i such that $\langle L(c_i)z, z \rangle > 0$. Suppose that there exists $z \in V \setminus \{0\}$ for all i such that $\langle L(c_i)z, z \rangle = 0$. Remark that $L(c_i)$ is positive semidefinite. Therefore

$$0 = \sum_{i=1}^{r} \langle L(c_i)z, z \rangle = \sum_{i=1}^{r} \langle c_i \circ z, z \rangle = \langle e \circ z, z \rangle = \langle z, z \rangle.$$

This implies that $z = 0$ which contradicts the hypothesis of $z \in V \setminus \{0\}$. Since w is invertible, the eigenvalues of w^2 are greater than zero. From everything that was exposed, we can now say that $L(w^2)$ is positive definite. It follows that $\Omega \subset \Gamma$. Let $y \in \Gamma$ and $y = \sum_{i=1}^{r} \lambda_i c_i$ its spectral decomposition. Since, by definition $\langle z, L(y)z\rangle > 0$ for all $z \in V \setminus \{0\}$, we have

$$\langle L(y)c_i, c_i\rangle = \lambda_i \langle c_i, c_i\rangle = \lambda_i \|c_i\|^2 > 0,$$

which implies that

$$\lambda_i = \frac{1}{\|c_i\|}\langle L(y)c_i, c_i\rangle > 0.$$

Therefore y is invertible and $y = w^2$ with

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

It follows that $\Gamma \subset \Omega$, and the result is proved because $\Gamma = (\mathcal{K}^*)^0 = \mathcal{K}^0$. ■

Remark 2.5.6. From Proposition 2.5.5, it immediately follows that $\mathcal{K}$ is the closure of $\{x^2 : x \in \mathcal{I}\}$.

Before proving the following proposition, we give some definitions. Let $g : V \mapsto V$ be a self-adjoint linear map. We say g is *positive semidefinite (definite)* if $\langle gx, x\rangle \geq 0$ (> 0) for all $x \in V$ $(V \setminus \{0\})$.

Proposition 2.5.7. *We have the following properties:*

(i) $P(x)\mathcal{K}^0 = \mathcal{K}^0, \; x \in \mathcal{I}$;

(ii) $P(x)$, *with* $x \in \mathcal{K}^0$, *is positive definite.*

Proof. (*i*) Since $\mathcal{K}^0$ is convex, $P(x)\mathcal{K}^0$ is also convex. Note that by Proposition 2.3.8 any $z = P(x)y \in P(x)\mathcal{K}^0$ is invertible, for $y \in \mathcal{K}^0$ and $x \in \mathcal{I}$. This means that $P(x)\mathcal{K}^0$ does not cross the border of $\mathcal{K}^0$. Thus, it is enough to prove that there is a common element to both sets: we have for invertible x, $x^2 = P(x)e$ belongs to $\mathcal{K}^0$ and to $P(x)\mathcal{K}^0$. Hence $P(x)\mathcal{K}^0 \subseteq \mathcal{K}^0$. Let $y \in \mathcal{K}^0$. We have that $P(x)y \in P(x)\mathcal{K}^0 \subseteq \mathcal{K}^0$. Therefore, $y = P(x)^{-1}P(x)y \in P(x)^{-1}\mathcal{K}^0$. It follows that $\mathcal{K}^0 \subseteq P(x)\mathcal{K}^0$ for x invertible. Thus, $P(x)\mathcal{K}^0 = \mathcal{K}^0$.

(*ii*) If $x \in \mathcal{K}^0$ then $x = u^2$, with $u \in \mathcal{I}$. Therefore

$$P(x) = P(u^2) = P(u)^2.$$

Since if u is invertible then $P(u)$ is invertible (Proposition 2.3.4), we can conclude that $P(x)$ is positive definite.

The proposition is proved. ■

Proposition 2.5.8. *$\mathcal{K}$ is a symmetric cone.*

Proof. By Proposition 2.5.3, $\mathcal{K}$ is self-dual. For $x, y \in \mathcal{K}^0$ we define

$$g := P(y^{1/2})P(x^{-1/2}).$$

By Proposition 2.5.7 and Remark 2.5.6, $P(x) \in \mathrm{Aut}(\mathcal{K})$ for invertible x. Therefore, $g \in \mathrm{Aut}(\mathcal{K})$ and

$$gx = P(y^{1/2})P(x^{-1/2})x = P(y^{1/2})e = y.$$

We conclude that for any x and y in $\mathcal{K}^0$ there exists $g \in \mathrm{Aut}(\mathcal{K})$ such that $gx = y$. Hence, $\mathcal{K}$ is homogeneous. Moreover, $\mathcal{K}$ is symmetric. ∎

We have proved that $\mathcal{K}$ is a symmetric cone. In fact, the converse also holds. However, its proof is beyond the scope of this thesis. We state the result below without proof.

Theorem 2.5.9 (Theorem III.3.1 in [15]). **(Jordan algebraic characterization of symmetric cones)** *A cone is symmetric if and only if it is the cone of squares of some Euclidean Jordan algebra.*

Proposition 2.5.10. *Let V be an Euclidean Jordan $\mathbb{R}$-algebra with rank r and $\mathcal{K}$ its cone of squares. If $x \in V$ is such that $x = \sum_{i=1}^{r} \lambda_i c_i$, where $c_1, \dots c_r$ is a Jordan frame, then $\lambda_i \geq 0 (> 0)$ for $i = 1, \dots k$ if and only if $x \in \mathcal{K}(\mathcal{K}^0)$.*

Proof. If $\lambda_i \geq 0$ then we can write $x = y^2$, with $y = \sum_{i=1}^{r} \sqrt{\lambda_i} c_i$, which means that $x \in \mathcal{K}$. In case that $\lambda_i > 0$ for $i = 1, \dots, r$, we know that x is invertible, and it follows that $x \in \mathcal{K}^0$. Conversely, if $x \in \mathcal{K}$ we have $x = y^2$, with $y \in V$. If we denote α_i the eigenvalues of y, we can write $x = \sum_{i=1}^{r} \alpha_i^2 c_i$ where α_i^2's are the eigenvalues of x, which are greater or equal than zero. If $x \in \mathcal{K}^0$ then y is invertible, therefore α_i^2's are greater than zero. The proposition is proved. ∎

We give the following two more properties for the quadratic representation.

Proposition 2.5.11. *$P(x)^{\frac{1}{2}} = P(x^{\frac{1}{2}})$ if $x \in \mathcal{K}$.*

Proof. We have $P(x^{1/2})e = 2L^2(x^{1/2})e - L(x)e = x$. Therefore, by the fundamental formula,

$$P(x) = P(P(x^{1/2})e) = P(x^{1/2})P(e)P(x^{1/2}) = P(x^{1/2})^2.$$

Since $L(x)$ is symmetric, $P(x)$ is also symmetric, and thus $P(x)^{1/2} = P(x^{1/2})$. ∎

Proposition 2.5.12 (Proposition III.4.2 in [15]). *Let V be a Euclidean Jordan algebra. Then*

$$\det(P(y)x) = \det(y)^2 \det(x), \;\; x, y \in V. \tag{2.21}$$

Sketch of the proof. By the fundamental formula we have

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

Since, by Proposition III.4.2 in [15],

$$\operatorname{Det}(P(x)) = (\det x)^{\frac{2n}{r}},$$

with $\dim V = n$ and $\operatorname{rank}(V) = r$, the identity (2.21) follows. ■

2.6 The natural barrier function

The function $B : \mathcal{K}^0 \mapsto \mathbb{R}$ is defined as

$$B(x) := -\log\det(x).$$

As the eigenvalues are positive for any $x \in \mathcal{K}^0$ (Proposition 2.5.10), $\det(x)$ is positive for any $x \in \mathcal{K}^0$. We conclude that B is well defined.

Since, by Proposition 2.5.10, at least one eigenvalue of the elements belonging to the boundary of the $\mathcal{K}^0$ is zero, it follows that

$$B(x) \to \infty \quad \text{as} \quad x \to \partial\mathcal{K}^0.$$

We call B the *natural barrier function* of the cone $\mathcal{K}$.

Below we compute the gradient and the Hessian of B at $x \in \mathcal{K}^0$. Before doing this we recall its definition. The function $f : U \subset V \mapsto \mathbb{R}$ is said to be *twice differentiable* if f is continuous differentiable and there exists a linear operator $H(x) : V \mapsto V$ such that

$$g(x+u) - g(x) - H(x)u = o(u),$$

where g is the gradient of f at x.

If it exists, $H(x)$ is said to be the Hessian of f at x. If the function $x \mapsto H(x)$ is continuous at x then $H(x)$ is self-adjoint. In this case, $H(x)$ is the self-adjoint operator such that

$$D_x^u D_x^v f(x) = \langle H(x)u, v\rangle.$$

We denote $H(x)$ by $D_x^2 f(x)$.

Proposition 2.6.1. *One has:*

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

(ii) $D_x^2 B(x) = P(x)^{-1}$.

Proof. If $x = \sum_{i=1}^r \lambda_i c_i$, then $\det(x) = \prod_{i=1}^r \lambda_i$. Therefore, we can write

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

Using tools that will be presented in Chapter 3, the gradient of $\log\det x$ turns out to be given by

$$\nabla \log\det x = \sum_{i=1}^{r} (D_t \log t)_{t=\lambda_i} c_i = \sum_{i=1}^{r} \lambda_i^{-1} c_i = x^{-1}.$$

Since by Proposition 2.3.8 -*(i)*

$$D_x^u x^{-1} = -P(x)^{-1} u$$

and

$$D_x^v \log\det x = \langle x^{-1}, v\rangle,$$

it follows that

$$D_x^u D_x^v B(x) = D_x^u \langle -x^{-1}, v\rangle = \langle P(x)^{-1} u, v\rangle,$$

and (ii) follows. ■

Proposition 2.6.1 - (ii) relates the quadratic representation $P(x)$ with the Hessian of the natural barrier function $B(x)$.

Example 2.6.2. Consider the Euclidean Jordan algebra of Example 2.4.3. Then one can easily see that:

(i) $\mathcal{K} = \mathbb{R}^n_+$;

(ii) $B(x) = -\sum_{i=1}^n \log(x_i)$, $x = (x_1; \ldots; x_n) \in \mathcal{K}^0$;

(iii) x is invertible if and only if $x_i \neq 0$ for every i.

In case this case

$$\nabla B(x) = -\left(\frac{1}{x_1}; \ldots; \frac{1}{x_n}\right).$$

Further,

$$D_x^u D_x^v B(x) = \sum_{i=1}^{n} \frac{u_i v_i}{x_i^2}, \quad u, v \in \mathbb{R}^n.$$

Hence

$$P(x) = \mathrm{diag}(x_1^2; \ldots; x_n^2),$$

in agreement with Example 2.3.1. □

Example 2.6.3. Let $(\mathbb{R}^{n+1}, \circ)$ be the Euclidean Jordan algebra defined in Example 2.4.4. Then:

(i) $\mathcal{K} = \{(x_0; \bar{x}) \in \mathbb{R}^{n+1} : x_0 \geq \|\bar{x}\|\}$;

(ii) $B(x) = -\log(x_0^2 - \|\bar{x}\|^2)$, for $x \in \mathcal{K}^0$;

(*iii*) x is invertible if and only if $x_0 \neq \|\bar{x}\|$ and

$$\nabla B(x) = -\frac{1}{\det x}(x_0; -\bar{x}).$$

Since $H(x) = P(x)^{-1} = P(x^{-1})$ and using the matrix representation of $P(x)$ in Example 2.3.2, we obtain

$$H(x) = \frac{1}{\det(x)^2}\begin{bmatrix} x^T x & -2x_0\bar{x}^T \\ -2x_0\bar{x} & \det(x)I + 2\bar{x}\bar{x}^T \end{bmatrix}.$$

This provides an easy way to obtain the Hessian. □

Example 2.6.4. Let $(\mathbf{S}^n, \circ)$ be the Euclidean Jordan algebra defined in Example 2.4.5. We have:

(*i*) $\mathcal{K}$ is the cone of the positive semidefinite matrices;

(*ii*) $B(X) = -\log \mathrm{Det} X$ is the barrier function of the cone of positive semidefinite matrices, where $\mathrm{Det}(X)$ has the "usual" meaning;

(*iii*) the matrix X is invertible in $(\mathbf{S}^n, \circ)$, if and only if X is invertible with respect to the usual matrix product.

We have (cf. [12])

$$\nabla B(X) = -X^{-1}.$$

and

$$D_X^U D_X^V B(X) = \langle X^{-1}UX^{-1}, V\rangle \quad U, V \in \mathbf{S}^n, X \in \mathcal{K}^0.$$

Further

$$P(X)^{-1}U = X^{-1}UX^{-1}$$

for $U \in \mathbf{S}^n, X \in \mathcal{K}^0$. □

In the preceding three examples we have encountered the most common symmetric cones and their natural barrier functions in optimization:

(*i*) The so-called linear cone is the non-negative orthant:

$$\mathbb{R}^n_+ := \{x \in \mathbb{R}^n : x_i \geq 0,\ i = 1, \ldots, n\}.$$

(*ii*) The second order cone:

$$\mathcal{L}^n := \{(x_0; \bar{x}) \in \mathbb{R}^{n+1} : x_0 \geq \|\bar{x}\|\}.$$

(*iii*) The positive semidefinite cone:

$$\mathbf{S}^n_+ := \{X \in \mathbf{S}^n : X \succeq 0\},$$

where $\succeq$ means that X is positive semidefinite.

2.7 Simple Jordan algebras

A vector space V is said to be the *direct sum* of the subspaces $V_1, \dots, V_k$ if $V = V_1 + \dots + V_k$ and $V_i \cap V_j = \{0\}$ for $i \neq j$. We then write

$$V = V_1 \oplus \dots \oplus V_k.$$

A Jordan algebra V is said to be the *direct sum* of the subalgebras $V_1, \dots, V_k$, symbolically $V = V_1 \oplus \dots \oplus V_k$, if V is the direct sum of the vector spaces $V_1, \dots, V_k$, and if $V_i \circ V_j = \{0\}$ holds for $i \neq j$.

Definition 2.7.1. Let V be $\mathbb{R}$-algebra. We say that D is an *ideal* of V if it is a vector subspace of V and for every a of D and every x of V the elements $x \circ a$ and $a \circ x$ belongs to D. □

Definition 2.7.2. We say that a (Euclidean Jordan) $\mathbb{R}$-algebra V is *simple* if the only ideals are $\{0\}$ and V. □

The ideals $\{0\}$ and V are called the *trivial ideals* of V. A non-trivial ideal D is called *minimal* if there exists no nonzero ideal E, such that E is a proper subset of D.

In the next proposition we use the notation

$$V_1 \circ V_2 := \{u_1 \circ u_2 : u_1 \in V_1, u_2 \in V_2\},$$

for any two subsets of V.

Proposition 2.7.3 (Proposition III.4.4 in [15])**.** *If V is a Euclidean Jordan algebra, then it is, in a unique way, a direct sum of simple ideals.*

Proof. Let D be a minimal ideal in V. We first show that the orthogonal complement

$$D^\perp := \{x \in V : \langle x, y\rangle = 0, \forall y \in D\},$$

is an ideal. Let x be in V and y in $D^\perp$, then for any z in D

$$\langle x \circ y, z\rangle = \langle y, x \circ z\rangle = 0,$$

because the inner product is associative and $x \circ z$ belongs to D. Hence $x \circ y$ belongs to $D^\perp$. Note that, if P_D is an orthogonal projection onto D, then, for every $y \in V$,

$$y = P_D y + (I - P_D)y.$$

Since $(I - P_D)y \in D^\perp$, this means that we can write $V = D \oplus D^\perp$. We have

$$D \circ D^\perp \subset D \cap D^\perp = \{0\},$$

because D and $D^{\perp}$ are ideals. Therefore V is the direct sum of the ideals D and $D^{\perp}$. Now we apply the same process to $D^{\perp}$. Since V is a finite-dimensional vector space, after a finite number of steps we obtain the desired decomposition. To prove the uniqueness we consider two decompositions of V as a direct sum of simple ideals:

$$\begin{aligned} V &= D_1 \oplus D_2 \oplus \cdots \oplus D_k \\ &= E_1 \oplus E_2 \oplus \cdots \oplus E_\ell. \end{aligned}$$

The intersection $D_i \cap E_1$ is an ideal, therefore by minimality of either it is $\{0\}$ or $E_1 = D_i$. It can not be $\{0\}$ for every i, since this would imply

$$\forall i, D_i \circ E_1 \subset D_i \cap E_1 = \{0\},$$

hence $V \circ E_1 = \{0\}$, and $E_1 = e \circ E_1 = \{0\}$. So $E_1 = D_i$ for some i. In the same way ones shows that every E_j equals D_i for some i. The proposition is proved. ■

The previous proposition immediately implies that any Euclidean Jordan algebra is, in a unique way, a direct sum of simple Euclidean Jordan algebras.

A symmetric cone $\mathcal{K}$ in a Euclidean space V is said to be *primitive* if there do not exist non-trivial subspaces V_1, V_2, and symmetric cones $\mathcal{K}_1 \subset V_1$, $\mathcal{K}_2 \subset V_2$ such that V is the direct sum of V_1 and V_2, and $\mathcal{K} = \mathcal{K}_1 \oplus \mathcal{K}_2$.

Proposition 2.7.4. *Any symmetric cone $\mathcal{K}$ is, in a unique way, the direct sum of primitive symmetric cones.*

Proof. Let V be an Euclidean Jordan algebra and $\mathcal{K}$ its associated symmetric cone. By Proposition 2.7.3 we can write

$$V = V_1 \oplus \cdots \oplus V_m.$$

Let $\mathcal{K}_i$ be the symmetric cone associated to V_i, with $i = 1, \ldots, r$. Since $x_i \circ x_j = 0$ for any $x_i \in V_i$ and $x_j \in V_j$ with $i \neq j$, we have

$$x_1^2 + \cdots + x_m^2 = (x_1 + \cdots + x_m)^2.$$

Hence

$$\mathcal{K} = \mathcal{K}_1 \oplus \cdots \oplus \mathcal{K}_m.$$

The proposition is proved. ■

We now are ready to give a fundamental result: it states that there are five simple Euclidean Jordan algebras and consequently there are five primitive symmetric cones. The proof is quite extensive and beyond the scope of this thesis. We refer to Chapter V in [15].

Theorem 2.7.5. *If V is a simple Euclidean Jordan algebra then it is isomorphic to one of the following:*

(i) The space $\mathbb{R}^{n+1}$ *with Jordan multiplication defined as follows: let* $x := (x_0; \bar{x})$ *and* $y := (y_0; \bar{y})$ *with* $\bar{x}, \bar{y} \in \mathbb{R}^n$, *and* $x_0, y_0 \in \mathbb{R}$, *then* $x \circ y := (x^T y; x_0\bar{y} + y_0\bar{x})$.

(ii) The set of real symmetric $n \times n$ *matrices with "*$\circ$*" defined by*

$$X \circ Y := (XY + YX)/2 \tag{2.22}$$

for symmetric matrices X *and* Y.

(iii) The set of complex Hermitian $n \times n$ *matrices with "*$\circ$*" defined by* (2.22).

(iv) The set of Hermitian $n \times n$ *matrices with quaternion entries and with "*$\circ$*" defined by* (2.22).

(v) The set of 3×3 *Hermitian matrices with octonion entries and with "*$\circ$*" defined by* (2.22).

As we mentioned before, the cone associated to the simple Euclidean Jordan algebra defined in (i) of Theorem 2.7.5 is the second order cone. For $A = \mathbb{R}, \mathbb{C}, \mathbb{H}$, we denote $\mathrm{Herm}(r, A)$ the vector space of Hermitian matrices with entries in A and rank r, where $\mathbb{H}$ denote the set of quaternions (for details see Section C.1). The primitive symmetric cones associated to these, are the cones of positive semidefinite matrices with entries in A. The cone associated to $\mathrm{Herm}(3, \mathbb{O})$, where $\mathbb{O}$ denotes the set of the octotions (see Section C.3) is the cone of 3×3 Hermitian positive semidefinite matrices with octonions entries. This cone is often referred to as "the exceptional 27-dimensional cone".

Note that the non-negative orthant $\mathbb{R}^n_+$ that appeared in several examples before, may be written as the direct sum of n copies of $\mathcal{K}(\mathrm{Herm}(1, \mathbb{R}))$, i.e.,

$$\mathbb{R}^n_+ = \mathcal{K}(\mathrm{Herm}(1, \mathbb{R})) \oplus \cdots \oplus \mathcal{K}(\mathrm{Herm}(1, \mathbb{R})).$$

Remark 2.7.6. Suppose that we can write V as direct sum of two simple Euclidean Jordan algebras, i.e, $V = V_1 \oplus V_2$. Let $x = x_1 + x_2 \in V$ and $y = y_1 + y_2 \in V$, with $x_1, y_1 \in V_1$ and $x_2, y_2 \in V_2$. Consider the spectral decompositions $x_1 = \sum_{i=1}^q \lambda_i(x_1)c_i$ and $x_2 = \sum_{i=1}^s \lambda_i(x_2)d_i$, where $c_1 \ldots, c_q$ and $d_1 \ldots, d_s$ are Jordan frames in V_1 and V_2, respectively. Note that $(c_1, \ldots, c_q, d_1, \ldots, d_s)$ is a Jordan frame in $V_1 \oplus V_2$. Then we have

$$x = \sum_{i=1}^{q} \lambda_i(x_1)c_i + \sum_{i=1}^{s} \lambda_i(x_2)d_i.$$

is a spectral decomposition of x. By Theorem 2.4.2 it follows that

$$\det(x) = \det(x_1)\det(x_2)$$

and

$$\mathrm{tr}\,(x) = \mathrm{tr}\,(x_1) + \mathrm{tr}\,(x_2)\,.$$

Moreover

$$\begin{aligned} P(x)y &= P(x_1+x_2)(y_1+y_2) \\ &= 2(x_1+x_2)\circ((x_1+x_2)\circ y_1)-(x_1+x_2)^2\circ y_1 \\ &\quad +2(x_1+x_2)\circ((x_1+x_2)\circ y_2)-(x_1+x_2)^2\circ y_2 \\ &= 2x_1\circ(x_1\circ y_1)-x_1^2\circ y_1+2x_2\circ(x_2\circ y_2)-x_2^2\circ y_2 \\ &= P(x_1)y_1+P(x_2)y_2, \end{aligned}$$

where the third equality follows by Proposition 2.7.3. □

2.8 Automorphisms

Let $(V,\circ)$ be a Euclidean Jordan algebra and $\mathcal{K}$ its associated symmetric cone. This section introduces some important properties of automorphisms of V and $\mathcal{K}$.

Recall that we denote the set of invertible linear maps from V to V by $\mathrm{GL}(V)$.

Definition 2.8.1. A map $g \in \mathrm{GL}(V)$ is called an *automorphism* of V if for every x and y in V, we have $g(x\circ y) = gx\circ gy$ or, equivalently, $gL(x)g^{-1} = L(gx)$. The set of automorphisms of V is denoted as $\mathrm{Aut}(V)$. □

In Section 2.5, we called $g \in \mathrm{GL}(V)$ an automorphism of the symmetric cone $\mathcal{K}$ if $g\mathcal{K} = \mathcal{K}$. We denote the automorphism group of a symmetric cone as $\mathrm{Aut}(\mathcal{K})$. So we have

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

We say that a linear map is *orthogonal* if $g^* = g^{-1}$. The set of *orthogonal automorphisms* that leave $\mathcal{K}$ invariant is denoted as $\mathrm{OAut}(\mathcal{K})$. So we have

$$\mathrm{OAut}(\mathcal{K}) = \{g \in \mathrm{GL}(V) : g^* = g^{-1} \text{ and } g\mathcal{K} = \mathcal{K}\}.$$

We would like to stress that $\mathrm{Aut}(V) \neq \mathrm{Aut}(\mathcal{K})$. For example, an element g of $\mathrm{Aut}(\mathcal{K})$ may not satisfy $g(x\circ y) = gx\circ gy$, since $\mathcal{K}$ is not closed under the Jordan product operation.

Proposition 2.8.2. *If* $g \in \mathrm{Aut}(\mathcal{K})$ *then* $g^* \in \mathrm{Aut}(\mathcal{K})$.

Proof. Let $g \in \mathrm{Aut}(\mathcal{K})$ and $x \in \mathcal{K}$, then for all $y \in \mathcal{K}$,

$$\langle x, g^*y\rangle = \langle gx, y\rangle \geq 0,$$

because $gx, y \in \mathcal{K}$ and $\mathcal{K}$ is self-dual. Therefore $g^*y \in \mathcal{K}$ which implies $g^* \in \mathrm{Aut}(\mathcal{K})$. ■

Proposition 2.8.3 (Proposition II.4.2 in [15])**.** *The trace and the determinant are invariant under* $\mathrm{Aut}(V)$*.*

Proof. If g is an automorphism of V and p is a polynomial in $\mathbb{R}[X]$, then $p(gx) = gp(x)$, and $p(gx) = 0$ if and only if $p(x) = 0$. Therefore, the minimal polynomial of x and gx are the same. It follows that x and gx have the same trace and determinant. ■

The next theorem establish a connection between the automorphism group of a symmetric cone and the automorphism group of a Jordan algebra.

Theorem 2.8.4. *We have*

$$\mathrm{Aut}(V) = \mathrm{OAut}(\mathcal{K}).$$

Proof. To prove that $\mathrm{OAut}(\mathcal{K})$ is a subset of $\mathrm{Aut}(V)$ we refer to [15, Theorem III.5.1], because it would need details that are far beyond this introduction to Euclidean Jordan algebras and symmetric cones. Anyway, we prove that $\mathrm{Aut}(V) \subset \mathrm{OAut}(\mathcal{K})$. Let $g \in \mathrm{Aut}(V)$ then for any $x \in V$, $gx^2 = (gx)^2 \in \mathcal{K}$, which implies that $g\mathcal{K} = \mathcal{K}$. Furthermore,

$$\langle gx, y\rangle = \mathrm{tr}\,(gx \circ y) = \mathrm{tr}\left(g(x \circ g^{-1}y)\right) = \mathrm{tr}\left(x \circ g^{-1}y\right) = \langle x, g^{-1}y\rangle.$$

Note that, the third equality follows from Proposition 2.8.3. Thus, $g^* = g^{-1}$. ■

Theorem 2.8.5 (Theorem IV.2.5 in [15])**.** *If* $\{c_1, \ldots, c_r\}$ *and* $\{d_1, \ldots, d_r\}$ *are two Jordan frames, then there exists* $g \in \mathrm{Aut}(V)$ *such that*

$$gc_i = d_i \quad (1 \le i \le r).$$

2.9 The Peirce decomposition in a Jordan algebra

Let c be an idempotent element in a Jordan algebra V. Note that we do not assume that V is Euclidean. By Proposition 2.4.6 the only possible eigenvalues of $L(c)$ are 1, $\frac{1}{2}$ and 0. One introduces the subspaces

$$V(c,\lambda) := \{x \in V : c \circ x = \lambda x\}, \qquad \lambda \in \{0, \tfrac{1}{2}, 1\}.$$

Each of these subspaces is the eigenspace corresponding to an eigenvalue of $L(c)$, because if $x \in V(c,\lambda)$ then $L(c)x = c \circ x = \lambda x$. Hence V is the direct sum of the corresponding subspaces $V(c,1)$, $V(c,\frac{1}{2})$ and $V(c,0)$. The decomposition

$$V = V(c,1) \oplus V(c,\tfrac{1}{2}) \oplus V(c,0) \tag{2.23}$$

is called the *Peirce decomposition* of V with respect to the idempotent c. If $x \in V(c,1)$ and $y \in V(c,\frac{1}{2})$ then $L(c)x = x$ and $L(c)y = \frac{1}{2}y$. Hence

$$\langle x, y\rangle = \langle L(c)x, y\rangle = \langle x, L(c)y\rangle = \langle x, \tfrac{1}{2}y\rangle,$$

and it follows that $\langle x, y\rangle = 0$. Therefore the decomposition (2.23) is orthogonal with respect to the inner product.

Example 2.9.1. Let $(V, \circ)$ be the Jordan algebra $(\mathbf{S}^n, \circ)$ defined in Example 2.2.7. Let $n = p + q$ then

$$c := \begin{bmatrix} I_p & 0 \\ 0 & 0 \end{bmatrix}$$

is a orthogonal projection from $\mathbb{R}^n$ to $\mathbb{R}^n$ and hence an idempotent. We now have,

$$\begin{aligned} V(c,1) &= \left\{ \begin{bmatrix} A & 0 \\ 0 & 0 \end{bmatrix} : A \text{ is a } p \times p \text{ symmetric matrix} \right\}, \\ V(c,0) &= \left\{ \begin{bmatrix} 0 & 0 \\ 0 & B \end{bmatrix} : B \text{ is a } q \times q \text{ symmetric matrix} \right\}, \\ V(c,\tfrac{1}{2}) &= \left\{ \begin{bmatrix} 0 & D \\ D^T & 0 \end{bmatrix} : D \text{ is a } p \times q \text{ matrix} \right\} \end{aligned}$$

and $p + q = n$. □

Proposition 2.9.2. *If $x \in V(c, 0)$ or $x \in V(c, 1)$ then $[L(x), L(c)] = 0$*

Proof. If we replace y and z in identity *(ii)* of the Proposition 2.2.2, we get

$$[L(x), L(c)] + [L(c), L(c \circ x)] + [L(c), L(x \circ c)] = 0.$$

Since $x \in V(c, 0)$ or $x \in V(c, 1)$ means that $x \circ c = 0$ or $x \circ c = x$, the result follows. ■

Proposition 2.9.3 (Proposition IV.1.1 in [15]). *Let V be a Jordan algebra and c an idempotent, the subspaces $V(c, 1)$ and $V(c, 0)$ are subalgebras of V. They are orthogonal in the sense that*

$$V(c, 1) \circ V(c, 0) = \{0\}.$$

Furthermore,

$$\begin{aligned} (V(c,1) + V(c,0)) \circ V(c,\tfrac{1}{2}) &\subset V(c,\tfrac{1}{2}), \\ V(c,\tfrac{1}{2}) \circ V(c,\tfrac{1}{2}) &\subset V(c,1) + V(c,0). \end{aligned}$$

Proof. Using identity (iii) of Proposition 2.2.2, with c instead of x and x instead of y, we obtain, for $x \in V(c, \lambda)$,

$$(L(c) - \lambda I)L(x)(2L(c) - I) = 0,$$

and applying it to $y \in V(c, \mu)$ $(\lambda, \mu = 0, \frac{1}{2}, 1)$, it follows

$$(2\mu - 1)(L(c) - \lambda I)(x \circ y) = 0.$$

If $\mu = 0$ or 1, then

$$L(c)(x \circ y) = \lambda(x \circ y),$$

which means that $(x \circ y)$ belongs to $V(c, \lambda)$. Taking $\lambda = 0$ or 1, and $\mu = \lambda$ this proves that $V(c, 1)$ and $V(c, 0)$ are subalgebras of V. Taking again $\lambda = 0$ or 1 and $\mu = 0$ or 1 but $\mu \neq \lambda$ and using Proposition 2.9.2 this proves that $V(c, 1)$ and $V(c, 0)$ are orthogonal. Again, for $\mu = 0$ or 1 and taking $\lambda = \frac{1}{2}$, we have

$$L(c)(x \circ y) = \tfrac{1}{2}(x \circ y),$$

which means that $x \circ y \in V(c, \frac{1}{2})$. Therefore

$$V(c, 1) \circ V(c, \tfrac{1}{2}) \subset V(c, \tfrac{1}{2}),$$

$$V(c, 0) \circ V(c, \tfrac{1}{2}) \subset V(c, \tfrac{1}{2})$$

and it follows

$$(V(c, 1) + V(c, 0)) \circ V(c, \tfrac{1}{2}) \subset V(c, \frac{1}{2}).$$

For the last assertion it is enough to show that if x belongs to $V(c, \frac{1}{2})$, then x^2 belongs to $V(c, 1) + V(c, 0)$, that is,

$$x^2 = a_0 + a_1,$$

with $a_0 \in V(c, 0)$ and $a_1 \in V(c, 1)$. In fact, with

$$a_0 = x^2 - c \circ x^2, \ a_1 = c \circ x^2,$$

we have

$$\begin{aligned} L(c)a_0 &= L(c)x^2 - L(c)L(c)x^2 \\ &= (I - L(c))a_1 \\ &= (L(x^2)L(c) - L(x^2 \circ c))c, \end{aligned}$$

and using identity (iii) of Proposition 2.2.2:

$$\begin{aligned} L(c)a_0 &= 2(L(x)L(c) - L(x \circ c))L(x)c \\ &= (L(x)L(c) - \tfrac{1}{2}L(x))x \\ &= L(x)(L(c) - \tfrac{1}{2}I)x = 0, \end{aligned}$$

because $x \in V(c, \frac{1}{2})$. Hence $a_0 \in V(c, 0)$. From

$$0 = L(c)a_0 = (I - L(c))a_1$$

follows that $L(c)a_1 = a_1$. To finish, note that for $x, y \in V(c, \frac{1}{2})$ we have

$$x^2, y^2, (x + y)^2 \in V(c, 0) + V(c, 1).$$

Since $V(c, 0) + V(c, 1)$ is vector space it follows that

$$x \circ y = \frac{1}{2}((x + y)^2 - x^2 - y^2) \in V(c, 0) + V(c, 1).$$

Everything is proved. ∎

Remark 2.9.4. From Proposition 2.9.3 and its proof we get that $V(c,0)$ and $V(c,1)$ are Jordan algebras. □

We say that a linear map $Q : V \mapsto V$ is an *orthogonal projection* if $Q^2 = Q$ and Q is self-adjoint, i.e., $Q^* = Q$. We say that two orthogonal projections, Q_1 and Q_2, are orthogonal with respect to each other if $Q_1Q_2 = 0$.

Proposition 2.9.5. *The orthogonal projections onto $V(c,1)$, $V(c,0)$ and $V(c,\frac{1}{2})$ are respectively $P(c)$, $P(e-c)$ and $I - P(c) - P(e-c)$. Moreover they are orthogonal with respect to each other.*

Sketch of the proof. We will just show that $P(c)$ is a orthogonal projection onto $V(c,1)$. The other cases are similar. Clearly $P(c)^2 = P(c^2) = P(c)$. $P(c)$ is self-adjoint. Let $y \in V$ and $y = y_1 + y_{\frac{1}{2}} + y_0$ its Peirce decomposition with respect to c, where $y_1 \in V(c,1)$, $y_{\frac{1}{2}} \in V(c,\frac{1}{2})$ and $y_0 \in V(c,0)$. Since $P(c)$ and $L(c)$ commute (see (2.8)), we have

$$L(c)P(c)y = P(c)L(c)y.$$

Thus,

$$P(c)L(c)y = P(c)(y_1 + \tfrac{1}{2}y_{\frac{1}{2}}) = 2y_1 - y_1 + \tfrac{1}{4}y_{\frac{1}{2}} - \tfrac{1}{4}y_{\frac{1}{2}} = y_1.$$

This proves that $P(c)y$ belongs to $V(c,1)$. Clearly, for $y \in V(c,1)$ we have $P(c)y = y$. For the orthogonality of the projections with respect to each other, it is just enough to verify that

$$P(c)P(e-c) = 0.$$

We can easily see that

$$P(c) = 2L^2(c) - L(c^2) = L(c)(2L(c) - I)$$

and

$$P(e-c) = (L(c) - I)(2L(c) - I).$$

Therefore,

$$\begin{aligned} P(c)P(e-c) &= (L(c)(2L(c) - I)(L(c) - I))(2L(c) - I) \\ &= (2L^3(c) - 3L(c) + L(c))(2L(c) - I). \end{aligned}$$

Since

$$2L^3(c) - 3L(c) + L(c) = 0,$$

by the proof of Proposition 2.4.6, it follows that the projections are orthogonal with respect to each other. For the other cases we just have to proceed analogously. ■

Lemma 2.9.6 (Lemma IV.1.3 in [15])**.** *If a and b are orthogonal idempotents, then $L(a)$ and $L(b)$ commute.*

Proof. This follows from identity (i) in Proposition 2.2.2 ■

We just gave a decomposition of the vector space V with respect to an idempotent. Now, we are able to get a finer decomposition, using a Jordan frame $\{c_1, \ldots, c_r\}$ of the Euclidean Jordan algebra $(V, \circ, \langle \cdot, \cdot \rangle)$.

For each c_i we have three Peirce spaces, $V(c_i, 0)$, $V(c_i, 1)$, $V(c_i, \frac{1}{2})$. We consider the following subspaces of V,

$$\begin{aligned} V_{ii} &:= V(c_i, 1) = \mathbb{R}c_i, \\ V_{ij} &:= V(c_i, \tfrac{1}{2}) \cap V(c_j, \tfrac{1}{2}), \end{aligned}$$

where the set $\mathbb{R}c_i := \{\alpha c_i : \alpha \in \mathbb{R}\}$. The equality $V_{ii} = \mathbb{R}c_i$ immediately follows: for all $y \in V_{ii}$ we have $y \circ c_i = y$, which means that c_i is the identity element in V_{ii}. But since c_i is primitive, i.e., it can not be written as the sum of two orthogonal idempotents, c_i is the only element of the only Jordan frame in V_{ii}. We conclude that, any element of V_{ii} can be written as αc_i, with $\alpha \in \mathbb{R}$.

We define

$$P_{ii} = P(c_i) \tag{2.24}$$

$$P_{ij} = 4L(c_i)L(c_j), \ i \neq j \tag{2.25}$$

Proposition 2.9.7. *The endomorphisms P_{ij} are orthogonal projections onto V_{ij} and they are orthogonal with respect to each other.*

Proof. We have seen in Proposition 2.9.5 that P_{ii} are orthogonal projections onto V_{ii}. Suppose now that $i \neq j$. If we replace x and y by c_i and c_j respectively, in the third item of Proposition 2.2.2, we get

$$2L(c_i)L(c_j)L(c_i) = L(c_i)L(c_j). \tag{2.26}$$

Then we have

$$P_{ij}^2 = 16L(c_i)L(c_j)L(c_i)L(c_j) = 8L(c_i)L(c_j)L(c_j) = 4L(c_i)L(c_j) = P_{ij}.$$

Since $L(x)$ for any $x \in V$ is self-adjoint then P_{ij} is self-adjoint. Let

$$W_{ij} := P_{ij}V = \{P_{ij}x : x \in V\}.$$

We want to prove that $W_{ij} = V_{ij}$. Let $y \in W_{ij}$, then there exists $x \in V$ such that $y = P_{ij}x$. We have,

$$L(c_i)y = 4L(c_i)L(c_i)L(c_j)x.$$

Therefore,

$$L(c_i)y = 2L(c_i)L(c_j)x = \frac{1}{2}y,$$

which means that $y \in V(c_i, \frac{1}{2})$. Analogously, we can prove that $y \in V(c_j, \frac{1}{2})$. Hence $y \in V_{ij}$. Now suppose that $y \in V_{ij}$, then easily

$$L(c_i)L(c_j)y = \frac{1}{4}y,$$

implying that $y \in W_{ij}$. Proving that the projections are orthogonal with respect to each other, resumes to prove what follows. For $i \neq j$, we have

$$\begin{aligned} P(c_i)P(c_j) &= (2L^2(c_i) - L(c_i))(2L^2(c_j) - L(c_j)) \\ &= 4L^2(c_i)L^2(c_j) - 2L^2(c_i)L(c_j) - 2L^2(c_j)L(c_i) + L(c_j)L(c_i) \\ &= 0, \end{aligned}$$

where the last equality follows by (2.26). Now, for $i \neq j$ and using (2.26) it follows

$$P_{kk}P_{ij} = P(c_k)4L(c_i)L(c_j) = (2L^2(c_k) - L(c_k))4L(c_i)L(c_j) = 0.$$

The last different case, for i, j and k all distinct, is

$$P_{ki}P_{ij} = 4L(c_k)L(c_i)4L(c_i)L(c_j) = 8L(c_k)L(c_i)L(c_j).$$

By the identity (2.26) we have

$$2L^2(c_i + c_j)L(c_k) = L(c_i + c_j)L(c_k),$$

which is equivalent to

$$2L(c_i)L(c_j)L(c_k) + 2L^2(c_i)L(c_k) + 2L^2(c_j)L(c_k) = L(c_i)L(c_k) + L(c_j)L(c_k).$$

From this, using (2.26), we conclude that $L(c_i)L(c_j)L(c_k) = 0$ and consequently $P_{ki}P_{ij} = 0$. Finally, the remaining cases follow from the ones we have proved. ■

The orthogonal projections defined in Proposition 2.9.5 may be written, after simple calculations, as follows:

$$\begin{aligned} P(c) &= L(c)(2L(c) - I), \\ P(e - c) &= (L(c) - I)(2L(c) - I), \\ I - P(c) - P(e - c) &= 4L(c)(I - L(c)), \end{aligned}$$

and $c \in V$ is an idempotent.

Thus, with respect to the Jordan frame, one can give a finer decomposition.

Theorem 2.9.8. *The vector space V has the following orthogonal direct sum decomposition:*

$$V = \bigoplus_{i \leq j} V_{ij}.$$

Proof. Let us fix a positive integer k such that $1 \leq k \leq r$. Using the fact that $e - c_k$ is an idempotent and $c_k = e - \sum_{i \neq k} c_i$, we can decompose the orthogonal projection

$P(e-c_k)$. Hence, by simple calculations:

$$\begin{aligned} P(e-c_k) &= (L(c_k)-I)(2L(c_k)-I) \\ &= \left(-\sum_{i\neq k} L(c_i)\right)\left(I-2\sum_{i\neq k} L(c_i)\right) \\ &= 2\sum_{i\neq k}\sum_{j\neq k} L(c_i)L(c_j) - \sum_{i\neq k} L(c_i) \\ &= \sum_{i<j,\ i,j\neq k} 4L(c_i)L(c_j) + \sum_{i\neq k} 2L^2(c_i) - L(c_i) \\ &= \sum_{i<j\ i,j\neq k} P_{ij} + \sum_{i\neq k} P(c_i). \end{aligned}$$

This means, by Propositions 2.9.5 and 2.9.7, that we can decompose

$$V(c_k,0) = \sum_{i<j,\ i,j\neq k} V_{ij} + \sum_{i\neq k} V_{ii}.$$

Analogously,

$$I - P(c_k) - P(e-c_k) = 4L(c_k)(I-L(c_k)) \tag{2.27}$$

$$= 4L(c_k)\sum_{i\neq k} L(c_i) \tag{2.28}$$

$$= \sum_{i\neq k} 4L(c_k)L(c_i). \tag{2.29}$$

Again, we have

$$V(c_k,\frac{1}{2}) = \sum_{i\neq k} V_{ki}.$$

Since, we have seen that

$$V = V(c_k,1) + V(c_k,\frac{1}{2}) + V(c_k,0),$$

we can now conclude that

$$V = \sum_{i\leq j} V_{ij}.$$

The Proposition 2.9.7 implies that the sum is direct. ■

Proposition 2.9.9 (Theorem IV.2.1 in [15]). *One has*

$$\begin{aligned} &V_{ij}\circ V_{ij} \subset V_{ii} + V_{jj}, \\ &V_{ij}\circ V_{jk} \subset V_{ik},\ \textit{if}\ i\neq k, \\ &V_{ij}\circ V_{kl} = \{0\},\ \textit{if}\ \{i,j\}\cap\{k,l\} = \emptyset. \end{aligned}$$

Sketch of the proof. By Proposition 2.9.3:

$$V_{ij} \circ V_{ij} \subset V_{ii} + V_{jj}.$$

For i, j, k distinct:

$$V_{ij} \subset V(c_i + c_j, 1) \subset V(c_k, 0),$$

therefore it follows from Proposition 2.9.3 that

$$V_{ij} \circ V_{jk} \subset V(c_k, 0) \circ V(c_k, \frac{1}{2}) \subset V(c_k, \frac{1}{2}),$$

and, interchanging i and k, we obtain

$$V_{ij} \circ V_{jk} \subset V_{ik}.$$

Finally, if $\{i, j\} \cap \{k, \ell\} = \emptyset$, then

$$V_{ij} \subset V(c_i + c_j, 1),$$

$$V_{k\ell} \subset V(c_k + c_\ell, 1) \subset V(c_i + c_j, 0),$$

and from Proposition 2.9.3 it follows that

$$V_{ij} \circ V_{k\ell} = \{0\}.$$

Everything is proved. ■

Proposition 2.9.10. *For $x \in V_{ij}$ with $i \neq j$ we have* $\mathrm{tr}\,(x) = 0$.

Proof. If $x \in V_{ij}$ then $x = 4L(c_i)L(c_j)x$. Hence, by associativity of the trace,

$$\mathrm{tr}\,(x) = 4\mathrm{tr}\,(c_i \circ (c_j \circ x)) = \mathrm{tr}\,((c_i \circ c_j) \circ x) = 0,$$

and the property is proved. ■

In view of the Theorem 2.9.8 we can decompose

$$x = \sum_{i=1}^{r} x_{ii} + \sum_{i<j} x_{ij},$$

with $x_{ij} \in V_{ij}$ and $i \leq j$ or equivalently,

$$x = \sum_{i=1}^{r} x_i c_i + \sum_{i<j} x_{ij} \tag{2.30}$$

with $x_i \in \mathbb{R}$, $i = 1, \ldots, r$ and $x_{ij} \in V_{ij}$, $i < j$. We call it the Peirce decomposition of x with respect to the Jordan frame $c_1, \ldots, c_r$.

Recall from Section 2.1 that the eigenvalues of x are the same as those of $L_0(x)$. But what can be said on the eigenvalues of $L(x)$, and of $P(x)$? A complete answer to this question is given by the following result.

Proposition 2.9.11. *For $x \in V$ and $x = \sum_{k=1}^r \lambda_k c_k$ its spectral decomposition we have*

$$L(x) = \sum_{i=1}^r \lambda_i P_{ii} + \sum_{i<j} \frac{\lambda_i + \lambda_j}{2} P_{ij},$$

$$P(x) = \sum_{i=1}^r \lambda_i^2 P_{ii} + \sum_{i<j} \lambda_i \lambda_j P_{ij}$$

Proof. Since, for $y \in V$, we can write $y = \sum_{i \le j} y_{ij}$ with $y_{ij} \in V_{ij}$ we also have $L(c_i)y_{ii} = y_{ii}$, $L(c_k)y_{ii} = 0$ for $k \ne i$, $L(c_i)y_{ij} = \frac{1}{2}y_{ij}$, $L(c_j)y_{ij} = \frac{1}{2}y_{ij}$ for $i \ne j$ and $L(c_k)y_{ij} = 0$ for $k \notin \{i, j\}$. Hence

$$\begin{aligned} L(x)y &= \sum_{k=1}^r \lambda_k L(c_k) y \\ &= \sum_{i=1}^r \sum_{k=1}^r \lambda_k L(c_k) y_i + \sum_{i<j} \sum_{k=1}^r \lambda_k L(c_k) y_{ij} \\ &= \sum_{i=1}^r \lambda_i y_i + \sum_{i<j} \lambda_i \frac{1}{2} y_{ij} + \lambda_j \frac{1}{2} y_{ij} \\ &= \sum_{i=1}^r \lambda_i P_{ii} y + \sum_{i<j} \frac{1}{2}(\lambda_i + \lambda_j) P_{ij} y. \end{aligned}$$

It follows that

$$L(x) = \sum_{i=1}^r \lambda_i P_{ii} + \sum_{i<j} \frac{1}{2}(\lambda_i + \lambda_j) P_{ij}.$$

The second expression of the theorem we easily get:

$$\begin{aligned} P(x) &= 2L(x)^2 - L(x^2) \\ &= 2\big(\sum_i \lambda_i L(c_i)\big)^2 - \sum \lambda_i^2 L(c_i) \\ &= \sum \lambda_i^2 \big(2L(c_i)^2 - L(c_i)\big) + 4 \sum_{i<j} \lambda_i \lambda_j L(c_i) L(c_j) \\ &= \sum_i \lambda_i^2 P_{ii} + \sum_{i<j} \lambda_i \lambda_j P_{ij}. \end{aligned}$$

Everything is proved. ∎

Corollary 2.9.12. *Let $x \in V$ and its spectral decomposition $x = \sum_{i=1}^r \lambda_i c_i$. Then the following statements hold.*

1. The eigenvalues of $L(x)$ have the form

$$\frac{\lambda_i + \lambda_j}{2} \qquad 1 \le i \le j \le r.$$

2. The eigenvalues of $P(x)$ have the form

$$\lambda_i\lambda_j, \qquad 1 \leq i \leq j \leq r.$$

2.10 Conclusion

The material presented in this chapter has been developed by algebraists and already discovered by some optimizers. However, most members of the optimization community are not familiar with it. We therefore intended to give a simple and not too deep exposition of Euclidean Jordan algebras and the associated symmetric cones.

The results without a reference, were composed by the author. In the other cases, if we included a proof, it was just the proof provided by the indicated reference.

In the next chapter we will frequently refer to the results of this chapter.

Chapter 3

Eigenvalues, spectral functions and their derivatives

3.1 Introduction

Eigenvalues are the key parameters when we generalize the interior-point methods for linear optimization to symmetric optimization. The reason is that the barrier functions can be written as a function that only depends on the eigenvalues of its argument. We will explore this characterization of the barrier functions in later chapters. Therefore, in this chapter we present some similarity properties and inequalities for eigenvalues.

The barrier functions used in interior-point methods are a subclass of spectral functions: functions which only depend on the eigenvalues of its argument. Thus we present some aspects of spectral functions, including their derivatives.

As before, we consider a n-dimensional Euclidean Jordan $\mathbb{R}$-algebra $(V, \circ, \langle\cdot,\cdot\rangle)$ with rank r and $\mathcal{K}$ will denote the associated symmetric cone, and the inner product is defined by $\langle x, y\rangle = \operatorname{tr}(x \circ y)$. For any x in V, let $\lambda(x) \in \mathbb{R}^r$ be the vector of the eigenvalues of x. We assume that the eigenvalues are always in non-increasing order, i.e., $\lambda_1(x) \geq \lambda_2(x) \geq \cdots \geq \lambda_r(x)$ and we denote this as $\lambda(x) \in \mathbb{R}^r_{\downarrow}$.

3.2 Similarity

One says that two elements x and y in V are *similar*, denoted as $x \sim y$, if and only if

$$x = gy \text{ for some } g \in \operatorname{OAut}(\mathcal{K}), \tag{3.1}$$

where $\operatorname{OAut}(\mathcal{K})$ stands for the set of orthogonal automorphisms that leave $\mathcal{K}$ invariant (c.f. [48]). Recall from Theorem 2.8.4 and Proposition 2.8.3 that x and gx have the same characteristic polynomial, so we can say that x and y are similar if and only if they share the same eigenvalues, including their multiplicities.

The following proposition establishes another characterization of similarity, using the linear map $L(x)$, $x \in V$.

Proposition 3.2.1 (Proposition 19 in [47]). *Two elements x and y of an Euclidean Jordan algebra are similar if and only if $L(x)$ and $L(y)$ are similar.*

Proof. As the eigenvalues of $L(x)$ are determined by the eigenvalues of x (see Corollary 2.9.12), the "only if" part is true.

Now assume x and y are not similar. Let $\lambda(x), \lambda(y) \in \mathbb{R}^r_{\downarrow}$. Let k be the smallest index such that $\lambda_k(x) \neq \lambda_k(y)$. W.l.o.g we can assume $\lambda_k(x) > \lambda_k(y)$. If $k = 1$, then $\lambda_k(x)$ is larger than all eigenvalues of $L(y)$ but it is an eigenvalue of $L(x)$. So $L(x)$ and $L(y)$ are not similar. In other case, we have $k > 1$ and $\lambda_i(x) = \lambda_i(y)$ for $i < k$. So the multiplicities for eigenvalues larger than $\lambda_k(x)$ are the same in $\lambda(x)$ and $\lambda(y)$. If $\lambda_k(x)$ is not eigenvalue of y then also $(\lambda_k(x) + \lambda_1(x))/2$ is not eigenvalue of $L(y)$. If $\lambda_k(x)$ is eigenvalue of y then its multiplicity is larger as eigenvalue of x than as eigenvalue of y. Hence the multiplicity of $(\lambda_k(x) + \lambda_1(x))/2$ is larger in the set of eigenvalues of $L(x)$ than of $L(y)$. Therefore $L(x)$ and $L(y)$ are not similar. This completes the proof. ■

In the following proposition we use the same arguments as in the proof of Proposition 3.2.1, as long as the eigenvalues of x and y are positive. If the eigenvalues are not positive, $P(x)$ and $P(y)$ may be similar but x and y do not. As example, let λ_1, λ_2 and $-\lambda_1, -\lambda_2$ be the eigenvalues of x and y, respectively. By Corollary 2.9.12, the eigenvalues of $P(x)$ and $P(y)$ are $\lambda_1^2, \lambda_1\lambda_2, \lambda_2^2$. Thus $P(x)$ and $P(y)$ are similar in opposition to x and y.

Proposition 3.2.2 (Corollary 20 in [47]). *Let x and y be two elements in the interior of $\mathcal{K}$. Then x and y are similar if and only if their quadratic representations $P(x)$ and $P(y)$ are similar.*

In the following we present more technical properties concerning similarity.

Proposition 3.2.3 (Proposition 21 in [47]). *Let V be Euclidean Jordan algebras and $x, s, z \in \mathcal{K}^0$. Define $\tilde{x} := P(z)x$ and $\tilde{s} := P(z^{-1})s$. Then*

(i) $P(x^{1/2})s \sim P(s^{1/2})x$

(ii) $P(x^{1/2})s \sim P(\tilde{x}^{1/2})\tilde{s}$

Proof. First note that by Proposition 2.5.7, $P(s^{1/2})x, P(x^{1/2})s \in \mathcal{K}^0$. By the fundamental formula we have

$$P(P(x^{1/2})s) = P(x)^{1/2}P(s)P(x)^{1/2},$$

which by Proposition B.1.1 is similar to

$$P(s)^{1/2}P(x)P(s)^{1/2} = P(P(s^{1/2})x)$$

and *(i)* follows from Proposition 3.2.2. Again, using the fundamental formula, we have

$$P(P(\tilde{x}^{1/2})\tilde{s}) = P(\tilde{x}^{1/2})P(\tilde{s})P(\tilde{x}^{1/2}),$$

which is similar to

$$P(\tilde{x})P(\tilde{s}).$$

Replacing $\tilde{x}$ and $\tilde{s}$, according to their definition, we obtain

$$\begin{aligned} P(\tilde{x})P(\tilde{s}) &= P(P(z)x)P(P(z^{-1})s) \\ &= P(z)P(x)P(z)P(z^{-1})P(s)P(z^{-1}) \\ &= P(z)P(x)P(s)P(z)^{-1}. \end{aligned}$$

Hence, $P(\tilde{x})P(\tilde{s})$ is similar to $P(x)P(s)$. Since $P(\tilde{x})P(\tilde{s})$ is similar to $P(\tilde{x}^{\frac{1}{2}})\tilde{s}$ and $P(x)P(s)$ is similar to $P(x^{1/2})s$, statement *(ii)* follows. ■

The following proposition is a new result. It plays an important role in the analysis of the algorithm that will be presented in Chapter 5.

Proposition 3.2.4. *Let $x, s \in \mathcal{K}^0$ and $w := P(x^{\frac{1}{2}})(P(x^{\frac{1}{2}})s)^{-\frac{1}{2}}$. Then*

$$(P(x^{1/2})s)^{1/2} \sim P(w)^{1/2}s.$$

Proof. By Proposition 3.2.2, the statement is equivalent to

$$P(P(x^{1/2})s)^{1/2} \sim P(P(w)^{1/2}s).$$

Therefore, by the fundamental formula,

$$P(P(x^{1/2})s)^{1/2} = (P(x)^{1/2}P(s)P(x)^{1/2})^{1/2} \sim (P(x)P(s))^{1/2}.$$

On the other hand

$$P(P(w^{1/2})s) = P(w)^{1/2}P(s)P(w)^{1/2} \sim P(w)P(s).$$

Thus, we obtain

$$\begin{aligned} P(w)P(s) &= P(x)^{1/2}P(P(x)^{1/2}s)^{-1/2}P(x)^{1/2}P(s) \\ &= P(x)^{1/2}(P(x)^{1/2}P(s)P(x)^{1/2})^{-1/2}P(x)^{1/2}P(s) \\ &\sim (P(x)^{1/2}P(s)P(x)^{1/2})^{-1/2}P(x)^{1/2}P(s)P(x)^{1/2} \\ &= (P(x)^{1/2}P(s)P(x)^{1/2})^{1/2} \\ &\sim (P(x)P(s))^{1/2}, \end{aligned}$$

and the proposition follows. ■

The point w in the previous proposition is known as the scaling point corresponding to x and s (see [16]), i.e.,

$$P(w)^{-1/2}x = P(w)^{1/2}s. \tag{3.2}$$

Later, in Section 5.4 we approach this topic.

Corollary 3.2.5. *We have*

$$P(x^{1/2})s \sim P(w)^{1/2}s \circ P(w)^{-1/2}x.$$

Proof. This follows from Proposition 3.2.4 and (3.2). ■

3.3 An important eigenvalue inequality

In this section, we prove the following inequality: for $x, s \in \mathcal{K}$ and $u := P(x)^{1/2}s$,

$$\prod_{i=1}^{k} \lambda_i(u) \leq \prod_{i=1}^{k} \lambda_i(x)\lambda_i(s) \quad k = 1, \ldots, r-1, \tag{3.3}$$

The property (3.3) is a generalization of the Theorem B.1.2 to Euclidean Jordan algebras. This inequality is of crucial importance to establish e-convexity of the barrier functions.

We first relate the minimum eigenvalue and the maximum eigenvalue of u with the minimum and the maximum eigenvalue, respectively, of x and s. It means that the eigenvalues of u are less dispersed than the component-wise product $\lambda(x)\lambda(s)$, where $\lambda(x) \in \mathbb{R}^r_\downarrow$ is the vector of eigenvalues of x. We will need it to prove the inequality (3.3).

Proposition 3.3.1. *Let $x, s \in \mathcal{K}$ and $u := P(x)^{1/2}s$, then*

$$\lambda_{\max}(u) \leq \lambda_{\max}(x)\lambda_{\max}(s)$$

and

$$\lambda_{\min}(u) \geq \lambda_{\min}(x)\lambda_{\min}(s).$$

Proof. By the fundamental formula,

$$P(u) = P(x)^{1/2}P(s)P(x)^{1/2},$$

which is similar to $P(x)P(s)$. Therefore, by Theorem B.1.2

$$\lambda_{\max}(P(u)) \leq \lambda_{\max}(P(x))\lambda_{\max}(P(s)),$$

which implies, by Corollary 2.9.12, that

$$\lambda^2_{\max}(u) \leq \lambda^2_{\max}(x)\lambda^2_{\max}(s).$$

Thus, the first inequality follows. The second one follows analogously. ■

We will prove the inequality (3.3) first for each of the five simple Euclidean Jordan algebras characterized by Theorem 2.7.5. In fact, the inequality is known to be true for $\mathrm{Herm}(n,\mathbb{R})$ and for $\mathrm{Herm}(n,\mathbb{C})$ and, to our knowledge, is not yet known for $V = \mathrm{Herm}(n,\mathbb{H})$ or $V = \mathrm{Herm}(3,\mathbb{O})$.

Theorem 3.3.2. *Let V be a simple Euclidean Jordan algebra, and $\mathcal{K}$ its cones of squares. Then for $x, s \in \mathcal{K}$ and $u := P(x)^{1/2}s$,*

$$\prod_{i=1}^{k} \lambda_i(u) \leq \prod_{i=1}^{k} \lambda_i(x)\lambda_i(s) \quad k = 1, \dots, r-1$$
$$\prod_{i=1}^{r} \lambda_i(u) = \prod_{i=1}^{r} \lambda_i(x)\lambda_i(s).$$

Proof. First we focus on the inequality. The equality is handled at the end. Suppose that $(V, \circ)$ is the simple Euclidean Jordan algebra described in Theorem 2.7.5-(i). In Example 2.4.4 we have seen that the elements of this Jordan algebra has only two eigenvalues. Hence, the result follows from Proposition 3.3.1. If $V = \mathrm{Herm}(n,\mathbb{R}), \mathrm{Herm}(n,\mathbb{C})$ or $\mathrm{Herm}(n,\mathbb{H})$, and

$$x \circ s = \frac{xs + sx}{2}$$

then, by simple calculations,

$$u = P(x)^{1/2}s = x^{1/2}sx^{1/2}.$$

For $V = \mathrm{Herm}(n,\mathbb{R})$ or $\mathrm{Herm}(n,\mathbb{C})$, the result follows by Theorem B.1.2. In the case of $V = \mathrm{Herm}(n,\mathbb{H})$ the result follows by Theorem C.2.5.

If $V = \mathrm{Herm}(3,\mathbb{O})$ and $x \circ s = \frac{xs+sx}{2}$, any $x \in V$ has only three eigenvalues (see Appendix C.4). Since $\lambda_3(u) \geq \lambda_3(x)\lambda_3(s)$ (Proposition 3.3.1) and $\det(u) = \det(x)\det(s)$ (Proposition 2.5.12), we have

$$\lambda_1(u)\lambda_2(u) \leq \lambda_1(x)\lambda_2(x)\lambda_1(s)\lambda_2(s).$$

Since by Theorem 2.7.5 any simple Euclidean Jordan algebra is isomorphic to one of these five, the inequality holds for any simple Euclidean Jordan algebra and any $x, s \in \mathcal{K}$. The equality is a direct consequence of

$$\det(P(x)^{1/2}s) = \det(x)\det(s),$$

according to Proposition 2.5.12. ∎

Theorem 3.3.3. *The inequality* (3.3) *is valid for any Euclidean Jordan algebra.*

Proof. For simplicity of notation we will prove the inequality for x instead of $x^{1/2}$, but the result follows in the same way. Suppose for the sake of simplicity that the

Euclidean Jordan algebra V is just the direct sum of two simple Euclidean Jordan algebras V_1 and V_2, i.e $V = V_1 \oplus V_2$. Let $\mathcal{K}$, $\mathcal{K}_1$ and $\mathcal{K}_2$ the symmetric cones associated with V, V_1 and V_2 respectively, where obviously $\mathcal{K}_1$ and $\mathcal{K}_2$ are primitive. Let $x = x_1 + x_2 \in \mathcal{K}$ and $s = s_1 + s_2 \in \mathcal{K}$ with $x_1, s_1 \in \mathcal{K}_1$ and $x_2, s_2 \in \mathcal{K}_2$. By Remark 2.7.6 we have

$$u := P(x)s = P(x_1)s_1 + P(x_2)s_2.$$

Let $u_1 := P(x_1)s_1$ and $u_2 := P(x_2)s_2$. Again, as is clear from Remark 2.7.6, the eigenvalues of u are all the eigenvalues of u_1 and u_2. Let the k largest eigenvalues of u be given by the k_1 largest eigenvalues of u_1 and by the k_2 largest eigenvalues of u_2, where $k = k_1 + k_2$. Hence, by Theorem 3.3.2, we have

$$\prod_{i=1}^{k} \lambda_i(u) = \prod_{i=1}^{k_1} \lambda_i(u_1) \prod_{i=1}^{k_2} \lambda_i(u_2) \leq \prod_{i=1}^{k_1} \lambda_i^2(x_1)\lambda_i(s_1) \prod_{i=1}^{k_2} \lambda_i^2(x_2)\lambda_i(s_2).$$

Note that $\lambda_1(s_1) \geq \cdots \geq \lambda_{k_1}(s_1)$ are the k_1 largest eigenvalues of s_1 and $\lambda_1(s_2) \geq \cdots \geq \lambda_{k_2}(s_2)$ are the k_2 largest eigenvalues of s_2, but both sets of eigenvalues may not be the k largest eigenvalues of s. Since they are eigenvalues of s we certainly have that

$$\prod_{i=1}^{k_1} \lambda_i(s_1) \prod_{i=1}^{k_2} \lambda_i(s_2) \leq \prod_{i=1}^{k} \lambda_i(s).$$

The same is valid for the eigenvalues of x. Thus

$$\prod_{i=1}^{k} \lambda_i(u) \leq \prod_{i=1}^{k_1} \lambda_i^2(x_1)\lambda_i(s_1) \prod_{i=1}^{k_2} \lambda_i^2(x_2)\lambda_i(s_2) \leq \prod_{i=1}^{k} \lambda_i^2(x)\lambda_i(s).$$

■

We proceed this section giving some more technical lemmas.

Lemma 3.3.4. *Let $x, s \in \mathcal{K}$. Then*

$$\lambda_{\min}(s)\lambda_{\max}(x) \leq \lambda_{\min}(s)\mathrm{tr}\,(x) \leq \mathrm{tr}\left(P(x)^{1/2}s\right) \leq \lambda_{\max}(s)\mathrm{tr}\,(x) \leq r\lambda_{\max}(s)\lambda_{\max}(x).$$

Proof. Since $P(x)$ is self-adjoint and $P(x)e = x^2$, we have

$$\mathrm{tr}\left(P(x)^{1/2}s\right) = \mathrm{tr}\left(P(x)^{1/2}s \circ e\right) = \mathrm{tr}\left(s \circ P(x)^{1/2}e\right) = \mathrm{tr}\,(x \circ s)\,.$$

Let $s = \sum_{i=1}^{r} \lambda_i(s)c_i$ be the spectral decomposition of s. So,

$$\mathrm{tr}\,(x \circ s) = \mathrm{tr}\left(\sum_{i=1}^{r} \lambda_i(s)c_i \circ x\right) = \sum_{i=1}^{r} \lambda_i(s)\mathrm{tr}\,(c_i \circ x)\,.$$

Since $c_i = c_i^2 \in \mathcal{K}$, by Lemma 2.5.7, $P(x)^{1/2}c_i \in \mathcal{K}$. This means that

$$\mathrm{tr}\,(x \circ c_i) = \mathrm{tr}\left(P(x)^{1/2}c_i\right) \geq 0.$$

We can now conclude that

$$\mathrm{tr}\,(x \circ s) \geq \lambda_{\min}(s) \sum_{i=1}^{r} \mathrm{tr}\,(c_i \circ x) = \lambda_{\min}(s)\mathrm{tr}\,(x) \geq \lambda_{\min}(s)\lambda_{\min}(x).$$

The other inequality follows in the same way. ■

The following lemma gives an explicit formula for the minimum and maximum eigenvalues of an element in V. In fact, it is a particular case of the Courant-Fischer's Theorem for Jordan algebras (Theorem 3.4.1 in [4]).

Lemma 3.3.5. *Let $a \in V$, then we obtain the smallest and the largest eigenvalue as*

$$\lambda_{\min}(a) = \min_y \frac{\langle y, a \circ y \rangle}{\langle y, y \rangle}, \qquad \lambda_{\max}(a) = \max_y \frac{\langle y, a \circ y \rangle}{\langle y, y \rangle}.$$

Lemma 3.3.6 (Lemma 14 in [47])**.** *Let $a, b \in V$, then we can bound the eigenvalues of $a + b$ as follows*

$$\lambda_{\min}(a+b) \geq \lambda_{\min}(a) - \|b\| \tag{3.4}$$

$$\lambda_{\max}(a+b) \leq \lambda_{\max}(a) + \|b\| \tag{3.5}$$

Proof. Recall that $\|\cdot\|$ denotes the norm induced by the inner product. Using Lemma 3.3.5 we may write

$$\begin{aligned}
\lambda_{\min}(a+b) &= \min_y \frac{\langle y, (a+b) \circ y \rangle}{\langle y, y \rangle} \\
&= \min_u \frac{\langle y, a \circ y \rangle + \langle y, b \circ y \rangle}{\langle y, y \rangle} \\
&\geq \min_y \frac{\langle y, a \circ y \rangle}{\langle y, y \rangle} + \min_y \frac{\langle y, b \circ y \rangle}{\langle y, y \rangle} \\
&= \lambda_{\min}(a) + \lambda_{\min}(b) \\
&\geq \lambda_{\min}(a) - \|b\|.
\end{aligned}$$

We have that

$$\|b\| = \sqrt{\mathrm{tr}\,(b^2)} = \sqrt{\lambda_1^2(b) + \cdots + \lambda_r^2(b)} \geq |\lambda_{\min}(b)|.$$

From here we obtain that

$$-\|b\| \leq \lambda_{\min}(b) \leq \|b\|.$$

Thus

$$\lambda_{\min}(a+b) \geq \lambda_{\min}(a) + \lambda_{\min}(b) \geq \lambda_{\min}(a) - \|b\|,$$

and (3.4) follows. The proof of (3.5) is similar. ■

Many inequalities of eigenvalues for symmetric matrices can be extended to Euclidean Jordan algebras. We dealt with the ones that are useful in this thesis. For matter of curiosity we finish the section with a generalization of the Hadamard inequality. Let $c_1, \ldots, c_r$ be a Jordan frame and $x = \sum_{i=1}^r x_i c_i + \sum_{i<j} x_{ij}$ be the Peirce decomposition of x with respect to the Jordan frame $c_1, \ldots, c_r$.

Proposition 3.3.7. *For x in $\mathcal{K}$ we have*

$$\det(x) \leq \prod_{i=1}^r \langle x, c_i \rangle,$$

where $c_1, \ldots, c_r$ is a Jordan frame.

Proof. Let $x = \sum_{j=1}^r \lambda_j(x) d_j$ be a spectral decomposition of x. Therefore

$$\langle x, c_i \rangle = \sum_{j=1}^r \lambda_j(x) \langle c_i, d_j \rangle.$$

Let $B = [b_{ij}]$ be a matrix such that $b_{ij} = \langle c_i, d_j \rangle$. Note that $b_{ij} \geq 0$ for all i and j, since c_i, d_j are in $\mathcal{K}$, and the sum of all elements in any column or row of B is 1. Hence

$$\begin{aligned}
\prod_{i=1}^r \langle x, c_i \rangle &= \prod_{i=1}^r \sum_{j=1}^r \lambda_j(x) b_{ij} \\
&\geq \prod_{i=1}^r \prod_{j=1}^r \lambda_j(x)^{b_{ij}} \\
&= \prod_{j=1}^r \lambda_j(x)^{\sum_{i=1}^r b_{ij}} \\
&= \prod_{j=1}^r \lambda_j(x) = \det(x),
\end{aligned}$$

where the inequality follows from the weighted arithmetic-geometric mean inequality (Lemma D.1.2). Note that, in case some $b_{ij} = 0$ we can remove it from the inequality to apply Lemma D.1.2. ■

3.4 Derivatives of eigenvalues

In this section we deduce formulas for some derivatives of eigenvalues. The formulas obtained in this section are, to our knowledge, new. To obtain these formulas we followed [29], where similar formulas for matrices were deduced. A thorough study about the differentiability of eigenvalues can be found in [4].

As before (see definition 2.3.7) we denoted $D_x f(x)$ as the derivative of f at x, where $f : V \mapsto V$ is a function whose domain has a non-empty interior and x is a point of the interior of the domain of f. The second derivative is denoted as $D_x^2 f(x)$. For simplicity of notation and in case $V = \mathbb{R}$ we will sometimes use $f'(t)$ and $f''(t)$ as the first and second derivative of f with respect to t, respectively, for $t \in \mathbb{R}$.

Let $b \in \mathbb{R}^r_\downarrow$. Since the coordinates of b are non-increasing, we can write

$$b_1 = \cdots = b_{k_1} > b_{k_1+1} = \cdots = b_{k_2} > b_{k_2+1} \cdots b_{k_d}, \qquad (k_d := r).$$

Thus d is the number of distinct values in the coordinates of the vector b. We define the corresponding partition $\{I_1, \dots, I_r\}$ of the index set $\{1, \dots, n\}$ such that

$$I_1 := \{1, 2, \dots, k_1\}, I_2 := \{k_1 + 1, \dots, k_2\}, \dots, I_d := \{k_{d-1} + 1, \dots, k_d\}, \qquad (3.6)$$

and we call these sets *blocks* of b. Below we will deal with the blocks of the vector $\lambda(x)$ for $x \in V$.

We say that an eigenvalue of x is *simple* if the correspondent block has size 1.

If for $x \in V$, its spectral decomposition is $x = \sum_{i=1}^r \lambda_i(x) c_i$ then $x \circ c_i = \lambda_i c_i$ for all i. We get this by multiplying, in Jordan product sense, both sides of $x = \sum_{i=1}^r \lambda_i(x) c_i$ by c_i.

Proposition 3.4.1 (Corollary 34 in [5]). *Let $1 \le k \le r$, $x = \sum_{i=1}^r \lambda_i(x) c_i \in V$ and $\lambda(x) \in \mathbb{R}^r_\downarrow$. Denote $S_k(x) = \sum_{i=1}^k \lambda_i(x)$. If $k = k_j$ for some $j \in \{1, \dots, d\}$ then S_k is differentiable with respect to x and $D_x S_k(x) = \sum_{i=1}^k c_i$.*

Proposition 3.4.2. *Under the assumptions of Proposition 3.4.1, we have*

$$D_x \left(\sum_{i \in I_\ell} \lambda_i(x) \right) = \sum_{i \in I_\ell} c_i$$

for $\ell = 1, \dots, d$.

Proof. Just notice that by Proposition 3.4.1 we have

$$D_x \left(\sum_{i \in I_{k_{j+1}}} \lambda_i(x) \right) = D_x S_{k_{j+1}}(x) - D_x S_{k_j}(x) = \sum_{i \in I_{k_{j+1}}} c_i,$$

thus proving the lemma. ∎

In fact, we can say that the sum of equal eigenvalues at x, i.e. $\sum_{i \in I_\ell} \lambda_i(x)$, is differentiable at x. Indeed, as we show below by an example, that if an eigenvalue is not simple then it might not be differentiable.

Example 3.4.3. Let V be the Euclidean Jordan algebra defined in Example 2.4.4. We have already seen that for $x \in V$,

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

Obviously

$$\lambda_1(x) = \lambda_2(x)$$

holds if and only if $\|\bar{x}\| = 0$, i.e, if and only if $\bar{x} = 0$. For that case we compute now the directional derivative of $\lambda_2(x)$ at x in the direction $u = (u_0; \bar{u})$:

$$\begin{aligned} D_x^u \lambda_2(x) &= \lim_{t \to 0} \frac{\lambda_2(x+tu) - \lambda_2(x)}{t} \\ &= \lim_{t \to 0} \frac{x_0 + tu_0 + \|\bar{x} + t\bar{u}\| - x_0}{t} \\ &= \lim_{t \to 0} \frac{tu_0 + \|t\bar{u}\|}{t} = \lim_{t \to 0} u_0 + \frac{|t|\|\bar{u}\|}{t}. \end{aligned}$$

Since we have

$$\lim_{t \to 0^-} \frac{|t|\|\bar{u}\|}{t} = -\|\bar{u}\|$$

and

$$\lim_{t \to 0^+} \frac{|t|\|\bar{u}\|}{t} = \|\bar{u}\|,$$

we conclude that $\lambda_2(x)$ is not differentiable at x if $\bar{u} \neq 0$. □

Corollary 3.4.4. *If $\lambda_i(x)$, with $1 \leq i \leq r$, is a simple eigenvalue then*

$$D_x \lambda_i(x) = c_i, \quad i = 1, \ldots, r.$$

Proof. This follows from Proposition 3.4.2, with $|I_\ell| = 1$. ■

The properties proved in the remaining section are not used in thesis. However, we present them here hoping that they will be useful in the future.

From now on we assume that x depends linearly in parameter t as follows:

$$x(t) = x_0 + tu, \quad x_0, u \in V, \ t \in \mathbb{R},$$

but we sometimes write x instead of $x(t)$ for simplicity of notation. Let

$$x(t) = \sum_{i=1}^{r} \lambda_i(x(t)) c_i$$

be the spectral decomposition of $x(t)$ and

$$u = \sum_{i=1}^{r} u_i c_i + \sum_{i<j} u_{ij}$$

the Peirce decomposition of u with respect to the Jordan frame $c_1, \ldots, c_r$ (see (2.30)).

Proposition 3.4.5. *Let $\lambda_i(x(t))$ be a simple eigenvalue of $x(t)$. Then*

$$D_t \lambda_i(x(t)) = u_i.$$

Proof. Applying the chain rule we have $D_t \lambda_i(x(t)) = \langle D_x \lambda_i(x(t)), x'(t) \rangle$. Therefore by Corollary 3.4.4,

$$D_t \lambda_i(x(t)) = \langle c_i, u \rangle = u_i.$$

The proposition is proved. ∎

Proposition 3.4.6. *Under the assumptions of Proposition 3.4.1, we have*

$$D_t \left(\sum_{i \in I_\ell} \lambda_i(x(t)) \right) = \sum_{i \in I_\ell} u_i,$$

for $\ell = 1, \ldots, d$.

Proof. This follows from Proposition 3.4.2. ∎

Lemma 3.4.7. *Let $c(t) \in V$ be an idempotent dependent on a real parameter t. If $c(t)$ is differentiable with respect to t then*

$$c(t) \circ c'(t) = \tfrac{1}{2} c'(t).$$

Proof. Since $c(t)^2 = c(t)$, we have $2c(t) \circ c'(t) = c'(t)$. The result follows. ∎

Lemma 3.4.8. *Let $c_i(t)$ and $c_j(t)$ be two orthogonal idempotents. If $c_i(t)$ and $c_j(t)$ are differentiable with respect to t then*

$$c_i'(t) \circ c_j(t) + c_i(t) \circ c_j'(t) = 0.$$

Proof. The proof follows from $c_i(t) \circ c_j(t) = 0$. ∎

Let $x = \sum_{i=1}^r \lambda_i(x) c_i$ be a spectral decomposition of x. If all the eigenvalues are simple, by Theorem 2.4.1 the Jordan frame $c_1, \ldots, c_r$ is unique with respect to x and every c_i is a polynomial in x. Hence, every c_i is differentiable with respect to x. The following theorem deduces a formula for $D_t c_i(x(t))$.

Theorem 3.4.9. *Let $x(t) := x_0 + tu$ and $x = \sum_{i=1}^r \lambda_i(x(t)) c_i$ be the spectral decomposition of x. Suppose that the eigenvalues of x are simple. Then*

$$D_t c_i(x(t)) = 4 \sum_{j \neq i} \frac{(c_i \circ u) \circ c_j}{\lambda_i - \lambda_j} = \sum_{j \neq i} \frac{u_{ij}}{\lambda_i - \lambda_j}.$$

Proof. During the proof we use c_j or $c_j(t)$ instead of $c_j(x(t))$. Clearly $x \circ c_i = \lambda_i c_i$. Differentiating this expression to t we get

$$x'(t) \circ c_i + x \circ c_i'(t) = \lambda_i'(t) c_i + \lambda_i c_i'(t),$$

which is equivalent to

$$\lambda_i'(t) c_i = u \circ c_i + (x - \lambda_i e) \circ c_i'(t).$$

Pre-multiplying by c_j we obtain

$$0 = c_j \circ (u \circ c_i) + c_j \circ ((x - \lambda_i e) \circ c_i'(t)) \ \text{ for } j \neq i. \tag{3.7}$$

From Lemma 3.4.8 we get that

$$(c_i'(t) \circ c_j(t)) \circ x + (c_i(t) \circ c_j'(t)) \circ x = 0,$$

for $i \neq j$, which, commutating x with c_i (because $c_i \in \mathbb{R}[x]$), is equivalent to

$$(c_i'(t) \circ x) \circ c_j + (x \circ c_j'(t)) \circ c_i = 0, \quad i \neq j. \tag{3.8}$$

On the other hand, we have $(x \circ c_j) \circ c_i = (x \circ c_i) \circ c_j$. Differentiating this expression with respect to t we get

$$(u \circ c_j) \circ c_i + (x \circ c_j') \circ c_i + (x \circ c_j) \circ c_i' = (u \circ c_i) \circ c_j + (x \circ c_i') \circ c_j + (x \circ c_i) \circ c_j',$$

where we used c_i' instead of $c_i'(t)$. Rearranging the terms in the last expression we get

$$(x \circ c_j'(t)) \circ c_i + (x \circ c_j) \circ c_i'(t) = (x \circ c_i'(t)) \circ c_j + (x \circ c_i) \circ c_j'(t). \tag{3.9}$$

From expressions (3.8) and (3.9) we obtain

$$-(x \circ c_i'(t)) \circ c_j + (x \circ c_j) \circ c_i'(t) = (x \circ c_i'(t)) \circ c_j + (x \circ c_i) \circ c_j'(t).$$

From here it follows

$$\begin{aligned}
2(x \circ c_i'(t)) \circ c_j &= (x \circ c_j) \circ c_i'(t) - (x \circ c_i) \circ c_j'(t) \\
&= \lambda_j c_j \circ c_i'(t) - \lambda_i c_i \circ c_j'(t) \\
&= \lambda_j c_j \circ c_i'(t) + \lambda_i c_i'(t) \circ c_j \quad \text{(Lemma 3.4.8)} \\
&= (\lambda_j + \lambda_i) c_i'(t) \circ c_j
\end{aligned}$$

Using the last expression in equation (3.7), we get

$$0 = c_j \circ (u \circ c_i) + \tfrac{1}{2}(\lambda_j + \lambda_i) c_i'(t) \circ c_j - (c_j \circ \lambda_i e) \circ c_i'(t) \ \text{ for } j \neq i,$$

hence

$$(\lambda_i - \lambda_j) c_i'(t) \circ c_j = 2 c_j \circ (u \circ c_i) \ \text{ for } j \neq i$$

which is equivalent to

$$c_i'(t) \circ c_j = 2\frac{c_j \circ (u \circ c_i)}{\lambda_i - \lambda_j} \text{ for } j \neq i.$$

Taking the sum over $j \neq i$, it follows that

$$\sum_{j\neq i} c_j \circ c_i'(t) = 2\sum_{j\neq i} \frac{c_j \circ (u \circ c_i)}{\lambda_i - \lambda_j}.$$

Note that $\sum_{j\neq i} c_j = e - c_i$. Therefore, by Lemma 3.4.7,

$$\sum_{j\neq i} c_j \circ c_i'(t) = (e - c_i) \circ c_i'(t) = c_i'(t) - c_i \circ c_i'(t) = \frac{1}{2}c_i'(t).$$

We can now conclude that

$$c_i'(t) = 4\sum_{j\neq i} \frac{c_j \circ (u \circ c_i)}{\lambda_i - \lambda_j}.$$

Since $P_{ij} = 4L(c_i)L(c_j)$ (see Proposition 2.9.7) and $P_{ij}u = u_{ij}$ the last equality of the theorem follows. ∎

Theorem 3.4.9 provides an easy way to compute the derivative of $c_i(x(t))$ with respect to t. See the example below.

Example 3.4.10. Let $(\mathbb{R}^{n+1}, \circ)$ be the Euclidean Jordan algebra with the Jordan product defined by

$$x \circ y = (x^T y; x_0\bar{y} + y_0\bar{x}),$$

and denoting $(x_0; x_1; \ldots; x_n) \in \mathbb{R}^{n+1}$ as $(x_0; \bar{x})$ with $\bar{x} = (x_1; \ldots; x_n)$. As we know from Example 2.6.3, the symmetric cone of this Jordan algebra is the second-order cone. Recall from Example 2.4.4 that, the spectral decomposition of $x \in \mathbb{R}^{n+1}$ is

$$x = \lambda_1(x)c_1(x) + \lambda_2(x)c_2(x),$$

where the eigenvalues are

$$\lambda(x) = \begin{bmatrix} \lambda_1(x) \\ \lambda_2(x) \end{bmatrix} = \begin{bmatrix} x_0 - \|\bar{x}\| \\ x_0 + \|\bar{x}\| \end{bmatrix}$$

and the Jordan frame is

$$c_1(x) = \frac{1}{2}\begin{bmatrix} 1 \\ -\frac{\bar{x}}{\|\bar{x}\|} \end{bmatrix}, \qquad c_2(x) = \frac{1}{2}\begin{bmatrix} 1 \\ \frac{\bar{x}}{\|\bar{x}\|} \end{bmatrix}.$$

Setting $x := x_0 + tu$, we differentiate $c_1(x(t))$ with respect to t. This gives

$$c_1'(t) = \frac{1}{2}\left(0; D_t\left(-\frac{\bar{x}}{\|\bar{x}\|}\right)\right) = \frac{1}{2}\left(0; \frac{-\bar{u}\|\bar{x}\| + \bar{x}\frac{\bar{x}^T\bar{u}}{\|\bar{x}\|}}{\|\bar{x}\|^2}\right). \tag{3.10}$$

On the other hand, if we apply directly Theorem 3.4.9, we get

$$\begin{aligned} c_1'(t) &= 4\frac{c_2 \circ (u \circ c_1)}{\lambda_1 - \lambda_2} \\ &= 4\frac{c_2 \circ \frac{1}{2}(u_0 - \frac{\bar{u}^T\bar{x}}{\|\bar{x}\|}; \bar{u} - u_0\frac{\bar{x}}{\|\bar{x}\|})}{-2\|\bar{x}\|} \\ &= 4\frac{\frac{1}{4}(u_0 - \frac{\bar{u}^T\bar{x}}{\|\bar{x}\|} + \frac{\bar{u}^T\bar{x}}{\|\bar{x}\|} - u_0\frac{\bar{x}^t\bar{x}}{\|\bar{x}\|}; \bar{u} - u_0\frac{\bar{x}}{\|\bar{x}\|} + (u_0 - \frac{\bar{u}^T\bar{x}}{\|\bar{x}\|})\frac{\bar{x}}{\|\bar{x}\|})}{-2\|\bar{x}\|} \\ &= \frac{(0; \bar{u} - \frac{\bar{u}^T\bar{x}}{\|\bar{x}\|}\frac{\bar{x}}{\|\bar{x}\|})}{-2\|\bar{x}\|} \\ &= \frac{1}{2}(0; -\frac{\bar{u}}{\|\bar{x}\|} + \frac{\bar{u}^T\bar{x}}{\|\bar{x}\|^3}\bar{x}), \end{aligned}$$

in agreement with (3.10). □

In the following property we use the twice differentiability of the eigenvalues. In fact, if the eigenvalues of $x \in V$ are simple then they are twice differentiable at x. This follows from Corollary 3.4.4 and because in this case the Jordan frame is differentiable at x.

Corollary 3.4.11. *Let $x(t) = \sum_{i=1}^r \lambda_i(x(t))c_i(x(t))$. If the eigenvalues are simple then*

$$D_t^2\lambda_i(x(t)) = 4\sum_{j\neq i}\frac{\operatorname{tr}((u \circ (c_j \circ u)) \circ c_i)}{\lambda_i - \lambda_j} = \sum_{j\neq i}\frac{\operatorname{tr}(u_{ij}^2)}{\lambda_i - \lambda_j}.$$

Proof. From Proposition 3.4.5 we get

$$D_t^2\lambda_i(x(t)) = D_t(u_i) = \langle c_i'(t), u\rangle. \tag{3.11}$$

Now, the result follows from Theorem 3.4.9. ■

We prove similar properties for eigenvalues which are not simple. However, its proof requires special attention because, we cannot guarantee differentiability for the eigenvalues at x. Aggregating all the idempotents in the same block, for $\lambda(x) \in \mathbb{R}^{\downarrow}$, we have, by Theorem 2.4.1, that $e_\ell := \sum_{i\in I_\ell} c_i$ with $\ell = 1\ldots,d$ is a unique complete system of idempotents and $e_\ell \in \mathbb{R}[x]$. Consequently e_ℓ, with $\ell = 1,\ldots,d$ are differentiable with respect to x. The Propositions 3.4.1 and 3.4.2 provide us the tools for the proof of the following result.

Theorem 3.4.12. *Let $x(t) = \sum_{i=1}^r \lambda_i(x(t))c_i(x(t))$. If the eigenvalues are not simple and $\lambda(x) \in \mathbb{R}^{\downarrow}$, then*

$$D_t\left(\sum_{i\in I_\ell} c_i(x(t))\right) = 4\sum_{i\in I_\ell}\sum_{j\notin I_\ell}\frac{(c_j \circ u) \circ c_i}{\lambda_i - \lambda_j} = \sum_{i\in I_\ell}\sum_{j\notin I_\ell}\frac{u_{ij}}{\lambda_i - \lambda_j},$$

for $\ell = 1,\ldots,d$.

Proof. Let $e_\ell := \sum_{i\in I_\ell} c_i$ and $e_m := \sum_{i\in I_m} c_i$, with $m \neq \ell$. Clearly

$$(x \circ e_\ell) \circ e_m = 0. \tag{3.12}$$

Differentiating both sides of (3.12) to t, we get

$$(x'(t) \circ e_\ell) \circ e_m + (x \circ e'_\ell(t)) \circ e_m + (x \circ e_\ell) \circ e'_m(t) = 0,$$

which is equivalent to

$$(u \circ e_\ell) \circ e_m + (x \circ e'_\ell(t)) \circ e_m + \sum_{i\in I_\ell} \lambda_i c_i \circ e'_m(t) = 0. \tag{3.13}$$

From Lemma 3.4.8 we get that

$$(e'_\ell(t) \circ e_m(t)) \circ x + (e_\ell(t) \circ e'_m(t)) \circ x = 0,$$

for $m \neq \ell$, which commutating x with e_ℓ and e_m, is equivalent to

$$(e'_\ell(t) \circ x) \circ e_m + (x \circ e'_m(t)) \circ e_\ell = 0, \quad m \neq \ell. \tag{3.14}$$

On the other hand, we have $(x \circ e_m) \circ e_\ell = (x \circ e_\ell) \circ e_m$. Differentiating this expression to t we get

$$(u \circ e_m) \circ e_\ell + (x \circ e'_m) \circ e_\ell + (x \circ e_m) \circ e'_\ell = (u \circ e_\ell) \circ e_m + (x \circ e'_\ell) \circ e_m + (x \circ e_\ell) \circ e'_m,$$

where we used e'_ℓ instead of $e'_\ell(t)$. Rearrange the last expression we get

$$(x \circ e'_m(t)) \circ e_\ell + (x \circ e_m) \circ e'_\ell(t) = (x \circ e'_\ell(t)) \circ e_m + (x \circ e_\ell) \circ e'_m(t). \tag{3.15}$$

Using the identity (3.14) in the identity (3.15) we obtain

$$-(x \circ e'_\ell(t)) \circ e_m + (x \circ e_m) \circ e'_\ell(t) = (x \circ e'_\ell(t)) \circ e_m + (x \circ e_\ell) \circ e'_m(t).$$

From here, it follows that

$$\begin{aligned}
2(x \circ e'_\ell(t)) \circ e_m &= (x \circ e_m) \circ e'_\ell(t) - (x \circ e_\ell) \circ e'_m(t) \\
&= \alpha_m e_m \circ e'_\ell(t) - \alpha_\ell e_\ell \circ e'_m(t) \\
&= \alpha_m e_m \circ e'_\ell(t) + \alpha_\ell e_m \circ e'_\ell(t) \quad \text{(Lemma 3.4.8)} \\
&= (\alpha_m + \alpha_\ell) e'_\ell(t) \circ e_m,
\end{aligned}$$

where we denoted $\alpha_m := \lambda_j$ for $j \in I_m$, because $\lambda_i = \lambda_j$ for all $i, j \in I_m$. Using the last expression in equation (3.13), we get

$$0 = (u \circ e_\ell) \circ e_m + \frac{1}{2}(\alpha_m + \alpha_\ell) e'_\ell(t) \circ e_m - \alpha_\ell e_m \circ e'_\ell(t) \quad \text{for } \ell \neq m.$$

Hence

$$(\alpha_\ell - \alpha_m) e'_\ell(t) \circ e_m = 2 e_m \circ (u \circ e_\ell) \quad \text{for } m \neq \ell$$

which is equivalent to

$$e_\ell'(t) \circ e_m = 2\frac{e_m \circ (u \circ e_\ell)}{\alpha_\ell - \alpha_m} \text{ for } m \neq \ell.$$

Taking the sum over $m \neq \ell$, it follows that

$$\sum_{m \neq \ell} e_m \circ e_\ell'(t) = 2 \sum_{m \neq \ell} \frac{e_m \circ (u \circ e_\ell)}{\alpha_m - \alpha_\ell}.$$

Regard that $\sum_{m \neq \ell} e_m = e - e_\ell$. Therefore, by Lemma 3.4.7,

$$\sum_{m \neq \ell} e_m \circ e_\ell'(t) = (e - e_\ell) \circ e_\ell'(t) = e_\ell'(t) - e_\ell \circ e_\ell'(t) = \frac{1}{2} e_\ell'(t).$$

We can now conclude that

$$e_\ell'(t) = 4 \sum_{m \neq \ell} \frac{e_m \circ (u \circ e_\ell)}{\alpha_\ell - \alpha_m}.$$

Replacing e_m by $\sum_{j \in I_m} c_j$ and e_ℓ by $\sum_{i \in I_\ell} c_i$ we obtain

$$\begin{aligned} D_t \sum_{i \in I_\ell} c_i(t) &= 4 \sum_{m \neq \ell} \sum_{i \in I_\ell} \sum_{j \in I_m} \frac{c_j \circ (u \circ c_i)}{\alpha_\ell - \alpha_m} \\ &= 4 \sum_{m \neq \ell} \sum_{i \in I_\ell} \sum_{j \in I_m} \frac{c_j \circ (u \circ c_i)}{\lambda_i - \lambda_j} \\ &= 4 \sum_{i \in I_\ell} \sum_{j \notin I_\ell} \frac{c_j \circ (u \circ c_i)}{\lambda_i - \lambda_j}. \end{aligned}$$

Since $P_{ij} = 4L(c_i)L(c_j)$ (see Proposition 2.9.7) and $P_{ij}u = u_{ij}$ the last equality of the theorem follows. ■

Corollary 3.4.13. *Under the assumptions of the Theorem 3.4.12 we have*

$$D_t^2 \left(\sum_{i \in I_\ell} \lambda_i(x(t)) \right) = 4 \sum_{i \in I_\ell} \sum_{j \notin I_\ell} \frac{\operatorname{tr}((c_j \circ (u \circ c_i)) \circ u)}{\lambda_i - \lambda_j} = \sum_{i \in I_\ell} \sum_{j \notin I_\ell} \frac{\operatorname{tr}\left(u_{ij}^2\right)}{\lambda_i - \lambda_j},$$

for $\ell = 1, \ldots, d$.

Proof. From Theorem 3.4.6 we get

$$D_t^2 \sum_{i \in I_\ell} \lambda_i(x(t)) = \langle D_t \left(\sum_{i \in I_\ell} c_i(t) \right), u \rangle, \tag{3.16}$$

for $l = 1, \ldots, d$. Since

$$\operatorname{tr}((c_j \circ (u \circ c_i)) \circ u) = \operatorname{tr}(P_{ij} u \circ u) = \operatorname{tr}\left(u_{ij}^2\right),$$

the result follows by Theorem 3.4.12. ■

In the properties presented in this section, i.e., for the derivatives of eigenvalues and Jordan frames, we had two versions, one supposing that the eigenvalues are simple and another supposing that they are not. These both versions appear naturally when we are deducing the formulas, which is coherent with the differentiability of the Jordan frame.

3.5 Spectral functions

We define spectral functions using the notion of similarity, as defined in the beginning of this chapter. Let F be a real-valued function on V such that

$$F(gx) = F(x),$$

for every x in its domain and every $g \in \mathrm{OAut}(\mathcal{K})$.

Given a permutation π of n elements, $\pi : \{1, \ldots, r\} \mapsto \{\pi(1), \ldots, \pi(r)\}$. Its permutation matrix P is the $r \times r$ matrix whose entries are 0 except the entries $(i, \pi(i))$ which are equal to 1.

Let $f : \mathbb{R}^r \mapsto \mathbb{R}$ and assume that for $x \in \mathrm{dom} f$ we have $Px \in \mathrm{dom} f$ for any permutation matrix P. If f has the property

$$f(x) = f(Px)$$

for any permutation matrix P and any $x \in \mathrm{dom} f$ we call the function f *symmetric*.

One may consider the eigenvalues of x in V as functions of x and define the eigenvalue map,

$$\lambda : V \mapsto \mathbb{R}^r_{\downarrow}.$$

It satisfies $\lambda(gx) = \lambda(x)$ since x and gx are similar, with $g \in \mathrm{OAut}(\mathcal{K})$.

Now, we have the following proposition.

Proposition 3.5.1. *The following two properties are equivalent:*

(i) For any element $x \in V$ and $g \in \mathrm{OAut}(\mathcal{K})$ we have $F(gx) = F(x)$.

(ii) $F = f\lambda$ for some symmetric function $f : \mathbb{R}^r \mapsto \mathbb{R}$. Here $f\lambda$ denotes f composed by λ.

Proof. $(ii) \implies (i)$: by statement (ii) we have $F(gx) = f(\lambda(gx))$. Since $gx \sim x$ we may write

$$F(gx) = f\lambda(gx) = f(\lambda(gx)) = f(\lambda(x)) = F(x).$$

Therefore, statement (i) follows.

$(i) \implies (ii)$: for $a \in \mathbb{R}^r$ define

$$f(a) := F(a_1 c_1 + \cdots + a_r c_r)$$

for some fixed Jordan frame $c_1, \dots, c_r$. We have that $c_{\pi(1)}, \dots, c_{\pi(r)}$ is a Jordan frame. By Theorem 2.8.5 there exists g such that $gc_i = c_{\pi(i)}$ and by (i)

$$F(g(a_1c_1 + \cdots + a_rc_r)) = F(a_1c_1 + \cdots + a_rc_r).$$

Therefore f is symmetric. Hence

$$f(\lambda(x)) = F(\lambda_1(x)c_1 + \cdots + \lambda_r(x)c_r) = F(x).$$

This completes the proof. ■

So every function that is invariant under $\mathrm{OAut}(K)$, can be decomposed as $F(x) = f(\lambda(x))$. We say that F is generated by f and we call F a *spectral function*. Hence, it is quite clear that F depends only on the eigenvalues of its argument. An important subclass of spectral functions is obtained when $f(a) = g(a_1) + \cdots + g(a_n)$, with $a \in \mathbb{R}^r$ for some univariate real function g. We call such symmetric functions *separable* and the respective spectral functions are called *separable spectral* functions.

3.6 Derivatives of spectral functions

In this section we present derivatives of spectral functions. The following result is fundamental to prove the remaining results of this section.

Proposition 3.6.1 (Corollary 24 in [5]). *For every $x, y \in V$, we have*

$$\|\lambda(x) - \lambda(y)\|_{\mathbb{R}^r} \leq \|x - y\|,$$

where $\lambda(x) \in \mathbb{R}^r$ and $\|\cdot\|_{\mathbb{R}^r}$ denotes the Euclidean norm in $\mathbb{R}^r$.

Let $x = \sum_{i=1}^r \lambda_i(x)c_i$ be a spectral decomposition of x. We may write $x = \sum_{\ell=1}^d \alpha_\ell(x)e_\ell$ where $e_\ell = \sum_{i \in I_\ell} c_i$ and $\alpha_\ell = \lambda_i$ for $i \in I_\ell$ and I_ℓ, $\ell = 1, \dots, d$ are the blocks as in (3.6) defined for the vector $\lambda(x) \in \mathbb{R}^r_\downarrow$. It follows that the complete system of orthogonal idempotents $e_1 \dots, e_\ell$ is unique with respect to x (see Section 2.4).

The following theorem gives explicitly the derivative of a separable spectral function. This result is adapted from a more general case (for spectral functions) obtained in [5].

Theorem 3.6.2. *Let $D \subset \mathbb{R}$ be an open set and $f : D \mapsto \mathbb{R}$. Let $U = \{x \in V : \lambda(x) \in D^r\}$ and $F : U \mapsto \mathbb{R}$ defined as $F(x) = \sum_{i=1}^r f(\lambda_i(x))$, with $x = \sum_{i=1}^r \lambda_i(x)c_i \in V$. If f is differentiable in D then F is differentiable in U and*

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

Proof. We want prove that

$$F(x+u) - F(x) - \langle \sum_{i=1}^{r} f'(\lambda_i(x))c_i, u\rangle = o(u).$$

We have

$$|F(x+u) - F(x) - \langle \sum_{i=1}^{r} f'(\lambda_i(x))c_i, u\rangle| = \left| \sum_{\ell=1}^{d} \sum_{i \in I_\ell} (f(\lambda_i(x+u)) - f(\lambda_i(x)) - f'(\lambda_i(x))\langle c_i, u\rangle) \right|, \quad (3.17)$$

where I_ℓ are the blocks as defined in (3.6). The right-hand side of (3.17) can be written as

$$\left| \sum_{\ell=1}^{d} \sum_{i \in I_\ell} \big(f(\lambda_i(x+u)) - f(\lambda_i(x)) - f'(\lambda_i(x))(\lambda_i(x+u) - \lambda_i(x)) + f'(\lambda_i(x))(\lambda_i(x+u) - \lambda_i(x) - \langle c_i, u\rangle)\big) \right|,$$

which is less than or equal to

$$\sum_{\ell=1}^{d} \left(\sum_{i \in I_\ell} |f(\lambda_i(x+u)) - f(\lambda_i(x)) - f'(\lambda_i(x))(\lambda_i(x+u) - \lambda_i(x))| + \left| \sum_{i \in I_\ell} f'(\lambda_i(x))(\lambda_i(x+u) - \lambda_i(x) - \langle c_i, u\rangle) \right| \right). \quad (3.18)$$

Since f is differentiable,

$$\sum_{\ell=1}^{d} \sum_{i \in I_\ell} |f(\lambda_i(x+u)) - f(\lambda_i(x)) - f'(\lambda_i(x))(\lambda_i(x+u) - \lambda_i(x))| = \sum_{i=1}^{r} o(\lambda_i(x+u) - \lambda_i(x)).$$

For simplicity of notation, let $\varepsilon := \lambda(x+u) - \lambda(x)$ and

$$h(\varepsilon_i) := f(\lambda_i(x) + \varepsilon_i) - f(\lambda_i(x)) - f'(\lambda_i(x))\varepsilon_i.$$

Since

$$\lim_{\|\varepsilon\| \to 0} \frac{|h(\varepsilon_i)|}{\|\varepsilon\|} \le \lim_{\|\varepsilon\| \to 0} \frac{|h(\varepsilon_i)|}{|\varepsilon_i|} = 0,$$

we have

$$\lim_{\|\varepsilon\|\to 0} \frac{\sum_{i=1}^r |h(\varepsilon_i)|}{\|\varepsilon\|} = 0.$$

Thus,

$$\sum_{i=1}^r |h(\varepsilon_i)| = o(\varepsilon).$$

The left-hand side of the last expression is also $o(u)$ because, by Proposition 3.6.1, $\|\varepsilon\| \le \|u\|$. The second summand of (3.18) can be rewritten:

$$\left| \sum_{i\in I_\ell} f'(\lambda_i(x))(\lambda_i(x+u) - \lambda_i(x) - \langle c_i, u\rangle) \right| =$$

$$|f'(\alpha_\ell(x))| \left| \sum_{i\in I_\ell} (\lambda_i(x+u) - \lambda_i(x) - \langle c_i, u\rangle) \right|,$$

with $\alpha_\ell(x) = \lambda_i(x)$, for $i \in I_\ell$. Therefore, by Proposition 3.4.2,

$$\left| \sum_{i\in I_\ell} f'(\lambda_i(x))(\lambda_i(x+u) - \lambda_i(x) - \langle c_i, u\rangle) \right| = o(u).$$

Thus, the theorem is proved. ■

Let $x = \sum_{i=1}^r \lambda_i(x) c_i$ be a spectral decomposition of x. Let f be a real valued function on an open subset D of $\mathbb{R}$ and assume that the eigenvalues of x are in D. We denote

$$[\lambda_j, \lambda_k]_f := \frac{f(\lambda_j) - f(\lambda_k)}{\lambda_j - \lambda_k}.$$

When $\lambda_j = \lambda_k$ the quotient is understood as a derivative: $f'(\lambda_j)$.

The following lemma is also a direct consequence of Theorem 5.6.1 in [4].

Lemma 3.6.3 (Lemma 1 in [28]). *If f is a continuous differentiable function in a suitable domain that contains all the eigenvalues of $x \in V$, then $G(x) = \sum_{i=1}^r f(\lambda_i(x))c_i$ is continuous differentiable at x and*

$$D_x G(x) = \sum_{i=1}^r [\lambda_i, \lambda_i]_f P_{ii} + \sum_{j<k} [\lambda_j, \lambda_k]_f P_{jk}, \tag{3.19}$$

with P_{jk}, $1 \le j \le k \le r$ as defined in (2.24) *and* (2.25).

Sketch of the proof. First we consider the special case where $f(\lambda_i) = f'(\lambda_i) = 0$ for each i. The right hand side of (3.19) as well as $G(x)$ is then zero. So, for proving that

$$\|G(x+u) - G(x) - D_x G(x) u\| = o(\|u\|)$$

it suffices to show that

$$\|G(x+u)\| = o(\|u\|). \tag{3.20}$$

We have that $x+u = \sum_{i=1}^{r}(\lambda_i + \varepsilon_i)c_i'$ for some Jordan frame $c_1', \ldots, c_r'$. Applying Proposition 3.6.1 we get

$$\sum_{i=1}^{r} \varepsilon_i^2 = \sum_{i=1}^{r}(\lambda_i + \varepsilon_i - \lambda_i)^2 \le \|(x+u) - x\|^2 = \|u\|^2.$$

The mean value theorem now gives, with some ε_i' between 0 and ε_i, for each i,

$$\begin{aligned} \|G(x+u)\|^2 &= \Big\| \sum_i f(\lambda_i + \varepsilon_i)c_i' - \sum_i f(\lambda_i)c_i' \Big\|^2 \\ &= \sum_i \varepsilon_i^2 f'(\lambda_i + \varepsilon_i')^2 \le \|u\|^2 \sum_i f'(\lambda_i + \varepsilon_i')^2 \end{aligned}$$

and (3.20) follows. In the case of a general continuous differentiable function f, we can find a polynomial p such that p and p' coincide with f and f' at λ_i. Then, by what we just proved, the lemma holds for $f-p$; it only remains to prove that it also holds for p. For this, it will clearly suffice to prove the lemma for the functions $x^m (m \in \mathbb{N})$. Explicitly, what remains to prove is that

$$(x+u)^m - x^m = \sum_i m\lambda_i^{m-1} u_i c_i + \sum_{j<k} \frac{\lambda_j^m - \lambda_k^m}{\lambda_j - \lambda_k} u_{jk} + o(\|u\|) \tag{3.21}$$

for all $m \in \mathbb{N}$. We prove this by induction on m. For $m = 1$ we have

$$u = \sum_i u_i c_i + \sum_{j<k} u_{jk} + o(\|u\|) = u + o(\|u\|).$$

By power-associativity we have

$$(x+u)^{m+1} - x^{m+1} = (x+u)\left((x+u)^m - x^m\right) + hx^m.$$

Using the induction hypothesis (3.21) this equals

$$(x+u)\left(\sum_i m\lambda_i^{m-1} u_i c_i + \sum_{j<k} \frac{\lambda_j^m - \lambda_k^m}{\lambda_j - \lambda_k} u_{jk}\right) + hx^m + o(\|u\|).$$

In this expression we write x and u in terms of the Peirce decomposition and use the rules $c_i c_j = \delta_{ij} c_i$, $c_i h_{jk} = \frac{1}{2}(\delta_{ij} + \delta_{ik})u_{jk}$. After simple computations it turns out to be equal to

$$\sum_i (m+1)\lambda_i h_i c_i + \sum_{j<k} \frac{\lambda_j^{m+1} - \lambda_k^{m+1}}{\lambda_j - \lambda_k} u_{jk} + o(\|u\|),$$

finishing the proof. ∎

3.7 Conclusion

In this chapter we proved an inequality concerning the product of eigenvalues (Theorem 3.3.2), which will be crucial for the proof of an inequality for barrier functions in Chapter 4. The similarity property given by Proposition 3.2.4 enters to our consideration during the analysis of the algorithm in Chapter 5.

We deduced derivative formulas for eigenvalues presented in Section 3.4.

Finally, we presented spectral functions and proved (see Proposition 3.5.1) that they only depend on the eigenvalues of its argument.

Chapter 4

Barrier functions

4.1 Introduction

A new class of barrier functions was introduced in [6–9]. Each barrier function in this class is generated by a univariate real function, and provides an interior-point method, whose iteration complexity for so-called large-update methods is better than for the primal-dual logarithmic barrier function. In this chapter we introduce this class of barrier functions in the framework of Euclidean Jordan algebras.

We derive the first and second derivatives of the aforementioned barrier functions when their argument depends linearly on a real parameter. We use the results on derivatives presented in the previous chapter.

4.2 Kernel functions

In this section we give the first steps in order to define the barrier functions, generated by kernel functions, for symmetric cones. Following [9], we call

$$\psi(t) : (0, +\infty) \mapsto [0, +\infty)$$

a *kernel function* if ψ is twice differentiable and the following conditions are satisfied.

(*i*) $\psi'(1) = \psi(1) = 0$;

(*ii*) $\psi''(t) > 0$, for all $t > 0$;

(*iii*) $\psi(t)$ is e-convex, (i.e. $\psi(e^t)$ is convex).

We say that ψ is *coercive* if

$$\lim_{t \to 0} \psi(t) = \lim_{t \to +\infty} \psi(t) = +\infty. \tag{4.1}$$

Clearly, this definition implies that $\psi(t)$ is nonnegative and strictly convex whereas $\psi(1) = 0$. As can be easily verified, any kernel function ψ is completely determined by its second derivative:

$$\psi(t) = \int_1^t \int_1^\xi \psi''(\zeta) d\zeta d\xi.$$

For the purpose of this work, and as in [9], we consider more conditions on the kernel functions. We require that ψ is three times continuous differentiable and

$$\begin{aligned} \psi'''(t) &< 0, && (4.2)\\ 2\psi''(t)^2 - \psi'(t)\psi'''(t) &> 0, \quad t < 1, && (4.3)\\ \psi''(t)\psi'(\beta t) - \beta\psi'(t)\psi''(\beta t) &> 0, \quad t > 1,\ \beta > 1. && (4.4) \end{aligned}$$

The e-convexity property of the kernel function admits different interpretations, which are described below. As observed by Glineur ([21]), condition 4.3 is also consequence of the self-concordant barrier conditions with complexity parameter 1.

Lemma 4.2.1 (Lemma 2.1.2 in [43]). *The following three properties are equivalent:*

(i) $\psi(t)$ *is e-convex;*

(ii) $\psi(\sqrt{t_1 t_2}) \leq \frac{1}{2}(\psi(t_1) + \psi(t_2))$ *for* $t_1, t_2 > 0$*;*

(iii) $\psi'(t) + t\psi''(t) \geq 0, \quad t > 0.$

Following [9], we say that $\psi(t)$ *exponential convex*, or shortly *e-convex* if and only if $\psi(e^t)$ is convex. This property has been proven to be very useful in the analysis of primal-dual algorithms based on kernel functions (see for example [9, 43]). Of course, Lemma 4.2.1 gives different interpretations of e-convexity.

In what follows we present some technical properties of the kernel functions.

Definition 4.2.2. Any kernel function that is coercive and satisfies the conditions (4.2)-(4.4) is called an *eligible kernel function.* □

In the sequel we assume that ψ is an eligible kernel function and present some technical lemmas for ψ.

We frequently use that if $K > 0$ then there are precisely two values of t for which $\psi(t) = K$. This follows since $\psi(t)$ is strictly convex and minimal at $t = 1$, with $\psi(1) = 0$. If these values are t_1 and t_2, and $t_1 \leq t_2$, then $t_1 < 1 < t_2$.

Lemma 4.2.3 (Lemma 3.1 in [9]). *Suppose that* $\psi(t_1) = \psi(t_2)$*, with* $t_1 \leq 1 \leq t_2$ *and* $\beta \geq 1$*. Then*

$$\psi(\beta t_1) \leq \psi(\beta t_2).$$

Equality holds if and only if $\beta = 1$ *or* $t_1 = t_2 = 1$*.*

Lemma 4.2.4 (Lemma 4.8 in [9]). *Suppose that* $\psi(t_1) = \psi(t_2)$, *with* $t_1 \leq 1 \leq t_2$. *Then* $\psi'(t_1) \leq 0$ *and* $\psi'(t_2) \geq 0$, *whereas*

$$-\psi'(t_1) \geq \psi'(t_2).$$

Lemma 4.2.5 (Lemma 2.6 in [9]). *We have*

$$\frac{1}{2}\psi''(t)(t-1)^2 < \psi(t) < \frac{1}{2}\psi''(1)(t-1)^2, \quad t > 1,$$

$$\frac{1}{2}\psi''(1)(t-1)^2 < \psi(t) < \frac{1}{2}\psi''(t)(t-1)^2, \quad t < 1.$$

4.3 Barrier functions based on kernel functions

Let the triple $(V, \circ, \langle , \rangle)$ be an n-dimensional Euclidean Jordan algebra with rank r, where "$\circ$" stands for the Jordan product and $\langle , \rangle$ the inner product given by $\langle x, y\rangle = \operatorname{tr}(x \circ y)$. Let x be an element in V and $x = \sum_{i=1}^{r} \lambda_i(x) c_i$ be its spectral decomposition, where $\lambda_i(x)$ with $i = 1 \ldots, r$ are the eigenvalues of x, such that $\lambda_r(x) \in \mathbb{R}^r_\downarrow$ and $c_1, \ldots, c_r$ is a Jordan frame for x. As before, $\mathcal{K}$ is the symmetric cone associated to V. Moreover, we assume that $x \in \mathcal{K}^0$, which means, by Proposition 2.5.10, that $\lambda_i(x) > 0$ for $i = 1, \ldots, r$.

We can now extend the kernel functions to Euclidean Jordan algebras. Let ψ be a kernel function as defined in the previous section. Since the eigenvalues of any element of a Euclidean Jordan algebra are real (see Theorem 2.4.1), we can extend these real functions to Euclidean Jordan algebras. We define the function $\phi : \mathcal{K}^0 \mapsto V$ as

$$\phi(x) := \sum_{i=1}^{r} \psi(\lambda_i(x)) c_i.$$

The *barrier function* induced by the kernel function is now defined by,

$$\Psi(x) := \operatorname{tr}(\phi(x)) = \sum_{i=1}^{r} \psi(\lambda_i(x)). \tag{4.5}$$

So, if we define, for $a \in \mathbb{R}^r$,

$$\bar{\Psi}(a) := \sum_{i=1}^{r} \psi(a_i)$$

then

$$\Psi(x) = (\bar{\Psi}\lambda)(x),$$

thus making clear that Ψ is a separable spectral function (cf. Section 3.5).

The next lemma establishes that e-convex functions preserve systems of inequalities such as (3.3).

Lemma 4.3.1 (Corollary 3.3.10 in [25]). *Let f be a real function and $\alpha_1, \ldots, \alpha_n$, $\beta_1, \ldots, \beta_n$ be $2n$ given nonnegative real numbers such that $\alpha_1 \geq \cdots \geq \alpha_n \geq 0$ and $\beta_1 \geq \cdots \geq \beta_n \geq 0$. If*

$$\prod_{i=1}^{k} \alpha_i \leq \prod_{i=1}^{k} \beta_i, \quad k = 1, \ldots, n-1, \tag{4.6}$$

$$\prod_{i=1}^{n} \alpha_i = \prod_{i=1}^{n} \beta_i \tag{4.7}$$

and $f(t)$ is e-convex on the interval $[\beta_n, \beta_1]$, then

$$\sum_{i=1}^{n} f(\alpha_i) \leq \sum_{i=1}^{n} f(\beta_i).$$

Suppose now that ψ is e-convex. Our next theorem presents an appealing property of barrier functions induced by e-convex kernel functions. It is a crucial result of the thesis.

Theorem 4.3.2. *Let $\Psi(v)$ be the function defined in (4.5) and $x, s \in \mathcal{K}^0$. Then*

$$\Psi((P(x)^{1/2}s)^{1/2}) \leq \frac{1}{2}(\Psi(x) + \Psi(s)). \tag{4.8}$$

Proof. Let $u = P(x)^{1/2}s$, then, using $\lambda_i(u^{1/2}) = \lambda_i^{1/2}(u)$,

$$\Psi(u^{1/2}) = \sum_{i=1}^{r} \psi(\lambda_i^{1/2}(u)).$$

Since by Theorem 3.3.2 the real numbers

$$\alpha_i := \lambda_i^{1/2}(u) > 0, \qquad i = 1, \ldots, r$$

and

$$\beta_i := \lambda_i^{1/2}(x)\lambda_i^{1/2}(s) > 0, \qquad i = 1, \ldots, r$$

satisfy inequality (4.6) and equality (4.7), and $\psi(t)$ is e-convex, we have

$$\sum_{i=1}^{r} \psi(\lambda_i^{1/2}(u)) \leq \sum_{i=1}^{r} \psi(\lambda_i^{1/2}(x)\lambda_i^{1/2}(s)).$$

By Lemma 4.2.1, it follows that

$$\sum_{i=1}^{r} \psi(\lambda_i^{1/2}(x)\lambda_i^{1/2}(s)) \leq \sum_{i=1}^{r} \big(\frac{1}{2}\psi(\lambda_i(x)) + \frac{1}{2}\psi(\lambda_i(s))\big).$$

Therefore

$$\Psi(u^{1/2}) \leq \frac{1}{2}(\Psi(x) + \Psi(s)).$$

The result is proved. ∎

4.4 Derivatives of the barrier function

Let $x(t) := x_0 + tu$ with $t \in \mathbb{R}$ and $u \in V$ and assume that $x(t) \in \mathcal{K}^0$. Let $x(t) = \sum_{i=1}^{r} \lambda_i(x(t))c_i$ be the spectral decomposition of $x(t)$.

Our following aim is to obtain expressions for $D_t\Psi(x(t))$ and $D_t^2\Psi(x(t))$. For the purpose we use Theorem 3.6.2 and Lemma 3.6.3, respectively. By Theorem 3.6.2 we have

$$D_t\Psi(x(t)) = \operatorname{tr}\left(D_x\Psi(x(t)) \circ x'(t)\right) = \operatorname{tr}\left(\sum_{i=1}^{r} \psi(\lambda_i(x(t)))c_i \circ u\right). \tag{4.9}$$

In order to obtain $D_t^2\Psi(x)$, we first rewrite (3.19):

$$D_xG(x) = \sum_{i=1}^{r} f'(\lambda_i)P_{ii} + \sum_{\substack{j<k\\ \lambda_j=\lambda_k}} f'(\lambda_j)P_{jk} + \sum_{\substack{j<k\\ \lambda_j\neq\lambda_k}} \frac{f(\lambda_j)-f(\lambda_k)}{\lambda_j-\lambda_k}P_{jk}.$$

By Theorem 3.6.2, $D_x\Psi(x) = \sum_{i=1}^{r} \psi'(\lambda_i(x))c_i$. Hence, applying (3.19) to $D_x\Psi(x)$ we get

$$D_x^2\Psi(x) = \sum_{i=1}^{r} \psi''(\lambda_i)P_{ii} + \sum_{\substack{j<k\\ \lambda_j=\lambda_k}} \psi''(\lambda_j)P_{jk} + \sum_{\substack{j<k\\ \lambda_j\neq\lambda_k}} \frac{\psi'(\lambda_j)-\psi'(\lambda_k)}{\lambda_j-\lambda_k}P_{jk}.$$

Since $D_tx(t) = u$, we have $D_t\Psi(x(t)) = \langle D_x\Psi(x(t)), u\rangle$. From here

$$D_t^2\Psi(x(t)) = \langle D_x^2\Psi(x(t))u, u\rangle,$$

i.e.,

$$D_t^2\Psi(x(t)) = \sum_{i=1}^{r} \psi''(\lambda_i)\langle P_{ii}u,u\rangle + \sum_{\substack{j<k\\ \lambda_j=\lambda_k}} \psi''(\lambda_j)\langle P_{jk}u,u\rangle + \sum_{\substack{j<k\\ \lambda_i\neq\lambda_j}} \frac{\psi'(\lambda_j)-\psi'(\lambda_k)}{\lambda_j-\lambda_k}\langle P_{jk}u,u\rangle.$$

We can easily verify that

$$\langle P_{ii}u, u\rangle = \langle P_{ii}u, P_{ii}u\rangle = \langle u_ic_i, u_ic_i\rangle = u_i^2,$$

and

$$\langle P_{ij}u, u\rangle = \operatorname{tr}\left(u_{ij}^2\right),$$

because P_{ii} and P_{jk} are orthogonal projections onto the Peirce spaces $V(c_i, 1)$ and V_{jk}, respectively. Therefore,

$$D_t^2\Psi(x(t)) = \sum_{i=1}^{r} \psi''(\lambda_i)u_i^2 + \sum_{\substack{j<k\\ \lambda_j=\lambda_k}} \psi''(\lambda_j)\operatorname{tr}\left(u_{jk}^2\right) + \sum_{\substack{j<k\\ \lambda_j\neq\lambda_k}} \frac{\psi'(\lambda_j)-\psi'(\lambda_k)}{\lambda_j-\lambda_k}\operatorname{tr}\left(u_{jk}^2\right). \tag{4.10}$$

We can get a upper bound for $D_t^2\Psi(x(t))$ that will allows us to work with a simpler expression.

Proposition 4.4.1. *One has*

$$D_t^2\Psi(x(t)) \leq \sum_{i=1}^{r} \psi''(\lambda_i)u_i^2 + \sum_{j<k} \psi''(\lambda_k)\mathrm{tr}\left(u_{jk}^2\right).$$

Proof. Note that by the well known mean value theorem, there exists $\beta \in (\lambda_k, \lambda_j)$ such that

$$\frac{\psi'(\lambda_j) - \psi'(\lambda_k)}{\lambda_j - \lambda_k} = \psi''(\beta).$$

But this can be bounded by $\psi''(\lambda_k)$ because Condition 4.2 implies that $\psi''(t)$ is monotonically decreasing, and we have assumed that for $j < k$ such that $\lambda_j \neq \lambda_k$ we have $\lambda_j > \lambda_k$. Since the second term of the right-hand side in equation (4.10) is for $\lambda_k = \lambda_j$, the result immediately follows. ∎

4.5 Conclusion

In this chapter we introduced a barrier function that is generated by an eligible kernel function. The inequality (4.8) is crucial in the analysis of the algorithm presented in the following chapter. Up till now this inequality was known only for the cone of positive semidefinite matrices and for the second-order cone.

Chapter 5

Interior-point methods based on kernel functions

5.1 Introduction

This chapter introduces interior-point methods for symmetric optimization based on the kernel functions introduced in the previous chapter. We first recall the definition of the symmetric optimization problem and the necessary and sufficient conditions for optimality. Then we present the so-called Nesterov-Todd (NT) direction adapted to the case of symmetric optimization. After defining the search direction that uses kernel functions, we present and analyze the algorithm. Later, the kernel functions discovered so far are listed. We conclude the chapter with some notes.

5.2 Symmetric optimization problem

As before, we consider a n-dimensional Euclidean Jordan $\mathbb{R}$-algebra $(V, \circ, \langle \cdot, \cdot \rangle)$ with rank r and $\mathcal{K}$ will denote the associated symmetric cone, and the inner product is defined by $\langle x, y \rangle = \operatorname{tr}(x \circ y)$. The function ψ will denote an eligible kernel function, and Ψ the associated barrier function, as defined in Section 4.3.

We consider the following primal-dual pair of optimization problems,

$$\min\{\langle c, x \rangle : \langle a_i, x \rangle = b_i,\ i = 1, \ldots, m,\ x \in \mathcal{K}\} \tag{5.1}$$

$$\max\{b^T y : \sum_{i=1}^{m} y_i a_i + s = c,\ s \in \mathcal{K},\ y \in \mathbb{R}^m\}, \tag{5.2}$$

where $c, a_i \in V$, for $i = 1, \ldots, m$, and $b \in \mathbb{R}^m$. We call $x \in \mathcal{K}$ *primal feasible* if $\langle a_i, x \rangle = b_i$ for $i = 1, \ldots, m$. $(y, s) \in \mathbb{R}^m \times \mathcal{K}$ is called *dual feasible* if $\sum_{i=1}^{m} y_i a_i + s = c$.

Let $A \in \mathbb{R}^{m\times n}$ be the matrix corresponding to the linear transformation that maps x to the m-vector whose i^{th} component is $\langle a_i, x\rangle$. Then the sets of primal and dual interior feasible solutions are given by

$$\mathcal{F}_p := \{x \in V : Ax = b, x \in \mathcal{K}^0\}, \tag{5.3}$$

$$\mathcal{F}_d := \{(y, s) \in \mathbb{R}^m \times V : A^T y + s = c,\ s \in \mathcal{K}^0,\ y \in \mathbb{R}^m\}.$$

We say that the optimization problems (5.1) and (5.2) are *strictly feasible* if $\mathcal{F}_p$ and $\mathcal{F}_d$ are nonempty, respectively.

5.3 Duality

In this section we recall the conditions for optimality.

We start to define *duality gap*: the difference between the primal and dual objective values at feasible solutions of (5.1) and (5.2).

Theorem 5.3.1 (Weak duality). *Let x and (y, s) be primal and dual feasible, respectively. One has*

$$\langle c, x\rangle - b^T y \geq 0,$$

i.e the duality gap is nonnegative at feasible solutions.

Proof. Replacing c by $A^T y + s$ and b by Ax, we have

$$\langle c, x\rangle - b^T y = \langle A^T y + s, x\rangle - (Ax)^T y = \langle s, x\rangle + \langle A^T y, x\rangle - (Ax)^T y.$$

Since $Ax, y \in \mathbb{R}^m$, $\langle Ax, y\rangle = (Ax)^T y$. Hence,

$$\langle A^T y, x\rangle = \langle y, Ax\rangle = (Ax)^T y.$$

It follows that

$$\langle c, x\rangle - b^T y = \langle s, x\rangle \geq 0,$$

where the inequality follows from the self-duality of $\mathcal{K}$. ∎

In the following we state the strong duality theorem.

Theorem 5.3.2 (Strong duality). *Let*

$$p^* := \inf\{\langle c, x\rangle : Ax = b, x \in \mathcal{K}\}$$

and

$$d^* := \sup\{b^T y : A^T y + s = c, s \in \mathcal{K}\}.$$

If there exists a strictly feasible solution $(\hat{y}, \hat{s})$ for (5.2) and d^* is finite, then $p^* = d^*$ and p^* is attained for some x such that $Ax = b$ and $x \in \mathcal{K}$.

We omit the proof because it is exactly the same as the proof of the conic duality theorem as given in [11]. We only mention that the proof does not use that V is a Euclidean Jordan algebra.

We use that V is a Euclidean Jordan algebra in the next result.

Lemma 5.3.3 (Lemma 2.2 in [18]). *Let* $(x,s) \in \mathcal{K} \times \mathcal{K}$ *and* $\langle x, s\rangle = 0$. *Then*

$$x \circ s = 0. \tag{5.4}$$

Proof. Since $s \in \mathcal{K}$, $s = z^2$ for some $z \in V$. So we may write, using the associativity of the inner product,

$$0 = \langle x, z^2\rangle = \langle z, x \circ z\rangle = \langle z, L(x)z\rangle.$$

We will derive from this that $L(x)z = 0$. As in the proof of Theorem 2.4.1, we can write $L(x) = \sum_{i=1}^{k} \lambda_i P_i$, where λ_i, $i = 1, \ldots, k$ are the eigenvalues of $L(x)$ and P_i, $i = 1, \ldots, k$ are non-zero orthogonal projections. So,

$$0 = \langle z, L(x)z\rangle = \langle z, \sum_{i=1}^{k} \lambda_i P_i z\rangle = \sum_{i=1}^{k} \lambda_i \langle z, P_i z\rangle = \sum_{i=1}^{k} \lambda_i \langle P_i z, P_i z\rangle = \sum_{i=1}^{k} \lambda_i \|P_i z\|^2.$$

Since $x \in \mathcal{K}$, by Proposition 2.5.10, the eigenvalues of x are nonnegative. This implies, by Corollary 2.9.12, that the eigenvalues of $L(x)$ are also nonnegative. Thus, $\lambda_i \|P_i z\|^2 = 0$ for all $i = 1, \ldots, k$. Hence, $\lambda_i = 0$ or $P_i z = 0$, for all $i = 1, \ldots, k$. Therefore we obtain $L(x)z = 0$. Furthermore,

$$\langle s, L(x)s\rangle = \langle z^2, x \circ z^2\rangle = \langle z^4, x\rangle = \langle z^3, x \circ z\rangle = \langle z^3, L(x)z\rangle = 0,$$

which using the same arguments as before, implies that $L(x)s = x \circ s = 0$. ■

Note that if $x \circ s = 0$ then obviously $\langle x, s\rangle = 0$. Thus, we have that for $(x,s) \in \mathcal{K} \times \mathcal{K}$, $\langle x, s\rangle = 0$ if and only if $x \circ s = 0$.

The following assumptions will be made through this manuscript.

Assumption 1 The vectors a_i are linearly independent.

Assumption 2 The problems (5.1) and (5.2) are strictly feasible.

Assumption 1 and 2 are not much restrictive. If the vectors a_i are not linearly independent, we can reduce the problem to a subset of a_i's which are linearly independent. For the case that the primal and/or dual problems are not strictly feasible we can embed both problems in a self-dual problem that is strictly feasible, as has been shown in e.g., [32, 49]. As a consequence of these assumptions there exist optimal solutions for (5.1) and (5.2) with duality gap 0.

5.4 Scaling

In this section we show the existence and the uniqueness of a scaling point w corresponding to any points $x, s \in \mathcal{K}^0$, such that $P(w)$ takes s into x. This was done for self-scaled cones in [38]. Faybusovich [16] also showed it using the framework of the Euclidean Jordan algebras.

Let the *geometric mean* $a\#b$ of $a, b \in \mathcal{K}^0$ be defined as

$$a\#b := P(a^{\frac{1}{2}})(P(a^{-\frac{1}{2}})b)^{\frac{1}{2}}.$$

We have $a\#b \in \mathcal{K}^0$, because, by Proposition 2.5.7, $P(a^{-\frac{1}{2}})b \in \mathcal{K}^0$ and also

$$P(a^{\frac{1}{2}})(P(a^{-\frac{1}{2}})b)^{\frac{1}{2}} \in \mathcal{K}^0.$$

Remark 5.4.1. The name geometric mean can be explained in the following way: let V be the Euclidean Jordan algebra given in Example 2.4.3. Since for $a, b \in V$,

$$P(a) = \text{Diag}(a_1^2; \ldots; a_r^2),$$

we have in that case

$$a\#b = \text{Diag}(a_1; \ldots; a_r)(a_1^{-1}b_1; \ldots; a_r^{-1}b_r)^{1/2} = (a_1^{1/2}b_1^{1/2}; \ldots; a_r^{1/2}b_r^{1/2}).$$

So, the components of the last vector $(a_ib_i)^{1/2}$, with $i = 1 \ldots, r$, are the geometric mean of the components a_i and b_i. □

Proposition 5.4.2 (Proposition 2.4 in [31]). *Let $a, b \in \mathcal{K}^0$. Then $a\#b$ is the unique solution which belongs to $\mathcal{K}^0$ of the following equation in x:*

$$P(x)a^{-1} = b. \tag{5.5}$$

Proof. With $x = a\#b$ using the fundamental formula (2.17) of the quadratic representation, we may write

$$\begin{aligned} P(x)a^{-1} &= P(P(a^{\frac{1}{2}})(P(a^{-\frac{1}{2}})b)^{\frac{1}{2}})a^{-1} \\ &= P(a^{\frac{1}{2}})P(P(a^{-\frac{1}{2}})b)^{\frac{1}{2}})P(a^{\frac{1}{2}})a^{-1} \\ &= P(a^{\frac{1}{2}})P(P(a^{-\frac{1}{2}})b)^{\frac{1}{2}})e \\ &= P(a^{\frac{1}{2}})P(a^{-\frac{1}{2}})b = b. \end{aligned}$$

Now suppose that $P(x)a^{-1} = P(y)a^{-1}$ for some $y \in \mathcal{K}^0$. Then

$$P(P(x)a^{-1}) = P(P(y)a^{-1}),$$

which by (2.17) implies that

$$P(x)P(a^{-1})P(x) = P(y)P(a^{-1})P(y).$$

By Proposition 2.5.7, $P(x)$, $P(y)$, $P(a^{-1})$ are positive definite. Therefore, using Lemma B.1.3, $P(x) = P(y)$. It follows that

$$x^2 = P(x)e = P(y)e = y^2.$$

By Theorem 2.4.2 we can write $x^2 = \sum_{i=1}^r \lambda_i c_i = y^2$. Since $x, y \in \mathcal{K}^0$, the eigenvalues of x and y are positive. Thus,

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

Everything is proved. ■

The geometric mean is commutative: $a\#b = b\#a$. This can be checked by verifying that $b\#a$ is also a solution of the quadratic equation (5.5).

We conclude that for given $(x, s) \in \mathcal{K}^0 \times \mathcal{K}^0$, there exists a unique $w \in \mathcal{K}^0$ such that

$$P(w)s = x,$$

namely

$$w = x\#s^{-1}.$$

The point w is called the *scaling point* of x and s. It coincides with the unique scaling point introduced by Nesterov and Todd [38] for self-scaled cones. Below we present a few examples.

Example 5.4.3. Let $V = \mathbb{R}^n$ and $\mathcal{K}$ be the linear cone. For any $x, s \in \mathcal{K}^0$, we saw in Remark 5.4.1 that

$$x\#s^{-1} = \left(\frac{x_1^{1/2}}{s_1^{1/2}}, \dots, \frac{x_r^{1/2}}{s_r^{1/2}} \right). \tag{5.6}$$

Thus, the unique scaling point w is given by (5.6). □

Example 5.4.4. Let $V = \mathbf{S}^n$ and $\mathcal{K}$ be the cone of real positive semidefinite matrices. For any $X, S \in \mathcal{K}^0$ we have $P(X)S = XSX$ (cf. Example 2.3.3). Thus,

$$\begin{aligned} W &= X\#S^{-1} \\ &= P(X^{1/2})(P(X^{-1/2})S^{-1})^{1/2} \\ &= P(X^{1/2})(X^{-1/2}S^{-1}X^{-1/2})^{1/2} \\ &= X^{1/2}(X^{1/2}SX^{1/2})^{-1/2}X^{1/2}. \end{aligned}$$

The point W coincides with scaling point given in [38]. □

Example 5.4.5. Let $V = \mathbb{R}^{n+1}$ and $\mathcal{K}$ be the second-order cone. Using the matrix representation $L(x)$ as given in Example 2.3.2, we can easily get that $L^2(x) = L(x^2)$. Thus

$$P(x)s = \left(2L^2(x) - L(x^2)\right) s = L(x^2)s = x^2 \circ s,$$

with $x, s \in V$. So, for any $x, s \in \mathcal{K}^0$ we have

$$\begin{aligned} w &= x \# s^{-1} \\ &= P(x^{1/2})(P(x^{-1/2})s^{-1})^{1/2} \\ &= x \circ (P(x^{-1/2})s^{-1})^{1/2} \\ &= x \circ P(x^{1/2})s)^{-1/2} \\ &= x \circ (x \circ s)^{-1/2}, \end{aligned}$$

where the fourth equality follows from Proposition 2.3.8-(ii). □

5.5 The central path

In this section, we introduce the concept of central path for the primal and dual problems (5.1) and (5.2). We will prove the existence and uniqueness of the central path.

Recall from Section 5.2 the definition of the symmetric optimization problem. As we have seen in Section 5.3, under assumptions 1 and 2, the conditions

$$\begin{cases} \langle a_i, x \rangle &= b_i, \quad i = 1, \ldots, m, \; x \in \mathcal{K} \\ \sum_{i=1}^m y_i a_i + s &= c \qquad\qquad\qquad\quad s \in \mathcal{K} \\ x \circ s &= 0. \end{cases}$$

are necessary and sufficient for optimality. We perturb the optimality conditions, more precisely, the so-called complementary condition, $x \circ s = 0$, by introducing a parameter $\mu > 0$, as follows

$$\begin{cases} \langle a_i, x \rangle &= b_i, \quad i = 1, \ldots, m, \; x \in \mathcal{K} \\ \sum_{i=1}^m y_i a_i + s &= c \qquad\qquad\qquad\quad s \in \mathcal{K} \\ x \circ s &= \mu e. \end{cases} \tag{5.7}$$

Now we will prove that (5.7) has a unique solution for each $\mu > 0$. The function

$$f_p^\mu(x) := \frac{\langle c, x \rangle}{\mu} - \log\det x, \quad x \in \mathcal{K}^0$$

is the so-called *primal logarithmic barrier* function.

This function is strictly convex. Just note that the Hessian of $-\log\det x$ is $P(x)^{-1}$ (Proposition 2.6.1), which is positive definite for $x \in \mathcal{K}^0$. This implies that $-\log\det x$ is strictly convex. Since $\langle c, x \rangle$ is linear in x the strictly convexity of f_p^μ follows.

The existence of the minimizer of f_p^μ is proved below.

Theorem 5.5.1. *If $\mathcal{F}_p$ and $\mathcal{F}_d$ are nonempty then f_p^μ has a unique minimizer, on the strictly feasible region $\mathcal{F}_p$.*

Proof. By hypothesis, there exist vectors $x^0 \in \mathcal{F}_p$ and $(y^0, s^0) \in \mathcal{F}_d$. Taking $K = f_p^\mu(x^0)$ and defining the level set of $f_p^\mu(x)$ by

$$\mathcal{L}_K = \{x \in \mathcal{F}_p : f_p^\mu(x) \leq K\},$$

we have that $x^0 \in \mathcal{L}_K$, so $\mathcal{L}_K$ is not empty. Since f_p^μ is strictly convex, if f_p^μ has a minimizer then it is unique. The existence of the minimizer can be proved by showing that $\mathcal{L}_K$ is compact. By continuity of f_p^μ, $\mathcal{L}_K$ is closed. It remains to prove that $\mathcal{L}_K$ is bounded. Let $x \in \mathcal{L}_K$. Using Proposition 5.3.1 we have

$$\langle c, x\rangle - b^T y^0 = \langle x, s^0\rangle,$$

so, in the definition of $f_p^\mu(x)$ we may replace $\langle c, x\rangle$ by $b^T y^0 + \langle x, s^0\rangle$:

$$f_p^\mu(x) = \frac{\langle c, x\rangle}{\mu} - \log\det x = \frac{1}{\mu} b^T y^0 + \frac{1}{\mu}\langle x, s^0\rangle - \log\det x.$$

Since

$$\langle x, s^0\rangle = \operatorname{tr}\left(x \circ s^0\right) = \operatorname{tr}\left(P(x)^{1/2} s^0\right) = \langle P(x)^{1/2} s^0, e\rangle,$$

$\operatorname{tr}(e) = r$ and $\det(P(x)^{1/2} s^0) = \det(x)\det(s^0)$ (cf. Proposition 2.5.12), it follows that

$$f_p^\mu(x) = \langle e, \frac{P(x)^{1/2} s^0}{\mu} - e\rangle - \log\det \frac{P(x)^{1/2} s^0}{\mu} + r - r\log\mu + \frac{1}{\mu} b^T y^0 + \log\det s^0,$$

or equivalently,

$$\langle e, \frac{P(x)^{1/2} s^0}{\mu} - e\rangle - \log\det \frac{P(x)^{1/2} s^0}{\mu} = f_p^\mu(x) - r + r\log\mu - \frac{1}{\mu} b^T y^0 - \log\det s^0.$$

Hence, using $f_p^\mu(x) \leq K$ and defining $\bar{K}$ by

$$\bar{K} := K - r + r\log\mu - \frac{1}{\mu} b^T y^0 - \log\det s^0,$$

we obtain

$$\langle e, \frac{P(x)^{1/2} s^0}{\mu} - e\rangle - \log\det \frac{P(x)^{1/2} s^0}{\mu} \leq \bar{K}. \tag{5.8}$$

Note that $\bar{K}$ does not depend on x. Now let the function $\psi : (0, +\infty) \mapsto \mathbb{R}$ be the function defined

$$\phi(t) := t - 1 - \log t.$$

If we define $\Psi(u) := \sum_{i=1}^r \psi(\lambda_i(u))$, then we may rewrite (5.8) as follows

$$\Psi\left(\frac{P(x)^{1/2} s^0}{\mu}\right) \leq \bar{K}.$$

In fact, ψ is an eligible kernel function (cf. [20]). Thus ψ is nonnegative in its domain which implies that

$$\psi\left(\lambda_i\left(\frac{P(x)^{1/2}s^0}{\mu}\right)\right) \leq \bar{K}.$$

also that there exist $t_1 < 1$ and $t_2 > 1$, such that

$$\psi(t_1) = \psi(t_2) = \bar{K}.$$

We conclude that

$$t_1 \leq \lambda_i\left(\frac{P(x)^{1/2}s^0}{\mu}\right) \leq t_2 \qquad i = 1, \ldots, r.$$

Adding these expressions, for all i, we obtain

$$rt_1 \leq \operatorname{tr}\left(\frac{P(x)^{1/2}s^0}{\mu}\right) \leq rt_2.$$

Using Lemma 3.3.4 we obtain

$$0 < \operatorname{tr}(x) \leq \frac{rt_2\mu}{\lambda_{\min}(s^0)},$$

where $\lambda_{\min}(s^0)$ denotes the minimum eigenvalue of s^0. Since $x \in \mathcal{K}^0$, the eigenvalues of x are positive (cf. Proposition 2.5.10), which implies,

$$\|x\| = \sqrt{\sum_{i=1}^{r} \lambda_i^2(x)} \leq \sqrt{\left(\sum_{i=1}^{r} \lambda_i(x)\right)^2} = \operatorname{tr}(x).$$

Therefore the level set $\mathcal{L}_K$ is bounded. We conclude that $\mathcal{L}_K$ is compact. ■

The existence and uniqueness of solution of the system (5.7) is guaranteed by the following proposition.

Proposition 5.5.2. *For each $\mu > 0$ there exists a unique solution of the system (5.7), denoted as $(x(\mu), y(\mu), s(\mu))$.*

Proof. Consider the following problem

$$\min_{x \in \mathcal{K}^0} \left\{ f_p^\mu(x) := \frac{\langle c, x\rangle}{\mu} - \log\det x : \langle a_i, x\rangle = b_i,\ i = 1, \ldots, r \right\},$$

i.e. the minimization of the primal logarithmic barrier function over $\mathcal{F}_p$. Since the function f_p^μ is strictly convex and we assumed that the problem (5.1) is strictly feasible, the constraint qualifications hold for this problem. So, the KKT (first order) optimality conditions for this problem, with feasible x, are

$$\nabla f_p^\mu(x) - \sum_{i=1}^m \hat{y}_i \nabla(\langle a_i, x\rangle - b_i) = 0,$$

which is equivalent to the system

$$\begin{aligned} &\frac{1}{\mu}c - x^{-1} - \sum_{i=1}^m \hat{y}_i a_i = 0 \\ &\langle a_i, x\rangle - b_i = 0 \quad i = 1, \ldots, m. \end{aligned}$$

See Proposition 2.6.1 for the gradient of $\log\det x$. Defining $s = c - \sum_{i=1}^m y_i a_i$ where $y_i = \mu\hat{y}_i$, this system is the same as (5.7), provided that $x \circ s = \mu e$. So it remains to show that $s = \mu x^{-1}$ if and only if $x \circ s = \mu e$. It is straightforward to see that $s = \mu x^{-1}$ implies $x \circ s = \mu e$. To show the converse, note that μx^{-1} is a solution of the equation $x \circ s = \mu e$. On the other hand, since $x \in \mathcal{K}^0$, by Corollary 2.9.12, $L(x)$ is a positive definite operator. Thus implying that $L(x)s = \mu e$ has a unique solution. It follows that KKT conditions are equivalent to the conditions (5.7). Since by Theorem 5.5.1 f_p^μ has a unique minimizer, this minimizer is given by the system (5.7). ∎

The solution of the perturbed system (5.7) defines a curve parameterized by μ, through the feasible region, which leads to the optimal set as $\mu \to 0$. This curve is called the *central path* and most interior-point methods approximately follow the central path to reach the optimal set.

5.6 The Nesterov-Todd direction

Nesterov and Todd in [38] introduced the so-called Nesterov-Todd (shortly NT) direction. It is defined as a triple of vectors $(\Delta x, \Delta y \Delta s) \in V \times V$ that is uniquely defined by the conditions:

$$\left\{\begin{array}{rcll} \langle a_i, \Delta x\rangle &=& 0, & i = 1, \ldots, m, \\ \sum_{i=1}^m \Delta y_i a_i + \Delta s &=& 0 & \\ \Delta x + (F''(w))^{-1}\Delta s &=& -x + \mu s^{-1}, & \end{array}\right.$$

where $F''(w)$ is the Hessian of the barrier function $F(x) := -\log\det x$ for $\mathcal{K}$ at $w = x \# s^{-1}$. Their theory applies to optimization over self-scaled cones, but as we have mentioned earlier, it was shown in [22] that a cone is self-scaled if and only if it is symmetric. Thus, using Proposition 2.6.1, i.e., the fact that $F''(x) = P(x)^{-1}$ for $x \in \mathcal{K}^0$, the NT direction in terms of Euclidean Jordan algebra can be defined by the system

$$\left\{\begin{array}{rcll} \langle a_i, \Delta x\rangle &=& 0, & i = 1, \ldots, m, \\ \sum_{i=1}^m \Delta y_i a_i + \Delta s &=& 0 & \\ \Delta x + P(w)\Delta s &=& -x + \mu s^{-1}, & \end{array}\right. \tag{5.9}$$

This system for defining the NT direction in the framework of Euclidean Jordan algebras was obtained first in [16]. Note that the first and the second equations imply that

$$\langle \Delta x, \Delta s \rangle = 0.$$

We will use the scaling point w to scale the NT direction. We set

$$v := \frac{1}{\sqrt{\mu}} P(w)^{-\frac{1}{2}} x = \frac{1}{\sqrt{\mu}} P(w)^{\frac{1}{2}} s. \tag{5.10}$$

Then the last equation from (5.9) can be written as,

$$\Delta x + P(w)\Delta s = -\sqrt{\mu} P(w)^{\frac{1}{2}} v + \frac{\mu}{\sqrt{\mu}} P(w)^{\frac{1}{2}} v^{-1}.$$

Applying $P(w)^{-\frac{1}{2}}$ to both sides of this equation, we can rewrite it in the form:

$$d_x + d_s = v^{-1} - v,$$

where

$$d_x = \frac{1}{\sqrt{\mu}} P(w)^{-\frac{1}{2}} \Delta x \tag{5.11}$$

and

$$d_s = \frac{1}{\sqrt{\mu}} P(w)^{\frac{1}{2}} \Delta s. \tag{5.12}$$

Thus, we can rewrite the system that defines the NT direction for symmetric cones (or self-scaled cones) as

$$\left\{ \begin{array}{rcll} \langle \bar{a}_j, d_x \rangle & = & 0 & j = 1, \ldots, m, \\ -\sum_{j=1}^m \Delta y_j \bar{a}_j & = & d_s & \\ d_x + d_s & = & v^{-1} - v & \end{array} \right. \tag{5.13}$$

where $\bar{a}_j = \frac{1}{\sqrt{\mu}} P(w)^{\frac{1}{2}} a_j$.

In the next section we will use the characterization (5.13) of the NT direction to derive a new direction.

5.7 A new search direction for symmetric optimization

We now are ready to generalize the search direction introduced in [41] for LO to symmetric optimization, using the framework of Euclidean Jordan algebras.

The so-called *dual logarithmic barrier* function is given by

$$f_d^\mu(y, s) := \frac{1}{\mu} b^T y + \log\det s, \qquad (y, s) \in \mathcal{F}_d.$$

Combining the primal and dual logarithmic barrier functions, we define the *primal-dual logarithmic barrier* function as

$$\begin{aligned} f^\mu(x,s) &:= f_p^\mu(x) - f_d^\mu(y,s) - r + r\log\mu \\ &= \frac{1}{\mu}\langle c,x\rangle - \frac{1}{\mu}b^Ty - \log\det x - \log\det s - r + r\log\mu \\ &= \frac{1}{\mu}\langle x,s\rangle - \log\det\frac{P(x)^{1/2}s}{\mu} - r. \end{aligned}$$

This can be written in terms of v as,

$$\begin{aligned} f^\mu(x,s) &= \langle P(w)^{\frac{1}{2}}v, P(w)^{-\frac{1}{2}}v\rangle - \log\det((P(w)^{\frac{1}{2}}v)\det((P(w)^{-\frac{1}{2}}v) - r \\ &= \langle v,v\rangle - \log(\det(w)\det(v)\det(w)^{-1}\det(v)) - r \\ &= \langle v^2,e\rangle - 2\log\det v - r \\ &= \langle v^2-e,e\rangle - 2\log\det v, \end{aligned}$$

which is called the v-scaled primal-dual logarithmic barrier function and will be denoted as $\Phi(v)$.

Remark 5.7.1. The function $\Phi(v)$ can be generated by a kernel function. If $v = \sum_{i=1}^r \lambda_i(v)c_i$ is a spectral decomposition of v, then

$$\langle v^2-e,e\rangle = \operatorname{tr}\left(v^2-e\right) = \operatorname{tr}\left(\sum_{i=1}^r \lambda_i^2(v)c_i - \sum_{i=1}^r c_i\right) = \sum_{i=1}^r(\lambda_i^2(v)-1)$$

and

$$\log\det v = \log\prod_{i=1}^r \lambda_i(v) = \sum_{i=1}^r \log\lambda_i(v).$$

Thus we obtain

$$\Phi(v) = \sum_{i=1} \psi(\lambda_i(v)),$$

with $\psi(t) = t^2 - 1 - 2\log t$. The funcion $\frac{1}{2}\psi(t)$ is an eligible kernel function (cf. [9]). □

In the case of LO it was observed that the gradient of $\Phi(v)$ and the right-hand side of the last equation in the system (5.13) coincide (cf. [9]). It is clear that here the same holds:

$$\frac{1}{2}\nabla\Phi(v) = v - v^{-1},$$

using Theorem 3.6.2. This coincidence motivated a new search direction, which is defined by the following system

$$\begin{cases} \langle \bar{a}_j, d_x\rangle = 0 & j = 1,\ldots,m, \\ \sum \Delta y_j \bar{a}_j + d_s = 0 & \\ d_x + d_s = -\nabla\Psi(v), & \end{cases} \tag{5.14}$$

where $\Psi(v)$ is the barrier function induced by a kernel function as defined in Section 4.3 and $\nabla\Psi(v)$ its gradient. At this point, it is worth to verify what happens to v when x and s are in the central path.

Proposition 5.7.2. *For primal and dual strictly feasible points x and s we have*

$$x \circ s = \mu e \iff v = e.$$

Proof. By the proof of the Proposition 5.5.2 we know that $x \circ s = \mu e$ is equivalent to $s = \mu x^{-1}$. By the definition (5.10) of v it follows that $s = \mu x^{-1}$ is equivalent to

$$\sqrt{\mu} P(w)^{-1/2} v = \frac{\mu}{\sqrt{\mu}} \left(P(w)^{1/2} v \right)^{-1}.$$

It follows that

$$P(w)^{-1/2} v = P(w)^{-1/2} v^{-1},$$

and we obtain $v^2 = e$. If $v = \sum_{i=1}^{r} \lambda_i(v) c_i$ then $v^2 = \sum_{i=1}^{r} \lambda_i^2(v) c_i = \sum_{i=1}^{r} c_i$. Hence, $\lambda_i^2(v) = 1$ which implies that $\lambda_i(v) = 1$, since $v \in \mathcal{K}$. Therefore $v = e$. The other implication follows analogously. ■

The system (5.14) defines uniquely the search direction. We prove this below.

Proposition 5.7.3. *Under Assumption 1 and Assumption 2, there is a unique solution to* (5.14).

Proof. It suffices to prove that the linear operator defined by the left-hand side of (5.14) is injective. Let

$$\begin{cases} \langle \bar{a}_j, d_x \rangle &= 0 \qquad j = 1, \ldots, m, \\ \sum \Delta y_j \bar{a}_j + d_s &= 0 \\ d_x + d_s &= 0. \end{cases}$$

Taking the inner product of the last equation with d_x, and using $\langle d_x, d_s \rangle = 0$, it follows

$$\|d_x\|^2 = 0,$$

implying $d_x = 0$. By the same reasoning it follows that $d_s = 0$. By Assumption 1 $\bar{a}_1, \ldots, \bar{a}_m$ are linearly independent, which implies that $\Delta y = 0$. Note that the system (5.14) is only well defined under Assumption 2. The result follows. ■

A solution of the system (5.14) returns d_x, Δy and d_s. The original search directions Δx and Δs are obtained using (5.11) and (5.12), i.e.,

$$\Delta x = \sqrt{\mu} P(w)^{\frac{1}{2}} d_x$$

and

$$\Delta s = \sqrt{\mu} P(w)^{-\frac{1}{2}} d_s.$$

5.8 The algorithm

The algorithm we propose in this section is adapted from [43].

We start to remark that since d_x and d_s are orthogonal, we will have

$$d_x = d_s = 0 \iff \nabla\Psi(v) = 0 \iff v = e \iff \Psi(v) = 0,$$

i.e., if and only if x and s belong to the central path. Hence, if

$$(x, y, s) \neq (x(\mu), y(\mu), s(\mu)),$$

then $(\Delta x, \Delta y, \Delta s)$ is nonzero.

By taking a step along the search direction, with a step size α defined by some line search rules, one constructs a new triple (x, y, s) according to

$$x_+ := x + \alpha\Delta x, \qquad y_+ := y + \alpha\Delta y, \qquad s_+ := s + \alpha\Delta s. \tag{5.15}$$

If necessary, this procedure is repeated until we find iterates (x, y, s) that are "close" enough to $(x(\mu), y(\mu), s(\mu))$. Then μ is reduced by the factor $1 - \theta$ and we apply the above method targeting at the new μ-center, and so on. This procedure is repeated until μ is small enough, say, until $r\mu \leq \epsilon$.

The closeness of (x, y, s) to $(x(\mu), y(\mu), s(\mu))$ is measured by the value of $\Psi(v)$, with τ as threshold value: if $\Psi(v) \leq \tau$, then we start a new *outer iteration*, performing an update of the barrier parameter; otherwise we enter an *inner iteration* by computing the search directions at the current iterates with respect to the current value of μ and apply (5.15) to get new iterates, see Figure 5.1.

When it stops, the algorithm returns a solution such that is τ-close (in the barrier function sense) to a point on the central path with $r\mu \leq \epsilon$. However, this does not directly imply that the accuracy of the solution (measured by the duality gap) is bounded by ϵ. We will discuss the accuracy of the solution in Section 5.12.

We can use as starting points (s^0, x^0) without loss of generality (cf. [32]).

The parameters τ, θ and the step size α should be chosen in such a way that the algorithm is "optimized" in the sense that the number of iterations required by the algorithm is as small as possible. The choice of the barrier update parameter θ plays an important role in both theory and practice of IPMs. Usually, if θ is a constant independent of the dimension r of the problem, for instance, $\theta = \frac{1}{2}$, then we call the algorithm a *large-update* (or *long-step*) method. If θ depends on the dimension of the problem, such as $\theta = \frac{1}{\sqrt{r}}$, then the algorithm is called a *small-update* (or *short-step*) method.

The choice of the step size α ($\alpha > 0$) is another crucial issue in the analysis of the algorithm. It has to be taken such that the closeness of the iterates to the current μ-center improves by a sufficient amount. In the theoretical analysis the step size α is usually given a value that depends on the closeness of the current iterates to the μ-center.

It is generally agreed that the total number of inner iterations required by the algorithm is an appropriate measure for its efficiency. This number will be referred

Generic Primal-Dual Algorithm for Symmetric Optimization

Input:
  A threshold parameter $\tau > 0$;
  an accuracy parameter $\epsilon > 0$;
  a fixed barrier update parameter θ, $0 < \theta < 1$;
  (x^0, s^0) and $\mu^0 = 1$ such that $\Psi(v^0) \leq \tau$.
begin
  $x := x^0$; $s := s^0$; $\mu := \mu^0$;
  while $r\mu \geq \epsilon$ **do** `%outer iteration`
  begin
    $\mu := (1-\theta)\mu$;
    while $\Psi(v) > \tau$ **do** `%inner iteration`
    begin
      $x := x + \alpha\Delta x$;
      $s := s + \alpha\Delta s$;
      $y := y + \alpha\Delta y$;
      $v := \frac{1}{\sqrt{\mu}}P(w)^{-\frac{1}{2}}x \left(= \frac{1}{\sqrt{\mu}}P(w)^{\frac{1}{2}}s\right)$;
    end
  end
end

Figure 5.1: Generic algorithm

as the *iteration complexity* of the algorithm; it will be described as a function of the rank of V, and the accuracy parameter ϵ.

As already pointed out, the barrier function $\Psi(v)$ not only serves to define the search direction, but also acts as a measure of closeness of the current iterates to the μ-center. In the analysis of the algorithm we also use the *norm-based proximity measure* $\delta(v)$ defined by

$$\delta(v) := \frac{1}{2}\|\nabla\Psi(v)\| = \frac{1}{2}\|d_x + d_s\|. \tag{5.16}$$

Note that since Ψ is strictly convex and minimal at $v = e$ we have

$$\Psi(v) = 0 \iff \delta(v) = 0 \iff v = e.$$

Thus, both measures are naturally determined by the kernel function.

5.9 Analysis of the algorithm

5.9.1 Growth behavior

At the start of each outer iteration of the algorithm, just before the μ-update, we have $\Psi(v) \leq \tau$. Due to the update of μ, the vector v is divided by the factor $\sqrt{1-\theta}$, with $0 < \theta < 1$, which in general leads to an increase in the value of $\Psi(v)$. Then, during the subsequent inner iterations, $\Psi(v)$ decreases until it passes the threshold τ again. Hence, during the course of the algorithm the largest values if $\Psi(v)$ occur just after the updates of μ. That is why we derive an estimate for the effect of a μ-update on the value of $\Psi(v)$. In other words, with $\beta = \frac{1}{\sqrt{1-\theta}} \geq 1$ we want to find an upper bound for $\Psi(\beta v)$ in terms of $\Psi(v)$.

We have the following result.

Theorem 5.9.1. *Let $\varrho : [0, +\infty) \mapsto [1, +\infty)$ be the inverse function of $\psi(t)$ for $t \geq 1$. Then we have for any vector $v \in \mathcal{K}^0$ and any $\beta \geq 1$:*

$$\Psi(\beta v) \leq r\psi\Big(\beta\varrho\Big(\frac{\Psi(v)}{r}\Big)\Big).$$

Proof. First we consider the case where $\beta > 1$. We consider the following maximization problem:

$$\max_v \{\Psi(\beta v) : \Psi(v) = z\},$$

where z is any nonnegative number. Let $v = \sum_{i=1}^r \lambda_i(v) c_i$ be the spectral decomposition of v. By Theorem 3.6.2 the first order optimality conditions for this maximization problem are

$$\sum_{i=1}^r \beta\psi'(\beta\lambda_i(v))c_i = \nu \sum_{i=1}^r \psi'(\lambda_i(v))c_i,$$

where ν denotes a Lagrange multiplier. This implies that

$$\beta\psi'(\beta\lambda_i(v)) = \nu\psi'(\lambda_i(v)), \qquad i = 1, \ldots, r. \tag{5.17}$$

The remaining of the proof is the same as in the proof of Theorem 3.2 in [9], but we include it here for the sake of completeness. Since $\psi'(1) = 0$ and $\beta\psi'(\beta) > 0$, we must have $\lambda_i(v) \neq 1$ for all i. We even may assume that $\lambda_i(v) > 1$ for all i. To see this, let z_i be such that $\psi(\lambda_i(v)) = z_i$. Given z_i, this equation has two solutions: $\lambda_i(v) = a_i^{(1)} < 1$ and $\lambda_i(v) = a_i^{(2)} > 1$. As a consequence of Lemma 4.2.3 we have $\psi(\beta v_i^{(1)}) \leq \psi(\beta a_i^{(2)})$. Since we are maximizing $\Psi(\beta v)$, it follows that we may assume $\lambda_i(v) = a_i^{(2)} > 1$. Thus we have shown that without loss of generality we may assume that $\lambda_i(v) > 1$ for all i. Note that then (5.17) implies $\beta\psi'(\beta\lambda_i(v)) > 0$ and $\psi'(\lambda_i(v)) > 0$, whence also $\nu > 0$. Now defining $g(t)$ as

$$g(t) := \frac{\psi'(t)}{\psi'(\beta t)}, \qquad t \geq 1,$$

We deduce from (5.17) that $g(\lambda_i(v)) = \frac{\beta}{\nu}$ for all i. However, by (4.4) we have $g'(t) > 0$ for $t > 1$. So $g(t)$ is strictly monotonically increasing. Hence it follows that all $\lambda_i(v)'s$ are mutually equal. Putting $\lambda_i(v) = t > 1$, for all i, we deduce from $\Psi(v) = z$ that $r\psi(t) = z$. This implies $t = \varrho\big(\frac{z}{r}\big)$. Hence the maximal value that $\Psi(v)$ can attain is given by

$$\Psi(\beta te) = r\psi(\beta t) = r\psi\Big(\beta\varrho\Big(\frac{z}{r}\Big)\Big) = r\psi\Big(\beta\varrho\Big(\frac{\Psi(v)}{r}\Big)\Big).$$

This proves the theorem if $\beta > 1$. For the case $\beta = 1$ it suffices to observe that both sides of the inequality in the theorem are continuous in β. ■

Remark 5.9.2. The bound of Theorem 5.9.1 is sharp. One may easily verify that if $v = \beta e$, with $\beta \geq 1$, then the bound holds with equality. □

As a result we have if $\Psi(v) \leq \tau$ and $\beta = \frac{1}{\sqrt{1-\theta}}$ then

$$L_\Psi(r, \theta, \tau) := r\psi\left(\frac{\varrho(\frac{\tau}{r})}{\sqrt{1-\theta}}\right) \tag{5.18}$$

is an upper bound for $\Psi(\beta v)$, the value of $\Psi(v)$ after the μ-update. This upper bound depends on the parameters θ, τ, on the rank of the Euclidean Jordan algebra and on the kernel function. Note that there is no relevant difference between the upper bound obtained in [9] for LO and for symmetric optimization.

5.9.2 Decrease of the barrier function during a inner iteration

We are going to compute a default value for the step size α in order to yield a new triple (x_+, y_+, s_+) as defined in (5.15). This results in a sufficient decrease of the barrier function during an inner iteration.

The analysis is sometimes different from the linear case and sometimes quite similar. If the analysis is different from the linear case we go into details, otherwise we avoid it. We use the quadratic representation and the similarity properties (as discussed in Section 3.2). Theorem 4.3.2 plays here its crucial role.

After a damped step we have

$$\begin{aligned} x_+ &:= x + \alpha\Delta x \\ &= x + \sqrt{\mu}\alpha P(w)^{\frac{1}{2}} d_x \\ &= P(w)^{\frac{1}{2}}(P(w)^{-\frac{1}{2}}x + \sqrt{\mu}\alpha d_x) \\ &= P(w)^{\frac{1}{2}}(\sqrt{\mu}v + \sqrt{\mu}\alpha d_x), \end{aligned}$$

and we get

$$x_+ = \sqrt{\mu}P(w)^{\frac{1}{2}}(v + \alpha d_x). \tag{5.19}$$

Analogously,

$$\begin{aligned} s_+ &:= s + \alpha \Delta s \\ &= P(w)^{-\frac{1}{2}}(P(w)^{\frac{1}{2}}s + \sqrt{\mu}\alpha d_s) \\ &= P(w)^{-\frac{1}{2}}(\sqrt{\mu}v + \sqrt{\mu}\alpha d_s), \end{aligned}$$

and we can write

$$s_+ = \sqrt{\mu}P(w)^{-\frac{1}{2}}(v + \alpha d_s). \tag{5.20}$$

Hence, defining v_+ as

$$v_+ := \frac{1}{\sqrt{\mu}}P(w_+)^{-\frac{1}{2}}x_+ = \frac{1}{\sqrt{\mu}}P(w_+)^{\frac{1}{2}}s_+$$

we have

$$v_+ = P(w_+)^{\frac{1}{2}}P(w)^{-\frac{1}{2}}(v + \alpha d_s) = P(w_+)^{-\frac{1}{2}}P(w)^{\frac{1}{2}}(v + \alpha d_x),$$

where, according to Proposition 5.4.2,

$$w_+ := P(x_+)^{\frac{1}{2}}((P(x_+)^{\frac{1}{2}}s_+)^{-\frac{1}{2}}).$$

The following step is key element in our analysis. We replace v_+ by an element which is similar to the new iterate v_+, thus simplifying the analysis.

Proposition 5.9.3. *One has*

$$v_+ \sim \left(P(v + \alpha d_x)^{\frac{1}{2}}(v + \alpha d_s)\right)^{\frac{1}{2}}.$$

Proof. By Proposition 3.2.4 we have

$$\sqrt{\mu}v_+ = P(w_+)^{\frac{1}{2}}s_+ \sim \left(P(x_+)^{\frac{1}{2}}s_+\right)^{\frac{1}{2}}.$$

Using (5.19) and (5.20), we get

$$\left(P(x_+)^{\frac{1}{2}}s_+\right)^{\frac{1}{2}} = \left(\mu P\left(P(w)^{\frac{1}{2}}(v + \alpha d_x)\right)^{\frac{1}{2}} P(w)^{-\frac{1}{2}}(v + \alpha d_s)\right)^{\frac{1}{2}}. \tag{5.21}$$

Since by Proposition 3.2.3 - *(ii)* (with $z = w^{\frac{1}{2}}$) the second term of (5.21) is similar to

$$\left(\mu P(v + \alpha d_x)^{\frac{1}{2}}(v + \alpha d_s)\right)^{\frac{1}{2}},$$

the result follows. Note that we used $P(w^{1/2}) = P(w)^{1/2}$ (Proposition 2.5.11). ■

In the case of LO, the two elements of Proposition (5.9.3) are equal. However, for our purpose the similarity property is enough, because it implies that

$$\Psi(v_+) = \Psi\left(\left(P(v+\alpha d_x)^{\frac{1}{2}}(v+\alpha d_s)\right)^{\frac{1}{2}}\right).$$

Therefore, by Theorem 4.3.2,

$$\Psi(v_+) \leq \frac{1}{2}(\Psi(v+\alpha d_x)+\Psi(v+\alpha d_s)).$$

Defining

$$f(\alpha) := \Psi(v_+) - \Psi(v),$$

we thus have $f(\alpha) \leq f_1(\alpha)$, where

$$f_1(\alpha) := \frac{1}{2}(\Psi(v+\alpha d_x)+\Psi(v+\alpha d_s)) - \Psi(v).$$

Note that $f_1(\alpha)$ gives an upper bound for the decrease of the barrier function. Working with f_1 instead of f has two advantages: f_1 is convex, whereas f is in general not convex and the derivatives of f_1, are easier to compute than those of f.

Obviously, $f(0) = f_1(0) = 0$. Using (4.9), the derivative of $f_1(\alpha)$ with respect to α is given by

$$f_1'(\alpha) = \frac{1}{2}(\mathrm{tr}\,(\Psi'(v+\alpha d_x)\circ d_x) + \mathrm{tr}\,(\Psi'(v+\alpha d_s)\circ d_s)\,.$$

This gives

$$\begin{aligned} f_1'(0) &= \frac{1}{2}\mathrm{tr}\,(\Psi'(v)\circ(d_x+d_s)) \\ &= -\frac{1}{2}\mathrm{tr}\,(\nabla\Psi(v)\circ\nabla\Psi(v)) \\ &= -\frac{1}{2}\|\nabla\Psi(v)\|^2 = -2\delta^2(v). \end{aligned}$$

Following Theorem 2.4.2 we can write $v+\alpha d_x = \sum_{i=1}^r \lambda_i(v+\alpha d_x)c_i$ and $v+\alpha d_s = \sum_{i=1}^r \lambda_i(v+\alpha d_s)b_i$. Let $d_x = \sum_{i=1}^r d_{xi}c_i + \sum_{i<j} d_{xij}$ be the Peirce decomposition of d_x with respect to the Jordan frame $c_1,\ldots,c_r$ and $d_s = \sum_{i=1}^r d_{si}f_i + \sum_{i<j} d_{sij}$ be the Peirce decomposition of d_s with respect to the Jordan frame $b_1,\ldots,b_r$. For simplicity of notation, below we use $\eta_i := \lambda_i(v+\alpha d_x)$ and $\gamma_i = \lambda_i(v+\alpha d_s)$, for $i = 1,\ldots,r$. Thus, if we differentiate f_1' with respect to α, using (4.10), we obtain

$$f_1''(\alpha) = g_1(\alpha) + g_2(\alpha), \tag{5.22}$$

where

$$g_1(\alpha) = \sum_{i=1}^r \psi''(\eta_i){d_{xi}}^2 + \sum_{\substack{i<j\\ \eta_i=\eta_j}} \psi''(\eta_i)\mathrm{tr}\left({d_{xij}}^2\right) + \sum_{\substack{i<j\\ \eta_i\neq\eta_j}} \frac{\psi'(\eta_i)-\psi'(\eta_j)}{\eta_i-\eta_j}\mathrm{tr}\left({d_{xij}}^2\right)$$

and

$$g_2(\alpha) = \sum_{i=1}^{r} \psi''(\gamma_i) d_{s\,i}^{\,2} + \sum_{\substack{i<j \\ \gamma_i=\gamma_j}} \psi''(\gamma_i) \mathrm{tr}\left(d_{s\,ij}^{\,2}\right) + \sum_{\substack{i<j \\ \gamma_i\neq\gamma_j}} \frac{\psi'(\gamma_i)-\psi'(\gamma_j)}{\gamma_i-\gamma_j} \mathrm{tr}\left(d_{s\,ij}^{\,2}\right).$$

Proposition 5.9.4. *We have*

$$\begin{aligned} f_1''(\alpha) \leq\ & \tfrac{1}{2}\left(\sum_{i=1}^{r} \psi''(\eta_i) d_{x\,i}^{\,2} + \sum_{i<j} \psi''(\eta_j) \mathrm{tr}\left(d_{x\,ij}^{\,2}\right)\right) \\ & + \tfrac{1}{2}\left(\sum_{i=1}^{r} \psi''(\gamma_i) d_{s\,i}^{\,2} + \sum_{i<j} \psi''(\gamma_j) \mathrm{tr}\left(d_{s\,ij}^{\,2}\right)\right). \end{aligned} \tag{5.23}$$

Proof. Using (5.22) and the above expressions for $g_1(\alpha)$ and $g_2(\alpha)$, this proposition follows from Proposition 4.4.1. ∎

In the following we want find the step size α that minimizes $f_1(\alpha)$. We already know that $f_1'(0) = -2\delta(v)^2 < 0$, which means that $f_1(\alpha)$ is monotonically decreasing in a neighborhood of $\alpha = 0$.

Using Proposition 5.9.4 we will get simpler upper bound for $f_1''(\alpha)$ than the one provided previously. Below we write δ instead of $\delta(v)$, with $\delta(v)$ as defined in (5.16).

Lemma 5.9.5. *One has* $f_1''(\alpha) \leq 2\delta^2 \psi''(\lambda_{\min}(v) - 2\alpha\delta)$.

Proof. Since d_x and d_s are orthogonal, (5.16) implies that

$$4\delta^2 = \|d_x + d_s\|^2 = \|d_x\|^2 + \|d_s\|^2.$$

Hence we have $\|d_x\| \leq 2\delta$ and $\|d_s\| \leq 2\delta$. Therefore, by Theorem 3.3.6

$$\eta_i = \lambda_i(v + \alpha d_x) \geq \lambda_{\min}(v + \alpha d_x) \geq \lambda_{\min}(v) - \|\alpha d_x\| \geq \lambda_{\min}(v) - 2\alpha\delta.$$

In the same way one can prove that

$$\gamma_i = \lambda_i(v + \alpha d_s) \geq \lambda_{\min}(v) - 2\alpha\delta.$$

Using the Peirce decomposition of d_x with respect to $c_1 \ldots, c_r$, we may write

$$\|d_x\|^2 = \langle d_x, d_x \rangle = \sum_{i=1}^{r} d_{x\,i}^{\,2} + \sum_{i<j} \mathrm{tr}\left(d_{x\,ij}^{\,2}\right).$$

Since $\psi''(t)$ is monotonically decreasing and because of (5.23) it follows that

$$\begin{aligned} f_1''(\alpha) &\leq \frac{1}{2}\psi''(\lambda_{\min}(v) - 2\alpha\delta)\left(\sum_{i=1}^{r}\left(d_{x\,i}^{\,2} + d_{s\,i}^{\,2}\right) + \sum_{i<j}\left(\mathrm{tr}\left(d_{s\,ij}^{\,2}\right) + \mathrm{tr}\left(d_{x\,ij}^{\,2}\right)\right)\right) \\ &= \frac{1}{2}\psi''(\lambda_{\min}(v) - 2\alpha\delta)(\|d_x\|^2 + \|d_s\|^2) \\ &= 2\delta^2\psi''(\lambda_{\min}(v) - 2\alpha\delta). \end{aligned}$$

This proves the lemma. ∎

Integrating twice the inequality in Lemma 5.9.5 with respect to α, we obtain an upper bound for $f_1(\alpha)$. In fact Lemma 5.9.5 is the same as Lemma 4.1 in [9]. As a consequence, from here on the analysis is similar to the case of LO. By integrating the inequality in Lemma 5.9.5 we get the next result.

Lemma 5.9.6. *$f_1'(\alpha) \leq 0$ holds certainly if α satisfies the inequality*

$$-\psi'(\lambda_{\min}(v) - 2\alpha\delta) + \psi'(\lambda_{\min}(v)) \leq 2\delta. \tag{5.24}$$

Lemma 5.9.6 means that for α satisfying the inequality (5.24), we have that $f_1(\alpha)$ is monotonically decreasing. We will obtain the largest step size that satisfies this inequality.

The proof of next lemma depends on the condition (4.2). We can see that the lemma is valid for symmetric optimization by just applying Lemma 4.3 in [9] to the vector $\lambda(v) \in \mathbb{R}^r$.

Lemma 5.9.7 (Lemma 4.3 in [9])**.** *Let $\rho : [0, +\infty) \mapsto (0, 1]$ denote the inverse function of the restriction of $-\frac{1}{2}\psi'(t)$ to the interval $(0, 1]$. Then the largest step size of α that satisfies (5.24) is given by*

$$\bar{\alpha} := \frac{1}{2\delta}(\rho(\delta) - \rho(2\delta)). \tag{5.25}$$

The next lemma follows from Lemma 5.9.7. We deduce a lower bound for $\bar{\alpha}$ which is easier to work with than $\bar{\alpha}$ itself.

Lemma 5.9.8 (Lemma 4.4 in [9])**.** *Let ρ and $\bar{\alpha}$ be as defined in Lemma 5.9.7. Then*

$$\bar{\alpha} \geq \frac{1}{\psi''(\rho(2\delta))}.$$

Note that an upper bound for $f_1(\alpha)$ is also an upper bound for $f(\alpha)$. Analogously to LO we can obtain an upper bound for $f_1(\alpha)$ and simplifying it using the technical Lemma D.1.1. This upper bound is given below.

Lemma 5.9.9 (Lemma 4.5 in [9])**.** *If the step size α is such that $\alpha \leq \bar{\alpha}$ then*

$$f(\alpha) \leq -\alpha\delta^2.$$

In the following we define

$$\tilde{\alpha} := \frac{1}{\psi''(\rho(2\delta))} \tag{5.26}$$

and we will use $\tilde{\alpha}$ as the default step size. By Lemma 5.9.8 we have $\bar{\alpha} \geq \tilde{\alpha}$. It turns out that the default step size is exactly the same as for the LO case.

The next theorem is an immediate consequence of Lemmas 5.9.8 and 5.9.9. It provides an estimate of the decrease of the barrier function when taking a damped step with size $\tilde{\alpha}$.

Theorem 5.9.10 (Theorem 4.6 in [9])**.** *With $\tilde{\alpha}$ being the default step size, as given by (5.26), one has*

$$f(\tilde{\alpha}) \leq -\frac{\delta^2}{\psi''(\rho(2\delta))}. \tag{5.27}$$

The following lemma depends on Condition (4.3). For its proof we refer to [9].

Lemma 5.9.11 (Lemma 4.7 in [9])**.** *The right hand side expression in (5.27) is monotonically decreasing in δ.*

We want to express the decrease of the barrier function during a inner iteration as a function of $\Psi(v)$. To this end we need a lower bound on $\delta(v)$ in terms of $\Psi(v)$. Such a bound is provided by the following theorem. The statement is exactly as in [9], but its proof is slightly different.

Theorem 5.9.12. *One has*

$$\delta(v) \geq \frac{1}{2}\psi'(\rho(\Psi(v))).$$

Proof. The statement in the lemma is obvious if $v = e$ since then $\delta(v) = \Psi(v) = 0$. Otherwise we have $\delta(v) > 0$ and $\Psi(v) > 0$. Let

$$v = \sum_{i=1}^{r} \lambda_i(v) c_i$$

be a spectral decomposition of v. To deal with nontrivial case we consider, for $\omega \geq 0$, the problem

$$z_\omega = \min_v \left\{ \delta(v)^2 = \frac{1}{4}\sum_{i=1}^{r} \psi'(\lambda_i(v))^2 : \Psi(v) = \omega \right\}. \tag{5.28}$$

Therefore the first order optimality conditions are

$$\frac{1}{2}\sum_{i=1}^{r} \psi'(\lambda_i(v))\psi''(\lambda_i(v))c_i = \sum_{i=1}^{r} \gamma\psi'(\lambda_i(v))c_i,$$

where $\gamma \in \mathbb{R}$. This implies that

$$\frac{1}{2}\psi'(\lambda_i(v))\psi''(\lambda_i(v)) = \gamma\psi'(\lambda_i(v)), \qquad i = 1, \ldots, r.$$

From this we conclude that we have either $\psi'(\lambda_i(v)) = 0$ or $\psi''(\lambda_i(v)) = 2\gamma$ for each i. Since $\psi''(t)$ is monotonically decreasing, this implies that all $\lambda_i(v)'s$ for which $\psi''(\lambda_i(v)) = 2\gamma$ have the same value. Denoting this value as t, and observing that all other eigenvalues have value 1 (since $\psi'(\lambda_i(v)) = 0$ for these eigenvalues), we conclude that,

$$v = tc_1 + \cdots + tc_k + c_{k+1} + \cdots + c_r,$$

where we supposed there are k eigenvalues with value t. Now $\Psi(v) = \omega$ implies $k\psi(t) = \omega$. From here on we proceed as in the proof of Theorem 4.9 in [9]. Given k, this uniquely determines $\psi(t)$, whence we have

$$4\delta(v)^2 = k(\psi'(t))^2, \qquad \psi(t) = \frac{\omega}{k}.$$

Note that the equation $\psi(t) = \frac{\omega}{k}$ has two solutions, one smaller than 1 and one larger than 1. By Lemma 4.2.4 the larger value gives the smallest value of $(\psi'(t))^2$. Since we are minimizing $\delta(v)^2$, we conclude that $t > 1$ (since $\omega > 0$). Hence we may write

$$t = \varrho\left(\frac{\omega}{k}\right),$$

where, as before, ϱ denotes the inverse function of $\psi(t)$ for $t \geq 1$. Thus we obtain that

$$4\delta(v)^2 = k(\psi'(t))^2, \quad t = \varrho\left(\frac{\omega}{k}\right). \tag{5.29}$$

The question now is which value of k minimizes $\delta(v)^2$. To investigate this, we take the derivative with respect to k of (5.29) extended to $k \in \mathbb{R}$. This gives

$$\frac{d4\delta(v)^2}{dk} = \psi'(t)^2 + 2k\psi'(t)\psi''(t)\frac{dt}{dk}. \tag{5.30}$$

From $\psi(t) = \frac{\omega}{k}$ we derive that

$$\psi'(t)\frac{dt}{dk} = -\frac{\omega}{k^2} = -\frac{\psi(t)}{k},$$

which gives

$$\frac{dt}{dk} = -\frac{\psi(t)}{k\psi'(t)}.$$

Substitution into (5.30) gives

$$\frac{d4\delta(v)^2}{dk} = \psi'(t)^2 - 2\psi(t)\psi''(t).$$

Defining $f(t) = \psi'(t)^2 - 2\psi(t)\psi''(t)$, we have $f(1) = 0$ and

$$f'(t) = 2\psi'(t)\psi''(t) - 2\psi'(t)\psi''(t) - 2\psi(t)\psi'''(t) = -2\psi(t)\psi'''(t) > 0.$$

We conclude that $f(t) > 0$ for $t > 1$. Hence $\frac{d\delta(v)^2}{dk} > 0$, so $\delta(v)^2$ increases when k increases. Since we are minimizing $\delta(v)^2$, at optimality we have $k = 1$. Also using that $\psi(t) \geq 0$, we obtain from (5.29) that

$$\min_v\{\delta(v) : \Psi(v) = \omega\} = \frac{1}{2}\psi'(t) = \frac{1}{2}\psi'(\varrho(\omega)) = \frac{1}{2}\psi'(\varrho(\Psi(v))).$$

This completes the proof of the theorem. ∎

Remark 5.9.13. The bound of Theorem 5.9.12 is sharp. One may easily verify that if v is such that all eigenvalues are equal to 1 except one eigenvalue which is greater than or equal to 1, then the bound holds with equality. □

Combining the results of Theorems 5.9.10 and 5.9.12 we obtain

$$f(\tilde{\alpha}) \leq -\frac{(\psi'(\varrho(\Psi(v))))^2}{4\psi''(\rho(\psi'(\varrho(\Psi(v)))))}. \tag{5.31}$$

This expresses the decrease of $\Psi(v)$ during an inner iteration completely in $\Psi(v)$, the first and second derivatives of ψ, and the inverse functions ρ and ϱ, of $-\frac{1}{2}\psi'$ restricted to $(0,1]$ and ψ restricted to $[1,+\infty)$, respectively.

5.9.3 Iteration bounds

After the update of μ to $(1-\theta)\mu$, we have, by Theorem 5.9.1 and (5.18),

$$\Psi(v) \leq L_\Psi(r,\theta,\tau) = r\psi\left(\frac{\varrho(\frac{\tau}{r})}{\sqrt{1-\theta}}\right) \tag{5.32}$$

We need to count how many inner iterations are required to return to the situation where $\psi(v) \leq \tau$. We denote the value $\Psi(v)$ after the μ-update as Ψ_0, and the subsequent values are denoted as Ψ_k, $k=1,2,\ldots$. The decrease during each inner iteration is given by (5.31). In the sequel we assume that the expression in the right-hand side expression of (5.31) satisfies

$$\frac{(\psi'(\varrho(\Psi(v))))^2}{4\psi''(\rho(\psi'(\varrho(\Psi(v)))))} \geq \kappa\Psi(v)^{1-\gamma} \tag{5.33}$$

for some positive constants κ and γ, with $\gamma \in (0,1]$.

Let us establish that such constants γ and κ do exist. At the start of each inner iteration we have $\Psi(v) > \tau > 0$. Since

$$\varrho : [0,+\infty) \mapsto [1,+\infty)$$

is increasing (as the inverse of an increasing function), $\varrho(\Psi(v)) > \varrho(\tau) > 1$. From here, we obtain $\psi'(\varrho(\Psi(v))) > \psi'(\varrho(\tau)) > 0$. Let $z := \psi'(\varrho(\tau))$. It follows, by definition of ρ that $\rho(z) > 0$. Using that $\psi''(t) > 0$, we conclude that the left-hand side expression of (5.33) is positive. By Lemma 5.9.11

$$\frac{z^2}{\psi''(\rho(z))} \tag{5.34}$$

is monotonically increasing, which implies that the left-hand side in (5.33) is greater than (5.34). Hence (5.33) certainly holds if $\gamma = 1$ and

$$\kappa = \frac{z^2}{4\psi''(\rho(z))}.$$

For each kernel function we should find constants γ and κ satisfying (5.33). This process is not straightforward and may vary, depending on the kernel function. For different cases and appropriate choices of γ and κ, see [9]. The next lemma makes clear that we want $\gamma \in (0, 1]$ as small as possible.

Lemma 5.9.14 (Lemma 5.1 in [9]). *If K denotes the number of inner iterations, we have*

$$K \leq \frac{\Psi_0^\gamma}{\kappa\gamma}. \tag{5.35}$$

The last lemma provides an estimate for the number of inner iterations in terms of Ψ_0 and the constants κ and γ. Recall that Ψ_0 is bounded above according to (5.32).

An upper bound for the total number of iterations is obtained by multiplying the upper bound in (5.35) for the number K by the number of barrier parameter updates, which is bounded above by (see Lemma II.17 in [45])

$$\frac{1}{\theta} \log \frac{r}{\epsilon}.$$

So the total number of iterations is bounded above by

$$\frac{\Psi_0^\gamma}{\kappa\gamma\theta} \log \frac{r}{\epsilon}.$$

5.10 Recipe to calculate a complexity bound

In this section we summarize the results of the previous sections by presenting a simple scheme to obtain iteration bounds for both large- and small-update methods. This recipe was introduced in [9] for LO. The amazing fact of this recipe, is that, for a given ψ, τ, θ and ϵ it returns an iteration bound for the IPM based on it.

We first state some notational conventions, that we use in this section. Let $f, g : \mathbb{R}_+ \mapsto \mathbb{R}_+$. We write $f(t) = O(g(t)$ if $f(t) \leq \nu g(t)$ for some positive constant ν and $f(t) = \Theta(g(t))$ if $\nu_1 g(t) \leq f(t) \leq \nu_2 g(t)$ for positive constants ν_1 and ν_2.

Step 1 Input a kernel function ψ; an update parameter θ, $0 < \theta < 1$; a threshold parameter τ; and an accuracy parameter ϵ.

Step 2 Solve the equation $-\frac{1}{2}\psi'(t) = s$ to get $\rho(s)$, the inverse function of $-\frac{1}{2}\psi'(t)$, $t \in (0, 1]$. If the equation is hard to solve, derive a lower bound for $\rho(s)$.

Step 3 Calculate the decrease of $\Psi(v)$ in terms of δ for the default step size $\tilde{\alpha}$ from

$$f(\tilde{\alpha}) \leq -\frac{\delta^2}{\psi''(\rho(2\delta))}.$$

Step 4 Solve the equation $\psi(t) = s$ to get $\varrho(s)$, the inverse function of $\psi(t)$, $t \geq 1$. If the equation is hard to solve, derive a lower bound and an upper bound for $\varrho(s)$.

Step 5 Derive a lower bound for $\delta(v)$ in terms of $\Psi(v)$ by using

$$\delta(v) \geq \frac{1}{2}\psi'(\varrho(\Psi(v))).$$

Step 6 Using the results of Step 3 and Step 4 find a valid inequality of the form

$$f(\tilde{\alpha}) \leq -\kappa\Psi(v)^{1-\gamma}$$

for some positive constants κ and γ, with $\gamma \in (0,1]$ and γ as small as possible.

Step 7 Calculate an upper bound for Ψ_0 from

$$\Psi_0 \leq L_\Psi(r,\theta,\tau) = r\psi\Big(\frac{\varrho(\frac{\tau}{r})}{\sqrt{1-\theta}}\Big).$$

Step 8 Derive an upper bound for the total number of iterations using that this number is bounded above by

$$\frac{\Psi_0^\gamma}{\theta\kappa\gamma}\log\frac{r}{\epsilon}.$$

Step 9 Set $\tau = O(r)$ and $\theta = \Theta(1)$ so as to calculate complexity bound for large-update methods, and set $\tau = O(1)$ and $\theta = \Theta(\frac{1}{\sqrt{r}})$ to obtain a complexity bound for small-update methods.

5.11 Examples

In Table 5.1 we present the kernel functions that up to now have been proposed and analyzed in the literature.

The table also shows the current iterations bounds for the corresponding algorithm. Most of these examples have been studied for LO. However, our work allows us to conclude that the iteration complexity associated to each function also applies to symmetric optimization. The best iteration complexity for large-update method is obtained for

$$\psi(t) = \frac{t^2-1}{2} + \frac{t^{1-q}-1}{q-1},$$

with $q = \log r$. In this case the iteration complexity is $O(\sqrt{r}\log r\log\frac{r}{\epsilon})$.

	kernel function	iteration complexity	references
1	$\frac{t^2-1}{2} - \log t$	$O(r \log \frac{r}{\epsilon})$	[3, 13, 50]
2	$\frac{t^2-1}{2} + \frac{t^{1-q}-1}{q(q-1)} - \frac{q-1}{q}(t-1), \quad q > 1$	$O(qr^{\frac{q+1}{2q}} \log \frac{r}{\epsilon})$	[42, 43]
3	$\frac{t^2-1}{2} + \frac{(e-1)^2}{e}\frac{1}{e^t-1} - \frac{e-1}{e}$	$O(r^{\frac{3}{4}} \log \frac{r}{\epsilon})$	[8]
4	$\frac{1}{2}(t - \frac{1}{t})^2$	$O(r^{\frac{2}{3}} \log \frac{r}{\epsilon})$	[41]
5	$\frac{t^2-1}{2} + e^{\frac{1}{t}-1} - 1$	$O(\sqrt{r} \log^2 r \log \frac{r}{\epsilon})$	[9]
6	$\frac{t^2-1}{2} - \int_1^t e^{\frac{1}{\xi}-1} d\xi$	$O(\sqrt{r} \log^2 r \log \frac{r}{\epsilon})$	[9]
7	$\frac{t^2-1}{2} + \frac{t^{1-q}-1}{q-1}, \quad q > 1$	$O(qr^{\frac{q+1}{2q}} \log \frac{r}{\epsilon})$	[40]
8	$\frac{t^{1+p}-1}{1+p} - \log t, \quad p \in [0,1]$	$O(r \log \frac{r}{\epsilon})$	[20]
9	$\frac{t^{1+p}-1}{1+p} + \frac{t^{1-q}-1}{q-1}, \quad q > 1, \quad p \in [0,1]$	$O(qr^{\frac{p+q}{q(1+p)}} \log \frac{r}{\epsilon})$	[20]
10	$t + \frac{1}{t} - 2$	$O(r \log \frac{r}{\epsilon})$	[6]

Table 5.1: Kernel functions

5.12 The accuracy of solution produced by the algorithm

As we mention in Section 5.8 when the algorithm stops, it returns an solution that is τ-closed (in the barrier function sense) to a point $(x(\mu), y(\mu), s(\mu))$ such that $\langle x(\mu), s(\mu)\rangle = \mu r \leq \epsilon$. But this does not directly imply that we got a solution (x, y, s) such that $\langle x, s\rangle \leq \epsilon$. As much as we know, this question is for the first time analyzed here.

In the following we deduce an upper bound for the duality gap. We can assume without loss of generality that, for $v \in \mathcal{K}$, $r_1 < r$ eigenvalues are bigger than one and the remaining eigenvalues are less than one. This means that $\lambda_1 = \lambda_{\max} > 1$. Using Lemma 4.2.5 we obtain

$$\Psi(v) \geq \sum_{i=1}^{r_1} \frac{1}{2}\psi''(1)(\lambda_i(v) - 1)^2 + \sum_{i=r_1+1}^{r} \frac{1}{2}\psi''(\lambda_i(v))(\lambda_i(v) - 1)^2.$$

Since ψ'' is monotonically decreasing,

$$\Psi(v) \geq \sum_{i=1}^{r} \frac{1}{2}\psi''(\lambda_1(v))(\lambda_i(v) - 1)^2.$$

Let $\nu := \psi''(\lambda_1(v))$. Hence

$$\Psi(v) \geq \frac{1}{2}\nu\|v - e\|^2,$$

which is equivalent to

$$\frac{\Psi(v)}{\nu} \geq \frac{1}{2}\|v\|^2 - \mathrm{tr}\,(v) + \frac{r}{2}.$$

Using the fact that $\mathrm{tr}\,(v) \leq \sqrt{r}\|v\|$ we obtain

$$\frac{\Psi(v)}{\nu} \geq \frac{1}{2}\|v\|^2 - \sqrt{r}\|v\| + \frac{r}{2}.$$

This implies that

$$\|v\| \leq \sqrt{r} + \sqrt{\frac{2\Psi(v)}{\nu}}.$$

Since the duality gap can be written in terms of v, i.e.,

$$\langle x, s\rangle = \mu\|v\|^2,$$

and for the solution produced by the algorithm $\Psi(v) < \tau$, then we have

$$\langle x, s\rangle = \mu\|v\|^2 \leq \mu r + \frac{2\mu\tau}{\nu} + 2\mu\sqrt{\frac{2r\tau}{\nu}}.$$

Note that for large update methods we do $\tau = O(n)$. If we assume that $\nu \geq 1$, then the algorithm finally reports a feasible solution such that $\langle x, s\rangle = O(\epsilon)$, using $\mu r \leq \epsilon$. For instance if $\tau = n$ then

$$\langle x, s\rangle \leq 7\epsilon.$$

To obtain this result we have assumed that $\nu > 1$. In fact, this is the case for the kernel functions presented in Table 5.1 numbered from 1 to 7. However, for the kernel functions numbers 8, 9 (with $p < 1$) and 10 we may not have a "good" lower bound for ν, i.e., a lower bound for ν such that we certainly have $\langle x, s\rangle = O(\epsilon)$. For instance, if $\psi(v) = t + \frac{1}{t} - 2$ (number 10), then $\psi''(t) = \frac{2}{t^3}$. It is obviously that if t increases then ν decreases and in this case (up to now) we could not guarantee the quality of the solution.

Note that if the solution produced by the algorithm is such that $\lambda_1(v) < 1$, then we certainly have a solution such that $\langle x, s\rangle = O(\epsilon)$, because we can make $\nu := \psi''(1)$.

5.13 Conclusions

The interior-point methods based on kernel functions for LO can be extended, and in many places, word by word, to symmetric optimization. Our main conclusion is that the recipe presented in Section 5.10 for symmetric optimization is exactly the same as for LO. As a consequence, we proved that computing the iteration complexity for symmetric optimization is as hard (or easy) as for LO. The recipe uses only the kernel function $\psi(t)$ as input.

Chapter 6

Conclusions

6.1 Final notes

The aim of this work was to generalize interior-point methods for LO based on kernel functions to symmetric optimization. To achieve this we needed to study Euclidean Jordan algebras and their relation with symmetric cones, as it is done in Chapter 2.

The barrier functions based on kernel functions are, in fact, separable spectral functions, whose fundamental properties were studied in Chapter 4.

Since spectral functions are functions depending on the eigenvalues of their argument we had to establish eigenvalues properties, and these were presented in Chapter 3. We further derived expressions for the derivatives of eigenvalues.

In Chapter 5 it was proved that interior-point methods based on kernel functions are generalizable to symmetric optimization and that getting complexity bounds for symmetric optimization is as hard (or easy) as for LO. In fact, it turns out that the same scheme for computing an iteration bound for LO can also be used for symmetric optimization.

The e-convexity property is shared by the so-called self-regular functions, as introduced in [43] and also by the class of eligible kernel functions, as introduced in [9]. In both cases, this property needs to be extended to the induced barrier function. As we could do this, we can conclude that the interior point methods based on the self-regular functions are also extendable to symmetric optimization.

6.2 Directions for further research

We list below some possible directions for further research:

- Does there exist a kernel function for which the complexity of large-update methods is the same as for small-update methods?

- Can we develop a different analysis of the Algorithm 5.8, in a such way that we can accept kernel functions, that do not satisfy Conditions (4.2)-(4.4)?
- The set of homogeneous cones is quite larger set than the set of symmetric cones. Can we extend the interior-point methods based on kernel functions to homogeneous cones?
- Theorem 3.3.2 was proved for each primitive symmetric cone separately and then for the direct sum of the primitives cones. Can this theorem be proved in an unified way for any symmetric cone?

Appendix A

Topological notions

When V is an n-dimensional vector space over $\mathbb{R}$, any vector in V has real coordinates with respect to a certain fixed basis $\{b_1, \dots, b_n\}$ of V. We can define a norm on V by

$$\|x\| := \sqrt{x_1^2 + \cdots + x_n^2}, \quad x \in V,$$

where $x_1, \dots, x_n$ are the coordinates of x with respect to the basis. This norm induces a metric $d : V \times V \mapsto \mathbb{R}_+$ defined as

$$d(x, y) := \|y - x\|, \quad x, y \in V.$$

Using this metric the ball, with radius ϵ, centered at $x \in V$, is given by

$$B_\epsilon(x) := \{y \in V : \|y - x\| < \epsilon\}.$$

Let U be a nonempty subset of V. A point $x \in U$ is said to be *interior* in U if there exists $\epsilon > 0$ such that $B_\epsilon(x) \subset U$. The set of all interior points of U is denoted U^0. We say that U is *open* if $U^0 = U$. We say that U is *closed* if $V \setminus U$ is open. The point $x \in V$ is called a *boundary* point of U if for each $\epsilon > 0$, $B_\epsilon(x)$ contains a point in U and a point not in U. The set of all boundary points is called the *boundary* of U and is denoted by ∂U.

Alternatively, we say that U is closed if it contains its boundary. The *closure* of U, denoted as $\mathrm{cl}(U)$, is the set $U \cup \partial U$. We say that U is *dense* in V if $\mathrm{cl}(U) = V$. It is well known that U is dense V if and only if every element $x \in V$ is the limit of a sequence of elements in U.

We say that U is *bounded* if there exists $L > 0$ such that

$$\|x\| \leq L, \quad \forall x \in U.$$

A set U is said to be *compact* if U is closed and bounded.

Appendix B

Some matrix properties

Let $\mathrm{Herm}(\mathbb{C}, n)$ be the set of complex $n \times n$ Hermitian matrices. Let $x^* := \bar{x}^T$, where $\bar{x}$ is the complex conjugate of x. We say that $A \in \mathrm{Herm}(\mathbb{C}, n)$ is *positive semidefinite (definite)* if $x^*Ax \geq 0$ for all $x, y \in \mathbb{C}^n$ ($x^*Ax > 0$ for all $x, y \in \mathbb{C}^n \setminus \{0\}$). Another characterization of positive semidefiniteness of the Hermitian matrix A can be given in terms of its eigenvalues: A is positive semidefinite (definite) if $\lambda_i(A) \geq 0$, ($\lambda_i(A) > 0$) $i = 1, \ldots, n$.

A matrix $A \in \mathrm{Herm}(\mathbb{C}, n)$ is said to be *similar* to a matrix $B \in \mathrm{Herm}(\mathbb{C}, n)$, if there exists a invertible matrix S, such that

$$A = S^{-1}BS.$$

We denote the similarity relation as $A \sim B$. We have the following properties:

Proposition B.1.1 (Section 5.2 in [33]). *Let $A, B \in \mathrm{Herm}(\mathbb{C}, n)$. One has*

(i) $AB \sim BA$,

(ii) if A, B are positive semidefinite then $AB \sim A^{1/2}BA^{1/2} \sim B^{1/2}AB^{1/2}$,

(iii) if B is invertible then $A \sim B^{-1}AB$.

Theorem B.1.2 (Theorem 9.H.1.a in [34]). *If A and B are $n \times n$ positive semidefinite Hermitian matrices, then*

$$\prod_{i=1}^{k} \lambda_i(AB) \leq \prod_{i=1}^{k} \lambda_i(A)\lambda_i(B), \qquad k = 1, \ldots, n-1,$$
$$\prod_{i=1}^{n} \lambda_i(AB) = \prod_{i=1}^{n} \lambda_i(A)\lambda_i(B).$$

Lemma B.1.3. *Let A, X, Y be positive definite symmetric matrices. If $XAX = YAY$ then $X = Y$.*

Proof. The equation $XAX = YAY$ implies

$$A^{1/2}XAXA^{1/2} = A^{1/2}YAYA^{1/2}.$$

This can be written as follows

$$(A^{1/2}XA^{1/2})(A^{1/2}XA^{1/2}) = (A^{1/2}YA^{1/2})(A^{1/2}YA^{1/2}),$$

or equivalently

$$(A^{1/2}XA^{1/2})^2 = (A^{1/2}YA^{1/2})^2.$$

Since X and Y are positive definite, the matrices $A^{1/2}XA^{1/2}$ and $A^{1/2}YA^{1/2}$ are also positive definite. Therefore,

$$A^{1/2}XA^{1/2} = A^{1/2}YA^{1/2},$$

implying that $X = Y$. ■

Appendix C

Matrices of quaternions and octonions

C.1 Quaternions

As usual, let $\mathbb{C}$ and $\mathbb{R}$ denote the field of the complex and real numbers, respectively. Let $\mathbb{H}$ be a four-dimensional vector space over $\mathbb{R}$ with an ordered basis, denoted by $1, i, j$ and k. A *real quaternion* or simply called *quaternion*, is a vector

$$x = x_0 + x_1 i + x_2 j + x_3 k \in \mathbb{H},$$

with real coefficients x_0, x_1, x_2, x_3. The elements of $\mathbb{H}$ are called (real) quaternions. The product of any two of quaternions is determined by the following product rules for the basis vectors:

$$i^2 = j^2 = k^2 = -1,$$

$$ij = -ji = k, \qquad jk = -kj = i, \qquad ki = -ik = j.$$

For any $x = x_0 + x_1 i + x_2 j + x_3 k \in \mathbb{H}$, we define $\Re x = x_0$, the *real part* of x; $\Im x = x_1 i + x_2 j + x_3 k$, the *imaginary part*; and $\bar{x} = x_0 - x_1 i - x_2 j - x_3 k$, the *conjugate* of x.

We may write

$$\mathbb{H} = \mathbb{R} + \mathbb{R}i + \mathbb{R}j + \mathbb{R}k = (\mathbb{R} + \mathbb{R}i) + (\mathbb{R} + \mathbb{R}i)j = \mathbb{C} + \mathbb{C}j. \tag{C.1}$$

The algebra $\mathbb{H}$, with product rules defined above, is associative and noncommutative.

C.2 Matrices of quaternions

Let U be a matrix with entries in $\mathbb{H}$. We call U a *quaternion matrix*. Following (C.1) we can represent $U = A + Bj$, with $A = A_1 + A_2 i$, $B = B_1 + B_2 i$ and $A_1, A_2, B_1, B_2 \in$

$\mathbb{R}^{n\times n}$. Let $\bar{U} := [\bar{u}_{ij}]$ be the conjugate matrix of U. We define

$$U^f := \begin{bmatrix} A & B \\ -\overline{B} & \overline{A} \end{bmatrix} \in \mathbb{C}^{2n\times 2n},$$

the complex matrix U^f is known as the *complex representation* of the matrix U.

The matrix $U^* := \bar{U}^T$ is called the *adjoint* of U. We say that U is *Hermitian* if $U^* = U$. Let $U \in \mathrm{Herm}(\mathbb{H}, n)$ an n-Hermitian matrix with quaternion entries.

Proposition C.2.1 (Theorem 3 in [30]). *Let $U, V \in \mathbb{H}^{n\times n}$. Then*

(i) U^f is Hermitian if and only if U is Hermitian;

(ii) $(UV)^f = U^f V^f$.

By Proposition C.2.1-(i) we can say that an Hermitian quaternion matrix is diagonalizable since any Hermitian complex matrix is diagonalizable.

Theorem C.2.2 (Theorem 3.1 in [26]). *Let $A \in \mathbb{H}^{n\times n}$. Then A is a diagonalizable quaternion matrix if and only if A^f is a diagonalizable complex matrix. Assume that A^f is diagonalizable. Let all eigenvalues of A^f be $\lambda_1, \overline{\lambda_1}, \lambda_2, \overline{\lambda_2}, \ldots, \lambda_n, \overline{\lambda_n}$, in which $\Im\lambda_i \geq 0$, $i = 1, \ldots, n$, and T be a nonsingular matrix such that*

$$T^{-1}A^f T = \begin{bmatrix} J & 0 \\ 0 & \overline{J} \end{bmatrix} = J^f,$$

where $J = \mathrm{Diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$. Let

$$S = \frac{1}{4}\begin{bmatrix} I_n & -jI_n \end{bmatrix}(T + Q_n^{-1}\overline{T}Q_n)\begin{bmatrix} I_n \\ jI_n \end{bmatrix},$$

where

$$Q_n = \begin{bmatrix} 0 & -I_n \\ I_n & 0 \end{bmatrix}.$$

Then S is a nonsingular quaternion matrix and $S^{-1}AS = J$.

We say that $A \in \mathrm{Herm}(\mathbb{H}, n)$ is a *positive semidefinite (definite)* matrix if all the eigenvalues are nonnegative (positive). Theorem 2.4.2 implies that the eigenvalues of any matrix $A \in \mathrm{Herm}(\mathbb{H}, n)$ are real numbers.

Proposition C.2.3. *If $A \in \mathrm{Herm}(\mathbb{H}, n)$ is positive semidefinite if and only if A^f is also positive semidefinite.*

Proof. Direct consequence of Theorem C.2.2 and of the fact that the eigenvalues are real numbers. ■

Proposition C.2.4. *Let $U := A^{1/2}BA^{1/2}$, with A and B $n \times n$ quaternion positive semidefinite matrices. Then U is similar to AB.*

Proof. By Theorem C.2.1,

$$U^f = (A^{1/2}BA^{1/2})^f = (A^{1/2})^f B^f (A^{1/2})^f.$$

Since A^f and B^f are matrices with complex entries, by Proposition B.1.1, U^f is similar to $A^f B^f = (AB)^f$. Thus, by Theorem C.2.2 AB is similar to U. ■

We extend the Theorem B.1.2 to set of quaternion positive semiefinite Hermitian matrices. As far as we know it was not done before.

Theorem C.2.5. *If A and B are $n \times n$ quaternion positive semidefinite Hermitian matrices and $U := A^{1/2}BA^{1/2}$ then*

$$\prod_{i=1}^{k} \lambda_i(U) \leq \prod_{i=1}^{k} \lambda_i(A)\lambda_i(B), \qquad k = 1, \ldots, n-1,$$
$$\prod_{i=1}^{n} \lambda_i(U) = \prod_{i=1}^{n} \lambda_i(A)\lambda_i(B).$$

Proof. We have, by Theorem C.2.2 that the eigenvalues of U^f are of the form

$$\lambda_1(U^f), \overline{\lambda_1(U^f)}, \lambda_2(U^f), \overline{\lambda_2(U^f)}, \ldots, \lambda_n(U^f), \overline{\lambda_n(U^f)}.$$

Analogously, the eigenvalues of A^f and B^f are

$$\lambda_1(A^f), \overline{\lambda_1(A^f)}, \lambda_2(A^f), \overline{\lambda_2(A^f)}, \ldots, \lambda_n(A^f), \overline{\lambda_n(A^f)},$$
$$\lambda_1(B^f), \overline{\lambda_1(B^f)}, \lambda_2(B^f), \overline{\lambda_2(B^f)}, \ldots, \lambda_n(B^f), \overline{\lambda_n(B^f)},$$

respectively. Since U^f is Hermitian, its eigenvalues are real numbers (Theorem 2.4.2). This implies that

$$\lambda_i(U^f) = \overline{\lambda_i(U^f)}, \quad \lambda_i(A^f) = \overline{\lambda_i(A^f)} \quad \text{and} \quad \lambda_i(B^f) = \overline{\lambda_i(B^f)},$$

for $i = 1, \ldots, n$. Therefore, Theorem C.2.2 implies that the eigenvalues of U, A, and B are

$$\lambda_1(U^f), \lambda_2(U^f), \ldots, \lambda_n(U^f),$$
$$\lambda_1(A^f), \lambda_2(A^f), \ldots, \lambda_n(A^f)$$

and

$$\lambda_1(B^f), \lambda_2(B^f), \ldots, \lambda_n(B^f),$$

respectively. Since, by Proposition B.1.1, $U^f = (A^{1/2})^f B^f (A^{1/2})^f$ is similar to $A^f B^f = (AB)^f$, applying Theorem B.1.2 to U^f, A^f and B^f we have

$$\prod_{i=1}^{k} \lambda_i^2(U^f) \leq \prod_{i=1}^{k} \lambda_i^2(A^f)\lambda_i^2(B^f).$$

From here we get that

$$\prod_{i=1}^{k} \lambda_i(U) \leq \prod_{i=1}^{k} \lambda_i(A)\lambda_i(B).$$

The equality follows analogously. ■

C.3 Octonions

The octonions, denoted as $\mathbb{O}$, are the eight-dimension vector space over the reals, spanned by the identity element 1 and seven imaginary units, which we label as $\{i, j, k, k\ell, j\ell, i\ell, \ell\}$. Each imaginary unit squares to -1,

$$i^2 = j^2 = k^2 = ... = \ell^2 = -1.$$

We omit here the remaining rules of multiplication between the imaginary units, because they are quite extensive and we do not need them. The *octonions* endowed with the mentioned multiplication rules are nonassociative and noncommutative.

C.4 The Albert Algebra

The *Albert algebra* denoted as $\mathrm{Herm}(\mathbb{O}, 3)$, is an algebra consisting of the 3×3 octonion Hermitian matrices. Recall the Theorem 2.7.5 which states that the Albert algebra is a Euclidean Jordan algebra, when defining for $A, B \in \mathrm{Herm}(\mathbb{O}, 3)$ the Jordan product,

$$A \circ B = \frac{AB + BA}{2}.$$

The Albert algebra it is well study by algebraist, and for $A \in \mathrm{Herm}(\mathbb{O}, 3)$ its characteristic polynomial is given by

$$\lambda^3 - (\mathrm{tr}\,(A))\lambda^2 + \sigma(A)A - (\det A) = 0,$$

with $\sigma(A) = \frac{1}{2}((\mathrm{tr}\,(A))^2 - \mathrm{tr}\left(A^2\right))$. For a deduction of the characteristic polynomial we refer to [14]. From everything that was exposed we can conclude that A has only three real eigenvalues.

Appendix D

Technical properties

Lemma D.1.1 (Lemma 3.12 in [42])**.** *Let $h(t)$ be a twice differentiable convex function with $h(0) = 0$, $h'(0) < 0$, and let $h(t)$ attain its (global) minimum at $t^* > 0$. If $h''(t)$ is increasing for $t \in [0, t^*]$, then*

$$h(t) \leq \frac{th'(0)}{2}, \quad 0 \leq t \leq t^*.$$

Lemma D.1.2 (Section 14 in [10])**.** *For $x_i \geq 0$ and $\alpha_i > 0$ such that $\alpha_1 + \cdots + \alpha_n = 1$, we have*

$$\sum_{i=1}^{n} \alpha_i x_i \geq \prod_{i=1}^{n} x_i^{\alpha_i}.$$

The last inequality is known as the *weighted arithmetic-geometric mean inequality.*

Bibliography

[1] F. Alizadeh. Interior point methods in semidefinite programming with applications to combinatorial optimization. *SIAM J. Optim.*, 5(1):13–51, 1995.

[2] F. Alizadeh and D. Goldfarb. Second-order cone programming. *Math. Program.*, 95(1, Ser. B):3–51, 2003.

[3] E. D. Andersen, J. Gondzio, C. Mészáros, and X. Xu. Implementation of interior-point methods for large scale linear programs. In *Interior point methods of mathematical programming*, volume 5 of *Appl. Optim.*, pages 189–252. Kluwer Acad. Publ., Dordrecht, 1996.

[4] M. Baes. *Spectral functions and smoothing techniques on Jordan algebras.* PhD thesis, Université Catholique de Louvain, 2006.

[5] M. Baes. Convexity and differentiability properties of spectral functions and spectral mappings on Euclidean Jordan algebras. *Linear Algebra Appl.*, 422: 664–700, 2007.

[6] Y. Q. Bai and C. Roos. A polynomial-time algorithm for linear optimization based on a new simple kernel function. *Optim. Methods Softw.*, 18(6):631–646, 2003.

[7] Y. Q. Bai, M. El Ghami, and C. Roos. A new efficient large-update primal-dual interior-point method based on a finite barrier. *SIAM J. Optim.*, 13(3):766–782, 2002.

[8] Y. Q. Bai, C. Roos, and M. El Ghami. A primal-dual interior-point method for linear optimization based on a new proximity function. *Optim. Methods Softw.*, 17(6):985–1008, 2002.

[9] Y. Q. Bai, M. El Ghami, and C. Roos. A comparative study of kernel functions for primal-dual interior-point algorithms in linear optimization. *SIAM J. Optim.*, 15(1):101–128, 2004.

[10] E. F. Beckenbach and R. Bellman. *Inequalities.* Springer-Verlag, Berlin, 1961.

[11] A. Ben-Tal and A. Nemirovski. *Lectures on modern convex optimization.* MPS/SIAM Series on Optimization. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2001.

[12] E. de Klerk. *Aspects of semidefinite programming*, volume 65 of *Applied Optimization.* Kluwer Academic Publishers, Dordrecht, 2002.

[13] D. den Hertog, C. Roos, and J.-Ph. Vial. A complexity reduction for the long-step path-following algorithm for linear programming. *SIAM J. Optim.*, 2(1): 71–87, 1992.

[14] T. Dray and C. A. Manogue. The exceptional Jordan eigenvalue problem. *Internat. J. Theoret. Phys.*, 38(11):2901–2916, 1999.

[15] J. Faraut and A. Korányi. *Analysis on symmetric cones.* Oxford Mathematical Monographs. The Clarendon Press Oxford University Press, New York, 1994.

[16] L. Faybusovich. A Jordan-algebraic approach to potential-reduction algorithms. *Math. Z.*, 239(1):117–129, 2002.

[17] L. Faybusovich. Euclidean Jordan algebras and interior-point algorithms. *Positivity*, 1(4):331–357, 1997.

[18] L. Faybusovich. Linear systems in Jordan algebras and primal-dual interior-point algorithms. *J. Comput. Appl. Math.*, 86(1):149–175, 1997.

[19] L. Faybusovich and R. Arana. A long-step primal-dual algorithm for the symmetric programming problem. *Systems Control Lett.*, 43(1):3–7, 2001.

[20] M. El Ghami. *New primal-dual interior-point methods based on kernel functions.* PhD thesis, TUDelft, 2005.

[21] F. Glineur. Improving complexity of structured convex optimization problems using self-concordant barriers. *European J. Oper. Res.*, 143(2):291–310, 2002.

[22] O. Güler. Barrier functions in interior point methods. *Math. Oper. Res.*, 21(4): 860–885, 1996.

[23] R. A. Hauser and O. Güler. Self-scaled barrier functions on symmetric cones and their classification. *Found. Comput. Math.*, 2(2):121–143, 2002.

[24] R. A. Hauser and Y. Lim. Self-scaled barriers for irreducible symmetric cones. *SIAM J. Optim.*, 12(3):715–723, 2002.

[25] R. A. Horn and C. R. Johnson. *Topics in matrix analysis.* Cambridge University Press, Cambridge, 1991.

[26] T. Jiang. Algebraic methods for diagonalization of a quaternion matrix in quaternionic quantum theory. *Journal of Mathematical Physics*, 46(5):052106, 2005.

[27] N. Karmarkar. A new polynomial-time algorithm for linear programming. *Combinatorica*, 4(4):373–395, 1984.

[28] A. Korányi. Monotone functions on formally real Jordan algebras. *Math. Ann.*, 269(1):73–76, 1984.

[29] P. Lancaster. On eigenvalues of matrices dependent on a parameter. *Numer. Math.*, 6:377–387, 1964.

[30] H. C. Lee. Eigenvalues and canonical forms of matrices with quaternion coefficients. *Proc. Roy. Irish Acad. Sect. A.*, 52:253–260, 1949.

[31] Y. Lim. Geometric means on symmetric cones. *Arch. Math. (Basel)*, 75(1):39–45, 2000.

[32] Z.-Q. Luo, J. F. Sturm, and S. Zhang. Conic convex programming and self-dual embedding. *Optim. Methods Softw.*, 14(3):169–218, 2000.

[33] H. Lütkepohl. *Handbook of matrices.* John Wiley & Sons Ltd., Chichester, 1996.

[34] A. W. Marshall and I. Olkin. *Inequalities: theory of majorization and its applications*, volume 143 of *Mathematics in Science and Engineering.* Academic Press Inc., New York, 1979.

[35] R. D. C. Monteiro and Y. Zhang. A unified analysis for a class of long-step primal-dual path-following interior-point algorithms for semidefinite programming. *Math. Programming*, 81(3, Ser. A):281–299, 1998.

[36] Evar D. Nering. *Linear algebra and matrix theory.* Wiley International Editions. John Wiley & Sons, Inc, second edition, 1970.

[37] Y. Nesterov and A. Nemirovskii. *Interior-point polynomial algorithms in convex programming*, volume 13 of *SIAM Studies in Applied Mathematics.* Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1994.

[38] Y. E. Nesterov and M. J. Todd. Self-scaled barriers and interior-point methods for convex programming. *Math. Oper. Res.*, 22(1):1–42, 1997.

[39] Y. E. Nesterov and M. J. Todd. Primal-dual interior-point methods for self-scaled cones. *SIAM J. Optim.*, 8(2):324–364, 1998.

[40] J. Peng, C. Roos, and T. Terlaky. A new and efficient large-update interior-point method for linear optimization. *Vychisl. Tekhnol.*, 6(4):61–80, 2001.

[41] J. Peng, C. Roos, and T. Terlaky. A new class of polynomial primal-dual methods for linear and semidefinite optimization. *European J. Oper. Res.*, 143(2):234–256, 2002.

[42] J. Peng, C. Roos, and T. Terlaky. Self-regular functions and new search directions for linear and semidefinite optimization. *Math. Program.*, 93(1, Ser. A):129–171, 2002.

[43] J. Peng, C. Roos, and T. Terlaky. *Self-regularity: a new paradigm for primal-dual interior-point algorithms.* Princeton Series in Applied Mathematics. Princeton University Press, Princeton, NJ, 2002.

[44] B. K. Rangarajan. Polynomial convergence of infeasible interior point methods over symmetric cones. *SIAM J. Optim.*, 16(4):1211–1229, March 2006.

[45] C. Roos, T. Terlaky, and J.-Ph. Vial. *Theory and algorithms for linear optimization.* Wiley-Interscience Series in Discrete Mathematics and Optimization. John Wiley & Sons Ltd., Chichester, 1997. (Second edition: Springer 2005).

[46] S. H. Schmieta and F. Alizadeh. Associative and Jordan algebras, and polynomial time interior-point algorithms for symmetric cones. *Math. Oper. Res.*, 26(3):543–564, 2001.

[47] S. H. Schmieta and F. Alizadeh. Extension of primal-dual interior point algorithms to symmetric cones. *Math. Program.*, 96(3, Ser. A):409–438, 2003.

[48] J. F. Sturm. Similarity and other spectral relations for symmetric cones. *Linear Algebra Appl.*, 312(1-3):135–154, 2000.

[49] J. F. Sturm. Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones. *Optim. Methods Softw.*, 11/12(1-4):625–653, 1999.

[50] M. J. Todd. Recent developments and new directions in linear programming. In *Mathematical programming (Tokyo, 1988)*, volume 6 of *Math. Appl. (Japanese Ser.)*, pages 109–157.

[51] S. J. Wright. *Primal-dual interior-point methods.* Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1997.

[52] Y. Ye. *Interior point algorithms.* Wiley-Interscience Series in Discrete Mathematics and Optimization. John Wiley & Sons Inc., New York, 1997.

Index

Summary

Jordan algebraic approach to symmetric optimization

In this thesis we present a generalization of interior-point methods for linear optimization based on kernel functions to symmetric optimization. It covers the three standard cases of conic optimization: linear optimization, second-order cone optimization and semi-definite optimization.

We give an introduction to Euclidean Jordan algebras and explain the connection between such algebras and symmetric cones.

We establish some properties of eigenvalues in Jordan algebras and prove that the barrier functions based on kernel functions are separable spectral functions that only depend on the eigenvalues of their arguments.

We propose an interior-point algorithm for symmetric optimization and derive its complexity bound.

Samenvatting

Jordan algebraïsche benadering van symmetrische optimalisering

Het onderwerp van dit proefschrift is een generalisatie van op kernfuncties gebaseerde inwendige punt methoden voor lineaire optimalisering naar symmetrische optimalisering. Het omvat de drie standaardgevallen van symmetrische optimalisering: lineare optimalisering, tweede-orde kegel optimalisering en semi-definiete optimalisering.

Wij geven een inleiding tot Euclidische Jordan algebras en verduidelijken de relatie tussen dergelijke algebras en symmetrische kegels.

Wij leiden enkele eigenschappen af van de eigenwaarden in Jordan algebras en bewijzen dat op kernfuncties gebaseerde barrière functies separable spectrale functies zijn die alleen afhangen van de eigenwaarden van hun argumenten.

Tenslotte introduceren wij een inwendige punt methode voor symmetrische optimalisering en leiden daarvoor een complexiteitsgrens af.

Curriculum Vitae

Manuel Vieira was born in Vila Nova de Foz Côa, Portugal on November 28, 1972.

He studied Applied Mathematics at the New University of Lisbon and graduated in 1996. During one year he worked as trainee at the National Electric Network. In 1998, he started to work as assistant teacher at the New University of Lisbon.

He got a master degree in Operational Research at the University of Lisbon, in 2002.

During the HPOPT 2002 Workshop, in Tilburg, he met Prof. dr. ir. Roos and they agreed to start his PhD in TUDelft in the following year.

He started his PhD in October 2003, at the Optimization Group, Department of Software Technology, Faculty of Electrical Engineering, Mathematics and Computer Science, TUDelft, under the supervision of Prof. dr. ir. Roos. During this period he got a LNMB (Dutch Network for the Mathematics of Operational Research) Diploma.

This PhD position at the Technical University of Delft was financially supported by the Portuguese Foundation for Science and Technology and by the New University of Lisbon.