CORE DISCUSSION PAPER
2006/13

Smoothing techniques in Euclidean Jordan algebras

Baes Michel

February 2006

Abstract

We extend the powerful smoothing techniques of Yu. Nesterov to the framework of Euclidean Jordan algebras. This study allows us to design a new scheme for minimizing the largest eigenvalue of an affine function on a Euclidean Jordan algebra. We prove that its complexity is in the order of $\mathcal{O}(1/\epsilon)$, where ϵ is the absolute tolerance on the value of the objective. Particularizing our result, we propose a new algorithm to minimize a sum of Euclidean norms and we perform its complete complexity analysis.

[1]CORE and INMA, Université catholique de Louvain, Belgium. E-mail: baes@core.ucl.ac.be

A research fellowship from the Belgian National Fund for Scientific Research is gratefully acknowledged by the author.
This text presents research results of the Belgian Program on Interuniversity Poles of Attraction initiated by the Belgian State, Prime Minister's Office, Science Policy Programming. The scientific responsibility is assumed by the author.

1 Introduction

Euclidean Jordan algebraic techniques are more and more used to generalize various results previously obtained in the framework of symmetric matrices. These techniques apply now in such different fields as statistics (e.g. [MN98]), positivity theory [GST04] or operation research – Leonid Faybusovich has initiated with [Fay97] this new field of investigation, which evolved rapidly by a large amount of contributions; among others, we can cite [Fay02], where potential-reduction methods have been extended to the Euclidean Jordan algebraic framework, and [Mur02, SA03], where Schmieta, Alizadeh, and Muramatsu have considered short- and long-step interior point methods, or [Ran05], where Rangarajan has developed an infeasible interior-point method. Among other adaptations, these extensions are performed by replacing the eigenvalues of symmetric matrices with the more general eigenvalues defined in the context of Euclidean Jordan algebras.

Some recent results of Nesterov tend to show that interior-point methods are not always the best procedures to solve some very large scale linear problems [Nes05a]. Whereas the number of iterations of these methods is *predictably* low, each of them requires so much work that performing the very first one might already be out of reach. Essentially, Nesterov has managed to combine the cheap iteration cost of subgradient methods and the efficiency of structural optimization in a very efficient method for solving some non-smooth optimization problem with a specific structure. In [Nes05a], he has designed a powerful scheme to minimize some piecewise linear function, and he extended it to solve some non-smooth problems involving symmetric matrices. Related problems have also been explored in the Master Thesis of Yu Qi[Qi05]. A natural question arises: can Euclidean Jordan algebras help to further extend this method ? We give a positive answer in this paper, and we particularize our study to the sum-of-norms problem.

The paper is organized as follows. In Section 2, we briefly recall how the smoothing techniques of Nesterov work. In Section 3, we present the few needed result from the theory of Euclidean Jordan algebras. Section 4 contains the main result of the paper, namely, the inequality (3), which allows us to estimate the complexity of smoothing techniques on Jordan algebras. We apply the obtained algorithm in Section 5 to solve the sum-of-norms problem, obtaining, up to our knowledge, the first theoretical complexity result for this problem.

2 Smoothing techniques in non-smooth convex optimization

The general problem of convex optimization can be formulated as follows. Given a convex function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ and a nonempty convex set $Q \subseteq \mathbb{R}^n$, find the minimal value f^* that f takes on Q, and, if possible, find a point of Q where this value is attained. On a finite-arithmetic computer, this goal is typically unreachable, and we content ourselves with an approximation of this minimal value: given an absolute tolerance $\epsilon > 0$, the problem consists in finding a point $\hat{x}$ in Q such that $f(\hat{x}) - f^* < \epsilon$.

The first methods designed for solving convex optimization problems that have been proposed and studied were the *subgradient schemes* (see in [Sho85], or in Chapter 2 and 3 of [Nes03] for an exposition). It has been proved, by resisting oracle techniques, that these methods cannot have a better complexity than $\mathcal{O}(1/\epsilon^2)$ in terms of number of iterations of the scheme [NY83]. Now, suppose that the objective function $f : \mathbb{R}^n \to \mathbb{R}$ is *smooth*, more precisely that its gradient exists and is Lipschitz continuous:

$$||\nabla f(x) - \nabla f(y)||_* \leq L||x - y|| \textit{ for every } x, y \in \operatorname{dom} f$$

where $|| \cdot ||$ is a norm of $\mathbb{R}^n$, $|| \cdot ||_*$ is the associated dual norm and the positive constant L the *gradient Lipschitz continuity constant*. In this case, the complexity analysis of subgradient schemes – we can actually call them *gradient schemes* in this case – shows that an approximate solution can be found in no more than $\mathcal{O}(\sqrt{L/\epsilon})$ iterations (see in Chapter 3 of [Nes03]).

Later appeared the interior-point methods [NN93]. In contrast with subgradient schemes, they explicitly exploit the structure of the problem: the construction of the *self-concordant barrier* needed in the algorithm mimics the mathematical description of the specific problem we have to solve. They have a complexity in the order of $\mathcal{O}(\sqrt{\nu}\ln(\nu/\epsilon))$ iterations, where ν is a structural parameter of the problem, usually a multiple of its dimension or of the number of constraints.

Subgradient schemes for non-smooth problems may seem completely outshone by interior-point methods. But the complexity of an iteration required by an interior-point method is much larger than the cost of a subgradient scheme iteration: indeed, interior-point methods typically require the resolution of a (typically sparse) linear system at each step, while subgradient methods only need vector manipulations (addition, computation of scalar products, projections on simple sets). Hence, very large-scale problems might be out of reach for interior-point methods.

The smoothing method of Nesterov [Nes05a] has been designed to potentially solve this issue, because, without affecting too severely the number of iterations, the iteration cost is much cheaper. It can be applied to optimization problems with the following very specific *structure* and performs at each iteration a cheap gradient-like step. We are given Q_1 and Q_2 two bounded convex set, respectively contained in the Euclidean vector spaces E_1 and E_2. The objective function, to be minimized over Q_1, is supposed to have the following form:

$$f(x) = \max_{u \in Q_2} \langle Ax, u \rangle - \hat{\phi}(u),$$

where $\hat{\phi}$ is a smooth convex function and A a linear operator from E_1 to E_2^*. We assume that an evaluation of f is not too costly, that is, that the maximization of $\langle Ax, u \rangle - \hat{\phi}(u)$ over Q_2 can be performed very efficiently, or even that a closed form of the solution exists.

The idea is to replace the non-smooth objective function f by a smooth approximation of it via a *prox-function* d_2 of Q_2, that is, a twice continuously differentiable function $d_2 : Q_2 \to \mathbb{R}$ whose minimal value is 0 and is attained in the relative interior of Q_2. We also require for a prox-function d_2 of Q_2 to be strongly convex on Q_2:

$$\text{for every } \; u \in Q_2 \; \text{ and } \; h \in E_2, \langle d_2''(u)h, h \rangle \geq \sigma_2 ||h||_{E_2}^2$$

for some norm $|| \cdot ||_{E_2}$ of E_2 and some *strong convexity constant* $\sigma_2 > 0$. We define for each parameter $\mu > 0$ the function:

$$f_\mu(x) := \max_{u \in Q_2} \langle Ax, u \rangle - \hat{\phi}(u) - \mu d_2(u).$$

This family of functions approaches f from below as μ goes to 0 and has a Lipschitz continuous gradient. We choose a norm $||\cdot||_{E_1}$ of E_1 and define

$$||A||_{E_1,E_2} := \max\{\langle Ax, u\rangle : ||x||_{E_1} \leq 1, ||u||_{E_2} \leq 1\}.$$

It can be proved (see Theorem 1 in [Nes05a]) that the Lipschitz constant of f'_μ equals $L_\mu := ||A||^2_{E_1,E_2}/(\mu\sigma_2)$. So, we can apply to it a cheap gradient-like scheme in order to minimize it.

This gradient-like scheme requires a prox-function d_1 of Q_1, whose strong convexity constant for the norm $||\cdot||_{E_1}$ will be denoted by σ_1 and its minimizer by x_0. The scheme updates at each step three sequences of points $(x_k)_{k\geq 0}$, $(y_k)_{k\geq 0}$, and $(z_k)_{k\geq 0}$. Letting $D_1 := \max_{x\in Q_1} d_1(x)$ and $D_2 := \max_{x\in Q_2} d_2(x)$, we put $\mu := \epsilon/(2D_2)$.

Algorithm 2.1 *For $k \geq 0$:*

1. *Compute $f'_\mu(x_k)$.*
2. *Find $y_k := \arg\min_{y\in Q_1}\left\{\langle f'_\mu(x_k), y - x_k\rangle + \frac{L_\mu}{2}||y - x_k||^2_{E_1}\right\}$.*
3. *Find $z_k := \arg\min_{y\in Q_1}\left\{\frac{L_\mu}{\sigma_1}d_1(y) + \sum_{i=1}^k \frac{i+1}{2}(\langle f'_\mu(x_i), y\rangle)\right\}$.*
4. *Let $x_{k+1} := \frac{k+1}{k+3}y_k + \frac{2}{k+3}z_k$.* ■

Theorem 2.1 (Theorem 3 in [Nes05a]) *For the sequence $(y_k)_{k\geq 0}$ generated by the algorithm, we have that $f(y_N) - f^* \leq \epsilon$ as soon as:*

$$N + 1 \geq 4||A||_{E_1,E_2}\sqrt{\frac{D_1D_2}{\sigma_1\sigma_2}} \cdot \frac{1}{\epsilon}.$$

■

In other words, this method is in $\mathcal{O}(1/\epsilon)$, which is the best known complexity for this class of non-smooth problems.

3 Euclidean Jordan algebras

In this work, we mostly deal with Euclidean Jordan algebras of finite dimension as they are defined in standard textbooks such as [BK66], [FK94], [HOS84] or [Koe99]. We briefly recall in this section the few needed basic results on these Jordan algebras. The reader can find in each of the above references the definitions we do not provide here.

Throughout the text, $\mathcal{J}$ denotes a Euclidean Jordan algebra (or, equivalently, formally real Jordan algebra) of finite dimension N and of rank r. Its unit element is denoted by e. To ease the writing, we drop the multiplication symbol between elements of $\mathcal{J}$.

We write $L(u)$ for the multiplication operator by an element $u \in \mathcal{J}$, so that $L(u)v := uv$ for all $v \in \mathcal{J}$. Two elements $u, v \in \mathcal{J}$ are said to *operator commute* when $L(u)L(v) = L(v)L(u)$. The quadratic operator is represented by $Q_u := 2L(u)^2 - L(u^2)$ and its polarization by $Q_{u,v} := (Q_{u+v} - Q_u - Q_v)/2$.

Theorem 3.1 (Unique subspaces spectral decomposition theorem)
Let $u \in \mathcal{J}$. There exists a system of idempotents $\{e_1, e_2, \ldots, e_s\}$ and distinct real numbers $\xi_1 > \ldots > \xi_s$ such that:

$$u = \sum_{i=1}^{s} \xi_i e_i.$$

This decomposition is unique *in the following sense: if there exists a system of idempotents $\{e'_1, \ldots, e'_k\} \in \mathcal{J}$ and some distinct real numbers $\eta_1, \ldots, \eta_k$ such that $u = \sum_{j=1}^{k} \eta_j e'_j$ then $k = s$ and, up to a renumbering, $\xi_j = \eta_j$ and $e_j = e'_j$ for all $1 \le j \le s$.*

A proof of this statement can be found in [FK94], Theorem III.1.1 or in [Koe99], Theorem VI.11. ■

Theorem 3.2 (Complete spectral decomposition theorem) *For every $u \in \mathcal{J}$, there exist a Jordan frame $\{c_1, \ldots, c_r\}$ and real numbers $\lambda_1(u) \ge \cdots \ge \lambda_r(u)$ such that:*

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

If there exists a Jordan frame $\{c'_1, \ldots, c'_r\}$ and real numbers $\eta_1 \ge \cdots \ge \eta_r$ for which $u = \sum_{i=1}^{r} \eta_i c'_i$, then $\eta_i = \lambda_i(u)$ for all i and $\sum_{\{j|\eta_j=\xi\}} c'_j = \sum_{\{j|\eta_j=\xi\}} c_j$ for every real number ξ.

See Theorem III.1.2 in [FK94] for a proof. ■

It can be proved with the help of Theorem 3.1 that the vector $\lambda(u) := (\lambda_1(u), \ldots, \lambda_r(u))^T$ is uniquely defined for every u of $\mathcal{J}$. The components of this vector are called the *eigenvalues* of u. By convention, we assume that they are always ordered decreasingly. In view of Proposition II.2.1 of [FK94], each function λ_i is continuous. The sum of the eigenvalues of u is called the *trace* of u and is denoted by $\mathrm{tr}(u)$.

Proposition 3.1 *The trace is a linear function. It is also associative: for all $u, v, w \in \mathcal{J}$, we can write $\mathrm{tr}((uv)w) = \mathrm{tr}(u(vw))$. In particular, $\mathrm{tr}(Q_u v) = \mathrm{tr}(u^2 v)$.*

This proposition merges results from Proposition II.2.1 and Proposition II.4.3 of [FK94]. ■

It is possible to obtain variational characterizations of the eigenvalues in Jordan algebras, similar to Fischer's formulas for Hermitian matrices (see [Hir70]). It is also possible to extend Ky Fan's inequalities in this framework (see [Bae04]). In this paper, we need the following basic characterization, which lies in fact at the intersection of these two results.

Proposition 3.2 *Let $K := \{v \in \mathcal{K}_\mathcal{J} | \mathrm{tr}(v) = 1\}$ be the Jordan algebraic extension of the standard simplex. For every element u of $\mathcal{J}$, we have:*

$$\lambda_1(u) = \max_{h \in K} \mathrm{tr}(uh).$$

■

The two spectral decomposition theorems allows us to construct the main object of interest in this paper, namely spectral functions. We mean by symmetric set of $\mathbb{R}^r$ a set that is invariant with respect to permutations of the components of its elements. Similarly, a symmetric function is here a function that remains unchanged under permutations of the components of its argument.

Definition 3.1 *Suppose that we are given a symmetric set $Q \subseteq \mathbb{R}^r$ and a symmetric function $f : Q \to \mathbb{R}$. The* spectral function generated by f *is the function F whose domain is $K := \{v \in \mathcal{J} | \lambda(v) \in Q\}$ and such that $F(v) := f(\lambda(v))$ for every $v \in K$.* ■

It is not difficult to deduce from Theorem 3.1 and from the required symmetry property of f that the definition of $F(v)$ does not depend on the particular complete spectral decomposition of v we have taken. The needed properties of this construction are exposed at the end of this section.

Proposition 3.3 *Two elements u, v of $\mathcal{J}$ operator commute if and only if there exist a Jordan frame $\{c_1, \ldots, c_r\}$ and two vectors $\gamma, \delta \in \mathbb{R}^r$ for which $u = \sum_{i=1}^r \gamma_i c_i$ and $v = \sum_{i=1}^r \delta_i c_i$.*

This is Theorem 27 of [SA03]. ■

Definition 3.2 *An element of $\mathcal{J}$ that has r different eigenvalues is called a* regular element.■

Definition 3.3 *The* Zariski topology *of $\mathbb{R}^N$ is the topology for which a set $A \subseteq \mathbb{R}^N$ is open if and only if there exists a polynomial $p : \mathbb{R}^N \to \mathbb{R}$ with real coefficients such that its set of roots is exactly $\mathcal{J} \backslash A$.* ■

Proposition 3.4 *Let $\mathcal{J}$ be a power-associative algebra. The set of regular elements of $\mathcal{J}$ is a Zariski nonempty open set of $\mathcal{J}$. Since $\mathcal{J}$ is an algebra over $\mathbb{R}$, this set is dense in $\mathcal{J}$ for the Euclidean topology.*

This statement is proved in [FK94], Proposition II.2.1. ■

The trace defines a scalar product represented here by $\langle u, v\rangle_{\mathcal{J}} := \mathrm{tr}(uv)$, or by $\langle u, v\rangle$ when there is no ambiguity about the considered scalar product. We denote the related norm by $||u||_{\mathcal{J}}$ or by $||u||$. The associativity of the trace is equivalent to the fact that $L(u)$ is self-adjoint with respect to the Jordan scalar product. The quadratic operator is self-adjoint too.

In the statement of the two next theorems the notation $A \circ B$ means the set $\{uv | u \in A, v \in B\}$ when subsets A and B belongs to $\mathcal{J}$.

Theorem 3.3 (First Pierce decomposition theorem) *Let c be an idempotent of $\mathcal{J}$. We define $\mathcal{J}_1(c) := Q_c \mathcal{J}$, $\mathcal{J}_{1/2}(c) := (I - Q_c - Q_{e-c})\mathcal{J} = 2Q_{c,e-c}\mathcal{J}$ and $\mathcal{J}_0(c) := Q_{e-c}\mathcal{J}$. Then:*

1. $\mathcal{J} = \mathcal{J}_1(c) \oplus \mathcal{J}_{1/2}(c) \oplus \mathcal{J}_0(c)$;
2. $\mathcal{J}_\alpha(c) = \{u \in \mathcal{J} | L(c)u = \alpha u\}$ *for* $\alpha = 1, 1/2, 0$;
3. $\mathcal{J}_1(c)$ *and* $\mathcal{J}_0(c)$ *are subalgebras of* $\mathcal{J}$ *and* $\mathcal{J}_0(c) \circ \mathcal{J}_1(c) = \{0\}$;

4. *$L(u)$ and $L(c)$ commute if and only if $u \in \mathcal{J}_0(c) \oplus \mathcal{J}_1(c)$;*

5. *$\mathcal{J}_{1/2}(c) \circ (\mathcal{J}_0(c) \oplus \mathcal{J}_1(c)) \subseteq \mathcal{J}_{1/2}(c)$;*

6. *$\mathcal{J}_{1/2}(c) \circ \mathcal{J}_{1/2}(c) \subseteq \mathcal{J}_0(c) \oplus \mathcal{J}_1(c)$;*

7. *if $u \in \mathcal{J}_{1/2}(c)$, then $\mathrm{tr}(u) = 0$.*

A proof of this statement can be found in [Koe99], Theorem III.8. ■

Theorem 3.4 (Second Pierce decomposition theorem)
Let $\{e_1, \ldots, e_n\}$ be a system of idempotents of $\mathcal{J}$. We put $\mathcal{J}_{ij} := Q_{e_i,e_j}\mathcal{J}$.
If $1 \leq i, j, k, l \leq n$, we have:

1. *$\mathcal{J}_{ii} = \mathcal{J}_1(c_i)$ and $\mathcal{J}_{ij} = \mathcal{J}_{1/2}(c_i) \cap \mathcal{J}_{1/2}(c_j) = \mathcal{J}_{ji}$ if $i \neq j$;*

2. *$\mathcal{J} = \bigoplus_{1 \leq i' \leq j' \leq n} \mathcal{J}_{i'j'}$;*

3. *$\mathcal{J}_{ij} \circ \mathcal{J}_{kl} = 0$, if $\{i, j\} \cap \{k, l\} = \emptyset$;*

4. *$\mathcal{J}_{ij} \circ \mathcal{J}_{jk} \subseteq \mathcal{J}_{ik}$, if i, j and k are different;*

5. *$\mathcal{J}_{ij} \circ \mathcal{J}_{ij} \subseteq \mathcal{J}_{ii} + \mathcal{J}_{jj}$;*

6. *$\mathcal{J}_{ii} \circ \mathcal{J}_{ij} \subseteq \mathcal{J}_{ij}$.*

This is Theorem IV.2.1 in [FK94]. ■

Corollary 3.1 *Let $\{e_1, \ldots, e_n\}$ be a system of idempotents, and let $h \in \mathcal{J}$. We denote $h_{ij} := Q_{e_i,e_j}h$ for every $1 \leq i, j \leq n$. Then $\mathrm{tr}(h_{ij}h) = 2\mathrm{tr}(h_{ij}^2)$ if $i \neq j$, and $\mathrm{tr}(h_{ii}h) = \mathrm{tr}(h_{ii}^2)$.*

Proof
From item 7 of Theorem 3.3 and item 1 of Theorem 3.4, we deduce that the subspaces $\mathcal{J}_{ij} := Q_{e_i,e_j}\mathcal{J}$ are orthogonal each other with respect to the Jordan scalar product. Hence,

$$\mathrm{tr}(h_{ii}h) = \langle h_{ii}, h \rangle = \langle h_{ii}, \sum_{j,k=1}^{r} h_{jk} \rangle = \langle h_{ii}, h_{ii} \rangle = \mathrm{tr}(h_{ii}^2).$$

If $i \neq j$, we have:

$$\mathrm{tr}(h_{ij}h) = \langle h_{ij}, h_{ij} + h_{ji} \rangle = 2\langle h_{ij}, h_{ij} \rangle = 2\mathrm{tr}(h_{ij}^2).$$

■

Proposition 3.5 (Eigenvalues and eigenspaces of $L(u)$) *Let $u = \sum_{i=1}^n \xi_i e_i$ be the decomposition of an element $u \in \mathcal{J}$ given by the unique eigenspaces spectral decomposition Theorem, and let $\mathcal{J}_{ij} := Q_{e_i,e_j}\mathcal{J}$ be the subspaces given by the second Pierce decomposition theorem for the system of idempotents $\{e_1, \ldots, e_n\}$.*

The eigenvalues of $L(u)$ are $\{(\xi_i + \xi_j)/2 | 1 \leq i \leq j \leq n\}$. The corresponding eigenspaces are $\{\mathcal{J}_{ij} | 1 \leq i \leq j \leq n\}$ respectively.

A proof can be found in Section V.5 of [Koe99]. ■

This result can be used to characterize the spectral decomposition of Q_u. In fact, we deduce here an interesting generalization of this decomposition practically for free.

Let $u, v \in \mathcal{J}$ be two elements that operator commute. From Proposition 3.3, we know that there exist a system of idempotents $\{e_1, \ldots, e_n\}$ and real numbers $\xi_1, \ldots, \xi_n$, $\xi'_1, \ldots, \xi'_n$ for which $u = \sum_{i=1}^n \xi_i e_i$ and $v = \sum_{i=1}^n \xi'_i e_i$, where we assume that the pairs (ξ_i, ξ'_i) are different.

Corollary 3.2 *With the above notation, the operator $Q_{u,v}$ has as eigenvalues*

$$\left\{ \frac{\xi_i \xi'_j + \xi_j \xi'_i}{2} \middle| 1 \leq i \leq j \leq n \right\}.$$

The eigenspace corresponding to $(\xi_i \xi'_j + \xi_j \xi'_i)/2$ is the direct sum of the subspaces $\mathcal{J}_{kl} := Q_{e_k, e_l} \mathcal{J}$ with $\xi_k \xi'_l + \xi_l \xi'_k = \xi_i \xi'_j + \xi_j \xi'_i$.

Proof
Let us fix $1 \leq i \leq j \leq n$. On the subspace $\mathcal{J}_{ij}$, the operator $Q_{u,v}$ reduces to the following:

$$\begin{aligned} Q_{u,v}|_{\mathcal{J}_{ij}} &= [L(u)L(v) + L(v)L(u) - L(uv)]|_{\mathcal{J}_{ij}} = [2L(v)L(u) - L(uv)]|_{\mathcal{J}_{ij}} \\ &= \frac{(\xi_i + \xi_j)(\xi'_i + \xi'_j)}{2} I - \frac{\xi_i \xi'_i + \xi_j \xi'_j}{2} I = \frac{\xi_i \xi'_j + \xi_j \xi'_i}{2} I, \end{aligned}$$

where I is the identity operator on $\mathcal{J}_{ij}$. The statement is hereby proved. In particular, if $u = v$, the eigenvalues of Q_u are $\{\xi_i \xi_j : 1 \leq i \leq j \leq n\}$. ■

We denote the (closed) cone of square elements of $\mathcal{J}$ by $\mathcal{K}_{\mathcal{J}}$. The following theorem summarizes the needed properties of this set.

Theorem 3.5 *For every $u \in \mathcal{K}_{\mathcal{J}}$, there exist $v \in \mathcal{K}_{\mathcal{J}}$ such that $v^2 = u$. We have $\mathcal{K}_{\mathcal{J}} = \{u \in \mathcal{J} | \lambda_r(u) \geq 0\}$. Moreover, for every $u \in \mathcal{J}$ and $v \in \mathcal{K}_{\mathcal{J}}$, the element $Q_u v$ is in $\mathcal{K}_{\mathcal{J}}$.*

See [FK94], Proposition III.2.2 for a demonstration. ■

Example 3.1 (Jordan spin algebra)
The *Jordan spin algebra*, or *spin factor*, or *quadratic terms algebra* is widely used in applications, ranging from statistics to relativistic mechanics. Optimizers utilize this algebra when they deal with second-order optimization problem, that is, optimization problems involving a convex quadratic objective to minimize on a convex quadratic set. We consider here the vector space $X := \mathbb{R}^{n+1}$, where $n \geq 1$. By convention, we denote by convention every vector $\bar{v}$ of X with an overline. The first component of $\bar{v}$ is written v_0, and the n-dimensional vector formed by its other components is written v, so that $\bar{v} = (v_0, v^T)^T$.

Consider an orthogonal basis $\{\bar{b}_0, \bar{b}_1, \ldots, \bar{b}_n\}$ of the vector space.We define the following multiplication in X.

$$\bar{u} \circ \bar{v} = \begin{pmatrix} u_0 \\ u \end{pmatrix} \circ \begin{pmatrix} v_0 \\ v \end{pmatrix} := \begin{pmatrix} u_0 v_0 + u^T v \\ u_0 v + v_0 u \end{pmatrix}, \quad \text{or} \quad L(\bar{u}) := \begin{pmatrix} u_0 & u^T \\ u & u_0 I_n \end{pmatrix}.$$

It is not difficult to prove that $\mathcal{S}_n := (X, \circ)$ is an Euclidean Jordan algebra. Its unit element is $\bar{e} := \bar{b}_0 = (1, 0, \ldots, 0)^T$, and that every idempotent $\bar{c}$ of $\mathcal{J}$ different from $\bar{e}$ has the form

$$\bar{c} = \frac{1}{2}\begin{pmatrix} 1 \\ u \end{pmatrix},$$

where u is an n-dimensional vector of Euclidean norm 1.

The trace of an element $\bar{u}$ is $\mathrm{tr}(\bar{u}) = 2u_0$, the determinant is $\det(\bar{u}) = u_0^2 - ||u||^2$, and the eigenvalues are $\lambda_1(\bar{u}) = u_0 + ||u||$ and $\lambda_2(\bar{u}) = u_0 - ||u||$; here $||\cdot||$ represents the Euclidean norm in $\mathbb{R}^n$. The quadratic operator can be written as:

$$Q_{\bar{u},\bar{v}} = \begin{pmatrix} u_0v_0 + u^Tv & u_0v^T + v_0u^T \\ u_0v + v_0u & uv^T + vu^T \end{pmatrix} + (u_0v_0 - u^Tv)\begin{pmatrix} 0 & 0 \\ 0 & I_n \end{pmatrix}.$$

■

Von Neumann's inequality, and especially the description of the equality case, plays an important role in the computation of subdifferentials of spectral function. In this work, we need it in the derivation of a bound for the Hessian of a prox-function; the description of the equality case is unnecessary. The interested reader can find a proof in [LKF03], although it only covers the case where $\mathcal{J}$ is a simple Jordan algebra. An alternative demonstration can be found in [Bae04].

Theorem 3.6 *Let $u, v \in \mathcal{J}$. We have:*

$$\sum_{i=1}^{r} \lambda_i(u)\lambda_i(v) \geq \mathrm{tr}(uv). \tag{1}$$

We have proved in [Bae04] and in [Bae05] that a symmetric function f transmits several properties to the spectral function F it generates. In this paper, we need to deal with their conjugate and their differentiability.

Let $f : \mathbb{R}^n \to \mathbb{R}$. Provided that $\mathbb{R}^n$ is endowed with a scalar product $\langle\cdot,\cdot\rangle$, we define *the conjugate function of f* as follows:

$$f^* : \mathbb{R}^n \to \mathbb{R}, \quad s \mapsto f^*(s) := \sup_{x \in \mathrm{dom}\, f} \langle s, x\rangle - f(x) = \sup_{x \in \mathbb{R}^n} \langle s, x\rangle - f(x).$$

It is possible to relate the conjugate of a symmetric function to the conjugate of the spectral function it generates. The following theorem has been proved in [Bae04] using the argument of the proof of Lewis [Lew96a], who has obtained the same result in the framework of Hermitian matrices.

Theorem 3.7 *Let Q be a nonempty symmetric set of $\mathbb{R}^r$, let $f : Q \to \mathbb{R}$ be a symmetric function and let F be the spectral function generated by f. Then F^* is the spectral function generated by f^*.* ■

The following theorem has been proved in [Bae04], following an idea of Lewis [Lew96b], and reproved independently by [SS04].

Theorem 3.8 *Let $Q \subseteq \mathbb{R}^r$ be an open symmetric set and $f : Q \to \mathbb{R}$ be a symmetric function. We define $K := \{v \in \mathcal{J} | \lambda(v) \in Q\}$ and $F : K \to \mathbb{R}$, $v \mapsto F(v) := f(\lambda(v))$. Let $u = \sum_{i=1}^r \lambda_i(u) c_i \in K$. If the function f is differentiable in $\lambda(u)$, then the function F is differentiable in u and*

$$F'(u) = \sum_{i=1}^{r} f_i'(\lambda(u)) c_i. \tag{2}$$

■

This last theorem is proved in [Bae05].

Theorem 3.9 *Let $Q \subseteq \mathbb{R}^r$ be an open symmetric set and $f : Q \to \mathbb{R}$ be a twice differentiable symmetric function. We define $K := \{v \in \mathcal{J} | \lambda(v) \in Q\}$ and $F : K \to \mathbb{R}$, $v \mapsto F(v) := f(\lambda(v))$. Let u be an element of K, with a complete spectral decomposition as follows: $u = \sum_{i=1}^r \lambda_i(u) c_i$, and a unique eigenspaces spectral decomposition as: $u = \sum_{\alpha=1}^s \lambda_{k_\alpha}(u) e_\alpha$. We denote $f''(\lambda(u)) = \mathrm{diag}(b) + B$. We have for every $h \in \mathcal{J}$:*

- ◇ $Q_{e_\alpha} F''(u) h = b_{k_\alpha} Q_{e_\alpha} h + \sum_{\beta=1}^s B_{k_\alpha k_\beta} \mathrm{tr}(Q_{e_\beta} h) e_\alpha$.
- ◇ $Q_{e_\alpha, e_\beta} F''(u) h = \dfrac{f'_{k_\alpha}(\lambda(u)) - f'_{k_\beta}(\lambda(u))}{\lambda_{k_\alpha}(u) - \lambda_{k_\beta}(u)} Q_{e_\alpha, e_\beta} h$.

Let $v, w \in \mathcal{J}$ and $v_{\alpha\beta} := Q_{e_\alpha, e_\beta} v$, $w_{\alpha\beta} := Q_{e_\alpha, e_\beta} w$. Then:

$$\begin{aligned} \langle F''(u) v, w \rangle &= \sum_{\alpha=1}^s b_{k_\alpha} \mathrm{tr}(v_{\alpha\alpha} w_{\alpha\alpha}) + \sum_{\alpha,\beta=1}^s B_{k_\alpha k_\beta} \mathrm{tr}(v_{\beta\beta}) \mathrm{tr}(w_{\alpha\alpha}) \\ &\quad + 2 \sum_{\alpha \neq \beta} \frac{f'_{k_\alpha}(\lambda(u)) - f'_{k_\beta}(\lambda(u))}{\lambda_{k_\alpha}(u) - \lambda_{k_\beta}(u)} \mathrm{tr}(v_{\alpha\beta} w_{\alpha\beta}). \end{aligned}$$

■

Corollary 3.3 *With the same notation as in the previous statement, if f is twice continuously differentiable, then F is twice continuously differentiable.* ■

4 An upper bound on the Hessian of power function

We generalize in this section to Euclidean Jordan algebras an inequality obtained recently by Nesterov [Nes05b] in the framework of symmetric matrices.

For every nonnegative integer k and every real r-dimensional vector λ, we let:

$$p_k(\lambda) := \lambda_1^k + \cdots + \lambda_r^k.$$

The spectral function generated by p_k is denoted by P_k:

$$\begin{aligned} P_k : \mathcal{J} &\to \mathbb{R} \\ u &\mapsto P_k(u) := \mathrm{tr}(u^k). \end{aligned}$$

The main result of this section is the following inequality.

For every integer $k \geq 2$, for every element $u = \sum_{i=1}^{r} \lambda_i(u)c_i$ of $\mathcal{J}$, and for every direction h of $\mathcal{J}$, we have:

$$\langle P_k''(u)h, h\rangle \leq k(k-1)\langle |u|^{k-2}h, h\rangle, \tag{3}$$

where $|u| := \sum_{i=1}^{r} |\lambda_i(u)|c_i$. ■

Its consequences will allow us to extend the smoothing techniques in the framework of Jordan algebras, and to determine a complexity bound for the obtained scheme.

Lemma 4.1 *Let p and q be two nonnegative integers. For every $u \in \mathcal{J}$, the operator $L(|u|^{p+q}) - Q_{u^p,u^q}$ is positive semidefinite. In other words, for every $h \in \mathcal{J}$, we have:*

$$\langle |u|^{p+q}h, h\rangle \geq \langle Q_{u^p,u^q}h, h\rangle.$$

Proof
Let us fix an element $u \in \mathcal{J}$, and let us consider one of its complete spectral decomposition $u = \sum_{i=1}^{r} \lambda_i(u)c_i$. For the sake of notational simplicity, we write λ for $\lambda(u)$. From Proposition 3.5 and Corollary 3.2, we know that $L(|u|^{p+q})$ and Q_{u^p,u^q} have the same eigenspaces, which are direct sums of the subspaces $\mathcal{J}_{ij} := Q_{c_i,c_j}\mathcal{J}$. The eigenvalues corresponding to $\mathcal{J}_{ij}$ are respectively $(|\lambda_i|^{p+q} + |\lambda_j|^{p+q})/2$ for $L(|u|^{p+q})$, and $(\lambda_j^p\lambda_i^q + \lambda_i^p\lambda_j^q)/2$ for Q_{u^p,u^q}. Observe that:

$$(|\lambda_i|^p - |\lambda_j|^p)(|\lambda_i|^q - |\lambda_j|^q) \geq 0,$$

so that:

$$|\lambda_i|^{p+q} + |\lambda_j|^{p+q} \geq |\lambda_j|^p|\lambda_i|^q + |\lambda_i|^p|\lambda_j|^q \geq \lambda_j^p\lambda_i^q + \lambda_i^p\lambda_j^q.$$

In other words, the eigenvalue of $L(|u|^{p+q}) - Q_{u^p,u^q}$ are nonnegative. ■

Proposition 4.1 *For every u and h of $\mathcal{J}$, the inequality (3):*

$$\langle P_k''(u)h, h\rangle \leq k(k-1)\langle |u|^{k-2}h, h\rangle$$

holds true for all $k \geq 2$.

Proof
Since the Hessian is continuous, it suffices to show the inequality for regular elements u, because they form a dense set in $\mathcal{J}$. Let us fix a regular element $u = \sum_{i=1}^{r} \lambda_i(u)c_i$ of $\mathcal{J}$, and let us compute $\langle P_k''(u)h, h\rangle$ using the formula for the Hessian.

We easily get:

$$[p_k'(\lambda)]_i = k\lambda_i^{k-1} \qquad \text{and} \qquad [p_k''(\lambda)]_{ij} = \delta_{ij}k(k-1)\lambda_i^{k-2},$$

where δ_{ij} is the Kronecker symbol. Let h be an element of $\mathcal{J}$, and let $h_{ij} := Q_{c_i,c_j}h$, so that $h = \sum_{i,j=1}^{r} h_{ij}$. The second Pierce decomposition of h with respect to the Jordan frame $\{c_1, \ldots, c_r\}$ is thus:

$$h = \sum_{i=1}^{r} h_{ii} + 2\sum_{i<j} h_{ij}.$$

We have by regularity of u:

$$\begin{aligned}\langle P_k''(u)h,h\rangle &= \sum_{i=1}^{r} k(k-1)\lambda_i^{k-2}\mathrm{tr}(h_{ii}^2) + 2\sum_{i\neq j} k\frac{\lambda_i^{k-1}-\lambda_j^{k-1}}{\lambda_i-\lambda_j}\mathrm{tr}(h_{ij}^2)\\ &= k\left(\sum_{i=1}^{r}(k-1)\lambda_i^{k-2}\mathrm{tr}(h_{ii}^2) + 2\sum_{i\neq j}\sum_{l=0}^{k-2}\lambda_i^l\lambda_j^{k-l-2}\mathrm{tr}(h_{ij}^2)\right)\\ &= k\left(\sum_{i=1}^{r}(k-1)\lambda_i^{k-2}\mathrm{tr}(h_{ii}^2) + 2\sum_{i\neq j}\sum_{l=0}^{k-2}\frac{\lambda_i^l\lambda_j^{k-l-2}+\lambda_j^l\lambda_i^{k-l-2}}{2}\mathrm{tr}(h_{ij}^2)\right).\end{aligned}$$

Observe now that, for every nonnegative integers p and q, we can write:

$$\langle Q_{u^p,u^q}h,h\rangle = \sum_{i,j=1}^{r}\frac{\lambda_i^p\lambda_j^q+\lambda_j^p\lambda_i^q}{2}\mathrm{tr}(h_{ij}h) = \sum_{i=1}^{r}\lambda_i^{p+q}\mathrm{tr}(h_{ii}^2) + \sum_{i\neq j}\frac{\lambda_i^p\lambda_j^q+\lambda_j^p\lambda_i^q}{2}\mathrm{tr}(h_{ij}^2).$$

With this relation, we can continue as follows:

$$\begin{aligned}\langle P_k''(u)h,h\rangle &= k\left(\sum_{i=1}^{r}(k-1)\lambda_i^{k-2}\mathrm{tr}(h_{ii}^2) + \sum_{l=0}^{k-2}\left(\langle Q_{u^l,u^{k-l-2}}h,h\rangle - \sum_{i=1}^{r}\lambda_i^{k-2}\mathrm{tr}(h_{ii}^2)\right)\right)\\ &= k\sum_{l=0}^{k-2}\langle Q_{u^l,u^{k-l-2}}h,h\rangle \leq k\sum_{l=0}^{k-2}\langle L(|u|^{k-2})h,h\rangle = k(k-1)\langle |u|^{k-2}h,h\rangle,\end{aligned}$$

where the inequality comes from Lemma 4.1. ■

The following corollaries are simple but very useful consequences of the previous proposition. Their proof follows closely those of [Nes05b].

Corollary 4.1 *Let $f:\mathbb{R}\to\mathbb{R}$ be a function that has a power series expansion*

$$f(t) = \sum_{k\geq 0} a_k t^k$$

such that all the coefficients a_k are nonnegative. Let us denote the domain of f by I, and the set containing all the elements of $\mathcal{J}$ that have their eigenvalues in I by K. We define $F: K\to\mathbb{R}$, $u\mapsto F(u) := \sum_{i=1}^r f(\lambda_i(u))$. For every $u\in K$ and all $h\in\mathcal{J}$, we have:

$$\langle F''(u)h,h\rangle \leq \sum_{i=1}^{r} f''(|\lambda_i(u)|)\lambda_i(h)^2.$$

Proof
By Proposition 4.1, we can write:

$$\langle F''(u)h,h\rangle = \sum_{k\geq 2} a_k\langle P_k''(u)h,h\rangle \leq \sum_{k\geq 2}\sum_{i=1}^{r} k(k-1)a_k\mathrm{tr}(|u|^{k-2}h^2)$$

The von Neumann inequality gives us $\mathrm{tr}(|u|^{k-2}h^2) \leq \sum_{i=1}^r |\lambda_i(u)|^{k-2}\lambda_i(h^2)$, from which we get:

$$\langle F''(u)h, h\rangle \leq \sum_{k\geq 2}\sum_{i=1}^{r} k(k-1)a_k|\lambda_i(u)|^{k-2}\lambda_i(h^2).$$

Now, since $f''(t) = \sum_{k\geq 2} k(k-1)a_k t^{k-2}$, we conclude:

$$\langle F''(u)h, h\rangle \leq \sum_{i=1}^{r} f''(|\lambda_i(u)|)\lambda_i(h^2).$$

■

Corollary 4.2 *Consider the function* $F : \mathcal{J} \to \mathbb{R}$, $u \mapsto F(u) := \sum_{i=1}^r \exp(\lambda_i(u))$, *and the function* $E(u) := \ln F(u)$. *Then*

$$\langle E''(u)h, h\rangle \leq \lambda_1(h^2)$$

for every u *and* h *of* $\mathcal{J}$.

Proof
A straightforward computation gives us:

$$\langle E''(u)h, h\rangle = \frac{\langle F''(u)h, h\rangle}{F(u)} - \frac{\langle F'(u), h\rangle^2}{F(u)^2} \leq \frac{\langle F''(u)h, h\rangle}{F(u)}.$$

Suppose preliminarily that $u \in \mathcal{K}_{\mathcal{J}}$. It is well-known that the coefficients of the power-series expansion of exp are positive. Using then the previous corollary, we can continue as follows:

$$\langle E''(u)h, h\rangle \leq \frac{\langle F''(u)h, h\rangle}{F(u)} \leq \frac{\sum_{i=1}^r \exp(|\lambda_i(u)|)\lambda_i(h^2)}{\sum_{i=1}^r \exp(\lambda_i(u))} \leq \lambda_1(h^2).$$

Now, observe that the element $u - Te$ is always in the cone of squares when T is smaller than $\lambda_r(u)$. Note also that $E(u - Te) = E(u) - T$. Hence, the above inequality holds true even for elements u that are not in $\mathcal{K}_{\mathcal{J}}$. ■

Corollary 4.3 *Let* $K := \{v \in \mathcal{K}_{\mathcal{J}} | \mathrm{tr}(v) = 1\}$ *be the Jordan algebraic extension of the standard simplex. The function* $d : K \to \mathbb{R}$, $v \mapsto d(v) := \sum_{i=1}^r \lambda_i(v)\ln\lambda_i(v)$ *satisfies, for all* $h \in \mathcal{J}$ *and all* $u \in K$, *the following inequality:*

$$\langle d''(u)h, h\rangle \geq ||h||_1^2,$$

where $||h||_1 := \sum_{i=1}^r |\lambda_i(h)|$ *is the norm generated by the* 1*-norm in* $\mathbb{R}^r$.

Proof
Let $\eta(\lambda) := \ln\sum_{i=1}^r \exp(\lambda_i)$ for every $\lambda \in \mathbb{R}^r$. The conjugate of the function η is $\delta(\lambda) := \sum_{i=1}^r \lambda_i \ln\lambda_i$ on the standard simplex

$$Q := \{\lambda \in \mathbb{R}^r | \sum_{i=1}^{r} \lambda_i = 1,\, \lambda_i \geq 0 \text{ for every } i\}.$$

The function d is then the conjugate of the spectral function E defined in the previous corollary. It is well-known (see Theorem 4.2.2 in [HUL93]) strong convexity and Lipschitz continuity of the gradient are dual notions. In other words, suppose that the function $f : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ is twice differentiable; then:

$$\langle f''(x)h, h\rangle \leq \frac{L}{2}||h||^2 \quad \forall\, x \in \operatorname{dom} f \text{ and } h \in \mathcal{J}$$

if and only if:

$$\langle f^{*\prime\prime}(x)h, h\rangle \geq \frac{1}{2L}||h||_*^2 \quad \forall\, x \in \operatorname{dom} f^* \text{ and } h \in \mathcal{J},$$

where $||\cdot||_*$ is the dual norm of $||\cdot||$. As the dual norm of $||h||_\infty := \sqrt{\lambda_1(h^2)}$ is the norm $||\cdot||_1$, we get that $\langle d''(u)h, h\rangle \geq ||h||_1^2$. ■

Let us consider the function:

$$\phi(x) := \max_{u\in K}\langle Ax, u\rangle - \langle b, u\rangle,$$

that maps $\mathbb{R}^m$ to $\mathbb{R}$; the set K is, like in the above corollary, the Jordan algebraic extension of the standard simplex. The linear application A maps $\mathbb{R}^m$ to $\mathcal{J}$, and the element b belongs to $\mathcal{J}$. The scalar products should be understood as Jordan scalar product. In view of Proposition 3.2, the function ϕ is exactly equal to $\lambda_1(Ax - b)$.

Using the prox-function $d_2(u) := \sum_{i=1}^r(\lambda_i(u)\ln\lambda_i(u)) + \ln r$ for K, we get:

$$\phi_\mu(x) := \max_{u\in K}\langle Ax, u\rangle - \langle b, u\rangle - \mu d_2(u) = \mu d_2^*((Ax - b)/\mu),$$

or, when $\mu > 0$,

$$\phi_\mu(x) = \mu \ln\left(\sum_{i=1}^r \exp(\lambda_i(Ax - b)/\mu)\right) - \mu\ln(r).$$

The above corollary ensures that the strong convexity constant σ_2 related to this smoothing equals 1 for the best possible norm (i.e. with the smallest unit ball), namely $||h||_{E_2} := \sum_{i=1}^r |\lambda_i(h)|$.

5 Sum of norms problem

The sum of norms problem can be formulated as follows. Given p real matrices $\{A_1, \dots, A_p\}$ of dimension $m \times n$ and p real m-dimensional vectors $\{b_1, \dots, b_p\}$, we need to minimize the function

$$f(x) := \sum_{j=1}^p ||A_j x - b_j||$$

over $Q_1 := \{x \in \mathbb{R}^n : ||x|| \leq R\}$, where $||\cdot||$ stands for the standard Euclidean norm of $E_1 := \mathbb{R}^n$ or over $\mathbb{R}^m$.

In this paper, we propose to solve this problem by using the smoothing function techniques in the Jordan spin algebra $\mathcal{J} := \mathcal{S}_m$. To do so, we define for all j the following elements:

$$\bar{A}_j = \begin{pmatrix} 0 \\ A_j \end{pmatrix} \qquad \text{and} \qquad \bar{b}_j = \begin{pmatrix} 0 \\ b_j \end{pmatrix}.$$

We also introduce the function:

$$\begin{aligned} \bar{f} : \mathbb{R}^n &\to \mathbb{R} \\ \bar{x} &\mapsto \bar{f}(x) := \sum_{j=1}^{p} \lambda_1(\bar{A}_j x - \bar{b}_j), \end{aligned}$$

where λ_1 is the largest eigenvalue of its argument in the Euclidean Jordan algebra $\mathcal{S}_m$. Observe that minimizing $\bar{f}$ over Q_1 is completely equivalent to the sum of norms problem.

Since λ_1 is the support function of the Jordan algebraic version of the standard simplex:

$$\Delta := \{\bar{u} \in \mathcal{S}_m | \lambda_1(\bar{u}) + \lambda_2(\bar{u}) = 1, \lambda_2(\bar{u}) \geq 0\} = \left\{ \bar{u} = \begin{pmatrix} 1/2 \\ u \end{pmatrix} \in \mathcal{J} : ||u|| \leq 1/2 \right\},$$

we can rewrite our function $\bar{f}$ as follows:

$$\bar{f}(\bar{x}) = \sum_{j=1}^{p} \lambda_1(\bar{A}_j x - \bar{b}_j) = \sum_{j=1}^{p} \max_{\bar{u}_j \in \Delta} \langle \bar{u}_j, \bar{A}_j x - \bar{b}_j \rangle_{\mathcal{J}}.$$

Now, we define

$$A := \begin{pmatrix} \bar{A}_1 \\ \vdots \\ \bar{A}_p \end{pmatrix} \qquad \text{and} \qquad b := \begin{pmatrix} \bar{b}_1 \\ \vdots \\ \bar{b}_p \end{pmatrix}.$$

Our expression of $\bar{f}$ then becomes $\bar{f}(x) = \max_{\bar{u} \in \Delta^p} \langle Ax, \bar{u} \rangle_{\mathcal{J}^p} - \langle b, \bar{u} \rangle_{\mathcal{J}^p}$, and our problem enters into the class of those for which smoothing techniques are applicable.

In the notation of Section 2, we let $|| \cdot ||_{E_1}$ be the *Euclidean norm*, and we let the prox-function for Q_1 be:

$$d_1(x) := \frac{||x||_{E_1}^2}{2}.$$

It is easy to compute that the constant σ_1 equals 1 and that $D_1 = \max\{d_1(x) | x \in Q_1\} = R^2/2$.

The space E_2 will be $\mathcal{J}^p$. For the set $Q_2 := \Delta^p$, we propose the following prox-function:

$$d_2(\bar{u}) := \sum_{j=1}^{p} ||\bar{A}_j^*|| \cdot [\lambda_1(\bar{u}_j) \ln(\lambda_1(\bar{u}_j)) + \lambda_2(\bar{u}_j) \ln(\lambda_2(\bar{u}_j)) + \ln 2],$$

and the following norm:

$$||\bar{u}||_{E_2} := \sqrt{\sum_{j=1}^{p} ||\bar{A}_j^*|| \cdot ||\bar{u}_j||_1^2}.$$

We have used the notation $||\cdot||_1$ to designate the spectral norm generated by the 1-norm on $\mathbb{R}^2$. The number $||\bar{A}_j^*||$ denotes here the maximum value that $\langle \bar{A}_j^* u_j, x\rangle$ can take when $||u_j||_1 \leq 1$ and $||x||_{E_1} \leq 1$. A straightforward computation shows that it equals $\sqrt{\lambda_{\max}(A_j^* A_j)}$, i.e. the maximal singular value of A_j.

We know from Corollary 4.3 that for every $\bar{h}_1, \ldots, \bar{h}_p \in \mathcal{J}$:

$$\sum_{j=1}^{p} ||\bar{A}_j^*|| \langle [\lambda_1(\bar{u}_j)\ln(\lambda_1(\bar{u}_j)) + \lambda_2(\bar{u}_j)\ln(\lambda_2(\bar{u}_j))]'' \bar{h}_j, \bar{h}_j\rangle \geq \sum_{j=1}^{p} ||\bar{A}_j^*|| \cdot ||\bar{h}_j||_1^2 = ||\bar{h}||_{E_2}^2.$$

Hence, we can take $\sigma_2 := 1$. Now, $D_2 = \max\{d_2(\bar{u}) | \bar{u} \in Q_2\} = \sum_{j=1}^{r} ||\bar{A}_j^*|| \ln 2$. It remains to compute the quantity $||A||_{E_1,E_2}$:

$$\begin{aligned}
||A||_{E_1,E_2} &= \max\{\langle Ax, \bar{u}\rangle_{\mathcal{J}^p} : ||x||_{E_1} \leq 1, \sum_{j=1}^{p} ||\bar{A}_j^*|| \cdot ||\bar{u}_j||_1^2 \leq 1\} \\
&= \max\left\{\sum_{j=1}^{p} \langle \bar{A}_j x, \bar{u}_j\rangle_{\mathcal{J}} : ||x||_{E_1} \leq 1, \sum_{j=1}^{p} ||\bar{A}_j^*|| \cdot ||\bar{u}_j||_1^2 \leq 1\right\} \\
&\leq \max\left\{\sum_{j=1}^{p} ||\bar{A}_j^*|| \cdot ||\bar{u}_j||_1 \cdot ||x||_{E_1} : ||x||_{E_1} \leq 1, \sum_{j=1}^{p} ||\bar{A}_j^*|| \cdot ||\bar{u}_j||_1^2 \leq 1\right\} \\
&\leq \max\left\{\sum_{j=1}^{p} ||\bar{A}_j^*|| \cdot ||\bar{u}_j||_1 : \sum_{j=1}^{p} ||\bar{A}_j^*|| \cdot ||\bar{u}_j||_1 \leq \sqrt{\sum_{j=1}^{p} ||\bar{A}_j^*||}\right\} = \sqrt{\sum_{j=1}^{p} ||\bar{A}_j^*||}.
\end{aligned}$$

The last inequality comes from the Cauchy-Schwarz relation:

$$\left(\sum_{j=1}^{p} ||\bar{A}_j^*|| \cdot ||\bar{u}_j||_1\right)^2 \leq \left(\sum_{j=1}^{p} ||\bar{A}_j^*||\right)\left(\sum_{j=1}^{p} ||\bar{A}_j^*|| \cdot ||\bar{u}_j||_1^2\right).$$

Letting $M := \sum_{j=1}^{p} ||\bar{A}_j^*||$, we can now conclude that the Algorithm 2.1 has the following rate of convergence:

$$\bar{f}(\bar{y}_N) - \bar{f}^* \leq \frac{4||A||}{N+1}\sqrt{\frac{D_1 D_2}{\sigma_1 \sigma_2}} = \frac{4\sqrt{\ln 2} MR}{N+1} = \mathcal{O}\left(\frac{RM}{N}\right).$$

If the matrices A_j are scaling matrices, that is, matrices of the form $A_j := m_j I$, Nesterov has shown that the same order of convergence can be predicted with the following smoothed version of f:

$$f_\mu(x) := \sum_{i=1}^{p} m_j \psi_\mu(||x - c_j||),$$

$$\text{with } \psi_\mu(t) = \begin{cases} t^2/2\mu & \text{if } 0 \leq t \leq \mu, \\ t - \mu/2 & \text{if } \mu \leq t. \end{cases}$$

Observe that the problem:

$$\min_{||x||\le R} \sum_{j=1}^{p} |\langle a_j, x\rangle - b_j|$$

is a particular case of the problem we have considered – it suffices to take $m = 1$. In this case, the constant M is the sum of Euclidean norms of the vectors a_j.

Acknowledgements

I warmly thank Yurii Nesterov for the many discussions that have motivated the research done in this paper.

References

[Bae04] M. Baes, *Spectral functions on Jordan algebras: differentiability and convexity properties*, CORE Discussion Paper 16, Center for Operation Research and Econometrics, Université catholique de Louvain, 2004.

[Bae05] ———, *Convexity and differentiability properties of spectral functions and spectral mappings on euclidean jordan algebras*, Submitted to Linear Algebras and Applications, 2005.

[BK66] H. Braun and M. Koecher, *Jordan-Algebren*, Grundlehren der mathematischen Wissenschaften, no. 128, Springer, 1966.

[Fay97] L. Faybusovich, *Linear systems in Jordan algebras and primal-dual interior-point algorithms*, Journal of Computational and Applied Mathematics **86** (1997), no. 1, 149–175.

[Fay02] ———, *A Jordan-algebraic approach to potential-reduction algorithms*, Mathematische Zeitschrift **239** (2002), no. 1, 117–129.

[FK94] J. Faraut and A. Koranyi, *Analysis on Symmetric Cones*, Oxford University Press, 1994.

[GST04] M. Gowda, R. Sznajder, and J. Tao, *Some P-properties for linear transformations on Euclidean Jordan algebras*, Linear Algebra and its Applications, Special issue on Positivity **393** (2004), 203–232.

[Hir70] U. Hirzebruch, *Der Min-Max-Satz von E. Fischer für formal-reelle Jordan-Algebren*, Mathematische Annalen **186** (1970), 65–69.

[HOS84] H. Hanche-Olsen and E. Stormer, *Jordan Operator Algebras*, Monographs and studies in mathematics, no. 21, Pitman, 1984.

[HUL93] J.-B. Hiriart-Urruty and C. Lemaréchal, *Convex Analysis and Minimization Algorithms II*, Grundlehren der mathematischen Wissenschaften, vol. 306, Springer, 1993.

[Koe99] M. Koecher, *The Minnesota Notes on Jordan Algebras and Their Applications*, Lecture Notes in Mathematics, no. 1710, Springer-Verlag, 1999.

[Lew96a] A. Lewis, *Convex analysis on the Hermitian matrices*, SIAM Journal on Optimization **6** (1996), no. 1, 164–177.

[Lew96b] ———, *Derivatives of spectral functions*, Mathematics of Operations Research **21** (1996), no. 3, 576–588.

[LKF03] Y. Lim, J. Kim, and L. Faybusovich, *Simultaneous diagonalization on simple Euclidean Jordan algebras and its applications*, Forum Mathematicum **15** (2003), no. 4, 639–644.

[MN98] H. Massam and E. Neher, *Estimation and testing for lattice conditional independence models on Euclidean Jordan algebras*, The Annals of Statistics **26** (1998), 1051–1082.

[Mur02] M. Muramatsu, *On commutative class of search directions for linear programming over symmetric cones*, Journal of Optimization Theory and Applications **112** (2002), no. 3, 595–625.

[Nes03] Y. Nesterov, *Introductory lectures on convex optimization: a basic course*, Applied Optimization, vol. 87, Kluwer Academic Publishers, 2003.

[Nes05a] ———, *Smooth minimization of non-smooth functions*, Mathematical Programming **103** (2005), no. 1, 127–152.

[Nes05b] ———, *Smoothing technique and its applications in semidefinite optimization*, Accepted for publication in Mathematical Programming (2005).

[NN93] Y. Nesterov and A. Nemirovski, *Interior Point Algorithms in Convex Programming*, Studies in Applied Mathematics, no. 13, SIAM, 1993.

[NY83] A. Nemirovski and D. Yudin, *Problem Complexity and Method Efficiency in Optimization*, John Wiley, 1983.

[Qi05] Y. Qi, *Smoothing approximations for two classes of convex eigenvalue optimization problems*, Ph.D. thesis, Department of Mathematics, National University of Singapore, 2005.

[Ran05] B. Rangarajan, *Polynomial convergence of infeasible-interior-point methods over symmetric cones*, SIAM Journal on Optimization (Accepted (2005)).

[SA03] S. Schmieta and F. Alizadeh, *Extension of Commutative Class of Interior Point Algorithms to Symmetric Cones*, Mathematical Programming **96** (2003), no. 3, 409–438.

[Sho85] N. Shor, *Minimization Methods for Non-Differentiable Functions*, Springer Series in Computational Mathematics, vol. 3, Springer, Berlin, 1985.

[SS04] D. Sun and J. Sun, *Löwner's Operator and Spectral Functions in Euclidean Jordan Algebras*, Tech. report, Department of Mathematics, Univertsity of Singapore, 2004.