Primal-dual potential reduction algorithm for symmetric programming problems with nonlinear objective funcions

Leonid Faybusovich *

June,2016

Abstract

We consider a primal-dual potential reduction algorithm for nonlinear convex optimization problems over symmetric cones. The same comlexity estimates as in the case of the linear objective function are obtained provided a certain nonlinear system of equations can be solved with a given accuracy. This generalizes the result of K. Kortanek,F.Potra and Y.Ye [7]. We further introduce a generalized Nesterov-Todd direction and show how it can be used to achieve a required accuracy (by solving the linearization of above mentioned nonlinear system) for a class of nonlinear convex functions satisfying scaling Lipschitz condition. This result is a far-reaching generalization of results of F. Potra, Y. Ye and J. Zhu [8],[9]. Finally, we show that a class of functions (which contains quantum entropy function) satisfies scaling Lipschitz condition.

Key words: quantum entropy, convex objective functions, optimization over symmetric cones

1 Introduction

A substantial amount of optimization problems arising in quantum information theory,quantum statistical physics ,information geometry , machine learning and other areas (for the impressive list of various applications see [1])requires dealing with the so-called quantum or von Neumann entropy which is the function of the form Tr(XLn(X)), where X is positive semi-definite complex Hermitian matrix. In [5] we introduced a class of functions (which contains the quantum entropy) which is compatible (in the sense of [10]) with the standard logarithmic barrier on the cone of invertible squares of a Euclidean Jordan algebra. This, in principle, allows one to use certain types of interior-point algorithms for solving optimization problems over symmetric cones involving quantum entropy.

In present paper we develop a primal-dual potential reduction algorithm for a class of nonlinear convex objective functions (which contains the quantum entropy) which provides an efficient way of dealing with nonlinear convex optimization problems over symmetric cones. In particular, the complexity estimates for the algorithm are (qualitatively) the same as in the case of linear objective functions [3]. Our results are far-reaching generalizations of those of K. Kortanek,F. Potra, Y. Ye and J. Zhu.

The plan of the paper is as follows.In section 2 we briefly recall relevant notations and results from the theory of Euclidean Jordan algebras (following [2]). In section 3 we formulate a version of a nonlinear complementarity problem based on the concept of a monotone manifold and describe a conceptual primal-dual algorithm for (approximately) solving it. In section 4 we develop a potential-reduction algorithm for a general nonlinear convex objective function for optimization problems over

*University of Notre Dame, Department of Mathematics (Email: leonid.faybusovich.1@nd.edu) This author is supported in part by Simmons Foundation Grant 275013.

symmetric cones. In section 5 a (generalized) primal-dual Nesterov-Todd direction is introduced. It is shown how to specify the potential reduction algorithm for a class of nonlinear convex objective functions satisfying the so-called scaled Lipschitz condition [9]. This section heavily relies on [3]. Finally, in section 6 we show that the class of functions introduced in [5] satisfies the scaling Lipschitz condition.

2 Jordan-algebraic Concepts

We adhere to the notation of an excellent book [2].

Let $\mathbf{F}$ be the field $\mathbf{R}$ or $\mathbf{C}$. A vector space V over $\mathbf{F}$ is called an algebra over $\mathbf{F}$ if a bilinear mapping $(x, y) \to xy$ from $V \times V$ into V is defined. For an element x in V let $L(x) : V \to V$ be the linear map such that

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

An algebra V over $\mathbf{F}$ is a Jordan algebra if

$$xy = yx, x(x^2 y) = x^2(xy), \forall x, y \in V.$$

In other words, Jordan algebra is always commutative but typically not associative. In an algebra V one defines x^n recursively by $x^n = x \cdot x^{n-1}$. An algebra V is said to be power associative if $x^p \cdot x^q = x^{p+q}$ for any $x \in V$ and integers p, q.

Proposition 2.1 *A Jordan algebra is power associative. Besides,*

$$[L(x^p), L(x^q)] = 0, \forall x \in V,$$

and any positive integers p and q. (In other words,the corresponding linear operators commute).

This is Proposition II.1.2 in [2]. We will always assume that the Jordan algebra has an identity element e (i.e. , $xe = x, \forall x \in V$).

Let V be a finite-dimensional power associative algebra over $\mathbf{F}$ with an identity element e, and let $F[X]$ denote the algebra over $\mathbf{F}$ of polynomials in one variable with coefficients in $\mathbf{F}$. For $x \in V$ we define

$$\mathbf{F}[x] = \{p(x) : p \in \mathbf{F}[X]\}.$$

A nonzero polynomial $p \in \mathbf{F}[X]$ of minimal possible degree such that $p(x) = 0$ is called the minimal polynomial of x. Given $x \in V$, let $m(x)$ be the degree of the minimal polynomial of x. We define the rank of V as

$$r = \max\{m(x) : x \in V\}.$$

An element x is called regular if $m(x) = r$.

Proposition 2.2 *The set of regular elements is open and dense in V.There exist polynomials $a_1, \ldots a_r$ on V such that the minimal polynomial of every regular element x is given by*

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

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

This is Proposition II.2.1 in [2]. The coefficient $a_1(x)$ is called the trace of x and is denoted $tr(x)$ (in particular, trace is linear). The coefficient $a_r(x)$ is called the determinant of x and is denoted $\det(x)$. An element x is said to be invertible if there exists an element $y \in \mathbf{F}[x]$ such that $xy = e$. The set $\lambda \in \mathbf{F}$ such that $x - \lambda e$ is not invertible is called the spectrum of x and is denoted $spec(x)$.

Given $x \in V$, we define

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

The map P is called the quadratic representation of V. We denote $DP(x)y$ by $2P(x, y)$. Here $DP(x)y$ is the Fréchet derivative of the map P at point $x \in V$ evaluated on $y \in V$. It is easy to see that

$$P(x, y) = L(x)L(y) + L(y)L(x) - L(xy), x, y \in V.$$

Proposition 2.3 *Let V be a finite-dimensional Jordan algebra over $\mathbf{F}$. An element $x \in V$ is invertible if and only if $P(x)$ is invertible. In this case*

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

This is Proposition II.3.1 in [2].

Proposition 2.4 *Let $\mathcal{J}$ be the (open) set of invertible elements in V. The map $x \to x^{-1} : \mathcal{J} \to \mathcal{J}$ is Fréchet differentiable and*

i)$D(x^{-1})u = -P(x^{-1})u, x \in \mathcal{J}, u \in V.$

ii) If x and y are invertible, then $P(x)y$ is invertible and $(P(x)y)^{-1} = P(x^{-1})y^{-1}$.

iii)

$$P(P(x)y) = P(x)P(y)P(x), \forall x, y \in V.$$

iv)

$$P(P(x)y, P(x)z) = P(x)P(y, z)P(x), \forall x, y, z \in V.$$

This is Proposition II.3.3 in [2]. A bilinear form β on V is called associative if

$$\beta(xy, z) = \beta(x, yz), \forall x, y, z \in V.$$

Proposition 2.5 *The symmetric bilinear forms* $\mathrm{Tr(L(xy))}$ *and* $\mathrm{tr(xy)}$ *are associative.*

This is Proposition II.4.3 in [2].

In case , where $\mathbf{F} = \mathbf{R}$ we consider an important class of Euclidean Jordan algebras. A Jordan algebra V over $\mathbf{R}$ is called Euclidean if $\mathrm{tr}(x^2) > 0, \forall x \in V \setminus \{0\}$. An element $c \in V$ is called idempotent if $c^2 = c$. Two idempotents are orthogonal if $cd = 0$. A system of idempotents $c_1, \ldots, c_k$ is a complete system of orthogonal idempotents If $c_i^2 = c_i, c_i c_j = 0, i \neq j$, and $c_1 + \ldots + c_k = e$.

Theorem 2.6 *Let V be an Euclidean Jordan algebra. Given $x \in V$,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*

$$x = \lambda_1 c_1 + \ldots + \lambda_k c_k.$$

In this case $spec(x) = \{\lambda_1, \ldots, \lambda_k\}, c_1, \ldots, c_k \in \mathbf{R}[x]$.

This is Theorem III.1.1 in [2].

An idempotent is primitive if it is non-zero and cannot be written as a sum of two non-zero idempotents. We say that $c_1, \ldots, c_m$ is a complete system of orthogonal primitive idempotents, or Jordan frame, if each c_j is primitive idempotent and if

$$c_j c_k = 0, j \neq k, c_1 + \ldots + c_m = e.$$

Note that in this case $m = r$ (rank of V).

Theorem 2.7 *Suppose V has rank r. Then for $x \in V$ 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 multiplicities) are uniquely determined by x. Furthermore,

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

This is Theorem III.1.2 in [2].

Given a function f which is defined at least on $spec(x)$, we can define

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

if $x = \sum_{i=1}^{r} \lambda_i c_i$. In particular,

$$\exp(x) = \sum_{i=1}^{r} \exp(\lambda_i) c_i, \ln x = \sum_{i=1}^{r} \ln \lambda_i c_i, \lambda_i > 0.$$

Let

$$\bar{\Omega} = \{x^2 : x \in V\}.$$

Theorem 2.8 *Let V be an Euclidean Jordan algebra.The interior Ω of $\bar{\Omega}$ is a symmetric (i.e. , self-dual, homogeneous) convex cone. Furthermore,Ω is the connected component of e in the set $\mathcal{J}$ of invertible elements, and also Ω is the set of elements x in V for which $L(x)$ is positive definite. In particular, the group of linear automorphisms $GL(\Omega)$ of Ω acts transitively on it. Moreover, $P(x) \in GL(\Omega)$ for any invertible x.*

This is Proposition III.2.2 in [2].

Let $c_1, \dots c_k$ be complete system of orthogonal idempotents. For each idempotent c, denote $V(c,0), V(c,1), V(c,1/2)$ the eigenspaces of $L(c)$ corresponding to eigenvalues $0, 1, 1/2$, respectively. Then $L(c_1), \dots, L(c_k)$ pairwise commute and

$$V = \bigoplus_{1 \le i \le j} V_{ij},$$

where $V_{ii} = V(c_i, 1), V_{ij} = V(c_i, 1/2) \cap V(c_j, 1/2)$. Such a decomposition of V corresponding to a complete system of orthogonal idempotents is called the Peirce decomposition. It is studied in detail in Section 1 of Chapter IV in [2]. A typical example of a Jordan algebra over a field $\mathbf{F}$ is the vector space of symmetric matrices over $\mathbf{F}$ with multiplication operation

$$A \cdot B = \frac{AB + BA}{2},$$

where on the right we have a usual matrix multiplication. In case when $\mathbf{F} = \mathbf{R}$ we get an exmple of an Euclidean Jordan algebra.

3 Nonlinear monotone complementarity problem

Let V be a Euclidean Jordan and Ω be the cone of invertible squares in V. Let M be a smooth submanifold in $V \times V$ such that $\dim M = \dim V$ and the following property holds:

$$(\xi, \eta) \in T_{(x,y)}(M) \Rightarrow \langle \xi, \eta \rangle \geq 0 \tag{1}$$

for any $(x, y) \in M$. Here $T_{(x,y)}(M)$ is the tangent space to M at a point $(x, y) \in M$. Note that $\langle \xi, \eta \rangle = tr(\xi\eta)$ stands for the canonical scalar product in V (see [2],p. 29). The nonlinear monotone complementarity problem is formulated as follows. Find

$$(x, y) \in M \cap (\bar{\Omega} \times \bar{\Omega}) \tag{2}$$

such that

$$\langle x, y \rangle = 0. \tag{3}$$

Here $\bar{\Omega}$ stands for the closure of Ω in V.

Let φ be a continuous convex function on $\bar{\Omega}$ which is smooth on Ω. Consider the following optimization problem:

$$\varphi(x) \to \min, x \in \mathcal{F} = (a + X) \cap \bar{\Omega}, \tag{4}$$

where X is a vector subspace in V , $a \in V$. We can associate with (4) the following submanifold in $V \times V$:

$$M = \{(x, \nabla\varphi(x) + \eta) : x \in (a + X) \cap \Omega, \eta \in X^{\perp}\}. \tag{5}$$

Here

$$X^{\perp} = \{\eta \in V : \langle \xi, \eta \rangle = 0, \forall \xi \in X\}$$

is the orthogonal complement to X in V, and $\nabla\varphi(x)$ is the gradient of $\varphi(x)$ at x, i.e.

$$D\varphi(x)h = \langle \nabla\varphi(x), h \rangle, \forall h \in V,$$

and $D\varphi(x)$ is the Fréchet derivative of φ at x. If $\mathcal{F} \neq \emptyset$, then $\dim M = \dim V$ and

$$T_{(x,y)}(M) = \{(\xi, H_{\varphi}(x)\xi + \zeta) : \xi \in X, \zeta \in X^{\perp}\}. \tag{6}$$

Here $H_{\varphi}(x)$ is the Hessian of φ at x,i.e.

$$D^2\varphi(x)(h_1, h_2) = \langle h_1, H_{\varphi}(x)h_2 \rangle, \forall h_1, h_2 \in V,$$

and $D^2\varphi(x)$ is the second Fréchet of φ at $x \in \Omega$. Since φ is convex,

$$\langle \xi, H_{\varphi}(x)\xi + \zeta \rangle = \langle \xi, H_{\varphi}(x)\xi \rangle \geq 0,$$

for $\xi \in X, \zeta \in X^{\perp}$, i.e. M is a monotone manifold. Let $(x, y) \in M \cap (\Omega \times \Omega)$ be such that

$$\langle x, y \rangle = 0.$$

Then x is an optimal solution to (4) (see e.g. [4]). Moreover, let $(x, y) \in M \cap (\Omega \times \Omega)$ and $\langle x, y \rangle \leq \epsilon$ for some $\epsilon > 0$. Then

$$\varphi(x) \leq \varphi(x^*) + \epsilon,$$

for an optimal solution x^* to the problem (4). Our main goal is to find a suboptimal solution to (4). Therefore, we will concentrate on monotone manifolds of the type (5).

Consider the potential function $\Phi_\rho : \Omega \times \Omega \to \mathbf{R}$:

$$\Phi_\rho(x, y) = \rho \ln\langle x, y \rangle - \ln \det x - \ln \det y. \tag{7}$$

Here $\rho \geq r$ is a real parameter and r is the rank of the Jordan algebra V (see [2], p. 28).

Lemma 3.1 *For any* $(x, y) \in \bar{\Omega} \times \bar{\Omega}$ *we have:*

$$\Phi_r(x, y) \geq r \ln r.$$

For a proof see [3], Lemma 2.3.

Lemma 3.2 *For any* $(x, y) \in \Omega \times \Omega$ *and* $\rho > r$*, we have:*

$$\langle x, y \rangle \leq \exp(\frac{\Phi_\rho(x, y)}{\rho - r}).$$

For a proof see [3],Lemma 2.4.

Proposition 3.3 *Suppose that we have a sequence* $(x_k, y_k) \in M \cap (\Omega \times \Omega), k = 0, 1, 2, \ldots$ *with the property*

$$\Phi_\rho(x_{k+1}, y_{k+1}) \leq \Phi_\rho(x_k, y_k) - \delta, \tag{8}$$

$k = 0, 1, \ldots,$ *for some fixed* $\delta > 0$ *and a fixed* $\rho > r$*. Then , given* $1 > \epsilon > 0$*, we have:*

$$\langle x_k, y_k \rangle \leq \epsilon$$

for

$$k \geq \frac{\rho - r}{\delta}(\frac{\Phi_\rho(x_0, y_0)}{\rho - r} + \ln(\frac{1}{\epsilon})).$$

Thus, to find a suboptimal solution to the problem (4) (with an arbitrary accuracy) ,it suffices to find a sequence with properties described in Proposition 3.3.

4 A primal-dual potential reduction algorithm

Given $x \in V$, we denote

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

Let

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

be the spectral decomposition of x in V (see [2],p.44). We denote

$$\|x\|_\infty = \max\{|\lambda_i| : i \in [1, r]\}.$$

Note that

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

In particular, $\|x\|_\infty \leq \|x\|$.

Recall that $P(x)$ is the quadratic representation of V (see [FK], p.32). Given $(x, y) \in \Omega$, let $(\xi, t) \in V \times V$ be such that

$$\tau = \max\{\|P(x)^{-\frac{1}{2}}\xi\|_\infty, \|P(y)^{-\frac{1}{2}}t\|_\infty\} < 1.$$

We have:

Lemma 4.1 *The pair* $(x + \xi, y + t) \in \Omega \times \Omega$ *and*

$$\Phi_\rho(x + \xi, y + t) - \Phi_\rho(x, y) \leq \Delta_1 + \Delta_2,$$

where

$$\Delta_1 = \rho(\frac{\langle \xi, y\rangle + \langle x, t\rangle}{\langle x, y\rangle}) - \langle x^{-1}, \xi\rangle - \langle y^{-1}, t\rangle, \tag{9}$$

$$\Delta_2 = \rho\frac{\langle \xi, t\rangle}{\langle x, y\rangle} + \frac{\|P(x)^{-\frac{1}{2}}\xi\|^2 + \|P(y)^{-\frac{1}{2}}t\|^2}{2(1-\tau)}. \tag{10}$$

For a proof see [3], p. 122. We will consider a specific descent direction for Φ_ρ.

Lemma 4.2 *Given* $(x, y) \in \Omega \times \Omega$, *there exists a unique* $z \in \Omega$ *such that*

$$y = P(z)^{-1}x. \tag{11}$$

For a proof see Lemma 3.2 in [3]. Introduce a Riemannian metric g on $\Omega \times \Omega$ as follows:

$$g(x, y)(\begin{bmatrix} \xi_1 \\ \eta_1 \end{bmatrix}, \begin{bmatrix} \xi_2 \\ \eta_2 \end{bmatrix}) = \langle \begin{bmatrix} \xi_1 \\ \eta_1 \end{bmatrix}, G(x, y) \begin{bmatrix} \xi_2 \\ \eta_2 \end{bmatrix}\rangle,$$

$(\xi_i, \eta_i) \in V \times V, i = 1, 2,$where

$$G(x, y) = \begin{bmatrix} P(z)^{-1} & I \\ I & P(z) \end{bmatrix},$$

$I : V \to V$ is the identity map. We can now introduce the generalized Nesterov-Todd direction (see [11]) as the negative gradient of the potential function Φ_ρ restricted to the submanifold $M \cap (\Omega \times \Omega)$ with respect to the metric g restricted to M. More precisely, the Nesterov-Todd direction $(\xi, \zeta) \in T_{(x,y)}(M)$ at $(x, y) \in M \cap (\Omega \times \Omega)$ is characterized as:

$$G(x, y) \begin{bmatrix} \xi \\ \zeta \end{bmatrix} = -\nabla\Phi_\rho(x, y) = \begin{bmatrix} \frac{-\rho y}{\langle x,y \rangle} + x^{-1} \\ -\frac{\rho x}{\langle x,y \rangle} + y^{-1} \end{bmatrix},$$

$$(\xi, \zeta) \in T_{(x,y)}(M).$$

The latter condition in light of (6) takes the form:

$$\zeta = H_\varphi(x)\xi + \eta, \xi \in X, \eta \in X^\perp.$$

Hence, we arrive at the following system of linear equations:

$$P(z)^{-1}\xi + H_\varphi(x)\xi + \eta = -\frac{\rho y}{\langle x, y\rangle} + x^{-1}, \tag{12}$$

$$\xi + P(z)H_\varphi(x)\xi + P(z)\eta = -\frac{\rho x}{\langle x, y\rangle} + y^{-1}, \tag{13}$$

$$\xi \in X, \eta \in X^\perp. \tag{14}$$

Note that (12) and (13) are equivalent (multiple (12) by $P(z)$ and use (11)).

Remark 4.3 *In case where φ is a convex linear-quadratic function, we arrive at the standard Nesterov-Todd direction considered in [3].* ■

In case of a general convex function φ we cannot move along the Nesterov-Todd direction as defined in (12),(13), and stay on the manifold M simaltaneously. Therefore, it is natural to consider the following iteration of the algorithm:

$$x \to x + \xi, y \to y + t, t = \nabla\varphi(x + \xi) - \nabla\varphi(x) + \eta, \tag{15}$$

Where $(\xi, \eta) \in X \times X^\perp$ satisfy (12). Alternatively, we can consider the following (nonlinear) system of equations with respect to (ξ, η).

$$P(z)^{-1}\xi + t = -\frac{\rho y}{\langle x, y\rangle} + x^{-1}, \tag{16}$$

$$t = \nabla\varphi(x + \xi) - \nabla\varphi(x) + \eta, \tag{17}$$

$$(\xi, \eta) \in X \times X^\perp. \tag{18}$$

It turns out that if we can solve (16)- (18) with a sufficiently small error in (16),(17), we can achieve a required decrease in (8).

Recalling (11), let

$$v = P(z)^{1/2}y = P(z)^{-1/2}x \in \Omega. \tag{19}$$

Consider a spectral decomposition

$$v = \sum_{i=1}^{r} v_i e_i, \tag{20}$$

and let

$$v_{\min} = \min\{v_i : i \in [1, r]\} > 0. \tag{21}$$

Let $(\xi, t) \in V \times V, \langle \xi, t \rangle \geq 0$ and

$$P(z)^{-1}\xi + t = \theta(-\frac{\rho y}{\langle x, y \rangle} + x^{-1}) + res. \tag{22}$$

Here $\theta > 0$ is a scaling parameter and $res \in V$ is an "error". Multiplying (22) by $P(z)^{1/2}$, we can rewrite it in the following equivalent form:

$$P(z)^{-1/2}\xi + P(z)^{1/2}t = \theta\Gamma(v) + P(z)^{1/2}res, \tag{23}$$

where

$$\Gamma(v) = -\frac{\rho v}{\langle v, v \rangle} + v^{-1}. \tag{24}$$

The following theorem is a generalization of Theorem 3.2 in [7], where the case $\Omega = \mathbf{R}^{\mathbf{n}}_{+}$ was considered.

Theorem 4.4 *Let $(\xi, t) \in V \times V$ be such that $\langle \xi, t \rangle \geq 0$ and (23) holds true, where*

$$\theta = \frac{\beta}{2}\frac{v_{\min}}{\|\Gamma(v)\|}, \|P(z)^{1/2}res\| \leq \frac{C\beta^2 v_{\min}}{2}, \tag{25}$$

and $0 < \beta < 1, C \geq 0$ are such that

$$1 - \beta C > 0, \beta(1 + \beta C) < 1. \tag{26}$$

Then

$$\Phi_\rho(x + \xi, y + t) - \Phi_\rho(x, y) \leq -\frac{\sqrt{3}}{4}\beta(1 - \beta C) + \frac{\beta^2}{4}(1 + \beta C)^2, \tag{27}$$

provided $2r \geq \rho \geq r + \sqrt{r}$. Moreover,

$$\max\{\|P(x)^{-1/2}\xi\|, \|P(y)^{-1/2}t\|\} \leq \frac{1}{2}. \tag{28}$$

In particular, $(x + \xi, y + t) \in \Omega \times \Omega$.

Remark 4.5 *Note that for a fixed $C \geq 0$, conditions (26) are satisfied and the right-hand side in (27) is negative for sufficiently small $\beta > 0$.* ∎

To prove Theorem 4.4 we need the following Lemma.

Lemma 4.6 *For $\rho \geq r + \sqrt{r}, (x, y) \in \Omega \times \Omega$, we have:*

$$\|\Gamma(v)\| v_{\min} \geq \frac{\sqrt{3}}{2}. \tag{29}$$

For a proof see Proposition 3.5 in [3]. We are in position to prove Theorem 4.4

Proof. First of all, by (23), we have:

$$\|P(z)^{-1/2}\xi + P(z)^{1/2}t\|^2 = \theta^2\|\Gamma(v)\|^2 + 2\theta\langle\Gamma(v), P(z)^{1/2}res\rangle + \langle P(z)res, res\rangle.$$

Consequently,

$$\langle P(z)^{-1}\xi, \xi\rangle + \langle P(z)t, t\rangle + 2\langle\xi, t\rangle \leq$$
$$\theta^2\|\Gamma(v)\|^2 + 2\theta\|\Gamma(v)\|\|P(z)^{1/2}res\| + \|P(z)^{1/2}res\|^2 =$$
$$(\theta\|\Gamma(v)\| + \|P(z)^{1/2}res\|)^2.$$

Since $\langle\xi, t\rangle \geq 0$ by the assumption, we obtain:

$$\langle P(z)^{-1}\xi, \xi\rangle + \langle P(z)t, t\rangle \leq (\theta\|\Gamma(v)\| + \|P(z)^{1/2}res\|)^2. \tag{30}$$

Note that

$$P(v) \geq v_{\min}^2 I. \tag{31}$$

See e.g. the reasoning on p. 125 in [3]. By (19):

$$P(v) = P(z)^{1/2}P(y)P(z)^{1/2} = P(z)^{-1/2}P(x)P(z)^{-1/2}. \tag{32}$$

Consequently, using (31):

$$P(y) = P(z)^{-1/2}P(v)P(z)^{-1/2} \geq v_{\min}^2 P(z)^{-1}, \tag{33}$$

$$P(x) = P(z)^{1/2}P(v)P(z)^{1/2} \geq v_{\min}^2 P(z), \tag{34}$$

where we used (31). Hence,

$$P(y)^{-1} \leq \frac{P(z)}{v_{\min}^2}, P(x)^{-1} \leq \frac{P(z)^{-1}}{v_{\min}^2}. \tag{35}$$

By (32),(27):

$$\langle P(x)^{-1}\xi, \xi\rangle + \langle P(y)^{-1}t, t\rangle \leq \frac{\langle P(z)^{-1}\xi, \xi\rangle + \langle P(z)t, t\rangle}{v_{\min}^2} \leq \tag{36}$$
$$(\frac{\theta\|\Gamma(v)\| + \|P(z)^{1/2}res\|}{v_{\min}})^2.$$

Using (25), we obtain from (36):

$$\|P(x)^{-1/2}\xi\|^2 + \|P(y)^{-1/2}t\|^2 \leq (\frac{\beta}{2} + \frac{C\beta^2}{2})^2. \tag{37}$$

By (37):

$$\max\{\|P(y)^{-1/2}t\|, \|P(x)^{-1/2}\xi\|\} \leq \frac{\beta(1+C\beta)}{2} \leq \frac{1}{2}, \tag{38}$$

where in the last inequality we used (26). We are now in position to use Lemma 4.1: We have (see (9):

$$\Delta_1 = \langle \frac{\rho y}{\langle x,y\rangle} - x^{-1}, \xi\rangle + \langle \frac{\rho x}{\langle x,y\rangle} - y^{-1}, t\rangle =$$
$$-\langle \Gamma(v), P(z)^{-1/2}\xi\rangle - \langle \Gamma(v), P(z)^{1/2}t\rangle =$$
$$-\langle \Gamma(v), \theta\Gamma(v) + P(z)^{1/2}res\rangle,$$

where $\Gamma(v)$ is defined in (24) and in the last equality we used (23). Hence,

$$\Delta_1 = -\theta\|\Gamma(v)\|^2 - \langle \Gamma(v), P(z)^{1/2}res\rangle \leq$$
$$-\theta\|\Gamma(v)\|^2 + \|\Gamma(v)\|\|P(z)^{1/2}res\| \leq$$
$$-\frac{\beta}{2}\|\Gamma(v)\|v_{\min} + \frac{C\beta^2}{2}\|\Gamma(v)\|v_{\min} =$$
$$-\frac{\beta}{2}\|\Gamma(v)\|v_{\min}(1-\beta C) \leq$$
$$-\frac{\beta}{2}\frac{\sqrt{3}}{2}(1-\beta C),$$

where in the last inequality we used Lemma 4.6 and the assumption $1-\beta C > 0$. To estimate Δ_2 in (10), we can take $\tau = \frac{1}{2}$, because of (38) (note that $\|x\|_\infty \leq \|x\|, x \in V$). We have:

$$\langle x,y\rangle = \langle v,v\rangle \geq rv_{\min}^2.$$

Hence,

$$\rho\frac{\langle \xi,t\rangle}{\langle x,y\rangle} \leq \frac{2r\langle \xi,t\rangle}{rv_{\min}^2} = \frac{2\langle \xi,t\rangle}{v_{\min}^2},$$

where we used $\rho \leq 2r, \langle \xi,t\rangle \geq 0$. Consequently, using (35):

$$\Delta_2 \leq \frac{2\langle \xi,t\rangle}{v_{\min}^2} + \frac{\langle P(z)^{-1}\xi,\xi\rangle + \langle P(z)t,t\rangle}{2v_{\min}^2(1-\tau)} =$$
$$\frac{2\langle \xi,t\rangle + \langle P(z)^{-1}\xi,\xi\rangle + \langle P(z)t,t\rangle}{v_{\min}^2} =$$
$$\frac{\|P(z)^{-1/2}\xi + P(z)^{1/2}t\|^2}{v_{\min}^2} = \frac{\|\theta\Gamma(v) + P(z)^{1/2}res\|^2}{v_{\min}^2},$$

where in the last equality we used (23). Therefore,

$$\Delta_2 \leq (\frac{\beta}{2} + \frac{C\beta^2}{2})^2 = \frac{\beta^2}{4}(1+C\beta)^2.$$

Applying Lemma 4.1, we obtain the estimate (27). ■

Consider the following nonlinear system:

$$P(z)^{-1/2}\xi + P(z)^{1/2}t = \theta\Gamma(v), \tag{39}$$

$$t = \nabla\varphi(x+\xi) - \nabla(x) + \eta, \tag{40}$$

$$\xi \in X, \eta \in X^{\perp}. \tag{41}$$

Note that $\langle \xi, t\rangle = \langle \xi, \nabla\varphi(x+\xi) - \nabla\varphi(x)\rangle \geq 0$, since φ is convex. If we can solve (39)-(41) with sufficiently small error in (39), (40) as described in Theorem 4.4 we can run the algorithm as described in Proposition 3.3.

Algorithm.

Given $\epsilon > 0, (x_0, y_0 = \nabla\varphi(x_0) + \zeta_0) \in \Omega \times \Omega, \zeta_0 \in X^{\perp}$. Further, assume that $0 < \beta < 1, C \geq 0, 1 - \beta C > 0, \beta(1+\beta C) \leq 1, \delta = \frac{\sqrt{3}}{4}\beta(1-\beta C) - \beta^2/4(1+\beta C)^2 > 0$.

begin

if $\langle x_k, y_k\rangle < \epsilon$, then terminate; else

define

$$z_k, v_k, \theta_k, v_{\min}(k)$$

$$y_k = P(z_k)^{-1}x_k, v_k = P(z_k)^{1/2}y_k = P(z_k)^{-1/2}x_k, v_{\min}(k)$$

is defined as the minimal eigenvalue of v_k

$$\theta_k = \frac{\beta}{2}\frac{v_{\min}(k)}{\|\Gamma(v_k)\|}$$

Find $\xi_k \in X, \eta_k \in X^{\perp}$ such that

$$P(z_k)^{-1/2}\xi_k + P(z_k)^{1/2}\left(\nabla\varphi(x_k+\xi_k) - \nabla\varphi(x_k) + \eta_k\right) = \theta_k\Gamma(v_k) + P(z_k)^{1/2}res_k,$$

$$\|P(z_k)^{1/2}res_k\| \leq \frac{C\beta^2 v_{\min}(k)}{2}.$$

set $x_{k+1} = x_k + \xi_k, y_{k+1} = y_k + \left(\nabla\varphi(x_k+\xi_k) - \nabla\varphi(x_k) + \eta_k\right)$.

End.

Theorem 1 guarantees that $(x_k, y_k) \in (\Omega \times \Omega) \cap M$ and $\phi_\rho(x_{k+1}, y_{k+1}) - \phi_\rho(x_k, y_k) \leq -\delta$ for $k = 0, 1, 2, \cdots$ until termination, provided $2r \geq \rho \geq r + \sqrt{r}$.

5 Generalized Nesterov-Todd direction

To solve nonlinear system (39)-(41) it is quite natural to use the Newton method, i.e. we need to linearize the nonlinear term $\nabla\varphi(x+\xi) - \nabla\varphi(x) \approx H_\varphi(x)\xi$. Hence we arrive at the linear system of the type (12),(13), (14) whose solution coincides with (scaling) Nesterov-Todd

Proposition 5.1 *Let $(\tilde{\xi}, \tilde{\eta})$ be the solution to the linear system*

$$P(z)^{-1/2}\tilde{\xi} + P(z)^{1/2}(H_{\varphi}(x)\tilde{\xi} + \tilde{\eta}) = \theta\Gamma(v), \tag{42}$$

$$\tilde{\xi} \in X, \tilde{\eta} \in X^{\perp}. \tag{43}$$

Then

$$P(z)^{-1/2}\tilde{\xi} + P(z)^{1/2}(\nabla\varphi(x+\tilde{\xi}) - \nabla\varphi(x) + \tilde{\eta}) = \theta\Gamma(v) + P(z)^{1/2}Res,$$

where

$$Res = \nabla\varphi(x+\xi) - \nabla\varphi(x) - H_{\varphi}(x)\tilde{\xi}. \tag{44}$$

Proof. Simple computation. ■

Suppose that the function φ satisfies the following property. There exists a monotonically increasing function $\psi : (0,1) \to (0,+\infty)$ such that

$$\|P(x)^{1/2}\left(\nabla\varphi(x+\xi) - \nabla\varphi(x) - H_{\varphi}(x)\xi\right)\| \leq \psi(\alpha) < H_{\varphi}(x)\xi, \xi >, \tag{45}$$

provided that $\|P(x)^{-1/2}\xi\| \leq \alpha, x \in \Omega$

This condition (for the case $\Omega = \mathbf{R}^n_+$ under the name scaled Lipschitz condition) has been introduced in [9]

Proposition 5.2 *Let $(\tilde{\xi}, \tilde{\eta})$ be the solution to the linear system (42) -(43) then*

$$< H_{\varphi}(x)\tilde{\xi}, \tilde{\xi} > \leq \frac{\theta^2\|\Gamma(v)\|^2}{4}. \tag{46}$$

Proof. Taking the scalar product of both sides of (42) with $P(z)^{-1/2}\tilde{\xi}$ and using $< \tilde{\xi}, \tilde{\eta} >= 0$, we obtain

$$< P(z)^{-1}\tilde{\xi}, \tilde{\xi} > + < H_{\varphi}(x)\tilde{\xi}, \tilde{\xi} >= \theta < \Gamma(v), P(z)^{-1/2}\tilde{\xi} >$$

Hence

$$< H_{\varphi}(x)\tilde{\xi}, \tilde{\xi} >= \theta < \Gamma(v), P(z)^{-1/2}\tilde{\xi} > -\|P(z)^{-1/2}\tilde{\xi}\|^2$$

$$\leq \theta\|\Gamma(v)\|\|P(z)^{-1/2}\tilde{\xi}\| - \|P(z)^{-1/2}\tilde{\xi}\|^2$$

$$\leq \frac{\theta^2\|\Gamma(v)\|^2}{4}$$

■

Theorem 5.3 *Let* $(\tilde{\xi}, \tilde{\eta})$ *be the solution to the linear system (42) -(43),* φ *satisfy (45),* $\theta = \frac{\beta}{2} \frac{v_{\min}}{\|\Gamma(v)\|}$*,* $c = \frac{\psi(\frac{1}{2})}{8}, 2r \geq \rho \geq r + \sqrt{r}$ *and choose* $1 > \beta > 0$ *such that*

$1 - \beta C > 0, \beta(1 + \beta C) < 1, -\frac{\sqrt{3}}{4}\beta(1 - \beta C) + \frac{\beta^2}{4}(1 + \beta C)^2 < 0.$

Then

$$\|P(z)^{-1/2}\tilde{\xi}\| < \frac{1}{2}$$

and

$$P(z)^{-1/2}\tilde{\xi} + P(z)^{1/2}(\nabla\varphi(x + \tilde{\xi}) - \nabla\varphi(x) + \tilde{\eta}) = \theta\Gamma(v) + P(z)^{1/2}res \tag{47}$$

where $\|P(z)^{1/2}res\| \leq \frac{C\beta^2 v_{\min}}{2}, res = \nabla\varphi(x + \xi) - \nabla\varphi(x) - H_\varphi(x)\tilde{\xi}$.

Remark 5.4 *If all assumptions of Theorem 5.3 are satisfied, then according to Theorem 1, one can use* $\xi = \tilde{\xi}, t = \nabla\varphi(x + \tilde{\xi}) - \nabla\varphi(x) + \tilde{\eta}$ *as the descent direction for* Φ_ρ *.* ■

Proof. The same reasoning as in the beginning of the proof of Theorem 1shows that

$$\|P(x)^{-1/2}\tilde{\xi}\|^2 + \|P(y)^{-1/2}\tilde{\eta}\|^2 \leq \frac{\beta^2}{4}$$

Consequently,

$$\max\{\|P(x)^{-1/2}\tilde{\xi}\|, \|P(y)^{-1/2}\tilde{\eta}\|\} \leq \frac{\beta}{2} < \frac{1}{2}.$$

But then according to (45)

$$\|P(x)^{1/2}\left(\nabla\varphi(x + \tilde{\xi}) - \nabla\varphi(x) - H_\varphi(x)\tilde{\xi}\right)\| \leq \psi\left(\frac{1}{2}\right) < H_\varphi(x)\xi, \xi >,$$

$$\leq \psi(\frac{1}{2})\frac{\theta^2\|\Gamma(v)\|^2}{4} = \psi\left(\frac{1}{2}\right)\frac{\beta^2 v_{\min}^2}{16} \tag{48}$$

where in the last inequality we used (46). By (34)

$$P(z) \leq \frac{P(x)}{v_{\min}^2}.$$

Hence,

$$< P(z)res, res > \leq \frac{< P(x)res, res >}{v_{\min}^2} \leq \psi\left(\frac{1}{2}\right)^2 \frac{\beta^4 v_{\min}^2}{16^2},$$

where in the last inequality we used (48) and Proposition 5.1. Consequently,

$$\|P(z)^{1/2}res\| \leq \psi(\frac{1}{2})\frac{\beta^2 v_{\min}}{16} = \frac{C\beta^2 v_{\min}}{2}.$$

■

Remark 5.5 *Theorem 5.3 is a generalization of Theorem 1 in [8], where the case* $\Omega = \mathbf{R}^n_+$ *was considered.* ■

6 A class of functions that satisfy scaling Lipschitz condition

We consider a class of functions on Ω introduced in [5]. A function $f : [0, \infty) \to \mathbf{R}$ is called matrix monotone if for any square symmetric matrices A, B with real entries such that $A \geq B \geq 0$,we have: $f(A) \geq f(B)$. Let $g : [0, \infty) \to \mathbf{R}$ be continuous function such that its derivative is matrix monotone on $[0, \infty)$. Given a Euclidean Jordan algebra V, consider $\varphi : \Omega \to \mathbf{R}, \varphi(x) = tr(g(x)), x \in \Omega$. We denote these class of functions by M.

It was shown in [5] that φ is at least three times Fréchet differentiable on Ω and

$$D^2\varphi(x)(h_1, h_2) = \gamma tr(h_1, h_2) + \int_0^{+\infty} tr\left((P(x + se)^{-1/2}h_1)(P(x + se)^{-1/2}h_2)\right) s^2 d\nu(s), \tag{49}$$

$$D^3\varphi(x)(h_1, h_2, h_3) = -2\int_0^{+\infty} tr\left((P(x + se)^{-1/2}h_1)(P(x + se)^{-1/2}h_2)(P(x + se)^{-1/2}h_3)\right) s^2 d\nu(s) \tag{50}$$

where $\gamma \geq 0$ is some constant (depending on g), ν is a positive measure on $(0, \infty)$ such that

$$\int_0^{\infty} \frac{s}{1 + s} d\nu(s) < +\infty,$$

and $h_1, h_2, h_3 \in V, x \in \Omega; e$ is the identity element in V.

Proposition 6.1 *For $x \in \Omega, h_1, h_2 \in V$, we have:*

$$|D^3\varphi(x)(h_1, h_1, h_2)| \leq 2D^2\varphi(x)(h_1, h_1) < P(x)^{-1}h_2, h_2 >^{1/2} . \tag{51}$$

Proof. We can assume without loss of generality that $\gamma = 0$. By Cauchy-Schwarz inequality

$$\begin{aligned} w &= \left|tr\left((P(x + se)^{-1/2}h_1)^2 P(x + se)^{-1/2}h_2\right)\right| \\ &= \left|\left\langle(P(x + se)^{-1/2}h_1)^2, P(x + se)^{-1/2}h_2\right\rangle\right| \\ &\leq \left|tr\left((P(x + se)^{-1/2}h_1)^4\right)^{1/2} tr\left((P(x + se)^{-1/2}h_2)^2\right)^{1/2}\right| \end{aligned}$$

Using elementary inequality

$$[tr(u^4)]^{1/2} \leq tr(u^2), \forall u \in V,$$

we obtain

$$w \leq \left|tr\left(P(x + se)^{-1/2}h_1\right)^2) tr\left((P(x + se)^{-1/2}h_2)^2\right)^{1/2}\right|$$

Now

$$tr(P(x + se)^{-1/2}h_2)^2 =< P(x + se)^{-1/2}h_2, P(x + se)^{-1/2}h_2 >$$

$$=< h_2, P(x+se)^{-1}h_2 >$$

$$\leq< h_2, P(x)^{-1}h_2 >$$

for and $s \geq 0$. This follows from

$$P(x+se)^{-1} \leq P(x)^{-1}, \forall s \geq 0 \tag{52}$$

(see the proof of theorem 3.2 in [5]).

Hence

$$\left| tr \left(\left[P(x+se)^{-1/2}h_1 \right]^2 P(x+se)^{-1/2}h_2 \right) \right| \leq tr \left[\left(P(x+se)^{-1/2}h_1 \right)^2 \right] < h_2, P(x)^{-1}h_2 >^{-1/2},$$

$$\forall s \geq 0.$$

Comparing this with (49), (50), we obtain (51).

■

Remark 6.2 *Proposition 6.1 in particular implies that φ is compatible (in the sense of [10], p. 159) with self-concordant barrier $-\ln \det x$ on Ω [6]. That was established in [5], but it does not lead as such to effective potential-reduction algorithms.* ■

Lemma 6.3 *Let $x \in \Omega, u \in V, \|P(x)^{-1/2}u\| < 1$. Then*

$$P(x+u)^{-1} \leq \frac{1}{(1 - \|P(x)^{-1/2}u)\|)^2}. \tag{53}$$

Proof. Obviously $x + u \in \Omega$. Furthermore,

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

Hence

$$P(x+u)^{-1} = P(x)^{-1/2}P(e + P(x)^{-1/2}u)^{-1}P(x)^{-1/2} \tag{54}$$

Let

$$P(x)^{-1/2}u = \sum_{i=1}^{r} \lambda_1 e_i$$

be a spectral decomposition and

$$V = \bigoplus_{1 \leq i \leq j \leq r} V_{ij}$$

the Peirce decomposition of V corresponding to the Jordan frame $e_1, \cdots, e_r$. Given $h \in V$, Let

$$h = \sum_{i \leq j} h_{ij}, h_{ij} \in V.$$

Then

$$< h, P(e + P(x)^{-1/2}u)^{-1}h >= \sum_{i \le j} \frac{\|h_{ij}\|^2}{(1+\lambda_i)(1+\lambda_j)}.$$

Let $\rho = \|P(x)^{-1/2}u\|$. Note that $\rho < 1$ by the assumption. We have

$$1 + \lambda_i = |1 + \lambda_i| \ge 1 - |\lambda_i| \ge 1 - \rho > 0$$

$$i \in [0, r]$$

Hence,

$$< h, P(e + P(x)^{-1/2}u)^{-1}h > \le \frac{1}{(1-\rho)^2} \sum_{i \le j} \|h_{ij}\|^2 = \frac{\|h\|^2}{(1-\rho)^2}$$

Hence,

$$P(e + P(x)^{-1/2}u)^{-1} \le \frac{1}{(1-\rho)^2} I$$

Consequently, $P(x+u)^{-1} \le \frac{P(x)^{-1}}{(1-\rho)^2}$.

■

Theorem 6.4 *Let $\varphi \in \mathcal{M}, x \in \Omega, h \in V,$ and $\|P(x)^{-1/2}h\| \le \alpha < 1$. Then*

$$\|P(x)^{1/2} \left(\nabla\varphi(x+h) - \nabla\varphi(x) - H_\varphi(x)h\right)\| \le \frac{1}{(1-\alpha)^3} < H_\varphi(x)h, h >,$$

i.e M belongs to the class of scaled Lipschitz functions (see (45)).

Proof. Obviously, $x + h \in \Omega$. Let $\mu = \nabla\varphi(x+h) - \nabla\varphi(x) - H_\varphi(x)h$. Consider the function

$$a(s) =< \nabla\varphi(x+sh), P(x)\mu >, s \in [0, 1].$$

Then

$$a'(s) =< H_\varphi(x+sh)h, P(x)\mu >= D^2\varphi(x+sh)(h, P(x)\mu),$$

$$a''(s) = D^3\varphi(x+sh)(h, h, P(x)\mu).$$

Using the Taylor expansion of a around 0, we obtain

$$a(1) = a(0) + a'(0) + \frac{a''(\bar{s})}{2}, \text{for some } 0 < \bar{s} < 1.$$

Hence,

$$< \nabla\varphi(x+h) - \nabla\varphi(x) - H_\varphi(x)h, P(x)\mu >=< \mu, P(x)\mu >$$
$$= \frac{1}{2} D^3\varphi(x+\bar{s}h)(h, h, P(x)\mu).$$

Hence, by Proposition 6.1

$$< \mu, P(x)\mu > \leq D^2\varphi(x+\bar{s}h)(h,h) < P(x+\bar{s}h)^{-1}P(x)\mu, P(x)\mu >^{1/2} . \tag{55}$$

By Lemma 6.3

$$P(x+\bar{s}h)^{-1} \leq \frac{P(x)^{-1}}{(1-\bar{s}\|P(x)^{-1/2}h\|)^2} \leq \frac{P(x)^{-1}}{(1-\alpha)^2} \tag{56}$$

By (55), (56):

$$< \mu, P(x)\mu >^{1/2} \leq \frac{D^2\varphi(x+\bar{s}h)(h,h)}{(1-\alpha)}. \tag{57}$$

It remains to estimate the right-hand side of (57). We have by Lemma 6.3

$$< P(x+\bar{s}h+se)^{-1}h, h > \leq \frac{1}{(1-\bar{s}\|P(x+se)^{1/2}h\|)^2} < P(x+se)^{-1}h, h > . \tag{58}$$

Furthermore,

$$\|P(x+se)^{-1/2}h\|^2 = < P(x+se)^{-1}h, h > \leq < P(x)^{-1}h, h >$$

by (52). Consequently,

$$\bar{s}\|P(x+se)^{-1/2}h\| \leq \|P(x)^{-1/2}h\| \leq \alpha < 1.$$

Thus,

$$\frac{1}{(1-\bar{s}P(x+se)^{-1/2}h)^2} \leq \frac{1}{(1-\alpha)^2}.$$

Therefore

$$< P(x+\bar{s}h+se)^{-1}h, h > \leq \frac{1}{(1-\alpha)^2} < P(x+se)^{-1}h, h >, s \geq 0$$

Recalling (50), we obtain

$$D^2\varphi(x+\bar{s}h)(h,h) \leq \frac{1}{(1-\alpha)^2} D^2\varphi(x)(h,h).$$

Combining this with (57) we obtain the result. ∎

The most natural examples of functions in $\mathcal{M}$ are as follows.

$$g_1(\lambda) = \lambda \ln \lambda, \ \lambda \geq 0, g_2(\lambda) = \lambda^p, \ 1 \leq p \leq 2.$$

Indeed,

$$g_1'(\lambda) = \ln\lambda + 1, \ g_2'(\lambda) = p\lambda^{p-1}.$$

Both functions $\ln\lambda$ and $\lambda^q, 0 \leq q \leq 1$ are known to be matrix monotone. The corresponding functions

$$\varphi_1(x) = \mathrm{tr}(x\ln x),\ \varphi_2(x) = \mathrm{tr}(x^p),\ 1 \leq p \leq 2$$

belong to the class $\mathcal{M}$ and arise in various applications typically for the Euclidean Jordan algebra of complex Hermitian matrices; φ_1 is known as quantum or von Neumann Entropy and arise in numerous applications ([1]). Tha function φ_2 appears in compressed sensing (and is known as nuclear norm for $p = 1$).

7 Concluding remarks

In present paper we analyzed a class of potential reduction primal-dual algorithms for convex optimization problems over symmetric cones. Based on generalized Nesterov-Todd direction, we developed far-reaching generalizations of the results of K. Kortanek, F. Potra, Y.Ye and J. Zhu. Obtained complexity estimates (in case of functions satisfying scaled Lipschitz condition) are (qualitatively) the same as in the case of linear-quadratic objective functions.

References

[1] V. Chandrasecaran and P. Shah: Conic geometric programming. Manuscript, arXiv:math/1310.099v2, October,2013.

[2] J. Faraut and A. Korányi: *Analysis on Symmetric Cones*, Cambridge University Press, Cambridge, UK, 1994.

[3] L. Faybusovich:*A Jordan-algebraic approach to potential-reduction algorithms. Mathematitsche Zeitschrift*, Vol. 239, 117-129,2002.

[4] L. Faybusovich and Y. Lu: *Jordan-algebraic aspects of nonconvex optimization over symmetric cones. : Appl. Math. Optim.*, Vol. 53,no 1, 67-77,2006.

[5] L. Faybusovich and T. Tsuchiya:*Matrix monotonicity and self-concordane:how to handle quantum entropy in optimization.* Preprint, August, 2014.

[6] L. Faybusovich: Jordan algebras and interior-point algorithms. *Positivity*, Vol. 1 (1997), pp. 331–357.

[7] K. Kortanek, F. Potra and Y. Ye: *: On some efficient interior point methods for nonlinear convex programming. Linear algebra and its applications*,Vol. 152, 169-189,1991.

[8] F. Potra and Y. Ye:*A quadratically convergent polynomial algorithm for solving entropy optimization problem. Siam J. Optimization*, Vol. 3, no 4,843-860, 1993.

[9] J. Zhu:*A path following algorithm for a class of convex programming problems. Methods and Models of Operations Research*, Vol. 36,359-377,1992.

[10] Yu. Nesterov and A. Nemirovskii: *Interior-Point Polynomial Algorithms in Convex Programming*. SIAM, Philadelphia, 1994.

[11] Y. Nesterov and M. Todd:*Self-scaled barriers and interior-point methods for the convex programmingMath. Oper. Res.*, Vol. 22, no.1, 1-42, 1997.