A Proximal Point Algorithm in Semidefinite Programming

Ronaldo Gregório and Paulo Roberto Oliveira[1]
poliveir@cos.ufrj.br, rgregor@cos.ufrj.br

May 27, 2005

**Abstract**

In this work, we propose a new proximal algorithm for Semidefinite Programming ($SDP$). It appears to be the first proximal method in the class that converts Semidefinite Programming in Nonlinear Programming. It is implementable, and each main iteration of the conceptual proximal point algorithm is replaced by a sequence of nonlinear programming problems in the positive octant of the real vector space $R^n$. We are motivated by results of the classical proximal algorithm extended to Riemannian manifolds with non positive sectional curvature. An important example of such manifold is the space of symmetric definite matrices, where the metrics is given by the Hessian of the standard barrier function $-\ln\det(X)$. Then, observing the obvious fact that proximal algorithms do not depend on the geodesics, we apply those ideas to develop a proximal point algorithm for ($SDP$).

# 1 Introduction

Consider the problem:

$$\{\min \;\; f(X); \;\; \text{s.a} \;\; X \succeq 0 \}, \tag{1}$$

where $f : S^n_+ \longrightarrow R$ is a $C^1$ geodesic convex function[2], with finite minimum in $S^n_{++}$. Let $d(X,Y)$ be the Riemannian distance between $X, Y \in S^n_{++}$ [23], given by

$$d(X,Y) = \left[\sum_{i=1}^{n} \ln^2\lambda_i \left(X^{-\frac{1}{2}} Y X^{-\frac{1}{2}}\right)\right]^{\frac{1}{2}} = \left\|\ln\left(X^{-\frac{1}{2}} Y X^{-\frac{1}{2}}\right)\right\| = \tag{2}$$

$$= \left(\text{Tr}\, \ln^2 X^{-\frac{1}{2}} Y X^{-\frac{1}{2}}\right)^{\frac{1}{2}}.$$

For each $X \in S^n_{++}$ define a function $\rho_X : S^n_{++} \longrightarrow R$ by

[1] PESC, COPPE-UFRJ.

[2] A function $f : M \longrightarrow R$ defined in a Riemannian manifold $M$ is said geodesic convex if, and only if the composition $f \circ \gamma$ is convex, for all geodesics $\gamma \in M$. We also have that $S^n_{++}$ with the metric given by the Hessian of the barrier $-\ln\det(X)$, is an example of Riemannian manifold with non positive sectional curvature [23].

$$\rho_X(Y) = \tfrac{1}{2}d^2(X,Y).$$

Given $\beta > 0$, the Moreau-Yosida regularization $f_\beta : S^n_{++} \longrightarrow R$ of $f$ can be written as

$$f_\beta(X) = \min_{Y \succ 0}\{f(Y) + \beta\rho_X(Y)\}. \tag{3}$$

$X_\beta = \arg\ f_\beta$ is the proximal point of $X$ with respect to $\beta, f$ and $\rho_X$. The regularization $\rho_X$ is geodesic strict convex and the objective function in the definition of $f_\beta$ is 1-coercive [6], so $X_\beta$ is well posed, i.e., there is a unique proximal point and its characterization is given by

$$\beta\nabla\rho_X(X_\beta) = -\nabla f(X_\beta). \tag{4}$$

When $f$ is not differentiable, the characterization writes

$$\beta\nabla\rho_X(X_\beta) \in -\partial f(f_\beta(X)),$$

where $\partial f(Y)$ is the *subdifferential* of $f$ at $Y$.

The Proximal point method in $S^n_{++}$ generates, for a initial point $X^0 \in S^n_{++}$, a minimizer sequence $(X^k)$ defined by

$$X^{k+1} = \arg\ f_{\beta^k}(X^k),$$

where

$$\sum_{k=0}^{\infty} \frac{1}{\beta^k} \longrightarrow \infty. \tag{5}$$

Applying the results established in [6], the global convergence of the proximal point algorithm is assured for geodesic convex functions relatively to the metric generated by the Hessian of de barrier $-ln\ det(X)$.

## 2 An algorithm in $R^n_{++}$

Here we work with an important property of the distance function. When we apply a nonsingular linear transformation on the arguments, that function belongs invariant. More details can be found in [21].

From the previous section, we know that the objective function in each iteration of the proximal method is given by

$$g_{\beta^k}(Y) = f(Y) + \frac{\beta^k}{2}\sum_{i=1}^{n} \ln^2\lambda_i\left((X^k)^{-\frac{1}{2}}Y(X^k)^{-\frac{1}{2}}\right). \tag{6}$$

Now, supposing fixed the iterate $X^k$, define the transformation $T_{X^k} : S^n_{++} \longrightarrow S^n_{++}$ by

$$T_{X^k}(Y) = (X^k)^{-\frac{1}{2}} Y (X^k)^{-\frac{1}{2}}.$$

Clearly, $T_{X^k}$ is linear and one-to-one. Besides, it is also an isometric function with respect to the Riemannian distance (2). Indeed, we know that $T_{X^k}$ is an isometry if, and only if $\langle d_Y T_{X^k} \cdot V, d_Y T_{X^k} \cdot U \rangle_{T_{X^k}(Y)} = \langle V, U \rangle_Y$, for vectors $U, V \in T_Y S^n_{++} = S^n$, where $d_Y T_{X^k} \cdot V$ is a differential of $T_{X^k}$ in the direction $V$. Observe that $d_Y T_{X^k} \cdot V = (X^k)^{-\frac{1}{2}} V (X^k)^{-\frac{1}{2}}$. Therefore

$$\langle d_Y T_{X^k} \cdot V, d_Y T_{X^k} \cdot U \rangle_{T_{X^k}(Y)} = \langle F''(T_{X^k}(Y)) d_Y T_{X^k} \cdot V, d_Y T_{X^k} \cdot U \rangle$$

$$= \left\langle (T_{X^k}(Y))^{-1} d_Y T_{X^k} \cdot V \, (T_{X^k}(Y))^{-1}, d_Y T_{X^k} \cdot U \right\rangle$$

$$= \langle (X^k)^{\frac{1}{2}} Y^{-1} (X^k)^{\frac{1}{2}} (X^k)^{-\frac{1}{2}} V (X^k)^{-\frac{1}{2}} (X^k)^{\frac{1}{2}} Y^{-1} (X^k)^{\frac{1}{2}}, (X^k)^{-\frac{1}{2}} U (X^k)^{-\frac{1}{2}} \rangle$$

$$= \langle (X^k)^{\frac{1}{2}} Y^{-1} V Y^{-1} (X^k)^{\frac{1}{2}}, (X^k)^{-\frac{1}{2}} U (X^k)^{-\frac{1}{2}} \rangle \qquad (\mathrm{Tr}(AB) = \mathrm{Tr}(BA))$$

$$= \langle Y^{-1} V Y^{-1}, U \rangle = \langle F''(Y) V, U \rangle = \langle V, U \rangle_Y.$$

Particularly, distances are invariant under isometries.

**Lemma 1** *Let be $Y \in S^n_{++}$ and $\delta > 0$. Then*

$$T_{X^k}(B_\delta(Y)) = B_\delta(T_{X^k}(Y)) \ [3].$$

**Proof.** Since any isometry is a homeomorphism, the inclusion $T_{X^k}(B_\delta(Y)) \subset B_\delta(T_{X^k}(Y))$ is satisfied. Now, let $Z \in B_\delta(T_{X^k}(Y))$, so $\delta > d\left(Z, T_{X^k}(Y)\right) = d\left((X^k)^{\frac{1}{2}} Z (X^k)^{\frac{1}{2}}, Y\right)$. Then $(X^k)^{\frac{1}{2}} Z (X^k)^{\frac{1}{2}} \in B_\delta(Y)$, for all $Z \in T_{X^k}\left(B_\delta(Y)\right)$. ■

At this point, we are interested in the function $\phi : T_{X^k}(S^n_{++}) = S^n_{++} \longrightarrow R$ by

$$\phi(Z) = f\left((X^k)^{\frac{1}{2}} Z (X^k)^{\frac{1}{2}}\right). \tag{7}$$

**Proposition 1** *If $\bar{Z}$ is a local minimum of $\phi$ then $\bar{Y} = (X^k)^{\frac{1}{2}} \bar{Z} (X^k)^{\frac{1}{2}}$ is a local minimum of $f$.*

**Proof.** There exists $\epsilon > 0$ so that $\phi(\bar{Z}) \leq \phi(Z)$ for any matrix $Z \in B_\epsilon(\bar{Z})$. Applying the definition of $\phi$ we have

$$f(\bar{Y}) = f\left((X^k)^{\frac{1}{2}} \bar{Z} (X^k)^{\frac{1}{2}}\right) = \phi(\bar{Z}) \leq \phi(Z) = f\left((X^k)^{\frac{1}{2}} Z (X^k)^{\frac{1}{2}}\right) = f(Y).$$

[3] For each $Y \in S^n_{++}$ and $\delta > 0$, $B_\delta(Y)$ is the open disc with center $Y$ and radius $\delta$.

Since $T_{X^k}^{-1}\left(B_\epsilon\left(\bar{Z}\right)\right) = B_\epsilon\left(\bar{Y}\right)$ (Lemma 1 applied to $T_{X^k}^{-1}$), the inequality is satisfied for all matrices $Y \in B_\epsilon(\bar{Y})$ [4]. ■

Thus, the main iteration of the Proximal Point Method can be replaced by

$$Z^{k+1} = \text{argmin}_{Z \succ 0}\left\{\phi(Z) + \frac{\beta^k}{2}\sum_{i=1}^{n}\ln^2\lambda_i\left(Z\right)\right\}, \tag{8}$$

and

$$X^{k+1} = (X^k)^{\frac{1}{2}} Z^{k+1} (X^k)^{\frac{1}{2}}. \tag{9}$$

Since (8) and (9) represent general similarity transformations (in particular, the eigenvalues are the same) we can choose a particular transformation that produces a diagonal matrix $Z$. Let $(Q_j, Z)$ be a Schur decomposition of $(X^k)^{-\frac{1}{2}} Y (X^k)^{-\frac{1}{2}}$, i.e.,

$$(Q_j)^T (X^k)^{-\frac{1}{2}} Y (X^k)^{-\frac{1}{2}} Q_j = Z$$

($Z$ is diagonal and $Z \succ 0$). It's easy to verify that the operator $T_{(X^k)(Q_j)} : S_{++}^n \longrightarrow S_{++}^n$ given by

$$T_{(X^k)(Q_j)}(Y) = Q_j^T (X^k)^{-\frac{1}{2}} Y (X^k)^{-\frac{1}{2}} Q_j$$

is also an isometry with respect to the metric (2). If we restrict the domain of $T_{(X^k)(Q_j)}$ to the set of the diagonal definite positive matrices, the local problem in now defined in $R_{++}^n$. However, it is not equivalent to the original problem. So, we propose to replace the iteration (8) by a sequence of problems in $R_{++}^n$, given by

$$Z_{j+1} = \text{argmin}_{Z \succ 0}\left\{\phi_j(Z) + \frac{\beta^k}{2}\sum_{i=1}^{n}\ln^2\lambda_i(Z)\right\}, \tag{10}$$

where $\phi_j(Z) = f\left((X^k)^{\frac{1}{2}} Q_j Z Q_j^T (X^k)^{\frac{1}{2}}\right)$

[4] $T_{X^k}^{-1}$ is also an isometry.

**Algorithm 1 (Modified Proximal Algorithm)** *given* $X^0 \succ 0$ *symmetric,* $\beta^0, \delta_1 > 0$*:*
$k \longleftarrow 0$*;*
*repeat*
. *given* $Q_0$ *orthogonal and* $Z_0 \succ 0$ *diagonal*
. $j \longleftarrow 0$*;*
. *begin*
. *1. compute*

$$Z_j^{aux} = \text{argmin}_{Z \succ 0} \left\{ f\left( \left(X^k\right)^{\frac{1}{2}} Q_j Z Q_j^T \left(X^k\right)^{\frac{1}{2}} \right) + \frac{\beta^k}{2} \sum_{i=1}^{n} \ln^2 \lambda_i(Z) \right\}^5$$

. *2. if* $Z_j^{aux} = Z_j$ *then (stop criterion)*
. $X^{k+1} = \left(X^k\right)^{\frac{1}{2}} Q_j Z_j^{aux} Q_j^T \left(X^k\right)^{\frac{1}{2}}$*;*
. *break;*
. *else*
. $[Q_{j+1}, Z_{j+1}] = Schur(Q_j Z_j^{aux} Q_j^T)$*;*
. *end;*
. $j \longleftarrow j+1$*, take* $Z_j$ *like initial point and return to 1.*
. *end;*
. $\beta^k \longleftarrow \beta^{k+1}$*;*
. $k \longleftarrow k+1$*;*
*until* $d\left(X^{k+1}, X^k\right) > \delta_1$.

**Lemma 2** *The iterates* $Z_{j+1}$ *and* $Z_j^{aux}$ *of the Algorithm (1) are similar.*

**Proof.** By construction, we have

$$Q_{j+1}^T Q_j Z_j^{aux} Q_j^T Q_{j+1} = Z_{j+1}.$$

As $Q_j$ and $Q_{j+1}$ are orthogonal matrices, then $Q_j^T Q_{j+1}$ is also orthogonal and nonsingular. The lemma is proved. ■

The next equality is a direct consequence of the Lemma 2,

$$f\left( \left(X^k\right)^{\frac{1}{2}} Q_j Z_j^{aux} Q_j^T \left(X^k\right)^{\frac{1}{2}} \right) + \frac{\beta^k}{2} \sum_{i=1}^{n} \ln^2 \lambda_i(Z_j^{aux}) =$$
$$= f\left( \left(X^k\right)^{\frac{1}{2}} Q_{j+1} Z_{j+1} Q_{j+1}^T \left(X^k\right)^{\frac{1}{2}} \right) + \frac{\beta^k}{2} \sum_{i=1}^{n} \ln^2 \lambda_i(Z_{j+1}).$$

[5] $Z_j^{aux}$ is a auxiliary variable that is computed in the $j$-nd iteration of the Algorithm 1.

**Proposition 2** *If* $Z_j^{aux} \neq Z_j$ *then*

$$g_{\beta^k}\left(\left(X^k\right)^{\frac{1}{2}} Q_{j+1} Z_{j+1} Q_{j+1}^T \left(X^k\right)^{\frac{1}{2}}\right) < g_{\beta^k}\left(\left(X^k\right)^{\frac{1}{2}} Q_j Z_j Q_j^T \left(X^k\right)^{\frac{1}{2}}\right). \tag{11}$$

**Proof.** It is a direct consequence of the construction of the Algorithm 1, Lemma 2 and the similarity properties. ■

We need to show that if $Z_j^{aux} = Z_j$ then $\left(X^k\right)^{\frac{1}{2}} Q_j Z_j^{aux} Q_j^T \left(X^k\right)^{\frac{1}{2}} = X^{k+1}$, i.e. the inner loop of the Algorithm 1 converges to $X^{k+1}$. This is a rsult to be obtained. The convergence of the external loop is assured by [6].

## 2.1 The cone of the diagonal definite positive matrices

We show in the following, problems in $R^n$ considered in our framework. Let $D \subset S_+^n$ be the set of the diagonal semidefinite positive matrices and $f : D \longrightarrow R$ a geodesic convex function. Consider the problem

$$\min \ f(X)$$

$$X \in D.$$

Given a starting point $X^0 \in \texttt{int}(D)$, we can apply the Algorithm 1 on this problem. So, take $k = 0$, $Q_0 = I$ and some matrix $Z_0 \in int(D)$ as the starting point for the inner loop, and set $j = 0$. Compute $Z_j^{aux}$ in the main iteration and update $Q_{j+1}$ and $Z_{j+1}$ through the Schur decomposition of $Q_j Z_j Q_j^T$, i.e.,

$$Q_{j+1}^T \left(Q_j Z_j^{aux} Q_j^T\right) Q_{j+1} = Z_{j+1}$$

Now, from step $j$ to $j+1$, we have that $Q_{j+1} = Q_j = I$ and $Z_{j+1} = Z_j^{aux}$. When we return to the step 1 and we compute $Z_{j+1}^{aux}$, we get $Z_{j+1}^{aux} = Z_{j+1}$. As an outcome, we have that $Z_{j+1}$ is the proximal point of $X^k$, i.e.

$$Z_{j+1} = X^{k+1}. \tag{12}$$

Let's observe that it is not necessary to update the orthogonal matrix,

$$Q_0 = \cdots = Q_j = Q_{j+1} = I. \tag{13}$$

In that case, we have assured the convergence of the Algorithm 1, but it is a very singular situation. It corresponds to the class of functions $f$ that only depends on the eigenvalues of the variable $X \in S_{++}^n$.

## 3 A well-posed problem

The step 1 of the Algorithm 1 is a minimization problem on the space of the diagonal definite positive matrices that is equivalent to a nonlinear programming problem in the positive octant of $R^n$. Clearly, the complexity study of the algorithm needs estimation of both inner and outer iterations. We do not take this in consideration, in this paper.

Define $\varphi_j : R^n_{++} \longrightarrow R$ by

$$\varphi_j(z) = f\left((X^k)^{\frac{1}{2}} Q_j Z Q_j^T (X^k)^{\frac{1}{2}}\right), \tag{14}$$

where $Z$ is the diagonal matrix constructed with the components of $z$, i.e., $Z = \mathrm{diag}(z)$. The regularization $\sum_{i=1}^{n} \ln^2 \lambda_i(Z)$ can be seen as a function from $R^n_{++}$ to $R$. Clearly, $\sum_{i=1}^{n} \ln^2 z_i$ is a barrier for $R^n_{++}$. Therefore, in terms of $z$, the step (1) can be rewritten as

$$z_{j+1} = \mathrm{argmin}_{z>0} \left\{ \varphi_j(z) + \frac{\beta^k}{2} \sum_{i=1}^{n} \ln^2 z_i \right\}. \tag{15}$$

Now, having supposed the differentiability of $f$, given a vector $v \in R^n$ and a small $t$, the directional derivative of $\varphi_j$ in the direction $v$ is:
$\frac{d\varphi_j}{dv}(z) =$
$\frac{d}{dt} \mid_{t=0} \varphi_j(z + tv) = \frac{d}{dt} \mid_{t=0} f\left((X^k)^{\frac{1}{2}} Q_j [Z + tV] Q_j^T (X^k)^{\frac{1}{2}}\right)$
$= Df\left((X^k)^{\frac{1}{2}} Q_j Z Q_j^T (X^k)^{\frac{1}{2}}\right) \bullet \left((X^k)^{\frac{1}{2}} Q_j V Q_j^T (X^k)^{\frac{1}{2}}\right),$
where $V = \mathrm{diag}(v)$. Therefore,

$$\nabla\varphi_j(z)_i = Df\left((X^k)^{\frac{1}{2}} Q_j Z Q_j^T (X^k)^{\frac{1}{2}}\right) \bullet \left((X^k)^{\frac{1}{2}} Q^k E_i (Q^k)^T (X^k)^{\frac{1}{2}}\right),$$

where $E_i = \mathrm{diag}(e_i)$ ($i$-th vector of the canonic base of $R^n$).

Considering $\sum_{i=1}^{n} \ln^2 z_i$, we see that it is not a strictly convex function, but it is coercive: take $\|z\| \longrightarrow \infty$, then $z_k \longrightarrow \infty$, for some $1 \leq k \leq n$. So, $\sum_{i=1}^{n} \ln^2 z_i \longrightarrow \infty$, too.

**Definition 1** *Let $S \subseteq R^n$ be closed convex and $f : S \longrightarrow R$. $f$ is said strongly quasiconvex function if, and only if, to each $x_1, x_2 \in S$, with $x_1 \neq$*

$x_2$,

$$f(tx_1 + (1-t)x_2) < \max\{f(x_1), f(x_2)\}, \tag{16}$$

*for all* $t \in (0,1)$.

The next result is an extension of the uniqueness solution theorem applied to strictly convex function problems.

**Theorem 1** *Let* $f : S \subseteq R^n \longrightarrow R$ *be strongly quasiconvex. Consider the problem*

$$\{\min\ f(x);\ x \in S\}. \tag{17}$$

*If* $\bar{x}$ *is a local solution of (17) then* $\bar{x}$ *is the unique global solution.*

**Proof.** See [2] ■

**Lemma 3** *The function* $f : R_{++} \longrightarrow R_+$, *defined by*

$$f(t) = \ln^2 t,$$

*is strongly quasiconvex in* $R_{++}$.

**Proof.** Take $t_1, t_2 \in R_{++}$ with $t_1 \neq t_2$, define $h(\alpha) = f(\alpha t_1 + (1-\alpha)t_2)$, $\alpha \in [0,1]$. $h$ is a continuous differentiable function in the open interval $(0,1)$, so it has maximum and minimum point in $[0,1]$. An interior solution of $h$ must satisfy $h'(\bar{\alpha}) = 0$. On the other hand,

$$h'(\alpha) = \frac{2\ln(\alpha t_1 + (1-\alpha)t_2)}{\alpha t_1 + (1-\alpha)t_2}(t_1 - t_2),$$

therefore

$$h'(\bar{\alpha}) = 0 \Longleftrightarrow \frac{2\ln(\bar{\alpha} t_1 + (1-\bar{\alpha})t_2)}{\bar{\alpha} t_1 + (1-\bar{\alpha})t_2}(t_1 - t_2) = 0 \Longleftrightarrow \bar{\alpha} t_1 + (1-\bar{\alpha})t_2 = 1.$$

Solving the equation for $\bar{\alpha}$ we have $\bar{\alpha} = (1-t_2)/(t_1 - t_2)$ and $h(\bar{\alpha}) = 0$. Then, $h(\bar{\alpha}) < \max\{h(0), h(1)\} = \max\{\ln^2 t_1, \ln^2 t_2\}$. We recall that $t_1 \neq t_2$ resulting that $\ln^2 t_1$ and $\ln^2 t_2$ are not simultaneously zero. To finish, we consider the case where $\bar{\alpha}$ is a minimum or inflection point; then the maximum point of $h$ will be at some extreme of the interval $[0,1]$. So,

$(i)$ $\alpha = 0 \Longrightarrow h(0) = \ln^2 t_2 \leq \max\{\ln^2 t_1, \ln^2 t_2\}$.

$(ii)$ $\alpha = 1 \Longrightarrow h(1) = \ln^2 t_1 \leq \max\{\ln^2 t_1, \ln^2 t_2\}$.

■

Notice that the barrier $\sum_{i=1}^{n} \ln^2 z_i$ is a sum of strongly quasiconvex functions and consequently is strongly quasiconvex, too. When we sum a strongly quasiconvex function with a convex function, the result is a strongly quasiconvex function. Indeed, given $f_1$ a convex function and $f_2$ a strongly quasiconvex function in $S \subseteq R^n$ and $x_1, x_2 \in S(x_1 \neq x_2)$, $t \in (0,1)$, $(f_1 + f_2)(tx_1 + (1-t)x_2) = f_1(tx_1 + (1-t)x_2) + f_2(tx_1 + (1-t)x_2) < tf_1(x_1) + (1-t)f_1(x_2) + \max\{f_2(x_1), f_2(x_2)\} < t\max\{f_1(x_1), f_1(x_2)\} + (1-t)\max\{f_1(x_1), f_1(x_2)\} + \max\{f_2(x_1), f_2(x_2)\} = \max\{(f_1 + f_2)(x_1), (f_1 + f_2)(x_2)\}$.

**Theorem 2** *The objective function of the problem (15) has only one minimum.*

**Proof.** The result is a direct consequence of the coercivity of $\sum_{i=1}^{n} \ln^2 z_i$, Lemma 3 and the definition of $\varphi$ in (14). ■

# 4 Applying Newton's method

Suppose that $\varphi_j$ is a strongly convex function of $C^2$ class. Let be $\alpha > 0$, satisfying $\alpha < \mu$, where $\mu = \lambda_{\min}\left(\varphi_j''(z)\right)$, $z \in R^n_{++}$. The following is true (a direct application of Weyl Theorem (see, e.g., [8]):

$$\lambda_{\min}\left(\varphi_j''(z) + \beta \operatorname{diag}\left(\frac{1-\ln z_i}{z_i^2}\right)\right) \geq \mu + \beta \min_{1\leq i\leq n}\left(\frac{1-\ln z_i}{z_i^2}\right) \geq \alpha. \tag{18}$$

So,

$$\beta \min_{1\leq i\leq n}\left(\frac{1-\ln z_i}{z_i^2}\right) \geq \alpha - \mu. \tag{19}$$

As $\alpha - \mu < 0$, the interesting part of this analysis corresponds to the case $\min_{1\leq i\leq n}(1-\ln z_i) < 0$, i.e. when $z_i > e$, for some $1 \leq i \leq n$. Then, from (19), we obtain

$$\beta \leq \min_{z_i > e,\ 1\leq i\leq n}\left(\frac{(\mu-\alpha)z_i^2}{\ln z_i - 1}\right).$$

The analysis of the function $h(t) = \frac{t^2}{\ln t - 1}$ for $t > e$, shows that $\bar{t} = e^{\frac{3}{2}} = \text{argmin}_{t>e}\, h(t)$ and the optimum value associated to $\bar{t}$ is given by $h(\bar{t}) = 2e^3$. Therefore, the choice

$$\beta \leq (\mu - \alpha)2e^3,$$

in particular,

$$\beta \leq \mu e^3 \tag{20}$$

satisfies (18). Observe that this condition accords with the hypothesis (5).

Finally, we point out that the barrier function $\sum_{i=1}^{n} \ln^2 z_i$ is strictly convex in the set $(0, e)^n$. That is an important property for classes of problems where the eigenvalues of $X$ belong to the open interval $(0, 1)$, for example Semidefinite Programming with the restriction $\text{Tr}X = 1$.

# 5 Linear $(SDP)$ problems

Consider the formulation of the Linear Semidefinite Programming $(SDP)$

$$\{\min \;\; C \bullet X; \;\; \text{s.a.} \;\; A_i \bullet X = b_i, \; i = 1, \cdots, m, \;\; X \succeq 0\}.$$

Applying the Algorithm 1 on this problem, the step 1 gets a solution $z_{j+1}$ of $(PP)_{SDP}$

$$\begin{gathered} \min \; C \bullet (X^k)^{\frac{1}{2}} Q_j Z Q_j^T (X^k)^{\frac{1}{2}} + \frac{\beta^k}{2} \sum_{i=1}^{n} \ln^2 z_i \\ \text{s.a} \; A_i \bullet (X^k)^{\frac{1}{2}} Q_j Z Q_j^T (X^k)^{\frac{1}{2}} = b_i, \; i = 1, \cdots, m \\ z > 0, \end{gathered}$$

where $Z = \text{diag}(z)$. As the objective function and the restrictions are linear, we can simplify some terms and rewrite above problem as $(PP)_{SDP}$

$$\begin{gathered} \min \; (c^j)^T z + \frac{\beta^k}{2} \sum_{i=1}^{n} \ln^2 z_i \\ \text{s.a} \; A^j z = b \\ z > 0, \end{gathered}$$

for some vector $c^j \in R^n$ and some matrix $A^j_{m \times n}$.

The necessary hypothesis to the application of Newton's method can be assured, for example, adding the regularization $-\mu^k \sum_{i=1}^{n} \ln z_i$. Then, we consider the new problem $(RPP)_{SDP}$:

$$\begin{array}{c} \min\ (c^j)^T z - \mu^k \sum_{i=1}^{n} \ln z_i + \dfrac{\beta^k}{2} \sum_{i=1}^{n} \ln^2 z_i \\ \text{s.a } A^j z = b \\ z > 0. \end{array}$$

The Lagrange function associated to $(RPP)_{SDP}$ is given by

$$L(z,y) = (c^j)^T z - \mu^k \sum_{i=1}^{n} \ln z_i + \frac{\beta^k}{2} \sum_{i=1}^{n} \ln^2 z_i - y^T (A^j z - b).$$

The $KKT$ conditions for $(RPP)_{PSD}$ are:

$$\begin{array}{c} c^j - \mu^j z^{-1} + \beta^k\ [\mathrm{diag}(z)]^{-1} \ln\ z - (A^j)^T y = 0 \\ b - A^j z = 0 \\ (z > 0). \end{array}$$

Now, in order to obtain Newton's direction, take $\bar{z} > 0$ a feasible point to $(RPP)_{PSD}$ and $\bar{y} \in R^n$. The Newton step must satisfy

$$\begin{array}{c} \left[\mu^k \bar{Z}^{-2} + \beta^k\ \mathrm{diag}\left(\frac{\bar{e} - \ln\ \bar{z}}{\bar{z}_i^2}\right)\right] \Delta z - (A^j)^T \Delta y = \\ -c^j + \mu^k \bar{z}^{-1} - \beta^k\ [\mathrm{diag}(\bar{z})]^{-1} \ln\ \bar{z} + (A^j)^T \bar{y} \end{array} \tag{21}$$

$$-A^j \Delta z = 0, \tag{22}$$

where $\bar{e}$ is the vector of $R^n$ with all components equal to 1. Define by $U = \mathrm{diag}\left(\frac{\mu^k + \beta^k (1 - \ln\ \bar{z}_i)}{\bar{z}_i^2}\right)$ the coefficient of $\Delta z$. Through the expression of the inverse of the coefficient matrix in (21) and (22), we see how the (logarithmic) regularization is essential. It is given by

$$\left[\begin{array}{cc} U^{-1}\left(I + (A^j)^T P^{-1} A^j U^{-1}\right) & -U^{-1}(A^j)^T P^{-1} \\ -P^{-1} A^j U^{-1} & P^{-1} \end{array}\right],$$

where $P = -A^j U (A^j)^T$.

**Lemma 4** *Suppose that the rows of the matrix $A^j$ of the $(RPP)_{PSD}$ are linearly independent. Then, given $\mu^k > 0$, satisfying*

$$\frac{\mu^k}{\beta^k} > \min_{\bar{z}_i > e,\ 1 \le i \le n} (\ln \bar{z}_i - 1),$$

*the condition for the application of Newton's method to $(RPP)_{PSD}$ is assured.*

**Proof.** Applying the same analysis of the previous section, we have

$$\mu^k + \min_{1\leq i\leq n}\beta^k(1-\ln\bar{z}_i) > 0.$$

If $\min_{1\leq i\leq n}(1-\ln\bar{z}_i) \geq 0$ then for all $\mu > 0$ Newton's method can be applied. If $\min_{1\leq i\leq n}(1-\ln\bar{z}_i) < 0$, i.e., if $\bar{z}_i > e$, for some $1 \leq i \leq n$ we must have

$$\frac{\mu^k}{\beta^k} > \min_{z_i>e,\ 1\leq i\leq n}(\ln\bar{z}_i - 1) > 0.$$

■

# 6 Conclusions

We have proposed a knew proximal algorithm for semidefinite programming, based on the local aproximation of the main iteration by a sequence of $R^n$ nonlinear problems. It appears to be the first in the nonlinear class os $SDP$ algorithms, in the proximal framework. There are two classes of problems where we expect to have the better application of the algorithm: problems defined in the "hypercube" $0 \leq X \leq I$, where the barrier we used is strict convex, and quadratic problems that can be represented, under some similarity transformation, by a sum that only depends on the eigenvalues. Also, we are working on the convergence theory, complexity results and numerical implementation.

# References


[1] Alizadeh, F., "Interior point methods in Semidefinite programming with applications to combinatorial optimization", *SIAM Journal on Optimization*, v. 5, n. 1, pp. 13-51, 1995.

[2] Bazaraa, M.S., Sherali, H.D., Shety, C.M, *Nonlinear Programming. Theory and Algorithms*. 2nd ed. New York, Wiley, 1993.

[3] Burer, S., Monteiro, R.D.C., "A nonlinear programming algorithm for solving semidefinite programs via low-rank factorization", *Mathematical Programming*, Series B, n. 95, pp. 329-357, 2003.

[4] Burer, S., Monteiro, R.D.C., Zhang, Y., "Solving a class of semidefinite programs via nonlinear programming", *Mathematical Programming*, v. 93, pp. 97-123, 2002.

[5] do Carmo, M.P., *Geometria riemanniana*. $2^a$ ed. Rio de janeiro, impa, 1988.

[6] Ferreira, O.P., Oliveira, P.R., "Proximal point algorithm on Riemannian manifolds", *Optimization*, v. 51, n. 2, pp. 257-270, 2002.

[7] Fukuda, M., Kojima, M., Murota, K., Nakata, K., "Exploiting sparsity in semidefinite programming via matrix completion I: General framework", *SIAM Journal on Optimization*, v. 11, pp. 647-674, 2001.

[8] Golub, G.H., Van Loan, C.F., *Matrix computations*. 3nd ed. Baltimore and London, The Johns Hopkins University Press, 1996.

[9] Güler, O., "Barrier functions in interior-point methods", *Mathematics of Operations Research*, v. 21, pp. 860-885, 1996.

[10] Helmberg, C., *An Interior Point Method for Semidefinite Programming and Max-Cut Bounds*. Ph.D. dissertation, Technischen Universität Graz, Graz, 1994.

[11] Helmberg, C., Rendl, F., "A spectral bundle method for semidefinite programming", *SIAM Journal on Optimization*, v. 10, pp. 673-696, 2000.

[12] Hiriart-Urruty, J-B., Lemaréchal, C., *Convex analysis and minimization algorithms II, advanced theory and bundle methods*. 1nd ed. Heidelberg, springer-verlag, 1993.

[13] Horn, R.A., Johnson, C.R., *Matrix Analysis*. 1nd ed. Cambridge, Cambridge University Press, 1985.

[14] Horn, R.A., Johnson, C.R., *Topics in Matrix Analysis*. 1nd ed. Cambridge, Cambridge University Press, 1991.

[15] Horner, S., Peinado, M., "Design and performance of parallel and distributed aproximation algorithms for max cut", *Journal of Parallel and Distributed computing*, v. 46, pp. 48-61, 1997.

[16] Monteiro, R.D.C., "First and second-order methods for semidefinite programming", *Mathematical Programming*, Series B, n. 97, pp. 209-244, 2003.

[17] Mosheye, L., Zibulevsk, M., *Penalty/Barrier Multiplier Algorithm for Semidefinite Programming*. Technical report, 1999.

[18] Nesterov, Y.E., Nemirovskii, A.S., *Interior-Point polynomial algorithms in convex Programming, vol. 13 of Studies in Applied Mathematics*. 1nd ed. Philadelphia, SIAM, 1994.

[19] Nesterov, Y.E., Nemirovskii, A.S., *Optimization over positive semidefinite matrices: Mathematical background and users manual*. Technical report, Central Economic and Mathematical Institute, USSR Acad. Sci. Moscow, USSR, 1990.

[20] Nesterov, Y.E., Nemirovskii, A.S., "Polynomial barrier methods in convex programming", *Ekonomika i Matem. Metody*, v. 24, pp. 1084-1091, 1988.

[21] Nesterov, Y.E., Todd, M.J., "On the Riemannian geometry defined by self-concordant barriers and interior-point methods", *Foundations of Computational Mathematics*, v. 2, pp. 333-361, 2002.

[22] Roos, T., Terlaky, T., Vial, J.-P., *Theory and Algorithms for Linear Optimization, An Interior Point Approach*. 1nd ed. New York, John Wiley and Sons, 1997.

[23] Rothaus, O.S., "Domains of positivity", *Abh. Math. Sem. Univ. Hamburg*, v. 24, pp. 189-235, 1960.

[24] Saigal, R., Vandenberghe, L., Wolkowicz, H., *Handbook of Semidefinite Programming*. 1nd ed. Boston-Dordrecht-London, Kluwer Academic Publishers, 2000.

[25] Vandenberghe, L., Boyd, S.,"Semidefinite Programming", *SIAM Journal on Optimization*, v. 38, n. 1, pp. 49-95, 1996.

[26] Vanderbei, R.J., Benson, Y., *On formulating semidefinite programming problems as smooth nonlinear optimization problems. Report orfe 99-01*, Princeton, Operations Research and Financial engineering, 1999.