A NEW SELF-CONCORDANT BARRIER FOR THE HYPERCUBE

E. A. Papa Quiroz and P. Roberto Oliveira

Programa de Engenharia de Sistemas e Computação -COPPE
Federal University of Rio de Janeiro
Rio de Janeiro, Brazil
erik@cos.ufrj.br, poliveir@cos.ufrj.br

June 2004

Abstract

In this paper we introduce a new barrier function $\sum_{i=1}^{n}(2x_i - 1)[\ln x_i - \ln(1 - x_i)]$ to solve the following optimization problem: min $f(x)$ subject to: $Ax = b$; $0 \leq x \leq e$. We show that this function is a $(3/2)n$-self-concordant barrier on the hypercube $[0, 1]^n$. We prove that the central path is well defined and that under an additional assumption on the objective function, which includes linear and convex quadratic problems, the primal central path converges to the analytic center of the optimal set (with respect to the given barrier). Finally, we present a new polynomial long step path following algorithm to solve the problem when the objective function is linear. For that algorithm we give an upper bound for the total number of Newton iterations to obtain an ϵ-optimal solution, with similar complexity to the classic logarithmic barrier, that is $O(n \ln(n\mu_0/\epsilon))$.

Keywords: Interior point methods, self-concordant function, barrier function, Newton methods.

1 Introduction

The logarithmic barrier function, proposed by Frisch [9] in 1955 and further developed by Fiacco and McCormick [8], plays a fundamental role to obtain polynomial interior point methods for certain classes of convex optimization problems, for instance linear programs, convex quadratic programs and semidefinite programs. Other barrier functions have been proposed, ensuring convergence properties for their algorithms, but unfortunately they were not polynomial, see [5], [7], [23]. Observe that all of them are not self-concordant barriers, property that was introduced by Nesterov and Nemirovskii [18]. Those authors develop a general approach that allows the using of interior point methods to solve convex program based on that class of barriers. They have proved that self-concordant barriers associated with Newton methods permit the development of performing algorithms in both points of view: theoretical and practical. Although the theoretical importance of that approach, actually the unique known self-concordant barrier is just the logarithmic function. So, a natural question arises: is it possible to obtain other self-concordant barriers with similar properties to the classic logarithmic barrier function? That is the motivation to the present work. The outline of this paper is as follows. In subsection 1.1 we give the notation used along this paper and in subsection 1.2 we review some facts on Riemannian manifolds. In section 2, we present the linear constrained convex optimization problem with bounded variables, studied in this paper, we impose some assumptions on that problem, and introduce a new barrier for the hypercube $[0,1]^n$, defined by $B(x) = \sum_{i=1}^{n}(2x_i - 1)[\ln x_i - \ln(1-x_i)]$. In section 3 we prove that the Riemannian manifold $\langle 0,1\rangle^n$ endowed with the metric $X^{-2}(I-X)^{-2}$, where $X = \mathtt{diag}(x_1, x_2, ..., x_n)$, which is the Hessian of the barrier B, has suitable geometric properties. In Section 4 we prove that the introduced barrier B is $(3/2)n$-self-concordant on the hypercube $[0,1]^n$. In Section 5 we propose a new interior penalized optimization problem to solve the original problem, and define central path, obtaining some feasibility properties. In Section 6 we prove that the primal central path, defined in the previous section, converges, and its limit point is the analytic center of the optimal set of the problem. In Section 7, we present an application for linear programming. For that class of problems, we present a new polynomial long step path following algorithm, deriving upper bounds to obtain an ϵ-solution. The complexity of our algorithm is $O(n\ln(n\mu_0/\epsilon))$, for the number de iterations, identical to the long step path following algorithm that uses the logarithmic barrier function.

1.1 Notation

The following notation is used throughout this paper:

1. $[0,1]^n = [0,1] \times [0,1] \times ... \times [0,1]$ (n times)
2. $\langle 0,1\rangle^n = \langle 0,1\rangle \times \langle 0,1\rangle \times ... \times \langle 0,1\rangle$ (n times)
3. Given a subset $M \subset I\!R^n$, $\mathtt{ri}(M)$ denotes the relative interior of M.
4. $\mathtt{bound}[0,1]^n$ is the boundary of $[0,1]^n$.
5. $\ln X = \mathtt{diag}(\ln x_1, ..., \ln x_n)$ and $\ln(I-X) = \mathtt{diag}(\ln(1-x_1), ..., \ln(1-x_n))$.
6. $C^\infty(< 0,1 >^n) = \{f :< 0,1 >^n \to I\!R$ such that f is a infinitely differentiable function $\}$.

1.2 Basic aspects of Riemannian geometry

In this subsection we introduce some fundamental properties and notation for Riemannian manifolds, that will be applied in section 3, where we study the geometric properties of our barrier. Those basic facts can be encountered, for example, in do Carmo [6].

Let S be a differential manifold. We denote by T_xS the tangent space of S at $x \in S$ and $TS = \bigcup_{x\in S} T_xS$. T_xS is a linear space and has the same dimension as S. Because we restrict ourselves to real manifolds, T_xS is isomorphic to $I\!R^n$. If S is endowed with a Riemannian metric g then S is a Riemannian manifold and we denoted it by (S, g). The inner product of two vectors $u, v \in T_xS$ is written $\langle u, v\rangle_x := g_x(u, v)$, where g_x is the metric at the point x. The norm of a vector $v \in T_xS$ is defined by $||v||_x := \sqrt{\langle v, v\rangle_x}$.

The metric can be used to define the length of a piecewise smooth curve $\alpha : [a, b] \to S$ joining p' to p through $L(\alpha) = \int_a^b \|\alpha'(t)\|dt$, where $\alpha(a) = p'$ and $\alpha(b) = p$. Minimizing this length functional over the set of all curves we obtain a Riemannian distance $d(p', p)$ which induces the original topology on S.

Given two vector fields V and W in S (a vector field V is an application of S in TS), the covariant derivative of W in the direction V is denoted by $\nabla_V W$. In this paper ∇ is the Levi-Civita connection associated to (S, g). This connection defines an unique covariant derivative D/dt, where for each vector field $V : S \to TS$, along a smooth curve $\alpha : [a, b] \to S$, another vector field is obtained, denoted by DV/dt. A curve α is a geodesic, starting from the point x with direction $v \in T_xS$, if $\alpha(0) = x$, $\alpha'(0) = v$ and

$$\frac{d^2\alpha_k}{dt^2} + \sum_{i,j=1}^{n} \Gamma_{ij}^k \frac{d\alpha_i}{dt}\frac{d\alpha_j}{dt} = 0, \quad k = 1, ..., n, \tag{1.1}$$

where Γ_{ij}^k are the Christoffel's symbols expressed by

$$\Gamma_{ij}^m = \frac{1}{2}\sum_{k=1}^{n}\left\{\frac{\partial}{\partial x_i} g_{jk} + \frac{\partial}{\partial x_j} g_{ki} - \frac{\partial}{\partial x_k} g_{ij}\right\} g^{km}, \tag{1.2}$$

(g^{ij}) denotes the inverse matrix of the metric $g = (g_{ij})$, and x_i and v_i are the coordinates of x and v, respectively. Now, suppose that V and W are represented by $V = \sum_{i=1}^n u^i V_i$, $W = \sum_{i=1}^n z^i V_i$, for some local basis V_i for T_xS, then $\nabla_{X_i} X_k = \sum_{j=1}^n \Gamma_{ik}^j X_j$. A Riemannian manifold is complete if its geodesics are defined for any value of $t \in I\!R$.

Given the vector fields V, W, Z on S, we denote by R the curvature tensor defined by

$$R(V, W)Z = \nabla_W\nabla_V Z - \nabla_V\nabla_W Z + \nabla_{[V,W]}Z, \tag{1.3}$$

where $[V, W] := WV - VW$ is the Lie bracket. If $R(V, W) = 0$, for all vector fields V, W, then S is called a null curvature Riemannian manifold.

2 Definition of the problem and the new barrier function

In this paper we are interested in solving:

$$\begin{array}{ll} (P) & \min\ f(x) \\ & Ax = b \\ & 0 \le x \le e \end{array} \tag{2.1}$$

where $f : I\!R^n \to I\!R$ is a twice continuously differentiable convex function, A is a $m \times n$ matrix with $m \leq n$, e is a n dimensional vector whose components are 1, b and c are $m-$and $n-$dimensional vectors respectively; the $n-$dimensional vector x is the variable. We shall observe that all the results obtained in this paper are easily extendable to general hypercubes, given by $\alpha \leq x \leq \beta$, for any n dimensional vectors α and β.
We denoted by

$$F = \{x \in I\!R^n : Ax = b, 0 \leq x \leq e\}$$

the feasible set of the problem (2.1) and

$$F^0 = \{x \in I\!R^n : Ax = b, 0 < x < e\}$$

its relative interior. We impose the following assumptions on the problem (2.1)

Assumptions:

1. The set F^0 is non-empty.
2. The objective function satisfies the following condition: there exists a vector subspace W of $I\!R^n$, such that $\mathtt{ker}(\nabla^2 f(x)) = W$, for all $x \in < 0, 1 >^n$.
3. The matrix A has full rank, that is, $\mathtt{rank}(A) = m$.

We point out that 1 is a standard assumption in interior point methods. The assumption 2, will be important to prove the convergence of the primal central path to the analytic center of the optimal set of the problem (2.1). This assumption is satisfied by self-concordant functions (see Nesterov and Nemirovskii [18], Corollary 2.1.1), in particular, the linear and quadratic convex objective functions. The assumption 3, could be discarded in our development, however we keep it since it will simplify the arguments.
Due to the fact that f is a continuous function on the compact set F, it achieves a global minimum point on F. Besides, the convexity of f implies that any local minimum is a global minimum. Thus, the set of optimal solutions of the problem (2.1), that we denote by $\mathtt{sol}(P)$, is a non-empty and bounded convex set.
The Wolfe dual problem formulation of (2.1) is:

$$\begin{array}{rl} \max & d(x,y,s,w) = f(x) - y^T(Ax-b) + w^T(x-e) - s^T x \\ & A^T y + s - w = \nabla f(x) \\ & w, s \geq 0 \\ & y \in I\!R^m;\ x, s, w \in I\!R^n \end{array} \tag{2.2}$$

A feasible point (x, y, s, w) for (2.2), with $(s, w) > 0$ is called an interior dual feasible solution.
Now, we define in the n-dimensional hypercube $[0, 1]^n$ the function:

$$B(x) = \sum_{i=1}^{n} (2x_i - 1)[\ln x_i - \ln(1 - x_i)]. \tag{2.3}$$

We have immediately the following properties:

- If $x \to \mathtt{bound}[0, 1]^n$ (x approaches the boundary) then $B(x) \to \infty$.
- $B(x) \geq 0$, for all $x \in [0, 1]^n$
- $B \in C^\infty(\langle 0, 1\rangle^n)$, and the second order derivative of B is $X^{-2}(I - X)^{-2}$.

Consequently, B is a strictly convex barrier function infinitely differentiable on $\langle 0,1\rangle^n$.

In the following section we show that the Hessian $X^{-2}(I-X)^{-2}$ of the barrier function B gives a very rich geometric structure on the manifold $\langle 0,1\rangle^n$, that is, $\langle 0,1\rangle^n$ is a complete Riemannian manifold with null curvature. The first one means, essentially, that there is a geodesic between any two points, so we can measure distances in a appropriate way, through the minimum length geodesic. The second one means that the Riemannian manifold and the Euclidean space have similar geometric properties.

3 Geometric properties: completeness and null curvature

This section is motivated by the following fact: if we consider the Riemannian manifold $I\!R^n_{++}$ endowed with the Hessian metric X^{-2} of the logarithmic barrier function $\sum_{i=1}^{n} \ln x_i$, it can be proved that R^n_{++} is complete with null curvature (see Oliveira and Cruz Neto [20]). The following two theorems show that our barrier also satisfies those properties. We notice that the null curvature is a general property of diagonal Riemannian metrics on $\langle 0,1\rangle^n$ (see Papa Quiroz and Oliveira [21]), but for the sake of completeness we will give the proof. For a general study of Riemannian geometry applied to optimization we refer the reader to the following references: Luenberger [14], Gabay [10], Karmarkar [13], Cruz Neto et al., [3], Udriste [26], Rapcsák [25], Nesterov and Todd [19], and their references.

Theorem 3.1 *The Riemannian manifold* $\left(\langle 0,1\rangle^n, X^{-2}(I-X)^{-2}\right)$ *is complete.*

Proof. By definition $\langle 0,1\rangle^n$ is complete if its geodesic curves are defined for any value of $t \in I\!R$. We will obtain the explicit geodesic curves $\alpha : [a,b] \to \langle 0,1\rangle^n$, and verify that property. As in our case $(g_{ij}) = X^{-2}(I-X)^{-2}$ and so $(g^{ij}) = X^2(I-X)^2$, we have that the Christoffel's symbols (1.2) are reduced to:

$$\Gamma^m_{ij} = -\frac{2\alpha_i - 1}{\alpha_i(1-\alpha_i)}\delta_{im}\delta_{ij}. \tag{3.1}$$

Substituting this expression in (1.1) we obtain that the geodesic equation is equivalent to:

$$\int \frac{1}{\alpha_i(1-\alpha_i)} d\alpha_i = a_i t + b_i, \qquad i = 1,...,n$$

for some constants a_i, and b_i in $I\!R$, such that $\alpha_i(0) = x_i$ and $\alpha_i'(0) = v_i$, $i = 1,...,n$. Solving the integral equation, and considering the initial conditions, gives

$$\alpha_i(t) = \frac{1}{2}\left\{1 + \tanh\left(\frac{1}{2}\frac{v_i}{p_i(1-p_i)}t + \frac{1}{2}\ln\frac{p_i}{1-p_i}\right)\right\},\ i = 1,...,n. \tag{3.2}$$

where $\tanh(z) = (e^z - e^{-z})/(e^z + e^{-z})$ is the hyperbolic tangent function. This geodesic curve is well defined for all $t \in I\!R$, so the manifold $\langle 0,1\rangle^n$ is complete. ■

Theorem 3.2 *The Riemannian manifold* $\left(\langle 0,1\rangle^n, X^{-2}(I-X)^{-2}\right)$ *has null curvature.*

Proof. Given $p \in \langle 0,1\rangle^n$, let $\{V_i\}_{i=1,...,n}$ be a basis for $T_p(\langle 0,1\rangle^n) = I\!R^n$, then, we can write $V = \sum_{i=1}^{n} u^i V_i$, $W = \sum_{j=1}^{n} v^j V_j$, $Z = \sum_{k=1}^{n} w^k V_k$. As the curvature tensor R is trilinear we have:

$$R(V,W)Z = \sum_{i,j,k} u^i v^j w^k R(V_i, V_j)V_k.$$

¿From the definition of R (see section 1.2) we obtain:

$$R(V_i,V_j)V_k = \nabla_{V_j}(\nabla_{V_i}V_k) - \nabla_{V_i}(\nabla_{V_j}V_k) + \nabla_{[V_i,V_j]}X_k$$

Now, for the Riemannian connection, we have $[V_i, V_j] = 0$, and

$$R(V_i,V_j)V_k = \nabla_{V_j}(\nabla_{V_i}V_k) - \nabla_{V_i}(\nabla_{V_j}V_k)$$

If $i = j$ then $R(V_i, V_j)X_k = 0$. For $i \neq j$, recall, from section 1.2, that

$$\nabla_{V_i}V_k = \sum_{j=1}^{n} \Gamma_{ik}^{j} V_j$$

Substituting the Christofell's symbols (3.1) we get

$$\nabla_{V_i}V_k = -\frac{2x_i - 1}{x_i(1-x_i)}\delta_{ik}V_i \tag{3.3}$$

Therefore:

$$\nabla_{V_j}(\nabla_{V_i}V_k) = \nabla_{V_j}\left(-\frac{2x_i - 1}{x_i(1-x_i)}\delta_{ik}V_i\right).$$

We have the following cases to analyze. If $i \neq k$, it is immediate that $\nabla_{V_j}(\nabla_{V_i}V_k) = 0$. Otherwise, if $i = k$, and $j \neq k$, we obtain, applying (3.3)

$$\nabla_{V_j}(\nabla_{V_k}V_k) = \nabla_{V_j}\left(-\frac{2x_k - 1}{x_k(1-x_k)}V_k\right) = \frac{2x_k - 1}{x_k(1-x_k)}\delta_{jk}\frac{2x_j - 1}{x_j(1-x_j)}V_j = 0$$

Thus

$$\nabla_{V_j}(\nabla_{V_i}V_k) = 0$$

In the same way we have

$$\nabla_{V_i}(\nabla_{V_j}V_k) = 0$$

Those results lead to:

$$R(V_i,V_j)X_k = 0, \quad i,j,k,l = 1,...n.$$

therefore $R(V,W)Z = 0$. Then, the Riemannian manifold $\langle 0,1\rangle^n$, endowed with the metric $X^{-2}(I-X)^{-2}$, has null curvature.

4 A self-concordant barrier

In this section, we show that the function B is a $(3/2)n$-self-concordant barrier for the hypercube $\langle 0,1\rangle^n$. Let recall the definition given by Nesterov and Nemirovskii, [18].

Definition 4.1 *Let E be a finite dimensional real vector space, $Q \subset E$ be a open nonempty convex subset of E and $B : Q \to I\!R$ a function. B is called self-concordant barrier on Q with parameter $c > 0$ (c-self-concordant barrier) if:*

1. *$B \in C^3$ is a barrier function on Q.*

2. *B is a convex function on Q.*

3. for any $x \in Q$ and $h \in E$:

$$|\nabla^3 B(x)[h,h,h]| \leq 2(\nabla^2 B(x)[h,h])^{\frac{3}{2}}$$

4. there exists $c > 0$ such that B satisfied:

$$|\nabla B(x)[h]| \leq c^{\frac{1}{2}}(\nabla^2 B(x)[h,h])^{\frac{1}{2}}$$

Theorem 4.1 *The function $B(x) = \sum_{i=1}^{n}(2x_i - 1)[\ln x_i - \ln(1-x_i)]$ is $(3/2)n$-self-concordant barrier on $\langle 0,1\rangle^n$.*

Proof. First, we consider the general term

$$b(z) = (2z-1)[\ln z - \ln(1-z)], \quad z \in \langle 0,1\rangle$$

The first, second and third order derivatives are, respectively:

$$b'(z) = 2\left[\ln z - \ln(1-z)\right] - \frac{1}{z} + \frac{1}{1-z}$$

$$b''(z) = \frac{1}{z^2(1-z)^2}$$

$$b'''(z) = \frac{2(2z-1)}{z^3(1-z)^3}$$

It easy to check that $b(z)$ satisfies the two first conditions, so we prove the remaining. As

$$\frac{|b'''(z)|}{2(b''(z))^{\frac{3}{2}}} = |2z-1| \leq 1.$$

we obtain that $b(z)$ satisfies the third condition (sometimes called 1-self-concordant condition). To prove the last condition, let $c = 3/2$, then:

$$\frac{|b'(z)|}{c^{\frac{1}{2}}(b''(z))^{\frac{1}{2}}} = \frac{|2\ln(\frac{z}{1-z}) + (2z-1)\frac{1}{z(1-z)}|}{c^{\frac{1}{2}}\frac{1}{z(1-z)}} = \frac{1}{c^{\frac{1}{2}}}\left|2z(1-z)\ln\left(\frac{z}{1-z}\right) + 2z - 1\right|.$$

Now, equaling to zero the derivative of the last expression, we get

$$(1-2z)\ln(\frac{z}{1-z}) + 2 = 0$$

We solve that equation by using Matlab, [15], obtaining the roots $z_1 = 0.0832217$ and $z_2 = 0.916779$, both leading to $c = (1.19968)^2$, which is smaller than $3/2$, as wanted. Due to the stability property with respect to the direct product, see Propositions (2.1.1) and (2.3.1) in [18], we obtain that the function B is a $(3/2)n$-self-concordant barrier on $\langle 0,1\rangle^n$ and, therefore, the proof is completed. ■

5 The Central Path

In this section we propose a new interior penalized problem to solve (2.1) and study the behavior of the (primal-dual) central path, obtained by its KKT optimality conditions. We will observe that a property of this path is its primal feasibility, with respect to (2.1), and dual infeasibility, with respect to the dual problem (2.2), by a term $\mu[\ln X(\mu) - \ln(I - X(\mu))]e$, that converges to zero when μ converges to zero.

To solve the problem (2.1) we propose a new penalized problem:

$$\begin{array}{ll} \min & \phi_B(x,\mu) = f(x) + \mu \sum_{i=1}^{n} (2x_i - 1)[\ln x_i - \ln(1-x_i)] \\ & Ax = b \\ & (0 < x < e) \end{array} \tag{5.1}$$

where $\mu > 0$ is a positive parameter. The first and second order derivatives of ϕ_B are:

$$g = g(x,\mu) := \nabla f(x) + 2\mu[\ln X - \ln(I-X)]e - \mu X^{-1}e + \mu(I-X)^{-1}e \tag{5.2}$$

$$H = H(x,\mu) := \nabla^2 f(x) + \mu X^{-2}(I-X)^{-2}. \tag{5.3}$$

Since $\phi_B(x,\mu)$ is strictly convex (because f is a convex function, and $X^{-2}(I-X)^{-2}$ is a strictly positive symmetric matrix) on the relative interior of the feasible set, and takes infinite values on the boundary of F, that function achieves the minimal value in its domain (for fixed μ) at a unique point. Let $x(\mu)$ that point, which is called the μ center. The necessary and sufficient first order KKT optimality conditions for $x(\mu)$ are:

$$A^T y + s - w = \nabla f(x) + 2\mu[\ln X - \ln(I-X)]e \tag{5.4}$$

$$Ax = b \tag{5.5}$$

$$Xs = \mu e \tag{5.6}$$

$$(I-X)w = \mu e \tag{5.7}$$

$$(s,w) \geq 0 \tag{5.8}$$

$$(0 < x < 1) \tag{5.9}$$

The unique solution of this system, denoted by $(x(\mu), y(\mu), s(\mu), w(\mu), \mu > 0)$, is called the primal-dual central path. Clearly, this path is primal feasible and dual infeasible with respect to the problems (2.1) and (2.2) respectively.

Lemma 5.1 *Let* $x(\mu_i) := \arg\min\{\phi_B(x,\mu_i) : \ Ax = b, 0 < x < e\}, i = 1,2,$ *and let* $0 < \mu_2 < \mu_1$*. Then the following is true:*

1. $\phi_B(x(\mu_2),\mu_2) < \phi_B(x(\mu_1),\mu_1)$
2. $\sum_{i=1}^{n} (2x_i(\mu_1) - 1)[\ln x_i(\mu_1) - \ln(1 - x_i(\mu_1))] \leq \sum_{i=1}^{n} (2x_i(\mu_2) - 1)[\ln x_i(\mu_2) - \ln(1 - x_i(\mu_2))]$
3. $f(x(\mu_2)) \leq f(x(\mu_1))$

Proof. Is sufficient to consider the fact that the barrier $B \geq 0$.

Lemma 5.2 *If* x^* *is an optimal solution to the problem (2.1), and* $\mu > 0$*, then*

$$f(x^*) \leq f(x(\mu)) \leq \phi_B(x(\mu),\mu)$$

Proof. Use the following fact: $f(x^*) \leq f(x(\mu))$, for any $\mu > 0$, and $B \geq 0$.

Proposition 5.1 *For all $r > 0$ the set $\{(x(\mu), s(\mu), w(\mu)) : \ \mu < r\}$ is bounded.*

Proof. As $x(\mu) \in \langle 0, 1\rangle^n$, the set $\{x(\mu)\}$ is bounded, particularly, when $\mu < r$. Now, we will prove that $\{(s(\mu), w(\mu)) : \ \mu < r\}$ is also bounded. We know that $(x(\mu), y(\mu), s(\mu), w(\mu))$ solves the equation

$$A^T y + s - w = \nabla f(x) + 2\mu[\ln X - \ln(I - X)]e$$

so, defining the Lagrangian

$$L(x, y, s, w) = f(x) - y^T(Ax - b) - s^T x + w^T(x - e) + 2\mu e^T[X \ln X + (I - X)\ln(I - X)]e$$

we have

$$\nabla_x L(x(\mu), y(\mu), s(\mu), w(\mu)) = 0 \qquad (5.10)$$

where ∇_x denotes the gradient of L with respect to the x variables. Now, observing that $L(., y(\mu), s(\mu), w(\mu)))$ is a strictly convex function on $\langle 0, 1\rangle^n$, (5.10) implies

$$x(\mu) \in \arg\min\{L(x, y(\mu), s(\mu), w(\mu))), \ x \in \langle 0, 1\rangle^n\}.$$

On the other hand, we have: $f(x(\mu)) - L(x(\mu), y(\mu), s(\mu), w(\mu))$
$= s(\mu)^T x(\mu) - w(\mu)^T(x(\mu) - e) - 2\mu e^T[X(\mu)\ln X(\mu) + (I - X(\mu))\ln(I - X(\mu))]e$
$= n\mu + n\mu - 2\mu e^T[X(\mu)\ln X(\mu) + (I - X(\mu))\ln(I - X(\mu))]e$
$\leq 2n\mu(1 + \ln 2)$
Then, due to assumption 1, we can take x^0 primal feasible, getting

$$\begin{aligned} f(x(\mu)) - 2n\mu(1 + \ln 2) &\leq L(x(\mu), y(\mu), s(\mu), w(\mu))) \\ &\leq L(x^0, y(\mu), s(\mu), w(\mu)) \\ &\leq f(x^0) - s(\mu)^T x^0 + w(\mu)^T(x^0 - e), \end{aligned}$$

the last inequality being a consequence of the feasibility of x^0, and $\sum_{i=1}^{n}[x_i^0 \ln x_i^0 + (1 - x_i^0)\ln(1 - x_i^0)] \leq 0$. Thus

$$f(x(\mu)) - 2n\mu(1 + \ln 2) \leq f(x^0) - s(\mu)^T x^0 + w(\mu)^T(x^0 - e) \qquad (5.11)$$

Since $0 < x_i^0 < 1, \ i = 1, .., n$, the quantity ξ defined by

$$\xi = \min\{x_i^0, \ 1 - x_i^0, \ i = 1, ..., n\}$$

is positive. Now, let f^* be the optimal value of the problem (2.1). From (5.11), and using $\mu \leq r$, we find:

$$\begin{aligned} ||s||_1 + ||w||_1 &\leq \frac{1}{\xi}(f(x^0) + 2n\mu(1 + \ln 2) - f(x(\mu))) \\ &\leq \tfrac{1}{\xi}(f(x^0) + 2n(1 + \ln 2)r - f^*) \end{aligned}$$

where $||.||_1$ is the 1-norm (for $u \in I\!R^n, ||u||_1 = \sum_{i=1}^{n} |u_i|$). Therefore the set $\{(x(\mu), s(\mu), w(\mu)) : \mu < r\}$ is bounded. ∎

We take notice of the fact that the functions $\ln(x_i(\mu)) - \ln(1 - x_i(\mu))$ being not bounded on the interval $\langle 0,1\rangle$, we are not able to prove the boundedness of the path $\{y(\mu), \mu > 0\}$.

In the remaining of the text we use the following notation:

$$(x^k, y^k, s^k, w^k) = (x(\mu_k), y(\mu_k), s(\mu_k), w(\mu_k))$$

The following theorem is a well known property for general positive barriers

Theorem 5.1 *Let (μ_k) be a sequence of positive real numbers such that $\mu_k \to 0$ as $k \to \infty$. Then*

1. *$\mu_k B(x^k, \mu_k) \to 0$, when $k \to \infty$.*
2. *Any cluster point of the sequence $\{x^k\}$ is an optimal solution of the problem (2.1)*

Proof. See Bazaara [2].

Corollary 5.1 *Let (x^k), $k \in M \subset I\!N$ be a subsequence such that $x^k \to x^*$, where x^* is an optimal solution of the problem (2.1). Then*

$$\mu_k[\ln x_i^k - \ln(1 - x_i^k))] \to 0,\ k \to \infty,\ i = 1, ..., n$$

Proof. Taking in mind that $b(x_i^k) = (2x_i^k - 1)(\ln x_i^k - \ln(1 - x_i^k)) \geq 0$, the previous theorem leads to

$$\mu_k \sum_{i=1}^{n} \Big[(2x_i^k - 1)(\ln x_i^k - ln(1 - x_i^k))\Big] \to 0,\ k \to \infty,$$

so

$$\mu_k b(x_i^k) \to 0,\ k \to \infty,\ i = 1, ..., n,$$

or, equivalently

$$2\mu_k[x_i^k \ln x_i^k + (1 - x_i^k)\ln(1 - x_i^k)] - \mu_k \ln x_i^k - \mu_k \ln(1 - x_i^k) \to 0,\ k \to \infty,\ i = 1, ..., n \quad (5.12)$$

The sequence $\gamma(x_i^k) := x_i^k \ln x_i^k + (1 - x_i^k)\ln(1 - x_i^k)$ is bounded. Now, as $\mu_k \to 0,\ k \to \infty,\ i = 1, ..., n$, we have

$$2\mu_k \gamma(x_i^k) \to 0,\ k \to \infty,\ i = 1, ..., n. \quad (5.13)$$

Subtracting the expression (5.13) from (5.12), as both are convergent sequences, it gives

$$-\mu_k \ln x_i^k - \mu_k \ln(1 - x_i^k) \to 0,\ k \to \infty,\ i = 1, ..., n$$

Besides, we have $x_i^k \in \langle 0,1\rangle^n$, then $-\mu_k \ln x_i^k > 0$ and $-\mu_k \ln(1 - \ln x_i^k) > 0$, $i = 1, ..., n$. So

$$\mu_k \ln x_i^k \to 0,\ k \to \infty,\ i = 1, ..., n \quad (5.14)$$

$$-\mu_k \ln(1 - x_i^k) \to 0,\ k \to \infty,\ i = 1, ..., n. \quad (5.15)$$

Adding (5.14) and (5.15) we conclude that

$$\mu_k[\ln x_i^k - \ln(1 - x_i^k))] \to 0,\ k \to \infty,\ i = 1, ..., n \quad \blacksquare$$

6 Convergence of the Primal Central Path

In this section we prove that the primal central path converges, that is, there exists the limit of $x(\mu)$ when μ goes to zero. Furthermore we are able to characterize this limit point: it is the analytic center of the optimal solution set $\texttt{sol}(P)$. To prove this fact we will use the assumption 2. Additionally, we prove (applying the implicit function theorem) that $(x(\mu), y(\mu), s(\mu), w(\mu))$, for $\mu > 0$, is continuously differentiable.

The analytic center of the optimal set is defined as the solution of

$$\begin{array}{l} \min \sum_{j\in J}(2x_i - 1)[\ln x_j - \ln(1-x_j)] \\ \qquad x \in \texttt{ri}(\texttt{sol}(P)) \end{array} \tag{6.1}$$

where $\texttt{ri}(\texttt{sol}(P))$ is the relative interior of the optimal solution set of the problem (2.1), and $J = \{j \in \{1, 2, ..., n\} : \exists\, x \in \texttt{sol}(P),\ 0 < x_j < 1\}$. Observe that, from the convexity of $\texttt{sol}(P)$, it follows that

$$\texttt{ri}(\texttt{sol}(P)) = \{x \in \texttt{sol}(P) :\ 0 < x_J < 1\}$$

If $\texttt{sol}(P)$ is an unitary set, the unique solution is an extreme point of $[0,1]^n$, then $J = \emptyset$, and, by convention $\sum_{j\in J}(2x_j - 1)[\ln x_j - \ln(1-x_j)] := 0$. Notice that the objective function in (6.1) is strictly convex on $\texttt{ri}(\texttt{sol}(P))$, and diverges on the relative boundary of $\texttt{sol}(P)$. Then, due to the property that $\texttt{sol}(P)$ is bounded, the analytic center is well defined.

At this point, we will prove that the primal central path converges to the analytic center of the optimal set. The arguments are similar to those used for the logarithmic barrier, in Graña Drummond and Svaiter [12].

Theorem 6.1 *The primal central path $x(\mu)$ converges to the analytic center of $\texttt{sol}(P)$*

Proof.

We are interest in the case when $J \neq \emptyset$ (If $J = \emptyset$, $\texttt{sol}(P)$ is a unitary set and the unique solution is an extreme point, so Theorem 5.1, 2, implies the convergence of $x(\mu)$ to that point). As the primal central path $\{x(\mu)\}$ is bounded, there exists a cluster point x^*. Let (μ_k) be the sequence of positive numbers such that

$$\mu_k \to 0, \quad \texttt{and} \quad x(\mu_k) \to x^* \ \texttt{when} \ k \to \infty$$

Denote $x^k = x(\mu_k)$. Let $\bar{x}$ be the solution of (6.1). We will prove that $0 < x^*_J < 1$ and $\sum_{j\in J}(2x^*_j - 1)[\ln x_j{}^* - \ln(1-x^*_j)] \le \sum_{j\in J}(2\bar{x}_j - 1)[\ln \bar{x}_j - \ln(1-\bar{x}_j)]$. The first condition implies that x^* is feasible, the second one implies that x^* is a minimum for (6.1). Defining the sequence $p^k = \bar{x} + x^k - x^*$ we immediately obtain that $p^k \to \bar{x}$ when $k \to \infty$. First, we will prove that for k large enough, p^k is an strictly feasible solution for problem (6.1). Clearly, x^k being feasible and x^* an optimal solution, (see Theorem 5.1), we have that $Ap^k = A(\bar{x} + x^k - x^*) = b$, for all k. Now, suppose $j \notin J$. We are going to show that $\bar{x}_j = x^*_j = 0$ $\texttt{or}$ $\bar{x}_j = x^*_j = 1$. Indeed, suppose $\bar{x}_j \neq x^*_j$, for some $j \notin J$. Clearly, $z(t) = x^* + t(\bar{x} - x^*)$ solves (P) for any $t \in [0,1]$.Therefore, $f(z(t)) = f((1-t)x^* + t\bar{x}) \le (1-t)f(x^*) + tf(x^*) = f(x^*)$. For the j index above, we have $z_j(t) = x^*_j + t(\bar{x}_j - x^*_j)$. If $\bar{x}_j = 0$, then, from the hypothesis, $x^*_j = 1$, and $z_j(t) = 1 - t$, for any $t \in [0,1]$. We can take $t = 1/2$, getting $z(1/2) = x^* + (\bar{x} - x^*)/2$ as a solution for P. Now, observe that $0 < z_j(1/2) < 1$, which means that $j \in J$, a contradiction. The proof is analogous if we let $\bar{x}_j = 1$ and $x^*_j = 0$.

Thus, $x_j^k = p_j^k$, for $j \notin J$, and $0 < p_j^k < 1$, because $0 < x_j^k < 1$. On the other hand, for $j \in J$, since $x_j^k \to x_j^*$ when $k \to \infty$, and $0 < \bar{x}_j < 1$, then $0 < p_j^k < 1$ for k large enough. We conclude that $0 < p^k < e$ for k large enough.
Now, we will prove that

$$f(p^k) = f(x^k).$$

Consider $f(x^* + t(\bar{x} - x^*))$ as a function of $t \in [0, 1]$. Since x^* and $\bar{x}$ belong to the convex set $\texttt{sol}(P)$ and also using the convexity of f, it follows that $f(x^* + t(\bar{x} - x^*))$ is constant. Hence,

$$\nabla f(x^*)^T(\bar{x} - x^*) = 0,$$

$$(\bar{x} - x^*)^T \nabla^2 f(x^*)(\bar{x} - x^*) = 0$$

Since the Hessian of f is a symmetric positive semi-definite matrix, it follows that

$$\nabla^2 f(x^*)(\bar{x} - x^*) = 0,$$

thus $\bar{x} - x^* \in \texttt{ker}(\nabla^2 f(x^*))$. In view of the assumption 2 (see section 2), it is true that $\bar{x} - x^* \in W$, where W is some subspace of $I\!R^n$. Additionally, by Taylor integral formula, we have

$$\nabla f(x^k) = \nabla f(x^*) + \int_0^1 \nabla^2 f(x^* + t(x^k - x^*))(x^k - x^*)dt.$$

We have also $\nabla^2 f(x^* + t(x^k - x^*))(x^k - x^*) \in \texttt{range}(\nabla^2 f(x^* + t(x^k - x^*)) = (\ker(\nabla^2 f(x^* + t(x^k - x^*)))^T)^\perp = (\ker(\nabla^2 f(x^* + t(x^k - x^*))))^\perp = W^\perp$, again, due to assumption 2. It follows that $\int_0^1 \nabla^2 f(x^* + t(x^k - x^*))(x^k - x^*)dt \in W^\perp$, so,

$$\nabla f(x^k)^T(\bar{x} - x^*) = 0. \tag{6.2}$$

Finally, using the differential characterization of convex functions, the definition of p^k, and (6.2), we get

$$\begin{aligned} f(p^k) &\geq f(x^k) + \nabla f(x^k)^T(p^k - x^k) \\ &= f(x^k) + \nabla f(x^k)^T(\bar{x} - x^*) \\ &= f(x^k). \end{aligned}$$

Thus

$$f(p^k) \geq f(x^k) \tag{6.3}$$

Similarly, interchanging the roles of x^k and p^k, we have

$$f(x^k) \geq f(p^k)$$

which, with (6.3), leads to

$$f(x^k) = f(p^k).$$

Now, since $Ap^k = b$ and $0 < p^k < e$, for k large enough, optimality property of x^k furnishes, for k sufficiently large,

$$f(x^k) + \mu_k \sum_{i=1}^n (2x_i^k - 1)[\ln x_i^k - \ln(1 - x_i^k)] \leq f(p^k) + \mu_k \sum_{i=1}^n (2p_i^k - 1)[\ln p_i^k - \ln(1 - p_i^k)].$$

Next, as $p_j^k = x_j^k$, for all $j \notin J$, and $f(x^k) = f(p^k)$, we have, also for k large enough:

$$\sum_{j \in J} (2x_j^k - 1)[\ln x_j^k - \ln(1 - x_j^k)] \leq \sum_{j \in J} (2p_j^k - 1)[\ln p_j^k - \ln(1 - p_j^k)]$$

Taking limits as k goes to infinity in the last inequality, we conclude that $0 < x_j^* < 1$, for $j \in J$ and

$$\sum_{j \in J} (2x_j^* - 1)[\ln x_j^* - \ln(1 - x_j^*)] \leq \sum_{j \in J} (2\bar{x}_j - 1)[\ln \bar{x}_j - \ln(1 - \bar{x}_j)]$$

Since $\bar{x}$ is unique, we conclude that $x^* = \bar{x}$, therefore all cluster points of the primal central path $\{x(\mu), \mu > 0\}$ are equal to analytic center of $\mathtt{sol}(P)$, and the limit of $(x(\mu))$ exists, and solves the problem (2.1). ■

7 Application to Linear Optimization

In this section we apply the new barrier to linear optimization problems. We present a long step path following algorithm and give an upper bound for the total number of Newton iterations, in order to obtain an ϵ-optimal solution, achieving a complexity of $O(n \ln(n\mu_0/\epsilon))$ iterations, where μ_0 is the starting algorithm parameter.

Consider the problem (2.1) when the objective is a linear function:

$$\begin{array}{rl} \min & c^T x \\ & Ax = b \\ & 0 \leq x \leq e \end{array} \tag{7.1}$$

The assumptions are the same as given in the section 2, but we should observe that, due to the linearity of the objective function, the assumption 2 is immediately satisfied. Besides, the function $c^T x$ is continuous on the compact set F, so, there exists a minimum point to the problem. Furthermore, the set of optimal solutions $\mathtt{sol}(P)$ is nonempty and bounded. The dual formulation of (2.1) is:

$$\begin{array}{rl} \max & d(y, s, w) = b^T y - w^T e \\ & A^T y + s - w = c \\ & w, s \geq 0 \\ & y \in \mathbb{R}^m \end{array} \tag{7.2}$$

where y, s are $m-$ and $n-$ dimensional vectors, respectively. Now, we can compute the duality gap. From (7.1) and (7.2), it is given by:

$$\mathtt{gap} := c^T x - (b^T y - w^T e) = e^T (Xs + (I - X)w). \tag{7.3}$$

It follows that $d(y, s, w) \leq z^* \leq c^T x$, where z^* denotes the optimal objective value in (7.1). It is also well known that if x^* is optimal in (7.1), then there exists (y^*, s^*, w^*) that is optimal for (7.2), with $c^T x^* = d(y^*, s^*, w^*) = z^*$. Those solutions are characterized by the Karush-Kuhn-Tucker (KKT) conditions:

$$A^T y + s - w = c \tag{7.4}$$

$$Ax = b \tag{7.5}$$

$$Xs = 0 \tag{7.6}$$

$$(I - X)w = 0 \tag{7.7}$$

$$(s, w) \geq 0 \tag{7.8}$$

$$0 \leq x \leq e \tag{7.9}$$

Now, we consider the barrier function for the primal problem (7.1):

$$\phi_B(x,\mu) = \frac{1}{\mu}(c^T x) + \sum_{i=1}^{n}(2x_i - 1)[\ln x_i - \ln(1-x_i)] \tag{7.10}$$

where μ is a positive parameter. The first and second order derivatives of ϕ_B are:

$$g = g(x,\mu) := \frac{c}{\mu} + 2[\ln X - \ln(I-X)]e - X^{-1}e + (I-X)^{-1}e \tag{7.11}$$

$$H = H(x,\mu) := X^{-2}(I-X)^{-2}. \tag{7.12}$$

Since $\phi_B(x,\mu)$ is strictly convex on the relative interior of the feasible set and takes infinite values on the boundary of F, that function achieves the minimal value in its domain (for fixed μ) at a unique point $x(\mu)$. The necessary and sufficient first order optimality conditions (KKT conditions) for $x(\mu)$ are:

$$\begin{aligned}
A^T y + s - w &= c + 2\mu[\ln X - \ln(I-X)]e && (7.13)\\
Ax &= b && (7.14)\\
Xs &= \mu e && (7.15)\\
(I-X)w &= \mu e && (7.16)\\
(s,w) &\geq 0 && (7.17)\\
(0 < x &< e) && (7.18)
\end{aligned}$$

The unique solution of this system is again denoted by $(x(\mu), y(\mu), s(\mu), w(\mu))$. The duality gap in this solution satisfies:

$$\texttt{gap}(\mu) := c^T x(\mu) - d(x(\mu), y(\mu), s(\mu), w(\mu)) = 2\mu(n - e^T[\ln X(\mu) - \ln(I - X(\mu))]x(\mu)).$$

We have proved that the term $\mu[\ln X - \ln(I-X)]e$ converges to zero and the primal central path converges to the analytic center of the solution set $\texttt{sol}(P)$ (see Corollary 5.1 and theorem 6.1), so the gap converges to zero when $\mu \to 0$. The following lemma states that the primal objective function decreases along the primal path and the dual objective increases along the dual path. In the proof we will use the following primal-dual barrier function for (7.2):

$$\phi_B^d(x,y,w,s,\mu) = \frac{1}{\mu}(-y^T b + w^T e) - \sum_{i=1}^{n}\ln s_i - \sum_{i=1}^{n}\ln w_i - 2\sum_{i=1}^{n}\ln(1-x_i)$$

and the dual problem

$$\begin{aligned}
\min \quad & \phi_B^d(x,y,w,s,\mu)\\
& A^T y + s - w = c + 2\mu[\ln X - \ln(I-X)]e
\end{aligned} \tag{7.19}$$

We observe that the barrier is only convex, then the minimum value can be achieved in different points.

Lemma 7.1 *If μ decreases, then the objective function $c^T x(\mu)$ of the primal problem (7.1) is monotonically decreasing, and the objective function $d(x(\mu), y(\mu), w(\mu))$ of the dual problem (7.2) is monotonically increasing.*

Proof. The first part is a classical result of Fiacco and McCormick [8] (see also Lemma 5.1). To prove the second part, first note that $\phi_B^d(x, y, w, s, \mu)$ is a convex function. The Karush-Kuhn-Tucker conditions for a minimizing point are:

$$\begin{aligned} A^T y + s - w &= c + 2\mu[\ln X - \ln(I - X)]e && (7.20)\\ A(\mu S^{-1}e) &= b && (7.21)\\ (I - \mu S^{-1})e &= \mu W^{-1}e && (7.22)\\ (I - X)^{-1}e &= \mu(X^{-1} + (I - X)^{-1})S^{-1}e && (7.23) \end{aligned}$$

where $s > 0$, $w > 0$ and $0 < x < e$. Indeed, the Lagrangian function for (7.19) is:
$L(x, y, w, s, v) =$

$$\frac{1}{\mu}(-y^T b + w^T e) - \sum_{i=1}^{n} \ln s_i - \sum_{i=1}^{n} \ln w_i - 2\sum_{i=1}^{n} \ln(1 - x_i) - v^T(A^T y + s - w - c - 2\mu[\ln X - \ln(I - X)]e),$$

from which we write the necessary and sufficient optimality KKT conditions:

$$\begin{aligned} A^T y + s - w - c - 2\mu[\ln X - \ln(I - X)]e = 0 && (7.24)\\ -b/\mu - Av = 0 && (7.25)\\ e/\mu - W^{-1}e + v = 0 && (7.26)\\ -S^{-1}e - v = 0 && (7.27)\\ 2(I - X)^{-1}e + 2\mu(X^{-1} + (I - X)^{-1})v = 0 && (7.28) \end{aligned}$$

where $s > 0$, $w > 0$ and $0 < x < e$. That system rewrites:

$$\begin{aligned} A^T y + s - w &= c + 2\mu[\ln X - \ln(I - X)]e && (7.29)\\ A(-\mu v) &= b && (7.30)\\ Ie - (-\mu v) &= \mu W^{-1}e && (7.31)\\ -\mu v &= \mu S^{-1}e && (7.32)\\ (I - X)^{-1}e &= -\mu(X^{-1} + (I - X)^{-1})v && (7.33) \end{aligned}$$

where $s > 0$, $w > 0$ and $0 < x < e$. Substituting (7.32) in (7.30), (7.31) and (7.33), one gets, respectively, $A(\mu S^{-1}e) = b$, $(I - \mu S^{-1})e = \mu W^{-1}e$ and $(I - X)^{-1}e = \mu(X^{-1} + (I - X)^{-1})S^{-1}e$. Thus, we achieve the KKT conditions (7.20), (7.21), (7.22), (7.23).
Now, we prove that the central path $(x(\mu), y(\mu), s(\mu), w(\mu))$ minimizes $\phi_B^d(x, y, w, s, \mu)$.
We know that the central path KKT conditions are:

$$\begin{aligned} A^T y + s - w &= c + 2\mu[\ln X - \ln(I - X)]e && (7.34)\\ Ax &= b && (7.35)\\ Xs &= \mu e && (7.36)\\ (I - X)w &= \mu e && (7.37)\\ (s, w) &\geq 0 && (7.38)\\ (0 < & \; x < 1) && (7.39) \end{aligned}$$

¿From (7.36) we have $x(\mu) = \mu S(\mu)^{-1}e$. Substituting this last expression in (7.35) and (7.37), it gives, respectively,

$$A(\mu S(\mu)^{-1}e) = b$$

and

$$(I - \mu S^{-1}(\mu))e = \mu W(\mu)^{-1}e.$$

Finally, refereing to (7.23), we get $(I - X(\mu))^{-1}e - \mu(X^{-1}(\mu) + (I - X(\mu))^{-1})S^{-1}(\mu)e = (I - X(\mu))^{-1}e - e - (I - X(\mu))^{-1}X(\mu)e = (I - X(\mu))^{-1}(I - X(\mu))e - e = 0$. Thus

$$(I - X(\mu))^{-1}e = \mu(X^{-1}(\mu) + (I - X(\mu))^{-1})S^{-1}(\mu)e$$

Therefore the central path satisfies the optimality conditions (7.20), (7.21), (7.22) and (7.23). Use again the classical result of Fiacco and McCormick to prove the second part of the lemma. ■

In the algorithm, approximate line search along projected Newton directions are carried out to minimize ϕ_B for fixed μ, i.e., the directions correspond to exact minimization of the quadratic approximation to ϕ_B on the affine space $Ax = b$. This means that the Newton direction $p = p(x, \mu)$ is determined as the solution of

$$g + Hp = A^T\lambda \tag{7.40}$$

$$Ap = 0 \tag{7.41}$$

seen as a linear system of equations in p and λ, where H is given in 7.12. As usual, we will use the $||q||_H = \sqrt{q^T H q}$, $q \in I\!R^n$ to measure distances between points. Clearly, $||.||_H$ defines a norm because H is positive definite.

7.1 Properties Near the Central Path

In this subsection we deal with some lemmas which are needed to obtain an upper bound for the total number for outer and inner iterations. The following two Lemmas are in essence due to Nesterov and Nemirovski [18](see also den Hertog [4], for a simplified version of the proofs).

Lemma 7.2 *Let $x \in \langle 0, 1\rangle^n$ and $d \in I\!R^n$ arbitrary. If $||d||_{H(x,\mu)} < 1$, then $x + d \in \langle 0, 1\rangle^n$.*

Lemma 7.3 *Let $x \in F^0$, p be the projected Newton direction and $x^+ = x + p$. If $||p||_H < 1$ then $x^+ \in F^0$ and*

$$||p||_{H(x^+,\mu)} \leq \frac{1}{(1 - ||p||_H)^2}||p||_H^2$$

For $||p||_H < \frac{3-\sqrt{5}}{2}$ this implies that $||p(x^+, \mu)|| < ||p||_H$, thus ensuring the convergence of Newton's method. For $||p||_H < \frac{1}{3}$, it holds that

$$||p||_{H(x^+,\mu)} \leq \frac{9}{4}||p||_H^2 \tag{7.42}$$

The following Lemma gives an upper bound for the difference in barrier value in a nearly centered point x and the center $x(\mu)$.

Lemma 7.4 *If the projected Newton direction p satisfies $||p||_H < \frac{1}{3}$, then*

$$\phi_B(x, \mu) - \phi_B(x(\mu), \mu) \leq \frac{||p||_H^2}{1 - (\frac{9}{4})^2||p||_H^2} \tag{7.43}$$

Proof. The barrier function is convex in x, whence

$$p^T g + \phi_B(x,\mu) \le \phi_B(x+p,\mu)$$

Now using $g = A^T\lambda - Hp$ and $Ap = 0$ we have

$$p^T g = -p^T Hp = -||p||_H^2$$

Substitution gives

$$\phi_B(x,\mu) - \phi_B(x+p,\mu) \le -p^T g = ||p||_H^2 \tag{7.44}$$

Now let $x^0 := x$, $p^0 := p$ and let $x^1, x^2, ...,$ denote the sequence of points obtained by repeating Newton steps $x^k = x^{k-1} + p^{k-1}$, starting at x^0, with $k = 0, 1, 2, ...$ By Lemma 7.3 we have

$$||p^k||_H \le \frac{4}{9}\left(\frac{9}{4}||p||_H\right)^{2^k}$$

So, using (7.44) we may write

$$\begin{aligned}\phi_B(x,\mu) - \phi_B(x(\mu),\mu) &= \sum_{k=0}^{\infty}\left(\phi_B(x^k,\mu) - \phi_B(x^{k+1},\mu)\right)\\ &\le \sum_{k=0}^{\infty} ||p^k||_H^2\\ &\le \sum_{k=0}^{\infty}\left(\tfrac{4}{9}\left(\tfrac{9}{4}||p||_H\right)^{2^k}\right)^2\\ &= (\tfrac{4}{9})^2 \sum_{k=0}^{\infty}\left(\tfrac{9}{4}||p||_H\right)^{2^{k+1}}\\ &\le \frac{||p||_H^2}{1-(\frac{9}{4})^2||p||_H^2} \quad \blacksquare\end{aligned}$$

The next Lemma gives an upper bound for the difference between the objective function values at a nearly centered point x and $x(\mu)$.

Lemma 7.5 *If the projected Newton direction p satisfies $||p||_H < \frac{1}{3}$ then*

$$|c^T x - c^T x(\mu)| \le \frac{1+\frac{9}{4}||p||_H}{1-\frac{9}{4}||p||_H}\mu\sqrt{\frac{3n}{2}}||p||_H \tag{7.45}$$

Proof. As $g = c/\mu + \nabla B(x)$, we have

$$c^T(x+p) - c^T x = c^T p = \mu(g^T p - \nabla B(x)^T p) = \mu(-||p||_H^2 - \nabla B(x)^T p).$$

Taking absolute values and using triangular inequality we derive

$$|c^T(x+p) - c^T x| \le \mu(||p||_H^2 + |\nabla B(x)^T p|)$$

Using the self-concordant barrier property $\left(|\nabla B(x)^T p| \le \sqrt{\frac{3}{2}n}||p||_H\right)$, it follows

$$|c^T(x+p) - c^T x| \le \mu(||p||_H^2 + \sqrt{\frac{3}{2}n}||p||_H)$$

$$\leq \sqrt{\tfrac{3}{2}n}\mu||p||_H(1+||p||_H)$$

Again, let $x := x^0$, $d := d^0$ e $x^1, x^2, ..., x^k$ denote the sequence of points obtained by repeating Newton steps, starting at x^0. We have

$$|c^T x(\mu) - c^T x| \leq \sum_{k=0}^{\infty} |c^T x^k - c^T x^{k+1}|$$

$$\leq \sqrt{\tfrac{3}{2}n}\mu \sum_{k=0}^{\infty} ||p^k||_H(1+||p^k||_H)$$

Using the property

$$||p^k||_H \leq \frac{4}{9}\left(\frac{9}{4}||p||_H\right)^{2^k} < \frac{9}{4}||p||_H$$

we obtain

$$|c^T x(\mu) - c^T x| \leq \sqrt{\frac{3}{2}n}\mu(1+\frac{9}{4}||p||_H)\sum_{k=0}^{\infty} ||p^k||_H$$

$$\leq \frac{1+\frac{9}{4}||p||_H}{1-\frac{9}{4}||p||_H}\sqrt{\tfrac{3}{2}n}\mu||p||_H \quad \blacksquare$$

We notice that if $||p||_H = 0$, then Lemmas 7.4 and 7.5 imply
$\phi_B(y,\mu) = \phi(x(\mu),\mu)$ and $c^T x = c^T x(\mu)$.

7.2 Long-step Path-Following Algorithm

Below we give a long step path following algorithm which we analyze in the next section. The scalar quantities τ, ϵ and θ are parameters that must be specified. In our case we use $\tau = 1/3$ because this condition guarantees the polynomial convergence of our algorithm. It is also required x^0, a starting interior solution of (7.1), and μ_0, an initial value of the barrier parameter. We now describe the algorithm for finding an ϵ-optimal solution.

Algorithm:

Imput

ϵ is the accuracy parameter;
$\tau = \frac{1}{3}$ is the proximity parameter;
θ is the reduction parameter $0 < \theta < 1$;
μ_0 is the initial barrier value;
x^0 is a given interior feasible point such that $||p(x^0,\mu_0)||_{H(x^0,\mu_0)} < \tau$

begin
    $x := x_0$; $\mu := \mu_0$;
    while $\mu > \frac{\epsilon}{6n}$ **do**
    begin (outer step)
        $\mu := (1-\theta)\mu$;
        while $||p||_H \geq \tau$ **do**
        begin(inner step)
            $\bar{\alpha} := argmin_{\alpha>0}\{\phi_B(x+\alpha p,\mu) : x+\alpha p \in F^0\}$
            $x := x + \bar{\alpha}\, p$
        end(inner step)
    end(outer step)
end.

The step length α used at inner iterations can be any value such that $x + \alpha p$ is an interior solution of (7.1), and a minimum decrease in $\phi_B(x, \mu)$ is ensured as given by Lemma 7.8, below. We also assume that the initial values (x^0, μ_0) satisfy the proximity criterion $||p||_H \leq \tau$. For finding an initial point that satisfies those assumptions we refer the reader to Renegar [24], Monteiro and Adler [16] and Guler et al.[11].

7.3 Complexity Analysis

In this section we obtain a complexity bound for the long step path following algorithm above. In the algorithm we will use $\tau = 1/3$. We start with a general self-concordant barrier property.

Lemma 7.6 *Let B be a ν-self-concordant barrier. Then for any $x, y \in \texttt{dom}B$, we have*

$$(y - x)^T \nabla B(x) \leq \nu \tag{7.46}$$

Proof. See Nesterov [17].

Lemma 7.7 *Let $\mu > 0$, and z^* the optimal cost for problem (7.1). We state that*

$$c^T x(\mu) - z^* \leq \frac{3}{2} n\mu \tag{7.47}$$

Proof. Due to previous Lemma, the property $3n/2$-self-concordant barrier for B, and the expression $g = c/\mu + \nabla B(x)$, leads to the aimed result. ■

The following theorem gives an upper bound for the number of outer iterations.

Theorem 7.1 *After at most*

$$\frac{1}{\theta} \ln \left(\frac{6n\mu_0}{\epsilon} \right) \tag{7.48}$$

outer iterations, the long step barrier algorithm ends with a primal solution verifying $c^T x - z^* \leq \epsilon$.

Proof. The algorithm stops when $\mu_K = (1 - \theta)^K \mu_0 \leq \epsilon/6n$. Equivalently, taking logarithms produces

$$-K \ln(1 - \theta) \geq \ln \frac{6n\mu_0}{\epsilon}$$

Since $\theta \leq -\ln(1 - \theta)$, this certainly holds if

$$K \geq \frac{1}{\theta} \ln \frac{6n\mu_0}{\epsilon}.$$

Now we prove that the difference between $c^T x$ and z^* is ϵ.

$$\begin{aligned} c^T x - z^* &= (c^T x - c^T x(\mu_k)) + (c^T x(\mu_k) - z^*) \\ &\leq \tfrac{1 + \frac{9}{4}||p||_H}{1 - \frac{9}{4}||p||_H} \mu_K \sqrt{\tfrac{3n}{2}} ||p||_H + \tfrac{3n}{2} \mu_K \\ &\leq \tfrac{3}{2} \mu_K (\tfrac{7}{3} \sqrt{n} + n) \end{aligned}$$

$$\leq \tfrac{\epsilon}{4n}(\tfrac{7}{3}\sqrt{n} + n)$$

$$\leq \epsilon,$$

where the first inequality is due to Lemma 7.5 and Lemma 7.7 respectively. ■

The following lemma adapts to our barrier a general self-concordant property.

Lemma 7.8 *Let* $\bar{\alpha} = (1 - ||p||_H)^{-1}$ *then*

$$\Delta = \phi_B(x, \mu) - \phi_B(x + \bar{\alpha}, \mu) \geq ||p||_H - \ln(1 + ||p||_H) > 0 \tag{7.49}$$

Proof.(den Hertog [4]).

The following theorem gives an upper bound for the total number of inner iterations in each outer iteration.

Theorem 7.2 *Each outer iteration requires at most*

$$\frac{22\theta}{(1-\theta)^2}\left(\frac{7}{3}\sqrt{\frac{3n}{2}} + \frac{3n}{2}\right) + \frac{22}{3} \tag{7.50}$$

inner iterations.

Proof. We denote the barrier parameter value in an arbitrary outer iteration by $\bar{\bar{\mu}}$, while the parameter value in the previous outer iteration is denoted by $\bar{\mu}$. The iterate at the beginning of the outer iteration is denoted by x. Hence x is centered with respect to $x(\bar{\mu})$, and $\bar{\bar{\mu}} = (1-\theta)\,\bar{\mu}$. By Lemma 7.8, during each inner iteration the decrease in the barrier function value is at least

$$\triangle = \frac{1}{3} - \ln(1 + \frac{1}{3}) > \frac{1}{22}.$$

Let N denote the number of inner iterations during one outer iteration. We have

$$\frac{N}{22} < N\triangle \leq \phi_B(x, \bar{\bar{\mu}}) - \phi_B(x(\bar{\bar{\mu}}), \bar{\bar{\mu}}). \tag{7.51}$$

Now, define the function

$$\Psi_B(x, \bar{\bar{\mu}}) := \phi_B(x, \bar{\bar{\mu}}) - \phi_B(x(\bar{\bar{\mu}}), \bar{\bar{\mu}})$$

According to the Mean Value Theorem there exists $\tilde{\mu} \in (\bar{\bar{\mu}}, \bar{\mu})$ such that

$$\Psi_B(x, \bar{\bar{\mu}}) = \Psi_B(x, \bar{\mu}) - \frac{d}{d\mu}\Psi_B(x, \mu)\Big|_{\mu=\tilde{\mu}}(\bar{\mu} - \bar{\bar{\mu}}) \tag{7.52}$$

Remembering that $\phi_B(x, \mu) = c^T x/\mu + B(x)$, we get

$$\frac{d}{d\mu}\phi_B(x, \mu) = -\frac{c^T x}{\mu^2},$$

and, denoting the derivative of $x(\mu)$ with respect to μ by x',

$$\frac{d}{d\mu}\phi_B(x(\mu), \mu) = -\frac{c^T x(\mu)}{\mu^2}$$

So

$$-\frac{d}{d\mu}\Psi_B(x,\mu)|_{\mu=\tilde{\mu}} = \left(\frac{c^Tx - c^Tx(\mu)}{\mu^2}\right)|_{\mu=\tilde{\mu}}$$
$$\le |\tfrac{c^Tx - c^Tx(\bar{\bar{\mu}})}{\bar{\bar{\mu}}^2}|$$

where the last inequality follows from fact that $\bar{\bar{\mu}} < \bar{\mu}$ and $c^Tx(\bar{\bar{\mu}}) \le c^Tx(\tilde{\mu})$.
Substituting this into (7.52) gives

$$\Psi_B(x,\bar{\bar{\mu}}) \le \Psi_B(x,\bar{\mu}) + \tfrac{c^Tx - c^Tx(\bar{\bar{\mu}})}{\bar{\bar{\mu}}^2}(\bar{\mu} - \bar{\bar{\mu}})$$

$$\le \Psi_B(x,\bar{\mu}) + \left(\frac{|c^Tx(\bar{\mu}) - c^Tx|}{\bar{\bar{\mu}}} + \frac{c^Tx(\bar{\mu}) - c^Tx(\bar{\bar{\mu}})}{\bar{\bar{\mu}}}\right)\frac{\bar{\mu} - \bar{\bar{\mu}}}{\bar{\bar{\mu}}} \tag{7.53}$$

Because $||p(x,\bar{\mu})||_{H(x,\bar{\mu})} < \frac{1}{3}$, Lemma 7.4 implies

$$\Psi_B(x,\bar{\mu}) \le \frac{1}{3}$$

Now, note that due to Lemma 7.5 and $||p||_H < \frac{1}{3}$:

$$|c^Tx - c^Tx(\bar{\mu})| \le \frac{7}{3}\sqrt{\frac{3}{2}n}\,\bar{\mu}$$

also we have

$$\frac{c^T(\bar{\mu}) - c^Tx(\bar{\bar{\mu}})}{\bar{\bar{\mu}}} \le \frac{c^Tx(\bar{\mu}) - z^*}{\bar{\bar{\mu}}} \le \frac{\frac{3}{2}n\,\bar{\mu}}{\bar{\bar{\mu}}}$$

Plugging all these upper bounds in (7.53) gives

$$\Psi_B(x,\bar{\bar{\mu}}) \le \frac{1}{3} + \left(\frac{7}{3}\sqrt{\frac{3}{2}n} + \frac{3}{2}n\right)\frac{\theta}{(1-\theta)^2}. \tag{7.54}$$

Substituting the last bound in (7.51) we have

$$N \le 22\left(\frac{1}{3} + \left(\frac{7}{3}\sqrt{\frac{3}{2}n} + \frac{3}{2}n\right)\frac{\theta}{(1-\theta)^2}\right)$$
$$\le \tfrac{22\theta}{(1-\theta)^2}\left(\tfrac{7}{3}\sqrt{\tfrac{3}{2}n} + \tfrac{3}{2}n\right) + \tfrac{22}{3} \quad \blacksquare$$

Finally, we can combine theorems 7.1 and 7.2, obtaining the total number of Newton iterations required for the algorithm:

Theorem 7.3 *An upper bound for the total number of Newton iterations is given by*

$$\left[\frac{22}{(1-\theta)^2}\left(\frac{7}{3}\sqrt{\frac{3}{2}n} + \frac{3}{2}n\right) + \frac{22}{3\theta}\right]\ln\left(\frac{6n\mu_0}{\epsilon}\right) \tag{7.55}$$

This makes clear that to obtain an $\epsilon-$optimal solution the algorithm needs $O(n\ln(n\mu_0/\epsilon))$ Newton iterations for the long-step variant ($0 < \theta < 1$).

8 Concluding Remarks

We introduced a new self-concordant barrier for the hypercube and we proved that the central path, under standard assumptions, converges to the analytic center of the optimal solution set. This barrier can be extended for the following class of Semidefinite Optimization. Let consider the problem: min $C \bullet X$ s.t: $A_i \bullet X = b_i$, $0 \preceq X \preceq I$, where $\bullet$ represents the matrix trace scalar product. That family is motivated by some problems, as the minimization of the sum of the largest eigenvalues of symmetric matrices, see Alizadeh [1]. Now, we propose the barrier:

$$B(X) = \mathtt{tr}\,[(2X - I)(\ln X - \ln (I - X))]$$

We are studying that barrier in Papa Quiroz et al., [22].

References

[1] F. Alizadeh, Interior Point Methods in Semidefinite Programming with Applications to Combinatorial Optimization

[2] M. S. Bazaara, H. D. Sherali, C. M. Shetty: Nonlinear Programming: Theory and Algorithms, 2nd ed., John Wiley and Sons, Inc., New York, 1993.

[3] J. X. Cruz Neto, L. L. Lima, P. R. Oliveira, Geodesic Algorithms in Riemannian Geometry, Balkan Journal of Geometry and its Applications, **3**, 2, pp. 89-100, 1998

[4] D. den Hertog, Interior Point Approach to Linear, Quadratic and Convex Programming, 1 ed. Kluwer Academic Publishers, 1992.

[5] D. den Hertog, C. Roos, T. Terlaky,Inverse Barrier Methods for Linear Programming, Report 91-27, Faculty of Technical Mathematics and Informatics,, Delf University of Technology, The Netherlands, 1991.

[6] M. P. do Carmo, Riemannian Geometry, Birkhausen, Boston, 1992.

[7] J. R. Eriksson, An iterative Primal-Dual Algorithm for Linear Programming, Report LiTH-MAT-R-1985-10, Departament of Mathematics, Linköping University, Linköping, Sweden, 1985.

[8] A. V. Fiacco, G. P. McCormick, Non Linear Programming: Sequential Unconstrained Minimization Techniques, 1 ed. New York, Wiley, 1968, Reprinted by SIAM Publication, 1990.

[9] K. R. Frisch, The Logarithm Potential Method of Convex Programming, Technical Report, University Institute of Economics, Oslo, Norway, 1955.

[10] D. Gabay, Minimizing a Differentiable Function over a Differentiable Manifold, Journal of Optimization Theory and Application, **37**, pp. 177-219, 1982.

[11] O. Güler, C. Roos, T. Terlaky, J.-Ph. Vial, Interior Point Approach to the Theory of Linear Programming, Technical Report 1992.3, Université de Geneve, Départament d'Economie Commerciale et Industrielle, Geneve, Switzerland, 1992.

[12] L. M. Graña Drummond, B. F. Svaiter, On Well Definedness of the Central Path, Journal of Optimization Theory and Applications, **102**, 2,pp. 223-237, 1999.

[13] N. Karmarkar, Riemannian Geometry Underlying Interior-Point Methods for Linear Programming, *Contemporary Mathematics*, v. 114, pp. 51-75, 1990.

[14] D. G. Luenberger, The Gradient Projection Method Along Geodesics, Management Science, **18**, 11, pp. 620-631, 1972.

[15] C. B. Moler, J. N. Little, S. Bangert: PC-Matlab Users Guide, The Math Works Inc., 20 N. Main St., Sherborn, Mass., 1983

[16] R. D. C. Monteiro, I.Adler, Interior Path Following Primal-Dual Algorithms, Part I:Linear Programming, Mathematical Programming **44**, 43-66.

[17] Y. Nesterov, Introductory Lectures on convex Programming, Volume I: Basic Course, July 2, 1998 (Published by Kluwer in 2001).

[18] Y. E. Nesterov, A. N. Nemirovskii, Interior-Point Polynomial Algorithms in Convex Programming, Philadelphia, Society for Industrial and Applied Mathematics, 1994.

[19] Y. E. Nesterov, M. J. Todd, On the Riemannian Geometry Defined by Self-Concordant Barrier and Interior-Point Methods, Foundations of Computational Mathematics, **2**, pp. 333-361, 2002.

[20] P. R. Oliveira, J. X. da Cruz Neto, A Unified View of Primal Methods Through Riemannian Metrics, Technical Report ES-363/95, PESC/COPPE, Federal University of Rio de Janeiro, 1995.

[21] E. A. Papa Quiroz, P. R. Oliveira, New Results on Linear Optimization Through Diagonal Metrics and Riemannian Geometry Tools, Technical Report, ES-645/04, PESC/COPPE, Federal University of Rio de Janeiro, 2004.

[22] E. A. Papa Quiroz, P. R. Oliveira, S. Silva Souza, A New Barrier for Semidefinite Programming, paper in preparation.

[23] R. A. Polyak, Modified Barrier Function, Mathematical Sciences Department, IBM Research Division, T.J. Watson Research Center, Yorktown Heights, New York, 1989.

[24] J. Renegar, A Polynomial-Time Algorithm, Based on Newton's Method, for linear programming, Mathematical Programming, **40**, pp.59-93, 1988.

[25] T. Rapcsák, Smooth Nonlinear Optimization in $I\!R^n$, Kluwer Academic Publishers,1996.

[26] C. Udriste, Convex Functions and Optimization Methods on Riemannian Manifolds, Kluwer Academic Publishers, 1997.