

Two Interior-Point Algorithms for a Class of Convex Programming Problems

Stefano Herzel ^{*} Michael J. Todd [†]

June 10, 1994

Abstract: This paper describes two algorithms for the problem of minimizing a linear function over the intersection of an affine set and a convex set which is required to be the closure of the domain of a strongly self-concordant barrier function. One algorithm is a path-following method, while the other is a primal potential-reduction method. We give bounds on the number of iterations necessary to attain a given accuracy.

^{*}Center for Applied Mathematics, Engineering and Theory Center, Cornell University, Ithaca, NY 14853. E-mail: herzel@macomb.tn.cornell.edu. Research supported in part by CNR Grant No. 106005\01\91352.

[†]School of Operations Research and Industrial Engineering, Cornell University, Ithaca, New York 14853. E-mail: miketodd@cs.cornell.edu. Research supported in part by NSF, AFOSR and ONR through NSF Grant DMS-8920550.

1 Introduction

The aim of this paper is to describe two algorithms for the problem of minimizing a linear function over the intersection of an affine set and a convex set. Of course, any convex programming problem can be expressed in this form; if necessary, add an additional variable to be minimized with the extra constraint that it be at least as great as the original objective function. The key restriction we make is that the convex set be the closure of the domain of a strongly self-concordant barrier function, as defined by Nesterov and Nemirovsky [6, 7]. In fact, this is not such a great restriction, as Nesterov and Nemirovsky show that many constraint sets that occur in practice arise in this way. Our algorithms are then polynomial; that is, given a suitable starting point, the number of iterations required to attain a certain accuracy ϵ is bounded by a polynomial function of a certain parameter associated with the barrier function times $\log(1/\epsilon)$. All iterates are feasible and lie in the interior of the convex set; this classifies our methods as being interior-point algorithms.

We chose this format for our convex programming problem to make it as close to the standard-form linear programming problem as possible – the only difference is that a convex set replaces the nonnegative orthant. This makes it easier to compare our methods, a path-following algorithm and a primal potential-reduction algorithm, to well-known interior-point methods for linear programming. However, it disguises the similarity of our algorithms to those of Nesterov and Nemirovsky [6, 7]. Herzel [5] describes how our path-following method can be viewed as a special case of their approach, although our analysis is different. Our approach can also be viewed as an extension of that of Freund and Todd [2], in which the convex set is defined by simple bounds, to a more general setting.

In Sections 2, 3, and 4, we define and investigate self-concordant functions and barriers and their convex conjugates. While the results are similar to those of Nesterov and Nemirovsky, our notation and development are influenced more by Renegar [8]. We give most of the proofs to make the paper self-contained. Included are propositions on the existence of minimizers for strongly self-concordant functions, on quadratic convergence in this case, and on error bounds for Taylor approximations of such functions and their gradients.

Section 5 describes the convex programming problem we study and its dual, while Section 6 gives results on projections. Section 7 defines and investigates the central trajectory. We show that if a primal solution is close to the central trajectory, then a feasible solution to the dual problem can be constructed and the associated duality gap suitably bounded. The path-following algorithm follows this trajectory closely. The potential-reduction method updates its lower bound on the optimal value only when it gets close to the central trajectory.

Sections 8 and 9 describe these two algorithms. The first is similar to algorithms of Gonzaga [3] and Roos and Vial [10] for linear programming or more generally Nesterov and Nemirovsky's path-following method [7], Chapter 3, and

uses the quadratic convergence result. The second extends the primal potential-reduction method of Gonzaga [4] and Freund [1]; see also Ye [11] and Nesterov and Nemirovsky [7], Chapter 4. It uses the Taylor approximation of the barrier function and the bound on the duality gap for points close to the central trajectory.

2 Self-Concordant Functions: the Set \mathcal{S}_a

In this section we describe a set of functions called \mathcal{S}_a that is a subset of the set of *self-concordant functions* studied by Nesterov and Nemirovsky [6, 7].

First we need to introduce some notation. Given $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{\infty\}$, we denote by \mathcal{D}_f the domain of f , that is

$$\mathcal{D}_f := \{x \in \mathbb{R}^n : f(x) < \infty\}.$$

Let $\nabla^k f(x)$ be the k -th derivative of f at x . When $k \leq 2$ we use the notation g_x and H_x for the gradient (with respect to the standard inner product) and the Hessian of f at x . We write $\nabla^k f(x)[h_1, h_2, \dots, h_k]$ for the value of the operator $\nabla^k f(x)$ when applied to $[h_1, h_2, \dots, h_k] \in \mathbb{R}^n \times \mathbb{R}^n \times \dots \times \mathbb{R}^n$. When $k \leq 2$ we use $g_x^T h$ instead of $\nabla f(x)[h]$ and $h_1^T H_x h_2$ instead of $\nabla^2 f(x)[h_1, h_2]$.

The following definition is just a slight modification of the definition of *strongly self-concordant functions* given by Nesterov and Nemirovsky [6, 7].

Definition 2.1 (The set \mathcal{S}_a) *The set \mathcal{S}_a , where $a > 0$, consists of the functions $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{\infty\}$ such that*

- *$f(x)$ is a C^3 function on its domain \mathcal{D}_f .*
- *\mathcal{D}_f is a nonempty open convex set.*
- *The Hessian matrix H_x is positive definite¹ for all $x \in \mathcal{D}_f$.*
- *For all $x \in \mathcal{D}_f$ and for all $h \in \mathbb{R}^n$, the following inequality holds:*

$$|\nabla^3 f(x)[h, h, h]| \leq 2a(h^T H_x h)^{3/2}. \quad (2.1)$$

- *the sets $B_t^f := \{x \in \mathcal{D}_f : f(x) \leq t\}$ are closed for all $t \in \mathbb{R}$.*

If $f \in \mathcal{S}_a$ then, for $k > 0$,

$$|\nabla^3(kf)(x)[h, h, h]| \leq \frac{1}{\sqrt{k}} 2a(h^T (kH_x) h)^{3/2};$$

¹Nesterov and Nemirovsky ([7], corollary 2.1.1) show that the condition of positive definiteness of H_x is equivalent to requiring strict convexity of f . However, for the sake of a more self-contained paper, we keep this seemingly less general definition.

therefore, by multiplying f by a suitable constant k we can always assume $a = 1$.

The condition of closedness of sets B_t^f for all $t \in \mathbb{R}$ is equivalent to asking that $f(x_i)$ diverges when $\{x_i\}_{i=1}^\infty$ is a sequence of points in \mathcal{D}_f converging to a boundary point.

Since the Hessian of f is positive definite everywhere, it can be used to associate a metric on \mathbb{R}^n to each element $x \in \mathcal{D}_f$ by defining the *local inner product*:

$$\langle u, v \rangle_x := u^T H_x v, \quad u, v \in \mathbb{R}^n, x \in \mathcal{D}_f \quad (2.2)$$

We denote the norm given by such an inner product by $\|\cdot\|_x$. Obviously we have

$$\|v\|_x = \|H_x^{1/2} v\|_2, \quad v \in \mathbb{R}^n, x \in \mathcal{D}_f. \quad (2.3)$$

Note that with respect to such a local norm, the gradient of f at $y \in \mathcal{D}_f$ is given by $H_x^{-1} g_y$ while the Hessian matrix is given by $H_x^{-1} H_y$.

The following theorem states that local norms of close points are close; it is due to Nesterov and Nemirovsky. Before stating it let's define the set

$$E_x := \{y \in \mathbb{R}^n : \|y - x\|_x < 1\}; \quad (2.4)$$

this is the open ellipsoid (ball in the local norm) of radius 1 centered at x .

Theorem 2.1 *Given $f \in \mathcal{S}_1$, $x \in \mathcal{D}_f$, then*

$$E_x \subseteq \mathcal{D}_f.$$

Moreover, for each $y \in E_x$ and for each $v \in \mathbb{R}^n$ the following relation holds:

$$-\|y - x\|_x \leq \frac{\|v\|_y}{\|v\|_x} - 1 \leq \frac{\|y - x\|_x}{1 - \|y - x\|_x}. \quad (2.5)$$

Proof. See Nesterov and Nemirovsky [7] Theorem 2.1.1. \square

Equation (2.5) is similar to a Lipschitz condition on the gradient of f : the next results show how it affects the Newton method when applied to search for a minimizer. We use the symbol n_x to denote the Newton step at x , i.e.

$$n_x := -H_x^{-1} g_x.$$

Before proving a proposition that shows a property closely related to quadratic convergence, we need to state a lemma, due to Renegar [8], that will be very useful in what follows.

Lemma 2.1 *If $f \in \mathcal{S}_1$ and $x, y \in \mathcal{D}_f$, then*

$$\sup_{v \neq 0} \left| 1 - \frac{\|v\|_y^2}{\|v\|_x^2} \right| = \|I - H_x^{-1} H_y\|_x \quad (2.6)$$

where the norm on the right hand side is the operator norm induced by $\|\cdot\|_x$. Moreover, if $y \in E_x$, then

$$\|I - H_x^{-1}H_y\|_x \leq \frac{1}{(1 - \|y - x\|_x)^2} - 1. \quad (2.7)$$

Proof. For any $v \in \mathbb{R}^n, v \neq 0$,

$$\begin{aligned} \left| 1 - \frac{\|v\|_y^2}{\|v\|_x^2} \right| &= \left| \frac{\|v\|_x^2 - \|v\|_y^2}{\|v\|_x^2} \right| \\ &= \frac{|\langle v, [I - H_x^{-1}H_y]v \rangle_x|}{\|v\|_x^2}. \end{aligned}$$

Equation (2.6) then follows by observing that the operator $I - H_x^{-1}H_y$ is self adjoint with respect to the local inner product $\langle \cdot, \cdot \rangle_x$. Equation (2.7) immediately follows from Theorem 2.1 and some easy algebra. \square

Proposition 2.1 *Let $f \in \mathcal{S}_1$ have a minimizer $z \in \mathcal{D}_f$. If $x \in \mathcal{D}_f$ is such that*

$$\|z - x\|_x < 1$$

and if y is given by

$$y := x + n_x,$$

then

$$\|z - y\|_x \leq \frac{\|z - x\|_x^2}{1 - \|z - x\|_x}. \quad (2.8)$$

Proof. Define $v := z - x$. Then

$$\begin{aligned} \|z - y\|_x &= \|v + H_x^{-1}g_x\|_x \\ &= \|v - H_x^{-1} \int_0^1 H_{x+tv} v dt\|_x \quad (\text{since } g_z = 0) \\ &= \left\| \int_0^1 [I - H_x^{-1}H_{x+tv}] v dt \right\|_x \\ &\leq \int_0^1 \|I - H_x^{-1}H_{x+tv}\|_x \|v\|_x dt \\ &\leq \|v\|_x \int_0^1 \left(\frac{1}{(1 - t\|v\|_x)^2} - 1 \right) dt, \end{aligned}$$

where the last inequality follows from Lemma 2.1.

The result then follows by observing that

$$\int_0^1 \left(\frac{1}{(1 - t\|v\|_x)^2} - 1 \right) dt = \frac{\|v\|_x}{1 - \|v\|_x} \quad (2.9)$$

and substituting this in the previous inequality. \square

Now we want to find an upper bound, in terms of the local norm, on the second order approximation: in order to do this we adapt an argument of Renegar [8].

Proposition 2.2 *If $f \in \mathcal{S}_1$, $x \in \mathcal{D}_f$ and $y \in E_x$, then*

$$\left| f(y) - \left[f(x) - \langle n_x, y - x \rangle_x + \frac{1}{2} \|y - x\|_x^2 \right] \right| \leq \frac{1}{3} \frac{\|y - x\|_x^3}{1 - \|y - x\|_x} \quad (2.10)$$

Proof. Let's set $v := y - x$ and observe that

$$\begin{aligned} f(y) - f(x) &= \int_0^1 \langle g_{x+tv}, v \rangle dt \\ &= \langle g_x, v \rangle + \int_0^1 \int_0^t \langle v, H_{x+sv} v \rangle ds dt \\ &= \langle g_x, v \rangle + \frac{1}{2} \|v\|_x^2 + \int_0^1 \int_0^t \langle v, [H_{x+sv} - H_x] v \rangle ds dt. \end{aligned}$$

Therefore

$$\begin{aligned} \left| f(y) - \left[f(x) - \langle n_x, v \rangle_x + \frac{1}{2} \|v\|_x^2 \right] \right| &\leq \int_0^1 \int_0^t | \langle v, [H_{x+sv} - H_x] v \rangle | ds dt \\ &\leq \int_0^1 \int_0^t | \langle v, [H_x^{-1} H_{x+sv} - I] v \rangle_x | ds dt \\ &\leq \|v\|_x^2 \int_0^1 \int_0^t \left(\frac{1}{(1 - s\|v\|_x)^2} - 1 \right) ds dt \end{aligned}$$

where the last inequality follows from Lemma 2.1. From equation (2.9) we get

$$\begin{aligned} \left| f(y) - \left[f(x) - \langle n_x, v \rangle_x + \frac{1}{2} \|v\|_x^2 \right] \right| &\leq \|v\|_x^3 \int_0^1 \frac{t^2}{1 - t\|v\|_x} dt \\ &\leq \frac{\|v\|_x^3}{1 - \|v\|_x} \int_0^1 t^2 dt \\ &\leq \frac{1}{3} \frac{\|v\|_x^3}{1 - \|v\|_x} \end{aligned}$$

from which the result follows. \square

In what follows we show that, if the Newton step is not too big, then f has a minimizer.

Proposition 2.3 *If $f \in \mathcal{S}_1$, $x \in \mathcal{D}_f$, and*

$$\|n_x\|_x \leq \frac{1}{9}, \quad (2.11)$$

then f has a minimizer $z \in \mathcal{D}_f$ and

$$\|z - x\|_x \leq 3\|n_x\|_x.$$

Proof. If $\|n_x\|_x = 0$ then x is the minimizer and the statement is true.

Let's suppose $\|n_x\|_x > 0$. For $\alpha \in (0, 1/\|n_x\|_x)$ and any y such that

$$\|y - x\|_x = \alpha\|n_x\|_x, \quad (2.12)$$

we have

$$f(y) - f(x) \geq -\langle n_x, y - x \rangle_x + \frac{1}{2}\|y - x\|_x^2 - \frac{1}{3} \frac{\|y - x\|_x^3}{1 - \|y - x\|_x} \quad (2.13)$$

$$\geq -\alpha\|n_x\|_x^2 + \frac{1}{2}\alpha^2\|n_x\|_x^2 - \frac{1}{3} \frac{\alpha^3\|n_x\|_x^3}{1 - \alpha\|n_x\|_x} \quad (2.14)$$

$$= \frac{\|n_x\|_x^2 \alpha [-5\alpha^2\|n_x\|_x + \alpha(6\|n_x\|_x + 3) - 6]}{6(1 - \alpha\|n_x\|_x)}, \quad (2.15)$$

where (2.13) follows from Proposition 2.2 and (2.14) from the Cauchy-Schwartz inequality, while (2.15) is just easy algebra.

Note that the vectors y satisfying (2.12) are on the boundary of the ellipsoid of center x and radius $\alpha\|n_x\|_x$; we want to determine a value of $\alpha \in (0, 1/\|n_x\|_x)$ such that the quantity on the right hand side of (2.15) is positive. The convexity of f would then imply the existence of a minimizer z inside the ellipsoid.

Let's observe that, for $\alpha \in (0, 1/\|n_x\|_x)$, the right hand side of (2.15) is positive if and only if

$$-5\alpha^2\|n_x\|_x + \alpha(6\|n_x\|_x + 3) - 6 > 0. \quad (2.16)$$

There exists some α satisfying (2.16) only if

$$\Delta := (3 + 6\|n_x\|_x)^2 - 120\|n_x\|_x > 0,$$

which holds when $\|n_x\|_x \leq 1/9$.

Therefore (2.16) holds for any $\alpha \in (0, 1/\|n_x\|_x)$ such that

$$\frac{3 + 6\|n_x\|_x - \sqrt{\Delta}}{10\|n_x\|_x} < \alpha < \frac{3 + 6\|n_x\|_x + \sqrt{\Delta}}{10\|n_x\|_x}. \quad (2.17)$$

Since $\Delta \geq (3 - 24\|n_x\|_x)^2 > 0$ for $\|n_x\|_x \leq 1/9$,

$$\bar{\alpha} := 3$$

satisfies (2.17) and $\bar{\alpha} \in (0, 1/\|n_x\|_x)$; thus the proof is complete. \square

Next we see another interesting property of the set \mathcal{S}_1 .

Proposition 2.4 *If $f \in \mathcal{S}_1$ is bounded below, then it has a minimizer $z \in \mathcal{D}_f$.*

Proof. We want to show that there exists $x \in \mathcal{D}_f$ such that $\|n_x\|_x \leq 1/9$; the result then follows from Proposition 2.3.

From (2.10) we know that, for all $y \in E_x$

$$\begin{aligned} \langle n_x, y - x \rangle_x - \frac{1}{2}\|y - x\|_x^2 - \frac{1}{3} \frac{\|y - x\|_x^3}{1 - \|y - x\|_x} &\leq f(x) - f(y) \\ &\leq f(x) - \inf_{w \in \mathcal{D}_f} f(w). \end{aligned} \quad (2.18)$$

Let's set

$$y := x + \frac{n_x}{2\|n_x\|_x}.$$

Obviously $y \in E_x$ and

$$\|y - x\|_x = \frac{1}{2};$$

using this in (2.18) we get

$$\frac{\|n_x\|_x}{2} \leq f(x) - \inf_{w \in \mathcal{D}_f} f(w) + \frac{1}{8} + \frac{1}{12}. \quad (2.19)$$

Let H be the subset of \mathcal{D}_f

$$H := \left\{ x \in \mathcal{D}_f : f(x) - \inf_{w \in \mathcal{D}_f} f(w) \leq \frac{1}{24} \right\};$$

note that it is nonempty because f is bounded below; moreover from (2.19) it follows that, for $x \in H$,

$$\|n_x\|_x \leq \frac{1}{12} + \frac{1}{4} + \frac{1}{6} = \frac{1}{2}.$$

Therefore, when $x \in H$, we can set $y := x + n_x$ and substitute into (2.18) to get

$$\frac{\frac{1}{2}\|n_x\|_x^2 - \frac{5}{6}\|n_x\|_x^3}{1 - \|n_x\|_x} < f(x) - \inf_{w \in \mathcal{D}_f} f(w). \quad (2.20)$$

Let's now consider a sequence $x_k \in H$ such that

$$\lim_{k \rightarrow \infty} f(x_k) = \inf_{w \in \mathcal{D}_f} f(w).$$

Then, since

$$\|n_{x_k}\|_{x_k} \leq \frac{1}{2} < \frac{3}{5},$$

from (2.20) it follows that

$$\lim_{k \rightarrow \infty} \|n_{x_k}\|_{x_k} = 0$$

and then, when k is big enough,

$$\|n_{x_k}\|_{x_k} < 1/9,$$

completing the proof. \square

Observe that, from the previous result, the barrier $f(x) := 1/x$ doesn't belong to the set \mathcal{S}_a for any a .

In the next section we investigate the properties of a subset of \mathcal{S}_1 which we call $\mathcal{S}_1(K)$.

3 Self-Concordant Barriers: the Set $\mathcal{S}_1(K)$

We now give a definition, similar to that of a *self-concordant barrier* in Nesterov and Nemirovsky, of a subset of \mathcal{S}_1 characterized by having the local norm of the Newton step uniformly bounded. Note that our K below corresponds to the square root of Nesterov and Nemirovsky's parameter θ . In applications, K is usually the square root of the dimension n or of the number of constraints defining the closure of \mathcal{D}_f ; see [7].

Definition 3.1 *The function f belongs to $\mathcal{S}_1(K)$, where $K > 0$, if*

- $f \in \mathcal{S}_1$,
- $|g_x^T h| \leq K(h^T H_x h)^{1/2}, \quad \forall x \in \mathcal{D}_f, h \in \mathbb{R}^n$.

The following proposition gives an interpretation of the Definition 3.1;

Proposition 3.1 *Let $f \in \mathcal{S}_1$; then $f \in \mathcal{S}_1(K)$ if and only if $\|n_x\|_x \leq K$ for all $x \in \mathcal{D}_f$.*

Proof. For the sufficiency note that

$$\begin{aligned} |g_x^T h| &= | \langle n_x, h \rangle_x | \\ &\leq \|n_x\|_x \|h\|_x. \end{aligned}$$

For the necessity, setting $h := H_x^{-1} g_x$, we get

$$K \|n_x\|_x = K \|h\|_x \geq g_x^T h = \|n_x\|_x^2.$$

The proof is complete. \square

The next proposition, due to Renegar [8], will be useful in what follows.

Proposition 3.2 *If $f \in \mathcal{S}_1(K)$ and $x, y \in \mathcal{D}_f$ then*

$$g_x^T(y - x) < K^2.$$

Proof. Let's consider the functional $\phi(t)$ defined, for $t \in \mathbb{R}$, by

$$\phi(t) := f((1-t)x + ty);$$

its domain is given by

$$\mathcal{D}_\phi = \{t \in \mathbb{R} : (1-t)x + ty \in \mathcal{D}_f\}.$$

We want to show that $\phi'(0) < K^2$. We can assume that $\phi'(0) > 0$; therefore, by convexity of ϕ , $\phi'(t) > 0$ for any $t \in [0, 1]$.

Since $f \in \mathcal{S}_1(K)$, it follows that

$$\phi''(t) \geq \left(\frac{\phi'(t)}{K} \right)^2.$$

Now we define

$$\gamma(t) := \frac{K^2 \phi'(0)}{K^2 - t \phi'(0)}.$$

It's easy to check that $\gamma(0) = \phi'(0)$ and

$$\gamma'(t) = \left(\frac{\gamma(t)}{K} \right)^2, \quad \forall t \in \left[0, \frac{K^2}{\phi'(0)} \right).$$

It follows that $\gamma(t) \leq \phi'(t)$ for any $t \in \mathcal{D}_\phi$ such that $0 \leq t < K^2/\phi'(0)$.

Since $\gamma(t) \rightarrow \infty$ when $t \rightarrow K^2/\phi'(0)$, we see that $K^2/\phi'(0)$ does not belong to \mathcal{D}_ϕ .

But since $1 \in \mathcal{D}_\phi$, then $\phi'(0) < K^2$, which is what we wanted to show. \square

From this result it follows that

Corollary 3.1 *If $f \in \mathcal{S}_1(K)$ then \mathcal{D}_f doesn't contain any line.*

Proof. Let's suppose instead that there exist $x \in \mathcal{D}_f$ and nonzero $v \in \mathbb{R}^n$ such that

$$x + tv \in \mathcal{D}_f, \quad \forall t \in \mathbb{R}$$

Then, setting $y_t := x + tv$ in Proposition 3.2 we get $tg_x^T v < K^2$, which holds for all $t \in \mathbb{R}$ if and only if $g_x^T v = 0$. By proceeding in the same way for $x := x + v$ we can show that $g_{x+v}^T v = 0$; but since f is convex this would imply that

$$H_{x+tv} = 0, \quad t \in (0, 1),$$

contradicting the hypothesis of positive definiteness of the Hessian. \square

Later we show how to modify the definitions given for \mathcal{S}_a and for $\mathcal{S}_1(K)$ in order to allow more general domains.

Another implication of Proposition 3.2 is that the derivative of f in the direction of a half line contained in \mathcal{D}_f tends to zero in the limit along such a half line.

Corollary 3.2 *If $f \in \mathcal{S}_1(K)$ and*

$$x + tv \in \mathcal{D}_f, \quad \forall t \geq 0$$

where $x \in \mathcal{D}_f$ and $v \in \mathbb{R}^n$, then

$$\lim_{t \rightarrow +\infty} g_{x+tv}^T v = 0.$$

Proof. From Proposition 3.2 it follows that

$$tg_{x+tv}^T v = g_{x+tv}^T(x + 2tv - (x + tv)) \leq K^2 \quad \forall t \geq 0,$$

therefore

$$\limsup_{t \rightarrow +\infty} g_{x+tv}^T v \leq 0. \quad (3.1)$$

But, again because of Proposition 3.2,

$$-tg_{x+tv}^T v = g_{x+tv}^T(x - (x + tv)) \leq K^2 \quad \forall t \geq 0,$$

which implies the same inequality as (3.1) for $-g_{x+tv}^T v$ and thus the result. \square

4 Conjugate Functions

In this section we define the convex conjugate function of $f \in \mathcal{S}_1$ and study some of its properties. The definition is the same as that given in Rockafellar [9].

Definition 4.1

$$f^*(s) := \sup_{x \in \mathcal{D}_f} s^T x - f(x). \quad (4.1)$$

First we want to identify the domain of f^* .

Proposition 4.1 *If $f \in \mathcal{S}_1$ and f^* is its conjugate function, then the domain of f^* is the range of the gradient of f , that is*

$$\mathcal{D}_{f^*} \equiv g(\mathcal{D}_f),$$

where

$$g(\mathcal{D}_f) := \{y \in \mathbb{R}^n : \exists x \in \mathcal{D}_f \text{ such that } g_x = y\}.$$

Proof. Let's define

$$F_s(x) := f(x) - s^T x.$$

Note that F_s belongs to \mathcal{S}_1 since f does. Observe that

$$\mathcal{D}_{f^*} := \{s : f^*(s) < \infty\} = \left\{s : \inf_x F_s(x) > -\infty\right\}.$$

Then, if $s = g_z$ for some $z \in \mathcal{D}_f$, it follows that

$$F_{g_z}(x) = f(x) - g_z^T x$$

is a convex function whose gradient vanishes at $x := z$ and which is therefore bounded below. This proves the inclusion

$$g(\mathcal{D}_f) \subseteq \mathcal{D}_{f^*}.$$

To prove the opposite inclusion let's note that if F_s belongs to \mathcal{S}_1 and is bounded below, then, from Proposition 2.4, it follows that

$$\text{there exists } z \in \mathcal{D}_f \text{ such that } \nabla F_s(z) = 0,$$

that is, $s = g_z$, concluding the proof. \square

We observe that \mathcal{D}_{f^*} is an open convex set.

Nesterov and Nemirovsky proved the following important result:

Theorem 4.1 *If $f \in \mathcal{S}_1$ then $f^* \in \mathcal{S}_1$.*

Proof. See Nesterov and Nemirovsky [7], Theorem 2.4.1. \square

From this theorem it follows that the Hessian of f^* is positive definite everywhere, so that for any $s \in \mathcal{D}_{f^*}$ we can define a local norm as follows:

$$\|v\|'_s := (v^T H_s^* v)^{1/2} \quad (4.2)$$

where H_s^* is the Hessian of f^* computed at s . Since $\mathcal{D}_{f^*} = g(\mathcal{D}_f)$, we find it convenient to use the notation

$$\|v\|_x^* := \|v\|'_s$$

where $s = g_x$.

To investigate the relationship between the two local norms, we recall (see Rockafellar [9] theorem 26.5) that

$$H_{g_x}^* = H_x^{-1}.$$

Therefore

$$\|v\|_x^{*2} = v^T H_x^{-1} v. \quad (4.3)$$

This equality will be frequently used in what follows.

Remark 4.1 Observe that, for $f \in \mathcal{S}_1(K)$, $\|g_x\|_x^* = \|n_x\|_x$, so that, from Proposition 3.1,

$$\|g_x\|_x^* \leq K.$$

Proposition 4.2 Let $f \in \mathcal{S}_1$ and $x \in \mathcal{D}_f$. Then, for all $h \in \mathbb{R}^n$,

$$\|h\|_x^* = \sup_{w \neq 0} \frac{h^T w}{\|w\|_x}. \quad (4.4)$$

Proof. From relation (4.3) we get

$$\begin{aligned} \|h\|_x^* &= \|H_x^{-1/2} h\|_2 \\ &= \sup_{v \neq 0} \frac{|v^T H_x^{-1/2} h|}{\|v\|_2}. \end{aligned}$$

Setting $w := H_x^{-1/2} v$ in the previous formula we get the result. \square

Proposition 4.2 states that $\|\cdot\|_x^*$ is the operator norm induced by $\|\cdot\|_x$; such a *dual* norm turns out to be useful for measuring the gradient of f ; see, e.g., Remark 4.1. The following proposition provides an upper bound on the error in a first order approximation of the gradient.

Proposition 4.3 Given $x, y \in \mathcal{D}_f$, where $f \in \mathcal{S}_1$, let's define

$$v := x - y.$$

If $\|v\|_y < 1$, then

$$\|g_x - g_y - H_y v\|_x^* \leq \frac{\|v\|_y^2}{(1 - \|v\|_y)^2}.$$

Proof. The result follows from the following chain of inequalities:

$$\begin{aligned} \|g_x - g_y - H_y v\|_x^* &= \left\| \int_0^1 (H_{y+tv} - H_y) v dt \right\|_x^* \\ &\quad \text{(by definition of } v) \\ &\leq \int_0^1 \|(H_{y+tv} - H_y) v\|_x^* dt \\ &\quad \text{(using the triangle inequality)} \\ &= \int_0^1 \sup_{h \neq 0} \frac{h^T (H_{y+tv} - H_y) v}{\|h\|_x} dt \\ &\quad \text{(by Proposition 4.2)} \end{aligned}$$

$$\begin{aligned}
&= \int_0^1 \sup_{h \neq 0} \frac{\langle (H_y^{-1} H_{y+tv} - I)v, h \rangle_y}{\|h\|_x} dt \\
&\quad \text{(by definition of } \langle \cdot, \cdot \rangle_y \text{)} \\
&\leq \int_0^1 \sup_{h \neq 0} \frac{\|(H_y^{-1} H_{y+tv} - I)v\|_y \|h\|_y}{\|h\|_x} dt \\
&\quad \text{(from Cauchy-Schwartz)} \\
&\leq \frac{1}{1 - \|v\|_y} \|v\|_y \int_0^1 \left(\frac{1}{(1 - t\|v\|_y)^2} - 1 \right) dt \\
&\quad \text{(from Theorem 2.1 and Lemma 2.1)} \\
&= \frac{\|v\|_y^2}{(1 - \|v\|_y)^2}.
\end{aligned}$$

The last equality follows from (2.9). \square

The results of this and of the previous sections can be extended in a straightforward way to the case when f is no longer a strictly convex but just a convex function, i.e., when the assumption that the Hessian of f is positive definite everywhere is relaxed. Indeed, we observe that, in this case, the null space of the Hessian matrix doesn't depend on the point where it is computed (cf. Nesterov, Nemirovsky [7], corollary 2.1.1) and in fact coincides with the lineality space of \mathcal{D}_f . Taking the quotient with respect to this lineality space allows us to recover all the previous results. For further details see Herzel [5].

5 The Convex Programming Problem

We apply the theory studied so far to the constructions of algorithms to solve the following convex programming problem:

$$\begin{aligned}
(P) \quad & \min_x \quad c^T x \\
& Ax = b \\
& x \in C,
\end{aligned}$$

where C is the closure of the domain of a function $f \in \mathcal{S}_1(K)$. Then \mathcal{D}_f coincides with $\text{int } C$; we normally use the former notation because we are interested in the behavior of f and its derivatives at points in this set.

We assume that A is a full row rank matrix of dimension m by n and that the vectors x, b, c have conforming dimensions.

We denote by \mathcal{A} the set of feasible points of (P) and by \mathcal{A}° the set $\{x \in \mathcal{D}_f : Ax = b\}$, which we assume nonempty. Moreover, we assume that the set of optimal solutions of (P) is nonempty and bounded.

The dual problem of (P) is given by (see Rockafellar [9], section 30)

$$\begin{aligned}
\max_y \min_{x \in C} \{c^T x - y^T (Ax - b)\} &= \max_y \{b^T y - \delta^*(A^T y - c)\} \\
&= \max_y \{b^T y - \delta^*(-s)\},
\end{aligned}$$

where $s := c - A^T y$ and $\delta^*(v) := \sup_{x \in C} v^T x$.

Therefore the problem dual to (P) is given by

$$\begin{aligned}
(D) \quad & \max_{y, s} \quad b^T y - \delta^*(-s) \\
& s = c - A^T y.
\end{aligned}$$

We denote the feasible region of problem (D) by \mathcal{A}^* , and we say that the pair (y, s) is feasible for problem (D) if and only if it satisfies

$$s = c - A^T y \tag{5.1}$$

$$\delta^*(-s) < \infty \tag{5.2}$$

Let's observe that, for x and (y, s) feasible, the *duality gap* is given by

$$c^T x - b^T y + \delta^*(-s) = x^T s + \delta^*(-s). \tag{5.3}$$

It is well known (Rockafellar [9], theorem 30.3, 30.4), that, under our hypothesis of boundedness and nonemptiness for the set of optimal solutions, the duality gap is equal to zero if and only if x is optimal for (P) and (y, s) is optimal for (D).

6 Projections in the Local Norms

We use the local metrics defined above to perform projections on the linear space $\ker A$. This approach was used in Freund and Todd [2] and we refer to them for the following results. (In [2], H_x was diagonal as well as positive definite, but the proofs only require the latter property.)

The projection in the Euclidean norm of a vector $v \in \mathbb{R}^n$ on $\ker A$ is the unique solution of

$$\begin{aligned}
\max_d \quad & v^T d - \frac{1}{2} \|d\|^2 \\
Ad &= 0.
\end{aligned}$$

Analogously, using the local norm $\|\cdot\|_x$, we get the following

Proposition 6.1 *If $x \in \mathcal{D}_f$ then there is a unique solution \bar{d} of the problem*

$$\begin{aligned} \max_d \quad & v^T d - \frac{1}{2} \|d\|_x^2 \\ \text{Ad} \quad & = 0. \end{aligned} \quad (6.1)$$

Moreover \bar{d} is the first component of the solution (\bar{d}, \bar{y}) of the system of linear equations

$$\begin{pmatrix} H_x & A^T \\ A & O \end{pmatrix} \begin{pmatrix} d \\ y \end{pmatrix} = \begin{pmatrix} v \\ 0 \end{pmatrix}, \quad (6.2)$$

and it satisfies

$$\|\bar{d}\|_x^2 = v^T \bar{d}; \quad (6.3)$$

$$\|\bar{d}\|_x = \|v - A^T \bar{y}\|_x^*; \quad (6.4)$$

$$\|\bar{d}\|_x \leq \|v\|_x^*. \quad (6.5)$$

Proof. See Freund and Todd [2], Theorem 2.1. \square

An analogue of the previous result for the dual space is given by

Proposition 6.2 *If $x \in \mathcal{D}_f$ then there is a unique solution \bar{y} to the problem*

$$\min_y \quad \frac{1}{2} \|v - A^T y\|_x^{*2}. \quad (6.6)$$

Moreover \bar{y} is the second component of the solution (\bar{d}, \bar{y}) of (6.2).

Proof. See Freund and Todd [2], Theorem 2.2. \square

Observe that (6.1) and (6.6) are dual problems with the same optimal value.

7 The Central Trajectory

The central trajectory for (P) is defined as the set of solutions, for $\mu > 0$, of the following problems :

$$\begin{aligned} \min_x \quad & c^T x + \mu f(x) \\ Ax \quad & = b, \end{aligned} \quad (7.1)$$

The following result shows that for any $\mu > 0$ (7.1) has a unique solution, i.e., the central trajectory exists.

Proposition 7.1 *For any $\mu > 0$, the problem (7.1) has a unique optimal solution.*

Proof. Since the objective function of (7.1) is convex, it is sufficient to show that if $O^+(\mathcal{D}_f)$ is the recession cone of \mathcal{D}_f (equal to that of C) and

$$v \in \ker(A) \cap O^+(\mathcal{D}_f),$$

then

$$\phi(t) := c^T(x + tv) + \mu f(x + tv)$$

is increasing for any t sufficiently big.

We have

$$\phi'(t) = c^T v + \mu g_{x+tv}^T v.$$

From the assumptions on (P) it follows that $c^T v > 0$ (otherwise the set of optimal points cannot be nonempty and bounded); moreover from corollary 3.2 we get

$$\lim_{t \rightarrow \infty} g_{x+tv}^T v = 0.$$

Therefore there exists $T \in \mathbb{R}$ such that

$$\forall t > T, \quad \phi'(t) > 0$$

and the proof is complete. \square

The first consequence of Proposition 7.1 is that the Karush-Kuhn-Tucker conditions are necessary and sufficient for optimality in (7.1); therefore the central trajectory for (P) is the set of x -components of the unique solutions for $\mu > 0$ to

$$Ax = b \tag{7.2}$$

$$A^T y + s = c \tag{7.3}$$

$$\mu g_x + s = 0. \tag{7.4}$$

Observe that everything is linear except for (7.4). We denote the unique solution by $(x(\mu), y(\mu), s(\mu))$ and call the set of such triples the primal-dual central trajectory. We try to trace the central trajectory by allowing s/μ to be just an approximation of $-g_x$. The following result states that if the approximation is good enough in the dual norm, then it is possible to derive an upper bound for the duality gap.

Theorem 7.1 *Given $s/\mu = -g_x + h$, where $x \in \mathcal{A}^\circ$ and $h \in \mathbb{R}^n$, if h satisfies*

$$\|h\|_x^* \leq \beta \leq 1/9,$$

then

$$x^T s + \delta^*(-s) \leq \mu \frac{K^2}{1 - 3\beta}.$$

Proof. For easier notation we set $\tau := s/\mu$.

Since $\|h\|_x^* < 1$, applying Theorem 2.1 to the function f^* we get

$$-\tau \in \mathcal{D}_{f^*}.$$

Therefore it follows from Proposition 4.1 that

there exists $\tilde{x} \in \mathcal{D}_f$ such that $\tau = -g_{\tilde{x}}$.

For an arbitrary $\hat{x} \in C$, we want to find an upper bound for $-g_{\tilde{x}}^T x + g_{\tilde{x}}^T \hat{x}$. This will give an upper bound on $x^T \tau + \delta^*(-\tau)$ and thus the theorem. To do that consider, for $\xi \in [0, 1]$, the function

$$\phi(\xi) := f(\tilde{x} + \alpha\xi(\hat{x} - x)),$$

where obviously $\alpha \in \mathfrak{K}$ has to be chosen so that the following relation holds:

$$y_\alpha := \tilde{x} + \alpha(\hat{x} - x) \in \mathcal{D}_f. \quad (7.5)$$

In such a case it is easy to check that $\phi \in \mathcal{S}_1(K)$.

To determine α satisfying (7.5) let's define

$$w_\alpha := x + \alpha(\hat{x} - x),$$

so that

$$y_\alpha = w_\alpha + \tilde{x} - x.$$

From Theorem 2.1 we know that

$$\{y : \|y - x\|_x < 1\} \subseteq \mathcal{D}_f.$$

Therefore, from the convexity of the set \mathcal{D}_f ,

$$\{y : \|y - w_\alpha\|_x < 1 - \alpha\} \subset \mathcal{D}_f,$$

so that y_α belongs to \mathcal{D}_f if the following holds:

$$\|y_\alpha - w_\alpha\|_x = \|\tilde{x} - x\|_x < 1 - \alpha. \quad (7.6)$$

To estimate $\|\tilde{x} - x\|_x$ we need to define

$$\psi(y) := f(y) - g_{\tilde{x}}^T(y - \tilde{x}).$$

It is easy to see that $\psi \in \mathcal{S}_1$ and that its minimizer is \tilde{x} . Let ν_x be the Newton step for ψ at x . Then

$$\begin{aligned} \nu_x &= -H_x^{-1}[g_x - g_{\tilde{x}}] \\ &= -H_x^{-1}[\tau + g_x] \\ &= -H_x^{-1}h, \end{aligned}$$

so that

$$\|\nu_x\|_x = \|H_x^{1/2}H_x^{-1}h\|_2 = \|H_x^{-1/2}h\|_2 = \|h\|_x^* < 1/9.$$

From Proposition 2.3 applied to the function ψ we get

$$\|\tilde{x} - x\|_x \leq 3\|\nu_x\|_x \leq 3\beta,$$

Therefore (7.6) holds when

$$\alpha < 1 - 3\beta.$$

Now we can use Proposition 3.2 to show that $\phi'(\xi)(\zeta - \xi) < K^2$ for $\xi, \zeta \in [0, 1]$. Then, setting $\xi := 0$ and $\zeta := 1$, we obtain

$$-g_{\hat{x}}^T x + g_{\hat{x}}^T \hat{x} < \frac{K^2}{\alpha}.$$

Since this holds for any $\alpha < 1 - 3\beta$ we deduce that

$$-g_{\hat{x}}^T x + g_{\hat{x}}^T \hat{x} \leq \frac{K^2}{1 - 3\beta}.$$

Also, since \hat{x} is arbitrary,

$$\begin{aligned} x^T \tau + \delta^*(-\tau) &= -g_{\hat{x}}^T x + \sup_{\hat{x}} g_{\hat{x}}^T \hat{x} \\ &\leq \frac{K^2}{1 - 3\beta}. \end{aligned}$$

The proof is complete. \square

Observe that when (x, y, s) lies on the primal-dual central trajectory, then we can set $\beta := 0$. From the previous theorem it then follows that the objective values of points $x(\mu)$ on the central trajectory of (P) converge to the optimal value when μ approaches zero. Moreover, the theorem shows that this remains true even if we have approximations to points on the central trajectory as long as $\beta \leq 1/9$.

Combining Propositions 6.1 and 6.2 and Theorem 7.1 we can establish a necessary condition on the projection of the gradient of the objective function of problem (7.1) to get a constant reduction of the duality gap.

Theorem 7.2 *Given $\hat{x} \in \mathcal{A}^o$ and $\hat{\mu} > 0$, if (\hat{d}, \hat{y}) is the solution to (6.1) when v is defined as*

$$v := c + \hat{\mu} g_{\hat{x}}, \tag{7.7}$$

and if we set

$$\hat{s} := c - A^T \hat{y}, \quad (7.8)$$

then

$$\|\hat{d}\|_{\hat{x}} = \|\hat{s} + \hat{\mu} g_{\hat{x}}\|_{\hat{x}}^*. \quad (7.9)$$

Moreover, if

$$\|\hat{d}/\hat{\mu}\|_{\hat{x}} = \|\hat{s}/\hat{\mu} + g_{\hat{x}}\|_{\hat{x}}^* \leq \beta \leq 1/9, \quad (7.10)$$

then (\hat{y}, \hat{s}) is feasible for the dual problem (D) and the duality gap satisfies

$$\hat{x}^T \hat{s} + \delta^*(-\hat{s}) \leq \hat{\mu} K^2 \frac{1}{1 - 3\beta}. \quad (7.11)$$

Proof. From Proposition 6.1 it follows that

$$\begin{aligned} \|\hat{d}\|_{\hat{x}} &= \|v - A^T \hat{y}\|_x^* \\ &= \|c + \hat{\mu} g_{\hat{x}} - A^T \hat{y}\|_x^* \\ &= \|\hat{s} + \hat{\mu} g_{\hat{x}}\|_{\hat{x}}^*, \end{aligned}$$

which is (7.9).

If (7.10) holds, then (7.11) follows immediately from Theorem 7.1.

Finally, from relations (7.8) and (7.11), the pair (\hat{y}, \hat{s}) is feasible for (D) . \square

When (7.10) holds we say that \hat{x} is β -close to $x(\hat{\mu})$.

The next result gives a sufficient condition for \hat{x} to be β -close to $x(\hat{\mu})$:

Proposition 7.2 *Let $\hat{x} \in \mathcal{A}^\circ$ and $\hat{\mu} > 0$. If \tilde{y} and \tilde{s} satisfy*

(i) $A^T \tilde{y} + \tilde{s} = c$, and

(ii) $\|\tilde{s}/\hat{\mu} + g_{\hat{x}}\|_{\hat{x}}^ \leq \beta$.*

then \hat{x} is β -close to $x(\hat{\mu})$.

Proof. From Propositions 6.1 and 6.2 it follows that \hat{y} as defined above is the optimal solution to the problem

$$\min_y \|c + \hat{\mu} g_{\hat{x}} - A^T y\|_{\hat{x}}^*.$$

Then, from the definition of \hat{s} ,

$$\begin{aligned} \|\hat{s} + \hat{\mu} g_{\hat{x}}\|_{\hat{x}}^* &= \|c + \hat{\mu} g_{\hat{x}} - A^T \hat{y}\|_{\hat{x}}^* \\ &\leq \|c + \hat{\mu} g_{\hat{x}} - A^T \tilde{y}\|_{\hat{x}}^*, \end{aligned}$$

and therefore, from (i) and (ii),

$$\|\hat{s} + \hat{\mu} g_{\hat{x}}\|_{\hat{x}}^* \leq \beta \hat{\mu},$$

completing the proof. \square

8 An Algorithm for Tracing the Central Trajectory

We approximate the central trajectory by performing steps in the direction of the gradient of the objective function of (7.1) projected on $\ker A$ with respect to the norm $\|\cdot\|_x$. This algorithm extends those of Gonzaga [3] and Roos and Vial [10] for linear programming, and is closely related to the general path-following method of Nesterov and Nemirovsky [7], Chapter 3.

Let's assume as given $\hat{x} \in \mathcal{A}^\circ$, $\hat{\mu} > 0$, and $\beta \leq 1/9$, where \hat{x} is β -close to $x(\hat{\mu})$.

Let (\hat{d}, \hat{y}) be the solution of the projection problem (6.1), with

$$v := c + \hat{\mu}g_{\hat{x}}.$$

From Proposition 6.1 it follows that

$$\|\hat{d}/\hat{\mu}\|_{\hat{x}} = \|\hat{s}/\hat{\mu} + g_{\hat{x}}\|_{\hat{x}}^* \leq \beta. \quad (8.1)$$

Now we set

$$\bar{d} := -\hat{d}/\hat{\mu}$$

and define the next iterate as

$$x := \hat{x} + \bar{d}. \quad (8.2)$$

Proposition 8.1 *Let x be given as in (8.2). Then $x \in \mathcal{A}^\circ$ and*

$$x \text{ is } \frac{\beta^2}{(1-\beta)^2}\text{-close to } x(\hat{\mu}).$$

Proof. Since $\|\bar{d}\|_{\hat{x}} \leq \beta < 1$, it follows that $x \in \mathcal{D}_f$; moreover, since $\bar{d} \in \ker A$, we also have $Ax = b$. Therefore $x \in \mathcal{A}^\circ$.

The second part of the proof is just an application of Proposition 7.2. In fact it is enough to show the existence of (y, s) such that

$$A^T y + s = c \quad \text{and} \quad \|s/\hat{\mu} + g_x\|_x^* \leq \frac{\beta^2}{1-2\beta^2}. \quad (8.3)$$

We show that (8.3) holds when $(y, s) = (\hat{y}, \hat{s})$.

From Proposition 6.1

$$H_{\hat{x}}\hat{d} + A^T\hat{y} = c + \hat{\mu}g_{\hat{x}},$$

so that

$$\begin{aligned} \hat{s}/\hat{\mu} &= (c - A^T\hat{y})/\hat{\mu} \\ &= -g_{\hat{x}} - H_{\hat{x}}(-\hat{d}/\hat{\mu}) \\ &= -g_{\hat{x}} - H_{\hat{x}}\bar{d}. \end{aligned}$$

Therefore

$$\begin{aligned}\|\hat{s}/\hat{\mu} + g_x\|_x^* &= \|g_x - g_{\hat{x}} - H_{\hat{x}}\bar{d}\|_x^* \\ &\leq \frac{\|\bar{d}\|_{\hat{x}}^2}{(1 - \|\bar{d}\|_{\hat{x}})^2}\end{aligned}$$

where the last inequality is given by Proposition 4.3. From (8.1) it follows that

$$\|\hat{s}/\hat{\mu} + g_x\|_x^* \leq \frac{\beta^2}{(1 - \beta)^2},$$

completing the proof. \square

The next proposition states how to decrease μ in order to stay feasible and β -close.

Proposition 8.2 *Given $\hat{x} \in \mathcal{A}^o$ and β -close to $x(\hat{\mu})$ for $\beta \leq 1/9$, if α is defined as*

$$\alpha := 1 - \frac{\beta(1 - 3\beta + \beta^2)}{(1 - \beta)^2(\beta + K)}, \quad (8.4)$$

and

$$\mu := \alpha\hat{\mu},$$

then, for x as given in (8.2), it follows that $x \in \mathcal{A}^o$ and x is β -close to $x(\mu)$.

Proof. From Proposition 8.1 it immediately follows that $x \in \mathcal{A}^o$. To show that x is β -close to $x(\mu)$ observe that

$$\begin{aligned}\|\hat{s}/\mu + g_x\|_x^* &= \|\hat{s}/(\alpha\hat{\mu}) + g_x\|_x^* \\ &= \left\| \frac{1}{\alpha} \left(\frac{\hat{s}}{\hat{\mu}} + g_x \right) - \left(\frac{1}{\alpha} - 1 \right) g_x \right\|_x^* \\ &\leq \frac{1}{\alpha} \left\| \frac{\hat{s}}{\hat{\mu}} + g_x \right\|_x^* + \left(\frac{1}{\alpha} - 1 \right) \|g_x\|_x^* \\ &\leq \frac{\beta^2}{\alpha(1 - \beta)^2} + \left(\frac{1}{\alpha} - 1 \right) K \\ &= \beta,\end{aligned}$$

where the last inequality is given by (8.3) with $s = \hat{s}$ and from Remark 4.1, while the last equality follows from the definition of α . \square

From the previous result it follows that, when $\beta := 1/9$,

$$\alpha = 1 - \frac{55}{64(9K + 1)} \leq 1 - \frac{1}{12K},$$

where the inequality follows from $K \geq 1$.

From these results we can finally construct the following algorithm:

Algorithm A

Step 0:

Set $\beta := 1/9$, $\alpha := 1 - 55/(64(9K + 1))$, $i := 0$,
 $\mu_0 > 0$ and x_0 β -close to $x(\mu_0)$.

Step 1:

Set $v_i := c + \mu_i g_{x_i}$, and compute
 $x_{i+1} := x_i - d_i/\mu_i$,
 where (d_i, y_i) is the solution to problem (6.1) with $v = v_i$, $\hat{x} = x_i$.

Step 2:

Set $\mu_{i+1} := \alpha\mu_i$
 and go to **Step 1**.

Algorithm A generates a sequence of points x_i feasible for (P), as well as a sequence of points (y_i, s_i) feasible for (D). Moreover their objective values converge to the optimal value of (P), as it is proved in the next theorem:

Theorem 8.1 *If $\{(x_i, y_i, s_i)\}$ are generated by Algorithm A, then*

(i) $x_i \in \mathcal{A}^o$, $(y_i, s_i) \in \mathcal{A}^*$;

(ii) x_i is β -close to $x(\mu_i)$;

(iii) the duality gap for x_i and (y_i, s_i) is bounded above by $\alpha^i \mu^0 K^2 (1 - 3\beta)^{-1}$.

Moreover, for any $M > 0$, in $O(MK \log(\mu_0 K))$ steps, algorithm A finds a point x_i such that

$$c^T x_i - c^T x^* < e^{-M}.$$

Proof. (i) and (ii) immediately follow from Proposition 8.2, while (iii) is a consequence of Theorem 7.2.

Moreover, since $\alpha \leq 1 - 1/12K$, it follows that

$$c^T x_i - c^T x^* \leq \left(1 - \frac{1}{12K}\right)^i \mu_0 K^2 (1 - g(\beta))^{-1}.$$

Therefore from the inequality $\log(1 - 1/12K) \leq -1/12K$ we get the last statement. \square

9 A Potential-Reduction Algorithm

In this section we generalize a primal potential-reduction algorithm studied by Gonzaga [4] and Freund [1]; see also Ye [11] and Freund and Todd [2].

We consider the problem

$$\min_{x,z} \phi(x, z) := q \log(c^T x - z) + f(x) \quad (9.1)$$

$$\text{s.t. } Ax = b \quad (9.2)$$

$$x \in C \quad (9.3)$$

$$z \leq z^*, \quad (9.4)$$

where q is a constant value and z is a lower bound for the unknown optimal value of (P), z^* . We refer to $\phi(x, z)$ as the *potential function*.

Let's assume we know $\hat{x} \in \mathcal{A}^o$ and $\hat{z} \leq z^*$. The gradient with respect to x of $\phi(x, z)$ computed at (\hat{x}, \hat{z}) is given by

$$\tilde{v} = \left(\frac{q}{c^T \hat{x} - \hat{z}} \right) c + g_{\hat{x}}. \quad (9.5)$$

We have the following result:

Proposition 9.1 *Let (\bar{d}, \bar{y}) be the solution to (6.1) for $v := \tilde{v}$.*

If $\|\bar{d}\|_{\hat{x}} \geq 1/9$ and $\alpha \in [0, 1)$, then

$$x(\alpha) := \hat{x} - \frac{\alpha \bar{d}}{\|\bar{d}\|_{\hat{x}}} \in \mathcal{A}^o. \quad (9.6)$$

Moreover $x(\alpha)$ satisfies the inequality

$$\phi(x(\alpha), \hat{z}) \leq \phi(\hat{x}, \hat{z}) - \frac{\alpha}{9} + \frac{\alpha^2}{2} + \frac{\alpha^3}{3(1-\alpha)}. \quad (9.7)$$

Proof. From Theorem 2.1 and since $\hat{x} \in \mathcal{A}^o, d \in \ker A$, it immediately follows that $x(\alpha) \in \mathcal{A}^o$.

To get inequality (9.7) we proceed as follows:

$$\begin{aligned} \phi(x(\alpha), \hat{z}) - \phi(\hat{x}, \hat{z}) &= q \log \left(1 - \frac{\alpha c^T \bar{d}}{\|\bar{d}\|_{\hat{x}}(c^T \hat{x} - \hat{z})} \right) + f(x(\alpha)) - f(\hat{x}) \\ &\leq -\frac{\alpha q c^T \bar{d}}{\|\bar{d}\|_{\hat{x}}(c^T \hat{x} - \hat{z})} + f\left(\hat{x} - \alpha \frac{\bar{d}}{\|\bar{d}\|_{\hat{x}}}\right) - f(\hat{x}) \\ &\quad \text{(by concavity of logarithm function)} \\ &\leq -\frac{\alpha q c^T \bar{d}}{\|\bar{d}\|_{\hat{x}}(c^T \hat{x} - \hat{z})} - \frac{\alpha g_{\hat{x}}^T \bar{d}}{\|\bar{d}\|_{\hat{x}}} + \frac{\alpha^2}{2} + \frac{\alpha^3}{3(1-\alpha)} \\ &\quad \text{(from Proposition 2.2)} \\ &= -\frac{\alpha}{\|\bar{d}\|_{\hat{x}}} \left(\frac{q}{c^T \hat{x} - \hat{z}} c + g_{\hat{x}} \right)^T \bar{d} + \frac{\alpha^2}{2} + \frac{\alpha^3}{3(1-\alpha)} \\ &= -\frac{\alpha}{\|\bar{d}\|_{\hat{x}}} \tilde{v}^T \bar{d} + \frac{\alpha^2}{2} + \frac{\alpha^3}{3(1-\alpha)} \end{aligned}$$

$$\begin{aligned}
&= -\alpha \|\bar{d}\|_{\hat{x}} + \frac{\alpha^2}{2} + \frac{\alpha^3}{3(1-\alpha)} \\
&\quad \text{(from (6.3) of Proposition 6.1)} \\
&\leq -\frac{\alpha}{9} + \frac{\alpha^2}{2} + \frac{\alpha^3}{3(1-\alpha)},
\end{aligned}$$

which completes the proof. \square

The previous result implies that, whenever $\|\bar{d}\|_{\hat{x}} \geq 1/9$ and α is small enough, it is possible to attain a reduction γ of the potential function. In particular, when $\alpha := 1/10$, we get $\gamma > 1/200$.

When $\|\bar{d}\|_{\hat{x}} < 1/9$ we can get a more accurate lower bound for z^* by using the following definitions:

$$\hat{\mu} := (c^T \hat{x} - \hat{z})/q \quad (9.8)$$

$$\hat{v} := \hat{\mu} \tilde{v} = c + \hat{\mu} g_{\hat{x}} \quad (9.9)$$

$$\hat{y} := \hat{\mu} \bar{y} \quad (9.10)$$

$$\hat{s} := c - A^T \hat{y}. \quad (9.11)$$

Therefore, denoting by $P_{\hat{x}}$ the projection operator on $\ker(A)$ with respect to the local norm $\|\cdot\|_{\hat{x}}$, we obtain

$$\hat{d} := P_{\hat{x}}(\hat{v}) = \hat{\mu} P_{\hat{x}}(\tilde{v}) = \hat{\mu} \bar{d},$$

that is

$$\|\hat{d}/\hat{\mu}\|_{\hat{x}} = \|\bar{d}\|_{\hat{x}} < 1/9.$$

Now we can apply Theorem 7.1 to conclude that the duality gap is bounded above by the quantity

$$\zeta := \hat{\mu} \frac{K^2}{1 - 3(1/9)} = 3\hat{\mu} K^2/2.$$

We can define a new lower bound as

$$z := c^T \hat{x} - \zeta. \quad (9.12)$$

Next we can prove that we actually get a constant reduction in the potential function.

Proposition 9.2 *Let $\hat{x} \in \mathcal{A}^o$, $\gamma > 0$, and (\bar{d}, \bar{y}) be the solution of the problem (6.1), where $v := \tilde{v}$; moreover let (\hat{y}, \hat{s}) , z be defined by (9.10), (9.11), (9.12).*

Then, when

$$\|\bar{d}\|_{\hat{x}} < 1/9,$$

z is a lower bound on z^ . Moreover, when $q \geq 3K^2/2 + \gamma$, then*

$$\phi(\hat{x}, z) \leq \phi(\hat{x}, \hat{z}) - \gamma.$$

Proof. From the definition of z it follows that

$$\begin{aligned} c^T \hat{x} - z &= \hat{\mu} \frac{3K^2}{2} \\ &= \frac{c^T \hat{x} - \hat{z}}{q} \frac{3K^2}{2}. \end{aligned}$$

Therefore, since

$$\phi(\hat{x}, z) - \phi(\hat{x}, \hat{z}) = q \log \left(\frac{c^T \hat{x} - z}{c^T \hat{x} - \hat{z}} \right),$$

the hypothesis on q implies

$$\begin{aligned} \phi(\hat{x}, z) - \phi(\hat{x}, \hat{z}) &\leq q \log \left(1 - \frac{\gamma}{q} \right) \\ &\leq -q \frac{\gamma}{q}, \end{aligned}$$

completing the proof. \square

The next result follows from the two previous propositions:

Theorem 9.1 *There is a positive value of γ (for example, $\gamma = 1/200$) such that, given $q \geq 3K^2/2 + \gamma$ and $\hat{x} \in \mathcal{A}^\circ$, $\hat{z} \leq z^*$, if we set $v := \tilde{v}$ in (9.5), compute (\bar{d}, \bar{y}) from the (6.1), and set \hat{y}, \hat{s} and z using (9.8)-(9.12), then:*

- (i) *when $\|\bar{d}\|_{\bar{x}} \geq 1/9$, it follows that ϕ can be reduced by at least γ by taking a step in the direction $-\bar{d}$;*
- (ii) *when $\|\bar{d}\|_{\bar{x}} < 1/9$, it follows that ϕ can be reduced by at least γ by substituting z for \hat{z} .*

Proof. Immediate from Propositions 9.1 and 9.2. \square

This theorem gives us an idea on how to construct an algorithm for solving (P). We suppose that we can start from a point x_0 that is optimal for

$$\begin{aligned} \min_{x \in \mathcal{D}_f} \quad & f(x) \\ & Ax = b. \end{aligned}$$

Moreover we assume that we know $z \leq z^*$ as well as the values for α and γ (one possibility is $\alpha := 1/10, \gamma := 1/200$).

Algorithm B

Step 0:

Set $q := 3K^2/2 + \gamma$ and $i := 0$.

Step 1:

Set $\hat{x} := x_i, \hat{z} := z_i$.

Set $v := \tilde{v}$ as in (9.5).

Compute (\bar{d}, \bar{y}) from problem (6.1).

Step 2:

if $\|\bar{d}\|_{\bar{x}} \geq 1/9$:

$\tilde{x} := \hat{x} - \alpha \bar{d} / \|\bar{d}\|_{\bar{x}},$

$x_{i+1} := \tilde{x}, z_{i+1} := \hat{z},$

and go to **Step 1**.

Step 3:

if $\|\bar{d}\|_{\bar{x}} < 1/9$:

compute (\hat{y}, \hat{s}, z) from (9.8)-(9.12),

set $x_{i+1} := \hat{x}, z_{i+1} := z$

and go to **Step 1**.

The next result shows that the objective function values of the sequence produced by algorithm B converge to the optimal value of (P) and gives an upper bound on the total number of steps to attain a given accuracy.

Theorem 9.2 *Algorithm B constructs a sequence of points in \mathcal{A}° and, for any $M > 0$, it determines a point x_i such that*

$$c^T x_i - z^* \leq e^{-M} (c^T x_0 - z_0)$$

after at most

$$\lceil \log(c^T x_0 - z_0) \rceil \frac{qM}{\gamma} \quad (9.13)$$

iterates, that is the number of iterates is $O(MK^2)$.

Proof. All the points in the sequence are feasible by Proposition 9.1. Moreover, since

$$\begin{aligned} q \log(c^T x_i - z_i) &= \phi(x_i, z_i) - f(x_i) \\ &\leq -\gamma i + \phi(x_0, z_0) - f(x_i) \\ &= -\gamma i + q \log(c^T x_0 - z_0) - f(x_i) + f(x_0), \end{aligned}$$

using $f(x_i) \geq f(x_0)$ yields

$$c^T x_i - z_i \leq (c^T x_0 - z_0) \exp\left(-\frac{\gamma i}{q}\right).$$

Therefore

$$\begin{aligned}
c^T x_i - z^* &\leq c^T x_i - z_i \\
&\leq (c^T x_0 - z_0) \exp\left(-\frac{\gamma^i}{q}\right) \\
&\leq e^{-M},
\end{aligned}$$

where the last inequality is verified when i is given by (9.13). \square

References

- [1] Freund, R.M. (1991), Polynomial-time algorithms for linear programming based only on primal scaling and projected gradients of a potential function, *Mathematical Programming* 51, 203-222.
- [2] Freund, R.M., Todd, M.J. (1992), Barrier Functions and Interior-Point Algorithms for Linear Programming with Zero-, One-, or Two-Sided Bounds on the Variables, Tech. Rep. No. 1016, School of Operations Research and Industrial Engineering, College of Engineering, Cornell University, Ithaca, NY (to appear in *Mathematics of Operations Research*).
- [3] Gonzaga, C.C. (1989), An algorithm for solving linear programming problems in $O(n^3 L)$ operations, in N. Megiddo, editor, *Progress in Mathematical Programming: Interior Point and Related Methods*, Springer Verlag, New York, pp. 1-28.
- [4] Gonzaga, C.C. (1990), Polynomial affine algorithms for linear programming, *Mathematical Programming* 49, 7-21.
- [5] Herzel, S. (1994), Una Teoria Generale per una Classe di Algoritmi per la Programmazione Convessa, Tesi di Dottorato di Ricerca, VI ciclo, Università di Milano, Italy.
- [6] Nesterov, Yu.E., Nemirovsky, A.S. (1989), Self-Concordant Functions and Polynomial Time Methods in Convex Programming, Report, Central Economic and Mathematical Institute, USSR Academy of Science, Moscow, Russia.
- [7] Nesterov, Yu.E., Nemirovsky, A.S. (1994), Interior Point Polynomial Algorithms in Convex Programming, *SIAM*, Philadelphia.
- [8] Renegar, J., Lecture Notes on the Efficiency of Newton's Method for Convex Optimization in Hilbert Spaces, available from the author at renegar@orie.cornell.edu.
- [9] Rockafellar, R.T. (1970), *Convex Analysis*, Princeton University Press, Princeton.

- [10] Roos, C., Vial, J.-P. (1992), A polynomial method of approximate centers for linear programming, *Mathematical Programming* 54, 295-305.
- [11] Ye, Y. (1991), An $O(n^3L)$ Potential Reduction Algorithm for Linear Programming, *Mathematical Programming* 50, 239-258.