Analysis of a Path Following Method for Nonsmooth Convex Programs*

Sanjay Mehrotra
Muhittin Gokhan Ozevin
Dept. of Industrial Engineering and Management Sciences
Northwestern University, Evanston, IL, 60208
email: [mehrotra, ozevin]@northwestern.edu

July 16, 2002

Abstract

Recently Gilbert, Gonzaga and Karas [7] constructed examples of ill-behaved central paths for convex programs. In this paper we show that under mild conditions the central path has sufficient smoothness to allow construction of a path-following interior point algorithm for non-differentiable convex programs. We show that starting from a point near the center of the first set an ϵ-optimal solution can be obtained in a finite number of iterations converging linearly.

Key Words: Convex Programming, Interior Point Methods, Non-smooth optimization

*The work of both authors was supported in part by NSF grants DMI-9500396 and DMI-9908038 and ONR grant N00014-01-1-0048.

1. Introduction

In this paper we focus on the nonsmooth convex optimization problem

$$(NSCP)\begin{cases} \textit{maximize} & c_0(x) \\ \textit{subject to} & c_i(x) \geq 0,\ i = 1, 2, \ldots, m, \end{cases}$$

where $c_i(x) : \mathbb{R}^n \rightarrow \mathbb{R}$ are concave functions. The optimal objective value of $(NSCP)$ is denoted by z^*. Let $P_z = \{x \in \mathbb{R}^n | c_0(x) \geq z,\ c_i(x) \geq 0,\ i = 1, 2, \ldots, m\}$ be the set obtained by restricting the set of feasible solutions of $(NSCP)$ with an additional constraint of the form $c_0(x) \geq z$, where z is a parameter. We assume that $m = O(n)$. In addition, we make the following assumptions:

A1 The set $intP_z = \{x \in \mathbb{R}^n | c_0(x) > z,\ c_i(x) > 0,\ i = 1, 2, \ldots, m\}$ is nonempty and bounded for all $z < z^*$.

A2 Without loss of generality we assume that $c_1(x)$ is of the form $r^2 - x^T x \geq 0$. This ensures the uniqueness of the center of P_z which is defined below. The constant r also appears in the definition of ρ in Lemma 4.1

For simplicity of notation let us define m identical functions $c_i(x) = c_0(x) - z$, for $i = m+1, \ldots, 2m$. A point $w(z) \in \mathbb{R}^n$ is called a center of P_z if it maximizes the *potential function*

$$F(x, z) = \sum_{i=1}^{2m} \ln c_i(x) \tag{1.1}$$

subject to $x \in intP_z$. Because of the strict concavity of $c_1(x) = r^2 - x^T x$, $F(x, z)$ is strictly concave in x on $intP_z$, and therefore $w(z)$ exists and it is unique. Let $\{w(z) | -\infty \leq z \leq z^*\}$ be the trajectory of centers. In this paper we generate $\{x^k\}$ that follows the trajectory of centers to an ϵ-optimal solution while $\{z^k\}$ approaches z^*.

The notion of central path defined by the maximizers of the log-barrier function has been fundamental in the development of practical algorithms for large scale linear and non-linear optimization problems. The log-barrier was first introduced by Frisch[5] for developing a sequential algorithm for non-linear programming, and its properties were studied by Fiacco and McCormick [4]. It became a fundamental tool for developing interior point methods after Karmarkar [13] introduced his polynomial time algorithm for linear programming. The method of centers based on following the central path was introduced by Huard [11]. Renegar [19] used the method of center framework to develop an algorithm that improved

the worst case iteration complexity bound of Karmarkar by $\sqrt{n}$. Subsequently Vaidya [21] used this setting to develop improved overall complexity results for linear programs. Mehrotra and Sun [14] extended these results to quadratic programs and subsequently to convex programming [15] under a "Hessian similarity" assumption. This assumption uses a curvature constant as part of the complexity results. Jarre [12], and Nesterov and Nemirovsky [16, 17] also developed similar algorithms for convex programming using alternative assumptions.
There is also some work on extending interior point methods for non-smooth convex programming. Sun and Qi [20] extended the results in Mehrotra and Sun [15] for C^1-convex programs. More recently Frenk, Sturm and Zhang [6] have proposed an algorithm for linearly constrained nondifferentiable convex programs. Their algorithm combines affine scaling and subgradient methods and prove its convergence using the log barrier function. Our analysis extends the results in Vaidya [21] and Mehrotra and Sun [15] to more general nonsmooth convex programs. Several proofs in the earlier work are reconstructed in the absence of differentiability assumption.

The work was partially motivated from the recent results of Gilbert, Gonzaga, and Karas [7]. They have shown that even for infinitely differentiable convex programs the central path can be ill-behaved. In particular, they constructed examples showing that central path may have infinitely many fixed length segments and infinite variation. This suggests that central path may not be followed in finite steps in a path following algorithm. In this paper we construct an analysis of a path following algorithm that shows that under some mild assumptions the central path can be followed even for nonsmooth convex programs in a finite number of steps in a path following setting. The presence of a spherical constraint (Assumption A2), or box constraints, give sufficient regularity to apply a Newton-like method. In our Newton-like method we construct a positive definite matrix using subgradient information, and approximate Hessians with identity matrices. Sufficient progress in a Newton-like method is possible under certain curvature assumptions (see Assumptions A6 and A7 below) on non-differentiable convex functions. A more advanced method may replace the identity matrices with suitable positive definite matrices obtained possibly using quasi-Newton type approximations, however we don't give a detailed analysis here.

Our results are also interesting because they show that a Newton-like method can be adapted for the nonsmooth case in the method of centers setting. This is important in practice because a barrier-type method is likely to exploit the sparsity properties of the problem data in comparison to the cutting plane based methods such as the volumetric barrier method [22, 1]. In comparison with the ACCPM method [9, 2, 8, 18] the presented approach allows for the possibility of including curvature information, whenever it can be approximated.

This paper is organized as follows. In the next section we state several additional assumptions and our basic algorithm. Section 3 is devoted to several utility lemmas which are frequently used in Section 4 to analyze the converge of our algorithm. Concluding remarks are made in Section 5.

2. Basics

Throughout this paper the following additional assumptions are made:

A3 $c_i(x),\ i = 0, 1, \ldots, m$, are possibly nonsmooth, concave functions from $\mathbb{R}^n$ to $\mathbb{R}$. They are also bounded on $intP_z$, i.e, there exist a positive scalar $\widehat{K}$ such that $\mid c_i(x) \mid \leq \widehat{K}$. Note that by Lemma 3.1.1 [10, Chapter 4] this assumption implies that $c_i(x),\ i = 0, 1, \ldots, m$ are locally Lipschitzian on $intP_z$.

A4 For every $x \in P_z$ we assume that we can evaluate $c_i(x)$ and an arbitrary element of the generalized gradient $\partial c_i(x)$, for $i = 0, 1, \ldots, m$, where the generalized gradient of a function f at x is a set defined as follows:

$$\partial f(x) = co\{h \mid h = \lim_{i \to \infty} \nabla f(x_i), x_i \to x, \nabla f(x_i) \text{ exists and converges}\}.$$

Note that since $c_i(\cdot)$ are finite-valued on $intP_z$ (Assumption A3) the generalized gradient $\partial c_i(\cdot)$ is a compact set on $intP_z$, for $i = 0, 1, \ldots, m$.

A5 The set $intP_{z^*-\epsilon}$ is sufficiently large to contain a sphere of radius $\underline{r}$.

A6 For $i = 0, 1, \ldots, m$, let $\overline{x_i^*}$ be the point that maximizes $c_i(x)$ over an arbitrary sphere in $intP_z$ of radius r and $\underline{x_i^*}$ be the point that minimizes $c_i(x)$ over the same sphere. We assume that there exist a strictly positive scalar L such that:

$$c_i(\overline{x_i^*}) - c_i(\underline{x_i^*}) \geq L\|r\|$$

Note that this assumption ensures that $c_i(\cdot)$ are not constant and they change sufficiently. It is used to lower bound the slacks in individual constraints defining P_{z^k} at the iterates generated by our algorithm.

A7 There exist finite curvature constants $\gamma_i > 0$ such that the inequality

$$-c_i(y) + c_i(x) + h_i^T(y - x) \leq \gamma_i \parallel y - x \parallel^2$$

holds for any $x, y \in intP_z$ and for any $h_i \in \partial c_i(x)$, $i = 0, 1, \ldots, m$.

Since $c_i(\cdot)$ are finite convex functions, they have the directional first-order expansion [10]

$$c_i(x + h) = c_i(x) + f'(x, h) + o(\|h\|), \text{ for } i = 0, 1, \ldots, m.$$

More strongly, Assumption A7 requires that $c_i(x + h) = c_i(x) + f'(x, h) + O(\|h\|^2)$.

2.1 The Algorithm

Let $F(x, z)$ be defined as in (1.1) and $w(z)$ be the corresponding center. Let the *normalized potential function* $f(x, z) = F(w(z), z) - F(x, z)$. Note that $w(z)$ is the center of P_z iff it minimizes $f(x, z)$ over P_z. If $c_i(x)$, $i = 0, 1, \ldots, m$ were twice continuously differentiable the gradient and Hessian of $f(x, z)$ with respect to x would be given as

$$\nabla f(x, z) = -\sum_{i=1}^{2m} \frac{\nabla c_i(x)}{c_i(x)}$$

and

$$\nabla^2 f(x, z) = \sum_{i=1}^{2m} \frac{1}{c_i(x)^2} \nabla c_i(x) \nabla c_i(x)^T - \sum_{i=1}^{2m} \frac{1}{c_i(x)} \nabla^2 c_i(x).$$

Since in our case $f(x, z)$ is not differentiable we approximate $\nabla^2 f(x, z)$ by

$$H(x) = \frac{1}{2} \sum_{i=1}^{2m} \frac{h_i h_i^T}{c_i(x)^2} + \sum_{i=1}^{2m} \frac{\gamma_i \parallel y - x \parallel^2}{c_i(x)}$$

where h_i is an element of $\partial c_i(x)$, for $i = 1, \ldots, 2m$. Note that the matrix $H(x)$ is symmetric and positive definite.

Algorithm.
Initialization: Let $x^0 \in intP_{z^0}$ and $z^0 < z^*$ be such that $f(x^0, z^0) \leq \upsilon$.

DO UNTIL $c(x^k) - z^k \leq \epsilon/4$
Evaluate an arbitrary element p^k of $\partial f(x^k, z)$. Determine a step direction d by solving

$$H(x^k) d = p^k$$

Let

$$x^{k+1} \leftarrow x^k + \frac{\beta^k d}{\sqrt{d^T H(x) d}}$$
$$z^{k+1} \leftarrow z^k + \frac{\alpha}{\sqrt{m}}(c(x^{k+1}) - z^k),$$

where β^k, α, and ϵ have suitable positive values.
LOOP

The algorithm starts from a feasible x^k close to $w(z^k)$; i.e., $f(x^k, z^k) \leq \upsilon$. It takes a partial Newton-like step to reduce $f(x, z^k)$ and gets to x^{k+1}, a point that is even closer to $w(z^k)$. Then we increase z^k to z^{k+1} by an amount that x^{k+1} still remains close to $w(z^{k+1})$; i.e. $f(x^{k+1}, z^{k+1}) \leq \upsilon$. This routine is repeated until $c_0(x^k)$ gets sufficiently close to the optimal objective z^* of $(NSCP)$. The stopping criterion $c_0(x^k) - z^k \leq \epsilon/4$ implies that $z^* - z^k \leq \epsilon$, which will be explained after Lemma 3.2. We will specify the parameters α, β^k and υ in our analysis of the algorithm.

3. Building Blocks for Convergence Analysis

Proposition 3.1 *Let $\partial f(x, z)$ be the generalized subgradient of $f(x, z)$ and $\partial c_i(x)$ be the generalized subgradients of $c_i(\cdot)$ at x, for $i = 1, 2, \ldots, 2m$. For all $x \in intP_z$ we have*

$$\partial f(x, z) = -\sum_{i=1}^{2m} \frac{\partial c_i(x)}{c_i(x)},$$

here $\sum$ should be understood as a set-sum.

Proof.

$$f(x, z) = \sum_{i=1}^{2m} \ln\left[\frac{c_i(w(z))}{c_i(x)}\right] \text{ and } \partial f(x, z) = -\partial\left(\sum_{i=1}^{2m} \ln c_i(x)\right)$$

From Theorem 4.1.1 in [10, Chapter 4] by induction it follows that

$$-\partial\left(\sum_{i=1}^{2m} \ln c_i(x)\right) = -\sum_{i=1}^{2m} \partial \ln c_i(x)$$

Since $c_i(x)$ are locally Lipschitzian functions and $\ln(x)$ is strictly differentiable for all $x > 0$, by Theorem 2.3.9 in [3], for all $x \in intP_z$, we have

$$\partial \ln c_i(x) = \frac{\partial c_i(x)}{c_i(x)}$$

Hence the lemma follows. ■

3.1 Preliminaries

Let

$$H(y) = \frac{1}{2} \sum_{i=1}^{2m} \frac{s_i s_i^T}{c_i(y)^2} + \sum_{i=1}^{2m} \frac{\gamma_i \parallel x - y \parallel^2}{c_i(y)}, \text{ where } s_i \in \partial c_i(y)$$

and let $E(y, \delta) = \{x | (x - y)^T H(y)(x - y) \leq \delta^2\}$ be the ellipsoid of radius δ around $y \in \ intP_z$. Note that since γ_i can be large, ellipsoids defined in this way can be very small. The following lemma bounds the relative variation of function values in $E(y, \delta)$.

Lemma 3.1 *If $x \in E(y, \delta)$, where $y \in intP_z$, and $\delta > 0$, then*

$$\frac{|c_i(y) - c_i(x)|}{c_i(y)} \leq \sqrt{2}\delta + \delta^2, \ i = 1, \dots, m \tag{3.1}$$

and

$$\frac{|c_i(y) - c_i(x)|}{c_i(y)} \leq \sqrt{2/m}\ \delta + \delta^2/m, \ i = m + 1, \dots, 2m. \tag{3.2}$$

Proof. By Assumption A7 we have $c_i(x) - c_i(y) \leq s_i^T(x - y) - \gamma_i \parallel x - y \parallel^2$. Then since $c_i(y) > 0$,

$$\frac{|c_i(x) - c_i(y)|}{c_i(y)} \leq \frac{| \ s_i^T(x - y) \ |}{c_i(y)} + \frac{\gamma_i \|x - y\|^2}{c_i(y)} \tag{3.3}$$

Now the definition of $E(y, \delta)$ implies that

$$\sum_{i=1}^{2m} \left(\frac{s_i^T(x - y)}{c_i(y)} \right)^2 \leq 2\delta^2, \qquad \sum_{i=1}^{2m} \frac{\gamma_i \|x - y\|^2}{c_i(y)} \leq \delta^2. \tag{3.4}$$

The lemma simply follows from (3.3) and (3.4). ■

Corollary 3.1 *Let $y \in \ intP_z$. Then for all $x \in E(y, 0.5)$, $|1 - c_i(x)/c_i(y)| < 1$ for $i = 1, \dots, m$. Furthermore, $E(y,\ 0.5) \subset intP_z$.*

Proof. If $y \in \ intP_z$ and $\delta = 0.5$, then Lemma 3.1 implies $|1 - c_i(x)/c_i(y)| < 1$ for all $x \in E(y, 0.5)$. Thus $c_i(x) > 0$, i.e., $E(y, 0.5) \subset \ intP_z$. ■

Lemma 3.2 *Let w be the center of the convex set P_z. Then for any $x \in P_z$ one has*

$$\frac{c(x) - z}{c(w) - z} \le 2. \tag{3.5}$$

Proof. Since $c_i(x)$ are concave functions on P_z and $c_i(w) > 0$, for any $g_i \in \partial c_i(w)$ one has

$$\frac{c_i(x)}{c_i(w)} \le 1 + \frac{1}{c_i(w)} g_i^T (x - w) \ for\ i = 1, \ldots, 2m.$$

Hence,

$$m\frac{c(x) - z}{c(w) - z} \le \sum_{i=1}^{2m} \frac{c_i(x)}{c_i(w)} \le \sum_{i=1}^{2m} \left[1 + \frac{g_i^T (x - w)}{c_i(w)}\right]. \tag{3.6}$$

By the definition of the center w, $0 \in \partial f(w, z)$. Therefore by Proposition 3.1 there exist $g_i \in \partial c_i(w)$, for $i = 1, \ldots, 2m$, such that $\sum_{i=1}^{2m} \frac{g_i}{c_i(w)} = 0$. Since inequality (3.6) holds for any $g_i \in \partial c_i(w)$, $i = 1, \ldots, 2m$, the lemma follows. ■

Remark. As a consequence of Lemma 3.2 and Lemma 3.1, for $0 < \delta \le 0.5$ and $x^k \in E(w(z^k), \delta)$ we have,

$$\begin{aligned} z^* - z^k &\le 2[c_0(w(z^k)) - z^k] \\ &\le 2\left(1 - \frac{\sqrt{2}\delta}{\sqrt{m}} - \frac{\delta^2}{m}\right)^{-1} [c_0(x^k) - z^k] \\ &< 4[c_0(x^k) - z^k]. \end{aligned}$$

Thus the stopping criterion $c_0(x^k) - z^k \le \epsilon/4$ implies $z^* - z^k \le \epsilon$.

3.2 Linear Approximation of the Normalized Potential Function

Let η be an arbitrary element of the generalized gradient $\partial f(y, z)$. Then $f(y, z) + \eta^T (x - y)$ is a first order approximation of $f(x, z)$, where x and $y \in P_z$. The next lemma bounds the error of this first order approximation on points in an ellipsoid $E(y, \delta)$.

Lemma 3.3 *Let $y \in intP_z$, and $x \in E(y, \delta)$, where $0 < \delta \le 0.5$. Let η be an arbitrary element of the generalized gradient $\partial f(y, z)$. Then one has*

$$f(x, z) - f(y, z) - \eta^T (x - y) \le \delta^2 (1 + \sqrt{2}\delta + \delta^2/2) + \frac{(\sqrt{2}\delta + \delta^2)^3}{3(1 - \sqrt{2}\delta - \delta^2)} \tag{3.7}$$

Proof. From Corollary 3.1, for $x \in E(y, \delta)$, $|1 - c_i(x)/c_i(y)| < 1$ for $i = 1, \ldots, 2m$. Therefore we have

$$f(x, z) - f(y, z) = -\sum_{i=1}^{2m} \ln \frac{c_i(x)}{c_i(y)} = \sum_{i=1}^{2m} \sum_{k=1}^{\infty} \frac{1}{k} \left[1 - \frac{c_i(x)}{c_i(y)}\right]^k \tag{3.8}$$

We can write the above sum as

$$\sum_{i=1}^{2m} \sum_{k=1}^{\infty} \frac{1}{k} \left[\frac{\kappa_i \|x - y\|^2 - h_i^T(x - y)}{c_i(y)}\right]^k \tag{3.9}$$

where h_i is an arbitrary element of $\partial c_i(y)$, and κ_i is a nonnegative scalar such that $c_i(x) - c_i(y) = h_i^T(y - x) - \kappa_i \parallel y - x \parallel^2$. Note that $0 \leq \kappa_i \leq \gamma_i$, for $i = 1, \ldots, 2m$.

Let $\phi \in \Re^{2m}$ and $\varphi \in \Re^{2m}$, where

$$\phi_i = \left|\frac{h_i^T(x - y)}{c_i(y)}\right| \text{ and } \varphi_i = \frac{\kappa_i \|x - y\|^2}{c_i(y)} \text{ for } i = 1, \ldots, 2m.$$

Note that since $x \in E(y, \delta)$, from the definition of the ellipsoid $E(y, \delta)$ we have $\sum_{i=1}^{2m} (\frac{\phi_i^2}{2} + \varphi_i) \leq \delta^2$, which in turn implies that

$$\|\phi\|_2 \leq \sqrt{2}\delta, \ \|\varphi\|_1 \leq \delta^2 \leq 0.25, \text{ and } \phi_i + \varphi_i < 1, \text{for } i = 1, \ldots, 2m. \tag{3.10}$$

We will first bound the third and higher powered terms in (3.8).
Now,

$$\begin{aligned}
\sum_{i=1}^{2m} \sum_{k=3}^{\infty} \frac{1}{k} \left[1 - \frac{c_i(x)}{c_i(y)}\right]^k &\leq \sum_{i=1}^{2m} \sum_{k=3}^{\infty} \frac{1}{k} (\phi_i + \varphi_i)^k \\
&\leq \sum_{i=1}^{2m} \frac{(\phi_i + \varphi_i)^3}{3} \sum_{k=0}^{\infty} (\phi_i + \varphi_i)^k \\
&= \sum_{i=1}^{2m} \frac{(\phi_i + \varphi_i)^3}{3} \frac{1}{1 - \phi_i - \varphi_i} \\
&\leq \frac{1}{3(1 - \sqrt{2}\delta - \delta^2)} \sum_{i=1}^{2m} (\phi_i + \varphi_i)^3 \\
&= \frac{1}{3(1 - \sqrt{2}\delta - \delta^2)} \|\phi + \varphi\|_3^3 \\
&\leq \frac{1}{3(1 - \sqrt{2}\delta - \delta^2)} (\|\phi\|_2 + \|\varphi\|_1)^3 \\
&\leq \frac{1}{3(1 - \sqrt{2}\delta - \delta^2)} (\sqrt{2}\delta + \delta^2)^3 \qquad (3.11)
\end{aligned}$$

Now we will establish the bounds on the first two terms in (3.8). The first inequality below uses Lemma 3.1.

$$
\begin{aligned}
& \sum_{i=1}^{2m}\sum_{k=1}^{2}\frac{1}{k}\left[\frac{\kappa_i\|x-y\|^2-h_i^T(x-y)}{c_i(y)}\right]^k \\
= \; & \sum_{i=1}^{2m}\frac{\kappa_i\|x-y\|^2-h_i^T(x-y)}{c_i(y)}+\frac{1}{2}\sum_{i=1}^{2m}\left[\frac{(h_i^T(x-y))^2+\kappa_i^2\|x-y\|^4-2\kappa_i\|x-y\|^2h_i^T(x-y)}{c_i(y)^2}\right] \\
\leq \; & \sum_{i=1}^{2m}\left[\frac{\gamma_i\|x-y\|^2-h_i^T(x-y)}{c_i(y)}+\frac{1}{2}\frac{(h_i^T(x-y))^2}{c_i(y)^2}\right]+(\frac{\delta^2}{2}+\sqrt{2}\delta)\sum_{i=1}^{2m}\frac{\kappa_i\|x-y\|^2}{c_i(y)} \\
\leq \; & -\sum_{i=1}^{2m}\frac{h_i^T(x-y)}{c_i(y)}+\frac{\gamma_i\|x-y\|^2}{c_i(y)}+\frac{1}{2}\frac{(h_i^T(x-y))^2}{c_i(y)}+\delta^3(\frac{\delta}{2}+\sqrt{2}) \\
\leq \; & \eta^T(x-y)+\delta^2(1+\sqrt{2}\delta+\delta^2/2),
\end{aligned} \tag{3.12}
$$

where $\eta=-\sum_{i=1}^{2m}\frac{h_i^T(x-y)}{c_i(y)}$. By Proposition 3.1 η is an element of $\partial f(y,z)$. Since h_i are arbitrary elements of $\partial c_i(y)$, (3.12) will hold for all elements of $\partial f(y,z)$. Combining (3.11) and (3.12) the lemma follows. ■

Remark. Let $x^k \in E(w(z^k),\delta)$, $0<\delta\leq 0.5$, be the solution after the kth iteration of the algorithm. By Assumption A5 $intP_{z^k}$ contains a ball of radius $\underline{r}$. Let x^* be the point that maximizes the potential function $F(x,z^k)$ over this ball. Due to Assumption A6, $F(x^*,z^k)$ should be bounded below, i.e., $\sum_{i=1}^{2m}\ln c_i(x^*)\geq -\widetilde{N}$, where $\widetilde{N}$ is some strictly positive scalar. Consequently we have $F(w(z^k),z^k)\geq -\widetilde{N}$ since $w(z^k)$ maximizes $F(x,z^k)$ over $intP_{z^k}$. Let i^* denote the tightest constraint at $w(z^k)$, i.e., $c_{i^*}(w(z^k))\leq c_i(w(z^k))$, for $i=1,2,\dots,2m$. In view of Assumption A3 $\ln c_i(w(z^k))$ is bounded from above by $\ln\widehat{K}$, $i=1,2,\dots,2m$, threfore we have $c_{i^*}(w(z^k))\geq e^{-(N+(2m-1)\ln\widehat{K})}$. Noting that $x^k\in E(w(z^k),\delta)$, Lemma 3.1 implies that $c_i(x^k)\geq(1-\sqrt{\delta}-\delta^2)e^{-(\tilde{N}+(2m-1)\ln\widehat{K})}$. So at each iteration k of our algorithm, $c_i(x^k)$ is sufficiently larger than zero. We will write $c_i(x)\geq\tau$, for $i=0,1,\dots,2m$, where $\tau=(1-\sqrt{\delta}-\delta^2)e^{-(\tilde{N}+(2m-1)\ln\widehat{K})}$. We will make use of the scalar τ in the sequel.

4. Analysis of the Algorithm

Lemma 4.1 *Let* $0 < \theta \leq 0.5$ *and let* $\rho = \tau/(r^2 \sum_{i=1}^{2m} \gamma_i)$. *If* $f(x,z) < 0.5\theta^2 - 0.3\theta^3$, *then* $x \in E(w, \theta/\sqrt{\rho})$.

Proof. By the definition of the center w, $0 \in \partial f(w,z)$. Therefore, from Proposition 3.1 we know that there exist $g_i \in \partial c_i(w)$, for $i = 1, \ldots, 2m$, such that $\sum_{i=1}^{2m} \frac{g_i}{c_i(w)} = 0$. Let $\kappa_i^{x,w}$ be nonnegative scalars such that $c_i(x) = c_i(w) + g_i^T(x-w) - \kappa_i^{x,w}\|x-w\|^2$, for $i = 1, 2, \ldots, 2m$. Let

$$H^*(w) = \frac{1}{2}\sum_{i=1}^{2m} \frac{g_i g_i^T}{c_i(x)^2} + \sum_{i=1}^{2m} \frac{\kappa_i^{x,w} \parallel y - x \parallel^2}{c_i(x)}$$

and $E^*(w,\theta) = \{x | (x-w)^T H^*(w)(x-w) \leq \theta^2\}$.

We know that

$$\begin{aligned}
\sum_{i=1}^{2m} \kappa_i^{x,w} \frac{\|x-w\|^2}{c_i(x)} &\geq \frac{\|x-w\|^2}{r^2} \text{ (since } \kappa_1^{x,w} = 1) \\
&\geq \frac{\sum_{i=1}^{2m} \gamma_i \|x-w\|^2}{r^2 \sum_{i=1}^{2m} \gamma_i} \\
&\geq \frac{\tau \sum_{i=1}^{2m} (\gamma_i \|x-w\|^2 / c_i(x))}{r^2 \sum_{i=1}^{2m} \gamma_i} \\
&= \rho \sum_{i=1}^{2m} \frac{\gamma_i \|x-w\|^2}{c_i(x)}
\end{aligned}$$

Relying on the last inequality above, $x \in E^*(w,\theta)$ implies that $x \in E(w, \theta/\sqrt{\rho})$.

Since $f(x,z)$ is strictly convex and $f(w,z) = 0$, its minimum over the region $\{x \mid x \in P_z,\ x \notin \text{int } E^*(w,\theta)\}$ occurs on the boundary of $E^*(w,\theta)$. Therefore to prove the lemma it suffices to show that $f(x,z) \geq 0.5\theta^2 - 0.3\theta^3$ for all points on the boundary of $E^*(w,\theta)$.

Now,

$$f(x) = -\sum_{i=1}^{2m} \ln\left(\frac{c_i(x)}{c_i(w)}\right) = -\sum_{i=1}^{2m} \ln\left(1 + \frac{g_i^T(x-w) - \kappa_i^{x,w} \parallel x - w \parallel^2}{c_i(w)}\right).$$

Consider a point x is on the boundary of $E^*(w,\theta)$. Note that for $\theta \leq 0.5$

$$\frac{| \, g_i^T(x-w) - \kappa_i^{x,w} \parallel x - w \parallel^2 |}{c_i(w)} < 1.$$

Using that $\ln(1+t) \leq t - t^2/4$, for $-1 \leq t \leq 1.5$, and that $\sum_{i=1}^{2m} \frac{g_i}{c_i(w)} = 0$, we get

$$
\begin{aligned}
f(x) &\geq \sum_{i=1}^{2m} \left\{ \frac{\kappa_i^{x,w} \parallel x-w \parallel^2}{c_i(w)} + 0.25 \left(\frac{g_i^T(x-w)}{c_i(w)} \right)^2 - 0.5 \frac{g_i^T(x-w)}{c_i(w)} \frac{\kappa_i^{x,w} \parallel x-w \parallel^2}{c_i(w)} \right\} \\
&\geq \frac{\theta^2}{2} - 0.5 \sum_{i=1}^{2m} \frac{g_i^T(x-w)}{c_i(w)} \frac{\kappa_i^{x,w} \parallel x-w \parallel^2}{c_i(w)} \;\; (\text{since } x \in E^*(w,\theta)\ ) \qquad (4.1)
\end{aligned}
$$

To establish the desired bound we need to get an upper bound on

$$
\sum_{i=1}^{2m} \frac{g_i^T(x-w)}{c_i(w)} \frac{\kappa_i^{x,w} \parallel x-w \parallel^2}{c_i(w)}.
$$

To this end, let $a_i = \frac{g_i^T(x-w)}{c_i(w)}$ and $b_i = \frac{\kappa_i^{x,w}\|x-w\|^2}{c_i(w)}$. Since $x \in E^*(w,\theta)$, from the definition of $E^*(w,\theta)$ we have $\sum_{i=1}^{2m} \frac{a_i^2}{2} + b_i \leq \theta^2$. Now consider the following optimization problem:

$$
\Psi^* = max \ \sum_{i=1}^{2m} a_i b_i
$$

$$
\text{subject to } \sum_{i=1}^{2m} \frac{a_i^2}{2} + b_i \leq \theta^2
$$

It can be verified that

$$
\Psi^* \leq \left(\frac{2}{3}\right)^{1.5} \theta^3, \text{ for } \theta \leq 0.7. \qquad (4.2)
$$

Combining (4.1) and (4.2) we get,

$$
f(x) \geq 0.5\theta^2 - 0.3\theta^3.
$$

Thus the lemma is valid. ■

Lemma 4.2 *Let $x \in E(\omega,\delta)$, where $0 < \delta \leq 0.25$. Let p be an arbitrary element of $\partial f(x,z)$. Let $h_i \in \partial c_i(x)$ such that $p = -\sum_{i=1}^{2m} \frac{h_i}{c_i(x)}$. Let $g_i \in \partial c_i(w)$ such that $\sum_{i=1}^{2m} \frac{g_i}{c_i(w)} = 0$. Let $\kappa_i^{x,w}$ and $\kappa_i^{w,x}$ be nonnegative scalars satisfying*

$$
\begin{aligned}
c_i(x) &= c_i(w) + g_i^T(x-w) - \kappa_i^{x,w}\|x-w\|^2 \quad \text{and} \\
c_i(w) &= c_i(x) + h_i^T(w-x) - \kappa_i^{w,x}\|x-w\|^2.
\end{aligned}
$$

Then we have

$$
p^T(x-w) \geq \left[(1-2\delta(\sqrt{2}+\delta))f(x,z) \sum_{i=1}^{2m} \left[\left(\frac{h_i^T(x-w)}{c_i(x)}\right)^2 + (\kappa_i^{x,w} + \kappa_i^{w,x}) \frac{\|x-w\|^2}{c_i(x)} \right]\right]^{1/2} \qquad (4.3)
$$

Proof. Since $f(\omega, z) = 0$ and $f(x, z)$ is convex, we have

$$0 \leq f(x,z) \leq f(x,z) - f(w,z) \leq p^T(x-w).$$

Hence (4.3) can be implied by showing that

$$p^T(x-w) \geq (1 - 2\delta(\sqrt{2}+\delta)) \sum_{i=1}^{2m} \left[\left(\frac{h_i^T(x-w)}{c_i(x)} \right)^2 + (\kappa_i^{x,w} + \kappa_i^{w,x}) \frac{\|x-w\|^2}{c_i(x)} \right] \geq 0.$$

Note that $h_i^T(x-w) = g_i^T(x-w) - (\kappa_i^{x,w} + \kappa_i^{w,x})\|x-w\|^2$, for $i = 1, \ldots, 2m$.

Now,

$$\begin{aligned} p^T(x-w) &= -\sum_{i=1}^{2m} \frac{h_i^T(x-w)}{c_i(x)} = -\sum_{i=1}^{2m} \left(\frac{g_i^T(x-w) - (\kappa_i^{x,w} + \kappa_i^{w,x})\|x-w\|^2}{c_i(x)} \right) \\ &= \sum_{i=1}^{2m} \frac{(\kappa_i^{x,w} + \kappa_i^{w,x})\|x-w\|^2}{c_i(x)} + \sum_{i=1}^{2m} g_i^T(x-w) \left(\frac{1}{c_i(w)} - \frac{1}{c_i(x)} \right) \\ &\quad \text{(subtracting } \sum_{i=1}^{2m} \frac{g_i^T(x-w)}{c_i(w)} = 0 \text{ from right hand side)} \end{aligned} \tag{4.4}$$

We have,

$$
\begin{aligned}
& g_i^T(x-w)\left(\frac{1}{c_i(w)}-\frac{1}{c_i(x)}\right) \\
= \quad & \frac{(c_i(x)-c_i(w))g_i^T(x-w)}{c_i(x)c_i(w)} \\
= \quad & \frac{c_i(x)}{c_i(w)}\frac{c_i(x)-c_i(w)}{g_i^T(x-w)}\left(\frac{g_i^T(x-w)}{c_i(x)}\right)^2 \\
= \quad & \frac{c_i(x)}{c_i(w)}\left(1-\frac{\kappa_i^{x,w}\|x-w\|^2}{g_i^T(x-w)}\right)\left(\frac{g_i^T(x-w)}{c_i(x)}\right)^2 \\
= \quad & \frac{c_i(x)}{c_i(w)}\left(\frac{g_i^T(x-w)}{c_i(x)}\right)^2-\frac{\kappa_i^{x,w}\|x-w\|^2}{c_i(x)}\left(\frac{g_i^T(x-w)}{c_i(w)}\right) \\
= \quad & \frac{c_i(x)}{c_i(w)}\left(\frac{h_i^T(x-w)+(\kappa_i^{x,w}+\kappa_i^{w,x})\|x-w\|^2}{c_i(x)}\right)^2-\frac{\kappa_i^{x,w}\|x-w\|^2}{c_i(x)}\left(\frac{g_i^T(x-w)}{c_i(w)}\right) \\
= \quad & \frac{c_i(x)}{c_i(w)}\left\{\left(\frac{h_i^T(x-w)}{c_i(x)}\right)^2+\frac{2(\kappa_i^{x,w}+\kappa_i^{w,x})\|x-w\|^2}{c_i(x)}\frac{h_i^T(x-w)}{c_i(x)}+\frac{(\kappa_i^{x,w}+\kappa_i^{w,x})^2\|x-w\|^4}{c_i(x)^2}\right\} \\
& -\frac{\kappa_i^{x,w}\|x-w\|^2}{c_i(x)}\frac{g_i^T(x-w)}{c_i(w)} \\
= \quad & \frac{c_i(x)}{c_i(w)}\left\{\left(\frac{h_i^T(x-w)}{c_i(x)}\right)^2+\frac{2(\kappa_i^{x,w}+\kappa_i^{w,x})\|x-w\|^2}{c_i(x)}\frac{(g_i^T(x-w)-(\kappa_i^{x,w}+\kappa_i^{w,x})\|x-w\|^2)}{c_i(x)}\right. \\
& \left.+\frac{(\kappa_i^{x,w}+\kappa_i^{w,x})^2\|x-w\|^4}{c_i(x)^2}\right\}-\frac{\kappa_i^{x,w}\|x-w\|^2}{c_i(x)}\frac{g_i^T(x-w)}{c_i(w)} \\
= \quad & \frac{c_i(x)}{c_i(w)}\left\{\left(\frac{h_i^T(x-w)}{c_i(x)}\right)^2+\frac{2(\kappa_i^{x,w}+\kappa_i^{w,x})\|x-w\|^2}{c_i(x)}\frac{g_i^T(x-w)}{c_i(x)}\right. \\
& \left.-\frac{(\kappa_i^{x,w}+\kappa_i^{w,x})^2\|x-w\|^4}{c_i(x)^2}\right\}-\frac{\kappa_i^{x,w}\|x-w\|^2}{c_i(x)}\frac{g_i^T(x-w)}{c_i(w)} \\
= \quad & \frac{c_i(x)}{c_i(w)}\left(\frac{h_i^T(x-w)}{c_i(x)}\right)^2+\frac{2(\kappa_i^{x,w}+\kappa_i^{w,x})\|x-w\|^2}{c_i(w)}\left(\frac{g_i^T(x-w)}{c_i(x)}-\frac{(\kappa_i^{x,w}+\kappa_i^{w,x})\|x-w\|^2}{2c_i(x)}\right) \\
& -\frac{\kappa_i^{x,w}\|x-w\|^2}{c_i(x)}\frac{g_i^T(x-w)}{c_i(w)} \\
= \quad & \frac{c_i(x)}{c_i(w)}\left(\frac{h_i^T(x-w)}{c_i(x)}\right)^2+\frac{(\kappa_i^{x,w}+2\kappa_i^{w,x})\|x-w\|^2}{c_i(w)}\frac{g_i^T(x-w)}{c_i(x)}-\frac{(\kappa_i^{x,w}+\kappa_i^{w,x})^2\|x-w\|^4}{c_i(x)c_i(w)} \\
\geq \quad & \frac{c_i(x)}{c_i(w)}\left(\frac{h_i^T(x-w)}{c_i(x)}\right)^2-\sqrt{2}\delta\frac{(\kappa_i^{x,w}+2\kappa_i^{w,x})\|x-w\|^2}{c_i(x)}-\frac{(\kappa_i^{x,w}+\kappa_i^{w,x})^2\delta^2}{\gamma_i}\frac{\|x-w\|^2}{c_i(x)} \\
\geq \quad & (1-2\sqrt{2}\delta-\delta^2)\left(\frac{h_i^T(x-w)}{c_i(x)}\right)^2-(\sqrt{2}\delta+2\delta^2)(\kappa_i^{x,w}+2\kappa_i^{w,x})\frac{\|x-w\|^2}{c_i(x)} \qquad (4.5)
\end{aligned}
$$

Last inequality above uses $\frac{c_i(x)}{c_i(w)} \geq (1-\sqrt{2}\delta-\delta^2)$ and $(\kappa_i^{x,w}+\kappa_i^{w,x})/\gamma_i \leq 2$ (since $\kappa_i^{x,w}, \kappa_i^{w,x} \leq \gamma_i$).

From (4.4) and (4.5) we obtain

$$\begin{aligned} p^T(x-w) &\geq \left[(1-2\delta(\sqrt{2}+\delta/2))\sum_{i=1}^{2m}\left(\frac{h_i^T(x-w)}{c_i(x)}\right)^2\right] \\ &\quad + \left[(1-2\delta(\sqrt{2}+\delta))\sum_{i=1}^{2m}\frac{(\kappa_i^{x,w}+\kappa_i^{w,x})\|x-w\|^2}{c_i(x)}\right] \\ &\geq (1-2\delta(\sqrt{2}+\delta))\sum_{i=1}^{2m}\left[\left(\frac{h_i^T(x-w)}{c_i(x)}\right)^2 + \frac{(\kappa_i^{x,w}+\kappa_i^{w,x})\|x-w\|^2}{c_i(x)}\right] \end{aligned}$$

■

Lemma 4.3 *Let $x \in E(w,\delta)$, where $0 < \delta \leq 0.25$, and β be a positive scalar such that $0 < \beta \leq 0.5$. Let p be an arbitrary element of $\partial f(x,z)$ and let x^+ denote the point that minimizes $p^T y$, over $E(x,\beta)$. Let $\rho = \tau/(r^2\sum_{i=1}^{2m}\gamma_i)$ as it was defined in Lemma 4.1. Then,*

$$f(x^+,z)-f(x,z) \leq -\beta\sqrt{2\rho(1-2\delta(\sqrt{2}+\delta))f(x,z)}+\beta^2(1+\sqrt{2}\beta+\beta^2/2)+\frac{(\sqrt{2}\beta+\beta^2)^3}{3(1-\sqrt{2}\beta-\beta^2)} \tag{4.6}$$

Proof. Lemma 3.3 gives

$$f(x^+,z)-f(x,z) \leq p^T(x^+-x)+\beta^2(1+\sqrt{2}\beta+\beta^2/2)+\frac{(\sqrt{2}\beta+\beta^2)^3}{3(1-\sqrt{2}\beta-\beta^2)}$$

Let $\overline{x}$ be the point where the straight line joining x to the center w intersects with the boundary of $E(x,\beta)$. Since $p^T x^+ \leq p^T\overline{x}$, one has

$$f(x^+,z)-f(x,z) \leq -\frac{\beta p^T(x-w)}{\sqrt{(x-w)^T H(x)(x-w)}}+\beta^2(1+\sqrt{2}\beta+\beta^2/2)+\frac{(\sqrt{2}\beta+\beta^2)^3}{3(1-\sqrt{2}\beta-\beta^2)} \tag{4.7}$$

From Lemma 4.2,

$$p^T(x-w) \geq \left[(1-2\delta(\sqrt{2}+\delta))f(x,z)\sum_{i=1}^{2m}\left[\left(\frac{h_i^T(x-w)}{c_i(x)}\right)^2+(\kappa_i^{x,w}+\kappa_i^{w,x})\frac{\|x-w\|^2}{c_i(x)}\right]\right]^{1/2} \tag{4.8}$$

Since $c_1(x) = r^2 - x^T x$, we have $\kappa_1^{x,w} = \kappa_1^{w,x} = 1$. Therefore,

$$\begin{aligned}
\sum_{i=1}^{2m} (\kappa_i^{x,w} + \kappa_i^{w,x}) \frac{\|x-w\|^2}{c_i(x)} &\geq 2\frac{\|x-w\|^2}{r^2} \\
&\geq \frac{2\sum_{i=1}^{2m} \gamma_i \|x-w\|^2}{r^2 \sum_{i=1}^{2m} \gamma_i} \\
&\geq \frac{2\tau \sum_{i=1}^{2m} (\gamma_i \|x-w\|^2 / c_i(x))}{r^2 \sum_{i=1}^{2m} \gamma_i} \\
&= 2\rho \sum_{i=1}^{2m} \frac{\gamma_i \|x-w\|^2}{c_i(x)} \qquad (4.9)
\end{aligned}$$

Now (4.8) and (4.9) yield

$$p^T(x-w) \geq \left[2\rho(1 - 2\delta(\sqrt{2}+\delta)) f(x,z)(x-w)^T H(x)(x-w) \right]^{1/2} \qquad (4.10)$$

The proof is completed by combining (4.10) and (4.7). ■

Corollary 4.1 *Let $\theta = 0.1\sqrt{\rho}$ and $\nu = 0.5\theta^2 - 0.3\theta^3$. If $f(x,z) \leq \nu$ and $\beta = \rho f(x,z)$, then $f(x^+, z) \leq \nu - 0.93(\rho\nu)^{1.5}$.*

Proof. Since $f(x) \leq 0.5\theta^2 - 0.3\theta^3$, where $\theta = 0.1\sqrt{\rho}$, by Lemma 4.1 we know that $x \in E(w, 0.1)$. Since $\rho \leq 1$, $\nu \leq 0.047$ and hence $\rho f(x,z) \leq 0.047$. Consequently, setting $\delta = 0.1$ and $\beta = \rho f(x,z)$ in Lemma 4.3, we get

$$\begin{aligned}
f(x^+, z) &\leq f(x,z) - \rho f(x,z)\sqrt{1.39\rho f(x,z)} + 1.007(\rho f(x,z))^2 + (\rho f(x,z))^3 \\
&\leq f(x,z) - 0.93(\rho f(x,z))^{1.5} \qquad (4.11)
\end{aligned}$$

The value of ρ is not known in advance. However, for any value of ρ in the interval $(0,1]$, the maximum of the right hand side of the inequality (4.11), subject to $f(x,z) \leq \nu$, occurs at $f(x,z) = \nu$. This implies that

$$f(x^+, z) \leq \nu - 0.93(\nu\rho)^{1.5}$$

for all $f(x,z) \leq [0,\nu)$ and for all $\rho \in (0,1]$. ■

Lemma 4.4 *Let $z^* > z^+ \geq z$. Also let w^+ and w be the unique minimizers of $f(x, z^+)$ and $f(x,z)$ over the sets P_{z^+} and P_z, respectively. Then we have*

$$c(w^+) - c(w) \leq z^+ - z.$$

Proof. Since w and w^+ are minimizing points, by Proposition 3.1 there exist $\zeta \in \partial F(w, z)$ and $\eta \in \partial F(w^+, z^+)$ such that

$$\zeta = \sum_{i=1}^{2m} \frac{g_i}{c_i(w)} = 0 \text{ and } \eta = \sum_{i=1}^{2m} \frac{g_i^+}{c_i(w^+)} = 0, \tag{4.12}$$

where $g_i \in \partial c_i(w)$ and $g_i^+ \in \partial c_i(w^+)$, $i = 1, \ldots, 2m$.

Since $\ln c_i(x)$ are concave , for $i = 1, \ldots, m$, we have

$$\begin{aligned} \sum_{i=1}^{m} \ln c_i(w^+) &\leq \sum_{i=1}^{m} \ln c_i(w) + \sum_{i=1}^{m} \frac{g_i^T(w^+ - w)}{c_i(w)} \text{ and} \\ \sum_{i=1}^{m} \ln c_i(w) &\leq \sum_{i=1}^{m} \ln c_i(w^+) + \sum_{i=1}^{m} \frac{g_i^{+T}(w - w^+)}{c_i(w^+)} \end{aligned}$$

Adding these two inequalities we get

$$\sum_{i=1}^{m} \frac{g_i^{+T}(w^+ - w)}{c_i(w^+)} - \sum_{i=1}^{m} \frac{g_i^T(w^+ - w)}{c_i(w)} \leq 0. \tag{4.13}$$

Recall that $g_i(x) = g_0(x)$, for $i = m+1, \ldots, 2m$. Now, from (4.12) we have

$$\begin{aligned} 0 &= (\eta - \zeta)^T(w^+ - w) \\ &= \sum_{i=1}^{m} \left(\frac{g_i^{+T}(w^+ - w)}{c_i(w^+)} - \frac{g_i^T(w^+ - w)}{c_i(w)} \right) + m \left(\frac{g_0^{+T}(w^+ - w)}{c_0(w^+) - z^+} - \frac{g_0^T(w^+ - w)}{c_0(w) - z} \right) \end{aligned} \tag{4.14}$$

As a consequence of (4.13) and (4.14) one has,

$$\begin{aligned} 0 &\leq \left(\frac{g_0^{+T}(w^+ - w)}{c_0(w^+) - z^+} - \frac{g_0^T(w^+ - w)}{c_0(w) - z} \right) \\ &= \frac{(c_0(w) - z)g_0^{+T}(w^+ - w) - (c_0(w^+) - z^+)g_0^T(w^+ - w)}{(c_0(w^+) - z^+)(c_0(w) - z)} \\ &= \frac{(c_0(w) - z)[(g_0^+ - g_0)^T(w^+ - w)]}{(c_0(w^+) - z^+)(c_0(w) - z)} \\ &\quad - \frac{[(c_0(w^+) - z^+) - (c_0(w) - z)]g_0^T(w^+ - w)}{(c_0(w^+) - z)(c_0(w) - z)} \end{aligned} \tag{4.15}$$

Observing that $(g_0^+ - g_0)^T(w^+ - w) \leq 0$ due to concavity of $c_0(\cdot)$, from (4.15) we get

$$\frac{[(c_0(w^+) - z^+) - (c_0(w) - z)]g_0^T(w^+ - w)}{(c_0(w^+) - z)(c_0(w) - z)} \leq 0 \tag{4.16}$$

(4.16) implies that

$$(c_0(w^+) - c_0(w))g_0^T(w^+ - w) \leq (z^+ - z)g_0^T(w^+ - w) \tag{4.17}$$

If $g_0^T(w^+ - w) < 0$ were true then by concavity of $c(\cdot)$ we would have $c_0(w^+) - c_0(w) < 0$. Since $z^+ - z \geq 0$ this would contradict with the inequality (4.17), hence $g_0^T(w^+ - w) \geq 0$.

For the case $g_0^T(w^+ - w) > 0$ dividing both sides by $g_0^T(w^+ - w)$ in (4.17), we conclude that

$$c_0(w^+) - c_0(w) \leq z^+ - z.$$

For $g_0^T(w^+ - w) = 0$, again by concavity of $c(\cdot)$, we have $c_0(w^+) - c_0(w) \leq 0$. Thereby the lemma is proved. ■

Lemma 4.5 *Let $z^+ = z + \frac{\alpha}{\sqrt{m}}(c(x^+) - z)$, where $\sqrt{m} > \alpha > 0$, and $x^+ \in$ int P_{z^+}. If $x^+ \in E(w, \delta)$, then one has*

$$f(x^+, z^+) \leq f(x^+, z) + \frac{m\alpha}{\sqrt{m} - \alpha}\left(\sqrt{\frac{2}{m}}\delta + \frac{\delta^2}{m}\right) + \frac{\alpha^2\left(1 + \sqrt{\frac{2}{m}}\delta + \frac{\delta^2}{m}\right)^2}{1 - \frac{\alpha}{\sqrt{m}}\left(1 + \sqrt{\frac{2}{m}}\delta + \frac{\delta^2}{m}\right)}.$$

Proof. From the definition of the normalized potential function the following equality readily holds true:

$$f(x^+, z^+) = f(x^+, z) + f(w, z^+) + m \ln\left[\frac{(c(x^+) - z)(c(w) - z^+)}{(c(x^+) - z^+)(c(w) - z)}\right] \tag{4.18}$$

First we will bound the last term in (4.18). Now note that

$$\begin{aligned}
\frac{(c(x^+) - z)(c(w) - z^+)}{(c(x^+) - z^+)(c(w) - z)} &= 1 + \frac{(z^+ - z)(c(w) - c(x^+))}{(c(x^+) - z^+)(c(w) - z)} \\
&= 1 + \frac{\alpha(c(x^+) - z)(c(w) - c(x^+))}{\sqrt{m}(c(x^+) - z^+)(c(w) - z)} \\
&= 1 + \frac{\alpha}{\sqrt{m} - \alpha}\frac{c(w) - c(x^+)}{c(w) - z} \\
&\leq 1 + \frac{\alpha}{\sqrt{m} - \alpha}\left(\sqrt{\frac{2}{m}}\delta + \frac{\delta^2}{m}\right) \quad \text{using (3.2)}
\end{aligned}$$

Noting that $\ln(1+t) \le t$ we then have

$$m \ln \left[\frac{(c(x^+) - z)(c(w) - z^+)}{(c(x^+) - z^+)(c(w) - z)} \right] \le \frac{m\alpha}{\sqrt{m} - \alpha} \left(\sqrt{\frac{2}{m}}\delta + \frac{\delta^2}{m} \right) \tag{4.19}$$

Next we will get an upper bound on $(c(w^+) - z^+)/(c(w) - z^+)$ in order to get a bound on the value of $f(w, z^+)$. Note that

$$\frac{c(w^+) - z^+}{c(w) - z^+} = \frac{c(w^+) - z}{c(w) - z} \left(1 + \frac{(z^+ - z)(c(w^+) - c(w))}{(c(w^+) - z)(c(w) - z^+)} \right).$$

Using Lemma 4.4 we have

$$\begin{aligned}
\frac{c(w^+) - z^+}{c(w) - z^+} &\le \frac{c(w^+) - z}{c(w) - z} + \frac{(z^+ - z)^2}{(c(w) - z)(c(w) - z^+)} \\
&= \frac{c(w^+) - z}{c(w) - z} + \frac{\alpha^2 (c(x^+) - z)^2}{m(c(w) - z)(c(w) - z^+)} \\
&= \frac{c(w^+) - z}{c(w) - z} + \frac{\alpha^2}{m} \left(\frac{c(x^+) - c(w)}{c(w) - z} + 1 \right) \left(\frac{c(x^+) - z}{c(w) - z - \frac{\alpha}{\sqrt{m}}(c(x^+) - z)} \right) \\
&= \frac{c(w^+) - z}{c(w) - z} + \frac{\alpha^2}{m} \left(\frac{c(x^+) - c(w)}{c(w) - z} + 1 \right)^2 \left(1 - \frac{\alpha(c(x^+) - z)}{\sqrt{m}(c(w) - z)} \right)^{-1} \\
&\le \frac{c(w^+) - z}{c(w) - z} + \frac{\frac{\alpha^2}{m} \left(1 + \sqrt{\frac{2}{m}}\delta + \frac{\delta^2}{m} \right)^2}{1 - \frac{\alpha}{\sqrt{m}} \left(1 + \sqrt{\frac{2}{m}}\delta + \frac{\delta^2}{m} \right)} \quad \text{using (3.2)} \qquad (4.20)
\end{aligned}$$

Now we are ready to bound $f(w, z^+)$:

$$\begin{aligned}
f(w, z^+) &= \sum_{i=1}^{m} \ln \frac{c_i(w^+)}{c_i(w)} + m \ln \left(\frac{c(w^+) - z^+}{c(w) - z^+} \right) \\
&\le \sum_{i=1}^{m} \left(\frac{c_i(w^+)}{c_i(w)} - 1 \right) + m \left(\frac{c(w^+) - z^+}{c(w) - z^+} - 1 \right) \\
&\le \sum_{i=1}^{m} \frac{c_i(w^+) - c_i(w)}{c_i(w)} + m \left(\frac{c(w^+) - z}{c(w) - z} + \frac{\frac{\alpha^2}{m} \left(1 + \sqrt{\frac{2}{m}}\delta + \frac{\delta^2}{m} \right)^2}{1 - \frac{\alpha}{\sqrt{m}} \left(1 + \sqrt{\frac{2}{m}}\delta + \frac{\delta^2}{m} \right)} - 1 \right) \\
&\le \sum_{i=1}^{2m} \frac{g_i^T (w^+ - w)}{c_i(w)} + \frac{\alpha^2 \left(1 + \sqrt{\frac{2}{m}}\delta + \frac{\delta^2}{m} \right)^2}{1 - \frac{\alpha}{\sqrt{m}} \left(1 + \sqrt{\frac{2}{m}}\delta + \frac{\delta^2}{m} \right)},
\end{aligned} \tag{4.21}$$

where g_i is an arbitrary element of $\partial c_i(w)$, for $i = 1, 2, \ldots, 2m$.
Since $\sum_{i=1}^{2m} \frac{g_i}{c_i(w)} \in \partial F(w, z)$ and $0 \in \partial F(w, z)$, we then have

$$f(w, z^+) \leq \frac{\alpha^2 \left(1 + \sqrt{\frac{2}{m}}\delta + \frac{\delta^2}{m}\right)^2}{1 - \frac{\alpha}{\sqrt{m}}\left(1 + \sqrt{\frac{2}{m}}\delta + \frac{\delta^2}{m}\right)} \tag{4.22}$$

The proof is completed by combining (4.18), (4.19) and (4.22). ■

Now we are ready for the proof of the following convergence theorem.

Theorem 4.1 *Let z^* be the optimal objective value of (NSCP). Let $\nu = 0.5\theta^2 - 0.3\theta^3$. Suitably choosing $\theta = 0.1/\sqrt{\rho}$, $\alpha = .25(\rho\nu)^{1.5}$, and $\beta = \rho f(x^k, z)$ at each iteration, the algorithm is well-defined. Furthermore, at each iteration k we have*

$$\frac{z^* - z^{k+1}}{z^* - z^k} \leq 1 - \frac{2.1}{\sqrt{m}}(\rho\nu)^{1.5}.$$

Proof. We first show that $f(x^k, z^k) \leq \nu$ by induction. This inequality is valid for $k = 0$ after the initialization step of the algorithm. Now we suppose that $f(x^k, z^k) \leq \nu$. Plugging in $f(x^k, z^k) < \nu = 0.5\theta^2 - 0.3\theta^3$, where $\theta = 0.1/\sqrt{\rho}$, in Lemma 4.1 we know that $x^k \in E(w^k, 0.1)$. Since $\beta = \rho f(x^k, z)$ from Corollary 4.1, we get

$$f(x^{k+1}, z^k) \leq \nu - 0.93(\rho\nu)^{1.5}.$$

Hence $x^{k+1} \in E(w^k, 0.1)$ by Lemma 4.1.

For $\alpha = .25(\rho\nu)^{1.5}$ and $m \geq 1$, we have

$$\frac{m\alpha}{\sqrt{m} - \alpha}\left(\sqrt{\frac{2}{m}}\delta + \frac{\delta^2}{m}\right) + \frac{\alpha^2 \left(1 + \sqrt{\frac{2}{m}}\delta + \frac{\delta^2}{m}\right)^2}{1 - \frac{\alpha}{\sqrt{m}}\left(1 + \sqrt{\frac{2}{m}}\delta + \frac{\delta^2}{m}\right)} \leq 0.93(\rho\nu)^{1.5}.$$

Lemma 4.5 now implies that $f(x^{k+1}, z^{k+1}) \leq \nu$. This completes the induction. Corollary 3.1 and Lemma 4.1 then ensure that all $x^k \in E(w^k, 0.1) \subseteq \text{int } P_{z^k}$, thus the algorithm is well-defined for the given choices of the parameters.

Now, since $x^k \in E(w^k, 0.1)$, using Lemma 3.2 and Lemma 3.1, we get

$$\begin{aligned} c_0(x^{k+1}) - z^k &\geq \left(1 - \frac{\sqrt{2}\delta}{\sqrt{m}} - \frac{\delta^2}{m}\right)[c_0(w^k) - z^k] \\ &> 0.84[c_0(w^k) - z^k] \geq 0.42(z^* - z^k). \end{aligned}$$

Thus

$$z^* - z^{k+1} = z^* - z^k - \frac{\alpha}{\sqrt{m}}[c_0(w^k) - z^k] \leq (1 - \frac{2.1}{\sqrt{m}}(\rho\nu)^{1.5})(z^* - z^k)$$

The proof is now complete. ■

Using Theorem 4.1, one can show that after the initialization step in the algorithm, an ϵ-optimal solution can be obtained in $(2.1)^{-1}(\rho\nu)^{-1.5}\sqrt{m}|\ln[\epsilon/(z^* - z^0)]|$ iterations, where $\nu = 0.5\theta^2 - 0.3\theta^3$ and $\theta = 0.1\sqrt{\rho}$. We conclude that the algorithm converges globally at a linear rate and total number of iterations is of the order of $O(\rho^{-3}\sqrt{m}|\ln\epsilon|)$. Note that ρ^{-3} can be exponentially large, so we have not shown a polynomial-time complexity. However, we have shown that starting from a near central solution the algorithm will generate an ϵ-optimal solution in a finite number of iterations.

5. Conclusions

We have shown that the trajectory of analytic centers can be followed by using a partial Newton-like subroutine for a class of nonsmooth convex programming problems. Our analysis indicates that the performance of the method depends on the curvature of the objective and constraint functions and on the size of the feasible set. For a complete description of the algorithm we also need to specify a method for generating a near-central starting point. The can be accomplished by a two phase algorithm as suggested in Mehrotra and Sun [14]. In the first phase we expand the feasible region to make any point trivially a near-center of the expanded convex set. We use an algorithm similar to the one specified in this paper to shrink the expanded set to the original set, and then use algorithm of this paper to solve the optimization problem.

The algorithm suggested in this paper replaces the second order information of non-smooth convex functions $c_i(\cdot)$ with an identity matrix. In certain situations it might be useful to use a suitably build approximation that proxies for the curvature of the convex function. For example, one may build such approximations using quasi-Newton updates at suitable points generated during the algorithm. The analysis of this paper remains valid for as long as the condition number of the "approximate Hessian" replacing the identity matrix are bounded, and the smallest eigen value of this approximated is sufficiently positive.

Acknowledgments

The work of both authors was supported in part by NSF grants DMI-9500396 and DMI-9908038 and ONR grant N00014-01-1-0048.

Bibliography

[1] K. M. Anstreicher, "Towards a practical volumetric cutting plane method for convex programming," *SIAM Journal of Optimization* 9, 1 (1998) 190-206.

[2] E.D. Andersen, J. E. Mitchell, C. Roos and T. Terlaky, "A homogenized cutting plane method to solve the convex feasibility problem," in: *Optimization Methods and Applications* (Kluwer, 1999).

[3] F.H. Clarke, *Optimization and Nonsmooth Analysis*, Classics in Applied Mathematics (Wiley, reprinted by SIAM, New York, 1983(1990)).

[4] A. V. Fiacco, G. P. McCormick, *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*, (John Wiley, New York, 1968).

[5] K. R. Frisch, "The logarithmic potential method of convex programming," Memorandum of May 13, 1955, University Institute of Economics, Oslo, Norway.

[6] J. B. G. Frenk, J. F. Sturm and S. Zhang, "An interior point subgradient method for linearly constrained nondifferentiable convex programming," *Optimization Methods and Software* 10, 2, (1998), 197-215.

[7] J. C. Gilbert, C. C. Gonzaga and E. Karas, "Examples of ill-behaved central paths in convex optimization," Rapport de recherche n° 4179 (2001).

[8] J.-L Goffin, Z.-Q Luo and Y. Ye, "Complexity analysis of an interior point cutting plane method for convex feasibility problems", *SIAM Journal on Optimization* 6, 3 (1996) 638-652.

[9] J-L Goffin amd J-P Vial, "Convex nondifferentiable optimization: A survey focused on the analytic center cutting plane method," HEC/Logilab Technical Report (1999).

[10] J.-B. Hirriart-Urruty, and C. Lemarèchal, *Convex Analysis and Minimization Algorithms I*, Grundlehren der Mathematischen Wissenschaften (Springer Verlag, Berlin, 1996).

[11] P. Huard, "Resolution of mathematical programming with nonlinear constraints by the method of centers," in: J. Abadie, ed., *Nonlinear Programming* (North-Holland, Amsterdam, 1967) pp. 207-219.

[12] F. Jarre, "The method of analytic centers for smooth convex programs," (Grottenthaler Verlag, Barnberg, 1989).

[13] N. Karmarkar,"A new polynomial-time algorithm for linear programming," *Combinatorica* 4 (1984) 373-395.

[14] S. Mehrotra and J. Sun, "An algorithm for convex quadratic programming that requires $O(n^{3.5}L)$ arithmetic operations," *Mathematics of Operations Research* 15 (1990) 342-363.

[15] S. Mehrotra and J. Sun, "An interior point algorithm for solving smooth conves programs based on Newton's method," in: J. Lagarias and M. Todd, eds., *Contemporary Mathematics, Vol. 114* (AMS, Providence,RI, 1990) pp. 265-284.

[16] J. E. Nesterov and A. S. Nemirovsky, "Polynomial-time barrier methods in convex programming," *Ekonomika i Matem. Metody* 24,7 (1988) 1084-1091 (In Russian; English trans. Matekon: Translations of Russian and East European Math. Economics.)

[17] J. E. Nesterov and A. S. Nemirovsky, *Interior-Point Polynomial Algorithms in Conves Programming*, SIAM Studies in Applied Mathematics (SIAM, 1994).

[18] Y. Nesterov and J.-PH Vial, "Homogenous analytical center cutting plane methods for convex problems and variational inequalities", *SIAM Journal on Optimization* 9, 3 (1999) 707-728.

[19] J. Renegar, "A polynomial-time algorithm based on Newton's method for linear programming," *Mathematical Programming* 40,1 (1988) 59-93.

[20] J. Sun and L. Qi, "An interior point algorithm of $O(\sqrt{m}|\ln\epsilon|)$ iterations for C^1-convex programming," *Mathematical Programming* 57 (1992) 239-257.

[21] P. M. Vaidya, "An algorithm for solving linear programming which requires $O(((m+n)n^2 + (m+n)^{1.5}n)L)$ arithmetic operations," *Proc. 19th ACM Annual Sympos. on Theory of Computing* (1987) pp. 29-38.

[22] P. M. Vaidya, "A new algorithm for minimizing convex functions over convex sets," *Mathematical Programming* 73 (1996), 291-341.