AN INTERIOR-POINT BASED SUBGRADIENT METHOD FOR NONDIFFERENTIABLE CONVEX OPTIMIZATION

J.B.G. Frenk[1], J.F. Sturm[2] and S. Zhang[1]

Final revision July 1998

ABSTRACT

We propose in this paper an algorithm for solving linearly constrained nondifferentiable convex programming problems. This algorithm combines the ideas of the affine scaling method with the subgradient method. It is a generalization of the dual and interior point method for min-max problems proposed by Sturm and Zhang [16]. In the new method, the search direction is obtained by projecting in a scaled space a subgradient of the objective function with a logarithmic barrier term. The stepsize choice is analogous to the stepsize choice in the usual subgradient method. Convergence of the method is established.

Key words: Nondifferentiable convex programming, affine scaling search direction, subgradient method.

AMS subject classification: 49M45, 90C25.

---

[1] Econometric Institute, Erasmus University Rotterdam, The Netherlands.

[2] Communications Research Laboratory, McMaster University, Hamilton, Canada. Supported by Netherlands Organization for Scientific Research (NWO).

1 Introduction

Consider the following problem

$$\text{(P)} \quad \begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & Ax = b \\ & x \geq 0 \end{array}$$

where $f(x)$ is a nondifferentiable convex function and the matrix $A \in \Re^{m\times n}$ has rank m. Furthermore, assume that there is x^0 such that $Ax^0 = b$ and $x^0 > 0$.

Typically, one encounters problems formulated as (P) in solving multistage stochastic linear programming problems (see e.g. [1]), or in the Lagrangian dual approach to solve complex combinatorial optimization problems. In two-stage stochastic programming, one is required to make a first-stage decision without knowing for certain the outcome of the future events. The objective is to minimize the expected cost, taking into account of all possible second-stage recourse problems. This objective function is convex and nondifferentiable in general. Moreover, the first-stage decision variables need to satisfy a set of linear restrictions. Clearly, such problem is represented by (P). Other examples of (P) include solving large and complex combinatorial optimization problems. In that case one often resorts to the branch-and-bound method. A popular and practical way to evaluate a bound is to use the so-called Lagrangian dual, which is again in the form of (P). The ability to solve the Lagrangian dual problem efficiently is crucial for the success of the branch-and-bound approach.

A classical method for solving (P) is the so called *subgradient method* (cf. Shor [14]). This method evaluates a subgradient of the function f at each iteration, and then takes a step along the negative of the subgradient. Since the problem has only linear constraints, one may easily compute the projection of the iterate onto the feasible region once infeasibility is detected. In this way, the method produces a sequence of feasible solutions, and under certain stepsize rules it can be proven that the sequence is indeed convergent [14].

Recently, intensive research has been done on the so-called *affine scaling* algorithm after the interior point methods became popular. Dikin [2] introduced for the first time the affine scaling algorithm for linear programming in 1967. This algorithm remained unknown in the west until it was rediscovered by several authors independently after the publication of the well known paper of Karmarkar [10]. For a thorough historical review on the development of the affine scaling method, we refer to a comprehensive survey by Tsuchiya [20]. Application of the affine scaling idea to convex quadratic programming was proposed by Dikin and Zorkaltsev [4] in 1980, before Karmarkar's algorithm. This method was rediscovered by Ye [21] and Ye and Tse [23]. Later, in Ye [22] the method is extended to solve nonconvex quadratic programs. Although nonconvex quadratic programming is known to be $\mathcal{NP}$-hard, Ye proved in [22] that the affine scaling method produces a sequence converging to a local optimum. Sun [17] proved convergence of the affine scaling method applied to convex quadratic programs without any nondegeneracy assumption, using a similar proof technique as in Tseng and Luo [18]. A drawback is that only very small stepsizes are allowed. Recently, Monteiro and Tsuchiya [11] gave a global convergence proof for the long step version of the (second-order) affine scaling algorithm for convex quadratic programs without any nondegeneracy assumptions. Solving nonquadratic but differentiable convex programming by an affine scaling algorithm was first proposed by Gonzaga and Carlos [7]. Monteiro and Wang [12] proved the convergence of some trust region affine scaling algorithms for linearly constrained smooth convex or concave programs.

In this paper, we focus our attention on nondifferentiable convex programs and we propose to combine the affine scaling method with the subgradient method. Naturally, by affine scaling the search direction guarantees feasibility if the stepsize is appropriately chosen since the affine scaling direction takes care of the local geometry of the constaints. Due to many fascinating properties of the affine scaling method, we believe that the method in combination with the subgradient method deserves detailed investigation in the context of linearly constrained nondifferentiable convex programming. The algorithm presented in this paper is a generalization of the dual and affine scaling algorithm for min-max problems proposed by Sturm and Zhang in [16]. In our analysis, we use a logarithmic barrier term to steer the iterates away from the boundary. This simplifies the proof considerably. Under this modification we prove the global convergence of the best recorded iterates. Additionally, if an acute angle condition is assumed in the neighborhood of the optimum, then the convergence of the whole sequence is proven.

The organization of the paper is as follows. In the next section the new method will be introduced together with the logarithmic barrier technique. The convergence of a subsequence to the optimum is proven in Section 3. Furthermore, we prove in Section 4 the convergence of the whole sequence under the acute angle condition. Finally, we conclude the paper in Section 5.

2 The affine scaling subgradient method

Let F denote the feasible set $F := \{x : \; Ax = b, \, x \geq 0\}$. We assume throughout this paper that the feasible region F is bounded. Moreover, we assume that the objective function f is Lipschitz continuous with L as a Lipschitz constant.

Introduce the logarithmic barrier function with parameter $\mu > 0$

$$f_\mu(x) := f(x) - \mu \sum_{i=1}^{n} \log x_i.$$

Observe that $f_\mu(x)$ is a strictly convex function in $\Re^n_+$, for which the subgradient set is given by

$$\partial f_\mu(x) = \partial f(x) - \mu X^{-1} e \tag{2.1}$$

where X stands for the diagonal matrix with elements taken from x and e is the all-one vector.

The concept of logarithmic barrier was introduced by Frisch [5] in 1955. Recently, Nesterov and Nemirovskii [13] systematically treated the theory and techniques of interior point methods, including the barrier technique, for convex programming. In the next lemma we show that the optimizer of the barrier function will be a nearly optimal solution of (P) if the parameter μ of the barrier is small.

Lemma 2.1 *If $x^* \in F$ is such that $f_\mu(x^*) = \min_{x\in F} f_\mu(x)$ then*

$$f(x^*) \le \min_{x\in F} f(x) + n\mu.$$

PROOF.

From the convexity of f_μ, it follows that

$$0 \in \partial f_\mu(x^*),$$

i.e., there exists $\eta \in \partial f(x^*)$ such that

$$\eta - \mu (X^*)^{-1} e = 0.$$

By the convexity of f, we obtain for $x^{**} \in \arg\min_{x\in F} f(x)$ that

$$\begin{aligned} \min_{x\in F} f(x) &\ge f(x^*) + \eta^T (x^{**} - x^*) \\ &= f(x^*) + \mu e^T (X^*)^{-1} (x^{**} - x^*) \\ &= f(x^*) - \mu (n - e^T (X^*)^{-1} x^{**}) \\ &\ge f(x^*) - n\mu. \end{aligned}$$

□

In this paper we shall minimize f_μ over F with a prefixed parameter $\mu > 0$. We shall fix $0 < \mu < \epsilon/n$ if an ϵ-optimal solution is desired.

Due to the strict convexity of f_μ and the fact that F is bounded, there is a unique minimizer for f_μ for each $\mu > 0$. Therefore a central path can be defined for variable $\mu > 0$ which will lead to an optimal solution as $\mu \to 0$. Clearly, the idea of path-following (cf. Gonzaga [6]) can be explored in this context. In the current paper, however, we will restrict ourselves to the case when μ is fixed.

Consider now an iterative point x. Assume that $Ax = b$ and $x > 0$.

Let

$$d \in \partial f_\mu(x).$$

Consider the following direction finding problem

$$\begin{array}{lll} \text{(AS)} & \text{minimize} & d^T \Delta x \\ & \text{subject to} & A\Delta x = 0 \\ & & \|X^{-1}\Delta x\| \le 1 \end{array}$$

where $\|\cdot\|$ denotes the Euclidean norm.

The solution for (AS) is the well known *affine scaling* direction. For linear programming, such a direction was first considered by Dikin [2] in 1967. Interestingly, (AS) has an analytical solution given by

$$\Delta x = -X \frac{P_{AX} X d}{\|P_{AX} X d\|} \tag{2.2}$$

where $P_{AX} y$ is the projection of a given vector y onto the null space of AX. Letting I denote the identity matrix, we have for $x > 0$ that

$$P_{AX} y = (I - XA^T (AX^2 A^T)^{-1} AX) y.$$

Given a stepsize t we define the next iterate as

$$x' = x + t\Delta x.$$

Clearly, if $0 < t < 1$ we will automatically retain the interior property: $Ax' = b$ and $x' > 0$.

As in the classical subgradient method (cf. Shor [14]), we require a pre-specified sequence of stepsizes $\{t_k : k = 0, 1, ...\}$, satisfying

- $0 < t_k \le \alpha < 1$ and $\lim_{k \to \infty} t_k = 0$
- $\sum_{k=0}^{\infty} t_k = +\infty$.

Let $x^{(0)} \in \Re^n_{++}$ (i.e. $x^{(0)} > 0$) and $Ax^{(0)} = b$.

The affine scaling subgradient method proceeds as follows:

$$x^{(k+1)} := x^{(k)} + t_k \Delta x^{(k)} \text{ for } k = 0, 1, ...$$

where $\Delta x^{(k)}$ is computed according to (2.2).

Since we assume that the feasible region F is bounded, there exists a unique minimum point of f_μ over F. We denote this minimum point to be x^*. We shall prove in the next two sections the convergence of the sequence $\{x^{(k)} : k = 1, 2, ...\}$.

3 Convergence of a subsequence

The whole convergence proof consists of three parts. First we show that the iterates will be contained in a compact subset of $\Re^n_{++}$, the open interior of $\Re^n_+$, due to the logarithmic barrier effect. Second, we prove by contradiction that at least a subsequence will converge to the optimal point. Finally, in the next section we will continue to prove that under an additional condition in fact the whole sequence converges.

Denote

$$p = \frac{P_{AX} X d}{\|P_{AX} X d\|}.$$

Therefore,

$$\Delta x = -Xp$$

and so

$$x' = X(e - tp). \tag{3.1}$$

We shall first prove that the sequence of iterates has no cluster point on the boundary of the nonnegative orthant $\Re^n_+$. By definition of p, there holds that

$$p = \frac{1}{\|P_{AX} X d\|} P_{AX}(X\bar{d} - \mu e) \tag{3.2}$$

for some $\bar{d} \in \partial f(x)$.

Dikin [3] first showed that the norm of $(AX^2A^T)^{-1}AX^2$, which is a weighted pseudoinverse of A^T, has a uniform upper bound which is independent of the positive vector x (see also Stewart [15]). Therefore, there is a positive constant c_1 such that

$$\|X^{-1} P_{AX} X \bar{d}\| = \left\|(I - A^T(AX^2A^T)^{-1}AX^2)\bar{d}\right\| \le c_1 \|\bar{d}\|, \tag{3.3}$$

for all $x \in \Re^n_{++}$. Remark that by the Lipschitz continuity of f, it follows that $\|\bar{d}\| \le L$. Therefore,

$$\|X^{-1}P_{AX}X\bar{d}\| \le c_1 L.$$

Hence, it follows from (3.2) that for all i with $x_i \to 0$,

$$p_i \to -\frac{\mu}{\|P_{AX}Xd\|}e_i^T P_{AX}e. \tag{3.4}$$

Here, e_i stands for the i-th column of the order n identity matrix. Below we shall prove that for these indices i with $x_i \to 0$ we also have $e_i^T P_{AX}e \to 1$. To this end, we introduce the following lemma concerning the continuity of the projection operation.

Lemma 3.1 *Let $x^{(k)} \in F \cap \Re^n_{++}$, $k = 1, 2, \ldots$ be a sequence with $\lim_{k\to\infty} x^{(k)} = \bar{x}$. Then*

$$\lim_{k\to\infty} P_{AX^{(k)}}e = P_{A\bar{X}}e.$$

PROOF.

The lemma is clearly true for the case $\bar{x} \in \Re^n_{++}$, simply using the analytical expression for projection. Now we consider the case in which $\bar{x}$ lies on the boundary of $\Re^n_{+}$, i.e. we can make a partition such that

$$\bar{x} = \begin{bmatrix} \bar{x}_I \\ \bar{x}_J \end{bmatrix},\ \bar{x}_I = 0,\ \bar{x}_J > 0.$$

We will decompose $x^{(k)}$ with respect to the same partition (I, J). Let $y^{(k)}$ be the optimal solution of the following problem,

$$\begin{array}{ll} \text{minimize} & \|y - e\|^2 \\ \text{subject to} & AX^{(k)}y = 0 \end{array}$$

i.e. $y^{(k)} = P_{AX^{(k)}}e$, and let $\bar{y}$ be the solution of

$$\begin{array}{ll} \text{minimize} & \|y - e\|^2 \\ \text{subject to} & A\bar{X}y = 0, \end{array}$$

so that $\bar{y} = P_{A\bar{X}}e$. From the fact that $\bar{x}_I = 0$, it easily follows that $\bar{y}_I = e_I$. It is known from Gonzaga and Tapia [8] that

$$\lim_{k\to\infty} y_J^{(k)} = \bar{y}_J. \tag{3.5}$$

It remains to show that $y_I^{(k)} \to \bar{y}_I$. As $\bar{y}_I = e_I$, we have

$$\left\|y^{(k)} - e\right\|^2 = \left\|y_I^{(k)} - \bar{y}_I\right\|^2 + \left\|y_J^{(k)} - e_J\right\|^2,$$

so that

$$\limsup_{k\to\infty} \left\| y_I^{(k)} - \bar{y}_I \right\|^2 = \limsup_{k\to\infty} \left\| y^{(k)} - e \right\|^2 - \|\bar{y}_J - e_J\|^2 , \tag{3.6}$$

where we used (3.5). Now consider a sequence $\hat{y}^{(1)}, \hat{y}^{(2)}, \ldots$ defined as

$$\hat{y}_I^{(k)} = e_I \quad \text{and} \quad \hat{y}_J^{(k)} = (X_J^{(k)})^{-1}(x_J^{(k)} - \bar{x}_J) + (X_J^{(k)})^{-1}\bar{X}_J\bar{y}_J.$$

By construction, we have

$$\lim_{k\to\infty} \hat{y}^{(k)} = \bar{y} \quad \text{and} \quad AX^{(k)}\hat{y}^{(k)} = 0.$$

By definition of $y^{(k)}$ and subsequently using the triangle inequality, we thus have

$$\left\| y^{(k)} - e \right\| \le \left\| \hat{y}^{(k)} - e \right\| \le \left\| \hat{y}_J^{(k)} - \bar{y}_J \right\| + \|\bar{y}_J - e_J\| . \tag{3.7}$$

Using (3.6), (3.7) and $\hat{y}^{(k)} \to \bar{y}$, it follows from (3.6) that

$$\limsup_{k\to\infty} \left\| y_I^{(k)} - \bar{y}_I \right\|^2 = \limsup_{k\to\infty} \left\| \hat{y}_J^{(k)} - \bar{y}_J \right\|^2 = 0.$$

This completes the proof.

$\square$

We remark that several facts concerning affine scaling projection matrices such as the result in Lemma 3.1 were studied in Tsuchiya [19].

By (3.4) and Lemma 3.1, we have

$$p_i \to -\frac{\mu}{\|P_{AX}Xd\|} \text{ if } x_i \to 0.$$

Since $x' = X(e - tp)$ (see (3.1)) and using the fact that the feasible region is compact, we conclude that there is a positive constant c_2, such that as soon as $x_i \le c_2$ it will follow $x_i' > x_i$. This leads to the following lemma:

Lemma 3.2 *There is a constant $c_3 > 0$ such that*

$$\inf_{k\ge 0} \min_{1\le i\le n} x_i^{(k)} \ge c_3.$$

PROOF.

Since the steplength t_k converges to zero, we can find k_0 such that for all $k \ge k_0$ we have $t_k \le \min\{c_2/2, 1/2\}$. Now we claim for any $k \ge k_0$ and any $1 \le i \le n$ that

$$x_i^{(k)} \ge \min\{c_2/2, x_i^{(k_0)}\}.$$

This inequality is clearly true when $k = k_0$. Now suppose that this inequality holds for $k = \bar{k} \geq k_0$. Consider $x_i^{(\bar{k})}$ with $1 \leq i \leq n$. If $c_2/2 \leq x_i^{(\bar{k})} < c_2$ then $x_i^{(\bar{k}+1)} > x_i^{(\bar{k})} \geq c_2/2$. If $x_i^{(\bar{k})} \geq c_2$ then due to (3.1) and $t_{k_0} \leq 1/2$ and $\|p\| = 1$ we have $x_i^{(\bar{k}+1)} \geq x_i^{(\bar{k})}/2 \geq c_2/2$. Inductively this shows the desired inequality, which in turn proves the lemma.

□

To prove the global convergence, we first show that there is at least a subsequence of iterates converging to the optimal point. Remember that the optimal point is unique due to the logarithmic barrier term. The proof is done by contradiction. Assume that the optimal point is x^* and that x^* is not a cluster point of the sequence produced by the algorithm. Note that the set of all cluster points is closed, and hence is compact due to the compactness of F. Let $\hat{x}$ be one of the cluster points which minimizes f_μ over the set of all cluster points. By the contradiction assumption, there is an infinite index set $\mathcal{K}$ such that

$$\lim_{k \in \mathcal{K}} x^{(k)} = \hat{x}$$

and $\hat{x} \neq x^*$.

We consider now one iterative solution $x^{(k)}$.

Denote the lower level set of f_μ at $x^{(k)}$ by

$$L_k := \{x \in F : \; f_\mu(x) \leq f_\mu(x^{(k)})\}.$$

Let $\bar{x} \in F$ be a point in the relative interior of L_k. Denote

$$\delta^{(k)}(\bar{x}) := (p^{(k)})^T (X^{(k)})^{-1} (x^{(k)} - \bar{x}) \tag{3.8}$$

and

$$\rho^{(k)}(\bar{x}) := \left\| \bar{X}^{-1} (\bar{x} - x^{(k)}) \right\|. \tag{3.9}$$

We will see in the sequel that $\delta^{(k)}(\bar{x})$ is bounded below by the difference between the function values at $x^{(k)}$ and $\bar{x}$. The quantity $\rho^{(k)}(\bar{x})$ measures the Euclidean distance between $x^{(k)}$ and $\bar{x}$ in the scaled space. As will be shown later, $\rho^{(k)}(\bar{x})$ will be reduced if $\delta^{(k)}(\bar{x})$ is relatively large. By the contradiction assumption, the sequence $x^{(k)}$ never reaches a certain neighborhood of x^*, and so $\delta^{(k)}(\bar{x})$ stays large for a properly chosen $\bar{x}$. This would imply that the distance $\rho^{(k)}(\bar{x})$ will eventually become negative, which is a contradiction.

For ease of notation, in case of no confusion we drop the argument $\bar{x}$ and simply write $\delta^{(k)}$ and $\rho^{(k)}$ respectively for $\delta^{(k)}(\bar{x})$ and $\rho^{(k)}(\bar{x})$.

We have the following lemma:

Lemma 3.3 *There is a positive constant* c_4 *such that*

$$\delta^{(k)} \geq c_4(f_\mu(x^{(k)}) - f_\mu(\bar{x})).$$

PROOF.

Using $x^{(k)}$, $\bar{x} \in F$ and the subgradient inequality, we have

$$\begin{aligned} \delta^{(k)} &= (p^{(k)})^T (X^{(k)})^{-1}(x^{(k)} - \bar{x}) \\ &= (d^{(k)})^T X^{(k)} P_{AX^{(k)}} (X^{(k)})^{-1}(x^{(k)} - \bar{x}) / \left\| P_{AX^{(k)}} X^{(k)} d^{(k)} \right\| \\ &= (d^{(k)})^T (x^{(k)} - \bar{x}) / \left\| P_{AX^{(k)}} X^{(k)} d^{(k)} \right\| \\ &\geq (f_\mu(x^{(k)}) - f_\mu(\bar{x})) / \left\| P_{AX^{(k)}} X^{(k)} d^{(k)} \right\|. \end{aligned} \tag{3.10}$$

Since the feasible region F is bounded, denote g to be a bound of F, i.e. for all $x \in F$ it holds that $\|x\| \leq g$.

Thus we have, using the triangle inequality,

$$\left\| P_{AX^{(k)}} X^{(k)} d^{(k)} \right\| \leq \left\| X^{(k)} d^{(k)} \right\| \leq \left\| X^{(k)} (\bar{d}^{(k)} - \mu (X^{(k)})^{-1} e) \right\| \leq gL + \mu\sqrt{n}. \tag{3.11}$$

From (3.10) and (3.11) it follows that

$$\delta^{(k)} \geq c_4(f_\mu(x^{(k)}) - f_\mu(\bar{x}))$$

with $c_4 = 1/(gL + \mu\sqrt{n})$.

□

Consider two consecutive iteration points $x^{(k)}$ and $x^{(k+1)}$. We have the following relation:

Lemma 3.4 *If* $\rho^{(k)} < 1$ *then*

$$(\rho^{(k+1)})^2 \leq (\rho^{(k)})^2 - 2t_k[\delta^{(k)} - (1 + \frac{1}{1-\rho^{(k)}})(\rho^{(k)})^2 - (1+\rho^{(k)})^2 t_k/2].$$

PROOF.

Since $x^{(k+1)} = x^{(k)} - t_k X^{(k)} p^{(k)}$ we have

$$\begin{aligned} (\rho^{(k+1)})^2 &= \left\| \bar{X}^{-1}(\bar{x} - x^{(k)} + t_k X^{(k)} p^{(k)}) \right\|^2 \\ &= (\rho^{(k)})^2 - 2t_k\delta^{(k)} - 2t_k(\bar{x} - x^{(k)})^T \bar{X}^{-1}(\bar{X}(X^{(k)})^{-1} - \bar{X}^{-1}X^{(k)})p^{(k)} \\ &\quad + t_k^2 \left\| \bar{X}^{-1} X^{(k)} p^{(k)} \right\|^2 \\ &= (\rho^{(k)})^2 - 2t_k\delta^{(k)} - 2t_k(\bar{x} - x^{(k)})^T \bar{X}^{-1}(\bar{X}(X^{(k)})^{-1} - I + I - \bar{X}^{-1}X^{(k)})p^{(k)} \\ &\quad + t_k^2 \left\| \bar{X}^{-1} X^{(k)} p^{(k)} \right\|^2. \end{aligned} \tag{3.12}$$

Notice that

$$\left\|\bar{X}^{-1}x^{(k)}-e\right\|_{\infty} \leq \left\|\bar{X}^{-1}x^{(k)}-e\right\| = \rho^{(k)}. \tag{3.13}$$

Therefore, using $\left\|p^{(k)}\right\| = 1$ it follows

$$\left\|\bar{X}^{-1}X^{(k)}p^{(k)}\right\| \leq \left\|\bar{X}^{-1}x^{(k)}\right\|_{\infty}\left\|p^{(k)}\right\| \leq 1+\rho^{(k)}. \tag{3.14}$$

Similarly, using $\left\|p^{(k)}\right\| = 1$ and the Cauchy-Schwartz inequality, we have

$$\begin{aligned} \mid (\bar{x}-x^{(k)})^T\bar{X}^{-1}(I-\bar{X}^{-1}X^{(k)})p^{(k)} \mid \ &\leq \ \left\|(I-\bar{X}^{-1}X^{(k)})\bar{X}^{-1}(\bar{x}-x^{(k)})\right\| \\ &\leq \ \left\|(I-\bar{X}^{-1}X^{(k)})e\right\|_{\infty}\left\|\bar{X}^{-1}(\bar{x}-x^{(k)})\right\| \\ &= \ \left\|\bar{X}^{-1}x^{(k)}-e\right\|_{\infty}\rho^{(k)} \\ &\leq \ (\rho^{(k)})^2, \end{aligned} \tag{3.15}$$

where the last inequality follows from (3.13).

Again, using (3.13) and noting

$$\left\|(X^{(k)})^{-1}\bar{x}-e\right\|_{\infty} \leq \rho^{(k)}/(1-\rho^{(k)})$$

we have

$$\begin{aligned} \mid (\bar{x}-x^{(k)})^T\bar{X}^{-1}(\bar{X}(X^{(k)})^{-1}-I)p^{(k)} \mid \ &\leq \ \left\|\bar{X}^{-1}(\bar{x}-x^{(k)})\right\|\cdot\left\|(X^{(k)})^{-1}\bar{x}-e\right\|_{\infty} \\ &\leq \ \frac{(\rho^{(k)})^2}{1-\rho^{(k)}}. \end{aligned} \tag{3.16}$$

Substituting the inequalities (3.14), (3.15) and (3.16) into (3.12) yields the desired result.

□

By our contradiction assumption, $\hat{x}$ is non-optimal. Therefore, there exists $\tilde{f}$ such that

$$f_\mu(x^*) < \tilde{f} < f_\mu(\hat{x}).$$

However, $\hat{x}$ attains the minimum value in f_μ among all cluster points by definition. Hence, there exists for any given $\lambda \in (0,1)$ an integer k_1 such that for all $k \geq k_1$ it holds that

$$f_\mu(x^{(k)}) \geq (1-\lambda)f_\mu(\hat{x}) + \lambda\tilde{f}. \tag{3.17}$$

Let

$$\bar{x}_\lambda := (1-\lambda)\hat{x} + \lambda x^*,$$

where $0 < \lambda < 1$.

From now on, denote

$$\delta_\lambda^{(k)} := \delta^{(k)}(\bar{x}_\lambda).$$

and

$$\rho_\lambda^{(k)} := \rho^{(k)}(\bar{x}_\lambda).$$

By Lemma 3.3 and (3.17) we obtain

$$\delta_\lambda^{(k)} \geq c_4(f_\mu(x^{(k)}) - f_\mu(\bar{x}_\lambda)) \geq c_4\lambda(\tilde{f} - f_\mu(x^*)) \tag{3.18}$$

for all $k \geq k_1$.

Applying the triangle inequality,

$$\begin{aligned} \rho_\lambda^{(k)} &= \left\| \bar{X}_\lambda^{-1}(\bar{x}_\lambda - x^{(k)}) \right\| \\ &\leq \left\| \bar{X}_\lambda^{-1}(\bar{x}_\lambda - \hat{x}) \right\| + \left\| \bar{X}_\lambda^{-1}(\hat{x} - x^{(k)}) \right\| \\ &= \lambda \sqrt{\sum_{i=1}^n \left(\frac{\hat{x}_i - x_i^*}{(1-\lambda)\hat{x}_i + \lambda x_i^*}\right)^2} + \left\| \bar{X}_\lambda^{-1}(\hat{x} - x^{(k)}) \right\|. \end{aligned} \tag{3.19}$$

As $\hat{x}$ is a limit point and $\hat{x} \neq x^*$, there is an unbounded set $K(\lambda) \subseteq \mathcal{K}$ of integers such that

$$\left\| \bar{X}_\lambda^{-1}(\hat{x} - x^{(k)}) \right\| \leq \lambda \sqrt{\sum_{i=1}^n \left(\frac{\hat{x}_i - x_i^*}{(1-\lambda)\hat{x}_i + \lambda x_i^*}\right)^2} \tag{3.20}$$

for all $k \in K(\lambda)$.

By Lemma 3.2 we know that $\hat{x}_i \geq c_3 > 0$ for $1 \leq i \leq n$. Therefore, the relations (3.19) and (3.20) imply that

$$\rho_\lambda^{(k)} = \mathcal{O}(\lambda)$$

for $k \in K(\lambda)$.

In particular, there exists a sufficiently small constant $\lambda_0 > 0$ such that when $k \in K(\lambda_0)$ it holds that

$$\left(1 + \frac{1}{1 - \rho_{\lambda_0}^{(k)}}\right)(\rho_{\lambda_0}^{(k)})^2 < \min\{\frac{1}{3}c_4\lambda_0(\tilde{f} - f_\mu(x^*)), 1\}. \tag{3.21}$$

Let k_1 be chosen according to (3.17) for $\lambda = \lambda_0$.

Because $\lim_{k\to\infty} t_k = 0$, there is $k_2 \in K(\lambda_0)$ with $k_2 \geq k_1$, such that for all $k \geq k_2$ we have

$$2t_k < \frac{1}{3}c_4\lambda_0(\tilde{f} - f_\mu(x^*)).$$

In particular, for $k \geq k_2$ and if (3.21) holds, then $\rho_{\lambda_0}^{(k)} < 1$ and so we have

$$\frac{(1+\rho_{\lambda_0}^{(k)})^2}{2}t_k < 2t_k < \frac{1}{3}c_4\lambda_0(\tilde{f} - f_\mu(x^*)). \tag{3.22}$$

This leads to the following lemma:

Lemma 3.5 *For all $k \geq k_2$ it holds that*

$$(\rho_{\lambda_0}^{(k+1)})^2 \leq (\rho_{\lambda_0}^{(k)})^2 - \frac{2}{3}t_k c_4\lambda_0(\tilde{f} - f_\mu(x^*)).$$

Proof.

For $k = k_2$, (3.18), (3.21), (3.22) hold. Combining these three relations with Lemma 3.4 it follows that

$$\begin{aligned} (\rho_{\lambda_0}^{(k+1)})^2 &\leq (\rho_{\lambda_0}^{(k)})^2 - 2t_k(1 - \frac{1}{3} - \frac{1}{3})c_4\lambda_0(\tilde{f} - f_\mu(x^*)) \\ &= (\rho_{\lambda_0}^{(k)})^2 - \frac{2}{3}t_k c_4\lambda_0(\tilde{f} - f_\mu(x^*)) \end{aligned} \tag{3.23}$$

for $k = k_2$.

Above inequality implies that $\rho_{\lambda_0}^{(k_2+1)} < \rho_{\lambda_0}^{(k_2)}$ and so (3.18), (3.21) and (3.22) hold for $k := k_2 + 1$ as well, and consequently (3.23) also holds for $k := k_2 + 1$. Applying this argument inductively we prove the lemma.

□

We assumed that $f_\mu(x^*) < \tilde{f} < f_\mu(\hat{x})$. Recursively applying Lemma 3.5 yields a contradiction since $\sum_{j=k_2}^{\infty} t_j = +\infty$. Hence, $f_\mu(\hat{x}) = f_\mu(x^*)$. Noting that f_μ is strictly convex, this shows that $\hat{x}$ has to be equal to x^*. To summarize, we have proven the following result.

Theorem 3.1 *It holds that*

$$\liminf_{k\to\infty} f_\mu(x^{(k)}) = f_\mu(x^*).$$

4 Global convergence

In this section we shall continue the convergence analysis and show that in fact the whole sequence of iterates is convergent under some conditions. Namely, we assume additionally that as $x^{(k)}$ gets close to x^*, the angle between the subgradient direction $d^{(k)}$ and $x^{(k)} - x^*$ is uniformly smaller than $\pi/2$. In mathematical terms, this means that

$$\liminf_{k\to\infty} (x^{(k)} - x^*)^T d^{(k)} / (\|x^{(k)} - x^*\| \cdot \|d^{(k)}\|) > 0.$$

This condition is also used in [14] to derive some convergence rate results for the classical subgradient method. We will call it the *acute angle condition* in this paper; it essentially requires that f_μ is not gully shaped.

Similar to (3.9), define

$$\rho_*(x) := \left\| (X^*)^{-1}(x^* - x) \right\|$$

and

$$\rho_*^{(k)} := \rho_*(x^{(k)}).$$

It follows that

$$\begin{aligned} (\rho_*^{(k+1)})^2 &= \left\| (X^*)^{-1}(x^* - x^{(k)} + t_k X^{(k)} p^{(k)}) \right\|^2 \\ &= (\rho_*^{(k)})^2 + 2t_k (x^* - x^{(k)})^T (X^*)^{-2} X^{(k)} p^{(k)} \\ &\quad + t_k^2 (p^{(k)})^T X^{(k)} (X^*)^{-2} X^{(k)} p^{(k)}. \end{aligned} \tag{4.1}$$

To simplify the expression, denote

$$\tilde{P}_{AX} := X^{-1} P_{AX} X.$$

Now we have,

$$\begin{aligned} X^{(k)} p^{(k)} &= X^{(k)} P_{AX^{(k)}} X^{(k)} d^{(k)} / \left\| P_{AX^{(k)}} X^{(k)} d^{(k)} \right\| \\ &= (X^{(k)})^2 \tilde{P}_{AX^{(k)}} d^{(k)} / \left\| P_{AX^{(k)}} X^{(k)} d^{(k)} \right\|. \end{aligned}$$

Hence,

$$(x^* - x^{(k)})^T (X^*)^{-2} X^{(k)} p^{(k)} = \frac{(x^* - x^{(k)})^T ((X^*)^{-1} X^{(k)})^2 \tilde{P}_{AX^{(k)}} d^{(k)}}{\left\| P_{AX^{(k)}} X^{(k)} d^{(k)} \right\|}. \tag{4.2}$$

By (3.11) we have

$$\left\| P_{AX^{(k)}} X^{(k)} d^{(k)} \right\| \le 1/c_4. \tag{4.3}$$

Remark that for all $y \in F$ it follows that

$$\begin{aligned} f_\mu(y) &\geq f_\mu(x^{(k)}) + (y - x^{(k)})^T d^{(k)} \\ &= f_\mu(x^{(k)}) + (y - x^{(k)})^T \tilde{P}_{AX^{(k)}} d^{(k)}. \end{aligned}$$

Hence, by definition we know that $\tilde{P}_{AX^{(k)}} d^{(k)}$ is a subgradient of f_μ over the feasible region F. Furthermore, by (3.3) we have

$$\left\| \tilde{P}_{AX^{(k)}} d^{(k)} \right\| / \left\| d^{(k)} \right\| \leq c_1$$

so that

$$\frac{(x^{(k)} - x^*)^T \tilde{P}_{AX^{(k)}} d^{(k)}}{\|x^{(k)} - x^*\| \cdot \left\| \tilde{P}_{AX^{(k)}} d^{(k)} \right\|} \geq \frac{1}{c_1} \frac{(x^{(k)} - x^*)^T d^{(k)}}{\|x^{(k)} - x^*\| \cdot \|d^{(k)}\|}.$$

Therefore, the acute angle condition also holds between the directions $\tilde{P}_{AX^{(k)}} d^{(k)}$ and $x^* - x^{(k)}$.

Now, from this acute angle condition it follows that there exist $\gamma > 0$ and a constant $0 < c_5 < 1$ such that if $x^{(k)}$ satisfies

$$\rho_*^{(k)} = \left\| (X^*)^{-1} x^{(k)} - e \right\| < \gamma$$

then

$$\begin{aligned} (x^{(k)} - x^*)^T ((X^*)^{-1} X^{(k)})^2 \tilde{P}_{AX^{(k)}} d^{(k)} &\geq c_5 (x^{(k)} - x^*)^T \tilde{P}_{AX^{(k)}} d^{(k)} \\ &= c_5 (x^{(k)} - x^*)^T d^{(k)} \\ &\geq c_5 (f_\mu(x^{(k)}) - f_\mu(x^*)) \end{aligned} \tag{4.4}$$

where the last inequality is the well known subgradient inequality of the convex function f_μ.

Assume for the moment that $\rho_*^{(k)} < \gamma$ is satisfied. Combining (4.4), (4.3) and (4.2) yields

$$(x^* - x^{(k)})^T (X^*)^{-2} X^{(k)} p^{(k)} \leq -c_4 c_5 (f_\mu(x^{(k)}) - f_\mu(x^*)).$$

Therefore, replacing the above inequality into (4.1) and using

$$\left\| (X^*)^{-1} x^{(k)} \right\|_\infty < 1 + \gamma \quad \text{and} \quad \left\| p^{(k)} \right\| = 1$$

we obtain

$$(\rho_*^{(k+1)})^2 \leq (\rho_*^{(k)})^2 - 2c_4 c_5 (f_\mu(x^{(k)}) - f_\mu(x^*)) t_k + (1 + \gamma)^2 t_k^2. \tag{4.5}$$

To carry on the analysis we use the following lemma:

Lemma 4.1 *For $x \in F$ and $\rho_*(x) < 1$ we have*

$$f_\mu(x) - f_\mu(x^*) \geq \frac{\mu}{2} \frac{\rho_*(x)^2}{1 + \rho_*(x)}.$$

PROOF.

Let $\eta \in \partial f(x^*)$ be such that

$$\eta - \mu(X^*)^{-1}e = 0.$$

Applying the subgradient inequality of $f(x)$ at x^* yields

$$\begin{aligned} f_\mu(x) - f_\mu(x^*) &= f(x) - f(x^*) - \mu \sum_{i=1}^n \log \frac{x_i}{x_i^*} \\ &\geq \eta^T(x - x^*) - \mu \sum_{i=1}^n \log \frac{x_i}{x_i^*} \\ &= \mu e^T(X^*)^{-1}(x - x^*) - \mu \sum_{i=1}^n \log \frac{x_i}{x_i^*}. \end{aligned}$$

It is well known that for any $\mid \alpha \mid < 1$ it holds (cf. [10]):

$$\log(1+\alpha) \leq \alpha - \frac{\alpha^2}{2(1+\mid \alpha \mid)}.$$

Therefore,

$$\begin{aligned} f_\mu(x) - f_\mu(x^*) &\geq \mu e^T(X^*)^{-1}(x - x^*) - \mu \sum_{i=1}^n \log(1 + \frac{x_i - x_i^*}{x_i^*}) \\ &\geq \frac{\mu}{2} \sum_{i=1}^n \frac{\left(\frac{x_i - x_i^*}{x_i^*}\right)^2}{1 + \mid \frac{x_i - x_i^*}{x_i^*} \mid} \\ &\geq \frac{\mu}{2} \frac{\rho_*(x)^2}{1 + \rho_*(x)}. \end{aligned}$$

□

By Theorem 3.1 we know that there exists k such that $\rho_*^{(k)} < \gamma < 1$. For such k, using (4.5) and Lemma 4.1 we obtain

$$(\rho_*^{(k+1)})^2 \leq (\rho_*^{(k)})^2 - \mu c_4 c_5 t_k (\rho_*^{(k)})^2 / (1 + \rho_*^{(k)}) + (1+\gamma)^2 t_k^2. \tag{4.6}$$

Now consider an arbitrary positive constant $\epsilon > 0$. For simplicity, assume $\epsilon < \gamma$.

Since $\lim_{k \to \infty} t_k = 0$, there is k_3 such that for all $k \geq k_3$ it holds that

$$(1+\gamma)^2 t_k < \mu c_4 c_5 (\epsilon/2)^2 / (1 + \epsilon/2) \tag{4.7}$$

and

$$(1+\gamma)^2 t_k^2 < \epsilon^2/2. \tag{4.8}$$

By Theorem 3.1 it follows that there is k_4 ($k_4 \geq k_3$) such that

$$\rho_*^{(k)} < \sqrt{2}\epsilon/2$$

for $k = k_4$.

To summarize, we have the following lemma:

Lemma 4.2 *For all $k \geq k_4$ it holds that*

$$\rho_*^{(k)} < \epsilon.$$

PROOF.

For $k = k_4$,

$$\rho_*^{(k)} < \sqrt{2}\epsilon/2 < \epsilon$$

by definition of k_4.

For all $k \geq k_4$ we have the following situation:

- If $\rho_*^{(k)} \leq \sqrt{2}\epsilon/2$, then by (4.6) and (4.8) it follows that

  $$(\rho_*^{(k+1)})^2 < (\rho_*^{(k)})^2 + (1+\gamma)^2 t_k^2 < \epsilon^2,$$

  and hence $\rho_*^{(k+1)} < \epsilon$.
- If $\rho_*^{(k)} > \sqrt{2}\epsilon/2 > \epsilon/2$, then by (4.6) and (4.7), and using $k_4 > k_3$ it follows that

  $$\rho_*^{(k+1)} < \rho_*^{(k)}.$$

Based on these two properties, the lemma follows easily by induction.

□

By Lemma 4.2, we have

$$\lim_{k\to\infty} \rho_*^{(k)} = 0,$$

and hence the following global convergence result is proven:

Theorem 4.1 *If the acute angle condition holds, then*

$$\lim_{k\to\infty} x^{(k)} = x^*.$$

5 Concluding remarks

Convex programming with a nondifferentiable objective and linear constraints can be encountered in the procedure of solving, e.g., stochastic programming problems or combinatorial optimization problems. The classical subgradient method requires an additional projection procedure in order to guarantee the feasibility of the iterates. In this paper we propose a new method which is a combination of the classical subgradient method and the affine scaling method belonging to the modern interior point methods. The convergence of the method is established. One of the advantages of the new method is that all the iterates will automatically stay in the interior of the feasible region. As is well known for the normal affine scaling algorithms, the step-length choice is crucial for a fast convergence. In practical implementations, a *long step* strategy can be adopted. We pose as an open question to decide the convergence status of the affine scaling subgradient method without adding the logarithmic barrier term to the original objective function. Finally, we remark that a path-following scheme can be established, so that one may trace down along the central path to the true optimal solution.

References

[1] J.R. Birge and F. Louveaux (1997), Introduction to Stochastic Programming, Springer-Verlag, New York.

[2] I.I. Dikin (1967), Iterative solutions of problems of linear and quadratic programming, *Soviet Mathematics Doklady* 8, 674-675.

[3] I.I. Dikin (1974), About the convergence of an iterative process (in Russian), *Upravlyaemye Sistemi* 12, 54-60.

[4] I.I. Dikin and V.I. Zorkaltsev (1980), Iterative solutions of mathematical programming problems, Nauka, Novosibirsk.

[5] K.R. Frisch (1955), The logarithmic potential method for convex programming, Institute of Economics, University of Oslo, Oslo, Norway.

[6] C.C. Gonzaga (1992), Path-following methods for linear programming, *SIAM Review* 34, 167-224.

[7] C.C. Gonzaga and L.A. Carlos (1990), A primal affine scaling algorithm for linearly constrained convex programs, Technical Report ES-238/90, Dept. of Systems Engineering and Computer Science, COPPE Federal University of Rio de Janeiro, Brazil.

[8] C.C. Gonzaga and R.A. Tapia (1997), On the convergence of the Mizuno-Todd-Ye algorithm to the analytic center of the solution set, *SIAM Journal on Optimization* 7, 47-65.

[9] J.-B. Hiriart-Urruty and C. Lemaréchal (1993), Convex analysis and minimization algorithms (vol. 1), Springer-Verlag, Berlin.

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

[11] R.D.C. Monteiro and T. Tsuchiya (1998), Global convergence of the affine scaling algorithm for convex programming, *SIAM Journal on Optimization* 8, 26-58.

[12] R.D.C. Monteiro and Y. Wang (1998), Trust region affine scaling algorithms for linearly constrained convex and concave programs, *Mathematical Programming* 80, 283-313.

[13] Y. Nesterov and A. Nemirovskii (1994), Interior-point polynomial algorithms in convex programming, SIAM Studies in Applied Mathematics, Philadelphia.

[14] N.Z. Shor (1985), Minimization methods for non-differentiable functions, Springer-Verlag, Berlin.

[15] G.W. Stewart (1989), On scaled projection and pseudoinverses, *Linear Algebra and Its Applications* 112, 189-193.

[16] J.F. Sturm and S. Zhang (1995), A dual and interior point approach to solve convex min-max problems, in D.-Z. Du and P.M. Pardalos eds., *Minimax and Applications*, 69-78, Kluwer.

[17] J. Sun (1993), A convergence proof for an affine-scaling algorithm for convex quadratic programming without nondegeneracy assumptions, *Mathematical Programming* 60, 69-79.

[18] P. Tseng and Z.Q. Luo (1992), On the convergence of the affine scaling algorithm, *Mathematical Programming* 56, 301-319.

[19] T. Tsuchiya (1992), Global convergence property of the affine scaling methods for primal degenerate linear programming problems, *Mathematics of Operations Research* 17, 527-557.

[20] T. Tsuchiya (1996), Affine scaling algorithm, in T. Terlaky ed., *Interior Point Methods of Mathematical Programming*, 35-82, Kluwer Academic Publishers.

[21] Y. Ye (1989), An extension of Karmarkar's algorithm and the trust region method for quadratic programming, in N. Megiddo ed., *Progress in Mathematical Programming: Interior-Point and Related Methods*, 49-63, Springer-Verlag.

[22] Y. Ye (1992), On affine scaling algorithms for nonconvex quadratic programming, *Mathematical Programming* 56, 285-300.

[23] Y. Ye and E. Tse (1989), An extension of Karmarkar's projective algorithm for convex quadratic programming, *Mathematical Programming* 44, 157-180.