An Infeasible Interior Proximal Method for Convex Programming Problems with Linear Constraints[1]

Nobuo Yamashita[2], Christian Kanzow[3], Tomoyuki Morimoto[2], and Masao Fukushima[2]

[2] Department of Applied Mathematics and Physics
Graduate School of Informatics
Kyoto University
Kyoto 606-8501
JAPAN

[3] Department of Mathematics
Center for Optimization and Approximation
University of Hamburg
Bundesstrasse 55
20146 Hamburg
GERMANY

October 23, 2000

Abstract. In this paper, we propose an infeasible interior proximal method for solving a convex programming problem with linear constraints. The interior proximal method proposed by Auslender and Haddou is a proximal method using a distance-like barrier function, and it has a global convergence property under mild assumptions. However this method is applicable only to problems whose feasible region has an interior point, because an initial point for the method must be chosen from the interior of the feasible region. The algorithm proposed in this paper is based on the idea underlying the infeasible interior point method for linear programming. This algorithm is applicable to problems whose feasible region may not have a nonempty interior, and it can be started from an arbitrary initial point. We establish global convergence of the proposed algorithm under appropriate assumptions.

[1]This research was supported in part by a Grant-in-Aid for Scientific Research from the Ministry of Education, Science, Sports, and Culture of Japan.

1 Introduction

Let $f : \Re^p \to (-\infty, +\infty]$ be a closed proper convex function and C be a polyhedral set in $\Re^p$ defined by

$$C := \{x \in \Re^p \,|\, A^T x \le b\},$$

where A is a $p \times m$ matrix, $b \in \Re^m$, and $m \ge p$. We consider the convex programming problem with linear constraints

$$(P) \qquad \min\{f(x) \,|\, x \in C\}.$$

In this paper, we propose an infeasible interior proximal method for solving (P). The proximal method, first proposed by Martinet [20], and subsequently studied by Rockafellar [26, 27], Güler [10], Lemaire [17, 18] and many others, is based on the concept of proximal mapping introduced by Moreau [21]. The proximal method for problem (P) with $C = \Re^p$ generates the sequence $\{x^k\}$ by the iterative scheme

$$x^k := \operatorname{argmin}\{f(x) + \lambda_k^{-1}\|x - x^{k-1}\|^2\}, \tag{1}$$

where $\{\lambda_k\}$ is a sequence of positive real numbers and $\|\cdot\|$ denotes the Euclidean norm. This method has a global convergence property under very mild conditions [10, 18, 27].

Many researchers have attempted to replace the quadratic term in (1) by distance-like functions [1, 3, 6, 7, 13, 14, 15, 28, 29]. For example, Censor and Zenios [6] proposed the proximal minimization with D-functions, which generates the sequence $\{x^k\}$ by

$$x^k := \operatorname{argmin}\{f(x) + \lambda_k^{-1} D(x, x^{k-1})\},$$

where the function D is the so-called Bregman's distance [5] or D-function. Teboulle [28] proposed the proximal method using φ-divergence introduced by Csiszár [8]. With φ-divergence, Iusem, Svaiter and Teboulle [13] proposed the entropy-like proximal method to solve the problem with nonnegative constraints

$$\min\{f(x) \,|\, x \in \Re^p_+\},$$

where $\Re^p_+ := \{u \in \Re^p \,|\, u_j \ge 0 \,\forall j = 1, \dots, p\}$. This method generates the sequence $\{x^k\}$ with initial point $x^0 \in \Re^p_{++}$ by

$$x^k := \operatorname{argmin}\{f(x) + \lambda_k^{-1}\tilde{d}_\varphi(x, x^{k-1}) \,|\, x \in \Re^p_+\},$$

where $\Re^p_{++} := \{u \in \Re^p \,|\, u_j > 0 \,\forall j = 1, \dots, p\}$, and the function $\tilde{d}_\varphi : \Re^p \times \Re^p \to \Re$ is defined by

$$\tilde{d}_\varphi(x, y) := \sum_{j=1}^{p} y_j \varphi(x_j / y_j)$$

with $\varphi : \Re_+ \to \Re$ being a differentiable strictly convex function. For the general linearly constrained problem, these methods can be applied to the dual problem, since the dual problem is represented as the problem with nonnegative constraints. Such dual methods are regarded as multiplier methods [26] for convex programming [4, 9, 13, 24, 30].

For the general linearly constrained convex programming problem, the interior point method is known to be efficient not only theoretically but also practically [22, 31, 32]. Recently, the interior proximal method, which enjoys some favorable properties of both proximal and interior point methods, has been proposed to solve problem (P), see, for example, [1, 2, 3, 29]. Auslender and Haddou [1] and Teboulle [29] proposed an algorithm based on the entropy-like distance. More recently, Auslender, Teboulle and Ben-Tiba [2, 3] proposed an algorithm based on a new function φ, which is constructed by adding the quadratic regularization term to the barrier term that enforces the generated sequence to remain in the interior of the feasible region. This method generates the sequence $\{x^k\}$ by

$$x^k := \operatorname{argmin}\{f(x) + \lambda_k^{-1} d_\varphi(L(x), L(x^{k-1})) \,|\, x \in C\}, \tag{2}$$

where $L(x) = b - A^T x$, and the distance-like function $d_\varphi : \Re^m \times \Re^m \to \Re$ is defined by

$$d_\varphi(u, v) := \sum_{j=1}^{m} v_j^2 \varphi(u_j/v_j). \tag{3}$$

It is worth mentioning that when the function φ is the logarithmic-quadratic kernel in the sense of [3], the function d_φ has the self-concordant property introduced by Nesterov and Nemirovski [22].

We note that, if the iteration (2) is started from an interior point of the feasible region, that is, $x^0 \in \operatorname{int} C$, then the generated sequence $\{x^k\}$ remains in the interior of the feasible region automatically. In [3], the global convergence of this algorithm was established under the following assumptions:

(H1) $\operatorname{dom} f \cap \operatorname{int} C \neq \emptyset$.

(H2) A has maximal row rank, i.e., rank $A = p$.

The restriction of this method is that one should start from an interior point of the feasible region, and hence problems whose feasible region have an empty interior may not be dealt with. Moreover, even if the underlying optimization problem has a nonempty interior, it is generally hard to find such a point. However, this is precisely what is required in order to start the algorithm from [3].

To remove this restriction, we propose in this paper an algorithm based on the idea of the infeasible interior point method for linear programming. We introduce a slack variable y such that $y \in \Re^m_+$ and $y \approx b - A^T x$. This allows us to start our algorithm from basically any initial point; moreover, the method can also be applied to problems whose feasible region may not have a nonempty interior. For the proposed algorithm we establish global convergence under appropriate assumptions.

This paper is organized as follows. In Section 2, we review some definitions and preliminary results that will be used in the subsequent analysis. In Section 3, we propose an infeasible interior proximal method, and show that it is well-defined. In Section 4, we establish global convergence of the proposed algorithm. In Section 5, we conclude the paper with some remarks.

All vector norms in this paper are Euclidean norms. The matrix norm is the corresponding spectral norm. The inner product in $\Re^p$ is denoted by $\langle \cdot, \cdot \rangle$.

2 Preliminaries

In this section, we review some preliminary results on the distance-like function defined by (3) that will play an important role in the subsequent analysis. We begin with the definition of the kernel φ that is used to define the distance-like function. Note that this definition is slightly more general than the one in [3].

Definition 2.1 *Φ denotes the class of closed proper convex functions $\varphi : \Re \to (-\infty, +\infty]$ satisfying the following conditions:*

(i) $\text{int}\,(\text{dom}\,\varphi) = (0, +\infty)$.

(ii) *φ is twice continuously differentiable on* $\text{int}\,(\text{dom}\,\varphi)$.

(iii) *φ is strictly convex on* $\text{dom}\,\varphi$.

(iv) $\lim\limits_{t\to 0^+} \varphi^{'}(t) = -\infty$.

(v) $\varphi(1) = \varphi^{'}(1) = 0$ *and* $\varphi^{''}(1) > 0$.

(vi) *There exists $\nu \in (\frac{1}{2}\varphi^{''}(1), \varphi^{''}(1))$ such that*

$$(1 - 1/t)(\varphi^{''}(1) + \nu(t-1)) \le \varphi^{'}(t) \le \varphi^{''}(1)(t-1) \;\; \forall t > 0.$$

Note that (vi) immediately implies $\lim\limits_{t\to+\infty} \varphi^{'}(t) = +\infty$. A few examples of functions $\varphi \in \Phi$ are shown below. These are given by Auslender et al. [3].

Example 2.1 *The following functions belong to the class Φ:*

$$\begin{array}{rcll} \varphi_1(t) & = & t\log t - t + 1 + \frac{\nu}{2}(t-1)^2, & \nu > 1,\ \text{dom}\,\varphi = [0, +\infty), \\ \varphi_2(t) & = & -\log t + t - 1 + \frac{\nu}{2}(t-1)^2, & \nu > 1,\ \text{dom}\,\varphi = (0, +\infty), \\ \varphi_3(t) & = & 2(\sqrt{t} - 1)^2 + \frac{\nu}{2}(t-1)^2, & \nu > 1,\ \text{dom}\,\varphi = [0, +\infty). \end{array}$$

The constant ν in these examples plays the role of the constant ν in Definition 2.1 (vi).

Based on a function $\varphi \in \Phi$, we define a distance-like function d_φ as follows.

Definition 2.2 *For a given $\varphi \in \Phi$, the distance-like function $d_\varphi : \Re^m \times \Re^m \to (-\infty, +\infty]$ is defined by*

$$d_\varphi(u, v) := \begin{cases} \sum_{j=1}^{m} v_j^2 \varphi(u_j/v_j), & \text{if } (u, v) \in \Re^m_{++} \times \Re^m_{++}, \\ +\infty, & \text{otherwise.} \end{cases} \tag{4}$$

From the strict convexity of φ and property (v), φ satisfies

$$\varphi(t) \geq 0 \ \ \forall t > 0, \quad \text{and} \quad \varphi(t) = 0 \iff t = 1.$$

Hence, d_φ satisfies

$$\begin{aligned} d_\varphi(u,v) &\geq 0 \quad \forall (u,v) \in \Re^m_{++} \times \Re^m_{++}, \qquad \text{and} \\ d_\varphi(u,v) &= 0 \iff u = v. \end{aligned} \tag{5}$$

Next, we introduce two technical results on nonnegative sequences of real numbers that will be needed in the subsequent analysis.

Lemma 2.1 *Let $\{v_k\}$ and $\{\beta_k\}$ be nonnegative sequences of real numbers satisfying*

(i) $v_{k+1} \leq v_k + \beta_k$,

(ii) $\sum_{k=1}^{\infty} \beta_k < \infty$.

Then the sequence $\{v_k\}$ converges.

Proof. See [23, Chapter 2]. □

Lemma 2.2 *Let $\{\lambda_j\}$ be a sequence of positive numbers, and $\{a_j\}$ be a sequence of real numbers. Let $\sigma_k := \sum_{j=1}^{k} \lambda_j$ and $b_k := \sigma_k^{-1} \sum_{j=1}^{k} \lambda_j a_j$. If $\sigma_k \to \infty$, then*

(i) $\liminf_{k\to\infty} a_k \leq \liminf_{k\to\infty} b_k \leq \limsup_{k\to\infty} b_k \leq \limsup_{k\to\infty} a_k$.

(ii) *If $\lim_{k\to\infty} a_k = a < \infty$, then $\lim_{k\to\infty} b_k = a$.*

Proof. (i) See [19, Lemma 3.5].
(ii) This result, originally given in [16], follows immediately from (i). □

3 Algorithm

In this section, we propose an infeasible interior proximal method for the solution of problem (P). In order to motivate this method, consider the following iterative scheme with a sequence $\{\delta^k\} \subset \Re^m_{++}$ such that $\delta^k \to 0$:

$$(x^k, y^k) := \operatorname{argmin}\{f(x) + \lambda_k^{-1} \hat{d}_k(y) \,|\, y - (b - A^T x) = \delta^k, x \in \Re^p, y \in \Re^m_{++}\}, \tag{6}$$

where

$$\hat{d}_k(y) := d_\varphi(y, y^{k-1}).$$

Since $y = b - A^T x + \delta^k \in \Re^m_{++}$ for any feasible point of (6), the feasible region of problem (6) may be identified with the set

$$C_k := \{x \in \Re^p \mid A^T x \leq b + \delta^k\}, \tag{7}$$

which is considered a perturbation of the original feasible region C. Moreover, if $C \neq \emptyset$, then $C \subset \text{int } C_k \neq \emptyset$ for all k. Since $\delta^k \to 0$ as $k \to \infty$, the sequence $\{C_k\}$ converges to the set C. The optimality conditions for (6) are given by

$$\begin{array}{rcl} \partial f(x) + Au & \ni & 0, \\ \lambda_k^{-1} \nabla \hat{d}_k(y) + u & = & 0, \\ y - (b - A^T x) & = & \delta^k, \end{array} \tag{8}$$

where $u = (u_1, \ldots, u_m)^T \in \Re^m$ denotes the vector of Lagrange multipliers. From the definition of $\hat{d}_k$ and (4), we have

$$\nabla \hat{d}_k(y) = (y_1^{k-1} \varphi'(y_1/y_1^{k-1}), \ldots, y_m^{k-1} \varphi'(y_m/y_m^{k-1}))^T. \tag{9}$$

Hence (8) can be rewritten as

$$\begin{array}{rcl} \partial f(x) - \lambda_k^{-1} \displaystyle\sum_{i=1}^m a_i y_i^{k-1} \varphi'(y_i/y_i^{k-1}) & \ni & 0, \\ y - (b - A^T x) & = & \delta^k, \end{array} \tag{10}$$

where a_i denotes the ith column of the matrix A. This means that solving (6) is equivalent to finding (x^k, y^k) that satisfies (10). The algorithm proposed below generates a sequence of points that satisfy conditions (10) approximately. To this end, we replace the subdifferential $\partial f(x)$ in (10) by the ϵ-subdifferential; recall that, for an arbitrary $\epsilon \geq 0$, the ϵ-subdifferential of f at a point x is defined by

$$\partial_\epsilon f(x) := \{g \in \Re^p \mid f(y) \geq f(x) + \langle g, y - x \rangle - \epsilon \, \forall y \in \Re^p\}.$$

Now we describe the algorithm.

Algorithm

Step 1. Choose $\varphi \in \Phi$, a positive sequence $\{\beta_k\}$ such that $\sum_{k=1}^\infty \beta_k < \infty$, a sequence $\{\lambda_k\}$ with $\lambda_k \in [\lambda_{\min}, \lambda_{\max}]$ for all k and some constants $0 < \lambda_{\min} \leq \lambda_{\max}$, and a parameter $\rho \in (0, 1)$. Choose initial points $x^0 \in \Re^p$ and $y^0 \in \Re^m_{++}$ such that $\delta^0 := y^0 - (b - A^T x^0) \in \Re^m_{++}$, and set $k := 1$.

Step 2. Terminate the iteration if a suitable stopping rule is satisfied.

Step 3. Choose $\epsilon_k \geq 0$ satisfying $\epsilon_k \leq \beta_{k+1} \lambda_k^{-1}$, $\eta_k \in (0, \rho]$, and set $\delta^k := \eta_k \delta^{k-1}$. Find $(x^k, y^k) \in \Re^p \times \Re^m_{++}$ and $g^k \in \Re^p$ such that

$$g^k \in \partial_{\epsilon_k} f(x^k), \tag{11}$$

$$g^k - \lambda_k^{-1} \sum_{i=1}^m a_i y_i^{k-1} \varphi'(y_i^k / y_i^{k-1}) = 0, \tag{12}$$

$$y^k - (b - A^T x^k) = \delta^k. \tag{13}$$

Step 4. Set $k := k+1$, and return to Step 2.

In the above algorithm, the initial point can be chosen arbitrarily as long as δ^0 belongs to $\Re^m_{++}$. Namely, we only have to choose y^0 big enough to guarantee $\delta^0 \in \Re^m_{++}$. Note that the choice of the sequences $\{\epsilon_k\}$, $\{\lambda_k\}$, and $\{\delta^k\}$ ensures $\sum_{k=1}^{\infty} \epsilon_k < \infty$, $\sum_{k=1}^{\infty} \lambda_k = \infty$, $\sum_{k=1}^{\infty} \|\delta^k\| < \infty$, and $\sum_{k=1}^{\infty} \lambda_k \epsilon_k < \infty$.

It is important to guarantee the existence of (x^k, y^k) and g^k satisfying (11)–(13) in Step 3. In fact, the next theorem shows that the proposed algorithm is well-defined under the following two assumptions:

(A1) $\mathrm{dom} f \cap C \neq \emptyset$.

(A2) A has maximal row rank, i.e., rank $A = p$.

Note that the first assumption is rather natural since otherwise the optimal value of problem (P) would be $+\infty$. We further stress that (A1) is somewhat weaker than the corresponding condition (H1) mentioned in the introduction. It implies, however, that the feasible set C is nonempty.

Also the second assumption is often satisfied, e.g., if we have nonnegativity constraints on the variables like in the Lagrange dual of an arbitrary constrained optimization problem. Condition (A2) can also be stated without loss of generality if we view (P) as the dual of a linear program since then the matrix A plays the role of the constraint matrix for the equality constraints in the primal formulation, so that linearly dependent rows can be deleted from A without changing the problem itself.

We next recall that the recession function F_∞ for a convex function $F : \Re^p \to (-\infty, +\infty]$ is given by

$$F_\infty(d) := \lim_{t \to +\infty} \frac{F(x+td) - F(x)}{t},$$

where $x \in \mathrm{dom}\, F$. Note that the value of $F_\infty(d)$ does not depend on the choice of $x \in \mathrm{dom}\, F$ on the right-hand side. We further recall that if $F_\infty(d) > 0$ for all $d \neq 0$, then the set of minimizers of F is nonempty and compact [25, Theorem 27.1 (d)].

Theorem 3.1 *Let assumptions (A1) and (A2) be satisfied. Then, for any* $x^{k-1} \in \mathrm{int}\, C_{k-1}$, $y^{k-1} \in \Re^m_{++}$, $\lambda_k > 0$, $\epsilon_k \geq 0$, $\delta^k \in \Re^m_{++}$, *there exist* $x^k \in \mathrm{int}\, C_k$, $y^k \in \Re^m_{++}$ *and* $g^k \in \Re^p$ *satisfying* (11)–(13).

Proof. Let $F_k : \Re^p \to (-\infty, +\infty]$ be defined by

$$F_k(x) := f(x) + \lambda_k^{-1} \hat{d}_k(b - A^T x + \delta^k),$$

and C_k be defined by (7). Since $C \subset C_k$, (A1) implies dom $F_k \neq \emptyset$. Let $S_k := \mathrm{argmin}_x F_k(x)$. We show that S_k is nonempty. For this purpose, it is sufficient to show that

$$(F_k)_\infty(d) > 0 \ \ \forall d \neq 0. \tag{14}$$

From the definition of F_k, we have

$$(F_k)_\infty(d) = f_\infty(d) + \lambda_k^{-1}(\hat{d}_k)_\infty(-A^T d). \tag{15}$$

Since $(\hat{d}_k)_\infty(-A^T d) = +\infty$ for all $d \neq 0$ from rank $A = p$ by (A2) and the properties of φ (conditions (iv) and (vi) of Definition 2.1), it follows from (15) that $(F_k)_\infty(d) = +\infty$ for all $d \neq 0$. Thus we have $S_k \neq \emptyset$, which means that there exists an $x \in \Re^p$ satisfying

$$0 \in \partial F_k(x) = \partial f(x) - \lambda_k^{-1} A \nabla_y \hat{d}_k(b - A^T x + \delta^k).$$

Since $\partial f(x) \subseteq \partial_\epsilon f(x)$ for all $\epsilon \geq 0$, there exist x^k, y^k and g^k satisfying

$$\begin{aligned} & g^k \in \partial_{\epsilon_k} f(x^k), \\ & \lambda_k g^k - A \nabla \hat{d}_k(y^k) = 0, \\ & y^k - (b - A^T x^k) = \delta^k. \end{aligned}$$

From (9), this completes the proof. □

Note that the strict convexity of $\hat{d}_k$ and assumption (A2) imply that F_k is also strictly convex. Therefore, if a minimizer of F_k exists, then it is unique. In order to get another interpretation of the vectors computed in Step 3 of the above algorithm, let us give the following two remarks.

Remark 3.1 *Let*

$$\epsilon\text{-argmin} f(x) := \{x \mid f(x) \leq \inf_y f(y) + \epsilon\}.$$

Then, it is easy to see that

$$0 \in \partial_\epsilon f(x) \iff x \in \epsilon\text{-argmin} f(x).$$

In fact, from the definition of the ϵ-subgradient, we have

$$0 \in \partial_\epsilon f(x) \iff f(y) \geq f(x) - \epsilon \ \forall y \iff \inf_y f(y) + \epsilon \geq f(x).$$

Remark 3.2 *Let $F : \Re^p \to (-\infty, +\infty]$ be defined by $F(x) = f(x) + h(x)$, where f and h are both convex functions. Even if h is differentiable, it is not necessarily true that*

$$\partial_\epsilon F(x) = \partial_\epsilon f(x) + \nabla h(x),$$

although $\partial F(x) = \partial f(x) + \nabla h(x)$ holds. It is known that

$$\partial_\epsilon F(x) = \bigcup_{\substack{\epsilon_1 \geq 0, \epsilon_2 \geq 0 \\ \epsilon = \epsilon_1 + \epsilon_2}} \{\partial_{\epsilon_1} f(x) + \partial_{\epsilon_2} h(x)\}$$

holds [*11*, Theorem 2.1]. *Therefore, we have in general*

$$\partial_\epsilon F(x) \supseteq \partial_\epsilon f(x) + \nabla h(x).$$

*If $(F)_\infty(d) > 0$ for all $d \neq 0$, then $S :=$*argmin *$F(x)$ is nonempty and compact. Thus, there exists an x such that $0 \in \partial F(x) = \partial f(x) + \nabla h(x)$. Since $\partial f(x) \subseteq \partial_\epsilon f(x)$ for all $\epsilon \geq 0$, there exists an x satisfying $0 \in \partial_\epsilon f(x) + \nabla h(x)$, namely, there exist x and g such that*

$$\begin{cases} g \in \partial_\epsilon f(x), \\ g + \nabla h(x) = 0. \end{cases}$$

From Remarks 3.1 and 3.2, it can be shown that

$$0 \in \partial_\epsilon f(x) + \nabla h(x)$$

implies $x \in \epsilon$-argmin $F(x)$. Thus x^k obtained in Step 3 belongs to ϵ_k-argmin $F_k(x)$.

4 Global Convergence

In this section, we establish global convergence of the proposed algorithm. To this end, we assume throughout this section that the sequence $\{x^k\}$ generated by the algorithm is infinite.

To begin with, we prove some key lemmas. The first lemma is a slight modification of Lemma 3.4 in [3].

Lemma 4.1 Let $\varphi \in \Phi$. For any $a, b \in \Re^m_{++}$ and $c \in \Re^m_{+}$, we have

$$\langle c - b, \Phi'(b/a)\rangle \le \frac{1}{2}\theta \left(\|c-a\|^2 - \|c-b\|^2\right),$$

where $\Phi'(b/a) := (a_1\varphi'(b_1/a_1), \ldots, a_m\varphi'(b_m/a_m))^T$ and $\theta := \varphi''(1)$.

Proof. By the definition of φ, we have $\varphi'(t) \le \varphi''(1)(t-1)$ for all $t > 0$. Letting $t = b_i/a_i$ and multiplying both sides by $a_i c_i$ yield

$$a_i c_i \varphi'(b_i/a_i) \le a_i c_i \varphi''(1)(b_i/a_i - 1) = c_i \varphi''(1)(b_i - a_i). \tag{16}$$

Moreover, since there exists $\nu \in (\frac{1}{2}\varphi''(1), \varphi''(1))$ such that $-\varphi'(t) \le -\varphi''(1)(1-1/t) - \nu(t-1)^2/t$ for all $t > 0$, letting $t = b_i/a_i$ again and multiplying both sides by $a_i b_i$ give

$$\begin{aligned} -a_i b_i \varphi'(b_i/a_i) &\le -a_i b_i \varphi''(1)(1 - a_i/b_i) - a_i b_i \nu (b_i/a_i - 1)^2/(b_i/a_i) \\ &= -a_i \varphi''(1)(b_i - a_i) - \nu(b_i - a_i)^2. \end{aligned} \tag{17}$$

Using the identity

$$\langle b - a, c - a\rangle = \frac{1}{2}\left(\|c-a\|^2 - \|c-b\|^2 + \|b-a\|^2\right),$$

adding the above two inequalities (16) and (17), and summing over $i = 1, \ldots, m$ yield

$$\begin{aligned} \langle c - b, \Phi'(b/a)\rangle &\le \varphi''(1)\langle b-a, c-a\rangle - \nu\|b-a\|^2 \\ &= \frac{1}{2}\varphi''(1)\left(\|c-a\|^2 - \|c-b\|^2\right) + \left(\frac{1}{2}\varphi''(1) - \nu\right)\|b-a\|^2 \\ &\le \frac{1}{2}\varphi''(1)\left(\|c-a\|^2 - \|c-b\|^2\right), \end{aligned}$$

where the last inequality follows from $\nu \ge \varphi''(1)/2$. □

Let $\Pi_C(x)$ denote the projection of x onto the set C, and let $[x]_+$ be the projection of x onto the nonnegative orthant.

Lemma 4.2 *Let* $\operatorname{dist}(x^k, C)$ *denote the distance between* x^k *and* $\Pi_C(x^k)$*. Then there exists a constant* $\alpha > 0$ *such that*

$$\operatorname{dist}(x^k, C) = \|\Pi_C(x^k) - x^k\| \le \alpha \|\delta^k\| \tag{18}$$

for all k*. Moreover,* $\operatorname{dist}(x^k, C) \to 0$ *as* $k \to \infty$*.*

Proof. From Hoffman's lemma [12], there exists a constant $\alpha > 0$ such that

$$\operatorname{dist}(x^k, C) = \|\Pi_C(x^k) - x^k\| \le \alpha \|[A^T x^k - b]_+\|$$

for all k. Hence, we have

$$\begin{aligned} \operatorname{dist}(x^k, C) &\le \alpha \|[A^T x^k - b]_+\| \\ &= \alpha \|[A^T x^k - b]_+ - [-y^k]_+\| \\ &\le \alpha \|A^T x^k - b + y^k\| \\ &= \alpha \|\delta^k\|, \end{aligned}$$

where the first equality follows from $y^k \in \Re^m_{++}$, the second inequality follows from the nonexpansiveness of the projection operator, and the last equality follows from $A^T x^k - b + y^k = \delta^k$. Finally (18) and the fact that $\delta^k \to 0$ imply $\operatorname{dist}(x^k, C) \to 0$. □

By using this lemma, we can also estimate the distance between y^k and $b - A^T \Pi_C(x^k)$.

Lemma 4.3 *The following statements hold:*

(i) There exists a positive constant β_1 *such that, for all* k*,*

$$\|b - A^T \Pi_C(x^k) - y^k\| \le \beta_1 \|\delta^k\|.$$

(ii) There exist positive constants β_2 *and* β_3 *such that, for all* k*,*

$$\|b - A^T \Pi_C(x^{k-1}) - y^k\|^2 \ge \|y^k - y^{k-1}\|^2 - \beta_2 \|\delta^{k-1}\| - \beta_3 \|\delta^{k-1}\| \, \|y^k - y^{k-1}\|.$$

Proof. (i) From Lemma 4.2 we have

$$\begin{aligned} \|b - A^T \Pi_C(x^k) - y^k\| &= \|b - A^T \Pi_C(x^k) - (b - A^T x^k + \delta^k)\| \\ &\le \|A^T (\Pi_C(x^k) - x^k)\| + \|\delta^k\| \\ &\le \alpha \|A\| \, \|\delta^k\| + \|\delta^k\| \end{aligned}$$

for all k. Setting $\beta_1 := \alpha \|A\| + 1$, we get the desired inequality.

(ii) From (i) we have

$$\begin{aligned} \|y^k - y^{k-1}\| &\le \|b - A^T \Pi_C(x^{k-1}) - y^k\| + \|b - A^T \Pi_C(x^{k-1}) - y^{k-1}\| \\ &\le \|b - A^T \Pi_C(x^{k-1}) - y^k\| + \beta_1 \|\delta^{k-1}\| \end{aligned}$$

for all k. Squaring both sides of the inequality, we get

$$\begin{aligned} &\|y^k - y^{k-1}\|^2 \\ &\quad \le \|b - A^T \Pi_C(x^{k-1}) - y^k\|^2 + \beta_1^2 \|\delta^{k-1}\|^2 + 2\beta_1 \|\delta^{k-1}\| \, \|b - A^T \Pi_C(x^{k-1}) - y^k\|. \end{aligned}$$

Hence, we obtain

$$
\begin{aligned}
&\|b - A^T\Pi_C(x^{k-1}) - y^k\|^2 \\
&\quad \geq \|y^k - y^{k-1}\|^2 - \beta_1^2\|\delta^{k-1}\|^2 - 2\beta_1\|\delta^{k-1}\|\,\|b - A^T\Pi_C(x^{k-1}) - y^k\| \\
&\quad \geq \|y^k - y^{k-1}\|^2 - \beta_1^2\|\delta^{k-1}\|^2 - 2\beta_1\|\delta^{k-1}\|\left(\|b - A^T\Pi_C(x^{k-1}) - y^{k-1}\| + \|y^k - y^{k-1}\|\right) \\
&\quad \geq \|y^k - y^{k-1}\|^2 - 3\beta_1^2\|\delta^{k-1}\|^2 - 2\beta_1\|\delta^{k-1}\|\,\|y^k - y^{k-1}\|,
\end{aligned}
$$

where the last inequality follows from (i). Since $\delta^k \to 0$, there exists $\beta_2 > 0$ such that

$$3\beta_1^2\|\delta^{k-1}\|^2 \leq \beta_2\|\delta^{k-1}\|$$

for all k. Setting $\beta_3 := 2\beta_1$, we have (ii). □

The next lemma shows the relations between two consecutive iterates, which are the basis for establishing global convergence of the algorithm.

Lemma 4.4 *Let* $\theta := \varphi''(1)$.

(i) For any $x \in C$, $y = b - A^T x \in \Re^m_+$, *and for all* k, *we have*

$$\lambda_k\left(f(x^k) - f(x)\right) \leq \frac{1}{2}\theta\left(\|y - y^{k-1}\|^2 - \|y - y^k\|^2\right) + \theta\|\delta^{k-1}\|\,\|y^{k-1} - y^k\| + \lambda_k\epsilon_k. \quad (19)$$

(ii) There exist positive constants γ_1, γ_2 *and* γ_3 *such that, for all* k, *we have*

$$
\begin{aligned}
f(\Pi_C(x^k)) - f(\Pi_C(x^{k-1})) \;\leq\; & \gamma_1\|\delta^{k-1}\|\,\|y^k - y^{k-1}\| - \gamma_2\|y^{k-1} - y^k\|^2 \\
& + \gamma_3\|\delta^{k-1}\| + |f(\Pi_C(x^k)) - f(x^k)| + \epsilon_k. \quad (20)
\end{aligned}
$$

Proof. (i) For any $x \in C$ and $y = b - A^T x \in \Re^m_+$, we have

$$
\begin{aligned}
&\lambda_k(f(x^k) - f(x)) \\
&\quad \leq \langle \lambda_k g^k, x^k - x\rangle + \lambda_k\epsilon_k \\
&\quad = \langle A\nabla\hat{d}_k(y^k), x^k - x\rangle + \lambda_k\epsilon_k \\
&\quad = \langle y + \delta^k - y^k, \nabla\hat{d}_k(y^k)\rangle + \lambda_k\epsilon_k \\
&\quad = \langle y - y^k, \nabla\hat{d}_k(y^k)\rangle + \langle \delta^k, \nabla\hat{d}_k(y^k)\rangle + \lambda_k\epsilon_k \\
&\quad \leq \frac{1}{2}\theta\left(\|y - y^{k-1}\|^2 - \|y - y^k\|^2\right) + \sum_{i=1}^m \delta_i^k y_i^{k-1}\varphi'(y_i^k/y_i^{k-1}) + \lambda_k\epsilon_k \\
&\quad \leq \frac{1}{2}\theta\left(\|y - y^{k-1}\|^2 - \|y - y^k\|^2\right) + \sum_{i=1}^m \delta_i^k y_i^{k-1}\varphi''(1)(y_i^k/y_i^{k-1} - 1) + \lambda_k\epsilon_k \\
&\quad = \frac{1}{2}\theta\left(\|y - y^{k-1}\|^2 - \|y - y^k\|^2\right) + \theta\langle \delta^k, y^k - y^{k-1}\rangle + \lambda_k\epsilon_k \\
&\quad \leq \frac{1}{2}\theta\left(\|y - y^{k-1}\|^2 - \|y - y^k\|^2\right) + \theta\|\delta^k\|\,\|y^k - y^{k-1}\| + \lambda_k\epsilon_k \\
&\quad \leq \frac{1}{2}\theta\left(\|y - y^{k-1}\|^2 - \|y - y^k\|^2\right) + \theta\|\delta^{k-1}\|\,\|y^k - y^{k-1}\| + \lambda_k\epsilon_k,
\end{aligned}
$$

where the first inequality follows from (11), the first equality follows from (12), the second equality follows from (13), the second inequality follows from the definition of $\nabla \hat{d}_k(y^k)$ and Lemma 4.1, the third inequality follows from the property (vi) of φ, the fourth equality is the definition of θ, the fourth inequality is a consequence of the Cauchy-Schwarz inequality, and the last inequality follows from the updating rules for δ^k.

(ii) Applying (19) with $(x, y) = (\Pi_C(x^{k-1}), b - A^T \Pi_C(x^{k-1}))$, we have

$$\begin{aligned}
&\lambda_k \left(f(x^k) - f(\Pi_C(x^{k-1})) \right) \\
&\leq \frac{1}{2}\theta \left(\|b - A^T \Pi_C(x^{k-1}) - y^{k-1}\|^2 - \|b - A^T \Pi_C(x^{k-1}) - y^k\|^2 \right) \\
&\quad + \theta \|\delta^{k-1}\| \, \|y^{k-1} - y^k\| + \lambda_k \epsilon_k
\end{aligned}$$

for all k. This implies

$$\begin{aligned}
& f(\Pi_C(x^k)) - f(\Pi_C(x^{k-1})) \\
&= f(\Pi_C(x^k)) - f(x^k) + f(x^k) - f(\Pi_C(x^{k-1})) \\
&\leq \frac{\theta}{2\lambda_k} \left(\|b - A^T \Pi_C(x^{k-1}) - y^{k-1}\|^2 - \|b - A^T \Pi_C(x^{k-1}) - y^k\|^2 \right) \\
&\quad + \frac{\theta}{\lambda_k} \|\delta^{k-1}\| \, \|y^{k-1} - y^k\| + |f(\Pi_C(x^k)) - f(x^k)| + \epsilon_k \\
&\leq \frac{\theta}{2\lambda_{\min}} \|b - A^T \Pi_C(x^{k-1}) - y^{k-1}\|^2 - \frac{\theta}{2\lambda_{\max}} \|b - A^T \Pi_C(x^{k-1}) - y^k\|^2 \\
&\quad + \frac{\theta}{\lambda_{\min}} \|\delta^{k-1}\| \, \|y^{k-1} - y^k\| + |f(\Pi_C(x^k)) - f(x^k)| + \epsilon_k \\
&\leq \frac{\theta \beta_1^2}{2\lambda_{\min}} \|\delta^{k-1}\|^2 - \frac{\theta}{2\lambda_{\max}} \left(\|y^k - y^{k-1}\|^2 - \beta_2 \|\delta^{k-1}\| - \beta_3 \|\delta^{k-1}\| \, \|y^k - y^{k-1}\| \right) \\
&\quad + \frac{\theta}{\lambda_{\min}} \|\delta^{k-1}\| \, \|y^{k-1} - y^k\| + |f(\Pi_C(x^k)) - f(x^k)| + \epsilon_k \\
&= \theta \Big(\frac{\beta_3}{2\lambda_{\max}} + \frac{1}{\lambda_{\min}} \Big) \|\delta^{k-1}\| \, \|y^k - y^{k-1}\| - \frac{\theta}{2\lambda_{\max}} \|y^k - y^{k-1}\|^2 \\
&\quad + \frac{\theta}{2} \Big(\frac{\beta_2}{\lambda_{\max}} + \frac{\beta_1^2}{\lambda_{\min}} \|\delta^{k-1}\| \Big) \|\delta^{k-1}\| + |f(\Pi_C(x^k)) - f(x^k)| + \epsilon_k,
\end{aligned}$$

where the second inequality follows from $\lambda_k \in [\lambda_{\min}, \lambda_{\max}]$, and the last inequality follows from Lemma 4.3. Since $\delta^k \to 0$, we have (ii). □

We finally introduce another technical lemma that will also be used in order to prove a global convergence theorem for the algorithm. To show this, we need the following further assumptions.

(A3) $\sum_{k=0}^{\infty} |f(x^k) - f(\Pi_C(x^k))| < \infty$.

(A4) $f_* := \inf\{f(x) \,|\, x \in C\} > -\infty$.

While (A4) is rather natural, we will include a short discussion on the necessity of (A3) at the end of this section. For now, we just mention that (A3) requires the initial point x^0 and the sequence $\{\Pi_C(x^k)\}$ to lie in $\operatorname{dom} f$. (The algorithm ensures $x^k \in \operatorname{dom} f$ for all $k \geq 1$.) This restriction can be relaxed to the weaker assumption that $\sum_{k=\bar{k}}^{\infty} |f(x^k) - f(\Pi_C(x^k))| < \infty$ for some $\bar{k} > 0$, which will lead to slight modifications in the proofs of the subsequent lemmas and theorems.

Lemma 4.5 *Suppose that (A1)–(A4) hold. Then*

$$\sum_{k=1}^{\infty} \|\delta^{k-1}\| \, \|y^k - y^{k-1}\| < \infty.$$

Proof. We only have to show that the sequence $\{\|y^k - y^{k-1}\|\}$ is bounded since then the assertion follows immediately from the fact that $\sum_{k=1}^{\infty} \|\delta^{k-1}\| < \infty$.

Assume the sequence $\{\|y^k - y^{k-1}\|\}$ is unbounded. Then there is a subsequence $\{\|y^k - y^{k-1}\|\}_{k \in K}$ such that $\|y^k - y^{k-1}\| \to \infty$ for $k \in K$, whereas the complementary subsequence $\{\|y^k - y^{k-1}\|\}_{k \notin K}$ is bounded (note that this complementary subsequence could be finite or even empty). Summing the inequalities (20) over $j = 1, 2, \ldots, k$ gives

$$\begin{aligned}
& f(\Pi_C(x^k)) - f(\Pi_C(x^0)) \\
\leq\ & \gamma_1 \sum_{j=1}^{k} \|\delta^{j-1}\| \, \|y^j - y^{j-1}\| - \gamma_2 \sum_{j=1}^{k} \|y^j - y^{j-1}\|^2 + \gamma_3 \sum_{j=1}^{k} \|\delta^{j-1}\| \\
& + \sum_{j=1}^{k} |f(x^j) - f(\Pi_C(x^j))| + \sum_{j=1}^{k} \epsilon_j \\
\leq\ & \gamma_1 \sum_{j=1}^{k} \|\delta^{j-1}\| \, \|y^j - y^{j-1}\| - \gamma_2 \sum_{\substack{j=1 \\ j \in K}}^{k} \|y^j - y^{j-1}\|^2 + \gamma_3 \sum_{j=1}^{k} \|\delta^{j-1}\| \\
& + \sum_{j=1}^{k} |f(x^j) - f(\Pi_C(x^j))| + \sum_{j=1}^{k} \epsilon_j \\
=\ & \sum_{\substack{j=1 \\ j \in K}}^{k} \|y^j - y^{j-1}\| (\gamma_1 \|\delta^{j-1}\| - \gamma_2 \|y^j - y^{j-1}\|) + \gamma_3 \sum_{j=1}^{k} \|\delta^{j-1}\| \\
& + \gamma_1 \sum_{\substack{j=1 \\ j \notin K}}^{k} \|\delta^{j-1}\| \, \|y^j - y^{j-1}\| + \sum_{j=1}^{k} |f(x^j) - f(\Pi_C(x^j))| + \sum_{j=1}^{k} \epsilon_j.
\end{aligned} \tag{21}$$

Now let us recall that we have $\sum_{k=1}^{\infty} \epsilon_k < \infty$, $\sum_{k=1}^{\infty} \|\delta^{k-1}\| < \infty$, and $\sum_{k=1}^{\infty} |f(x^k) - f(\Pi_C(x^k))| < \infty$ by (A3). Furthermore, the definition of the index set K also implies

$$\sum_{\substack{j=1 \\ j \notin K}}^{\infty} \|\delta^{j-1}\| \, \|y^j - y^{j-1}\| < \infty$$

since the sequence $\{\|y^j - y^{j-1}\|\}_{j \notin K}$ is bounded and

$$\sum_{\substack{j=1 \\ j \notin K}}^{\infty} \|\delta^{j-1}\| < \infty.$$

On the other hand, we have

$$\sum_{\substack{j=1\\ j\in K}}^{k} \|y^j - y^{j-1}\| \left(\gamma_1\|\delta^{j-1}\| - \gamma_2\|y^j - y^{j-1}\|\right) \to -\infty$$

for $k \to \infty$ since the term in brackets eventually becomes less than a negative constant due to the fact that $\|\delta^{j-1}\| \to 0$ for $j \in K$ and $\{\|y^j - y^{j-1}\|\}_{j\in K}$ is unbounded. Therefore, taking the limit $k \to \infty$ in (21), we see that $\{f(\Pi_C(x^k))\}$ is not bounded from below or $f(\Pi_C(x^0)) = \infty$, which contradicts (A4). This completes the proof. □

We are now in the position to state our first global convergence result. It deals with the behavior of the sequence $\{f(x^k)\}$.

Theorem 4.1 *Suppose that (A1)–(A4) hold. Then we have* $\lim_{k\to\infty} f(x^k) = f_*$*, i.e.,* $\{x^k\}$ *is a minimizing sequence.*

Proof. Throughout this proof, we use the notation $\sigma_k := \sum_{j=1}^k \lambda_j$. Note that $\sigma_k \to \infty$ since $\{\lambda_j\}$ is bounded from below by the positive number $\lambda_{\min}$.

Summing the inequalities (19) over $j = 1, \ldots, k$, we obtain for any $x \in C$ and $y = b - A^T x \in \Re^m_+$

$$\begin{aligned} & -\sigma_k f(x) + \sum_{j=1}^k \lambda_j f(x^j) \\ \le\ & \frac{1}{2}\theta\left(\|y - y^0\|^2 - \|y - y^k\|^2\right) + \theta \sum_{j=1}^k \|\delta^{j-1}\|\,\|y^j - y^{j-1}\| + \sum_{j=1}^k \lambda_j \epsilon_j \\ \le\ & \frac{1}{2}\theta\|y - y^0\|^2 + \theta \sum_{j=1}^k \|\delta^{j-1}\|\,\|y^j - y^{j-1}\| + \sum_{j=1}^k \lambda_j \epsilon_j. \end{aligned}$$

This can be rewritten as

$$\sigma_k^{-1} \sum_{j=1}^k \lambda_j f(x^j) \le f(x) + \frac{1}{2}\theta\sigma_k^{-1}\|y - y^0\|^2 + \theta\sigma_k^{-1} \sum_{j=1}^k \|\delta^{j-1}\|\,\|y^j - y^{j-1}\| + \sigma_k^{-1} \sum_{j=1}^k \lambda_j \epsilon_j. \quad (22)$$

Since Lemma 2.2 implies that

$$\liminf_{k\to\infty} f(x^k) \le \liminf_{k\to\infty} \sigma_k^{-1} \sum_{j=1}^k \lambda_j f(x^j)$$

and

$$0 \le \limsup_{k\to\infty} \sigma_k^{-1} \sum_{j=1}^k \lambda_j \epsilon_j \le \limsup_{k\to\infty} \epsilon_k,$$

we have for any $x \in C$ and $y = b - A^T x \in \Re^m_+$

$$\liminf_{k\to\infty} f(x^k)$$

$$
\begin{aligned}
&\le \liminf_{k\to\infty} \sigma_k^{-1} \sum_{j=1}^{k} \lambda_j f(x^j) \\
&\le \liminf_{k\to\infty} \left\{ f(x) + \frac{1}{2}\theta\sigma_k^{-1}\|y-y^0\|^2 + \theta\sigma_k^{-1}\sum_{j=1}^{k}\|\delta^{j-1}\|\,\|y^j-y^{j-1}\| + \sigma_k^{-1}\sum_{j=1}^{k}\lambda_j\epsilon_j \right\} \\
&\le \limsup_{k\to\infty} \left\{ f(x) + \frac{1}{2}\theta\sigma_k^{-1}\|y-y^0\|^2 + \theta\sigma_k^{-1}\sum_{j=1}^{k}\|\delta^{j-1}\|\,\|y^j-y^{j-1}\| + \sigma_k^{-1}\sum_{j=1}^{k}\lambda_j\epsilon_j \right\} \\
&\le f(x) + \lim_{k\to\infty}\frac{1}{2}\theta\sigma_k^{-1}\|y-y^0\|^2 + \limsup_{k\to\infty}\theta\sigma_k^{-1}\sum_{j=1}^{k}\|\delta^{j-1}\|\,\|y^j-y^{j-1}\| + \limsup_{k\to\infty}\epsilon_k.
\end{aligned}
$$

It then follows from $\sigma_k \to \infty$, $\epsilon_k \to 0$ and Lemma 4.5 that

$$\liminf_{k\to\infty} f(x^k) \le f(x) \ \ \forall x \in C.$$

We therefore have

$$\liminf_{k\to\infty} f(x^k) \le f_*. \tag{23}$$

Moreover we obviously have

$$f(\Pi_C(x^k)) \ge f_* \tag{24}$$

for all k. Since (A3) implies $|f(x^k) - f(\Pi_C(x^k))| \to 0$, it follows from (23) and (24) that $\liminf_{k\to\infty} f(\Pi_C(x^k)) = f_*$. We now show that the entire sequence $\{f(\Pi_C(x^k))\}$ converges to f_*.

Since f_* is finite by (A4), Lemma 4.4 (ii) yields

$$
\begin{aligned}
f(\Pi_C(x^k)) - f_* \ &\le \ f(\Pi_C(x^{k-1})) - f_* + \gamma_1\|\delta^{k-1}\|\,\|y^k - y^{k-1}\| \\
&\quad + \gamma_3\|\delta^{k-1}\| + |f(\Pi_C(x^k)) - f(x^k)| + \epsilon_k.
\end{aligned}
$$

By Lemma 4.5, we have $\sum_{k=1}^{\infty}\|\delta^{k-1}\|\,\|y^k - y^{k-1}\| < \infty$. By (A3), we also have $\sum_{k=1}^{\infty}|f(\Pi_C(x^k)) - f(x^k)| < \infty$. Furthermore, we have $\sum_{k=1}^{\infty}\epsilon_k < \infty, \sum_{k=1}^{\infty}\|\delta^{k-1}\| < \infty$. Hence Lemma 2.1 shows that the nonnegative sequence $\{f(\Pi_C(x^k)) - f_*\}$ converges. Since f_* is just a constant, this means that the sequence $\{f(\Pi_C(x^k))\}$ converges. But we already know that $\liminf_{k\to\infty} f(\Pi_C(x^k)) = f_*$, i.e., there is a subsequence of $\{f(\Pi_C(x^k))\}$ converging to f_*. Hence the entire sequence $\{f(\Pi_C(x^k))\}$ converges to f_*. Then (A3) implies that the entire sequence $\{f(x^k)\}$ also converges to f_*. □

We next state our second global convergence result. It deals with the behavior of the sequence $\{x^k\}$ itself.

Theorem 4.2 *Suppose that (A1)–(A4) hold. Let us denote by*

$$X_* := \{x \in C \mid f(x) = f_*\}$$

the solution set of problem (P). Then the following statements hold:

(i) *If $X_* = \emptyset$, then $\lim_{k\to\infty}\|x^k\| = \infty$.*

(ii) *If $X_* \neq \emptyset$, then the entire sequence $\{x^k\}$ converges to a solution of problem (P).*

Proof. (i) Suppose that $X_* = \emptyset$. We show that every accumulation point of the sequence $\{x^k\}$ is a solution of problem (P). Then the assumption $X_* = \emptyset$ immediately implies that the sequence $\{x^k\}$ cannot have a bounded subsequence, so that we have $\lim_{k\to\infty} \|x^k\| = \infty$.

Therefore, let x^* be an accumulation point of $\{x^k\}$ and $\{x^k\}_{k\in K}$ be a corresponding subsequence converging to x^*. Since $\lim_{k\to\infty} f(x^k) = f_*$ by Theorem 4.1 and f is lower semicontinuous at x^*, it then follows that

$$f(x^*) \leq \liminf_{\substack{k\to\infty \\ k\in K}} f(x^k) = \lim_{k\to\infty} f(x^k) = f_*.$$

On the other hand, x^* is feasible for problem (P) because of Lemma 4.2. Hence the inequality $f(x^*) \geq f_*$ holds. Consequently we obtain $f(x^*) = f_*$, i.e., the accumulation point x^* is a solution of problem (P).

(ii) Now assume that the solution set X_* is nonempty, and let x^* be an arbitrary point in X_*. Since $f(\Pi_C(x^k)) \geq f_*$ for all k, we have

$$f(\Pi_C(x^k)) - f(x^k) \geq f_* - f(x^k).$$

Substituting $x^* \in X_*$ and $y^* = b - A^T x^*$ for x and y, respectively, in (19), we have

$$\|y^* - y^k\|^2 \leq \|y^* - y^{k-1}\|^2 + 2\theta^{-1}\lambda_k(f_* - f(x^k) + \epsilon_k) + 2\|\delta^{k-1}\|\,\|y^k - y^{k-1}\|.$$

Together with the previous inequality and $\lambda_k \leq \lambda_{\max}$, we obtain

$$\begin{aligned} & \|y^* - y^k\|^2 \\ \leq\ & \|y^* - y^{k-1}\|^2 + 2\theta^{-1}\lambda_k(f(\Pi_C(x^k)) - f(x^k)) + 2\theta^{-1}\lambda_k\epsilon_k + 2\|\delta^{k-1}\|\,\|y^k - y^{k-1}\| \\ \leq\ & \|y^* - y^{k-1}\|^2 + 2\theta^{-1}\lambda_{\max}|f(\Pi_C(x^k)) - f(x^k)| + 2\theta^{-1}\lambda_k\epsilon_k + 2\|\delta^{k-1}\|\,\|y^k - y^{k-1}\|. \end{aligned}$$

Using (A3), $\sum_{k=1}^{\infty} \lambda_k\epsilon_k < \infty$, Lemma 4.5, and Lemma 2.1, it follows that the sequence $\{\|y^* - y^k\|\}$ converges. Since $\|y^* - y^k\| = \|A^T(x^k - x^*) - \delta^k\|$, we have

$$\|y^* - y^k\| - \|\delta^k\| \leq \|A^T(x^k - x^*)\| \leq \|y^* - y^k\| + \|\delta^k\|.$$

Therefore $\{\|A^T(x^k - x^*)\|\}$ converges. Since the matrix A has maximal row rank by (A2), the sequence $\{\|x^k - x^*\|\}$ also converges. In particular, $\{x^k\}$ is bounded, and hence it contains a subsequence converging to some point x^∞. Since $x^\infty \in C$ by Lemma 4.2 and $f(x^k) \to f_*$ by Theorem 4.1, we must have $x^\infty \in X_*$. It then follows that the whole sequence $\{x^k\}$ converges to x^∞, since $\{\|x^k - x^*\|\}$ is convergent for any $x^* \in X_*$, as shown above. □

We note that none of our global convergence results assumes the existence of a strictly feasible point, in contrast to several related papers like [1, 2, 3] which consider feasible proximal-like methods.

We close this section with a short discussion regarding (A3). To this end, we first observe that this assumption is automatically satisfied if $\delta^k = 0$ for all k, i.e., if the algorithm generates feasible iterates. We further note that (A3) also holds if, for example, the function f is Lipschitzian; this follows from Lemma 4.2 since

$$\sum_{k=1}^{\infty} |f(x^k) - f(\Pi_C(x^k))| \leq L \sum_{k=1}^{\infty} \|x^k - \Pi_C(x^k)\| \leq \alpha L \sum_{k=1}^{\infty} \|\delta^k\| < \infty,$$

where $L > 0$ denotes the Lipschitz constant of f. In particular, (A3) is therefore satisfied for linear programs.

On the other hand, it should also be mentioned that (A3) is restrictive in the sense that it requires all projected points $\Pi_C(x^k)$ (or at least those for sufficiently large iterates k) to belong to the domain of f, since otherwise (A3) is certainly not satisfied. However, if, for example, an accumulation point of the sequence $\{x^k\}$ is in the interior of $\mathrm{dom} f$, then Lemma 4.2 automatically implies that f is finite-valued at the projected points $\Pi_C(x^k)$; moreover, it is known that, in this case, f is locally Lipschitzian around the accumulation point, so that (A3) is satisfied on such a subsequence.

To include a further motivation for the necessity of an assumption like (A3), consider the two-dimensional example shown in Figure 1: The feasible set C is just a line with $f(x) = 0$ for all $x \in C$. We further have $f(x) = \infty$ above this line, while $f(x) < 0$ below that line. In this example, f is a closed proper convex function with $C \cap \mathrm{dom} f \neq \emptyset$ and $C \cap \mathrm{int}\,\mathrm{dom}\, f = \emptyset$. The solution set X_* of this example is nonempty as every feasible point is a solution, and the optimal value is obviously $f_* = 0$. On the other hand, the sequence of function values $\{f(x^k)\}$ may tend to $-\infty$ or, depending on the precise behavior of the function f, at least to a negative number. Note that this happens although the iterates x^k get arbitrarily close to the feasible set C, i.e., to the solution set. Hence the statements of Theorems 4.1 and 4.2 do not hold for this example. The reason is that (A3) is not satisfied.

Figure 1: Counterexample illustrating the necessity of (A3)

5 Conclusion

In this paper, we proposed an infeasible interior proximal algorithm for convex programming problems with linear constraints, which can be started from an arbitrary initial point, and is applicable even if there is no interior of the feasible region. Nice global convergence properties were shown under suitable assumptions.

Among the possible future projects is an extension of the proposed algorithm to the solution of variational inequality problems. Furthermore, the method is still rather conceptual, and one should devise a practical procedure to execute step 3 in order to get an implementable algorithm. Another important subject is the rate of convergence of the algorithm.

References

[1] A. Auslender and M. Haddou, "An interior proximal method for convex linearly constrained problems and its extension to variational inequalities," *Mathematical Programming*, 71, 1995, 77–100.

[2] A. Auslender, M. Teboulle, and S. Ben-Tiba, "A logarithmic-quadratic proximal method for variational inequalities," *Computational Optimization and Applications*, 12, 1999, 31–40.

[3] A. Auslender, M. Teboulle, and S. Ben-Tiba, "Interior proximal and multiplier methods based on second order homogeneous kernels," *Mathematics of Operations Research*, 24, 1999, 645–668.

[4] A. Ben-Tal, and M. Zibulevsky, "Penalty-barrier methods for convex programming problems," *SIAM Journal on Optimization*, 7, 1997, 347–366.

[5] L. Bregman, "The relaxation method of finding the common points of convex sets and its application to the solution of problems in convex programming," *U.S.S.R. Computational Mathematics and Mathematical Physics*, 7, 1967, 200–217.

[6] Y. Censor and S.A. Zenios, "The proximal minimization algorithm with D-functions," *Journal of Optimization Theory and Applications*, 73, 1992, 451–464.

[7] G. Chen, and M. Teboulle, "Convergence analysis of a proximal-like minimization algorithm using Bregman functions," *SIAM Journal on Optimization*, 3, 1993, 538–543.

[8] I. Csiszár, "Information-type measures of difference of probability distributions and indirect observations," *Studia Scientiarum Mathematicarum Hungarica*, 2, 1967, 299–318.

[9] J. Eckstein, "Nonlinear proximal point algorithms using Bregman functions, with applications' to convex programming," *Mathematics of Operations Research*, 18, 1993, 202–226.

[10] O. Güler, "On the convergence of the proximal point algorithm for convex minimization," *SIAM Journal on Control and Optimization*, 29, 1991, 433–419.

[11] J.B. Hiriart-Urruty, "ϵ-subdifferential calculus," in *Convex Analysis and Optimization*, J.-P. Aubin and R.B. Vinter (eds.), Research Notes in Mathematics Series 57, Pitman, London, 1982, 43–92.

[12] A.J. Hoffman, "On approximate solutions of system of linear inequalities," *Journal of the National Bureau of Standards*, 49, 1952, 263–265.

[13] A.N. Iusem, B. Svaiter, and M. Teboulle, "Entropy-like proximal methods in convex programming," *Mathematics of Operations Research*, 19, 1994, 790–814.

[14] A. Iusem and M. Teboulle, "Convergence rate analysis of nonquadratic proximal and augmented Lagrangian methods for convex and linear programming," *Mathematics of Operations Research*, 20, 1995, 657–677.

[15] K.C. Kiwiel, "Proximal minimization methods with generalized Bregman functions," *SIAM Journal on Control and Optimization*, 35, 1997, 1142–1168.

[16] K. Knopp, *Infinite Sequences and Series*, Dover Publications, Inc., New York, 1956.

[17] B. Lemaire, "The proximal algorithm," in *New Methods in Optimization and Their Industrial Uses, Proc. Symp. Pau and Paris 1987,* J. P. Penot (ed.), International Series of Numerical Mathematics 87, Birkhäuser-Verlag, Basel, 1989, 73–87.

[18] B. Lemaire, "About the convergence of the proximal method," in *Advances in Optimization, Proceedings Lambrecht 1991*, Lecture Notes in Economics and Mathematical Systems 382, Springer-Verlag, 1992, 39–51.

[19] B. Lemaire, "On the convergence of some iterative methods for convex minimization," in *Recent Developments in Optimization,* R. Durier and C. Michelot (eds.), Lecture Notes in Economics and Mathematical Systems, 429, Springer-Verlag, 1995, 252–268.

[20] B. Martinet, "Perturbation des méthodes d'opimisation," *R.A.I.R.O, Analyse Numérique*, 12, 1978, 153–171.

[21] J.J. Moreau, "Proximité et dualité dans un espace Hilbertien, " *Bulletin de la Societé de Mathematique de France*, 93, 1965, 273–299.

[22] Y. Nesterov and A. Nemirovski, *Interior Point Polynomial Algorithms in Convex Programming*, SIAM Publications, Philadelphia, PA, 1994.

[23] B.T. Polyak, *Introduction to Optimization*, Optimization Software Inc., New York, 1987.

[24] R.A. Polyak and M. Teboulle, "Nonlinear rescaling and proximal-like methods in convex optimization," *Mathematical Programming*, 76, 1997, 265–284.

[25] R.T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, NJ, 1970.

[26] R.T. Rockafellar, "Augmented Lagrangians and applications of the proximal point algorithm in convex programming," *Mathematics of Operations Research*, 1, 1976, 97–116.

[27] R.T. Rockafellar, "Monotone operators and the proximal point algorithm," *SIAM Journal on Control and Optimization*, 14, 1976, 877–898.

[28] M. Teboulle, "Entropic proximal mappings with application to nonlinear programming," *Mathematics of Operations Research*, 17, 1992, 670–693.

[29] M. Teboulle, "Convergence of proximal-like algorithms," *SIAM Journal on Optimization*, 7, 1997, 1069–1083.

[30] P. Tseng and D. Bertsekas, "On the convergence of the exponential multiplier method for convex programming," *Mathematical Programming*, 60, 1993, 1–19.

[31] T. Terlaky (ed.), *Interior Point Methods of Mathematical Programming*, Kluwer Academic Publishers, Boston, 1996.

[32] S.J. Wright, *Primal-Dual Interior-Point Methods*, SIAM, Philadelphia, 1997.