RAIRO Operations Research
RAIRO Oper. Res. **41** (2007) 367–380
DOI: 10.1051/ro:2007032

RESCALED PROXIMAL METHODS FOR LINEARLY CONSTRAINED CONVEX PROBLEMS *

Paulo J.S. Silva [1] and Carlos Humes Jr. [1]

Abstract. We present an inexact interior point proximal method to solve linearly constrained convex problems. In fact, we derive a primal-dual algorithm to solve the KKT conditions of the optimization problem using a modified version of the rescaled proximal method. We also present a pure primal method. The proposed proximal method has as distinctive feature the possibility of allowing inexact inner steps even for Linear Programming. This is achieved by using an error criterion that bounds the subgradient of the regularized function, instead of using ϵ-subgradients of the original objective function. Quadratic convergence for LP is also proved using a more stringent error criterion.

Keywords. Interior proximal methods, Linearly constrained convex problems.

Mathematics Subject Classification. 90C25, 90C33.

1. Introduction

The idea of proximal methods can be traced back to Martinet [9, 10] and Rockafellar [12]. In the context of optimization, the classical proximal method replaces a minimization problem by a sequence of better behaved problems with a quadratic regularization term added to the objective function.

Many generalizations of this classical proximal algorithm were proposed in the last years. One of the main objectives was to replace the squared Euclidean norm

Received January 4, 2006. Accepted January 22, 2007.

* *The authors were partially supported by CNPq grants 304691/2002-0 and 307105/2003-2 respectively, and PRONEX - Optimization.*

[1] Instituto de Matemática e Estatística, Universidade de São Paulo, Brazil;
pjssilva@ime.usp.br, chumes@usp.br

by coercive regularizations that were able to implicitly deal with simple constraints, giving raise to interior point proximal methods [2–5, 7, 8, 13–16]. This effort followed the success of interior point methods for optimization, particularly Linear Programming (LP). For a classical survey on interior point methods for LP see [6].

Initially, interior point proximal methods were formulated for box constrained problems. More recently these methods were generalized for linearly constrained problems [1, 7, 16, 19]. In particular, exact versions of the proximal algorithm were applied to solve Linear Programming (LP) problems resulting on quadratic convergent methods.

Our intent in this paper is to use the rescaled proximal method from [13, 14] to linearly constrained optimization problems. This will be done in such a way that allows for inexact proximal subproblems even for LP, maintaining quadratic convergence. This objective is attained in four steps. First, we observe that the Karush-Kuhn-Tucker (KKT) conditions of a linearly constrained problem have a special structure that can be exploited by a specialized proximal method. Second, we present the detailed proximal method and show its convergence. Next, we apply the method to linearly constrained problems. Finally, we show a convergence rate result.

From now on, we use the following notation: $\mathbb{R}^n$ denotes the set of all n-dimensional vectors of real numbers, $\langle \cdot, \cdot \rangle$ denotes the usual (Euclidean) inner product, and $\| \cdot \|_2$ is the induced (Euclidean) norm. If $f : \mathbb{R}^n \mapsto \mathbb{R} \cup \{-\infty, +\infty\}$ is a function, we call the set $\operatorname{dom} f \doteq \{x \mid f(x) < \infty\}$ its *effective domain*. We say that f is *proper* if $\operatorname{dom} f \neq \emptyset$ and $f(x) > -\infty,\ \forall x \in \mathbb{R}^n$. If f is convex and $\epsilon \geq 0$, we say that $\gamma \in \mathbb{R}^n$ is an *ϵ-subgradient* of f at x if, for all $y \in \mathbb{R}^n$, $f(y) \geq f(x) + \langle \gamma, y - x \rangle - \epsilon$. The set of all ϵ-subgradients at x is called *ϵ-subdifferential* at x, and it is denoted by $\partial_\epsilon f(x)$. If $\epsilon = 0$, we omit the ϵ and simply say *subgradient, subdiffential* and use the symbol ∂f. Let $\{x^k\}$ be a sequence in $\mathbb{R}^n$, we write $x^k \to \bar{x}$ to state that the sequence converge to the vector $\bar{x}$. If $\mathcal{K} \subseteq \mathbb{N}$ is an infinite set of indexes, then $\{x^k\}_{\mathcal{K}}$ denote the subsequence of $\{x^k\}$ with indexes in $\mathcal{K}$ and $x^k \to_{\mathcal{K}} \bar{x}$ says that the subsequence converges to $\bar{x}$.

2. Preliminaries

Consider the following convex problem constrained to a polyhedron in the standard form:

$$
\begin{array}{ll}
\min & f(x) \\
\text{s.t.} & x \in \mathcal{F} \stackrel{\text{def}}{=} \{x \in \mathbb{R}^n \mid Ax = b, x \geq 0\},
\end{array}
\tag{1}
$$

where f is a proper lower semi-continuous convex function.

The associated KKT conditions are:

$$\begin{aligned} 0 &\in \partial f(x) - A'y + u \\ Ax &= b \\ x &\geq 0 \\ u &\leq 0 \\ \langle u, x \rangle &= 0. \end{aligned} \tag{2}$$

Or, using the variational inequalities notation,

$$0 \in \begin{bmatrix} \partial f(x) - A'y \\ Ax - b \end{bmatrix} + \begin{bmatrix} N_{\mathbb{R}^n_+}(x) \\ 0 \end{bmatrix}. \tag{3}$$

The variational inequality formulation opens up the opportunity to use a proximal method based only on box constraints. In this paper, we are specially interested on applying the Rescaling Proximal Method for Variational Inequalities (RPMVI) described in [14]. However, a naive, direct application of this idea may raise questions as:

(1) It is not clear why it is necessary to add a regularization term for the dual variables, y, since they are unconstrained.
(2) Let $\{x^k\}$ be a proximal sequence computed by the RPMVI method proposed in [14]. Let $\bar{x}$ be a cluster point of such sequence, *i.e.*, there is an infinite set $\mathcal{K} \subseteq \mathbb{N}$ such that $x^k \to_{\mathcal{K}} \bar{x}$. The convergence theorem in [14] depends on the technical assumption that, for each $i = 1, \dots, n$, either $x_i^{k-1} \to_{\mathcal{K}} \bar{x}_i$, or we can easily show that $0 \in T(\bar{x})_i$, where T is the operator that defines the variational inequality [14] Assumption 2.3.2.

    This assumption is usually difficult to verify. In [14], the authors show that it holds only for optimization problems or if they use a special stepsize choice. The KKT Equation (3) can not be restated as a simple optimization problem. Hence, to apply the results from [14], we would be forced to apply the special stepsize selection, which is not completely satisfactory due to its stringent character. We could also use double regularizations methods [13], but this choice would constraint the class of admissible regularizations. It is not clear if such restrictions to the proximal methods are really required.

Although there may be sensible answers to these points, we present below a detour that is able to avoid such issues.

3. The partial rescaled proximal method

Let us consider a variational problem inspired by (3),

$$0 \in T(x, y) + \begin{bmatrix} N_B(x) \\ 0 \end{bmatrix}, \tag{4}$$

where T is a (possibly set-valued) maximal monotone operator on $\mathbb{R}^{n+m}$ and B is a box, that is $B \stackrel{\text{def}}{=} ([a_1, b_1] \times \ldots \times [a_n, b_n]) \cap \mathbb{R}^n$, where $-\infty \leq a_i < b_i \leq +\infty$ for all $i = 1, \ldots, n$.

As discussed above, we will present a variation of the RPMVI [14] that has as distinctive feature the regularization of only the constrained variable x. Since we are interested in coercive proximal methods, we must assume from now on that:

Assumption 3.1.

$$\operatorname{dom} T \cap \operatorname{int} B \times \mathbb{R}^m \neq \emptyset.$$

This condition also guarantees that the sum in (4) is a maximal monotone operator.

Before presenting the modified proximal algorithm, we introduce the class of coercive distances it employes:

Definition 3.1. *For $i = 1, \ldots, n$, a function $d_i : \mathbb{R} \times (a_i, b_i) \to (-\infty, \infty]$ is called a* rescaled distance *if it presents the following properties:*

(3.1.1) For all $z_i \in (a_i, b_i)$, $d_i(\cdot, z_i)$ is closed and strictly convex, with its minimum at z_i. Moreover, $\operatorname{int} \operatorname{dom} d_i(\cdot, z_i) = (a_i, b_i)$.

(3.1.2) d_i is differentiable with respect to the first variable on $(a_i, b_i) \times (a_i, b_i)$, and this partial derivative is continuous at all points of the form $(x_i, x_i) \in (a_i, b_i) \times (a_i, b_i)$. Moreover, we will use the notation

$$d_i'(x_i, z_i) \stackrel{\text{def}}{=} \frac{\partial d_i}{\partial x_i}(x_i, z_i).$$

(3.1.3) For all $z_i \in (a_i, b_i)$, $d_i(\cdot, z_i)$ is essentially smooth [11] Chapter 26.

(3.1.4) There exist $L, \epsilon > 0$ such that if either $-\infty < a_i < z_i \leq x_i < a_i + \epsilon$ or $b_i - \epsilon < x_i \leq z_i < b_i < +\infty$, then $|d_i'(x_i, z_i)| \leq L\,|x_i - z_i|$.

This definition has appeared in [13] and is a slight generalization of [14] Assumption 2.1, as it does not explicit require twice-differentiable distances. The relation between both assumptions becomes clear if we identify the above distances with the ones in [14] Assumption 2.1, divided by the convenient second derivative. In particular, such distances encompasses second order homogeneous kernels and rescaled Bregman distances, as described in [14] Section 2.2.

We can now present the modified proximal algorithm:

Partial Rescaling Proximal Method (PRPM)

(1) **Initialization:** *Let $k = 0$. Choose a scalar $\underline{\alpha} > 0$, $\{\beta_k\}$ a positive real sequence converging to zero, and an initial iterate $(x^0, y^0) \in \operatorname{int} B \times \mathbb{R}^m$. Finally, let D be a separable distance defined by*

$$D(x, x^k) \stackrel{def}{=} \sum_{i=1}^{n} d_i(x_i, x_i^k), \tag{5}$$

where each d_i is a rescaled distance.

(2) **Iteration:**

(a) *Choose α^k such that each $\alpha_i^k \in [\underline{\alpha}, \infty)$, $i = 1, \ldots, n$.*

(b) *Find x^{k+1} and e^{k+1} such that*

$$\begin{bmatrix} e^{k+1} \\ 0 \end{bmatrix} \in T(x^{k+1}, y^{k+1}) + \begin{bmatrix} \operatorname{diag}(\alpha^k)^{-1} \nabla_1 D(x^{k+1}, x^k) \\ 0 \end{bmatrix}, \tag{6}$$

where $\nabla_1 D(\cdot, \cdot)$ denotes the partial derivative of D with respect to the first variable, and

$$\left| e_i^k \right| \leq \frac{1}{\alpha_i^k} \left| d_i'(x^{k+1}, x^k) \right| + \beta_k. \tag{7}$$

(c) *Let $k = k + 1$, and repeat the iteration.*

The algorithm above takes into account that a component of the original variational inequality may be an easily solvable equation. In such cases, it dismisses the need for a regularization in the respective variables. This structure appears in (3), where the lower part of the system is linear. We will further explore this fact in the next section.

However, let us first turn our attention to analyze the algorithm convergence. The first step is to adapt [14] Assumption 2.3.2, to our present context.

Assumption 3.2. *Define*

$$\gamma^k = \begin{bmatrix} \gamma_x^k \\ \gamma_y^k \end{bmatrix} \overset{def}{=} \begin{bmatrix} e^k - \operatorname{diag}(\alpha^{k-1})^{-1} \nabla_1 D(x^k, x^{k-1}) \\ 0 \end{bmatrix}. \tag{8}$$

Let $\bar{x}$ be an accumulation point of the x component in a sequence computed by the PRPM, i.e., *there is an infinite set $\mathcal{K} \subseteq \mathbb{N}$ such that $x^k \to_{\mathcal{K}} \bar{x}$. Denote by $\gamma_{x_i}^k$ the i-th component of γ_x^k. Then, either $\gamma_{x_i}^k \to_{\mathcal{K}} 0$ or there is an infinite set $\mathcal{K}' \subseteq \mathcal{K}$ such that $x_i^{k-1} \to_{\mathcal{K}'} \bar{x}_i$.*

Using the above definitions, we can easily show that the PRPM is a special instance of the RPMVI from [14] that explores its permissive error criterion. Therefore its convergence results still hold.

Theorem 3.2. *Let $\{(x^k, y^k)\}$ be a sequence computed by the PRPM with Assumptions 3.1 and 3.2 holding, then all of its limit points are solutions to the variational inequality problem (4).*

Proof. Let us extend the PRPM to regularize also the dual variables using the classical distance $d_{n+i}(y_i, z_i) \overset{\text{def}}{=} \frac{1}{2}(y_i - z_i)^2$ applied to the dual variables y. Consider also the dual stepsizes $\alpha_{n+i}^k \overset{\text{def}}{=} 1$ for all k and the dual error vectors $e_{n+i}^{k+1} = y_i^{k+1} - y_i^k$. Now it is easy to see that PRPM is a special instance of RPMVI and that the proposed result is only a special case Theorem 2.6 from [14]. □

4. Application to linearly constrained convex programming

We turn back our attention to the original problem (1). We will derive a primal-dual interior proximal method to solve its KKT conditions based on the PRPM. Afterwards, we will show how to derive a pure primal method for a problem with special constraint structure.

4.1. A primal-dual method

In order to apply the results above to solve the linearly constrained problem (1), we define L as its Lagrangian considering only the equality constraints. Identifying $T(x, y)$ with $\big(\partial_x L(x, y), -\partial_y L(x, y)\big)$, the KKT conditions for (1) are clearly a particular case of the variational inequality (4). Additionally, it is a well known result that such operator is maximal monotone, see Corollary 37.5.2 from [11].

Applying the PRPM to find KKT pairs, each iteration needs to find (x^{k+1}, y^{k+1}) such that

$$\begin{aligned} e^{k+1} &\in \partial f(x^{k+1}) - A' y^{k+1} + \operatorname{diag}(\alpha^k)^{-1} \nabla_1 D(x^{k+1}, x^k) \\ Ax^{k+1} &= b, \end{aligned} \tag{9}$$

where,

$$\left| e_i^{k+1} \right| \le \frac{1}{\alpha_i^k} \left| d_i'(x_i^{k+1}, x_i^k) \right|. \tag{10}$$

Note that the error criterion above is equivalent to (7) with β_k identically zero. It is also possible to choose more permissive $\{\beta_k\}$ [14], however it would clutter the proofs without presenting any particular advantage in the present context.

It is easy to recognize the above iteration as a variation of a classical primal-dual interior point method for convex programming under linear constraints[18]. However, the above algorithm presents the following important distinctions:

1. The barrier is replaced by the coercive generalized distance D.
2. The update of the barrier parameter is replaced by the update of the distance center, typical of proximal methods and, possibly, α^k.
3. The error criterion that must be achieved to change centers is now controlled by (10) instead of using a measure of the distance to the central path.

Let us verify the conditions required by Theorem 3.2, the convergence theorem for the PRPM. First, considering (3), it can be easily seen that Assumption 3.1 is implied by a Slater type constraint qualification, that is obligatory in interior point methods:

Assumption 4.1.

$$\exists x \in \operatorname{ri} \operatorname{dom} f, \ \textit{such that } Ax = b, \ x > 0.$$

Since the constraint set is a polyhedron, this assumption also asserts that the KKT conditions are necessary and sufficient for optimality, see Corollary 28.3.1 from [11].

On the other hand, Assumption 3.2 automatically holds for linearly constrained optimization problems. We prove this fact with the following three lemmas.

Lemma 4.1. *For all feasible x and for all iterations of the PRPM,*

$$f(x) \geq f(x^k) + \langle \gamma_x^k, x - x^k \rangle,$$

i.e., γ_x^k can be seen as a subgradient of the objective function among the feasible points.

Proof. Remembering (8), the definition of γ^k, we see that $\gamma_x^k \in \partial\big(f - \langle A'y^k, \cdot\rangle\big)(x^k)$. Therefore

$$\begin{aligned} f(x) &= f(x) + \langle y^k, b - Ax \rangle && [Ax = b] \\ &= f(x) - \langle A'y^k, x \rangle + \langle b, y^k \rangle \\ &\geq f(x^k) - \langle A'y^k, x^k \rangle + \langle \gamma_x^k, x - x^k \rangle + \langle b, y^k \rangle \\ &= f(x^k) + \langle y^k, b - Ax^k \rangle + \langle \gamma_x^k, x - x^k \rangle \\ &= f(x^k) + \langle \gamma_x^k, x - x^k \rangle. && [Ax^k = b] \qquad \square \end{aligned}$$

This allow us to prove the following extensions Lemmas 3.4 and 3.6 from [14].

Lemma 4.2. *In all iterations of the PRPM, $\gamma_{x_i}^k (x_i^{k-1} - x_i^k) \geq 0$.*

Proof. We have

$$\begin{aligned} \gamma_{x_i}^k (x_i^{k-1} - x_i^k) &= \left(e_i^k - \frac{1}{\alpha_i^{k-1}} d_i'(x_i^k, x_i^{k-1}) \right) (x_i^{k-1} - x_i^k) \\ &\geq \underbrace{-\frac{1}{\alpha_i^{k-1}} d_i'(x_i^k, x_i^{k-1})(x_i^{k-1} - x_i^k)}_{\geq 0} - \left|e_i^k\right| \left|x_i^{k-1} - x_i^k\right| \\ &= \frac{1}{\alpha_i^{k-1}} \left(\left|d_i'(x_i^k, x_i^{k-1})\right| - \alpha_i^{k-1} \left|e_i^k\right| \right) \left|x_i^{k-1} - x_i^k\right| \\ &\geq \frac{1}{\alpha_i^{k-1}} \left|0\right| \left|x_i^{k-1} - x_i^k\right| \qquad \text{[Due to (10)]} \\ &= 0. \qquad \square \end{aligned}$$

Lemma 4.3. $\{f(x^k)\}$ *is non-increasing and hence convergent if f is bounded below on $\mathcal{F}$. In this case, we also have*

$$\left|\gamma_{x_i}^k\right| \left|x_i^{k-1} - x_i^k\right| \to 0 \quad \forall i = 1, \ldots, n.$$

Proof. From Lemma 4.1,

$$f(x^{k-1}) \geq f(x^k) + \langle \gamma_x^k, x^{k-1} - x^k \rangle = f(x^k) + \sum_{i=1}^{n} \gamma_{x_i}^k (x_i^{k-1} - x_i^k),$$

whence the result follows from the non-negativity of $\gamma_{x_i}^k (x_i^{k-1} - x_i^k)$. □

Note that, if f is bounded below on $\mathcal{F}$, Assumption 3.2 is direct a consequence of the last lemma. Thus, Theorem 3.2 implies the optimality of all accumulation points of the sequence $\{(x^k, y^k)\}$ defined by (9)-(10). Actually, we can strength this result as:

Theorem 4.4. *Suppose that the problem* (1) *conforms Assumption 4.1 and that f is bounded below on the feasible set $\mathcal{F}$. Let $\{(x^k, y^k)\}$ be a sequence computed by* (9)-(10). *Then, if $\{x^k\}$ has a limit point, $\{f(x^k)\}$ converges to the optimal value and all its limit points will be minimizers of* (1). *A condition that ensures the boundedness of $\{x^k\}$ is the boundedness of the solution set, or any other level set of f restricted to $\mathcal{F}$.*

Proof. Let $\bar{x}$ be a limit point of $\{x^k\}$, *i.e.* $x^k \to_{\mathcal{K}} \bar{x}$, for some infinite set $\mathcal{K} \subseteq \mathbb{N}$. Since the PRPM is a special instance of the RPMVI, we may apply [14] Lemma 2.5, to see that $\{\gamma^k\}$ is bounded. Then, we may assume without loss of generality, that there is $\bar{\gamma}_x \in \mathbb{R}^n$ such that $\gamma_x^k \to_{\mathcal{K}} \bar{\gamma}_x$.

Using Lemma 4.1 we have that

$$\forall x \in \mathcal{F},\ f(x) \geq f(x^k) + \langle \gamma_x^k, x - x^k \rangle.$$

Taking limits, and remembering that f is l.s.c. we have:

$$\begin{aligned} \forall x \in \mathcal{F},\ f(x) &\geq \lim_{k\to\infty,\ k\in\mathcal{K}} f(x^k) + \langle \bar{\gamma}_x, x - \bar{x} \rangle \\ &\geq f(\bar{x}), \end{aligned}$$

where the last inequality comes from the signal structure for $\bar{\gamma}$ given in [14] Lemma 2.4. Therefore, $\bar{x}$ is a minimizer of f in $\mathcal{F}$. Letting $x = \bar{x}$ above, it also follows that, $f(x^k) \to f(\bar{x})$.

Finally, as Lemma 4.3 states that $\{f(x^k)\}$ is non-increasing, $\{x^k\}$ is included in a level set of f plus the indicator function of $\{x \geq 0 \mid Ax = b\}$. Since the boundedness of this set, or any other level set of the sum, is equivalent to the boundedness of the optimal solution set, the result follows. □

We end this section addressing an important issue. The theorems that guarantee that the proximal iterations are well defined are usually based on the fact that all variables are being regularized. This is not the case in the PRPM. Therefore we must prove that the PRPM iterations can be done, at least in the linearly constrained optimization case:

Proposition 4.5. *Suppose that* (1) *conforms to Assumption 4.1. Let* $(\bar{x}, \bar{y})$ *be a KKT pair of*

$$\begin{array}{ll} \min & f(x) + \sum_{i=1}^{n} \frac{1}{\alpha_i^k} d_i(x_i, x_i^k) \\ s.t. & Ax = b. \end{array} \tag{11}$$

Then, $(\bar{x}, \bar{y})$ *solves* (9) *with zero error. Moreover, such KKT pairs always exist if* f *is bounded below on* $\mathcal{F}$*.*

Proof. If we write down the KKT conditions for (11) we immediately recognize (9).

Moreover, since the constraints are all affine and Assumption 4.1, asserts that there is a point in $\operatorname{ri}\operatorname{dom}(f(\cdot)+\sum_{i=1}^{n} \frac{1}{\alpha_i^k} d_i(\cdot, x_i^k))$ that is feasible, KKT is necessary and sufficient for optimality [11].

Now, let l be a lower bound of f on F. Since each d_i is only finite at the positive orthant for a give $\zeta \in \mathbb{R}$ the level set:

$$\left\{ x \in \mathbb{R}^n \mid Ax = b,\ f(x) + \sum_{i=1}^{n} \frac{1}{\alpha_i^k} d_i(x_i, x_i^k) \le \zeta \right\} =$$

$$\left\{ x \ge 0 \mid Ax = b,\ f(x) + \sum_{i=1}^{n} \frac{1}{\alpha_i^k} d_i(x_i, x_i^k) \le \zeta \right\}$$

$$\subset \left\{ x \ge 0 \mid Ax = b,\ \sum_{i=1}^{n} \frac{1}{\alpha_i^k} d_i(x_i, x_i^k) \le \zeta - l \right\}.$$

The last level set is a level set of $\sum_{i=1}^{n} \frac{1}{\alpha_i^k} d_i(x_i, x_i^k)$ which is bounded as this function attains unique minimum at x^k by Definition 3.1.1. Therefore (11) admits solutions and hence KKT pairs. □

4.2. A PURE PRIMAL METHOD

Using the primal-dual method it is also possible derive a pure primal method for problem with a special constraint structure. Let us consider a problem in the form.

$$\begin{array}{ll} \min & f(x) \\ \text{s.t.} & Ax \ge b \\ & x \ge 0. \end{array} \tag{12}$$

Adding slack variables, s, the above problem is equivalent to

$$\begin{array}{ll} \min & f(x) \\ \text{s.t.} & Ax - s = b \\ & x \ge 0 \\ & s \ge 0, \end{array}$$

which is constrained by a standard polyhedron and therefore is in the form described by (1). We may then apply the primal-dual method described in the last section. The respective iteration asks to solve approximately

$$0 \in \begin{bmatrix} \partial f(x^{k+1}) & -A'y^{k+1} & +\operatorname{diag}(\alpha_x^k)^{-1}\nabla D(x^{k+1}, x^k) \\ 0 & +y^{k+1} & +\operatorname{diag}(\alpha_s^k)^{-1}\nabla \widetilde{D}(s^{k+1}, s^k) \\ & b - Ax^{k+1} + s^{k+1} & \end{bmatrix},$$

where D regularizes the original variables and $\widetilde{D}$ the slack variables. It is interesting to note that the last two groups of equations can be forced to hold by defining $s^{k+1} \stackrel{\text{def}}{=} Ax^{k+1} - b$ and $y^{k+1} \stackrel{\text{def}}{=} \operatorname{diag}(\alpha_s^k)^{-1}\nabla \widetilde{D}(s^{k+1}, s^k)$. Substituting this definitions back in the first group, we need to solve approximately

$$0 \in \partial f(x^{k+1}) + \operatorname{diag}(\alpha_s^k)^{-1}A'\nabla \widetilde{D}(Ax^{k+1} - b, Ax^k - b) + \operatorname{diag}(\alpha_x^k)^{-1}\nabla D(x^{k+1}, x^k).$$

This is equivalent to minimize approximately

$$f_k(x) \stackrel{\text{def}}{=} f(x) + \sum_{j=1}^{m} \frac{1}{(\alpha_s^k)_j}\tilde{d}_j\big((Ax - b)_j, (Ax^k - b)_j\big) + \sum_{i=1}^{n} \frac{1}{(\alpha_x^k)_i} d_i(x_i, x_i^k).$$

The acceptance criteria (10) turns out to be

$$\left|(\partial f_k(x^{k+1}))_i\right| \leq \frac{1}{(\alpha_x^k)_i}\left|d_i'(x_i^{k+1}, x_i^k)\right|. \tag{13}$$

The method above can be viewed as a variation of the algorithm proposed in [2] Section 4. This is not the case as its error criterion is based on the subgradient of f and not in ϵ-subgradients. Error criteria based on ϵ-subgradients are particularly useful in situations where it is difficult to compute f or its subgradients. Good examples are multiplier methods or situations where f is non-smooth. On the other hand, if the subgradients of the objective function are readily available, a criterion based on the subdifferential itself may be preferable. This is clearly the case in Linear Programming, as the only ϵ-subgradient of f is actually its gradient. Under such conditions, a criterion as in [2], Section 4, would require exact solutions of the proximal subproblems, while (10) or (13) still allow for approximate steps.

Finally, it should be clear that any Linear Programming problem can be reduced to the form (12) using standard textbook techniques.

5. Rate of convergence for Linear Programming

The convergence rate of proximal methods for LP has been studied by many authors. Rockafellar proved in [12], that a classical proximal algorithm, based on the squared Euclidean norm, converges in finite many steps. Tseng and Bertsekas showed in [17] that the exponential multiplier method with rescaling is super-linearly convergent. This last result was somewhat extended for general φ-divergences by Iusem and Teboulle [7]. In [1,16], Auslender, Haddou, and Teboulle have proved that methods based on a special φ-divergence may be quadratically convergent if the stepsizes converge fast enough to zero. In [2], Section 6, Auslender *et al.* proved quadratic convergence if the regularizations used are based on second order homogeneous kernels with an extra quadratic term. Finally, Silva and Eckstein [13] generalized the quadratic convergence for double regularizations, encompassing kernels based on Bregman distances. All these results required the exact solution of the proximal subproblems.

In this section, we demonstrate that the primal-dual method described by (9)-(10) is also quadratic convergent for Linear Programming even with inexact proximal steps. In order to do this, we will use two extra assumptions. The first one limits the class of generalized distances, imposing an upper bound on their value. The second strengths the acceptance criterion (10), but still allows for inexact subproblems.

We assume from now on that problem (1) have the form

$$\begin{array}{ll} \min & \langle c, x \rangle \\ \text{s.t.} & Ax = b \\ & x \geq 0. \end{array} \tag{14}$$

Assumption 5.1. *For $i = 1, \ldots, n$, let d_i be a rescaled distance. There must be $C > 0$ such that for all $x_i, z_i > 0$,*

$$d_i(x_i, y_i) \leq C(x_i - y_i)^2.$$

Two examples of distances that obey the above condition, with C equal 1 and 2 respectively, are the rescaled Bregman distances [14, Section 2.2.1] based on the kernels $x \log(x)$ and $x - \sqrt{x}$.

The second assumption aims to enforce that at each iteration the objective function decreases "enough". As stated in Proposition 4.5, if (9) is solved exactly, x^{k+1} would minimize

$$f(x) + \sum_{i=1}^{n} \frac{1}{\alpha_i^k} d_i(x_i, x_i^k)$$

among all the feasible solutions. Hence, between iterations the objective function would decrease $\sum_{i=1}^{n} \frac{1}{\alpha_i^k} d_i(x_i^{k+1}, x_i^k)$. Assumption 5.2 requires that at least a fraction of this bound should be attained in any inexact iteration.

Assumption 5.2. *Let $\beta \in [0, 1)$. For a given k, define $\epsilon_k(x) \stackrel{def}{=} \beta \sum_i 1/\alpha_i^k d_i(x_i, x_i^k)$. The pair (x^{k+1}, y^{k+1}) computed by the primal dual method (9)-(10) should also conform to*

$$0 \in c - A'y^{k+1} + \operatorname{diag}(\alpha_k)^{-1} \partial_\epsilon D(x^{k+1}, x^k), \quad \epsilon \leq \epsilon_k(x^{k+1}). \tag{15}$$

Note that in the above assumption the ϵ-subgradient operator is applied to the generalized distance, rather than to the objective function. Thus inexact proximal steps are allowed.

We proceed to prove quadratic convergence rate for Linear Programming. We begin stating an auxiliary lemma.

Lemma 5.1. *Let x^* be any solution of* (14). *Let $\{(x^k, y^k)\}$ be a sequence computed by the primal-dual method* (9)-(10) *with Assumptions 4.1, 5.1, and 5.2 holding. Then,*

$$\langle c, x^{k+1}\rangle + (1-\beta)\sum_i \frac{1}{\alpha_i^k} d_i(x_i^{k+1}, x_i^k) \le \langle c, x^*\rangle + \sum_i \frac{1}{\alpha_i^k} d_i(x_i^*, x_i^k).$$

Proof. Equation (15) states that x^{k+1} is at least an $\epsilon_k(x^{k+1})$-minimum of $h(x) \stackrel{\text{def}}{=} \langle c, x\rangle - \langle A'y^{k+1}, x\rangle + \sum_i \frac{1}{\alpha_i^k} d_i(x, x_i^k)$, therefore:

$$\langle c, x^{k+1}\rangle - \langle A'y^{k+1}, x^{k+1}\rangle + \sum_i \frac{1}{\alpha_i^k} d_i(x^{k+1}, x_i^k) \le$$
$$\langle c, x^*\rangle - \langle A'y^{k+1}, x^*\rangle + \sum_i \frac{1}{\alpha_i^k} d_i(x_i^*, x_i^k) + \epsilon_k(x^{k+1}).$$

As x^* and x^{k+1} are both feasible, $\langle A'y^{k+1}, x^*\rangle = \langle A'y^{k+1}, x^{k+1}\rangle$, and so

$$\begin{aligned}\langle c, x^{k+1}\rangle + \sum_{i=1}^n \frac{1}{\alpha_i^k} d_i(x_i^{k+1}, x_i^k) \le&\\ &\langle c, x^*\rangle + \sum_{i=1}^n \frac{1}{\alpha_i^k} d_i(x_i^*, x_i^k) + \epsilon_k(x^{k+1})\\ &= \langle c, x^*\rangle + \sum_{i=1}^n \frac{1}{\alpha_i^k} d_i(x_i^*, x_i^k) + \beta \sum_{i=1}^n \frac{1}{\alpha_i^k} d_i(x_i^{k+1}, x_i^k). \quad \square\end{aligned}$$

We can now prove the following convergence rate result:

Theorem 5.2. *Let $X^* \neq \emptyset$ be the solution set of (14) and f^* the optimal value. Let $\{(x^k, y^k)\}$ a sequence computed by the primal-dual method* (9)-(10) *with Assumptions 4.1, 5.1, and 5.2 holding. If $\{x^k\}$ has a limit point, then the distance of x^k to the solution set converges at least quadratically to zero and $\langle c, x^k\rangle$ converges at least quadratically to f^*.*

Proof. Theorem 4.4 already proves convergence. We need to focus only on the convergence rate estimates. Let $\mathcal{D}(x, X^*)$ denote the distance between a point $x \in \mathbb{R}^n$ and the solution set X^*.

Lemma 6.1 from [2] shows that there is a $\mu > 0$, depending only on the problem data, such that

$$\mathcal{D}(x^k, X^*) \le \mu\big(\langle c, x^k\rangle - f^*\big),\ \forall k = 1, 2, \dots \tag{16}$$

Let $\bar{x}^k$ be the point in X^* that achieves the minimum in the definition of $\mathcal{D}(x^k, X^*)$. Lemma 5.1 asserts that

$$\begin{aligned}\langle c, x^{k+1}\rangle &\leq \langle c, x^{k+1}\rangle + (1-\beta)\sum_{i=1}^{n} \frac{1}{\alpha_i^k} d_i(x_i^{k+1}, x_i^k) \\ &\leq f^* + \sum_{i=1}^{n} \frac{1}{\alpha_i^k} d_i(\bar{x}_i^k, x_i^k) \\ &\leq f^* + \sum_{i=1}^{n} \frac{1}{\underset{\sim}{\alpha}} d_i(\bar{x}_i^k, x_i^k),\end{aligned}$$

where $\underset{\sim}{\alpha}$ is the lower bound to α_i^k presented in PRPM. Using Assumption 5.1 and (16), it follows that:

$$\begin{aligned}\langle c, x^{k+1}\rangle - f^* &\leq \frac{C}{\underset{\sim}{\alpha}} \sum_i (x_i^k - \bar{x}_i^k)^2 \\ &\leq \frac{\mu^2 C}{\underset{\sim}{\alpha}} (\langle c, x^k\rangle - f^*)^2. \end{aligned} \tag{17}$$

Therefore $\langle c, x^k\rangle$ converges to f^* at least quadratically. Moreover, as for any $x \in X^*$ we have

$$\langle c, x^k\rangle - f^* = \langle c, x^k - x\rangle \leq \|c\|_2 \|x^k - x\|_2.$$

We may then combine (16) and (17) to conclude that:

$$\mathcal{D}(x^{k+1}, X^*) \leq \frac{\mu^3 C}{\underset{\sim}{\alpha}} \|c\|_2^2 \mathcal{D}(x^k, X^*)^2. \qquad \square$$

This new result on the quadratic convergence of an inexact proximal method induces future computational experiments. Particularly, the search for efficient practical procedures to ensure Assumption 5.2 should be investigated.

References

[1] A. Auslender and M. Haddou, An inteirior proximal method for convex linearly constrained problems. *Math. Program.* **71** (1995) 77–100.

[2] A. Auslender, M. Teboulle and S. Ben-Tiba, Interior proximal and multiplier methods based on second order homogeneous kernels. *Math. Oper. Res.* **24** (1999) 645–668.

[3] A. Auslender, M. Teboulle and S. Ben-Tiba, A logarithmic-quadratic proximal method for variational inequalities. *Comput. Optim. Appl.* **12** (1999) 31–40.

[4] Y. Censor and J. Zenios, The proximal minimization algorithms with D-functions. *J. Optim. Theor. App.* **73** (1992) 451–464.

[5] J. Eckstein, Nonlinear proximal point algorithms using Bregman functions, with applications to convex programming. *Math. Oper. Res.* **18** (1993) 202–226.

[6] C.C. Gonzaga, Path-following methods for linear programming. *SIAM Rev.* **34** (1992) 167–224.
[7] A.N. Iusem and M. Teboulle, Convergence rate analysis of nonquadratic proximal methods for convex and linear programming. *Math. Oper. Res.* **20** (1995) 657–677.
[8] A.N. Iusem, M. Teboulle and B. Svaiter, Entropy-like proximal methods in covex programming. *Math. Oper. Res.* **19** (1994) 790–814.
[9] B. Martinet, Regularisation d'inequations variationelles par approximations successives. *Rev. Fr. Inf. Rech. Oper.* (1970) 154–159.
[10] B. Martinet, Determination approch d'un point fixe d'une application pseudo-contractante. *C.R. Acad. Sci. Paris* **274A** (1972) 163–165.
[11] R.T. Rockafellar, *Convex Analysis.* Princeton University Press (1970).
[12] R.T. Rockafellar, Monotone operators and the proximal point algorithm. *SIAM J. Control Optim.* **14** (1976) 887–898.
[13] P.J.S. Silva and J. Eckstein, Double regularizations proximal methods, with complementarity applications. *Comput. Optim. Appl.* **33** (2006) 115–116.
[14] P.J.S. Silva, J. Eckstein and C. Humes Jr, Rescaling and stepsize selection in proximal methods using generalized distances. *SIAM J. Optim.* **12** (2001) 238–261.
[15] M. Teboulle, Entropic proximal methods with aplications to nonlinear programming. *Math. Oper. Res.* **17** (1992) 670–690.
[16] M. Teboulle, Convergence of proximal-like algorithms. *SIAM J. Optim.* **7** (1997) 1069–1083.
[17] P. Tseng and D. Bertesekas, On the convergence of the exponential multiplier method for convex programming. *Math. Program.* **60** (1993) 1–19.
[18] Stephen J. Wright, *Primal-Dual Interior-Point Methods.* SIAM (1997).
[19] N. Yamashita, C. Kanzow, T. Morimoto and M. Fukushima, An infeasible interior proximal method for convex programming problems with linear constraints. *Journal Nonlinear Convex Analysis* **2** (2001) 139–156.

To access this journal online:
www.edpsciences.org