On the convergence rate of a proximal multiplier algorithm applied to solve a convex program with separable structure

Orlando Sarmiento Chumbes
Universidade Federal do Rio de Janeiro - UFRJ
Programa de Engenharia de Sistemas e Computação, Caixa Postal 68511
CEP 21941-972, Rio de Janeiro - RJ, Brasil
osarmiento@cos.ufrj.br

Erik Alex Papa Quiroz
Universidad Nacional Mayor de San Marcos - UNMSM
Departamento de Ciencias Matemáticas, Lima, Perú
Av. Germán Amézaga s/n. Edificio Jorge Basadre, Ciudad Universitaria, Lima - Perú
erikpapa@gmail.com

Paulo Roberto Oliveira
Universidade Federal do Rio de Janeiro - UFRJ
Programa de Engenharia de Sistemas e Computação, Caixa Postal 68511
CEP 21941-972, Rio de Janeiro - RJ, Brasil
poliveir@cos.ufrj.br

ABSTRACT

This paper proposes an analysis on the convergence rate of a proximal method called Proximal Multiplier Algorithm with Proximal Distances (PMAPD), proposed by the authors, applied to solve convex problems with separable structure. This method unified the works of Chen and Teboulle (PCPM method), Kyono and Fukushima (NPCPMM method) and Auslender and Teboulle ((EPDM) method, applied to convex problems with linear coupling constraints) and extended the convergence properties for the class of $\varphi-$divergence distances. In this work, we present the global convergence result of the (PMAPD) algorithm and we prove that, under mild assumptions, its iterations generated converge linearly to the unique optimal solution of the problem.

KEYWORDS. Proximal multiplier method; Separable structure; Proximal distances.

Main Area: PM Mathematical Programming.

RESUMO

Este artigo propõe uma análise sobre a taxa de convergência de um método proximal chamado Algoritmo Multiplicador Proximal com Distâncias Proximais (AMPDP), proposto pelos autores, aplicado para resolver problemas convexos com estrutura separável. Este método unificou os trabalhos de Chen e Teboulle (método PCPM), Kyono e Fukushima (Método NPCPMM), Auslender e Teboulle (Método (EPDM) aplicado a problemas convexos com restrições de acoplamento lineares) e estendeu as propriedades de convergência para a classe de distâncias $\varphi-$divergências. Neste trabalho, apresentamos o resultado de convergência global do algoritmo (AMPDP) e provamos que, sob hipóteses adequadas, suas iterações geradas convergem linearmente para a única solução ótima do problema.

PALAVRAS CHAVE. Método multiplicador proximal; Estrutura separável; Distâncias Proximais.

Área principal: PM Programação Matemática.

1. Introduction

In this paper, we are interesting in solving the following separable convex optimization problem:

$$(CP) \quad \min\{f(x) + g(z) \,:\, Ax + Bz = b,\, x \in \bar{\mathtt{C}},\, z \in \bar{\mathtt{K}}\},$$

where $\mathtt{C} \subset I\!R^n$ and $\mathtt{K} \subset I\!R^p$ are nonempty open convex sets, $\bar{\mathtt{C}}$ and $\bar{\mathtt{K}}$ denote the closure (in the euclidean topology) of $\mathtt{C}$ and $\mathtt{K}$ respectively, $f : I\!R^n \to (-\infty, +\infty]$ and $g : I\!R^p \to (-\infty, +\infty]$ are closed proper convex functions and $A \in I\!R^{m\times n}$, $B \in I\!R^{m\times p}$, $b \in I\!R^m$.
The Lagrangian $L(x, z, y)$ for (CP) is defined by $L(x, z, y) : I\!R^n \times I\!R^p \times I\!R^m \to (-\infty, +\infty]$,

$$L(x, z, y) = (f + \delta_{\bar{\mathtt{C}}})(x) + (g + \delta_{\bar{\mathtt{K}}})(z) + \langle y, Ax + Bz - b\rangle.$$

where $\delta_{\mathtt{X}}$ denotes the indicator function of a subset X, $\langle \cdot, \cdot\rangle$ denotes the canonical inner product, $(f + \delta_{\bar{\mathtt{C}}})^*$, $(g + \delta_{\bar{\mathtt{K}}})^*$ are the conjugate functions of $f + \delta_{\bar{\mathtt{C}}}$ and $g + \delta_{\bar{\mathtt{K}}}$, respectively, and y is the Lagrangian multiplier associated with the constraint $Ax + Bz = b$.
In recent decades, a great interest has emerged in studying the separable structure of the problem (CP). This model has been found in various optimization problems. For example, in Telecommunications, see Mahey et al. (1997); in Management Electricity, see Lenoir (2008); and in computer science (to solve matrix completion problems), see example 2 of Goldfarb et al. (2013).
In the paper Sarmiento et al. (2015), the authors we consider the problem (CP) and proposed an algorithm that will be presented in Section 3. This method, called Proximal Multiplier Algorithm with Proximal Distances (PMAPD), is an extension of the (PCPM) and (NPCPMM) methods (see Chen and Teboulle (1994), Kyono and Fukushima (2000), respectively), and includes the class of phi-divergence distances (see Subsection 3.3 of Auslender and Teboulle (2006)), which to our knowledge has not yet been studied in this context. Moreover, the (EPDM) method (applied to convex problems with linear coupling constraints) is a particular case of our method when in our algorithm we consider exact iterations and we use the regularized log-quadratic distance (see Section 2 of Auslender and Teboulle (2001)).
The main result of this paper is to show that the sequence generated by (PMAPD) algorithm converges linearly to the unique optimal solution of the problem (CP). The outline of this paper is as follows: In Section 2, we will give some results in convex analysis and we will present the class of proximal distances that we will use along the paper. In Section 3, we will present the Proximal Multiplier Algorithm with Proximal Distances (PMAPD), and we will show its property of global convergence. In Section 4, we will analyze the convergence rate of the (PMAPD) algorithm. Finally, we state some final conclusions in Section 5.

2. Some results in convex analysis and proximal distance

Throughout the paper $I\!R^n$ is the Euclidean space endowed with the canonical inner product $\langle\cdot,\cdot\rangle$ and the norm of x given by $||x|| := \langle x, x\rangle^{1/2}$. For a matrix $M \in I\!R^{m\times n}$ we define $\|M\| := \max_{||x||\le 1} \|Mx\|$. Given an extended real valued function $f : I\!R^n \to I\!R \cup \{\pm\infty\}$ we denote its domain by $dom f := \{x \in I\!R^n : f(x) < +\infty\}$ and its epigraph $epi\, f := \{(x, \beta) \in I\!R^n \times I\!R : f(x) \le \beta\}$. f is said to be proper, if $dom f \neq \emptyset$ and for all $x \in dom f$, we have $f(x) > -\infty$. Also denote by $ri(X)$ the relative interior set of $X \subset I\!R^n$ and $\partial_\epsilon f$ is the $\epsilon-$subdifferential of f defined by $\partial_\epsilon f(u) = \{p \in I\!R^n : f(v) \ge f(u) + \langle p, v - u\rangle - \epsilon, \forall\, v \in dom f\}$.
Finally, f is a lower semicontinuous function if for each $x \in I\!R^n$ we have that all $\{x^l\}$ such that $\lim_{l\to+\infty} x^l = x$ implies that $f(x) \le \liminf_{l\to+\infty} f(x^l)$. It is easy to prove that the lower semicontinuity of f is equivalent to the closedness of the lower level set $L_f(\alpha) = \{x \in I\!R^n : f(x) \le \alpha\}$, for each $\alpha \in I\!R$. Recall that if f is a proper convex function, then f is closed if and only if f is lower semi-continuous.

2.1. Proximal Distances

In this subsection, we present a variant of the definition of the proximal distance and induced proximal distance, introduced by Auslender and Teboulle (2006).

Definition 1 *A function* $d : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}_+ \cup \{+\infty\}$ *is called proximal distance with respect to an open nonempty convex set* $C \subset \mathbb{R}^n$ *if for each* $y \in C$ *it satisfies the following properties:*

(i) $d(\cdot, y)$ *is proper, closed, convex on* $\mathbb{R}^n$ *and continuously differentiable on* C*;*

(ii) dom $d(\cdot, y) \subset \bar{C}$ *and dom* $\partial_1 d(\cdot, y) = C$*, where* $\partial_1 d(\cdot, y)$ *denotes the classical subgradient map of the function* $d(\cdot, y)$ *with respect to the first variable;*

(iii) $d(\cdot, y)$ *is coercive on* $\mathbb{R}^n$ *(i.e.,* $\lim_{||u||\to\infty} d(u, y) = +\infty$*).*

(iv) $d(y, y) = 0$.

We denote by $\mathcal{D}(C)$ *the family of functions satisfying this definition.*

Definition 2 *Given* $d \in \mathcal{D}(C)$*, a function* $H : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}_+ \cup \{+\infty\}$ *is called the induced proximal distance to* d *if there exists* $\gamma \in (0, 1]$ *with* H *a finite valued on* $C \times C$ *and such that for each* $a, b \in C$*, we have*

(Ii) $H(a, a) = 0$.

(Iii) $\langle c - b, \nabla_1 d(b, a)\rangle \le H(c, a) - H(c, b) - \gamma H(b, a), \quad \forall\, c \in C.$

We write $(d, H) \in \mathcal{F}(C)$ to the proximal and induced proximal distance that satisfies the premises of Definition 2.
We denote $(d, H) \in \mathcal{F}(\bar{C})$ if there exists H such that:

(Iiii) H is finite valued on $\bar{C} \times C$ satisfying (Ii) and (Iii), for each $c \in \bar{C}$.

(Iiv) For each $c \in \bar{C}$, $H(c, \cdot)$ has level bounded sets on C.

Finally, we write $(d, H) \in \mathcal{F}_+(\bar{C})$ if

(Iv) $(d, H) \in \mathcal{F}(\bar{C})$.

(Ivi) $\forall\, y \in \bar{C}$ and $\forall\, \{y^k\} \subset C$ bounded with $\lim_{k\to+\infty} H(y, y^k) = 0$, we have $\lim_{k\to+\infty} y^k = y$.

(Ivii) $\forall\, y \in \bar{C}$ and $\forall\, \{y^k\} \subset C$ such that $\lim_{k\to+\infty} y^k = y$, we obtain $\lim_{k\to+\infty} H(y, y^k) = 0$.

Several examples of proximal distances which satisfy the above definitions, for example Bregman distances, proximal distances based on $\varphi-$divergences, self-proximal distances, and distances based on second order homogeneous proximal distances, were given by Auslender and Teboulle (2006).
The following additional conditions on H will be useful to prove the convergence of (PMAPD) algorithm.
Given $(d, H) \in \mathcal{F}_+(\bar{C})$, H satisfies the following condition:

(Iviii) $\forall\, c \in \bar{C}$ and $\forall\, \{y^k\} \subset C$ such that $\lim_{k\to+\infty} y^k = y$, we have $\lim_{k\to+\infty} H(c, y^k) = H(c, y)$.

Some examples of proximal distances which satisfy this condition, were showed by Sarmiento, Quiroz and Oliveira (2015).

3. The (PMAPD) algorithm

In the proposed algorithm we used the class of proximal distances $(d_0, H_0) \in \mathcal{F}_+(\bar{\mathtt{C}})$, $(d_0', H_0') \in \mathcal{F}_+(\bar{\mathtt{K}})$, satisfying the condition (Iviii) and given $\mu > 0, \mu' > 0$ we defined the following functions:

$$\begin{aligned} d(x,y) &= d_0(x,y) + (\mu/2)\left\|x-y\right\|^2, \quad H(x,y) = H_0(x,y) + (\mu/2)\left\|x-y\right\|^2, &(1)\\ d'(x,y) &= d_0'(x,y) + (\mu'/2)\left\|x-y\right\|^2, \quad H'(x,y) = H_0'(x,y) + (\mu'/2)\left\|x-y\right\|^2. &(2)\end{aligned}$$

It is easy to check that $(d, H) \in \mathcal{F}_+(\bar{\mathtt{C}})$ and $(d', H') \in \mathcal{F}_+(\bar{\mathtt{K}})$ (for the same value of γ and γ' respectively) and both satisfy the condition (Iviii).
The algorithm, which will be called Proximal Multiplier Algorithm with Proximal Distances (PMAPD) is as follows:

(PMAPD) Algorithm

Step 0. Choose two pairs $(d_0, H_0) \in \mathcal{F}_+(\bar{\mathtt{C}})$, $(d_0', H_0') \in \mathcal{F}_+(\bar{\mathtt{K}})$ satisfying the condition (Iviii) and define (d, H), (d', H') given by (1) and (2) respectively. Take three sequences $a_k \geq 0$, $b_k \geq 0$ and $\lambda_k > 0$ and choose an arbitrary starting point $(x^0, z^0, y^0) \in \mathtt{C} \times \mathtt{K} \times I\!R^m$.

Step 1. For $k = 0, 1, 2, \ldots$, calculate $p^{k+1} \in I\!R^m$ by

$$p^{k+1} = y^k + \lambda_k(Ax^k + Bz^k - b). \tag{3}$$

Step 2. Find $(x^{k+1}, \upsilon^{k+1}) \in \mathtt{C} \times I\!R^n$ and $(z^{k+1}, \xi^{k+1}) \in \mathtt{K} \times I\!R^p$ such that

$$\upsilon^{k+1} \in \partial_{a_k} f^k(x^{k+1}), \qquad \upsilon^{k+1} + \lambda_k^{-1}\nabla_1 d(x^{k+1}, x^k) = 0, \tag{4}$$
$$\xi^{k+1} \in \partial_{b_k} g^k(z^{k+1}), \qquad \xi^{k+1} + \lambda_k^{-1}\nabla_1 d'(z^{k+1}, z^k) = 0. \tag{5}$$

where the functions $f^k : I\!R^n \to (-\infty, +\infty]$ and $g^k : I\!R^p \to (-\infty, +\infty]$ are defined by $f^k(x) = f(x) + \langle p^{k+1}, Ax\rangle$ and $g^k(z) = g(z) + \langle p^{k+1}, Bz\rangle$, respectively.

Step 3. Compute

$$y^{k+1} = y^k + \lambda_k(Ax^{k+1} + Bz^{k+1} - b). \tag{6}$$

Stopping criterion: If $x^{k+1} = x^k$, $z^{k+1} = z^k$ and $y^{k+1} = y^k$ then stop. Otherwise to do $k := k+1$, and go to Step 1.
In the paper Sarmiento et al. (2015), the following results were obtained

Theorem 3.1 *Let $d_0 \in \mathcal{D}(C)$ and $d_0' \in \mathcal{D}(K)$ be proximal distances that satisfy the premises of Definition 1. Suppose that the problem (CP) has an optimal solution (x^*, z^*) and a corresponding Lagrange multiplier y^* and there exist $x \in ri(dom\, d(\cdot, \upsilon)) \cap ri(dom\, f)$ and $z \in ri(dom\, d'(\cdot, \upsilon')) \cap ri(dom\, g)$ such that $Ax + Bz = b$. Then, for any $(x^k, z^k, y^k) \in C \times K \times I\!R^m$, $\lambda_k > 0$, there exists a unique point $(x^{k+1}, z^{k+1}) \in C \times K$ satisfying (4) and (5).*

Theorem 3.2 *Let $(d_0, H_0) \in \mathcal{F}_+(\bar{C})$, $(d_0', H_0') \in \mathcal{F}_+(\bar{K})$ be a proximal and induced proximal distance satisfying the condition (Iviii). Suppose that the assumptions of Theorem 3.1 are satisfied and $\{a_k\}$, $\{b_k\}$ are sequences nonnegative such that $\sum_{k=0}^{\infty}(a_k + b_k) < \infty$. Let $\{(x^k, z^k, y^k)\}$ be a sequence generated by (PMAPD) algorithm. If $\{\lambda_k\}$ satisfies*

$$\eta < \lambda_k < \bar{c} - \eta$$

for some $\eta \in (0, \bar{c}/2)$ with $\bar{c} := \min\{\frac{\sqrt{\gamma\mu}}{2\|A\|}, \frac{\sqrt{\gamma'\mu'}}{2\|B\|}\}$ where γ, γ' are constant defined in Definition 2, (Iii) and μ, μ' are positive constant defined in (1) and (2) respectively, then the sequence $\{(x^k, z^k, y^k)\}$ globally converges to (x^*, z^*, y^*), with (x^*, z^*) optimal for (CP) and y^* be a corresponding Lagrange multiplier.

4. Convergence rate of (PMAPD) algorithm

In this section, we will analyze the global convergence rate of (PMAPD) algorithm. Goldfarb et al. (2013) presented alternating linearization algorithms based on an alternating direction augmented Lagrangian approach for minimizing the sum of two convex functions. They showed that their basic method require at most $O(1/\epsilon)$ iterations to obtain an ϵ-optimal solution, while their accelerated version require at most $O(1/\sqrt{\epsilon})$ iterations, with little change in the computational effort required at each iteration. Independently to the works of Goldfarb et al. (2013), Chen and Teboulle (1994) developed a proximal decomposition method for convex minimization problems and they showed that the iterations of their algorithm converge linearly (where they use Euclidean distance). It is thus natural to ask if our (PMAPD) method, where we use proximal distances, can get the linear convergence as the works of Chen and Teboulle (1994) and Goldfarb et al. (2013). Below, we answer the question positively but to prove this result we need some additional assumptions on the problem's data. Before that, remember the next results.

Remark 1 *Let T be a set valued maximal monotone operator on $I\!R^n$. Following Rockafellar (1976), we say that the mapping T^{-1} is Lipschitz continuous at the origin with modulus $a \geq 0$, if there exists a unique solution $\bar{u}$ such that $0 \in T(\bar{u})$ and for some $\tau > 0$, we have $\|u - \bar{u}\| \leq a\,\|v\|$, whenever $v \in T(u)$ and $\|u\| \leq \tau$.*
Remember that, the Lagrangian $L(x, z, y)$ is a closed convex-concave function. Therefore, the set-valued subdifferential mapping S on $I\!R^n \times I\!R^p \times I\!R^m$ given by

$$S(x, z, y) = \partial_{x,z} L(x, z, y) \times \partial_y(-L(x, z, y)) \tag{7}$$

is maximal monotone, see Rockafellar (1976).

Assumptions. Throughout the section we assume the following assumptions:

($\mathcal{T}_1$) S^{-1} is Lipschitz continuous at the origin with modulus $a \geq 0$.

($\mathcal{T}_2$) The sequence $\{(x^k, z^k, y^k)\}$ is generated by (PMAPD) algorithm under the approximate criterion

$$\|x^{k+1} - \bar{x}^{k+1}\| \leq \eta_k \|x^{k+1} - x^k\|, \quad \|z^{k+1} - \bar{z}^{k+1}\| \leq \eta_k \|z^{k+1} - z^k\| \tag{8}$$

where $\bar{x}^{k+1}, \bar{z}^{k+1}$ denote the points obtained in Step 2 of (PMAPD) algorithm, when $a_k = b_k = 0\ \forall k$, and $\eta_k \geq 0$ with $\sum_{k=0}^{+\infty} \eta_k < +\infty$.

The Assumptions ($\mathcal{T}_1$) and ($\mathcal{T}_2$) were suggested by Rockafellar (see Rockafellar (1976), p. 100), to derive the rate of convergence of the proximal method of multipliers. It was used in the work of Chen and Teboulle (1994) and will also be used here to derive the rate of convergence for (PMAPD) algorithm.
Also we assume assumptions to the proximal distances $(d_0, H_0) \in \mathcal{F}_+(\bar{\mathrm{C}})$, $(d_0', H_0') \in \mathcal{F}_+(\bar{\mathrm{K}})$.

Remark 2 *We consider the proximal distances d, d' with $\mu = \mu' = 1$, i.e.,*

$$d(x, y) := d_0(x, y) + (1/2)\|x - y\|^2, \quad d'(x, y) := d_0'(x, y) + (1/2)\|x - y\|^2.$$

($\mathcal{T}_3$) Let (x^*, z^*) be an optimal solution for (CP). Then for some $k_0 \in I\!N$, the points $\bar{x}^{k+1}$, $\bar{z}^{k+1}$ satisfy

$$d_0(\bar{x}^{k+1}, x^k) \geq d_0(x^*, x^k), \quad d_0'(\bar{z}^{k+1}, z^k) \geq d_0'(z^*, z^k), \ \ \forall k \geq k_0.$$

$(\mathcal{T}_4)$ The functions $\nabla_1 d_0(\cdot,u), \nabla_1 d_0'(\cdot,v)$ are locally Lipschitz continuous with modulus α_1, α_2 on $\mathtt{C}$ and $\mathtt{K}$ respectively, that is, for any $x_0 \in \mathtt{C}$, $z_0 \in \mathtt{K}$ there exist $\alpha_1 > 0, \alpha_2 > 0$ and $r_1 > 0, r_2 > 0$ such that

$$\|\nabla_1 d_0(x,u) - \nabla_1 d_0(\bar{x},u)\| \leq \alpha_1 \|x-\bar{x}\|, \ \forall x,\bar{x} \in B_{r_1}(x_0), \ \forall u \in \bar{\mathtt{C}},$$
$$\|\nabla_1 d_0'(z,v) - \nabla_1 d_0'(\bar{z},v)\| \leq \alpha_2 \|z-\bar{z}\|, \ \forall z,\bar{z} \in B'_{r_2}(z_0), \ \forall v \in \bar{\mathtt{K}},$$

where $B_{r_1}(x_0) := \{x \in \mathtt{C} \ : \ \|x-x_0\| < r_1\}$ and $B'_{r_2}(z_0) := \{z \in \mathtt{K} \ : \ \|z-z_0\| < r_2\}$.

Remark 3 *With regard to the assumptions we make the following comments:*

1. *Note that for problem (CP) we have*

$$S^{-1}(v_1,v_2,v_3) = \arg\min_{x,z}\max_{y}\{L(x,z,y) - \langle x,v_1\rangle - \langle z,v_2\rangle + \langle y,v_3\rangle\}$$

and therefore Assumption $(\mathcal{T}_1)$, considering Remark 1, can be interpreted in terms of the problem's data as: there exists a unique saddle point w^* such that for some $\tau > 0$, we have $\|w-w^*\| \leq a\,\|v\|$, whenever $\|v\| \leq \tau$ and $w=(x,z,y) \in S^{-1}(v_1,v_2,v_3)$.

2. *Note that the use of the same η_k for the approximation criterion (8) is just to simplify notation in the analysis below. In fact, if one chooses different sequences $\eta_k^i \geq 0$, $\sum_{k=0}^{+\infty}\eta_k^i < +\infty$, $i=1,2$, then one should simply define $\eta_k = \max\{\eta_k^1,\eta_k^2\}$ in (8).*
Observe also, that if one has different k_0^i, $i=1,2$ in Assumption $(\mathcal{T}_3)$, then one should simply define $k_0 = \max\{k_0^1,k_0^2\}$.

3. *It is clear that φ-divergence proximal distances and second order homogeneous proximal distances satisfy the Assumption $(\mathcal{T}_4)$ using their definitions, i.e.,*

$$d_\varphi(x,y) := \sum_{i=1}^{n} y_i\varphi(\frac{x_i}{y_i}) \ \ \textit{with} \ \ \varphi \in C^2(I\!R_{++}),$$
$$d_\varphi(x,y) := \sum_{i=1}^{n} y_i^2\varphi(\frac{x_i}{y_i})$$

with $\varphi(t) = \mu p(t) + \frac{\nu}{2}(t-1)^2$, $\nu \geq \mu p''(1) > 0, p \in C^2(I\!R_{++})$, respectively. We note that in both cases, $\nabla_1 d_\varphi \in C^1(I\!R^n_{++})$. Therefore, $\nabla_1 d_\varphi(\cdot,y)$ is locally Lipschitz continuous.

Remark 4 *Given $\bar{x}^{k+1},\bar{z}^{k+1}$ defined in $(\mathcal{T}_2)$. We define $\bar{y}^{k+1} = y^k + \lambda_k(A\bar{x}^{k+1} + B\bar{z}^{k+1} - b)$.*

The subsequent convergence rate analysis follows a line of argument similar to that given by Chen and Teboulle (1994). Before proving our convergence rate result, we need some previous results.

Lemma 4.1 (*Chen and Teboulle (1994), Lemma* 3.1) *Let $F : I\!R^m \to (-\infty,+\infty]$ be a closed proper convex function, $\tau > 0$ and define: $u^{k+1} = \arg\min_{u\in I\!R^m}\{F(u) + (1/(2\tau))\|u-u^k\|^2\}$. Then for any integer $k \geq 0$,*

$$2\tau[F(u^{k+1}) - F(u)] \leq \|u^k-u\|^2 - \|u^{k+1}-u\|^2 - \|u^{k+1}-u^k\|^2, \ \forall\, u \in I\!R^m.$$

Lemma 4.2 *Let $F : I\!R^n \to I\!R \cup (-\infty,+\infty]$ be a closed proper convex function and $d_0 \in \mathcal{D}(\mathtt{C})$, define*

$$\bar{v}^{k+1} := \arg\min\{F(v) + (1/\lambda_k)d(v,v^k)\},$$

where $d(x,y) := d_0(x,y) + (1/2)\|x-y\|^2$. Then, for any integer $k \geq 0$,

$$2\lambda_k[F(\bar{v}^{k+1}) - F(v)] \leq \|v^k-v\|^2 - \|\bar{v}^{k+1}-v\|^2 - \|\bar{v}^{k+1}-v^k\|^2 - 2d_0(\bar{v}^{k+1},v^k) + 2d_0(v,v^k).$$

Proof: Let $\psi_k(v) := F(v) + (1/\lambda_k)d_0(v, v^k) + (1/2\lambda_k)\|v - v^k\|^2$. By definition of $\bar{v}^{k+1}$ we have $0 \in \partial\psi_k(\bar{v}^{k+1})$, since $F(\cdot)$ and $d_0(\cdot, v^k)$ are convex functions, then ψ_k is strongly convex with modulus $(1/\lambda_k)$ (see, Rockafellar (1976), Proposition 6), it follows that

$$2\lambda_k[\psi_k(v) - \psi_k(\bar{v}^{k+1})] \geq \|\bar{v}^{k+1} - v\|^2, \quad \forall\, v, \tag{9}$$

so, from definition of ψ_k, we have

$$2\lambda_k \left[F(\bar{v}^{k+1}) + \frac{1}{\lambda_k}d_0(\bar{v}^{k+1}, v^k) - F(v) - \frac{1}{\lambda_k}d_0(v, v^k)\right] \leq \|v^k - v\|^2 - \|\bar{v}^{k+1} - v\|^2 - \|\bar{v}^{k+1} - v^k\|^2.$$

Therefore

$$2\lambda_k[F(\bar{v}^{k+1}) - F(v)] \leq \|v^k - v\|^2 - \|\bar{v}^{k+1} - v\|^2 - \|\bar{v}^{k+1} - v^k\|^2 - 2d_0(\bar{v}^{k+1}, v^k) + 2d_0(v, v^k). \quad \blacksquare$$

In the next result, we establish two fundamental estimates relating the exact and inexact iterates from an optimal solution.

Lemma 4.3 *Let $(d_0, H_0) \in \mathcal{F}(\bar{\mathrm{C}})$, $(d'_0, H'_0) \in \mathcal{F}(\bar{\mathrm{K}})$ be a proximal and induced proximal distance that satisfy the premises of Definition 2 and let $\{(x^k, z^k, y^k)\}$ be the sequence generated by $(PMAPD)$ algorithm. Then, for any $k \geq 0$*

$$\begin{array}{rcl} (i)\ \|\bar{x}^{k+1} - x^*\|^2 + \|\bar{z}^{k+1} - z^*\|^2 & \leq & \|x^k - x^*\|^2 + \|z^k - z^*\|^2 - \{\|\bar{x}^{k+1} - x^k\|^2 + \|\bar{z}^{k+1} - z^k\|^2\} \\ & & -2\lambda_k \left\langle p^{k+1} - y^*, A\bar{x}^{k+1} + B\bar{z}^{k+1} - b\right\rangle \\ & & -2[d_0(\bar{x}^{k+1}, x^k) + d'_0(\bar{z}^{k+1}, z^k)] + 2[d_0(x^*, x^k) + d'_0(z^*, z^k)]; \end{array}$$

$$\begin{array}{rcl} (ii)\ \|\bar{y}^{k+1} - y^*\|^2 & \leq & \|y^k - y^*\|^2 - \{\|p^{k+1} - \bar{y}^{k+1}\|^2 + \|p^{k+1} - y^k\|^2\} \\ & & -2\lambda_k\{\left\langle y^* - \bar{y}^{k+1}, A\bar{x}^{k+1} + B\bar{z}^{k+1} - b\right\rangle + \left\langle \bar{y}^{k+1} - p^{k+1}, Ax^k + Bz^k - b\right\rangle\}. \end{array}$$

Proof: (i) From Step 2 of (PMAPD) algorithm the sequences $\{\bar{x}^k\}$, $\{\bar{z}^k\}$ are obtained when $a_k = b_k = 0$, so,

$$\begin{array}{rcl} \bar{x}^{k+1} & = & \arg\min\{f^k(x) + \delta_{\bar{\mathrm{C}}}(x) + (1/\lambda_k)d(x, x^k)\}, \\ \bar{z}^{k+1} & = & \arg\min\{g^k(z) + \delta_{\bar{\mathrm{K}}}(z) + (1/\lambda_k)d'(z, z^k)\}. \end{array}$$

from Remark 2 and applying Lemma 4.2 twice with the choice $F(\cdot) := (f^k + \delta_{\bar{\mathrm{C}}})(\cdot)$, $F(\cdot) := (g^k + \delta_{\bar{\mathrm{K}}})(\cdot)$ at the optimal point $x = x^*$ and $z = z^*$ respectively, we obtain

$$\begin{array}{rcl} 2\lambda_k[f^k(\bar{x}^{k+1}) - f^k(x^*)] & \leq & \|x^k - x^*\|^2 - \|\bar{x}^{k+1} - x^*\|^2 - \|\bar{x}^{k+1} - x^k\|^2 \\ & & -2d_0(\bar{x}^{k+1}, x^k) + 2d_0(x^*, x^k), \\ 2\lambda_k[g^k(\bar{z}^{k+1}) - g^k(z^*)] & \leq & \|z^k - z^*\|^2 - \|\bar{z}^{k+1} - z^*\|^2 - \|\bar{z}^{k+1} - z^k\|^2 \\ & & -2d'_0(\bar{z}^{k+1}, z^k) + 2d'_0(z^*, z^k), \end{array}$$

adding the above two inequalities and from definition of Lagrangian L, we obtain

$$\begin{array}{rcl} 2\lambda_k[L(\bar{x}^{k+1}, \bar{z}^{k+1}, p^{k+1}) - L(x^*, z^*, p^{k+1})] & \leq & \|x^k - x^*\|^2 + \|z^k - z^*\|^2 \\ & & -(\|\bar{x}^{k+1} - x^*\| + \|\bar{z}^{k+1} - z^*\|^2) \\ & & -(\|\bar{x}^{k+1} - x^k\| + \|\bar{z}^{k+1} - z^k\|^2) \\ & & -2[d_0(\bar{x}^{k+1}, x^k) + d'_0(\bar{z}^{k+1}, z^k)] \\ & & +2[d_0(x^*, x^k) + d'_0(z^*, z^k)]. \end{array} \tag{10}$$

Since (x^*, z^*, y^*) is a saddle point for Lagrangian $L(x, z, y)$, we also have

$$2\lambda_k[L(x^*, z^*, y^*) - L(\bar{x}^{k+1}, \bar{z}^{k+1}, y^*)] \leq 0. \tag{11}$$

Adding the inequalities (10) and (11) after rearranging terms, we get

$$\begin{aligned} \|\bar{x}^{k+1}-x^*\|^2+\|\bar{z}^{k+1}-z^*\|^2 \quad &\leq \quad \|x^k-x^*\|^2+\|z^k-z^*\|^2-\{\|\bar{x}^{k+1}-x^k\|^2+\|\bar{z}^{k+1}-z^k\|^2\} \\ &\quad -2\lambda_k\left\langle p^{k+1}-y^*, A\bar{x}^{k+1}+B\bar{z}^{k+1}-b\right\rangle \\ &\quad -2[d_0(\bar{x}^{k+1},x^k)+d_0'(\bar{z}^{k+1},z^k)]+2[d_0(x^*,x^k)+d_0'(z^*,z^k)]. \end{aligned}$$

On the other hand, note that Step 1 of (PMAPD) algorithm and from definition of $\bar{y}^{k+1}$ we obtain: $p^{k+1} = \arg\min\{-L(x^k,z^k,y)+(1/(2\lambda_k))\|y-y^k\|^2\}$. Then, using Lemma 4.1 twice with the choice $\tau=\lambda_k, F(y)=-L(x^k.z^k,y)$ and $F(y)=-L(\bar{x}^{k+1},\bar{z}^{k+1},y)$ respectively, we obtain

$$\begin{aligned} 2\lambda_k[L(x^k,z^k,\bar{y}^{k+1})-L(\bar{x}^k,z^k,p^{k+1})] &\leq \|y^k-\bar{y}^{k+1}\|^2-\|p^{k+1}-\bar{y}^{k+1}\|^2-\|p^{k+1}-y^k\|^2, \\ 2\lambda_k[L(\bar{x}^{k+1},\bar{z}^{k+1},y^*)-L(\bar{x}^{k+1},\bar{z}^{k+1},\bar{y}^{k+1})] &\leq \|y^k-y^*\|^2-\|\bar{y}^{k+1}-y^*\|^2-\|\bar{y}^{k+1}-y^k\|^2. \end{aligned}$$

Adding both inequalities and after rearranging terms, we obtain (ii). ■

Proposition 4.4 *Suppose that the assumptions of Lemma 4.3 are satisfied and suppose that* $\{\lambda_k\}$ *satisfies* $\eta<\lambda_k<\bar{c}-\eta$ *(for all k), for some* $\eta\in(0,\bar{c}/2)$ *with* $\bar{c}:=\min\{\frac{\sqrt{\gamma}}{2\|A\|},\frac{\sqrt{\gamma'}}{2\|B\|}\}$ *where* γ,γ' *are positive constants related to the* d *and* d'*, respectively, in Definition 2. Let* (x^*,z^*) *an optimal solution for (PC) and* y^* *be a corresponding Lagrange multiplier. Then, for any* $k\geq 0$

$$\begin{aligned} \|\bar{w}^{k+1}-w^*\|^2 \quad &\leq \quad \|w^k-w^*\|^2-D\{\|\bar{x}^{k+1}-x^k\|^2+\|\bar{z}^{k+1}-z^k\|^2+\|p^{k+1}-\bar{y}^{k+1}\|^2 \\ &\quad +\|p^{k+1}-y^k\|^2\}-2\{d_0(\bar{x}^{k+1},x^k)-d_0(x^*,x^k)+d_0'(\bar{z}^{k+1},z^k)-d_0'(z^*,z^k)\} \end{aligned} \tag{12}$$

where $D:=\min\{1-4(\bar{c}-\eta)^2\|A\|^2, 1-4(\bar{c}-\eta)^2\|B\|^2\}$.

Proof: We denote $w=(x,z,y)$ with associated norm $\|w\|^2=\|x\|^2+\|z\|^2+\|y\|^2$. Adding the inequalities $(i)-(ii)$ of Lemma 4.3, we obtain

$$\begin{aligned} \|\bar{w}^{k+1}-w^*\|^2 \quad &\leq \quad \|w^k-w^*\|^2-\{\|\bar{x}^{k+1}-x^k\|^2+\|\bar{z}^{k+1}-z^k\|^2\}-\{\|p^{k+1}-\bar{y}^{k+1}\|^2 \\ &\quad +\|p^{k+1}-y^k\|^2\}-2\{d_0(\bar{x}^{k+1},x^k)+d_0'(\bar{z}^{k+1},z^k)\} \\ &\quad +2\{d_0(x^*,x^k)+d_0'(z^*,z^k)\}+\phi \end{aligned} \tag{13}$$

where $\phi:=2\lambda_k\left\langle \bar{y}^{k+1}-p^{k+1}, A(\bar{x}^{k+1}-x^k)+B(\bar{z}^{k+1}-z^k)\right\rangle$.
Since $\bar{y}^{k+1}=y^k+\lambda_k(A\bar{x}^{k+1}+B\bar{z}^{k+1}-b)$ and $p^{k+1}=y^k+\lambda_k(Ax^k+Bz^k-b)$, using the inequality $(r+q)^2\leq 2(r^2+q^2)$, we obtain

$$\phi\leq 4\lambda_k^2\|A\|^2\|\bar{x}^{k+1}-x^k\|^2+4\lambda_k^2\|B\|^2\|\bar{z}^{k+1}-z^k\|^2. \tag{14}$$

From (13) and (14), we get

$$\begin{aligned} \|\bar{w}^{k+1}-w^*\|^2 \quad &\leq \quad \|w^k-w^*\|^2-(1-4\lambda_k^2\|A\|^2)\|\bar{x}^{k+1}-x^k\|^2-(1-4\lambda_k^2\|B\|^2)\|\bar{z}^{k+1}-z^k\|^2 \\ &\quad -\{\|p^{k+1}-\bar{y}^{k+1}\|^2+\|p^{k+1}-y^k\|^2\}-2\{d_0(\bar{x}^{k+1},x^k)+d_0'(\bar{z}^{k+1},z^k)\} \\ &\quad +2\{d_0(x^*,x^k)+d_0'(z^*,z^k)\}. \end{aligned} \tag{15}$$

Note that from assumption for λ_k, we have $\lambda_k<\bar{c}-\eta,$ with $\bar{c}:=\min\{\frac{\sqrt{\gamma}}{2\|A\|},\frac{\sqrt{\gamma'}}{2\|B\|}\}$, and together with the definition of γ and γ' (see Definition 2) and $\eta\in(0,\bar{c}/2)$, we obtain

$$0<1-4(\bar{c}-\eta)^2\|A\|^2<1-4\lambda_k^2\|A\|^2, \quad 0<1-4(\bar{c}-\eta)^2\|B\|^2<1-4\lambda_k^2\|B\|^2.$$

Therefore, considering $D:=\min\{1-4(\bar{c}-\eta)^2\|A\|^2, 1-4(\bar{c}-\eta)^2\|B\|^2\}>0$, in the inequality (15), we obtain

$$\begin{aligned} \|\bar{w}^{k+1}-w^*\|^2 \quad &\leq \quad \|w^k-w^*\|^2-D\{\|\bar{x}^{k+1}-x^k\|^2+\|\bar{z}^{k+1}-z^k\|^2+\|p^{k+1}-\bar{y}^{k+1}\|^2 \\ &\quad +\|p^{k+1}-y^k\|^2\}-2\{d_0(\bar{x}^{k+1},x^k)-d_0(x^*,x^k)+d_0'(\bar{z}^{k+1},z^k)-d_0'(z^*,z^k)\}. \end{aligned}$$

■

Lemma 4.5 *If the Assumption* $(\mathcal{T}_2)$ *holds, then*

$$\|w^{k+1} - \bar{w}^{k+1}\| \le \delta_k \|w^{k+1} - w^k\|$$
$$\text{where } \delta_k := \eta_k \max\{\sqrt{1 + 2\lambda_k^2 \|A\|^2}, \sqrt{1 + 2\lambda_k^2\|B\|^2}\}. \tag{16}$$

Proof: We have

$$\begin{aligned} y^{k+1} &= y^k + \lambda_k(Ax^{k+1} + Bz^{k+1} - b), \quad \|x^{k+1} - \bar{x}^{k+1}\| \le \eta_k \|x^{k+1} - x^k\|, \\ \bar{y}^{k+1} &= y^k + \lambda_k(A\bar{x}^{k+1} + B\bar{z}^{k+1} - b), \quad \|z^{k+1} - \bar{z}^{k+1}\| \le \eta_k\|z^{k+1} - z^k\|, \end{aligned}$$

then using the inequality $(r+q)^2 \le 2(r^2 + q^2)$, we obtain

$$\|y^{k+1} - \bar{y}^{k+1}\|^2 \le 2(\eta_k\lambda_k)^2(\|A\|^2\|x^{k+1} - x^k\|^2 + \|B\|^2\|z^{k+1} - z^k\|^2).$$

Therefore from assumption $(\mathcal{T}_2)$ and the definition of δ_k given in the lemma,

$$\begin{aligned} \|w^{k+1} - \bar{w}^{k+1}\|^2 &= \|x^{k+1} - \bar{x}^{k+1}\|^2 + \|z^{k+1} - \bar{z}^{k+1}\|^2 + \|y^{k+1} - \bar{y}^{k+1}\|^2 \\ &\le \eta_k^2(\|x^{k+1} - x^k\| + \|z^{k+1} - z^k\|^2) \\ &\quad +2(\eta_k\lambda_k)^2(\|A\|^2\|x^{k+1} - x^k\|^2 + \|B\|^2\|z^{k+1} - z^k\|^2) \\ &\le \eta_k^2\{(1 + 2\lambda_k^2\|A\|^2)\|x^{k+1} - x^k\|^2 + (1 + 2\lambda_k^2\|B\|^2)\|z^{k+1} - z^k\|^2\} \\ &\le \delta_k^2\|w^{k+1} - w^k\|^2. \end{aligned}$$

■

We can now state and prove our convergence rate result.

Theorem 4.6 *Let* $(d_0, H_0) \in \mathcal{F}_+(\bar{C})$, $(d_0', H_0') \in \mathcal{F}_+(\bar{K})$ *be a proximal and induced proximal distance satisfying the condition (Iviii). Let* $\{(x^k, z^k, y^k)\}$ *be a bounded sequence generated by (PMAPD) algorithm and suppose that the assumptions of Lemma 4.3 and* $(\mathcal{T}_1) - (\mathcal{T}_4)$ *hold and* λ_k *satisfies* $\eta < \lambda_k < \bar{c} - \eta$ *(for all k), for some* $\eta \in (0, \bar{c}/2)$ *with* $\bar{c} := \min\{\frac{\sqrt{\gamma}}{2\|A\|}, \frac{\sqrt{\gamma'}}{2\|B\|}\}$. *Then,* $\{w^k\}$ *converges linearly to the unique optimal solution* $w^* := (x^*, z^*, y^*)$*, that is, there exists an integer* $\bar{k}$ *such that, for all* $k \ge \bar{k}$

$$\|w^{k+1} - w^*\| \le \theta_k \|w^k - w^*\| \tag{17}$$

where $\theta_k \le \frac{\sqrt{a^2N+D}+a\sqrt{N}}{2\sqrt{a^2N+D}} < 1$ *with* D *defined in (12) and* $N := \max\{4(\bar{c} - \eta)^2(\|A^TA\|^2 + \|B^TA\|^2) + 2\alpha^2\eta^{-2}, 4(\bar{c} - \eta)^2(\|A^TB\|^2 + \|B^TB\|^2) + 2(\alpha')^2\eta^{-2}\}$.

Proof: Under our assumptions, $\{w^k\}$ is bounded and considering

$$a_k = \eta_k\|x^{k+1} - x^k\|, \;\; b_k = \eta_k\|z^{k+1} - z^k\|$$

we obtain $\sum_{k=1}^{+\infty}(a_k + b_k) < +\infty$, and therefore Theorem 3.2 holds and $\{w^k\}$ converges to w^*.
We now establish the rate of convergence.
From Step 2 of (PMAPD) algorithm, when $a_k = b_k = 0$,

$$\begin{aligned} -\lambda_k^{-1}\nabla_1 d(\bar{x}^{k+1}, x^k) \in \partial f^k(\bar{x}^{k+1}), \\ -\lambda_k^{-1}\nabla_1 d'(\bar{z}^{k+1}, z^k) \in \partial g^k(\bar{z}^{k+1}), \end{aligned}$$

furthermore,

$$\begin{aligned} 0 &\in \partial f(\bar{x}^{k+1}) + A^Tp^{k+1} + \lambda_k^{-1}\nabla_1 d(\bar{x}^{k+1}, x^k) \\ &= \partial f(\bar{x}^{k+1}) + A^T\bar{y}^{k+1} - A^T(\bar{y}^{k+1} - p^{k+1}) + \lambda_k^{-1}\nabla_1 d(\bar{x}^{k+1}, x^k) \end{aligned} \tag{18}$$

$$\begin{array}{rcl} 0 & \in & \partial g(\bar{z}^{k+1}) + B^T p^{k+1} + \lambda_k^{-1}\nabla_1 d'(\bar{z}^{k+1}, z^k) \\ & = & \partial g(\bar{z}^{k+1}) + B^T \bar{y}^{k+1} - B^T(\bar{y}^{k+1} - p^{k+1}) + \lambda_k^{-1}\nabla_1 d'(\bar{z}^{k+1}, z^k). \end{array} \tag{19}$$

From Remark 4, we have $\bar{y}^{k+1} = y^k + \lambda_k(A\bar{x}^{k+1} + B\bar{z}^{k+1} - b)$, then

$$-\lambda_k^{-1}(\bar{y}^{k+1} - y^k) = b - A\bar{x}^{k+1} - B\bar{z}^{k+1}. \tag{20}$$

Furthermore, since

$$\begin{array}{c} \partial_x L(\bar{x}^{k+1}, \bar{z}^{k+1}, \bar{y}^{k+1}) = \partial f(\bar{x}^{k+1}) + A^T\bar{y}^{k+1}, \\ \partial_z L(\bar{x}^{k+1}, \bar{z}^{k+1}, \bar{y}^{k+1}) = \partial g(\bar{z}^{k+1}) + B^T\bar{y}^{k+1}, \\ \partial_y(-L(\bar{x}^{k+1}, \bar{z}^{k+1}, \bar{y}^{k+1})) = b - A\bar{x}^{k+1} - B\bar{z}^{k+1}. \end{array}$$

then from (18), (19), (20) and Definition of S, see (7), we obtain

$$(\pi_k, \sigma_k, \xi_k) \in S(\bar{x}^{k+1}, \bar{z}^{k+1}, \bar{y}^{k+1})$$

where

$$\pi_k := A^T(\bar{y}^{k+1} - p^{k+1}) - \lambda_k^{-1}\nabla_1 d(\bar{x}^{k+1}, x^k), \tag{21}$$

$$\sigma_k := B^T(\bar{y}^{k+1} - p^{k+1}) - \lambda_k^{-1}\nabla_1 d'(\bar{z}^{k+1}, z^k), \tag{22}$$

$$\xi_k := -\lambda_k^{-1}(\bar{y}^{k+1} - y^k). \tag{23}$$

From Step 1 of (PMAPD) algorithm $p^{k+1} = y^k + \lambda_k(Ax^k + Bz^k - b)$, and since $\bar{y}^{k+1} = y^k + \lambda_k(A\bar{x}^{k+1} + B\bar{z}^{k+1} - b)$ then by subtracting, we obtain

$$\bar{y}^{k+1} - p^{k+1} = \lambda_k(A(\bar{x}^{k+1} - x^k) + B(\bar{z}^{k+1} - z^k)). \tag{24}$$

Substituting (24) in (21)-(23), we get

$$\pi_k = \lambda_k A^T(A(\bar{x}^{k+1} - x^k) + B(\bar{z}^{k+1} - z^k)) - \lambda_k^{-1}\nabla_1 d(\bar{x}^{k+1}, x^k) \tag{25}$$

$$\sigma_k = \lambda_k B^T(A(\bar{x}^{k+1} - x^k) + B(\bar{z}^{k+1} - z^k)) - \lambda_k^{-1}\nabla_1 d'(\bar{z}^{k+1}, z^k) \tag{26}$$

$$\xi_k = -\lambda_k^{-1}(\bar{y}^{k+1} - y^k). \tag{27}$$

On the other hand, let (x^∞, z^∞) be an optimal solution of (CP) with y^∞ be a corresponding Lagrange multiplier such that $w^k = (x^k, z^k, y^k)$ converges to $w^\infty = (x^\infty, z^\infty, y^\infty)$, we have

$$\|\bar{w}^{k+1} - w^\infty\| \le \|\bar{w}^{k+1} - w^{k+1}\| + \|w^{k+1} - w^\infty\|,$$

since $\|w^{k+1} - w^\infty\| \to 0$ and by Lemma 4.5, $\|\bar{w}^{k+1} - w^{k+1}\| \to 0$ $(k \to +\infty)$, then

$$\|\bar{w}^{k+1} - w^\infty\| \to 0, \quad (k \to +\infty)$$

using this result, considering the assumption $(\mathcal{T}_3)$ and taking the limit on both sides of (12), we obtain

$$\|\bar{x}^{k+1} - x^k\| \to 0, \quad \|\bar{z}^{k+1} - z^k\| \to 0, \quad \|p^{k+1} - \bar{y}^{k+1}\| \to 0, \quad \|p^{k+1} - y^k\| \to 0. \tag{28}$$

From Assumption $(\mathcal{T}_4)$, $\nabla_1 d_0(\cdot, u)$ is locally Lipchitz continuous, and since $\|\bar{x}^{k+1} - x^\infty\| \to 0$ and $\|x^k - x^\infty\| \to 0$ with $x^\infty \in C$ then, there exist $\alpha_1 > 0$, $k_1', k_1'' \in I\!N$ such that

$$\|\nabla_1 d_0(\bar{x}^{k+1}, x^k) - \nabla_1 d_0(x^k, x^k)\| \le \alpha_1\|\bar{x}^{k+1} - x^k\|, \ \ \forall k \ge k_1''' := \max\{k_1', k_1''\},$$

therefore,

$$\begin{array}{rcl} \|\nabla_1 d(\bar{x}^{k+1}, x^k)\| & = & \|\nabla_1 d_0(\bar{x}^{k+1}, x^k) + (\bar{x}^{k+1} - x^k)\| \\ & = & \|\nabla_1 d_0(\bar{x}^{k+1}, x^k) - \nabla_1 d_0(x^k, x^k) + (\bar{x}^{k+1} - x^k)\| \\ & \le & \alpha\|\bar{x}^{k+1} - x^k\|, \ \ \forall k \ge k_1''' := \max\{k_1', k_1''\} \end{array} \tag{29}$$

where $\alpha = \alpha_1 + 1$.
Analogously, there exists $\alpha_2 > 0$, $k_2', k_2'' \in I\!N$ such that

$$\|\nabla_1 d'(\bar{z}^{k+1}, z^k)\| \leq \alpha' \|\bar{z}^{k+1} - z^k\|, \ \ \forall k \geq k_2''' := \max\{k_2', k_2''\}, \tag{30}$$

where $\alpha' = \alpha_2 + 1$.
Thus, using (28), (29) and (30) in (25)-(27) with $\eta < \lambda_k < \bar{c} - \eta$, we obtain

$$(\pi_k, \sigma_k, \xi_k) \to 0, \ \ (k \to +\infty).$$

That is, there exist $\ddot{k}$ such that $\|(\pi_k, \sigma_k, \xi_k)\| < \tau$ for all $k \geq \ddot{k}$ and using the Assumption $(\mathcal{T}_1)$ and the facts that $0 \in S(x^*, z^*, y^*)$ and $(\pi_k, \sigma_k, \xi_k) \in S(\bar{x}^{k+1}, \bar{z}^{k+1}, \bar{y}^{k+1})$, with the choice

$$w^* = (x^*, z^*, y^*) \quad v = (\pi_k, \sigma_k, \xi_k),$$

we obtain

$$\|\bar{w}^{k+1} - w^*\| \leq a\|(\pi_k, \sigma_k, \xi_k)\| \quad \forall\, k \geq \ddot{k}. \tag{31}$$

We will estimate the right side of inequality (31). Using the definition of (π_k, σ_k, ξ_k), the inequality $(r+q)^2 \leq 2(r^2+q^2)$ and (29) - (30), we obtain

$$\begin{array}{rcl}
\|\pi_k\|^2 & \leq & (4\lambda_k^2\|A^TA\|^2 + 2\alpha^2\lambda_k^{-2})\|\bar{x}^{k+1} - x^k\|^2 + 4\lambda_k^2\|A^TB\|^2\|\bar{z}^{k+1} - z^k\|^2, \ \ \forall k \geq k_1''' \\
\|\sigma_k\|^2 & \leq & 4\lambda_k^2\|B^TA\|^2\|\bar{x}^{k+1} - x^k\|^2 + (2(\alpha')^2\lambda_k^{-2} + 4\lambda_k^2\|B^TB\|^2)\|\bar{z}^{k+1} - z^k\|^2, \ \ \forall k \geq k_2''' \\
\|\xi_k\|^2 & \leq & 2\lambda_k^{-2}(\|p^{k+1} - \bar{y}^{k+1}\|^2 + \|p^{k+1} - y^k\|^2).
\end{array}$$

Therefore, for all $k \geq k_3''' := \max\{k_1''', k_2'''\}$, we obtain

$$\begin{array}{rcl}
\|(\pi_k, \sigma_k, \xi_k)\| & = & \|\pi_k\|^2 + \|\sigma_k\|^2 + \|\xi_k\|^2 \\
& \leq & D_k\{\|\bar{x}^{k+1} - x^k\|^2 + \|\bar{z}^{k+1} - z^k\|^2 + \|p^{k+1} - \bar{y}^{k+1}\|^2 + \|p^{k+1} - y^k\|^2\}
\end{array} \tag{32}$$

where

$$D_k := \max\{4\lambda_k^2(\|A^TA\|^2 + \|B^TA\|^2) + 2\alpha^2\lambda_k^{-2}; 4\lambda_k^2(\|A^TB\|^2 + \|B^TB\|^2) + 2(\alpha')^2\lambda_k^{-2}; 2\lambda_k^{-2}\},$$

since $\eta < \lambda_k < \bar{c} - \eta$, it is clear that $D_k \geq M$ with $M := 2(\bar{c} - \eta)^{-2} > 0$. Moreover, since $\alpha > 1$, $\alpha' > 1$ and considering the definition of $N := \max\{4(\bar{c} - \eta)^2(\|A^TA\|^2 + \|B^TA\|^2) + 2\alpha^2\eta^{-2}, 4(\bar{c} - \eta)^2(\|A^TB\|^2 + \|B^TB\|^2) + 2(\alpha')^2\eta^{-2}\}$, we obtain that $D_k < N$, and therefore

$$0 < M \leq D_k < N. \tag{33}$$

Now, from (31), (32) and (33), for all $k \geq \hat{k} := \max\{\ddot{k}, k_3'''\}$, we obtain

$$\begin{array}{rcl}
\|\bar{w}^{k+1} - w^*\|^2 & \leq & a^2 D_k\{\|\bar{x}^{k+1} - x^k\|^2 + \|\bar{z}^{k+1} - z^k\|^2 + \|p^{k+1} - \bar{y}^{k+1}\|^2 + \|p^{k+1} - y^k\|^2\} \\
& < & a^2 N\{\|\bar{x}^{k+1} - x^k\|^2 + \|\bar{z}^{k+1} - z^k\|^2 + \|p^{k+1} - \bar{y}^{k+1}\|^2 + \|p^{k+1} - y^k\|^2\}
\end{array} \tag{34}$$

furthermore, from (12) and Assumption $(\mathcal{T}_3)$, for all $k \geq k_0$, we obtain

$$\begin{array}{rcl}
\|\bar{w}^{k+1} - w^*\|^2 & \leq & \|w^k - w^*\|^2 - D\{\|\bar{x}^{k+1} - x^k\|^2 + \|\bar{z}^{k+1} - z^k\|^2 + \|p^{k+1} - \bar{y}^{k+1}\|^2 \\
& & + \|p^{k+1} - y^k\|^2\}
\end{array} \tag{35}$$

thus, for $\check{k} := \max\{\hat{k}, k_0\}$, multiplying by D and a^2N in the inequalities (34) and (35), respectively and after adding, we obtain

$$a^2N\|\bar{w}^{k+1} - w^*\|^2 + D\|\bar{w}^{k+1} - w^*\|^2 \leq a^2N\|w^k - w^*\|^2 \quad \forall k \geq \check{k}.$$

Defining $\nu := a\sqrt{N}/\sqrt{a^2N+D}$, the latter inequality reduces to

$$\|\bar{w}^{k+1} - w^*\| \leq \nu\|w^k - w^*\| \quad \forall k \geq \check{k}. \tag{36}$$

But, from Lemma 4.5, we have

$$\begin{array}{rcl} \|w^{k+1} - \bar{w}^{k+1}\| & \leq & \delta_k\|w^{k+1} - w^k\| = \delta_k\|(w^{k+1} - w^*) + (w^* - w^k)\| \\ & \leq & \delta_k\|w^{k+1} - w^*\| + \delta_k\|w^k - w^*\|. \end{array} \tag{37}$$

Therefore from (36) and (37), we obtain

$$\begin{array}{rcl} \|w^{k+1} - w^*\| & = & \|(w^{k+1} - \bar{w}^{k+1}) + (\bar{w}^{k+1} - w^*)\| \\ & \leq & \delta_k\|w^{k+1} - w^*\| + \delta_k\|w^k - w^*\| + \nu\|w^k - w^*\|, \quad \forall k \geq \check{k}, \end{array}$$

wich proved (17) with $\theta_k = \frac{\nu+\delta_k}{1-\delta_k}$. Since $\delta_k \to 0$ and

$$\frac{\sqrt{a^2N+D} + a\sqrt{N}}{2\sqrt{a^2N+D}} > \frac{a\sqrt{N}}{\sqrt{a^2N+D}} = \nu,$$

for some $\bar{k} \geq \check{k}$, we have

$$1 > \frac{\sqrt{a^2N+D} + a\sqrt{N}}{2\sqrt{a^2N+D}} \geq \theta_k.$$

■

5. Conclusions

In this paper, we presented the global convergence result of the proximal multiplier method using regularized proximal distances (PMAPD) and we proved under mild assumptions that its iterations generated converge linearly to the unique optimal solution of the problem.

Acknowledgements

This research, for the first author, was supported by CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior). Part of the research of the second author was supported by the Postdoctoral Scholarship CAPES-FAPERJ Edital PAPD-2011.

References

Auslender, A. and Teboulle, M. (2001), Entropic proximal decomposition methods for convex programs and variational inequalities, *Math. Program.*, Ser. A, vol. 91, p. 33-47.

Auslender, A. and Teboulle, M. (2006), Interior gradient and proximal methods for convex and conic optimization, *SIAM J. Optim.*, vol. 16, 3, p. 697-725.

Chen, G. and Teboulle, M. (1994), A proximal-based decomposition method for convex minimization problems, *Mathematical Programming*, Serie A, 1, vol. 64, p. 81-101.

Goldfarb, D., Ma, S. and Scheinberg, K. (2013), Fast alternating linearization methods for minimizing the sum of two convex functions, *Math. Program.* 141, no. 1-2, Ser. A, p. 349-382.

Kyono, M. and Fukushima, M. (2000), Nonlinear proximal decomposition method for convex programming, *JOTA*, vol. 106, N. 2, p. 357-372.

Lenoir, A. Modèles et algorithmes pour la planification de production à moyen terme en environnement incertain. Application de méthodes de décomposition proximales, *Ph.D. Thesis*, Université Blaise Pascal - Clermont-Ferrand II, 2008.

Mahey, P., Ouorou, A., LeBlanc, L. and Chifflet, J. (1997), A new proximal decomposition algorithm for routing in telecommunication networks, *Networks*, vol. 31, p. 227-238.

Rockafellar, R.T. Convex Analysis, *Princeton University Press*, Princeton New Jersey, 1970.

Rockafellar, R.T. (1976), Augmented Lagrangians and Applications of the Proximal Point Algorithm in Convex Programming, *Mathematics of Operations Research*, vol. 1, p. 97-116.

Rockafellar, R.T. (1976), Monotone operators and the proximal point algorithm, *SIAM Journal on Control and Optimization*, vol. 14, N. 5, p. 877-898.

Sarmiento, O., Quiroz, E.A.P. and Oliveira, P.R. (2015), A proximal multiplier method for separable convex minimization, *accepted for publication in Optimization Journal*.