

Convergence rate of a proximal multiplier algorithm for separable convex minimization

O. Sarmiento*, E. A. Papa Quiroz† and P. R. Oliveira‡

Systems Engineering and Computer Science Program, COPPE, Federal University of Rio de Janeiro, postal address 68511, Rio de Janeiro, RJ 21941-972, Brazil.

Department of Mathematical Sciences, San Marcos National University, Lima, Perú.

(June 2015)

This paper analyze the convergence rate of a proximal algorithm called Proximal Multiplier Algorithm with Proximal Distances (PMAPD), proposed by us in [20], to solve convex minimization problems with separable structure. We prove that, under mild assumptions, its primal-dual sequences converge linearly to the optimal solution for a class of proximal distances.

Keywords: Proximal multiplier method; separable structure; proximal distances, linear rate of convergence.

AMS Subject Classification:

1. Introduction

Consider the following separable convex optimization problem:

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

where $\mathbf{C} \subset \mathbb{R}^n$ and $\mathbf{K} \subset \mathbb{R}^p$ are nonempty open convex sets, $\bar{\mathbf{C}}$ and $\bar{\mathbf{K}}$ denote the closure (in the euclidean topology) of \mathbf{C} and \mathbf{K} respectively, $f : \mathbb{R}^n \rightarrow (-\infty, +\infty]$ and $g : \mathbb{R}^p \rightarrow (-\infty, +\infty]$ are closed proper convex functions and $A \in \mathbb{R}^{m \times n}$, $B \in \mathbb{R}^{m \times p}$, $b \in \mathbb{R}^m$.

The Lagrangian for (CP) is a function $L : \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m \rightarrow (-\infty, +\infty]$, defined by

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

where δ_X denotes the indicator function of a subset X , $\langle \cdot, \cdot \rangle$ denotes the canonical inner product. The variable y is called the Lagrangian multiplier associated with the constraint $Ax + Bz = b$.

*Email: osarmiento@cos.ufrj.br

†Corresponding author. Email: erikpapa@gmail.com

‡Email: poliveir@cos.ufrj.br

Observe that the Lagrangian $L(x, z, y)$ is a closed convex-concave function. Therefore, the set-valued subdifferential mapping S on $\mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m$ given by

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

is maximal monotone, see Rockafellar [18]. Remember that, a pair (x^*, z^*) is optimal for (CP) and y^* is an optimal Lagrangian multiplier if and only if

$$L(x^*, z^*, y) \leq L(x^*, z^*, y^*) \leq L(x, z, y^*), \quad \forall (x, z) \in \mathbb{R}^{n \times p}, \forall y \in \mathbb{R}^m,$$

that is, if and only if $0 \in S(x^*, z^*, y^*)$, (in the sequel, 0 denotes a zero element in $\mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m$).

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. [15], Management Electricity, see Lenoir [13], and in computer science to solve matrix completion problems, see Example 2 of Goldfarb et al. [9].

Various decomposition methods that exploit the special structure of the problem have been proposed. Some examples of such methods are: alternating directions method of multipliers [3, 7, 8, 14], partial inverse of Spingarn method [6, 15, 22, 23], predictor corrector proximal multipliers methods of Chen and Teboulle and its extensions [1, 5, 12]. See also the recent works [10, 11] and references therein.

In a recent paper, see [20], we proposed the following inexact iteration called Proximal Multiplier Algorithm with Proximal Distances (PMAPD) to solve the problem (CP): Given $(x^k, z^k, y^k) \in \mathbb{C} \times \mathbb{K} \times \mathbb{R}^m$, find $(x^{k+1}, v^{k+1}) \in \mathbb{C} \times \mathbb{R}^n$ and $(z^{k+1}, \xi^{k+1}) \in \mathbb{K} \times \mathbb{R}^p$ such that

$$\begin{aligned} p^{k+1} &= y^k + \lambda_k(Ax^k + Bz^k - b) \\ v^{k+1} &\in \partial_{a_k} f^k(x^{k+1}), \quad v^{k+1} + \lambda_k^{-1} \nabla_1 d(x^{k+1}, x^k) = 0, \\ \xi^{k+1} &\in \partial_{b_k} g^k(z^{k+1}), \quad \xi^{k+1} + \lambda_k^{-1} \nabla_1 d'(z^{k+1}, z^k) = 0. \\ y^{k+1} &= y^k + \lambda_k(Ax^{k+1} + Bz^{k+1} - b) \end{aligned}$$

where the functions $f^k : \mathbb{R}^n \rightarrow (-\infty, +\infty]$ and $g^k : \mathbb{R}^p \rightarrow (-\infty, +\infty]$ 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, with p^{k+1} as above, $\partial_{a_k} f^k$ and $\partial_{b_k} g^k$ are the a_k -subdifferential and b_k -subdifferential of f^k and g^k respectively (i.e., are the usual ϵ -subdifferential with $\epsilon = a_k$ and $\epsilon = b_k$ defined in Section 2), d, d' are regularized proximal distances, see Subsection 2.1 and Section 3, and ∇_1 denote the gradient with respect to the first variable.

The (PMAPD) is an extension of the (PCPM) and (NPCPMM) methods (see [5] and [12] respectively) and includes the class of φ -divergence distances, see Subsection 3.3 of [2]. Observe also that the (EPDM) studied in [1] is a particular case of our algorithm when we consider exact iterations and we use the regularized log-quadratic distance, see Section 2 of [1].

In [20] we proved the global convergence of the (PMAPD) and we gave some computational experiments of the proposed algorithm. In this paper we are interesting in its rate of convergence.

The main contribution of this paper is that, under some natural assumptions in proximal point methods, we prove two types of linear rate of convergence:

- If the primal-dual induced proximal distance \hat{H} , defined in Remark 5, satisfies

$\hat{H}(w, w') := (\hat{\alpha}/2) \|w - w'\|^2$, with $\hat{\alpha} > 0$, then there exists an integer \bar{k} such that

$$\|w^{k+1} - w^*\| \leq \theta_k \|w^k - w^*\|,$$

for all $k \geq \bar{k}$, where θ is a certain constant between $(0, 1)$.

- If $\hat{H}(w, w') < (\hat{\beta}/2) \|w - w'\|^2$, with $\hat{\beta} > 0$, then for the exact iteration of the algorithm $\bar{w}^k = (\bar{x}^k, \bar{z}^k, \bar{y}^k)$ there exists an integer \hat{k} such that

$$\hat{H}(w^*, \bar{w}^{k+1}) \leq \bar{\theta} \hat{H}(w^*, \bar{w}^k),$$

for all $k \geq \hat{k}$ with $0 < \bar{\theta} < 1$.

We can show that there exists a large class of proximal distances satisfying the above conditions, see Remark 7 and Remark 8 respectively. Thus we can conclude that we obtain the linear rate of convergence of (PMAPD) for a large class of proximal distances.

The paper is organized as follows: In Section 2 we give some results in convex analysis and we present the class of proximal distances that we will use along the paper. In Section 3 we introduce the Proximal Multiplier Algorithm with Proximal Distances (PMAPD) and we show the results of global convergence of the iterations to the solution of the (CP) obtained in [20]. In Section 4 we analyze the convergence rate of the (PMAPD) and we obtain the linear convergence for a class of proximal distances. Finally we present our conclusions and give some directions for future research.

2. Some results in convex analysis and proximal distance

Throughout the paper \mathbb{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 $M \in \mathbb{R}^{m \times n}$ we define $\|M\| := \max_{\|x\| \leq 1} \|Mx\|$. Given an extended real valued function $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{\pm\infty\}$ we denote its domain by $\text{dom} f := \{x \in \mathbb{R}^n : f(x) < +\infty\}$ and its epigraph $\text{epi} f := \{(x, \beta) \in \mathbb{R}^n \times \mathbb{R} : f(x) \leq \beta\}$. f is said to be proper, if $\text{dom} f \neq \emptyset$ and for all $x \in \text{dom} f$, we have $f(x) > -\infty$. Also denote by $\text{ri}(X)$ the relative interior set of $X \subset \mathbb{R}^n$ and $\partial_\epsilon f$ is the ϵ -subdifferential of f defined by $\partial_\epsilon f(u) = \{p \in \mathbb{R}^n : f(v) \geq f(u) + \langle p, v - u \rangle - \epsilon, \forall v \in \text{dom} f\}$.

Finally, f is a lower semicontinuous function if for each $x \in \mathbb{R}^n$ we have that all $\{x^l\}$ such that $\lim_{l \rightarrow +\infty} x^l = x$ implies that $f(x) \leq \liminf_{l \rightarrow +\infty} f(x^l)$. It is easy to prove that the lower semicontinuity of f is equivalent to the closedness of the lower level set $L_f(\alpha) = \{x \in \mathbb{R}^n : f(x) \leq \alpha\}$, for each $\alpha \in \mathbb{R}$. Recall that if f is a proper convex function, then f is closed if f is lower semicontinuous.

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 [2].

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

- (i) $d(\cdot, y)$ is proper, closed, convex and continuously differentiable on \mathbf{C} ;
- (ii) $\text{dom } d(\cdot, y) \subset \bar{\mathbf{C}}$ and $\text{dom } \partial_1 d(\cdot, y) = \mathbf{C}$, 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., $\lim_{\|u\| \rightarrow \infty} d(u, y) = +\infty$).
- (iv) $d(y, y) = 0$, which clearly implies $\nabla_1 d(y, y) = 0$.

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

Associated to a proximal distance to be an induced proximal distance which we define as follows:

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

- (Ii) $H(a, a) = 0$.
- (Iii) $\langle c - b, \nabla_1 d(b, a) \rangle \leq H(c, a) - H(c, b) - \gamma H(b, a), \quad \forall c \in \mathbf{C}$.

We write $(d, H) \in \mathcal{F}(\mathbf{C})$ to the proximal and induced proximal distance that satisfies the premises of Definition 2.

We also denote $(d, H) \in \mathcal{F}(\bar{\mathbf{C}})$ if there exists H such that:

- (Iiii) H is finite valued on $\bar{\mathbf{C}} \times \mathbf{C}$ satisfying (Ii) and (Iii), for each $c \in \bar{\mathbf{C}}$.
- (Iiv) For each $c \in \bar{\mathbf{C}}$, $H(c, \cdot)$ has level bounded sets on \mathbf{C} .

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

- (Iv) $(d, H) \in \mathcal{F}(\bar{\mathbf{C}})$.
- (Ivi) $\forall y \in \bar{\mathbf{C}}$ and $\forall \{y^k\} \subset \mathbf{C}$ bounded with $\lim_{k \rightarrow +\infty} H(y, y^k) = 0$, we have $\lim_{k \rightarrow +\infty} y^k = y$.
- (Ivii) $\forall y \in \bar{\mathbf{C}}$ and $\forall \{y^k\} \subset \mathbf{C}$ such that $\lim_{k \rightarrow +\infty} y^k = y$, we obtain $\lim_{k \rightarrow +\infty} H(y, y^k) = 0$.

Several examples of proximal distances which satisfy the above definitions, for example Bregman distances, proximal distances based on φ -divergences, self-proximal distances, and distances based on second order homogeneous proximal distances, were given by Auslender and Teboulle [2].

The following additional conditions on H will be useful to prove the convergence of (PMAPD) algorithm.

Given $(d, H) \in \mathcal{F}_+(\bar{\mathbf{C}})$, H satisfies the following condition

- (Iviii) $\forall c \in \bar{\mathbf{C}}$ and $\forall \{y^k\} \subset \mathbf{C}$ such that $\lim_{k \rightarrow +\infty} y^k = y$, we obtain

$$\lim_{k \rightarrow +\infty} H(c, y^k) = H(c, y).$$

Some examples of proximal distances which satisfy this condition, were showed by Sarmiento, Papa Quiroz and Oliveira [20], Section 7.

3. The Algorithm

In the proposed algorithm we used the class of proximal distances $(d_0, H_0) \in \mathcal{F}_+(\bar{\mathbf{C}})$, $(d'_0, H'_0) \in \mathcal{F}_+(\bar{\mathbf{K}})$, satisfying the condition (Iviii) and given $\mu > 0, \mu' > 0$

we defined the following functions:

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

$$H(x, y) = H_0(x, y) + (\mu/2) \|x - y\|^2, \quad (2)$$

$$d'(x, y) = d'_0(x, y) + (\mu'/2) \|x - y\|^2, \quad (3)$$

$$H'(x, y) = H'_0(x, y) + (\mu'/2) \|x - y\|^2. \quad (4)$$

It is easy to check that $(d, H) \in \mathcal{F}_+(\bar{\mathcal{C}})$ and $(d', H') \in \mathcal{F}_+(\bar{\mathcal{K}})$ (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:

The (PMAPD) Algorithm

Step 0. Choose two pairs $(d_0, H_0) \in \mathcal{F}_+(\bar{\mathcal{C}})$, $(d'_0, H'_0) \in \mathcal{F}_+(\bar{\mathcal{K}})$ satisfying the condition (Iviii) and define (d, H) , (d', H') given by (1)-(2) and (3)-(4) respectively. Take three sequences $a_k \geq 0$, $b_k \geq 0$ and $\lambda_k > 0$ and choose an arbitrary starting point $(x^0, z^0, y^0) \in \mathcal{C} \times \mathcal{K} \times \mathbb{R}^m$.

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

$$p^{k+1} = y^k + \lambda_k (Ax^k + Bz^k - b). \quad (5)$$

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

$$v^{k+1} \in \partial_{a_k} f^k(x^{k+1}), \quad v^{k+1} + \lambda_k^{-1} \nabla_1 d(x^{k+1}, x^k) = 0, \quad (6)$$

$$\xi^{k+1} \in \partial_{b_k} g^k(z^{k+1}), \quad \xi^{k+1} + \lambda_k^{-1} \nabla_1 d'(z^{k+1}, z^k) = 0. \quad (7)$$

where the functions $f^k : \mathbb{R}^n \rightarrow (-\infty, +\infty]$ and $g^k : \mathbb{R}^p \rightarrow (-\infty, +\infty]$ 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

$$y^{k+1} = y^k + \lambda_k (Ax^{k+1} + Bz^{k+1} - b). \quad (8)$$

Stop criterium: 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 [20], the following results were obtained

THEOREM 3.1 *Let $d_0 \in \mathcal{D}(\mathcal{C})$ and $d'_0 \in \mathcal{D}(\mathcal{K})$. Suppose that the problem (CP) has an optimal solution (x^*, z^*) and a corresponding Lagrange multiplier y^* and there exists $x \in \text{ri}(\text{dom } d(\cdot, v)) \cap \text{ri}(\text{dom } f)$ and $z \in \text{ri}(\text{dom } d'(\cdot, v')) \cap \text{ri}(\text{dom } g)$ such that $Ax + Bz = b$, then for any $(x^k, z^k, y^k) \in \mathcal{C} \times \mathcal{K} \times \mathbb{R}^m$, $\lambda_k > 0$, there exists a unique point $(x^{k+1}, z^{k+1}) \in \mathcal{C} \times \mathcal{K}$ satisfying (6) and (7).*

THEOREM 3.2 *Let $(d_0, H_0) \in \mathcal{F}_+(\bar{\mathcal{C}})$, $(d'_0, H'_0) \in \mathcal{F}_+(\bar{\mathcal{K}})$ 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 $\{(x^k, z^k, y^k)\}$ 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 (3) 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 analyze the global convergence rate of the (PMAPD) algorithm. We start given some remarks

Remark 1 The Lagrangian $L(x, z, y)$ is a closed convex-concave function. Therefore, the set-valued subdifferential mapping S on $\mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m$ given by

$$S(x, z, y) = \partial_{x,z}L(x, z, y) \times \partial_y(-L(x, z, y)) \quad (9)$$

is maximal monotone, see Rockafellar [18].

Remark 2 Let T be a set valued maximal monotone operator on \mathbb{R}^n . Following Rockafellar [19], 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 $\|u\| \leq \tau$.

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

$$\|x^{k+1} - \bar{x}^{k+1}\| \leq \eta_k \|x^{k+1} - x^k\|, \quad \|z^{k+1} - \bar{z}^{k+1}\| \leq \eta_k \|z^{k+1} - z^k\| \quad (10)$$

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 [18], p. 100), to derive the rate of convergence of the proximal method of multipliers. It was used in the work of Chen and Teboulle [5] and will also be used here to derive the rate of convergence for (PMAPD) algorithm.

Also we assume an assumption to the proximal distances $(d_0, H_0) \in \mathcal{F}_+(\bar{\mathcal{C}})$, $(d'_0, H'_0) \in \mathcal{F}_+(\bar{\mathcal{K}})$.

- (\mathcal{T}_3) The functions $\nabla_1 d_0(\cdot, u)$ and $\nabla_1 d'_0(\cdot, v)$ satisfy the following statement.
For any $x_0 \in \bar{\mathcal{C}}$, $z_0 \in \bar{\mathcal{K}}$ there exists $\alpha_1 > 0, \alpha_2 > 0$ and $r_1 > 0, r_2 > 0$ such that

$$\begin{aligned} \|\nabla_1 d_0(x, u) - \nabla_1 d_0(\bar{x}, u)\| &\leq \alpha_1 \|x - \bar{x}\|, \quad \forall x, \bar{x} \in B_{r_1}(x_0) \cap \mathcal{C}, \quad \forall u \in \mathcal{C}, \\ \|\nabla_1 d'_0(z, v) - \nabla_1 d'_0(\bar{z}, v)\| &\leq \alpha_2 \|z - \bar{z}\|, \quad \forall z, \bar{z} \in B'_{r_2}(z_0) \cap \mathcal{K}, \quad \forall v \in \mathcal{K}, \end{aligned}$$

where $B_{r_1}(x_0) := \{x \in \mathbb{R}^n : \|x - x_0\| < r_1\}$ and $B'_{r_2}(z_0) := \{z \in \mathbb{R}^p : \|z - z_0\| < r_2\}$.

Remark 3 With respect to the introduced 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 and Remark 2, 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 (10) 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 (10).

(3) Note that φ -divergence proximal distances and second order homogeneous proximal distances satisfy the Assumption (\mathcal{T}_3) for any $x_0 \in \mathbf{C}$ and $z_0 \in \mathbf{K}$. Indeed, from Definition

$$d_\varphi(x, y) := \sum_{i=1}^n y_i \varphi\left(\frac{x_i}{y_i}\right) \text{ with } \varphi \in C^2(\mathbb{R}_{++}),$$

$$d_\varphi(x, y) := \sum_{i=1}^n y_i^2 \varphi\left(\frac{x_i}{y_i}\right)$$

with $\varphi(t) = \mu p(t) + \frac{\nu}{2}(t-1)^2$, $\nu \geq \mu p''(1) > 0$, $p \in C^2(\mathbb{R}_{++}^n)$, respectively. We note that in both cases, $\nabla_1 d_\varphi \in C^1(\mathbf{C})$ and $\nabla_1 d'_\varphi \in C^1(\mathbf{K})$ with $\mathbf{C} = \mathbb{R}_{++}^n$ and $\mathbf{K} = \mathbb{R}_{++}^p$. Then, $\nabla_1 d_\varphi(\cdot, y)$ and $\nabla_1 d'_\varphi(\cdot, y)$ are locally Lipschitz continuous on \mathbf{C} and \mathbf{K} respectively. Therefore, (\mathcal{T}_3) holds for any $x_0 \in \mathbf{C}$ and $z_0 \in \mathbf{K}$. Also if we consider h Bregman functions such that $h \in C^2$, then by Definition of Bregman distance, $\nabla_1 D_h(\cdot, y) = \nabla h(\cdot) - \nabla h(y)$ is continuously differentiable, then $\nabla_1 D_h(\cdot, y)$ and $\nabla_1 D'_h(\cdot, y)$ are locally Lipschitz continuous on \mathbf{C} and \mathbf{K} respectively. Therefore, (\mathcal{T}_3) holds for any $x_0 \in \mathbf{C}$ and $z_0 \in \mathbf{K}$.

Remark 4 Given $\bar{x}^{k+1}, \bar{z}^{k+1}$ defined in Assumption (\mathcal{T}_2) . We define

$$\begin{aligned} \bar{y}^{k+1} &= y^k + \lambda_k (A\bar{x}^{k+1} + B\bar{z}^{k+1} - b) \\ &= \arg \min \{-L(\bar{x}^{k+1}, \bar{z}^{k+1}, y) + (1/(2\lambda_k))\|y - y^k\|^2\}. \end{aligned}$$

The subsequent convergence rate analysis follows a line of argument similar to that given by Chen and Teboulle in [5].

Before proving our convergent rate result, we need some previous results.

LEMMA 4.1 ([5], Lemma 3.1) *Let $F : \mathbb{R}^m \rightarrow (-\infty, +\infty]$ a closed proper convex function, $\tau > 0$ and define:*

$$u^{k+1} = \arg \min_{u \in \mathbb{R}^m} \{F(u) + (1/(2\tau))\|u - u^k\|^2\}.$$

Then for any integer $k \geq 0$,

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

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

LEMMA 4.2 *Let $(d_0, H_0) \in \mathcal{F}(\bar{\mathcal{C}})$, $(d'_0, H'_0) \in \mathcal{F}(\bar{\mathcal{K}})$ and let $\{(x^k, z^k, y^k)\}$ be the sequence generated by (PMAPD) algorithm. Then for any $k \geq 0$*

$$(i) \quad H(x^*, \bar{x}^{k+1}) + H'(z^*, \bar{z}^{k+1}) \leq H(x^*, x^k) + H'(z^*, z^k) - (\gamma H(\bar{x}^{k+1}, x^k) + \gamma' H'(\bar{z}^{k+1}, z^k)) \\ - \lambda_k \langle p^{k+1} - y^*, A\bar{x}^{k+1} + B\bar{z}^{k+1} - b \rangle$$

$$(ii) \quad \frac{1}{2} \|\bar{y}^{k+1} - y^*\|^2 \leq \frac{1}{2} \|y^k - y^*\|^2 - \frac{1}{2} (\|p^{k+1} - \bar{y}^{k+1}\|^2 + \|p^{k+1} - y^k\|^2) \\ - \lambda_k (\langle y^* - \bar{y}^{k+1}, A\bar{x}^{k+1} + B\bar{z}^{k+1} - b \rangle + \langle \bar{y}^{k+1} - p^{k+1}, Ax^k + Bz^k - b \rangle).$$

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,

$$\bar{x}^{k+1} = \arg \min \{f^k(x) + \delta_{\bar{\mathcal{C}}}(x) + (1/\lambda_k)d(x, x^k)\}, \\ \bar{z}^{k+1} = \arg \min \{g^k(z) + \delta_{\bar{\mathcal{K}}}(z) + (1/\lambda_k)d'(z, z^k)\}.$$

Then,

$$f^k(x) \geq f^k(x^{k+1}) + \langle -\lambda_k^{-1} \nabla_1 d(\bar{x}^{k+1}, x^k), x - \bar{x}^{k+1} \rangle, \quad \forall x \in \text{dom } f \\ g^k(z) \geq f^k(z^{k+1}) + \langle -\lambda_k^{-1} \nabla_1 d'(\bar{z}^{k+1}, z^k), z - \bar{z}^{k+1} \rangle, \quad \forall z \in \text{dom } g$$

so from Definition 2 (Iii), we obtain

$$\lambda_k (f^k(\bar{x}^{k+1}) - f^k(x)) \leq \langle \nabla_1 d(\bar{x}^{k+1}, x^k), x - \bar{x}^{k+1} \rangle \\ \leq H(x, x^k) - H(x, \bar{x}^{k+1}) - \gamma H(\bar{x}^{k+1}, x^k). \quad (11)$$

and

$$\lambda_k (g^k(\bar{z}^{k+1}) - g^k(z)) \leq \langle \nabla_1 d'(\bar{z}^{k+1}, z^k), z - \bar{z}^{k+1} \rangle \\ \leq H'(z, z^k) - H'(z, \bar{z}^{k+1}) - \gamma H'(\bar{z}^{k+1}, z^k). \quad (12)$$

adding the inequalities (11) and (12), using the definition of Lagrangian L and taking $(x, z) = (x^*, z^*)$, we obtain

$$\lambda_k [L(\bar{x}^{k+1}, \bar{z}^{k+1}, p^{k+1}) - L(x^*, z^*, p^{k+1})] \leq H(x^*, x^k) + H'(z^*, z^k) \\ - H(x^*, \bar{x}^{k+1}) - H'(z^*, \bar{z}^{k+1}) \\ - \gamma H(\bar{x}^{k+1}, x^k) - \gamma' H'(\bar{z}^{k+1}, z^k). \quad (13)$$

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

$$2\lambda_k [L(x^*, z^*, y^*) - L(\bar{x}^{k+1}, \bar{z}^{k+1}, y^*)] \leq 0. \quad (14)$$

Adding the inequalities (13) and (14) after rearranging terms, we get

$$H(x^*, \bar{x}^{k+1}) + H'(z^*, \bar{z}^{k+1}) \leq H(x^*, x^k) + H'(z^*, z^k) - (\gamma H(\bar{x}^{k+1}, x^k) + \gamma' H'(\bar{z}^{k+1}, z^k)) \\ - \lambda_k \langle p^{k+1} - y^*, A\bar{x}^{k+1} + B\bar{z}^{k+1} - b \rangle.$$

On the other hand, note that Step 1 of (PMAPD) algorithm can be written equivalently as:

$$p^{k+1} = \arg \min\{-L(x^k, z^k, y) + (1/(2\lambda_k))\|y - y^k\|^2\},$$

and from definition of \bar{y}^{k+1} (see Remark 4)

$$\bar{y}^{k+1} = \arg \min\{-L(\bar{x}^{k+1}, \bar{z}^{k+1}, y) + (1/(2\lambda_k))\|y - y^k\|^2\},$$

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

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

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

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

Remark 5 Throughout the rest of this paper, we denote $w^k = (x^k, z^k, y^k)$, $w^* = (x^*, z^*, y^*)$, $w = (x, z, y)$, $s = (l, q, r) \in \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m$ and define the function $\hat{H} : X \times X \rightarrow \mathbb{R}_+ \cup \{+\infty\}$ by

$$\hat{H}(w, s) = \hat{H}((x, z, y), (l, q, r)) = H(x, l) + H'(z, q) + \frac{1}{2}\|y - r\|^2, \quad (15)$$

where $X = \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m$, H and H' are defined in (2) and (4) respectively.

PROPOSITION 4.3 *Suppose that the assumptions of Theorem 3.2 are satisfied. Let $\{w^k\}$ and $\{p^k\}$ be sequences generated by the (PMAPD); let (x^*, z^*) be an optimal solution of (CP), and let y^* be a corresponding Lagrange multiplier, then we have, for each k ,*

$$\begin{aligned} \hat{H}(w^*, \bar{w}^{k+1}) &\leq \hat{H}(w^*, w^k) - D\{\|\bar{x}^{k+1} - x^k\|^2 + \|\bar{z}^{k+1} - z^k\|^2 + \|p^{k+1} - \bar{y}^{k+1}\|^2 \\ &\quad + \|p^{k+1} - y^k\|^2\} \end{aligned} \quad (16)$$

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

Proof. Adding the inequalities (i) – (ii) of Lemma 4.2, we obtain

$$\begin{aligned} \hat{H}(w^*, \bar{w}^{k+1}) &\leq \hat{H}(w^*, w^k) - \gamma H(\bar{x}^{k+1}, x^k) - \gamma' H'(\bar{z}^{k+1}, z^k) \\ &\quad - \frac{1}{2}(\|p^{k+1} - \bar{y}^{k+1}\|^2 + \|p^{k+1} - y^k\|^2) \\ &\quad + \lambda_k \langle \bar{y}^{k+1} - p^{k+1}, A(\bar{x}^{k+1} - x^k) + B(\bar{z}^{k+1} - z^k) \rangle. \end{aligned} \quad (17)$$

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

$$\begin{aligned} \hat{H}(w^*, \bar{w}^{k+1}) &\leq \hat{H}(w^*, w^k) - \gamma H(\bar{x}^{k+1}, x^k) - \gamma' H'(\bar{z}^{k+1}, z^k) \\ &\quad - \frac{1}{2}(\|p^{k+1} - \bar{y}^{k+1}\|^2 + \|p^{k+1} - y^k\|^2) \\ &\quad + 2\lambda_k^2(\|A\|^2\|\bar{x}^{k+1} - x^k\|^2 + \|B\|^2\|\bar{z}^{k+1} - z^k\|^2). \end{aligned} \quad (18)$$

From (17) and (18), using the definitions of H and H' given in (2) and (4) respectively, we get

$$\begin{aligned}\hat{H}(w^*, \bar{w}^{k+1}) &\leq \hat{H}(w^*, w^k) - (\gamma H_0(\bar{x}^{k+1}, x^k) + (1/2)(\gamma\mu - 4\lambda_k^2\|A\|^2)\|\bar{x}^{k+1} - x^k\|^2) \\ &\quad - (\gamma' H'_0(\bar{z}^{k+1}, z^k) + (1/2)(\gamma'\mu' - 4\lambda_k^2\|B\|^2)\|\bar{z}^{k+1} - z^k\|^2) \\ &\quad - (1/2)(\|p^{k+1} - \bar{y}^{k+1}\|^2 + \|p^{k+1} - y^k\|^2).\end{aligned}\tag{19}$$

By simple algebra, using $\lambda_k < \bar{c} - \eta$, with $\bar{c} := \min\{\frac{\sqrt{\mu\gamma}}{2\|A\|}, \frac{\sqrt{\mu'\gamma'}}{2\|B\|}\}$ and $\eta \in (0, \bar{c}/2)$ one can verify that

$$\begin{aligned}0 &< \mu\gamma - 4(\bar{c} - \eta)^2\|A\|^2 < \mu\gamma - 4\lambda_k^2\|A\|^2, \\ 0 &< \mu'\gamma' - 4(\bar{c} - \eta)^2\|B\|^2 < \mu'\gamma' - 4\lambda_k^2\|B\|^2.\end{aligned}$$

Therefore, considering $D := \min\{\frac{\mu\gamma}{2} - 2(\bar{c} - \eta)^2\|A\|^2, \frac{\mu'\gamma'}{2} - 2(\bar{c} - \eta)^2\|B\|^2, \frac{1}{2}\} > 0$, in the inequality (19), we obtain

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

□

LEMMA 4.4 *If the Assumption (\mathcal{T}_2) holds, then*

$$\|w^{k+1} - \bar{w}^{k+1}\| \leq \delta_k \|w^{k+1} - w^k\|$$

$$\text{where } \delta_k := \eta_k \max\{\sqrt{1 + 2\lambda_k^2\|A\|^2}, \sqrt{1 + 2\lambda_k^2\|B\|^2}\}.\tag{20}$$

Proof. We have

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

then

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

so, using the inequality $(r + q)^2 \leq 2(r^2 + q^2)$, we obtain

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

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

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

□

We can now state and prove our convergent rate results.

THEOREM 4.5 *Let $(d_0, H_0) \in \mathcal{F}_+(\bar{\mathcal{C}})$, $(d'_0, H'_0) \in \mathcal{F}_+(\bar{\mathcal{K}})$ satisfying the condition (Iviii) with $\hat{H}(w, w') := (\hat{\alpha}/2) \|w - w'\|^2$. Let $\{(x^k, z^k, y^k)\}$ be a sequence generated by (PMAPD) algorithm and suppose that the assumptions of Theorem 3.2 and $(\mathcal{T}_1) - (\mathcal{T}_3)$ hold. 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 $k \geq \bar{k}$*

$$\|w^{k+1} - w^*\| \leq \theta_k \|w^k - w^*\|$$

where $\theta_k := \frac{\delta_k}{1-\delta_k} < \frac{\rho}{1-\rho} < 1$ for $\rho \in (0, 1/2)$ or $\theta_k \leq \frac{\sqrt{a^2 \hat{\alpha} N + 2D} + a\sqrt{N\hat{\alpha}}}{2\sqrt{a^2 \hat{\alpha} N + 2D}} < 1$ with D defined in (16) and $N := \max\{4(\bar{c} - \eta)^2 (\|A^T A\|^2 + \|B^T A\|^2) + 2(\alpha_1 + \mu)^2 \eta^{-2}, 4(\bar{c} - \eta)^2 (\|A^T B\|^2 + \|B^T B\|^2) + 2(\alpha_2 + \mu')^2 \eta^{-2}, 2\eta^{-2}\}$.

Proof. Under our assumptions, $\{w^k\}$ satisfy the assumptions of Theorem 3.2 therefore, $\{w^k\}$ 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,

$$\begin{aligned}
0 &\in \partial f(\bar{x}^{k+1}) + A^T p^{k+1} + \lambda_k^{-1} \nabla_1 d(\bar{x}^{k+1}, x^k) \\
&= \partial f(\bar{x}^{k+1}) + A^T \bar{y}^{k+1} - A^T (\bar{y}^{k+1} - p^{k+1}) + \lambda_k^{-1} \nabla_1 d(\bar{x}^{k+1}, x^k)
\end{aligned} \tag{21}$$

$$\begin{aligned}
0 &\in \partial g(\bar{z}^{k+1}) + B^T p^{k+1} + \lambda_k^{-1} \nabla_1 d'(\bar{z}^{k+1}, z^k) \\
&= \partial g(\bar{z}^{k+1}) + B^T \bar{y}^{k+1} - B^T (\bar{y}^{k+1} - p^{k+1}) + \lambda_k^{-1} \nabla_1 d'(\bar{z}^{k+1}, z^k).
\end{aligned} \tag{22}$$

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

$$-\lambda_k^{-1} (\bar{y}^{k+1} - y^k) = b - A\bar{x}^{k+1} - B\bar{z}^{k+1}. \tag{23}$$

Furthermore, since

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

then from (21), (22), (23) and Definition of S, see (9), we obtain

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

where

$$\pi_k := A^T(\bar{y}^{k+1} - p^{k+1}) - \lambda_k^{-1} \nabla_1 d(\bar{x}^{k+1}, x^k), \quad (24)$$

$$\sigma_k := B^T(\bar{y}^{k+1} - p^{k+1}) - \lambda_k^{-1} \nabla_1 d'(\bar{z}^{k+1}, z^k), \quad (25)$$

$$\xi_k := -\lambda_k^{-1}(\bar{y}^{k+1} - y^k). \quad (26)$$

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

$$\bar{y}^{k+1} - p^{k+1} = \lambda_k(A(\bar{x}^{k+1} - x^k) + B(\bar{z}^{k+1} - z^k)). \quad (27)$$

Substituting (27) in (24)-(26), we get

$$\pi_k = \lambda_k A^T(A(\bar{x}^{k+1} - x^k) + B(\bar{z}^{k+1} - z^k)) - \lambda_k^{-1} \nabla_1 d(\bar{x}^{k+1}, x^k) \quad (28)$$

$$\sigma_k = \lambda_k B^T(A(\bar{x}^{k+1} - x^k) + B(\bar{z}^{k+1} - z^k)) - \lambda_k^{-1} \nabla_1 d'(\bar{z}^{k+1}, z^k) \quad (29)$$

$$\xi_k = -\lambda_k^{-1}(\bar{y}^{k+1} - y^k). \quad (30)$$

On the other hand, let (x^∞, z^∞) a 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

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

since $\|w^{k+1} - w^\infty\| \rightarrow 0$ and by Lemma 4.4, $\|\bar{w}^{k+1} - w^{k+1}\| \rightarrow 0$ ($k \rightarrow +\infty$), then

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

using this result, considering Definition 2 (Ivii) para H and H' , we obtain

$$\hat{H}(w^\infty, \bar{w}^k) \rightarrow 0 \quad \text{and} \quad \hat{H}(w^\infty, w^k) \rightarrow 0,$$

then taking the limit on both sides of (16), we obtain

$$\begin{aligned}\|\bar{x}^{k+1} - x^k\| &\rightarrow 0, \quad \|\bar{z}^{k+1} - z^k\| \rightarrow 0 \\ \|p^{k+1} - \bar{y}^{k+1}\| &\rightarrow 0, \quad \|p^{k+1} - y^k\| \rightarrow 0.\end{aligned} \quad (31)$$

We note that

$$\begin{aligned}\|\nabla_1 d(\bar{x}^{k+1}, x^k)\| &= \|\nabla_1 d_0(\bar{x}^{k+1}, x^k) + \mu(\bar{x}^{k+1} - x^k)\| \\ &= \|\nabla_1 d_0(\bar{x}^{k+1}, x^k) - \nabla_1 d_0(x^k, x^k) + \mu(\bar{x}^{k+1} - x^k)\|.\end{aligned}$$

From assumption (\mathcal{T}_3) , $\nabla_1 d_0(\cdot, u)$ is locally Lipschitz continuous, and since $\|\bar{x}^{k+1} - x^\infty\| \rightarrow 0$ and $\|x^k - x^\infty\| \rightarrow 0$ with $x^\infty \in \bar{\mathcal{C}}$ then, there exists $\alpha_1 > 0$, $k'_1, k''_1 \in \mathbb{N}$ such that

$$\|\nabla_1 d_0(\bar{x}^{k+1}, x^k) - \nabla_1 d_0(x^k, x^k)\| \leq \alpha_1 \|\bar{x}^{k+1} - x^k\|, \quad \forall k \geq k_1''' := \max\{k'_1, k''_1\},$$

therefore,

$$\begin{aligned}\|\nabla_1 d(\bar{x}^{k+1}, x^k)\| &= \|\nabla_1 d_0(\bar{x}^{k+1}, x^k) + \mu(\bar{x}^{k+1} - x^k)\| \\ &= \|\nabla_1 d_0(\bar{x}^{k+1}, x^k) - \nabla_1 d_0(x^k, x^k) + \mu(\bar{x}^{k+1} - x^k)\| \\ &\leq \|\nabla_1 d_0(\bar{x}^{k+1}, x^k) - \nabla_1 d_0(x^k, x^k)\| + \mu \|\bar{x}^{k+1} - x^k\| \\ &\leq (\alpha_1 + \mu) \|\bar{x}^{k+1} - x^k\|, \quad \forall k \geq k_1''' := \max\{k'_1, k''_1\}.\end{aligned}$$

Analogously, from Assumption (\mathcal{T}_3) and since $\|\bar{z}^{k+1} - z^\infty\| \rightarrow 0$, $\|z^k - z^\infty\| \rightarrow 0$ with $z^\infty \in \bar{\mathcal{K}}$ then, there exists $\alpha_2 > 0$, $k'_2, k''_2 \in \mathbb{N}$ such that

$$\|\nabla_1 d'_0(\bar{z}^{k+1}, z^k) - \nabla_1 d'_0(z^k, z^k)\| \leq \alpha_2 \|\bar{z}^{k+1} - z^k\|, \quad \forall k \geq k_2''' := \max\{k'_2, k''_2\},$$

thus,

$$\begin{aligned}\|\nabla_1 d'(\bar{z}^{k+1}, z^k)\| &= \|\nabla_1 d'_0(\bar{z}^{k+1}, z^k) + \mu'(\bar{z}^{k+1} - z^k)\| \\ &= \|\nabla_1 d'_0(\bar{z}^{k+1}, z^k) - \nabla_1 d'_0(z^k, z^k) + \mu'(\bar{z}^{k+1} - z^k)\| \\ &\leq \|\nabla_1 d'_0(\bar{z}^{k+1}, z^k) - \nabla_1 d'_0(z^k, z^k)\| + \mu' \|\bar{z}^{k+1} - z^k\| \\ &\leq (\alpha_2 + \mu') \|\bar{z}^{k+1} - z^k\|, \quad \forall k \geq k_2''' := \max\{k'_2, k''_2\}.\end{aligned}$$

Therefore, we have

$$\begin{aligned}\|\nabla_1 d(\bar{x}^{k+1}, x^k)\| &\leq (\alpha_1 + \mu) \|\bar{x}^{k+1} - x^k\|, \quad \forall k \geq k_1''' := \max\{k'_1, k''_1\} \\ \|\nabla_1 d'(\bar{z}^{k+1}, z^k)\| &\leq (\alpha_2 + \mu') \|\bar{z}^{k+1} - z^k\|, \quad \forall k \geq k_2''' := \max\{k'_2, k''_2\},\end{aligned} \tag{32}$$

where $\alpha_1, \alpha_2, \mu, \mu'$ are positive constants. Thus, using (31), (32) in (28)-(30) with $\eta < \lambda_k < \bar{c} - \eta$, we obtain

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

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

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

we obtain

$$\|\bar{w}^{k+1} - w^*\| \leq a \|(\pi_k, \sigma_k, \xi_k)\| \quad \forall k \geq \bar{k}. \quad (33)$$

We note that if $a = 0$ then $\bar{w}^{k+1} = w^* \quad \forall k \geq \bar{k}$, and by Lemma 4.4, we obtain

$$\begin{aligned} \|w^{k+1} - w^*\| &= \|w^{k+1} - \bar{w}^{k+1}\| \\ &\leq \delta_k \|w^{k+1} - w^k\| \\ &= \delta_k \|(w^{k+1} - w^*) + (w^* - w^k)\| \\ &\leq \delta_k \|w^{k+1} - w^*\| + \delta_k \|w^k - w^*\| \end{aligned}$$

thus,

$$\|w^{k+1} - w^*\| \leq \theta_k \|w^k - w^*\|, \quad \text{with } \theta_k = \frac{\delta_k}{1 - \delta_k}.$$

Since $\delta_k \rightarrow 0$ (see (20) and the fact that $\eta_k \rightarrow 0$) then there exists $\bar{k} \in \mathbb{N}$ such that

$$\theta_k = \frac{\delta_k}{1 - \delta_k} < \frac{\rho}{1 - \rho} < 1, \quad \forall k \geq \bar{k}$$

for $\rho \in (0, 1/2)$.

Now, consider the case when $a > 0$.

We will estimate the right side of inequality (33). Using the definition of (π_k, σ_k, ξ_k) and the inequality $(r + q)^2 \leq 2(r^2 + q^2)$, we obtain

$$\begin{aligned} \|\pi_k\|^2 &\leq 2\{[\lambda_k(\|A^T A\| \|\bar{x}^{k+1} - x^k\| + \|A^T B\| \|\bar{z}^{k+1} - z^k\|)]^2 + \lambda_k^{-2} \|\nabla_1 d(\bar{x}^{k+1}, x^k)\|^2\} \\ &\leq 2\{2[\lambda_k^2 \|A^T A\|^2 \|\bar{x}^{k+1} - x^k\|^2 + \lambda_k^2 \|A^T B\|^2 \|\bar{z}^{k+1} - z^k\|^2] + \lambda_k^{-2} \|\nabla_1 d(\bar{x}^{k+1}, x^k)\|^2\} \\ &= 4\lambda_k^2 [\|A^T A\|^2 \|\bar{x}^{k+1} - x^k\|^2 + \|A^T B\|^2 \|\bar{z}^{k+1} - z^k\|^2] + 2\lambda_k^{-2} \|\nabla_1 d(\bar{x}^{k+1}, x^k)\|^2 \\ \|\sigma_k\|^2 &\leq 4\lambda_k^2 [\|B^T A\|^2 \|\bar{x}^{k+1} - x^k\|^2 + \|B^T B\|^2 \|\bar{z}^{k+1} - z^k\|^2] + 2\lambda_k^{-2} \|\nabla_1 d'(\bar{z}^{k+1}, z^k)\|^2 \\ \|\xi_k\|^2 &\leq \lambda_k^{-2} \|\bar{y}^{k+1} - y^k\|^2 = \lambda_k^{-2} (\|\bar{y}^{k+1} - p^{k+1}\| + \|p^{k+1} - y^k\|)^2 \\ &\leq 2\lambda_k^{-2} (\|\bar{y}^{k+1} - p^{k+1}\|^2 + \|p^{k+1} - y^k\|^2), \end{aligned}$$

thus, from (32), we obtain

$$\begin{aligned} \|\pi_k\|^2 &\leq 4\lambda_k^2 (\|A^T A\|^2 \|\bar{x}^{k+1} - x^k\|^2 + \|A^T B\|^2 \|\bar{z}^{k+1} - z^k\|^2) + 2(\alpha_1 + \mu)^2 \lambda_k^{-2} \|\bar{x}^{k+1} - x^k\|^2, \quad \forall k \geq k_1''' \\ &= (4\lambda_k^2 \|A^T A\|^2 + 2(\alpha_1 + \mu)^2 \lambda_k^{-2}) \|\bar{x}^{k+1} - x^k\|^2 + 4\lambda_k^2 \|A^T B\|^2 \|\bar{z}^{k+1} - z^k\|^2, \quad \forall k \geq k_1''' \\ \|\sigma_k\|^2 &\leq 4\lambda_k^2 \|B^T A\|^2 \|\bar{x}^{k+1} - x^k\|^2 + (2(\alpha_2 + \mu')^2 \lambda_k^{-2} + 4\lambda_k^2 \|B^T B\|^2) \|\bar{z}^{k+1} - z^k\|^2, \quad \forall k \geq k_2''' \\ \|\xi_k\|^2 &\leq 2\lambda_k^{-2} (\|p^{k+1} - \bar{y}^{k+1}\|^2 + \|p^{k+1} - y^k\|^2). \end{aligned}$$

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

$$\begin{aligned} \|(\pi_k, \sigma_k, \xi_k)\| &= \|\pi_k\|^2 + \|\sigma_k\|^2 + \|\xi_k\|^2 \\ &\leq D_k \{\|\bar{x}^{k+1} - x^k\|^2 + \|\bar{z}^{k+1} - z^k\|^2 + \|p^{k+1} - \bar{y}^{k+1}\|^2 + \|p^{k+1} - y^k\|^2\} \end{aligned} \quad (34)$$

where

$$D_k := \max\{4\lambda_k^2(\|A^T A\|^2 + \|B^T A\|^2) + 2(\alpha_1 + \mu)^2\lambda_k^{-2}; 4\lambda_k^2(\|A^T B\|^2 + \|B^T B\|^2) + 2(\alpha_2 + \mu')^2\lambda_k^{-2}; 2\lambda_k^{-2}\},$$

since $\eta < \lambda_k < \bar{c} - \eta$, we note that D_k is positive and limited.

In fact, since $D_k \geq 2\lambda_k^{-2}$, then $D_k \geq M$ with $M := 2(\bar{c} - \eta)^{-2} > 0$.

On the other hand, since $\eta < \lambda_k < \bar{c} - \eta$, then

$$\begin{aligned} 4\lambda_k^2(\|A^T A\|^2 + \|B^T A\|^2) + 2(\alpha_1 + \mu)^2\lambda_k^{-2} &< 4(\bar{c} - \eta)^2(\|A^T A\|^2 + \|B^T A\|^2) + 2(\alpha_1 + \mu)^2\eta^{-2} \\ 4\lambda_k^2(\|A^T B\|^2 + \|B^T B\|^2) + 2(\alpha_2 + \mu')^2\lambda_k^{-2} &< 4(\bar{c} - \eta)^2(\|A^T B\|^2 + \|B^T B\|^2) + 2(\alpha_2 + \mu')^2\eta^{-2} \\ 2\lambda_k^{-2} &< 2\eta^{-2} \end{aligned}$$

considering

$$\begin{aligned} N := \max\{4(\bar{c} - \eta)^2(\|A^T A\|^2 + \|B^T A\|^2) + 2(\alpha_1 + \mu)^2\eta^{-2}, 4(\bar{c} - \eta)^2(\|A^T B\|^2 + \|B^T B\|^2) \\ + 2(\alpha_2 + \mu')^2\eta^{-2}, 2\eta^{-2}\}, \end{aligned}$$

we obtain that $D_k < N$, and therefore

$$0 < M \leq D_k < N. \quad (35)$$

Now, from (33), (34) and (35), for all $k \geq \hat{k} := \max\{\hat{k}, k_3'''\}$, we obtain

$$\begin{aligned} \|\bar{w}^{k+1} - w^*\|^2 &\leq a^2 D_k \{\|\bar{x}^{k+1} - x^k\|^2 + \|\bar{z}^{k+1} - z^k\|^2 + \|p^{k+1} - \bar{y}^{k+1}\|^2 + \|p^{k+1} - y^k\|^2\} \\ &< a^2 N \{\|\bar{x}^{k+1} - x^k\|^2 + \|\bar{z}^{k+1} - z^k\|^2 + \|p^{k+1} - \bar{y}^{k+1}\|^2 + \|p^{k+1} - y^k\|^2\} \end{aligned} \quad (36)$$

furthermore, from (16), for all $k \geq 0$, we obtain

$$\begin{aligned} \hat{H}(w^*, \bar{w}^{k+1}) &\leq \hat{H}(w^*, w^k) - D \{\|\bar{x}^{k+1} - x^k\|^2 + \|\bar{z}^{k+1} - z^k\|^2 + \|p^{k+1} - \bar{y}^{k+1}\|^2 \\ &\quad + \|p^{k+1} - y^k\|^2\} \end{aligned} \quad (37)$$

thus, multiplying by D and $a^2 N$ in the inequalities (36) and (37), respectively and after adding, we obtain

$$D\|\bar{w}^{k+1} - w^*\|^2 + a^2 N \hat{H}(w^*, \bar{w}^{k+1}) \leq a^2 N \hat{H}(w^*, w^k) \quad \forall k \geq \hat{k}. \quad (38)$$

From assumptions, when $\hat{H}(w, w') = (\hat{\alpha}/2)\|w - w'\|^2$.

Defining $\nu := a\sqrt{\hat{\alpha}N}/\sqrt{a^2\hat{\alpha}N + 2D}$, the inequality in (59) reduces to

$$\|\bar{w}^{k+1} - w^*\| \leq \nu \|w^k - w^*\| \quad \forall k \geq \hat{k}. \quad (39)$$

But, from Lemma 4.4, we have

$$\begin{aligned} \|w^{k+1} - \bar{w}^{k+1}\| &\leq \delta_k \|w^{k+1} - w^k\| = \delta_k \|(w^{k+1} - w^*) + (w^* - w^k)\| \\ &\leq \delta_k \|w^{k+1} - w^*\| + \delta_k \|w^k - w^*\|. \end{aligned} \quad (40)$$

Therefore from (39) and (40), we obtain

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

then

$$\|w^{k+1} - w^*\| \leq \theta_k \|w^k - w^*\|, \quad \forall k \geq \hat{k} \quad \text{with} \quad \theta_k = \frac{\nu + \delta_k}{1 - \delta_k}.$$

Since $\delta_k \rightarrow 0$ and

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

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

$$\|w^{k+1} - w^*\| \leq \theta_k \|w^k - w^*\|, \quad \forall k \geq \bar{k}$$

with

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

□

Remark 6 Similarly at the analysis of the method proposed by Chen and Teboulle, Lemma 4.2, which allows us get the estimate (16), is the key in the above analysis.

Remark 7 In Theorem 4.5, the condition

$$\hat{H}(w, w') = (\hat{\alpha}/2) \|w - w'\|^2, \quad (41)$$

is fundamental to obtain the linear rate of convergence. Note that we have a large class of induced proximal distances satisfying this condition. In fact, consider for example

- **Induced proximal distances by some Bregman distance.**

When $\mathbf{C} = \mathbb{R}^n$ and $\mathbf{K} = \mathbb{R}^m$ with $h(x) = \frac{1}{2} \sum_{i=1}^n x_i^2$, $h'(z) = \frac{1}{2} \sum_{i=1}^n z_i^2$, we obtain $H(x, x') := d_h(x, x') = \frac{1}{2} \|x - x'\|^2$ and $H'(z, z') := d'_h(z, z') = \frac{1}{2} \|z - z'\|^2$, see Section 3.2 in [2]. Furthermore, in Section 7, example (a) of [20] was showed that $(d_h, H) \in \mathcal{F}_+(\mathbb{R}^n)$ and $(d'_h, H') \in \mathcal{F}_+(\mathbb{R}^m)$ satisfying the condition (Iviii) of Definition 2.1. Note that in this case the (PMAPD) algorithm is exactly the PCPM algorithm (proposed to solve the problem (CP) with $B = -I$, $b = 0$, see [5]) and (41) holds with $\hat{\alpha} = 1$.

- **Induced proximal distances by the Second-order homogeneous distances.**

Let $\varphi : \mathbb{R} \rightarrow \mathbb{R} \cup \{+\infty\}$ be a closed, convex, proper function such that $\text{dom} \varphi \subset \mathbb{R}_+$ and $\text{dom} \partial \varphi = \mathbb{R}_{++}$. We suppose in addition that φ is $C^2(\mathbb{R}_{++})$, strictly convex, and nonnegative on \mathbb{R}_{++} with $\varphi(1) = \varphi'(1) = 0$. We denote

by Φ the class of such kernels and $\bar{\Phi}$ the subclass of these kernels satisfying

$$\varphi''(1)(1 - 1/t) \leq \varphi'(t) \leq \varphi''(1)(t - 1), \quad \forall t > 0.$$

Let $\varphi(t) = \mu p(t) + \frac{\nu}{2}(t-1)^2$ with $\nu \geq \mu p''(1) > 0$, $p \in \bar{\Phi}$ and let the associated proximal distance be defined by

$$d_\varphi(x, y) = \sum_{j=1}^n y_j^2 \varphi\left(\frac{x_j}{y_j}\right).$$

The use of φ -divergence proximal distances is particularly suitable for handling polyhedral constraints. Let $C = \{x \in \mathbb{R}^n : Ax < b\}$, where A is an (m, n) matrix of full rank m ($m \geq n$). Particularly important cases include $C = \mathbb{R}_{++}^n$ or $C = \{x \in \mathbb{R}_{++}^n : a_i < x_i < b_i \forall i = 1, \dots, n\}$, with $a_i, b_i \in \mathbb{R}$. In Section 7, example (c) of [20] was showed that for $H(x, y) = \bar{\eta} \|x - y\|^2$ with $\bar{\eta} = 2^{-1}(\nu + \mu p''(1))$, we have $(d_\varphi, H) \in \mathcal{F}_+(\mathbb{R}_+^n)$ and analogously we can obtain $(d'_\varphi, H') \in \mathcal{F}_+(\mathbb{R}_+^p)$ satisfying the condition (Iviii) of Definition 2.1. Now, we note that for $\bar{\eta} = 1/2$ (for example when $1/2 \leq \nu < 1$ and $\mu = \frac{1-\nu}{p''(1)}$) the condition (41) holds with $\hat{\alpha} = 1$.

It is interesting to notice that considering the exact version of (PMAPD) Algorithm, one is also able to obtain a linear rate of convergence result but with respect to function \hat{H} . Before, we give the following result.

PROPOSITION 4.6 *Suppose that the assumptions of Theorem 3.2 are satisfied. Let $\{\bar{w}^k\}$ and $\{\bar{p}^k\}$ be sequences generated by the exact version of (PMAPD) algorithm. Let (x^*, z^*) be an optimal solution of (CP), and let y^* be a corresponding Lagrange multiplier, then we have, for each k ,*

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

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

Proof. Since a similar result to Lemma 4.2 can be obtained for the sequences $\{\bar{w}^k\}$ and $\{\bar{p}^k\}$, then the proof is analogous to the proof of Proposition 4.3. \square

THEOREM 4.7 *Let $(d_0, H_0) \in \mathcal{F}_+(\bar{\mathcal{C}})$, $(d'_0, H'_0) \in \mathcal{F}_+(\bar{\mathcal{K}})$ satisfying the condition (Iviii) and $\hat{H}(w, w') < (\hat{\beta}/2) \|w - w'\|^2$, for some $\hat{\beta} > 0$. Let $\{(\bar{x}^k, \bar{z}^k, \bar{y}^k)\}$, $\{\bar{p}^k\}$ the sequences generated by the exact version of (PMAPD) algorithm and suppose that the assumptions of Theorem 3.2 and $(\mathcal{T}_1) - (\mathcal{T}_3)$ hold. Then, $\{\bar{w}^k\} := \{(\bar{x}^k, \bar{z}^k, \bar{y}^k)\}$ converges linearly to the unique optimal solution $w^* := (x^*, z^*, y^*)$, that is, there exists an integer \hat{k} such that*

$$\hat{H}(w^*, \bar{w}^{k+1}) \leq \bar{\theta} \hat{H}(w^*, \bar{w}^k), \quad \forall k \geq \hat{k}$$

with $\bar{\theta} = (a^2 N)/(a^2 N + 2D/\hat{\beta}) < 1$

Proof. From Step 2 of the exact version of (PMAPD) algorithm, we have

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

furthermore,

$$0 \in \partial f(\bar{x}^{k+1}) + A^T \bar{y}^{k+1} - A^T(\bar{y}^{k+1} - \bar{p}^{k+1}) + \lambda_k^{-1}\nabla_1 d(\bar{x}^{k+1}, \bar{x}^k) \quad (42)$$

$$0 = \partial g(\bar{z}^{k+1}) + B^T \bar{y}^{k+1} - B^T(\bar{y}^{k+1} - \bar{p}^{k+1}) + \lambda_k^{-1}\nabla_1 d'(\bar{z}^{k+1}, \bar{z}^k). \quad (43)$$

From Step 4 of the exact version of (PMAPD) algorithm, we have $\bar{y}^{k+1} = \bar{y}^k + \lambda_k(A\bar{x}^{k+1} + B\bar{z}^{k+1} - b)$, then

$$-\lambda_k^{-1}(\bar{y}^{k+1} - \bar{y}^k) = b - A\bar{x}^{k+1} - B\bar{z}^{k+1}. \quad (44)$$

Furthermore, since

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

then from (42), (43), (44) and Definition of S, see (9), we obtain

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

where

$$\pi_k := A^T(\bar{y}^{k+1} - \bar{p}^{k+1}) - \lambda_k^{-1}\nabla_1 d(\bar{x}^{k+1}, \bar{x}^k), \quad (45)$$

$$\sigma_k := B^T(\bar{y}^{k+1} - \bar{p}^{k+1}) - \lambda_k^{-1}\nabla_1 d'(\bar{z}^{k+1}, \bar{z}^k), \quad (46)$$

$$\xi_k := -\lambda_k^{-1}(\bar{y}^{k+1} - \bar{y}^k). \quad (47)$$

From Step 1 and 4 of exact version of (PMAPD) algorithm, $\bar{p}^{k+1} = \bar{y}^k + \lambda_k(A\bar{x}^k + B\bar{z}^k - b)$ and $\bar{y}^{k+1} = \bar{y}^k + \lambda_k(A\bar{x}^{k+1} + B\bar{z}^{k+1} - b)$ then by subtracting, we obtain

$$\bar{y}^{k+1} - \bar{p}^{k+1} = \lambda_k(A(\bar{x}^{k+1} - \bar{x}^k) + B(\bar{z}^{k+1} - \bar{z}^k)). \quad (48)$$

Substituting (48) in (45)-(47), we get

$$\pi_k = \lambda_k A^T(A(\bar{x}^{k+1} - \bar{x}^k) + B(\bar{z}^{k+1} - \bar{z}^k)) - \lambda_k^{-1}\nabla_1 d(\bar{x}^{k+1}, \bar{x}^k) \quad (49)$$

$$\sigma_k = \lambda_k B^T(A(\bar{x}^{k+1} - \bar{x}^k) + B(\bar{z}^{k+1} - \bar{z}^k)) - \lambda_k^{-1}\nabla_1 d'(\bar{z}^{k+1}, \bar{z}^k) \quad (50)$$

$$\xi_k = -\lambda_k^{-1}(\bar{y}^{k+1} - \bar{y}^k). \quad (51)$$

From assumption (\mathcal{T}_3), $\nabla_1 d_0(\cdot, u)$ is locally Lipschitz continuous, and since $\|\bar{x}^{k+1} - \bar{x}^k\| \rightarrow 0$ with $\bar{x}^k \in \bar{\mathcal{C}}$ then, there exists $\alpha_1 > 0$, $k'_1, k''_1 \in \mathbb{N}$ such that

$$\|\nabla_1 d_0(\bar{x}^{k+1}, \bar{x}^k) - \nabla_1 d_0(\bar{x}^k, \bar{x}^k)\| \leq \alpha_1 \|\bar{x}^{k+1} - \bar{x}^k\|, \quad \forall k \geq k_1''' := \max\{k'_1, k''_1\},$$

therefore,

$$\begin{aligned}
\|\nabla_1 d(\bar{x}^{k+1}, \bar{x}^k)\| &= \|\nabla_1 d_0(\bar{x}^{k+1}, \bar{x}^k) + \mu(\bar{x}^{k+1} - \bar{x}^k)\| \\
&\leq \|\nabla_1 d_0(\bar{x}^{k+1}, \bar{x}^k) - \nabla_1 d_0(\bar{x}^k, \bar{x}^k)\| + \mu\|\bar{x}^{k+1} - \bar{x}^k\| \quad (52) \\
&\leq (\alpha_1 + \mu)\|\bar{x}^{k+1} - \bar{x}^k\|, \quad \forall k \geq k_1''' := \max\{k_1', k_1''\}.
\end{aligned}$$

Analogously, from Assumption (\mathcal{T}_3) and since $\|\bar{z}^{k+1} - z^\infty\| \rightarrow 0$ with $z^\infty \in \bar{K}$ then, there exists $\alpha_2 > 0, k_2', k_2'' \in \mathcal{N}$ such that

$$\|\nabla_1 d'_0(\bar{z}^{k+1}, \bar{z}^k) - \nabla_1 d'_0(\bar{z}^k, \bar{z}^k)\| \leq \alpha_2\|\bar{z}^{k+1} - \bar{z}^k\|, \quad \forall k \geq k_2''' := \max\{k_2', k_2''\},$$

thus,

$$\|\nabla_1 d'(\bar{z}^{k+1}, \bar{z}^k)\| \leq (\alpha_2 + \mu')\|\bar{z}^{k+1} - \bar{z}^k\|, \quad \forall k \geq k_2''' := \max\{k_2', k_2''\}. \quad (53)$$

Then, from (52) and (53) in (49)-(51) with $\eta < \lambda_k < \bar{c} - \eta$, it is clear that,

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

That is, there exists \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 $0 \in S(x^*, z^*, y^*)$ e $(\pi_k, \sigma_k, \xi_k) \in S(\bar{x}^{k+1}, \bar{z}^{k+1}, \bar{y}^{k+1})$, with the choice

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

we obtain

$$\|\bar{w}^{k+1} - w^*\| \leq a\|(\pi_k, \sigma_k, \xi_k)\| \quad \forall k \geq \ddot{k}. \quad (54)$$

We note that if $a = 0$ then $\bar{w}^{k+1} = w^* \forall k \geq \ddot{k}$, then the (PMAPD) algorithm would not have asymptotic convergence, therefore we consider $a > 0$.

We will estimate the right side of inequality (54). Using the definition of (π_k, σ_k, ξ_k) and the inequality $(r + q)^2 \leq 2(r^2 + q^2)$, we obtain

$$\begin{aligned}
\|\pi_k\|^2 &\leq 4\lambda_k^2[\|A^T A\|^2\|\bar{x}^{k+1} - \bar{x}^k\|^2 + \|A^T B\|^2\|\bar{z}^{k+1} - \bar{z}^k\|^2] + 2\lambda_k^{-2}\|\nabla_1 d(\bar{x}^{k+1}, \bar{x}^k)\|^2 \\
\|\sigma_k\|^2 &\leq 4\lambda_k^2[\|B^T A\|^2\|\bar{x}^{k+1} - \bar{x}^k\|^2 + \|B^T B\|^2\|\bar{z}^{k+1} - \bar{z}^k\|^2] + 2\lambda_k^{-2}\|\nabla_1 d'(\bar{z}^{k+1}, \bar{z}^k)\|^2 \\
\|\xi_k\|^2 &\leq 2\lambda_k^{-2}(\|\bar{y}^{k+1} - \bar{p}^{k+1}\|^2 + \|\bar{p}^{k+1} - \bar{y}^k\|^2),
\end{aligned}$$

thus, from (52), we obtain

$$\begin{aligned}
\|\pi_k\|^2 &\leq (4\lambda_k^2\|A^T A\|^2 + 2(\alpha_1 + \mu)^2\lambda_k^{-2})\|\bar{x}^{k+1} - \bar{x}^k\|^2 + 4\lambda_k^2\|A^T B\|^2\|\bar{z}^{k+1} - \bar{z}^k\|^2, \quad \forall k \geq k_1''' \\
\|\sigma_k\|^2 &\leq 4\lambda_k^2\|B^T A\|^2\|\bar{x}^{k+1} - \bar{x}^k\|^2 + (2(\alpha_2 + \mu')^2\lambda_k^{-2} + 4\lambda_k^2\|B^T B\|^2)\|\bar{z}^{k+1} - \bar{z}^k\|^2, \quad \forall k \geq k_2''' \\
\|\xi_k\|^2 &\leq 2\lambda_k^{-2}(\|\bar{p}^{k+1} - \bar{y}^{k+1}\|^2 + \|\bar{p}^{k+1} - \bar{y}^k\|^2).
\end{aligned}$$

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

$$\begin{aligned} \|(\pi_k, \sigma_k, \xi_k)\| &= \|\pi_k\|^2 + \|\sigma_k\|^2 + \|\xi_k\|^2 \\ &\leq D_k \{\|\bar{x}^{k+1} - \bar{x}^k\|^2 + \|\bar{z}^{k+1} - \bar{z}^k\|^2 + \|\bar{p}^{k+1} - \bar{y}^{k+1}\|^2 + \|\bar{p}^{k+1} - \bar{y}^k\|^2\} \end{aligned} \quad (55)$$

where

$$\begin{aligned} D_k &:= \max\{4\lambda_k^2(\|A^T A\|^2 + \|B^T A\|^2) + 2(\alpha_1 + \mu)^2 \lambda_k^{-2}; 4\lambda_k^2(\|A^T B\|^2 + \|B^T B\|^2) \\ &\quad + 2(\alpha_2 + \mu')^2 \lambda_k^{-2}; 2\lambda_k^{-2}\}, \end{aligned}$$

From the proof of Theorem 4.5, we have

$$0 < M \leq D_k < N. \quad (56)$$

where $M = 2(\bar{c} - \eta)^{-2} > 0$ and

$$\begin{aligned} N &:= \max\{4(\bar{c} - \eta)^2(\|A^T A\|^2 + \|B^T A\|^2) + 2(\alpha_1 + \mu)^2 \eta^{-2}, 4(\bar{c} - \eta)^2(\|A^T B\|^2 + \|B^T B\|^2) \\ &\quad + 2(\alpha_2 + \mu')^2 \eta^{-2}, 2\eta^{-2}\}, \end{aligned}$$

Now, from (54), (55) and (56), for all $k \geq \hat{k} := \max\{\hat{k}, k_3'''\}$, we obtain

$$\|\bar{w}^{k+1} - w^*\|^2 < a^2 N \{\|\bar{x}^{k+1} - \bar{x}^k\|^2 + \|\bar{z}^{k+1} - \bar{z}^k\|^2 + \|\bar{p}^{k+1} - \bar{y}^{k+1}\|^2 + \|\bar{p}^{k+1} - \bar{y}^k\|^2\} \quad (57)$$

furthermore, from Proposition 4.6, for all $k \geq 0$, we have

$$\begin{aligned} \hat{H}(w^*, \bar{w}^{k+1}) &\leq \hat{H}(w^*, \bar{w}^k) - D \{\|\bar{x}^{k+1} - \bar{x}^k\|^2 + \|\bar{z}^{k+1} - \bar{z}^k\|^2 + \|\bar{p}^{k+1} - \bar{y}^{k+1}\|^2 \\ &\quad + \|\bar{p}^{k+1} - \bar{y}^k\|^2\} \end{aligned} \quad (58)$$

thus, multiplying by D and $a^2 N$ in the inequalities (57) and (58), respectively and after adding, we obtain

$$D\|\bar{w}^{k+1} - w^*\|^2 + a^2 N \hat{H}(w^*, \bar{w}^{k+1}) \leq a^2 N \hat{H}(w^*, \bar{w}^k) \quad \forall k \geq \hat{k}. \quad (59)$$

From assumptions, when $\hat{H}(w, w') < (\hat{\beta}/2)\|w - w'\|^2$.

Defining $\bar{\theta} := \frac{a^2 N}{a^2 N + 2D/\hat{\beta}}$, the inequality in (59) reduces to

$$\hat{H}(w^*, \bar{w}^{k+1}) \leq \bar{\theta} \hat{H}(w^*, \bar{w}^k) \quad \forall k \geq \hat{k}.$$

□

Remark 8 In Theorem 4.7, the condition

$$\hat{H}(w, w') < (\hat{\beta}/2)\|w - w'\|^2, \quad (60)$$

is fundamental to obtain the linear rate of convergence. Note also that, we have a broad class of induced proximal distances satisfying this condition. In fact, consider the following cases

- In addition to the examples given in Remark 7 a broader class of induced proximal distances by the Second-order homogeneous proximal dis-

tances satisfy this condition, because we have $H(x, y) = \bar{\eta} \|x - y\|^2$ with $\bar{\eta} = 2^{-1}(\nu + \mu p''(1))$ for any $\nu \geq \mu p''(1) > 0$, then we can obtain (60) taking $\hat{\beta}$ sufficiently large.

- Other class of induced proximal distances that satisfy the condition (41) are the induced proximal distances by double regularization distances, introduced by Silva and Eckstein [21]. In [4], Remark 4.2, Burachik and Dutta showed that double regularization distances, defined by

$$\hat{d}(x, y) = \sum_{i=1}^n d_i(x, y) + \mu \|x - y\|^2$$

where $d_i(x, y)$ satisfy conditions given in [4], verify the definition of proximal distances. Moreover, if $\mu > 1$, then it is proved in [21], Lemma 3.3 and 3.4, that $\langle c - b, \nabla_1 \hat{d}(b, a) \rangle \leq H(c, a) - H(c, b) - \gamma H(b, a)$ with $H(x, y) = \frac{\mu+1}{2} \|x - y\|^2$ and $\gamma := \frac{\mu-1}{2}$. From Remark 4.2 of [4], it is not difficult to prove that $(d_\varphi, H) \in \mathcal{F}_+(\mathbb{R}_+^n)$ as well as, it is clear that the condition (Iviii) of Definition 2.1 is satisfied. Therefore, we can obtain (60) taking $\hat{\beta}$ sufficiently large.

5. Some Discussions

- (1) Recent progress in splitting methods exhibits explicit linear convergence rate under standard assumptions, see for example the works of Auslender and Teboulle [2], Goldfarb et al. [9] and Chen and Teboulle [5]. Of fact, Auslender and Teboulle, [2], developed an interior gradient method for convex conic problems using proximal distances, and were able to exhibit an $O(1/k^2)$ global convergence rate estimate for function values motivated from the work of Nesterov [16]. Goldfarb et al., [9], 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. On the other hand, Chen and Teboulle, [5], developed a proximal decomposition method for convex minimization problems and showed that the iterations of their algorithm converge linearly.

We should observe that the convergence analysis of Auslender and Teboulle [2], and Goldfarb et al. [9] are with respect to the objective function values and not with respect to the sequence of points generated by the algorithms. Motivated from the work of Chen and Teboulle [5], in this paper we prove that it is possible, under mild conditions, to obtain the same linear rate of convergence of the (PMAPD) for a broad class of proximal distances.

- (2) In this paper we assume three general assumptions given by $(\mathcal{T}_1) - (\mathcal{T}_3)$, see Section 4. The first and second are standard assumptions and were suggested by Rockafellar (see [18], p. 100), to derive the rate of convergence of the proximal method of multipliers. It was used in the work of Chen and Teboulle [5] and it also is used here to derive the rate of convergence for (PMAPD) algorithm. The other one is with respect to the proximal distance, which is satisfied by φ -divergence proximal distances, second order homogeneous proximal distances and some Bregman distances, when the optimal point

(x^*, z^*) is an interior point of the set of constraints, see Remark 3.

Conclusions and future researches

In this paper, we show that the inexact proximal multiplier method using regularized proximal distances called (PMAPD), proposed by us in [20], for solving convex minimization problems with a separable structure, converges linearly to an optimal solution of the problem in two cases.

In the first case, the condition $\hat{H}(w, w') = (\hat{\alpha}/2)\|w - w'\|^2$ is fundamental to obtain the linear rate of convergence to the inexact version of (PMAPD). This condition is satisfied by a broad class of induced proximal distances, as for example: a) induced proximal distances by the Second-order homogeneous proximal

distances $d_\varphi(x, y) = \sum_{j=1}^n y_j^2 \varphi(x_j/y_j)$, where $\varphi(t) = \mu p(t) + \frac{\nu}{2}(t-1)^2$, $p \in \Phi_2$

with $\nu \geq \mu p''(1) > 0$ (to satisfy the condition to $\hat{H}(\cdot, \cdot)$, here should be taken $\nu + \mu p''(1) = 1$). b) induced proximal distances by Bregman distances with the kernel function $h(x) = \frac{1}{2} \sum_{i=1}^n x_i^2$.

In the second case, the condition $\hat{H}(w, w') < (\hat{\beta}/2)\|w - w'\|^2$ is fundamental to obtain the rate of convergence to the exact version of (PMAPD), and this is satisfied by a large class of induced proximal distances, as for example, induced proximal distances by the Second-order homogeneous proximal distances (here may be taken $\nu + \mu p''(1) > 0$ arbitrarily) and induced proximal distances by double regularization distances.

A future research may include to find a variant of the (PMAPD) method to obtain superlinear rate of convergence, as also, to perform computational experiments and comparison with other methods in the literature. The extension of the (PMAPD) to solve more general problems, such as, quasiconvex optimization and variational inequalities are desirable.

Acknowledgements

This research, for the first author, was supported by CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico). Part of the research of the second author was supported by the PosDoctoral Scholarship CAPES-FAPERJ Edital PAPD-2011. The authors wish to thank at Systems Engineering and Computer Science Program, COPPE, Federal University of Rio de Janeiro for its backing.

References

- [1] A. Auslender, M. Teboulle, *Entropic proximal decomposition methods for convex programs and variational inequalities*, Math. Program., Ser. A, vol. 91, (2001), pp. 33-47.
- [2] A. Auslender, M. Teboulle, *Interior gradient and proximal methods for convex and conic optimization*, SIAM J. Optim., vol. 16, 3, (2006), pp. 697-725.
- [3] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and distributed computation, Numerical Methods*, Prentice Hall, (1989), Englewood Cliffs, New Jersey.
- [4] R. Burachik and J. Dutta, *Inexact proximal point methods for variational inequality problems*, SIAM J. Optim. 20, (2010), pp. 2653-2678.
- [5] G. Chen and M. Teboulle, *A proximal-based decomposition method for convex minimization problems*, Mathematical Programming, Serie A, 1, vol. 64, (1994), pp. 81-101.

- [6] J. Chifflet, P. Mahey and V. Reiner, *Proximal decomposition for multicommodity flow problems with convex costs*, Telecommun. Syst. vol. 3, (1994), pp. 1-10.
- [7] D. Gabay, *Chapter ix applications of the method of multipliers to variational inequalities*, Studies in mathematics and its applications, vol. 15, (1983), pp. 299-331.
- [8] D. Gabay and B. Mercier, *A Dual algorithm for the solution of nonlinear variational problems via finite element approximations*, Computers and Mathematics with Applications, vol. 2, (1976), pp. 17-40.
- [9] D. Goldfarb, S. Ma and K. Scheinberg, *Fast alternating linearization methods for minimizing the sum of two convex functions*, Math. Program. 141, (2013), no. 1-2, Ser. A, pp. 349-382.
- [10] D. Han, X. Yuan and W. Zhang, *An augmented Lagrangian based parallel splitting method for separable convex minimization with applications to image processing*, Mathematics of Computation, (2014), vol. 83, no 289, pp. 2263-2291.
- [11] B. He, M. Tao and X. Yuan, *A splitting method for separable convex programming*, IMA Journal of Numerical Analysis, 35(1), (2015), pp. 394-426.
- [12] M. Kyono and M. Fukushima, *Nonlinear proximal decomposition method for convex programming*, JOTA, vol. 106, N 2, (2000), pp. 357-372.
- [13] A. Lenoir, *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.
- [14] P. L. Lions and B. Mercier, *Splitting algorithms for the sum of two nonlinear operators*, SIAM Journal on Numerical Analysis, vol. 16, (1979), pp. 964-979.
- [15] P. Mahey, A. Ouorou, L. LeBlanc and J. Chifflet, *A new proximal decomposition algorithm for routing in telecommunication networks*, Networks, vol. 31, (1997), pp. 227-238.
- [16] Y. Nesterov, *On an approach to the construction of optimal methods of minimization of smooth convex functions*, Ekonom. i Mat. Metody, 24, (1988), pp. 509-517.
- [17] R. T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton New Jersey, 1970.
- [18] R. T. Rockafellar, *Augmented Lagrangians and Applications of the Proximal Point Algorithm in Convex Programming*, Mathematics of Operations Research, vol. 1, (1976), pp. 97-116.
- [19] R. T. Rockafellar, *Monotone operators and the proximal point algorithm*, SIAM Journal on Control and Optimization, vol. 14, N. 5, (1976), pp. 877-898.
- [20] O. Sarmiento, E. A. Papa Quiroz and P. R. Oliveira, *A proximal multiplier method for separable convex minimization*, Optimization, 2015, (In Press).
- [21] P. J. S. Silva and J. Eckstein, *Double-regularization proximal methods, with complementarity applications*, Comput. Optim. Appl., 33, (2006), pp. 115-156.
- [22] J. E. Spingarn, *Partial Inverse of a Monotone Operator*, Applied Mathematics and Optimization 10, (1983), pp. 247-265.
- [23] J. E. Spingarn, *Applications of the method of partial inverse to convex programming decomposition*, Math. Program. vol. 32, (1985), pp. 199-223.