Iteration-Complexity of a Linearized Proximal Multiblock ADMM Class for Linearly Constrained Nonconvex Optimization Problems

Jefferson G. Melo * Renato D.C. Monteiro †

April 14, 2017

Abstract

This paper analyzes the iteration-complexity of a class of linearized proximal multiblock alternating direction method of multipliers (ADMM) for solving linearly constrained nonconvex optimization problems. The subproblems of the linearized ADMM are obtained by partially or fully linearizing the augmented Lagrangian with respect to the corresponding minimizing block variable. The derived complexity bounds do not depend on the specific forms of the actual linearizations but only on some Lipschitz constants which quantify the approximation errors. Iteration-complexity is then established by showing that the linearized ADMM class is a subclass of a general non-Euclidean ADMM for which a general iteration-complexity analysis is also obtained. Both ADMM classes allow the choice of a relaxation parameter in the interval $(0, 2)$ as opposed to being equal to one as in many of the previous papers on this topic.

2000 Mathematics Subject Classification: 47J22, 49M27, 90C25, 90C26, 90C30, 90C60, 65K10.

Key words: Multiblock ADMM, nonconvex program, pointwise iteration-complexity, first-order methods.

1 Introduction

This paper considers the following linearly constrained problem

$$\min \left\{ \sum_{i=1}^{p} f_i(x_i) : \sum_{i=1}^{p} A_i x_i = b,\ x_i \in \mathbb{R}^{n_i}, i = 1, \cdots, p \right\} \tag{1}$$

where $f_i : \mathbb{R}^{n_i} \to (-\infty, \infty]$, $i = 1, \cdots, p$, are proper lower semicontinuous functions, $A_i \in \mathbb{R}^{d\times n_i}$, $i = 1, \cdots, p$, and $b \in \mathbb{R}^d$.

Optimization problems such as (1) appear in many important applications such as distributed matrix factorization, distributed clustering, sparse zero variance discriminant analysis, tensor decomposition, matrix completion, and asset allocation (see, e.g., [1, 6, 23, 35, 36, 38]). Recently, some

*Instituto de Matemática e Estatística, Universidade Federal de Goiás, Campus II- Caixa Postal 131, CEP 74001-970, Goiânia-GO, Brazil. (E-mail: jefferson@ufg.br). The work of this author was partially supported by CNPq Grants 406250/2013-8, 444134/2014-0 and 406975/2016-7.

†School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA, 30332-0205. (email: monteiro@isye.gatech.edu). The work of this author was partially supported by NSF Grant CMMI-1300221.

variants of the alternating direction method of multipliers (ADMM) have been successfully applied to solve some instances of the previous problem despite the lack of convexity.

An important ADMM class for solving (1) is the proximal ADMM which recursively computes a sequence $\{(x_1^k, \cdots, x_p^k, \lambda^k)\}$ as

$$x_i^k = \operatorname{argmin}_{x_i} \left\{ \mathcal{L}_\beta(x_1^k, \cdots, x_{i-1}^k, x_i, x_{i+1}^{k-1}, \cdots, x_p^{k-1}, \lambda^{k-1}) + \frac{1}{2}\|x_i - x_i^{k-1}\|_{H_i}^2 \right\} \tag{2}$$

for every $i = 1, \cdots, p$, and

$$\lambda^k = \lambda^{k-1} - \theta\beta \left(\sum_{i=1}^{p} A_i x_i^k - b \right) \tag{3}$$

where $\beta > 0$ is a penalty parameter, $\theta > 0$ is an relaxation parameter, $H_i \in \mathbb{R}^{n_i \times n_i}$, $i = 1, \cdots, p$, are symmetric and positive semidefinite matrices, and

$$\mathcal{L}_\beta(x_1, \cdots, x_p, \lambda) := \sum_{i=1}^{p} [f_i(x_i) - \langle \lambda, A_i x_i - b \rangle] + \frac{\beta}{2} \left\| \sum_{i=1}^{p} A_i x_i - b \right\|^2 \tag{4}$$

is the augmented Lagrangian function for problem (1). The above class of algorithms and some of its variants have been studied in the context of convex optimization (see for example [27, 3, 9, 12, 16, 18]) and nonconvex optimization (see for example [13, 20, 21, 10, 32, 34, 37]).

Our goal in this paper is to study an extension of the above proximal ADMM class whose subproblems are obtained by partially or entirely linearizing the partial augmented Lagrangian $x_i \mapsto \mathcal{L}_\beta(x_1^k, \cdots, x_{i-1}^k, x_i, x_{i+1}^{k-1}, \cdots, x_p^{k-1}, \lambda^{k-1})$ in (2) so as to make them easily solvable. Our analysis is quite general in that it applies to a large class of linearizations without the need to know the specific linearizations performed. Under the assumption that A_p is full row rank (actually, a weaker condition on A_p is assumed) and $f_p : \mathbb{R}^{n_p} \to \mathbb{R}$ is a differentiable function whose gradient is Lipschitz continuous, complexity bounds for the linearized ADMM class are obtained which, relative to the approximation, depend on the Lipschitz constants of the gradients of the linearized components of the partial augmented Lagrangian functions. By considering an extended notion of subdifferential for nonconvex functions (see for example [29, 31]), we establish an $\mathcal{O}(\rho^{-2})$-pointwise iteration-complexity for the linearized ADMM to obtain $(x_1, \cdots, x_p, \lambda, r_1, \cdots, r_{p-1})$ satisfying

$$r_i \in \partial f_i(x_1, \cdots, x_p) - A_i^* \lambda, \quad i = 1, \cdots, p-1, \tag{5}$$

$$\max \left\{ \left\| \sum_{i=1}^{p} A_i x_i - b \right\|, \|r_1\|, \cdots, \|r_{p-1}\|, \left\| \nabla f_p(x_p) - A_p^* \lambda \right\| \right\} \le \rho. \tag{6}$$

The above result applies for any arbitrary choice of $\theta \in (0, 2)$ as long as the Lipschitz constant of the gradient of the linearized component of the p-th partial augmented Lagrangian is $o(\beta)$. On the other hand, when the latter condition does not hold, it is shown that the complexity result above also holds if θ is chosen sufficiently small. The complexity analysis of the linearized ADMM is derived by considering a more general class of non-Euclidean ADMMs whose subproblems consist of replacing the proximal term in (2) with a suitable Bregman distance proximal term. By showing that the first class is a special instance of the latter one and by studying the complexity of the latter one, we will be able to obtain the complexity analysis of the linearized ADMM. We believe that this indirect approach gives a meaningful illustration of the usefulness of non-Euclidean ADMMs.

Previous most related works. We split our discussion into the following two cases: (i) the functions $f_1, \ldots, f_p$ are convex; and (ii) $f_1, \ldots, f_p$ and A_p satisfy the conditions following problem (1) and the assumption mentioned in the previous paragraph.

Case (i): We first consider case (i) with two blocks, i.e., $p = 2$. The ADMM class (2) with $p = 2$ and $H_i = 0$, $i = 1, 2$, corresponds to the standard ADMM which was introduced in [7, 8]. Complexity analysis for the latter method was first carried out in [28] where its pointwise (resp., ergodic) iteration-complexity is obtained for any $\theta \in (0, 1)$ (resp., $\theta \in (0, 1]$). Subsequently, several papers have obtained pointwise and/or ergodic complexity bounds for more general subclasses of the proximal ADMM (see for example [3, 9, 12, 16, 18]). For arbitrary choices of penalty parameter $\beta > 0$ and positive semidefinite matrices H_1 and H_2, the most general results in these papers establish pointwise (resp., ergodic) complexity bounds for any $\theta \in (0, (1 + \sqrt{5})/2)$ (resp., $\theta \in (0, (1 + \sqrt{5})/2]$). Moreover, iteration-complexity results for other ADMM classes are studied for example in [4, 5, 11, 14, 24, 30] and references therein.

We now consider case (i) with multiple blocks, i.e., $p > 2$. The proximal ADMM (2) with $H_i = 0$, $i = 1, \cdots, p$, is the multiblock version of the standard ADMM, which may not converge as shown in [2]. Convergence of the latter method has been established under the assumption that all (or, all but one) functions f_i, $i = 1, \cdots, p$, are strongly convex and β lies in a certain range (see for instance [15, 19, 25, 26]). Variants of the multiblock proximal ADMM with established iteration-complexity bound have been proposed in the literature (see for example [17, 27] and references cited therein).

Case (ii): In contrast to case (i), we discuss the two-block and multiblock versions of (2) at the same time since their approach and analysis are quite similar. Recently, there have been a lot of interest on the study of ADMM variants for nonconvex problems (see, e.g., [13, 20, 21, 22, 32, 33, 34, 37, 10]). Papers [13, 22, 32, 33, 34, 37] establish convergence of the generated sequence to a stationary point of (1) under conditions which guarantee that a certain potential function associated with the augmented Lagrangian (4) satisfies the Kurdyka-Lojasiewicz property. However, these papers do not study the iteration complexity of the proximal ADMM although their theoretical analysis are generally half-way towards accomplishing such goal. Paper [20] analyzes the convergence of variants of the ADMM for solving nonconvex consensus and sharing problems and establishes the iteration complexity of ADMM for the consensus problem. Paper [21] studies the iteration-complexity of two linearized variants of the multiblock proximal ADMM applied to a more general problem than (1) where a coupling term is also present in its objective function. It is worth mentioning though that both linearizations considered are with respect to the p-th block and, in contrast to this paper, linearizations with respect to the other blocks are not discussed. Paper [10] studies the iteration-complexity of the subclass of the proximal ADMM in which $p = 2$, $H_2 = \tau I$ for $\tau \geq 0$ sufficiently large, and the relaxation parameter θ is arbitrarily chosen in the interval $(0, 2)$.

Our paper is organized as follows. Subsection 1.1 contains some notation and basic results used in the paper. Section 2 describes our assumptions, introduces the linearized proximal ADMM class and states its corresponding convergence rate result (Theorem 2.3). Section 3 contains two subsections. Subsection 3.1 presents a non-Euclidean proximal ADMM and states its main convergence rate result (Theorem 3.3). Subsection 3.2 provides the proof of Theorem 2.3. Section 4 is dedicated to the proof of Theorem 3.3. The appendix contains an auxiliary lemma and the proof of a technical result.

1.1 Notation and basic results

The domain of a function $f : \mathbb{R}^s \to (-\infty, \infty]$ is the set $\operatorname{dom} f := \{x \in \mathbb{R}^s : f(x) < +\infty\}$. Moreover, f is said to be proper if $f(x) < \infty$ for some $x \in \mathbb{R}^{n_2}$.

Lemma 1.1. *Let $S \in \mathbb{R}^{n\times p}$ be a non-zero matrix and let σ_S^+ denote the smallest positive eigenvalue of SS^*. Then, for every $u \in \mathbb{R}^p$, there holds*

$$\|\mathcal{P}_{S^*}(u)\| \leq \frac{1}{\sqrt{\sigma_S^+}}\|Su\|.$$

We next recall some definitions and results of subdifferential calculus [29, 31].

Definition 1.2. *Let $h : \mathbb{R}^s \to (-\infty, \infty]$ be a proper lower semi-continuous function.*

(i) The Fréchet subdifferential of h at $x \in \operatorname{dom} h$, denoted by $\hat{\partial} h(x)$, is the set of all elements $u \in \mathbb{R}^s$ satisfying

$$\liminf_{y\neq x \; y\to x} \frac{h(y) - h(x) - \langle u, y - x\rangle}{\|y - x\|} \geq 0.$$

When $x \notin \operatorname{dom} h$, we set $\hat{\partial} h(x) = \emptyset$.

(ii) The limiting subdifferential of h at $x \in dom\, h$, denoted by $\partial h(x)$, is defined as

$$\partial h(x) = \{u \in \mathbb{R}^s : \exists\, x^k \to x, h(x^k) \to h(x), u^k \in \hat{\partial} h(x^k), \text{ with } u^k \to u\}.$$

(iii) A critical (or stationary) point of h is a point $x \in \operatorname{dom} h$ satisfying $0 \in \partial h(x)$.

The following result gives some properties of the limiting subdifferential.

Proposition 1.3. *Let $h : \mathbb{R}^s \to (-\infty, \infty]$ be a proper lower semi-continuous function.*

(a) If $x \in R^s$ is a local minimizer of h, then $0 \in \partial h(x)$;

(b) If $p : \mathbb{R}^s \to \mathbb{R}$ is a continuously differentiable function, then $\partial(h + p)(x) = \partial h(x) + \nabla p(x)$.

2 Linearized proximal ADMM and its convergence rate

For the sake of simplicity, we describe and analyze the ADMM variants discussed in this paper in the context of problem (1) with $p = 3$. Note that this clearly applies to the context in which $p \leq 2$. The generalization of our analysis to the context in which $p > 3$ is straightforward and follows by using similar arguments. Hence, hereafter, we consider the linearly constrained problem

$$\min\{f_1(x_1) + f_2(x_2) + g(y) : A_1x_1 + A_2x_2 + By = b,\ x_1 \in \mathbb{R}^{n_1}, x_2 \in \mathbb{R}^{n_2}, y \in \mathbb{R}^q\} \tag{7}$$

where $f_1 : \mathbb{R}^{n_1} \to (-\infty, \infty]$, $f_2 : \mathbb{R}^{n_2} \to (-\infty, \infty]$ and $g : \mathbb{R}^q \to (-\infty, \infty]$ are proper lower semicontinuous functions, $A_1 \in \mathbb{R}^{d\times n_1}$, $A_2 \in \mathbb{R}^{d\times n_2}$, $B \in \mathbb{R}^{d\times q}$ and $b \in \mathbb{R}^d$. Note that the above formulation of (1) with $p = 3$ replaces the function f_3 and matrix A_3 by g and B, respectively, in order to clearly distinguish the third block from the first two blocks.

This section describes the assumptions made on problem (7), states a linearized proximal ADMM for solving (7) and states its corresponding convergence rate result (Theorem 2.3). The proof of Theorem 2.3 uses the fact that the linearized proximal ADMM is a special case of a non-Euclidean proximal ADMM whose convergence analysis is studied in Subsection 3.1 and Section 4. Using this fact, the proof of Theorem 2.3 is then given in Subsection 3.2.

We starting by recalling the definition of critical points of (7).

Definition 2.1. *A quadruple* $(x_1^*, x_2^*, y^*, \lambda^*) \in \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \times \mathbb{R}^q \times \mathbb{R}^d$ *is a critical point of problem* (7) *if*

$$0 \in \partial f_1(x_1^*) - A_1^*\lambda^*, \quad 0 \in \partial f_2(x_2^*) - A_2^*\lambda^*, \quad 0 = \nabla g(y^*) - B^*\lambda^*, \quad 0 = A_1x_1^* + A_2x_2^* + By^* - b.$$

Under some mild conditions, it can be shown that if (x_1^*, x_2^*, y^*) is a global minimum of (7), then there exists λ^* such that $(x_1^*, x_2^*, y^*, \lambda^*)$ is a critical point of (7).

The augmented Lagrangian associated with problem (7) and with penalty parameter $\beta > 0$ is defined as

$$\mathcal{L}_\beta(x_1, x_2, y, \lambda) := f_1(x_1) + f_2(x_2) + g(y) - \langle \lambda, A_1x_1 + A_2x_2 + By - b\rangle + \frac{\beta}{2}\|A_1x_1 + A_2x_2 + By - b\|^2. \quad (8)$$

We assume that problem (7) satisfies the following set of conditions:

(A0) The functions f_1 and f_2 are proper lower semicontinuous;

(A1) $B \neq 0$ and $\mathrm{Im}(B) \supset \{b\} \cup \mathrm{Im}(A_1) \cup \mathrm{Im}(A_2)$;

(A2) $g : \mathbb{R}^q \to \mathbb{R}$ is differentiable everywhere on $\mathbb{R}^q$ and there exists $L_g \geq 0$ such that

$$\|\mathcal{P}_{B^*}(\nabla g(y')) - \mathcal{P}_{B^*}(\nabla g(y))\| \leq L_g\|y' - y\| \quad \forall y, y' \in \mathbb{R}^q;$$

(A3) there exists $\mu_g \in \mathbb{R}$ such that the function $g(\cdot) - \mu_g\|\cdot\|^2/2$ is convex, or equivalently,

$$g(y') - g(y) - \langle \nabla g(y), y' - y\rangle \geq \frac{\mu_g}{2}\|y' - y\|^2 \quad \forall y, y' \in \mathbb{R}^q;$$

(A4) there exists $\bar{\beta} \geq 0$ such that $v(\bar{\beta}) > -\infty$ where

$$v(\beta) := \inf_{(x_1, x_2, y)} \left\{ f_1(x_1) + f_2(x_2) + g(y) + \frac{\beta}{2}\|A_1x_1 + A_2x_2 + By - b\|^2 \right\} \quad \forall \beta \in \mathbb{R}.$$

Some comments about the above assumptions are in order. First, due to the generality of (**A0**), problem (7) may include an extra constraint of the form $x_i \in X_i$ for some $i \in \{1, 2\}$ where X_i is a closed set since this constraint can be incorporated into f_i by adding to it the indicator function of X_i. Second, (**A1**) implies that for every $(x_1, x_2) \in \mathbb{R}^{n_1} \times \mathbb{R}^{n_2}$, there exists $y \in \mathbb{R}^q$ such that (x_1, x_2, y) satisfies the (linear) constraint of (7). Moreover, the extra condition that $B \neq 0$ is very mild since otherwise (7) would be much simpler to solve. Third, if $\nabla g(\cdot)$ is L-Lipschitz continuous, then (**A2**) with $L_g = L$ and (**A3**) with $\mu_g = -L$ obviously hold. However, conditions (**A2**) and (**A3**) combined are generally weaker than the condition that $\nabla g(\cdot)$ be L-Lipschitz continuous.

Next we introduce a general way of describing linearizations of the augmented Lagrangian function in a unified manner. We believe that the definition below, although somewhat difficult to digest at first, considerably simplifies the notation in our presentation and treats different linearizations in a unified manner.

Definition 2.2. *Given a proper lower semi-continuous function $\phi : \mathbb{R}^s \to (-\infty, \infty]$, a pair of functions (ϕ_l, ϕ_n) is said to be a $\bar{\tau}$–Lipschitz decomposition of ϕ if the following conditions hold:*

(a) $\phi_l : \mathbb{R}^s \to \mathbb{R}$ *is differentiable on* $\mathrm{co}(\mathrm{dom}\,\phi)$*;*

(b) $\nabla \phi_l$ *is $\bar{\tau}$-Lipchitz continuous on* $\mathrm{co}(\mathrm{dom}\,\phi)$*;*

(c) $\phi_n : \mathbb{R}^s \to (-\infty, \infty]$ *is lower semi-continuous and* $\phi_n(x) = \phi(x) - \phi_l(x)$ *for every* $x \in \mathrm{dom}\,\phi$.

Moreover, for a given $z \in \mathrm{dom}\,\phi$, the sum $\ell_{\phi_l}(\cdot\,; z) + \phi_n$ is then referred to as a $\bar{\tau}$-Lipschitz linearization of ϕ at z.

The algorithm stated below is a variant of the proximal ADMM where partial augmented Lagrangian functions are replaced by corresponding Lipschitz linearizations. To illustrate this, assume that f_i, $i = 1, 2$, in problem (7) is of the form $f_i = \tilde{f}_i + \delta_{X_i}$ where $\tilde{f}_i : \mathbb{R}^{n_i} \to \mathbb{R}$ is differentiable with $L_{\tilde{f}_i}$-Lipschitz continuous gradient and δ_{X_i} is the indicator function of the set X_i. Then the partial augmented Lagrangian functions $\phi_1 = \mathcal{L}_\beta(\cdot, x_2, y, \lambda)$ and $\phi_2 = \mathcal{L}_\beta(x_1, \cdot, y, \lambda)$ have the following natural Lipschitz linearizations $\ell_{\phi_{i,l}}(\cdot; \tilde{x}_i) + \phi_{i,n}$, $i = 1, 2$, where:

(a) $\phi_{i,l} = \tilde{f}_i$ leading to a $L_{\tilde{f}_i}$-Lipschitz linearization of ϕ_i;

(b) $\phi_{i,n} = f_i$ leading to a $(\beta \|A_i\|^2)$-Lipschitz linearization of ϕ_i;

(c) $\phi_{i,n} = \delta_{X_i}$ leading to a $(L_{\tilde{f}_i} + \beta \|A_i\|^2)$-Lipschitz linearization of ϕ_i.

Clearly, similar Lipschitz linearizations can be described with respect to $\phi = \mathcal{L}_\beta(x_1, x_2, \cdot, \lambda)$ in which f_i, $\tilde{f}_i$, δ_{X_i} and A_i are replaced by g, g, $\delta_{\mathbb{R}^q}$ and B, respectively. Hence, since $\delta_{\mathbb{R}^q}$ is the zero function, the linearization of type (c) in this case corresponds to linearizing the entire ϕ.

Linearized Proximal ADMM

---

(0) Let $\bar{\beta}$ be as in **(A4)**, an initial point $(x_1^0, x_2^0, y^0, \lambda^0) \in \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \times \mathbb{R}^q \times \mathbb{R}^d$ and scalars $\bar{\tau}_1, \bar{\tau}_2, \bar{\tau} \geq 0$ be given and let $\tau_i > \bar{\tau}_i$, $i = 1, 2$. Choose scalars $\tau > \bar{\tau}$, $\beta \geq \bar{\beta}$ and a stepsize parameter $\theta \in (0, 2)$ such that

$$\bar{\delta}_1 := \frac{(\tau - \bar{\tau}) + \mu_g + \beta \sigma_B}{4} - \frac{3\gamma_\theta[(\tau + \bar{\tau})^2 + L_g^2]}{\beta \sigma_B^+} > 0, \tag{9}$$

where σ_B (resp., σ_B^+) denotes the smallest eigenvalue (resp., positive eigenvalue) of $B^* B$, and γ_θ is given by

$$\gamma_\theta := \frac{\theta}{(1 - |\theta - 1|)^2}. \tag{10}$$

Set k=1;

(1) let $(\phi_{1,l}^k, \phi_{1,n}^k)$ be a $\bar{\tau}_1$–Lipschitz decomposition of $\mathcal{L}_\beta(\cdot, x_2^{k-1}, y^{k-1}, \lambda^{k-1})$, and compute an optimal solution $x_1^k \in \mathbb{R}^{n_1}$ of the subproblem

$$\min_{x_1 \in \mathbb{R}^{n_1}} \left\{ \ell_{\phi_{1,l}^k}(x_1; x_1^{k-1}) + \phi_{1,n}^k(x_1) + \frac{\tau_1}{2} \|x_1 - x_1^{k-1}\|^2 \right\}; \tag{11}$$

let $(\phi_{2,l}^k, \phi_{2,n}^k)$ be a $\bar{\tau}_2$–Lipschitz decomposition of $\mathcal{L}_\beta(x_1^k, \cdot, y^{k-1}, \lambda^{k-1})$, and compute an optimal solution $x_2^k \in \mathbb{R}^{n_2}$ of the subproblem

$$\min_{x_2 \in \mathbb{R}^{n_2}} \left\{ \ell_{\phi_{2,l}^k}(x_2; x_2^{k-1}) + \phi_{2,n}^k(x_2) + \frac{\tau_2}{2}\|x_2 - x_2^{k-1}\|^2 \right\}; \tag{12}$$

let (ϕ_l^k, ϕ_n^k) be a $\bar{\tau}$–Lipschitz decomposition of $\mathcal{L}_\beta(x_1^k, x_2^k, \cdot, \lambda^{k-1})$, and compute an optimal solution $y^k \in \mathbb{R}^q$ of the subproblem

$$\min_{y \in \mathbb{R}^q} \left\{ \ell_{\phi_l^k}(y; y^{k-1}) + \phi_n^k(y) + \frac{\tau}{2}\|y - y^{k-1}\|^2 \right\}; \tag{13}$$

(2) set

$$\lambda^k = \lambda^{k-1} - \theta\beta \left[A_1 x_1^k + A_2 x_2^k + B y^k - b \right] \tag{14}$$

and $k \leftarrow k+1$, and go to step (1).

end

We now make a few remarks about the linearized proximal ADMM. First, the difficulty of solving (11)–(13) is clearly related to the choice of the Lipschitz linearizations of the partial augmented Lagrangians. If a Lipschitz linearization of type (c) is used (see the discussion after Definition 2.2) then solving (11)–(12) is equivalent to computing a projection of a point onto X_i while the solution of (13) is trivial as long as the evaluation of the gradient of ϕ_l^k is easy. Second, Lipschitz linearizations of types (b) and (c) for the partial augmented Lagrangian $\mathcal{L}_\beta(x_1, x_2, \cdot, \lambda)$ generally does not allow us to choose an arbitrary $\theta \in (0, 2)$ due to restrictive nature of condition (9). Indeed, these two Lipschitz linearizations have the property that their $\bar{\tau}$ constant satisfy $\bar{\tau} = \mathcal{O}(\beta)$ and hence it is not possible to make the second fraction in (9) small by choosing β large. We can however choose θ, and hence γ_θ, small in order to make the second fraction in (9) small.

We now state the main convergence rate result for the linearized ADMM whose proof is postponed to Subsection 3.2. Its main conclusion is that the linearized ADMM generates a quadruple $(\bar{x}_1, \bar{x}_2, \bar{y}, \bar{\lambda})$ which satisfies the optimality conditions of Definition 2.1 within an error of $\mathcal{O}(1/\sqrt{k})$. Its statement uses the quantity $\bar{\eta}_0$ defined as

$$\bar{\eta}_0 := \frac{(\tau - \bar{\tau}) + \mu_g + \beta\sigma_B}{4(\tau + \bar{\tau})^2}\|B^*\lambda^0 - \nabla g(y^0)\|^2. \tag{15}$$

Note that $\bar{\eta}_0 \geq 0$ due to (9).

Theorem 2.3. *Let $(x_1^0, x_2^0, y^0, \lambda^0) \in \operatorname{dom} f_1 \times \operatorname{dom} f_2 \times \mathbb{R}^q \times \mathbb{R}^d$ be given and define*

$$\Delta\bar{\mathcal{L}}_0 := \mathcal{L}_\beta(x_1^0, x_2^0, y^0, \lambda^0) - v(\beta) + \bar{\eta}_0 \tag{16}$$

$$\bar{\delta}_2 := \left(\frac{\beta\theta\bar{\eta}_0}{\Delta\bar{\mathcal{L}}_0} + \frac{3\theta\gamma_\theta[L_g^2 + (\tau + \bar{\tau})^2]}{\sigma_B^+ \bar{\delta}_1} \right)^{-1} \tag{17}$$

where $v(\beta)$, $\bar{\delta}_1$ and $\bar{\eta}_0$ are as in **(A4)**, (9) *and* (15)*, respectively. Also, for every $k \geq 1$, define the quantities $\bar{R}_1^k$, $\bar{R}_2^k$ and $\hat{\lambda}^k$ as*

$$\bar{R}_1^k := \beta A_1^* A_2 \Delta x_2^k + \beta A_1^* B \Delta y^k - \tau_1 \Delta x_1^k + \Delta\phi_{1,l}^k, \quad \bar{R}_2^k := \beta A_2^* B \Delta y^k - \tau_2 \Delta x_2^k + \Delta\phi_{2,l}^k \tag{18}$$

and

$$\hat{\lambda}^k := \lambda^{k-1} - \beta \left(A_1 x_1^k + A_2 x_2^k + By^k - b \right) \tag{19}$$

where

$$\Delta x_1^k := x_1^k - x_1^{k-1}, \quad \Delta x_2^k := x_2^k - x_2^{k-1}, \quad \Delta y^k := y^k - y^{k-1} \tag{20}$$

and

$$\Delta \phi_{1,l}^k := \nabla \phi_{1,l}^k(x_1^k) - \nabla \phi_{1,l}^k(x_1^{k-1}), \quad \Delta \phi_{2,l}^k := \nabla \phi_{2,l}^k(x_2^k) - \nabla \phi_{2,l}^k(x_2^{k-1}).$$

Then, the following statements hold:

a) $\Delta \bar{\mathcal{L}}_0 \geq 0$*;*

b) for every $k \geq 1$*,*

$$\bar{R}_i^k \in \partial f_i(x_i^k) - A_i^* \hat{\lambda}^k, \quad i = 1, 2,$$

and there exists $j \leq k$ *such that*

$$\|\bar{R}_1^j\| \leq 2 \left(\frac{\beta \|A_1^* A_2\|}{\sqrt{\tau_2 - \bar{\tau}_2}} + \frac{\beta \|A_1^* B\|}{\sqrt{\delta_1}} + \frac{\tau_1 + \bar{\tau}_1}{\sqrt{\tau_1 - \bar{\tau}_1}} \right) \sqrt{\frac{\Delta \bar{\mathcal{L}}_0}{k}},$$

$$\|\bar{R}_2^j\| \leq 2 \left(\frac{\beta \|A_2^* B\|}{\sqrt{\delta_1}} + \frac{\tau_2 + \bar{\tau}_2}{\sqrt{\tau_2 - \bar{\tau}_2}} \right) \sqrt{\frac{\Delta \bar{\mathcal{L}}_0}{k}},$$

$$\|\nabla g(y^j) - B^* \hat{\lambda}^j\| \leq (\tau + \bar{\tau}) \sqrt{\frac{2 \Delta \bar{\mathcal{L}}_0}{\bar{\delta}_1 k}},$$

$$\|A_1 x_1^j + A_2 x_2^j + By^j - b\| \leq \frac{1}{\beta \theta} \sqrt{\frac{2 \Delta \bar{\mathcal{L}}_0}{\bar{\delta}_2 k}}.$$

3 Proximal ADMM with Bregman distances

This section contains two subsections. Subsection 3.1 presents the non-Euclidean proximal ADMM (NEP-ADMM) studied in this paper and states its main convergence rate result (Theorem 3.3) whose proof is given in Section 4. Subsection 3.2 provides the proof of Theorem 2.3 based on the fact that the linearized proximal ADMM is a particular instance of the non-Euclidean proximal ADMM (see Proposition 3.5).

3.1 The Non-Euclidean Proximal ADMM

The main goal of this subsection is to present the NEP-ADMM and its main convergence rate result.

We start by introducing a class of distance generating functions (and its corresponding Bregman distances) which is suitable for our presentation in this paper.

Definition 3.1. *For given set* $Z \subset \mathbb{R}^s$ *and scalars* $m \leq M$*, we let* $\mathcal{D}_Z(m, M)$ *denote the class of real-valued functions* w *which are differentiable on* Z *and satisfy*

$$w(z') - w(z) - \langle \nabla w(z), z' - z \rangle \geq \frac{m}{2} \|z - z'\|^2 \quad \forall z, z' \in Z, \tag{21}$$

$$\|\nabla w(z) - \nabla w(z')\| \leq M \|z - z'\| \quad \forall z, z' \in Z. \tag{22}$$

A function $w \in \mathcal{D}_Z(m, M)$ with $m \geq 0$ is referred to as a distance generating function and its associated Bregman distance $dw : \mathbb{R}^s \times Z \to \mathbb{R}$ is defined as

$$(dw)(z'; z) := w(z') - w(z) - \langle \nabla w(z), z' - z \rangle \quad \forall (z', z) \in \mathbb{R}^s \times Z. \tag{23}$$

For notation simplicity, for every $z \in Z$, the function $(dw)(\,\cdot\,; z)$ will be denoted by $(dw)_z$ so that

$$(dw)_z(z') = (dw)(z'; z) \quad \forall (z', z) \in \mathbb{R}^s \times Z.$$

Clearly,

$$\nabla (dw)_z(z') = -\nabla (dw)_{z'}(z) = \nabla w(z') - \nabla w(z) \quad \forall z, z' \in Z, \tag{24}$$

We now state the non-Euclidean proximal ADMM based on the class of distance generating functions introduced in Definition 3.1.

Non-Euclidean Proximal ADMM (NEP-ADMM)

---

(0) Define $Z_i := \operatorname{dom} f_i$ for $i = 1, 2$ and let $\bar{\beta}$ be as in **(A4)**. Let an initial point $(x_1^0, x_2^0, y^0, \lambda^0) \in Z_1 \times Z_2 \times \mathbb{R}^q \times \mathbb{R}^d$ and scalars $M_i \geq m_i > 0$, $i = 1, 2$, be given. Choose scalars $M \geq m > 0$, $\beta \geq \bar{\beta}$ and a stepsize parameter $\theta \in (0, 2)$ such that

$$\delta_1 := \frac{m + \mu_g + \beta\sigma_B}{4} - \frac{3\gamma_\theta(M^2 + L_g^2)}{\beta\sigma_B^+} > 0 \tag{25}$$

where σ_B, σ_B^+ and γ_θ are as in step 0 of the linearized proximal ADMM. Set $k = 1$ and go to step 1.

(1) Choose $w_1^k \in \mathcal{D}_{Z_1}(m_1, M_1)$ and compute an optimal solution $x_1^k \in \mathbb{R}^{n_1}$ of

$$\min_{x_1 \in \mathbb{R}^{n_1}} \left\{ \mathcal{L}_\beta(x_1, x_2^{k-1}, y^{k-1}, \lambda^{k-1}) + (dw_1^k)_{x_1^{k-1}}(x_1) \right\}. \tag{26}$$

Also, choose $w_2^k \in \mathcal{D}_{Z_2}(m_2, M_2)$ and an optimal solution $x_2^k \in \mathbb{R}^{n_2}$ of

$$\min_{x_2 \in \mathbb{R}^{n_2}} \left\{ \mathcal{L}_\beta(x_1^k, x_2, y^{k-1}, \lambda^{k-1}) + (dw_2^k)_{x_2^{k-1}}(x_2) \right\}. \tag{27}$$

(2) Choose $w^k \in \mathcal{D}_{\mathbb{R}^q}(m, M)$ and compute an optimal solution $y^k \in \mathbb{R}^q$ of

$$\min_{y \in \mathbb{R}^q} \left\{ \mathcal{L}_\beta(x_1^k, x_2^k, y, \lambda^{k-1}) + (dw^k)_{y^{k-1}}(y) \right\}. \tag{28}$$

(3) Set

$$\lambda^k = \lambda^{k-1} - \theta\beta \left[A_1 x_1^k + A_2 x_2^k + B y^k - b \right], \tag{29}$$

$k \leftarrow k + 1$, and go to step (1).

end

Some comments about the NEP-ADMM stated above are in order. First, it follows from (25) that $m + \mu_g + \beta\sigma_B > 0$ and hence that the objective function of (28) is $(m + \mu_g + \beta\sigma_B)$–strongly convex which in turn implies that y^k is uniquely determined. Second, it is always possible to choose m, M, β and θ so that (25) is satisfied since the first fraction in (25) can be made positive by choosing either m sufficiently large or β sufficiently large if $\sigma_B > 0$, while the second fraction in (25) can be made smaller than the first one by choosing either β sufficiently large or $\theta > 0$ sufficiently close to zero. Third, as will be shown in Subsection 2.3, the linearized proximal ADMM can be viewed as an instance of the NEP-ADMM when the distance generating functions w_1^k, w_2^k, and w^k are properly chosen. Fourth, the use of variable distance generating functions (or variable metrics) is not only interesting in its own right but allows us to treat linearized ADMMs in a unified manner.

The next result describes a set of inclusions/equations satisfied by the sequence generated by the NEP-ADMM.

Lemma 3.2. *Consider the sequence $\{(x_1^k, x_2^k, y^k, \lambda^k)\}$ generated by the NEP-ADMM and let $\hat{\lambda}^k$ be as in* (19)*. Moreover, for every $k \geq 1$, define*

$$R_1^k := \beta A_1^* A_2 \Delta x_2^k + \beta A_1^* B \Delta y^k - \Delta w_1^k \text{ and } R_2^k := \beta A_2^* B \Delta y^k - \Delta w_2^k. \tag{30}$$

where Δx_2^k and Δy^k are as in (20) *and*

$$\Delta w_i^k := \nabla w_i^k(x_i^k) - \nabla w_i^k(x_i^{k-1}),\ i = 1, 2.$$

Then, for every $k \geq 1$, we have:

$$R_i^k \in \partial f_i(x_i^k) - A_i^* \hat{\lambda}^k \quad i = 1, 2, \tag{31}$$

$$0 = \left[\nabla g(y^k) - B^* \hat{\lambda}^k\right] + \Delta w^k, \tag{32}$$

$$0 = \left[A_1 x_1^k + A_2 x_2^k + B y^k - b\right] + \frac{1}{\theta\beta} \Delta\lambda^k \tag{33}$$

where $\Delta w^k := \nabla w^k(y^k) - \nabla w^k(y^{k-1})$ and $\Delta\lambda^k := \lambda^k - \lambda^{k-1}$.

Proof. The optimality conditions (see Proposition 1.3) for (26), (27) and (28) imply that

$$\begin{aligned} 0 &\in \partial f_1(x_1^k) - A_1^*[\lambda_{k-1} - \beta(A_1 x_1^k + A_2 x_2^{k-1} + B y^{k-1} - b)] + \Delta w_1^k, \\ 0 &\in \partial f_2(x_2^k) - A_2^*[\lambda_{k-1} - \beta(A_1 x_1^k + A_2 x_2^k + B y^{k-1} - b)] + \Delta w_2^k, \\ 0 &= \nabla g(y^k) - B^*[\lambda^{k-1} - \beta(A_1 x_1^k + A_2 x_2^k + B y^k - b)] + \Delta w^k, \end{aligned}$$

respectively. These relations combined with (19) immediately yield (31) and (32). Relation (33) follows immediately from (29). □

The convergence rate bounds for the NEP-ADMM stated in Theorem 3.3 below are expressed in terms of the following quantity. For a given $(y^0, \lambda^0) \in \mathbb{R}^q \times \mathbb{R}^d$, define

$$\begin{aligned} \eta_0 = \eta_0(y^0, \lambda^0) := \min_{(\Delta y^0, \Delta\lambda^0)} &\frac{c_1}{2}\|B^* \Delta\lambda^0\|^2 + \frac{m + \mu_g + \beta\sigma_B}{4}\|\Delta y^0\|^2 \\ &\text{s.t. } M\Delta y^0 + \frac{\theta - 1}{\theta} B^* \Delta\lambda^0 = B^* \lambda^0 - \nabla g(y^0) \end{aligned} \tag{34}$$

where

$$c_1 := \frac{2|\theta - 1|}{\beta\theta(1 - |\theta - 1|)\sigma_B^+} \geq 0. \tag{35}$$

We now present the convergence rate result for the NEP-ADMM whose proof is given in Section 4. Its main conclusion is that the NEP-ADMM generates a quadruple $(\bar{x}_1, \bar{x}_2, \bar{y}, \bar{\lambda})$ which satisfies the optimality conditions of Definition 2.1 within an error of $\mathcal{O}(1/\sqrt{k})$.

Theorem 3.3. *Let $(x_1^0, x_2^0, y^0, \lambda^0) \in \operatorname{dom} f_1 \times \operatorname{dom} f_2 \times \mathbb{R}^q \times \mathbb{R}^d$ be given and define*

$$\Delta\mathcal{L}_0 = \Delta\mathcal{L}_0(\beta) := \mathcal{L}_\beta(x_1^0, x_2^0, y^0, \lambda^0) - v(\beta) + \eta_0 \tag{36}$$

where η_0 and $v(\beta)$ are as in (34) *and* **(A4)**, *respectively. Consider δ_1 as in* (25) *and define*

$$\delta_2 = \left(\frac{\beta\theta\eta_0}{\Delta\mathcal{L}_0} + \frac{3\theta\gamma_\theta(L_g^2 + M^2)}{\sigma_B^+ \delta_1} \right)^{-1}. \tag{37}$$

Moreover, let $\hat{\lambda}^k$ and R_i^k, $i = 1, 2$, be as in (19) *and* (30), *respectively. Then, the following statements hold:*

a) $\Delta\mathcal{L}_0 \geq 0$;

b) for every $k \geq 1$, the inclusions (31)–(33) *hold and there exists $j \leq k$ such that*

$$\|R_1^j\| \leq 2\left(\frac{\beta\|A_1^* A_2\|}{\sqrt{m_2}} + \frac{\beta\|A_1^* B\|}{\sqrt{\delta_1}} + \frac{M_1}{\sqrt{m_1}} \right)\sqrt{\frac{\Delta\mathcal{L}_0}{k}}, \quad \|R_2^j\| \leq 2\left(\frac{\beta\|A_2^* B\|}{\sqrt{\delta_1}} + \frac{M_2}{\sqrt{m_2}} \right)\sqrt{\frac{\Delta\mathcal{L}_0}{k}},$$

$$\|\nabla g(y^j) - B^*\hat{\lambda}^j\| \leq M\sqrt{\frac{2\Delta\mathcal{L}_0}{\delta_1 k}}, \qquad \|A_1 x_1^j + A_2 x_2^j + By^j - b\| \leq \frac{1}{\beta\theta}\sqrt{\frac{2\Delta\mathcal{L}_0}{\delta_2 k}}.$$

3.2 Proof of Theorem 2.3

The main goal of this subsection is to present the proof of Theorem 2.3 which is based on showing that the linearized proximal ADMM is an instance of the NEP-ADMM and then using Theorem 3.3.

The next result shows how to obtain distance generating functions from Lipschitz decompositions of a proper lower semi-continuous function.

Proposition 3.4. *Let (ϕ_l, ϕ_n) be a $\bar{\tau}$-Lipschitz decomposition of a proper lower semi-continuous function ϕ where $\bar{\tau} \geq 0$ and, for some $\tau > \bar{\tau}$, define*

$$w_l := \frac{\tau}{2}\|\cdot\|^2 - \phi_l.$$

Then, the following statements hold:

(a) $w_l \in D_Z(m_l, M_l)$ where $Z := \operatorname{dom}\phi$ and

$$m_l = \tau - \bar{\tau}, \quad M_l := \tau + \bar{\tau};$$

(b) for every $z, z' \in \operatorname{dom}\phi$, we have

$$\phi_l(z) + (dw_l)_{z'}(z) = \ell_{\phi_l}(z; z') + \frac{\tau}{2}\|z - z'\|^2.$$

Proof. Definition 2.1 implies that $\nabla\phi_l$ is $\bar\tau$-Lipschitz continuous on $\mathrm{co}(Z)$ and hence

$$(d\phi_l)_{z'}(z) \le \frac{\bar\tau}{2}\|z'-z\|^2 \quad \forall z,z' \in \mathrm{co}(Z).$$

This inequality together with the definition of w_l then imply that

$$(dw_l)_{z'}(z) = \frac{\tau}{2}\|z'-z\|^2 - (d\phi_l)_{z'}(z) \ge \frac{\tau-\bar\tau}{2}\|z'-z\|^2 = \frac{m_l}{2}\|z'-z\|^2 \quad \forall z,z' \in \mathrm{co}(Z).$$

Since $\nabla w_l(z) = \tau z - \nabla\phi_l(z)$, the $\bar\tau$–Lipschitz continuity of $\nabla\phi_l$ on $\mathrm{co}(Z)$ also implies that

$$\|\nabla w_l(z) - \nabla w_l(z')\| \le \tau\|z-z'\| + \|\nabla\phi_l(z) - \nabla\phi_l(z')\| \le (\tau+\bar\tau)\|z-z'\| \quad \forall z,z' \in \mathrm{co}(Z).$$

Hence (a) follows. (b) follows immediately from the following relation:

$$(dw_l)_{z'}(z) = \frac{\tau}{2}\|z-z'\|^2 - (d\phi_l)_{z'}(z) = \frac{\tau}{2}\|z-z'\|^2 - \left[\phi_l(z) - \ell_{\phi_l}(z;z')\right]. \qquad \square$$

The next result shows that the linearized proximal ADMM is an instance of the NEP-ADMM.

Proposition 3.5. *The linearized proximal ADMM is an instance of the NEP-ADMM where $m_i = \tau_i - \bar\tau_i, M_i = \tau_i + \bar\tau_i$, $i = 1,2$, $m = \tau - \bar\tau, M = \tau + \bar\tau$ and the distance generating functions are given by*

$$w_i := \frac{\tau_i}{2}\|\cdot\|^2 - \phi_i, \quad i = 1,2, \text{ and } w := \frac{\tau}{2}\|\cdot\|^2 - \phi.$$

Proof. From step 1 of the linearized proximal ADMM we see that, for each $k \ge 1$, the pairs (ϕ_l^k, ϕ_n^k) and $(\phi_{i,l}^k, \phi_{i,n}^k)$, i=1,2, are $\bar\tau$–Lipschitz and $\bar\tau_i$–Lipschitz decomposition of the partial augmented Lagrangians $\mathcal{L}_y^k := \mathcal{L}_\beta(x_1^k, x_2^k, \cdot, \lambda^{k-1})$, $\mathcal{L}_{x_1}^k := \mathcal{L}_\beta(\cdot, x_2^{k-1}, y^{k-1}, \lambda^{k-1})$ and $\mathcal{L}_{x_2}^k := \mathcal{L}_\beta(x_1^k, \cdot, y^{k-1}, \lambda^{k-1})$, respectively. Hence, using Proposition 3.4 consecutively with $\phi = \mathcal{L}_y^k$, τ and $\bar\tau$ as above, and $\phi = \mathcal{L}_{x_i}^k, \bar\tau = \bar\tau_i$, $\tau = \tau_i$, $i = 1,2$, we see that $\bar\delta_1$ as in (9) corresponds exactly to δ_1 given in (25). Moreover, the subproblems (11), (12) and (13) correspond to (26), (27) and (28), respectively. Hence, since (14) is the same as (29), the proof is concluded. $\square$

We end this subsection by presenting the proof of Theorem 2.3.

Proof of Theorem 2.3. Proposition 3.5 shows that the linearized proximal ADMM is an instance of the NEP-ADMM. Moreover, by considering the scalars $m_i, M_i, i = 1,2$, and m, M as in Proposition 3.5, we see that $\bar\delta_1$ as in (9) corresponds to δ_1 given in (25). Also, from (15), (16), (34) and (36), we see that $\eta_0 \le \bar\eta_0$, $\Delta\mathcal{L}_0 \le \Delta\bar{\mathcal{L}}_0$ and $\bar\eta_0/\Delta\bar{\mathcal{L}}_0 \le \eta_0/\Delta\mathcal{L}_0$ which in turn implies that $\bar\delta_2 \ge \delta_2$. Clearly by choosing w_i, $i = 1,2$, as in Proposition 3.5, we also see that $\bar R_i^k$ as in (18) corresponds to R_i^k given in (30), $i = 1,2$, respectively. Hence, altogether show that Theorem 2.3 follows from Theorem 3.3.

4 Proof of Theorem 3.3

The main goal of this section is to provide the proof of Theorem 3.3. First we need some technical lemmas. The first one provides a recursive relation for the sequence $\{\Delta\lambda^k\}$.

Lemma 4.1. *Let $\Delta y^0 \in \mathbb{R}^q$ and $\Delta\lambda^0 \in \mathbb{R}^d$ be such that*

$$M\Delta y^0 + \frac{\theta - 1}{\theta} B^*\Delta\lambda^0 = B^*\lambda^0 - \nabla g(y^0) \tag{38}$$

and define $\Delta w^0 := M\Delta y^0$. Then, for every $k \geq 1$, we have

$$B^*\Delta\lambda^k = (1-\theta)B^*\Delta\lambda^{k-1} + \theta u^k, \tag{39}$$

where

$$u^k := \Delta g^k + (\Delta w^k - \Delta w^{k-1}), \quad \Delta g^k := \nabla g^k(y^k) - \nabla g^k(y^{k-1}) \quad \forall k \geq 1, \tag{40}$$

$\Delta\lambda^k$ and Δw^k are as in Lemma 3.2.

Proof. Using (19) and (33) we easily see that

$$\theta\hat{\lambda}^k := \lambda^k + (\theta - 1)\lambda^{k-1}, \quad \forall k \geq 1.$$

This expression together with (32) then imply that

$$B^*\lambda^k = (1-\theta)B^*\lambda^{k-1} + \theta[\nabla g(y^k) + \Delta w^k], \quad \forall k \geq 1. \tag{41}$$

Hence, in view of (40), relation (39) holds for every $k \geq 2$. Now, note that (38) is equivalent to the relation $\theta(\nabla g(y^0) + \Delta w^0) = \theta B^*\lambda^0 + (1-\theta)B^*\Delta\lambda^0$. Hence using this relation, (40) and (41) both with $k = 1$, we have

$$\begin{aligned} B^*\Delta\lambda^1 &= -\theta B^*\lambda^0 + \theta\left[\nabla g(y^1) + \Delta w^1\right] \\ &= -\theta B^*\lambda^0 + \theta\left[\nabla g(y^0) + \Delta w^0 + u^1\right] \\ &= -\theta B^*\lambda^0 + \theta B^*\lambda^0 + (1-\theta)B^*\Delta\lambda^0 + \theta u^1. \end{aligned}$$

Hence (39) also holds for $k = 1$. □

The next lemma describes how the sequence $\{(x_1^k, x_2^k, y^k, \lambda^k)\}$ affects the value of the augmented Lagrangian function defined in (8).

Lemma 4.2. *For every $k \geq 1$, we have*

(a) $\mathcal{L}_\beta(x_1^k, x_2^{k-1}, y^{k-1}, \lambda^{k-1}) - \mathcal{L}_\beta(x_1^{k-1}, x_2^{k-1}, y^{k-1}, \lambda^{k-1}) \leq -(dw_1^k)_{x_1^{k-1}}(x_1^k)$*;*

(b) $\mathcal{L}_\beta(x_1^k, x_2^k, y^{k-1}, \lambda^{k-1}) - \mathcal{L}_\beta(x_1^k, x_2^{k-1}, y^{k-1}, \lambda^{k-1}) \leq -(dw_2^k)_{x_2^{k-1}}(x_2^k)$*;*

(c) $\mathcal{L}_\beta(x_1^k, x_2^k, y^k, \lambda^{k-1}) - \mathcal{L}_\beta(x_1^k, x_2^k, y^{k-1}, \lambda^{k-1}) \leq -\left(2m + \mu_g + \beta\sigma_B\right)\|\Delta y^k\|^2/2$*;*

(d) $\mathcal{L}_\beta(x_1^k, x_2^k, y^k, \lambda^k) - \mathcal{L}_\beta(x_1^k, x_2^k, y^k, \lambda^{k-1}) = \|\Delta\lambda^k\|^2/(\theta\beta)$*.*

Proof. (a) and (b) follows directly from (26) and (27), respectively.

(c) Observe that the objective function of (28) can be written as

$$\mathcal{L}_\beta(x_1^k, x_2^k, \cdot, \lambda^{k-1}) + (dw^k)_{y^{k-1}}(\cdot) = g + \psi \tag{42}$$

where $\psi = q + w^k$ for some quadratic function q whose Hessian is $\beta B^* B$. Hence, $\mu_\psi \geq m + \beta\sigma_B > -\mu_g$ where the second inequality is due to (25). Since $\mu_g + \mu_\psi > 0$, it follows from (42) and Lemma A.1 with $\mu = \mu_g$, $y = y^{k-1}$ and $\bar{y} = y^k$ that

$$
\begin{aligned}
\mathcal{L}_\beta(x_1^k, x_2^k, y^{k-1}, \lambda^{k-1}) &= (g+\psi)(y^{k-1}) \geq (g+\psi)(y^k) + (d\psi)_{y^{k-1}}(y^k) + \frac{\mu_g}{2}\|\Delta y^k\|^2 \\
&= \mathcal{L}_\beta(x_1^k, x_2^k, y^k, \lambda^{k-1}) + 2(dw^k)_{y^{k-1}}(y^k) + \frac{\beta}{2}\|B\Delta y^k\|^2 + \frac{\mu_g}{2}\|\Delta y^k\|^2 \\
&\geq \mathcal{L}_\beta(x_1^k, x_2^k, y^k, \lambda^{k-1}) + \frac{2m + \mu_g + \beta\sigma_B}{2}\|\Delta y^k\|^2.
\end{aligned}
$$

(d) This statement follows from (29), the identity $\Delta\lambda^k = \lambda^k - \lambda^{k-1}$ and the fact that (8) implies that

$$
\mathcal{L}_\beta(x_1^k, x_2^k, y^k, \lambda^k) = \mathcal{L}_\beta(x_1^k, x_2^k, y^k, \lambda^{k-1}) - \langle \lambda^k - \lambda^{k-1}, A_1 x_1^k + A_2 x_2^k + B y^k - b\rangle. \qquad \square
$$

Our goal now is to show that a certain sequence associated with $\{\mathcal{L}_\beta(x_1^k, x_2^k, y^k, \lambda^k)\}$ is monotonically decreasing, namely, the sequence $\{\hat{\mathcal{L}}_k\}$ defined as

$$
\hat{\mathcal{L}}_k := \mathcal{L}_\beta(x_1^k, x_2^k, y^k, \lambda^k) + \eta_k \qquad \forall k \geq 0, \tag{43}
$$

where

$$
\eta_k := \frac{c_1}{2}\|B^*\Delta\lambda^k\|^2 + \frac{m + \mu_g + \beta\sigma_B}{4}\|\Delta y^k\|^2 \qquad \forall k \geq 1, \tag{44}
$$

η_0 and c_1 are as in (34) and (35), respectively. Before establishing the monotonicity property of the sequence $\{\hat{\mathcal{L}}_k\}$, we state some technical results. The first one describes an upper bound on $\hat{\mathcal{L}}_k - \hat{\mathcal{L}}_{k-1}$ in terms quantities related to $\{\Delta x_1^k\}$, $\{\Delta x_2^k\}$, $\{\Delta\lambda^k\}$ and $\{\Delta y^k\}$.

Lemma 4.3. *For every $k \geq 1$,*

$$
\hat{\mathcal{L}}_k - \hat{\mathcal{L}}_{k-1} \leq -(dw_1^k)_{x_1^{k-1}}(x_1^k) - (dw_2^k)_{x_2^{k-1}}(x_2^k) + \Theta_\lambda^k + \Theta_y^k, \tag{45}
$$

where

$$
\Theta_\lambda^k := \frac{1}{\beta\theta}\|\Delta\lambda^k\|^2 + \frac{c_1}{2}\left(\|B^*\Delta\lambda^k\|^2 - \|B^*\Delta\lambda^{k-1}\|^2\right) \tag{46}
$$

and

$$
\Theta_y^k := -\left(\frac{m + \mu_g + \beta\sigma_B}{4}\right)\left(\|\Delta y^k\|^2 + \|\Delta y^{k-1}\|^2\right), \tag{47}
$$

where c_1 is defined in (35).

Proof. The lemma follows by adding the inequalities given in statements (a), (b), (c) and (d) of the previous lemma and using the definition of $\hat{\mathcal{L}}_k$ in (43). $\square$

The next result provides an upper bound for Θ_λ^k in terms of the sequence $\{\Delta y^k\}$.

Lemma 4.4. *Consider Θ_λ^k as in* (46) *and u^k as in* (40)*. Then,*

$$
\Theta_\lambda^k \leq \frac{\gamma_\theta}{\beta\sigma_B^+}\|u_k\|^2 \leq \frac{\gamma_\theta}{\beta\sigma_B^+}\left[3(L_g^2 + M^2)\|\Delta y^k\|^2 + 3M^2\|\Delta y^{k-1}\|^2\right] \qquad \forall k \geq 1.
$$

where γ_θ and Δy^0 are defined in (10) *and Lemma 4.1, respectively.*

Proof. The proof of the first inequality is the same as the one given in Lemma 3.5 of [10]. We now prove the second inequality. Note that due to the Lipschitz continuity of ∇w^k, non-expansiveness of the projection operator, and the fact that $\Delta w^0 = M\Delta y^0$ (see Lemma 4.1), we obtain $\|\mathcal{P}(\Delta w^k)\| \leq M\|\Delta y^k\|$ for all $k \geq 0$. The latter relation, assumption **(A2)**, the fact that $u^k \in \operatorname{Im} B^*$ (see (39)), and relation (40), imply that

$$
\begin{aligned}
\|u^k\|^2 &= \|\mathcal{P}_{B^*}(u^k)\|^2 = \|\mathcal{P}_{B^*}\left[\Delta g_k + (\Delta w^k - \Delta w^{k-1})\right]\|^2 \\
&\leq \left[L_g\|\Delta y^k\| + M(\|\Delta y^k\| + \|\Delta y^{k-1}\|)\right]^2 \leq 3\left[L_g^2\|\Delta y^k\|^2 + M^2\left(\|\Delta y^k\|^2 + \|\Delta y^{k-1}\|^2\right)\right]
\end{aligned}
$$

where the last two inequalities follow from the triangle inequality for norms and the relation $(s_1 + s_2 + s_3)^2 \leq 3(s_1^2 + s_2^2 + s_3^3)$ for $s_1, s_2, s_3 \in \mathbb{R}$. □

The next proposition shows that the sequence $\{\hat{\mathcal{L}}_k\}$ is decreasing and bounded below.

Proposition 4.5. *The following statements hold:*

(a) for every $k \geq 1$,

$$\hat{\mathcal{L}}_k - \hat{\mathcal{L}}_{k-1} \leq -(dw_1^k)_{x_1^{k-1}}(x_1^k) - (dw_2^k)_{x_2^{k-1}}(x_2^k) - \delta_1(\|\Delta y^k\|^2 + \|\Delta y^{k-1}\|^2);$$

(b) the sequence $\{\hat{\mathcal{L}}_k\}$ given in (43) satisfies $\hat{\mathcal{L}}_k \geq v(\beta)$ for every $k \geq 1$;

(c) for every $k \geq 1$,

$$\sum_{j=1}^{k}\left[(dw_1^j)_{x_1^{j-1}}(x_1^j) + (dw_2^j)_{x_2^{j-1}}(x_2^j) + \delta_1(\|\Delta y^j\|^2 + \|\Delta y^{j-1}\|^2)\right] \leq \Delta\mathcal{L}_0$$

*where $v(\beta)$, δ_1 and $\Delta\mathcal{L}_0$ are as in (**A4**), (25) and (16), respectively.*

Proof. (a) It follows from Lemma 4.4 that

$$\Theta_\lambda^k \leq \frac{3\gamma_\theta\left(L_g^2 + M^2\right)}{\beta\sigma_B^+}\left(\|\Delta y^k\|^2 + \|\Delta y^{k-1}\|^2\right)$$

and hence, in view of (25) and (47), we have

$$
\begin{aligned}
\Theta_\lambda^k + \Theta_y^k &\leq -\left[\frac{m + \mu_g + \beta\sigma_B}{4} - \frac{3\gamma_\theta(L_g^2 + M^2)}{\beta\sigma_B^+}\right]\left(\|\Delta y^k\|^2 + \|\Delta y^{k-1}\|^2\right) \\
&= -\delta_1(\|\Delta y^k\|^2 + \|\Delta y^{k-1}\|^2).
\end{aligned}
$$

Hence, (a) follows by combining the above estimate with (45). The proof of (b) is given in Appendix B, and the proof of (c) follows from (a), (b) and the definition of $\Delta\mathcal{L}_0$ in (16). □

Proposition 4.6. *Let $\Delta\mathcal{L}_0$, δ_1 and δ_2 be as in* (36), (25) *and* (37)*, respectively. Then, for every $k \geq 1$, we have*

$$\sum_{j=1}^{k}\left[(dw_1^j)_{x_1^{j-1}}(x_1^j) + (dw_2^j)_{x_2^{j-1}}(x_2^j) + \delta_1\|\Delta y^j\|^2 + \delta_2\|\Delta\lambda^j\|^2\right] \leq 2\Delta\mathcal{L}_0 \tag{48}$$

and there exists $j \leq k$ such that

$$\|\Delta x_1^j\| \leq \sqrt{\frac{4\Delta\mathcal{L}_0}{km_1}}, \quad \|\Delta x_2^j\| \leq \sqrt{\frac{4\Delta\mathcal{L}_0}{km_2}}, \quad \|\Delta y^j\| \leq \sqrt{\frac{2\Delta\mathcal{L}_0}{k\delta_1}}, \quad \|\Delta\lambda^j\| \leq \sqrt{\frac{2\Delta\mathcal{L}_0}{k\delta_2}}. \tag{49}$$

Proof. It follows from Proposition 4.5(c) that

$$\sum_{j=1}^{k}(\|\Delta y^j\|^2 + \|\Delta y^{j-1}\|^2) \leq \frac{\Delta\mathcal{L}_0}{\delta_1} \tag{50}$$

and that in order to prove (48), it suffices to show that

$$\sum_{j=1}^{k}\|\Delta\lambda^j\|^2 \leq \frac{\Delta\mathcal{L}_0}{\delta_2}. \tag{51}$$

Then, in the remaining part of the proof we will show that (51) holds. By rewriting (46), we have

$$\|\Delta\lambda^k\|^2 = \beta\theta\left[\frac{c_1}{2}\left(\|B^*\Delta\lambda^{k-1}\|^2 - \|B^*\Delta\lambda^k\|^2\right) + \Theta_\lambda^k\right] \qquad \forall k \geq 1,$$

where $\Delta\lambda^0$ is such that the pair $(\Delta y^0, \Delta\lambda^0)$ is a solution of (34). Hence, using (34) and Lemmas 4.4, we obtain

$$\begin{aligned}
\sum_{j=1}^{k}\|\Delta\lambda^j\|^2 &\leq \beta\theta\left[\frac{c_1}{2}\|B^*\Delta\lambda^0\|^2 + \sum_{j=1}^{k}\Theta_j^1\right] \leq \beta\theta\eta_0 + \frac{\theta\gamma_\theta}{\sigma_B^+}\sum_{j=1}^{k}\|u^j\|^2 \\
&\leq \beta\theta\eta_0 + \sum_{j=1}^{k}\frac{3\theta\gamma_\theta L_g^2}{\sigma_B^+}\|\Delta y^j\|^2 + \frac{3\theta\gamma_\theta M^2}{\sigma_B^+}\sum_{j=1}^{k}(\|\Delta y^j\|^2 + \|\Delta y^{j-1}\|^2) \\
&\leq \frac{\beta\theta\eta_0\Delta\mathcal{L}_0}{\Delta\mathcal{L}_0} + \frac{3\theta\gamma_\theta L_g^2\Delta\mathcal{L}_0}{\delta_1\sigma_B^+} + \frac{3\theta\gamma_\theta M^2\Delta\mathcal{L}_0}{\delta_1\sigma_B^+} \\
&\leq \left[\frac{\beta\theta\eta}{\Delta\mathcal{L}_0} + \frac{3\theta\gamma_\theta(L_g^2 + M^2)}{\delta_1\sigma_B^+}\right]\Delta\mathcal{L}_0 = \frac{\Delta\mathcal{L}_0}{\delta_2}
\end{aligned}$$

where the fourth inequality is due to (50). Hence, (51) follows from the above estimate, proving (48). To end the proof of the proposition, combine (48) with the fact that $(dw_i^k)(z;z') \geq m_i\|z - z'\|/2$, $i = 1,2$, due to $w_i^k \in \mathcal{D}_{Z_i}(m_i, M_i)$, $i = 1,2$ (see steps 1 and 2 of the NEP-ADMM). □

We end this section by presenting the proof of Theorem 3.3.

Proof of Theorem 3.3. (a) holds due to Proposition 4.5(c). Lemma 3.2 shows that the first statement of (b) holds. Now, it follows from (30), (32), (33) and the fact that $w_i^k \in \mathcal{D}_{Z_i}(m_i, M_i)$, $i = 1, 2$, and $w^k \in \mathcal{D}_Z(m, M)$, that

$$\|R_1^k\| \le \beta\|A_1^* A_2\|\|\Delta x_2^k\| + \beta\|A_1^* B\|\|\Delta y^k\| + M_1\|\Delta x_1^k\|,$$
$$\|R_2^k\| \le \beta\|A_2^* B\|\|\Delta y^k\| + M_2\|\Delta x_2^k\|,$$
$$\|\nabla g(y^k) - B^*\hat{\lambda}^k\| \le M\|\Delta y^k\|, \quad \|A_1 x_1^k + A_2 x_2^k + By^k - b\| = \frac{1}{\beta\theta}\|\Delta\lambda^k\|.$$

Hence, to end the proof, just combine the above relations with (49). □

A Auxiliary Result

This section presents an auxiliary result used in Lemma 4.2(c).

Lemma A.1. *Assume that $g, \psi : \mathbb{R}^q \to \mathbb{R}$ are lower semicontinuous function such that for some $\mu \in \mathbb{R}$, $g(\cdot) - \mu\|\cdot\|^2/2$ is convex and $\psi(\cdot) + \mu\|\cdot\|^2/2$ is strongly convex and differentiable. Then, the problem*

$$\min\{(g + \psi)(y) : y \in \mathbb{R}^q\} \tag{52}$$

has a unique optimal solution $\bar{y}$ and

$$(g + \psi)(y) \ge (g + \psi)(\bar{y}) + (d\psi)_{\bar{y}}(y) + \frac{\mu}{2}\|y - \bar{y}\|^2 \quad \forall y \in \mathbb{R}^q. \tag{53}$$

Proof. Define $\tilde{g} := g - \mu\|\cdot\|^2/2$ and $\tilde{\psi} := \psi + \mu\|\cdot\|^2/2$. Clearly, $\tilde{g}$ is a proper lower semi-continuous convex function and $\tilde{\psi}$ is a strongly convex function. Since $g + \psi = \tilde{g} + \tilde{\psi}$, we conclude that the objective function of (52) is strongly convex, and hence that the first statement of the lemma follows. Moreover, we have

$$0 \in \partial(g + \psi)(\bar{y}) = \partial(\tilde{g} + \tilde{\psi})(\bar{y}) = \partial\tilde{g}(\bar{y}) + \nabla\tilde{\psi}(\bar{y})$$

and hence

$$\tilde{g}(y) \ge \tilde{g}(\bar{y}) - \langle\nabla\tilde{\psi}(\bar{y}), y - \bar{y}\rangle \quad \forall y \in \mathbb{R}^q.$$

On the other hand, the definition of $d\tilde{w}$ implies that

$$\tilde{\psi}(y) = \tilde{\psi}(\bar{y}) + \langle\nabla\tilde{\psi}(\bar{y}), y - \bar{y}\rangle + d\tilde{\psi}_{\bar{y}}(y) \quad \forall y \in \mathbb{R}^q.$$

Adding the above two relations, and using the fact that $g + \psi = \tilde{g} + \tilde{\psi}$ and noting that

$$(d\tilde{\psi})_{\bar{y}}(y) = (d\psi)_{\bar{y}}(y) + \frac{\mu}{2}\|y - \bar{y}\|^2,$$

we conclude that (53) holds. □

Next we present the proof of Proposition 4.5(b).

B Proof of Proposition 4.5(b)

We want to prove that $\hat{\mathcal{L}}_k \geq v(\beta)$ for every $k \geq 1$. Assume for contradiction that there exists an index $k_0 \geq 0$ such that $\hat{\mathcal{L}}_{k_0+1} < v(\beta)$. Since $\{\hat{\mathcal{L}}_k\}$ is decreasing (see Proposition 4.5(a)), we obtain

$$\sum_{k=1}^{j}(\hat{\mathcal{L}}_k - v(\beta)) \leq \sum_{k=1}^{k_0}(\hat{\mathcal{L}}_k - v(\beta)) + (j-k_0)(\hat{\mathcal{L}}_{k_0+1} - v(\beta)) \quad \forall j > k_0$$

and hence

$$\lim_{j\to\infty}\sum_{k=1}^{j}(\hat{\mathcal{L}}_k - v(\beta)) = -\infty.$$

On the other hand, it follows from (8), (29), (43) and **(A4)** that

$$\begin{aligned}
\hat{\mathcal{L}}_k &= \mathcal{L}_\beta(x_1^k, x_2^k, y^k, \lambda^k) + \eta_k \geq \mathcal{L}_\beta(x_1^k, x_2^k, y^k, \lambda^k) \\
&= f_1(x_1^k) + f_2(x_2^k) + g(y^k) + \frac{\beta}{2}\|A_1x_1^k + A_2x_2^k + By^k - b\|^2 + \frac{1}{\beta\theta}\langle\lambda^k, \lambda^k - \lambda^{k-1}\rangle \\
&\geq v(\beta) + \frac{1}{2\beta\theta}\left(\|\lambda^k\|^2 - \|\lambda^{k-1}\|^2 + \|\lambda^k - \lambda^{k-1}\|^2\right) \geq v(\beta) + \frac{1}{2\beta\theta}\left(\|\lambda^k\|^2 - \|\lambda^{k-1}\|^2\right)
\end{aligned}$$

and hence that

$$\sum_{k=1}^{j}(\hat{\mathcal{L}}_k - v(\beta)) \geq \frac{1}{2\beta\theta}\left(\|\lambda^j\|^2 - \|\lambda^0\|^2\right) \geq -\frac{1}{2\beta\theta}\|\lambda^0\|^2 \quad \forall j \geq 1,$$

which yields the desired contradiction. □

References

[1] B. P. W. Ames and M. Hong. Alternating direction method of multipliers for penalized zero-variance discriminant analysis. *Comput. Optim. Appl.*, 64(3):725–754, 2016.

[2] C. Chen, B. He, Y. Ye, and X. Yuan. The direct extension of admm for multi-block convex minimization problems is not necessarily convergent. *Math. Programming*, 155(1):57–79, 2016.

[3] Y. Cui, X. Li, D. Sun, and K. C. Toh. On the convergence properties of a majorized ADMM for linearly constrained convex optimization problems with coupled objective functions. *J. Optim. Theory Appl.*, 169(3):1013–1041, 2016.

[4] W. Deng and W. Yin. On the global and linear convergence of the generalized alternating direction method of multipliers. *J. Sci. Comput.*, pages 1–28, 2015.

[5] E. X. Fang, B. He, H. Liu, and X. Yuan. Generalized alternating direction method of multipliers: new theoretical insights and applications. *Math. Prog. Comp.*, 7(2):149–187, 2015.

[6] P. A. Forero, A. Cano, and G. B. Giannakis. Distributed clustering using wireless sensor networks. *IEEE J. Selected Topics Signal Process.*, 5(4):707–724, 2011.

[7] D. Gabay and B. Mercier. A dual algorithm for the solution of nonlinear variational problems via finite element approximation. *Comput. Math. Appl.*, 2:17–40, 1976.

[8] R. Glowinski and A. Marroco. Sur l'approximation, par éléments finis d'ordre un, et la résolution, par penalisation-dualité, d'une classe de problèmes de dirichlet non linéaires. 1975.

[9] M. L. N. Gonçalves, J. G. Melo, and R. D. C. Monteiro. Extending the ergodic convergence rate of the proximal ADMM. *Arxiv preprint: https://arxiv.org/abs/1611.02903*, 2016.

[10] M. L. N. Goncalves, J. G. Melo, and R. D. C. Monteiro. Convergence rate bounds for a proximal ADMM with over-relaxation stepsize parameter for solving nonconvex linearly constrained problems. *Arxiv Preprint: https://arxiv.org/abs/1702.01850*, 2017.

[11] M. L. N. Gonçalves, J. G. Melo, and R. D. C. Monteiro. Improved pointwise iteration-complexity of a regularized ADMM and of a regularized non-euclidean HPE framework. *SIAM J. Optim.*, 27(1):379–407, 2017.

[12] Y. Gu, B. Jiang, and H. Deren. A semi-proximal-based strictly contractive Peaceman-Rachford splitting method. *Arxiv preprint: https://arxiv.org/abs/1506.02221*, 2015.

[13] K. Guo, D. R. Han, and T. T. Wu. Convergence of alternating direction method for minimizing sum of two nonconvex functions with linear constraints. *Int. J. Comput. Math. "DOI: 10.1080/00207160.2016.1227432"*, 2016.

[14] W. W. Hager, M. Yashtini, and H. Zhang. An $O(1/k)$ convergence rate for the variable stepsize Bregman operator splitting algorithm. *SIAM J. Numer. Anal.*, 54(3):1535–1556, 2016.

[15] D. Han and X. Yuan. A note on the alternating direction method of multipliers. *J. Optim. Theory Appl.*, 155(1):227–238, 2012.

[16] B. He, F. Ma, and X. Yuan. On the step size of symmetric alternating directions method of multipliers. *Preprint: http://www.optimization-online.org*, 2015.

[17] B. He, M. Tao, and X. Yuan. Convergence rate analysis for the alternating direction method of multipliers with a substitution procedure for separable convex programming. *Math. Oper. Res. "DOI:10.1287/moor.2016.0822"*, 2017.

[18] B. He and X. Yuan. On the $\mathcal{O}(1/n)$ convergence rate of the Douglas-Rachford alternating direction method. *SIAM Journal on Numer. Anal.*, 50(2):700–709, 2012.

[19] M. Hong and Z.-Q. Luo. On the linear convergence of the alternating direction method of multipliers. *Math. Programming*, 162(1):165–199, 2017.

[20] M. Hong, Z.-Q. Luo, and M. Razaviyayn. Convergence analysis of alternating direction method of multipliers for a family of nonconvex problems. *SIAM J. Optim.*, 26(1):337–364, 2016.

[21] B. Jiang, T. Lin, S. Ma, and S. Zhang. Structured nonconvex and nonsmooth optimization: algorithms and iteration complexity analysis. *Arxiv Preprint: https://arxiv.org/abs/1605.02408*, 2016.

[22] G. Li and T. K. Pong. Global convergence of splitting methods for nonconvex composite optimization. *SIAM J. Optim.*, 25(4):2434–2460, 2015.

[23] A.P. Liavas and N.D. Sidiropoulos. Parallel algorithms for constrained tensor factorization via the alternating direction method of multipliers. *Arxiv Preprint: https://arxiv.org/abs/1409.2383*, 2014.

[24] T. Lin, S. Ma, and S. Zhang. An extragradient-based alternating direction method for convex minimization. *Found. Comput. Math.*, pages 1–25, 2015.

[25] T. Lin, S. Ma, and S. Zhang. On the global linear convergence of the admm with multiblock variables. *SIAM J. Optim.*, 25(3):1478–1497, 2015.

[26] T. Lin, S. Ma, and S. Zhang. On the sublinear convergence rate of multi-block admm. *J. Oper. Res. China*, 3(3):251–274, 2015.

[27] T. Lin, S. Ma, and S. Zhang. Iteration complexity analysis of multi-block admm for a family of convex minimization without strong convexity. *J. Sci. Comput.*, 69(1):52–81, 2016.

[28] R. D. C. Monteiro and B. F Svaiter. Iteration-complexity of block-decomposition algorithms and the alternating direction method of multipliers. *SIAM J. Optim.*, 23(1):475–507, 2013.

[29] B.S. Mordukhovich. *Variational analysis and generalized differentiation I: basic theory*. Grundlehren der mathematischen Wissenschaften. Springer, Berlin,, 2006.

[30] Y. Ouyang, Y. Chen, G. Lan, and E. Pasiliao Jr. An accelerated linearized alternating direction method of multipliers. *SIAM J. Imaging Sci.*, 8(1):644–681, 2015.

[31] R. T. Rockafellar and R. J.-B. Wets. *Variational analysis*. Springer, Berlin, 1998.

[32] F. Wang, W. Cao, and Z. Xu. Convergence of multi-block Bregman ADMM for nonconvex composite problems. *Arxiv preprint: https://arxiv.org/abs/1505.03063*, 2015.

[33] F. Wang, Z. Xu, and H. K. Xu. Convergence of Bregman alternating direction method with multipliers for nonconvex composite problems. *Arxiv preprint:https://arxiv.org/abs/1410.8625*, 2014.

[34] W. Wang, Y. Yin and J. Zeng. Global convergence of ADMM in nonconvex nonsmooth optimization. *Arxiv preprint: https://arxiv.org/abs/1511.06324*, 2015.

[35] Z. Wen, X. Peng, X. Liu, X. Sun, and X. Bais. Asset allocation under the Basel accord risk measures. *Arxiv preprint: https://arxiv.org/abs/1308.1321*, 2013.

[36] Y. Xu, W. Yin, Z. Wen, and Y. Zhang. An alternating direction algorithm for matrix completion with nonnegative factors. *Frontiers Math. China*, 7(2):365–384, 2012.

[37] L. Yang, T. K. Pong, and X. Chen. Alternating direction method of multipliers for a class of nonconvex and nonsmooth problems with applications to background/foreground extraction. *SIAM J. Imaging Sci.*, 10(1):74–110, 2017.

[38] R. Zhang and J. T. Kwok. Asynchronous distributed admm for consensus optimization. *Proceedings of the 31st International Conference on Machine Learning*, 2014.