Iteration-complexity of a Jacobi-type non-Euclidean ADMM for multi-block linearly constrained nonconvex programs

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

May 13, 2017

Abstract

This paper establishes the iteration-complexity of a Jacobi-type non-Euclidean proximal alternating direction method of multipliers (ADMM) for solving multi-block linearly constrained nonconvex programs. The subproblems of this ADMM variant can be solved in parallel and hence the method has great potential to solve large scale multi-block linearly constrained nonconvex programs. Moreover, our analysis allows the Lagrange multiplier to be updated with a relaxation parameter in the interval $(0, 2)$.

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

Key words: Jacobi multiblock ADMM, nonconvex program, iteration-complexity, first-order methods, non-Euclidean distances.

1 Introduction

This paper considers the following linearly constrained optimization problem

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

where $f_i : \mathbb{R}^{n_i} \to (-\infty, \infty]$, $i = 1,\dots,p$, are proper lower semicontinuous functions, $A_i \in \mathbb{R}^{d\times n_i}$, $i = 1,\dots,p$, 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, 24, 39, 40, 42]). Recently, some 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.

*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.

In this paper we analyze the Jacobi-type proximal ADMM for solving (1), which recursively computes a sequence $\{(x_1^k, \cdots, x_p^k, \lambda^k)\}$ as

$$x_i^k = \mathrm{argmin}_{x_i} \left\{ \mathcal{L}_\beta(x_1^{k-1}, \ldots, x_{i-1}^{k-1}, x_i, x_{i+1}^{k-1}, \ldots, x_p^{k-1}, \lambda^{k-1}) + (dw_i)(x_i, x_i^{k-1}) \right\}, \quad i = 1, \ldots, p, \quad (2)$$

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

where $\beta > 0$ is a penalty parameter, $\theta > 0$ is a relaxation parameter, dw_i is a Bregman distance, and

$$\mathcal{L}_\beta(x_1, \ldots, x_p, \lambda) := \sum_{i=1}^p f_i(x_i) - \left\langle \lambda, \sum_{i=1}^p A_i x_i - b \right\rangle + \frac{\beta}{2} \left\| \sum_{i=1}^p A_i x_i - b \right\|^2 \quad (3)$$

is the augmented Lagrangian function for problem (1). An important feature of this ADMM variant is that the subproblems (2) can be solved in parallel and hence the method has great potential to solve large scale multi-block linearly constrained nonconvex programs. Under the assumption that A_p is full row rank and $f_p : \mathbb{R}^{n_p} \to \mathbb{R}$ is a differentiable function whose gradient is Lipschitz continuous, we establish an $\mathcal{O}(\rho^{-2})$ iteration-complexity bound for the Jacobi-type ADMM (2) to obtain $(x_1, \ldots, x_p, \lambda, r_1, \ldots, r_{p-1})$ satisfying

$$r_i \in \partial f_i(x_1, \ldots, x_p) - A_i^* \lambda, \quad i = 1, \cdots, p-1, \quad (4)$$

$$\max \left\{ \left\| \sum_{i=1}^p A_i x_i - b \right\|, \|r_1\|, \cdots, \|r_{p-1}\|, \left\| \nabla f_p(x_p) - A_p^* \lambda \right\| \right\} \le \rho \quad (5)$$

where ∂f_i denotes the limiting subdifferential (see for example [32, 34]).

We briefly discuss in this paragraph the development of ADMM in the convex setting. The standard ADMM (i.e., where $p = 2$, $w_i \equiv 0$ for $i = 1, \ldots, 2$ and x_2^k is obtained as above but with x_1^{k-1} replaced by x_1^k) was introduced in [7, 8] and its complexity analysis was first carried out in [31]. Since then several papers have obtained iteration-complexity results for various ADMM variants (see for example [2, 9, 12, 16, 18, 3, 5, 11, 14, 25, 33]). Multiblock ADMM variants have also been extensively studied (see for example [15, 19, 26, 27, 28, 17, 4, 37, 23]). In particular, papers [17, 4, 37, 23] study the convergence and/or complexity of Jacobi-type ADMM variants.

Recently, there have been a lot of interest on the study of ADMM variants for nonconvex problems (see, e.g., [13, 20, 21, 22, 35, 36, 38, 41, 10, 30, 29]). Papers [13, 22, 35, 36, 38, 41] 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 (3) 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. Paper [10] studies the iteration-complexity of a proximal ADMM for the two block optimization problem, i.e., $p = 2$, and the relaxation parameter θ is arbitrarily chosen in the interval $(0, 2)$, contrary to the previous related literature where this parameter is considered as one or at most

$(\sqrt{5}+1)/2$. Paper [30] analyzes the iteration-complexity of a multi-block proximal ADMM via a general linearization scheme. Finally, while the authors were in the process of finalizing this paper, they have learned of the recent paper [29] which studies the asymptotic convergence of a Jacobi-type linearized ADMM for solving non-convex problems. The latter paper though does not deal with the issue of iteration-complexity and considers the case of $\theta = 1$ only.

Our paper is organized as follows. Subsection 1.1 contains some notation and basic results used in the paper. Section 2 describes our assumptions and contains two subsections. Subsection 2.1 introduces the concept of distance generating functions (and its corresponding Bregman distances) considered in this paper, and formally states the non-Euclidean Jacobi-type ADMM. Section 2.2 is devoted to the convergence rate analysis of the latter method. Our main convergence rate result is in this subsection (Theorem 2.11). The appendix contains proofs of some results stated in the paper.

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 $x \in \mathbb{R}^s$.

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 [32, 34].

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

2 Jacobi-type non-Euclidean proximal ADMM and its convergence rate

We start by recalling the definition of critical points of (1).

Definition 2.1. *An element* $(x_1^*, \ldots, x_p^*, \lambda^*) \in \mathbb{R}^{n_1} \times \ldots \times \mathbb{R}^{n_p} \times \mathbb{R}^d$ *is a critical point of problem* (1) *if*

$$0 \in \partial f_i(x_i^*) - A_i^* \lambda^*, \quad i = 1, \ldots, p, \quad \sum_{i=1}^{p} A_i x_i^* = b.$$

Under some mild conditions, it can be shown that if $(x_1^*, \ldots, x_p^*)$ is a global minimum of (1), then there exists λ^* such that $(x_1^*, \ldots, x_p^*, \lambda^*)$ is a critical point of (1).

The augmented Lagrangian associated with problem (1) and with penalty parameter $\beta > 0$ is defined as

$$\mathcal{L}_\beta(x_1, \ldots, x_p, \lambda) := \sum_{i=1}^{p} f_i(x_i) - \left\langle \lambda, \sum_{i=1}^{p} A_i x_i - b \right\rangle + \frac{\beta}{2} \left\| \sum_{i=1}^{p} A_i x_i - b \right\|^2 . \tag{6}$$

We assume that problem (1) satisfies the following set of conditions:

(A0) The functions f_i, $i = 1, \ldots, p-1$, are proper lower semicontinuous;

(A1) $A_p \neq 0$ and $\mathrm{Im}(A_p) \supset \{b\} \cup \mathrm{Im}(A_1) \cup \ldots \cup \mathrm{Im}(A_{p-1})$;

(A2) $f_p : \mathbb{R}^{n_p} \to \mathbb{R}$ is differentiable with gradient L_p-Lipschitz continuous.

(A3) there exists $\bar{\beta} \geq 0$ such that $v(\bar{\beta}) > -\infty$ where

$$v(\beta) := \inf_{(x_1, \ldots, x_p)} \left\{ \sum_{i=1}^{p} f_i(x_i) + \frac{\beta}{2} \left\| \sum_{i=1}^{p} A_i x_i - b \right\|^2 \right\} \quad \forall \beta \in \mathbb{R}.$$

2.1 The non-Euclidean proximal Jacobi ADMM

In this subsection, we introduce a class of distance generating functions (and its corresponding Bregman distances) which is suitable for our study. We also formally describe the non-Euclidean proximal Jacobi ADMM for solving problem (1).

Definition 2.2. *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*

$$w(z') - w(z) - \langle \nabla w(z), z' - z \rangle \geq \frac{m}{2} \|z - z'\|^2 \quad \forall z, z' \in Z, \tag{7}$$

$$\|\nabla w(z) - \nabla w(z')\| \leq M \|z - z'\| \quad \forall z, z' \in Z. \tag{8}$$

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*

$$(dw)(z'; z) := w(z') - w(z) - \langle \nabla w(z), z' - z \rangle \quad \forall (z', z) \in \mathbb{R}^s \times Z. \tag{9}$$

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,

$$\nabla (dw)_z(z') = -\nabla (dw)_{z'}(z) = \nabla w(z') - \nabla w(z) \quad \forall z, z' \in Z, \tag{10}$$

We now state the non-Euclidean proximal Jacobi ADMM based on the class of distance generating functions introduced in Definition 2.2. In its statement and in some technical results, we denote the block of variables $(x_1, \ldots, x_{i-1})$ simply by $x_{<i}$ and the block of variables $(x_{i+1}, \ldots, x_p)$ simply by $x_{>i}$. Hence, the whole vector $(x_1, \ldots, x_p)$ can also be denoted as $(x_{<i}, x_i, x_{>i})$ when there is a need to emphasize the i-th block. For convenience, we also extend the above for notation for $i = 1$ and $i = p$. Hence, $(x_{<1}, x_1, x_{>1}) = (x_{<p}, x_p, x_{>p}) = (x_1, \ldots, x_p)$.

Non-Euclidean Proximal Jacobi ADMM (NEPJ-ADMM)

---

(0) Define $Z_i := \operatorname{dom} f_i$ for $i = 1, \ldots, p$, and let $\bar\beta$ be as in **(A3)**. Let an initial point $(x_1^0, \ldots, x_p^0, \lambda^0) \in Z_1 \times \ldots \times Z_p \times \mathbb{R}^d$. Choose scalars $\alpha > 0$, $\beta \geq \bar\beta$, $M_i \geq m_i > 0$, $i = 1, \ldots, p$, and a stepsize parameter $\theta \in (0, 2)$ such that

$$\delta_i := \frac{m_i}{4} - \left(\frac{p - 2 + \alpha}{2} + \frac{2\gamma_\theta (p+1)}{\sigma_{A_p}^+} \|A_p^*\|^2 \right) \beta \max_{1 \leq l \leq p-1} \|A_l\|^2 > 0, \quad i = 1, \ldots, p-1$$

$$\delta_p := \frac{m_p}{4} - \left(\frac{\beta (p-1) \|A_p\|^2}{2\alpha} + \frac{\gamma_\theta (p+1)(L_p^2 + 2M_p^2)}{\beta \sigma_{A_p}^+} \right) > 0 \tag{11}$$

where σ_{A_p} (resp., $\sigma_{A_p}^+$) denotes the smallest eigenvalue (resp., positive eigenvalue) of $A_p^* A_p$, and γ_θ is given by

$$\gamma_\theta := \frac{\theta}{(1 - |\theta - 1|)^2}. \tag{12}$$

Set $k = 1$ and go to step 1.

(1) For each $i = 1, \ldots, p$, choose $w_i^k \in \mathcal{D}_{Z_i}(m_i, M_i)$ and compute an optimal solution $x_i^k \in \mathbb{R}^{n_i}$ of

$$\min_{x_i \in \mathbb{R}^{n_i}} \left\{ \mathcal{L}_\beta (x_{<i}^{k-1}, x_i, x_{>i}^{k-1}, \lambda^{k-1}) + (dw_i^k)_{x_i^{k-1}}(x_i) \right\}. \tag{13}$$

(2) Set

$$\lambda^k = \lambda^{k-1} - \theta\beta \left[\sum_{i=1}^p A_i x_i^k - b \right], \tag{14}$$

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

end

Some comments about the NEPJ-ADMM are in order. First, it is always possible to choose the constants m_i, i=1,...,p, sufficiently large so as to guarantee that δ_i, i=1,...,p, are strictly positive. Second, one of the main features of NEPJ-ADMM is that its subproblems (13) are completely independent of one another. As a result, they can all be solved in parallel which shows the potential of NEPJ-ADMM as a suitable ADMM variant to solve large instance of (1). Third, as in the papers [10, 30], NEPJ-ADMM allows the choice of a relaxation parameter $\theta \in (0, 2)$.

2.2 Convergence Rate Analysis of the NEPJ-ADMM

This subsection is dedicated to the convergence rate analysis of the NEPJ-ADMM.

We first present some technical lemmas which are useful to prove our main result (Theorem 2.11). To simplify the notation, we denote by x^k the vector $(x_1^k, \ldots, x_p^k)$ generated by the NEPJ-ADMM.

Lemma 2.3. *Consider the sequence $\{(x^k, \lambda^k)\}$ generated by the NEPJ-ADMM. For every $k \geq 1$, define*

$$\hat{\lambda}^k := \lambda^{k-1} - \beta \left(\sum_{i=1}^{p} A_i x_i^k - b \right) \tag{15}$$

and

$$R_i^k := - \sum_{j=1, j \neq i}^{p} \beta A_i^* A_j \Delta x_j^k + \Delta w_i^k, \qquad i = 1, \ldots, p, \tag{16}$$

where

$$\Delta x_i^k := x_i^k - x_i^{k-1}, \qquad \Delta w_i^k := \nabla w_i^k(x_i^k) - \nabla w_i^k(x_i^{k-1}), \qquad i = 1, \ldots, p. \tag{17}$$

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

$$0 \in \partial f_i(x_i^k) - A_i^* \hat{\lambda}^k + R_i^k \quad i = 1, \ldots, p \tag{18}$$

$$0 = \left[\sum_{i=1}^{p} A_i x_i^k - b \right] + \frac{1}{\theta \beta} \Delta \lambda^k \tag{19}$$

where $\Delta \lambda^k := \lambda^k - \lambda^{k-1}$.

Proof. The optimality conditions (see Proposition 1.3) for (13) imply that

$$0 \in \partial f_i(x_i^k) - A_i^* \left[\lambda_{k-1} - \beta \left(A_i x_i^k + \sum_{j=1, j \neq i}^{p} A_j x_j^{k-1} - b \right) \right] + \Delta w_i^k, \quad i = 1, \ldots, p.$$

This relation combined with (15) and (16) immediately yield (18). Relation (19) follows directly from (14). □

Next result presents a recursive relation involving the displacements $\Delta \lambda^k$ and $\Delta \lambda^{k-1}$.

Lemma 2.4. *Consider the sequence $\{(x^k,\lambda^k)\}$ generated by the NEPJ-ADMM and define*

$$R_p^0 = A_p^*\lambda^0 - \nabla f_p(x_p^0), \quad \Delta\lambda^0 = 0. \tag{20}$$

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

$$A_p^*\Delta\lambda^k = (1-\theta)A_p^*\Delta\lambda^{k-1} + \theta u^k, \tag{21}$$

where

$$u^k := \Delta f_p^k + \Delta R_p^k, \quad \Delta f_p^k := \nabla f_p^k(x_p^k) - \nabla f_p^k(x_p^{k-1}), \quad \Delta R_p^k := R_p^k - R_p^{k-1} \quad \forall k \geq 1, \tag{22}$$

$\Delta\lambda^k$ and R_p^k are as in Lemma 2.3.

Proof. From (15) and (19), we obtain the following relation

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

Using this relation and (18) with $i = p$, we have

$$A_p^*\lambda^k = (1-\theta)A_p^*\lambda^{k-1} + \theta[\nabla f_p(x_p^k) + R_p^k], \quad \forall k \geq 1. \tag{23}$$

Hence, in view of (22), relation (21) holds for every $k \geq 2$. Now, note that (20) is equivalent to $\nabla f_p(x_p^0) + R_p^0 = A_p^*\lambda^0$. This relation combined with (22) and (23), both with $k = 1$, yield

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

Hence, in view of $\Delta\lambda^0 = 0$, relation (21) also holds for $k = 1$. □

Next we consider an auxiliary result to be used to compare consecutive terms of the sequence $\{\mathcal{L}_\beta(x^k,\lambda^k)\}$. See the comments immediately before the NEPJ-ADMM about the notation used hereafter.

Lemma 2.5. *For every $y^0 = (y_1^0,\ldots,y_p^0)$, $y = (y_1,\ldots,y_p) \in \operatorname{dom} f_1 \times \ldots \times \operatorname{dom} f_p$, $\lambda \in \mathbb{R}^d$ and $i = 2,\ldots,p$, we have*

$$\begin{aligned} \mathcal{L}_\beta(y_{<i},y_i,y_{>i}^0,\lambda) - \mathcal{L}_\beta(y_{<i},y_i^0,y_{>i}^0,\lambda) = \mathcal{L}_\beta(y_{<i}^0,y_i,y_{>i}^0,\lambda) - \mathcal{L}_\beta(y_{<i}^0,y_i^0,y_{>i}^0,\lambda) \\ + \beta\sum_{j=1}^{i-1}\langle A_i\Delta y_i, A_j\Delta y_j\rangle. \end{aligned}$$

Proof. It is easy to see thatf the gradient of the function

$$y_{<i} \mapsto \mathcal{L}_\beta(y_{<i},y_i,y_{>i}^0) - \mathcal{L}_\beta(y_{<i},y_i^0,y_{>i}^0) \tag{24}$$

is given by

$$\beta[A_1\cdots A_{i-1}]^*A_i\Delta y_i$$

and its Hessian equal to zero everywhere in $\operatorname{dom} f_1 \times \ldots \times \operatorname{dom} f_{i-1}$. Hence, the function given in (24) is affine. The conclusion of the lemma now follows by noting that

$$\langle [A_1 \cdots A_{i-1}]^* A_i \Delta y_i, \Delta y_{<i} \rangle = \sum_{j=1}^{i-1} \langle A_i \Delta y_i, A_j \Delta y_j \rangle.$$

□

The next result compares consecutive terms of the sequence $\{\mathcal{L}_\beta(x^k, \lambda^k)\}$.

Lemma 2.6. *For every $k \geq 1$, we have*

$$\mathcal{L}_\beta(x^k, \lambda^k) - \mathcal{L}_\beta(x^{k-1}, \lambda^{k-1}) \leq \sum_{1 \leq j < i \leq p} \beta \langle A_i \Delta x_i^k, A_j \Delta x_j^k \rangle - \sum_{i=1}^{p} \frac{m_i}{2} \|\Delta x_i^k\|^2 + \frac{1}{\theta\beta} \|\Delta \lambda^k\|^2.$$

Proof. First note that (13) together with the fact that $w_i^k \in \mathcal{D}_{Z_i}(m_i, M_i)$ imply that

$$\mathcal{L}_\beta(x_{<i}^{k-1}, x_i^k, x_{>i}^{k-1}, \lambda^{k-1}) - \mathcal{L}_\beta(x^{k-1}, \lambda^{k-1}) \leq -m_i \|\Delta x_i^k\|^2/2, \qquad i = 1, \ldots, p.$$

Hence, using Lemma 2.5 with $y^0 = x^{k-1}$, $y = x^k$ and $\lambda = \lambda^{k-1}$, we see that

$$\begin{aligned}
&\mathcal{L}_\beta(x_{<i}^{k}, x_i^k, x_{>i}^{k-1}, \lambda^{k-1}) - \mathcal{L}_\beta(x_{<i}^{k}, x_i^{k-1}, x_{>i}^{k-1}, \lambda^{k-1}) \\
&\qquad = \mathcal{L}_\beta(x_{<i}^{k-1}, x_i^k, x_{>i}^{k-1}, \lambda^{k-1}) - \mathcal{L}_\beta(x^{k-1}, \lambda^{k-1}) + \beta \sum_{j=1}^{i-1} \langle A_i \Delta x_i^k, A_j \Delta x_j^k \rangle \\
&\qquad \leq -\frac{m_i}{2} \|\Delta x_i^k\|^2 + \beta \sum_{j=1}^{i-1} \langle A_i \Delta x_i^k, A_j \Delta x_j^k \rangle.
\end{aligned}$$

Hence

$$\begin{aligned}
\mathcal{L}_\beta(x^k, \lambda^{k-1}) - \mathcal{L}_\beta(x^{k-1}, \lambda^{k-1}) &= \sum_{i=1}^{p} \left[\mathcal{L}_\beta(x_{<i}^{k}, x_i^k, x_{>i}^{k-1}, \lambda^{k-1}) - \mathcal{L}_\beta(x_{<i}^{k}, x_i^{k-1}, x_{>i}^{k-1}, \lambda^{k-1}) \right] \\
&\leq -\sum_{i=1}^{p} \frac{m_i}{2} \|\Delta x_i^k\|^2 + \beta \sum_{i=2}^{p} \sum_{j=1}^{i-1} \langle A_i \Delta x_i^k, A_j \Delta x_j^k \rangle. \qquad (25)
\end{aligned}$$

On the other hand, due to $\Delta\lambda^k = \lambda^k - \lambda^{k-1}$ and (14), we have

$$\mathcal{L}_\beta(x^k, \lambda^k) - \mathcal{L}_\beta(x^k, \lambda^{k-1}) = -\left\langle \lambda^k - \lambda^{k-1}, \sum_{i=1}^{p} A_i x_i^k - b \right\rangle = \frac{1}{\beta\theta} \|\Delta\lambda^k\|^2.$$

To conclude the proof, just add the last relation and (25). □

Lemma 2.6 is essential to show that a certain sequence $\{\hat{\mathcal{L}}_k\}$ associated to $\{\mathcal{L}_\beta(x^k, \lambda^k)\}$ is monotonically decreasing. This sequence is defined as

$$\hat{\mathcal{L}}_k := \mathcal{L}_\beta(x^k, \lambda^k) + \eta_k \qquad \forall k \geq 0, \tag{26}$$

where

$$\eta_0 := \frac{m_p}{4M_p^2} \|A_p^* \lambda^0 - \nabla f_p(x_p^0)\|^2 \tag{27}$$

$$\eta_k := \sum_{i=1}^{p} \frac{m_i}{4} \|\Delta x_i^k\|^2 + \frac{c_1}{2} \|A_p^* \Delta \lambda^k\|^2 \qquad \forall k \geq 1, \tag{28}$$

$$c_1 := \frac{2|\theta - 1|}{\beta\theta(1 - |\theta - 1|)\sigma_B^+} \geq 0. \tag{29}$$

Before establishing the monotonicity property of the sequence $\{\hat{\mathcal{L}}_k\}$, we first present an upper bound on $\hat{\mathcal{L}}_k - \hat{\mathcal{L}}_{k-1}$ in terms of some quantities related to $\Delta x_1^k, \ldots, \Delta x_p^k$, and $\Delta \lambda^k$.

Lemma 2.7. *For any $k \geq 1$, there holds*

$$\hat{\mathcal{L}}_k - \hat{\mathcal{L}}_{k-1} \leq \sum_{i=1}^{p-1} \left(\frac{(p-2+\alpha)\beta\|A_i\|^2}{2} - \frac{m_i}{4} \right) \|\Delta x_i^k\|^2 + \Theta_\lambda^k + \Theta_p^k \tag{30}$$

where

$$\Theta_\lambda^k := \frac{1}{\beta\theta} \|\Delta \lambda^k\|^2 + \frac{c_1}{2} \left(\|A_p^* \Delta \lambda^k\|^2 - \|A_p^* \Delta \lambda^{k-1}\|^2 \right) \tag{31}$$

$$\Theta_p^k := \left(\frac{(p-1)\beta\|A_p\|^2}{2\alpha} - \frac{m_p}{4} \right) \left(\|\Delta x_p^k\|^2 + \|\Delta x_p^{k-1}\|^2 \right) \tag{32}$$

and $\Delta \lambda^0 = 0$, $\Delta x_p^0 = R_p^0 / M_p$ (see Lemma 2.4).

Proof. From Lemma 2.6 and definitions of $\hat{\mathcal{L}}_k$ and Θ_λ^k, we obtain

$$\hat{\mathcal{L}}_k - \hat{\mathcal{L}}_{k-1} \leq \sum_{1 \leq j < i < p} \beta \|A_i \Delta x_i^k\| \|A_j \Delta x_j^k\| + \sum_{i=1}^{p-1} \beta \|A_i \Delta x_i^k\| \|A_p \Delta x_p^k\| - \sum_{i=1}^{p-1} \frac{m_i}{2} \|\Delta x_i^k\|^2 + \Theta_\lambda^k \tag{33}$$

$$- \frac{m_p}{4} \left(\|\Delta x_p^k\|^2 + \|\Delta x_p^{k-1}\|^2 \right) \tag{34}$$

$$\leq \sum_{1 \leq j < i < p} \left(\frac{\beta}{2} \|A_i \Delta x_i^k\|^2 + \frac{\beta}{2} \|A_j \Delta x_j^k\|^2 \right) + \sum_{i=1}^{p-1} \frac{\alpha\beta}{2} \|A_i \Delta x_i^k\|^2 + \frac{(p-1)\beta}{2\alpha} \|A_p \Delta x_p^k\|^2 \tag{35}$$

$$- \sum_{i=1}^{p-1} \frac{m_i}{2} \|\Delta x_i^k\|^2 + \Theta_\lambda^k - \frac{m_p}{4} \left(\|\Delta x_p^k\|^2 + \|\Delta x_p^{k-1}\|^2 \right) \tag{36}$$

$$\leq \sum_{i=1}^{p-1} \left(\frac{(p-2+\alpha)\beta\|A_i\|^2}{2} - \frac{m_i}{2} \right) \|\Delta x_i^k\|^2 + \Theta_\lambda^k + \Theta_p^k \tag{37}$$

where the inequalities are due to Cauchy-Schwarz inequality and by using the relation $2s_1 s_2 \leq t s_1^2 + (1/t) s_2^2$, $s_1, s_2 \in \mathbb{R}$ for $t = 1$ and $t = \alpha$, respectively.

□

The next result compares Θ_λ^k with $\|u_k\|$, defined in (31) and (22), respectively, and provides an upper bound for both elements in terms of $(\Delta x_1^k, \ldots, \Delta x_p^k)$.

Lemma 2.8. *Consider Θ_λ^k as in (31) and u^k as in (22). Then,*

$$\Theta_\lambda^k \leq \frac{\gamma_\theta}{\beta\sigma_{A_p}^+}\|u_k\|^2 \tag{38}$$

$$\leq \frac{\gamma_\theta(p+1)}{\beta\sigma_{A_p}^+}\left[\sum_{j=1}^{p-1}\beta^2\|A_p^*A_j\|^2(\|\Delta x_j^k\| + \|\Delta x_j^{k-1}\|)^2 + \left(L_p^2 + M_p^2\right)\left(\|\Delta x_p^k\| + \|\Delta x_p^{k-1}\|\right)^2\right].$$

where γ_θ is as in (12) and $\Delta x_i^0 = 0, i = 1, \ldots, p-1$, and $\Delta x_p^0 = R_p^0/M_p$ (see Lemma 2.4).

Proof. The proof of this lemma is given in Appendix A. □

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

Proposition 2.9. *Let $\Delta x_i^0 = 0, i = 1, \ldots, p-1$, and $\Delta x_p^0 = R_p^0/M_p$. Then, the following statements hold:*

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

$$\hat{\mathcal{L}}_k - \hat{\mathcal{L}}_{k-1} \leq -\sum_{i=1}^{p}\delta_i(\|\Delta x_i^k\|^2 + \|\Delta x_i^{k-1}\|^2);$$

(b) the sequence $\{\hat{\mathcal{L}}_k\}$ given in (26) satisfies $\hat{\mathcal{L}}_k \geq v(\beta)$ for every $k \geq 0$;

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

$$\sum_{j=1}^{k}\sum_{i=1}^{p}\delta_i(\|\Delta x_i^j\|^2 + \|\Delta x_i^{j-1}\|^2) \leq \hat{\mathcal{L}}_0 - v(\beta)$$

*where $v(\beta)$ and δ_i are as in (**A3**) and (11), respectively.*

Proof. (a) It follows from (30), Lemma 2.8, (11) and (32) that

$$\hat{\mathcal{L}}_k - \hat{\mathcal{L}}_{k-1} \leq -\sum_{i=1}^{p-1}\left[\frac{m_i}{4} - \left(\frac{p-2+\alpha}{2} + \frac{2\gamma_\theta(p+1)}{\sigma_{A_p}^+}\|A_p^*\|^2\right)\beta\max_{1\leq j\leq p-1}\|A_i\|^2\right]\left(\|\Delta x_i^k\|^2 + \|\Delta x_i^{k-1}\|^2\right)$$

$$-\left[\frac{m_p}{4} - \left(\frac{\beta(p-1)\|A_p\|^2}{2\alpha} + \frac{\gamma_\theta(p+1)(L_p^2 + 2M_p^2)}{\beta\sigma_{A_p}^+}\right)\right]\left(\|\Delta x_p^k\|^2 + \|\Delta x_p^{k-1}\|^2\right)$$

$$= -\sum_{i=1}^{p}\delta_i\left(\|\Delta x_i^k\|^2 + \|\Delta x_i^{k-1}\|^2\right),$$

proving (a). The proof of (b) is given Appendix B. The proof of (c) follows immediately from (a) and (b). □

Next proposition presents some convergence rate bounds for the displacements Δx_i^k, $i=1,\dots,p$, and $\Delta\lambda^k$ in terms of some initial parameters. Our main result will follow easily from this proposition, due to the fact that the residual generated by $(x^k,\hat\lambda^k)$ in order to satisfy the Lagrangian system (2.1) (see Lemma 2.3) can be controlled by these displacements.

Proposition 2.10. *Let δ_i, i=1,...,p, be as in* (11) *and define*

$$\delta_\lambda := \left[\frac{\theta\gamma_\theta(p+1)}{\sigma^+_{A_p}\min\limits_{1\le l\le p}\delta_l}\left(2\beta^2\|A_p^*\|^2\max_{1\le l\le p-1}\|A_l\|^2+L_p^2+2M_p^2\right)\right]^{-1} \tag{39}$$

where $\Delta\mathcal{L}_0 := \hat{\mathcal{L}}_0 - v(\beta)$ (see (26) and (**A3**)). *Then, for every $k\ge 1$, we have*

$$\sum_{j=1}^k\left\{\left[\sum_{i=1}^p\delta_i\left(\|\Delta x_i^j\|^2+\|\Delta x_i^{j-1}\|^2\right)\right]+\delta_\lambda\|\Delta\lambda^j\|^2\right\}\le 2\Delta\mathcal{L}_0 \tag{40}$$

and there exists $j\le k$ such that

$$\|\Delta x_i^j\|\le\sqrt{\frac{2\Delta\mathcal{L}_0}{k\delta_i}},\ i=1,\dots,p,\quad \|\Delta\lambda^j\|\le\sqrt{\frac{2\Delta\mathcal{L}_0}{k\delta_\lambda}}. \tag{41}$$

Proof. It follows from Proposition 2.9(c) that

$$\sum_{j=1}^k\sum_{i=1}^p\left(\|\Delta x_i^j\|^2+\|\Delta x_i^{j-1}\|^2\right)\le\frac{\Delta\mathcal{L}_0}{\min\limits_{1\le i\le p}\delta_i} \tag{42}$$

and that in order to prove (40), it suffices to show that

$$\sum_{j=1}^k\|\Delta\lambda^j\|^2\le\frac{\Delta\mathcal{L}_0}{\delta_\lambda}. \tag{43}$$

Then, in the remaining part of the proof we will show that (43) holds. By rewriting (31), we have

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

Hence, due to $\Delta\lambda^0=0$ and Lemma 2.8, we obtain

$$\begin{aligned}\sum_{j=1}^k\|\Delta\lambda^j\|^2 &\le \beta\theta\sum_{j=1}^k\Theta_\lambda^j\le\frac{\theta\gamma_\theta}{\sigma^+_{A_p}}\sum_{j=1}^k\|u^j\|^2\\ &\le\frac{\theta\gamma_\theta(p+1)}{\sigma^+_{A_p}}\left[2\beta^2\|A_p^*\|^2\max_{1\le l\le p-1}\|A_l\|^2\sum_{j=1}^k\sum_{i=1}^{p-1}\left(\|\Delta x_i^j\|^2+\|\Delta x_i^{j-1}\|^2\right)\right]\\ &\quad+\frac{\theta\gamma_\theta(p+1)}{\sigma^+_{A_p}}(L_p^2+2M_p^2)\sum_{j=1}^k\left(\|\Delta x_p^j\|^2+\|\Delta x_p^{j-1}\|^2\right).\\ &\le\frac{\theta\gamma_\theta(p+1)}{\sigma^+_{A_p}}\left(2\beta^2\|A_p^*\|^2\max_{1\le l\le p-1}\|A_l\|^2+L_p^2+2M_p^2\right)\frac{\Delta\mathcal{L}_0}{\min\limits_{1\le i\le p}\delta_i}\end{aligned}$$

where the fourth inequality is due to (42). It is now to verify that the previous estimate and (39) imply (43), which in turn implies (40). ☐

We now present the main convergence rate result for the NEPJ-ADMM. Its main conclusion is that the NEPJ-ADMM generates an element $(\bar{x}_1, \ldots, \bar{x}_p, \bar{\lambda})$ which satisfies the optimality conditions of Definition 2.1 within an error of $\mathcal{O}(1/\sqrt{k})$.

Theorem 2.11. *Let $\Delta\mathcal{L}_0 := \mathcal{L}_\beta(x^0, \lambda^0) - v(\beta) + \eta_0$ where η_0 and $v(\beta)$ are as in (27) and* **(A3)***, respectively. Let $\hat{\lambda}^k$ and R_i^k, $i = 1, \ldots, p$, be as in (15) and (16), respectively. Consider δ_i, $i = 1, \ldots, p$, as in (11) and let δ_λ be as in (39). Then, the following statements hold:*

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

b) for every $k \geq 1$,

$$0 \in \partial f_i(x_i^k) - A_i^* \hat{\lambda}^k + R_i^k \qquad i = 1, \ldots, p,$$

and there exists $j \leq k$ such that

$$\|R_i^j\| \leq \left[\sum_{l=1, l\neq i}^{p} \beta \|A_i^* A_l\| + M_i \right] \sqrt{\frac{2\Delta\mathcal{L}_0}{k \min\limits_{1\leq l\leq p} \delta_l}}, \qquad i = 1, \ldots, p,$$

$$\left\| \sum_{i=1}^{p} A_p x_i^j - b \right\| \leq \frac{1}{\beta\theta} \sqrt{\frac{2\Delta\mathcal{L}_0}{k\delta_\lambda}}.$$

Proof. (a) holds due to Proposition 2.9(c). Lemma 2.3 shows that the first statement of (b) holds. Now, it follows from (16), (19) and the fact that $w_i^k \in \mathcal{D}_{Z_i}(m_i, M_i)$, $i = 1, \ldots, p$, that

$$\|R_i^k\| \leq \sum_{l=1, l\neq i}^{p} \beta \|A_i^* A_l\| \|\Delta x_l^k\| + M_i \|\Delta x_i^k\|, \qquad i = 1, \ldots, p,$$

$$\left\| \sum_{i=1}^{p} A_i x_i^k - b \right\| = \frac{1}{\beta\theta} \|\Delta\lambda^k\|.$$

Hence, to end the proof, just combine the above relations with (41). ☐

A Proof of Lemma 2.8

Let us first prove the first inequality (38). Assumption **(A1)** clearly implies that

$$\Delta\lambda_k = -\beta\theta \left(\sum_{i=1}^{p} A_i x_i - b \right) \in \operatorname{Im}(A_p).$$

Hence, it follows from Lemma 1.1 that

$$\|\Delta\lambda_k\| = \|\mathcal{P}_{A_p}(\Delta\lambda_k)\| \leq \frac{1}{\sqrt{\sigma_{A_p}^+}} \|A_p^* \Delta\lambda_k\|.$$

Thus, in view of (21) and (31), we have

$$\begin{aligned}\Theta_\lambda^k &\le \frac{1}{\beta\theta\sigma^+_{A_p}}\|A_p^*\Delta\lambda_k\|^2 + \frac{c_1}{2}\left(\|A_p^*\Delta\lambda_k\|^2 - \|A_p^*\Delta\lambda_{k-1}\|^2\right)\\ &= \left(\frac{1}{\beta\theta\sigma^+_{A_p}} + \frac{c_1}{2}\right)\|(1-\theta)A_p^*\Delta\lambda_{k-1} + \theta u_k\|^2 - \frac{c_1}{2}\|A_p^*\Delta\lambda_{k-1}\|^2.\end{aligned}$$

Note that if $\theta = 1$, then (29) implies that $c_1 = 0$ and the above inequality proves the first inequality of the lemma. We will now prove the first inequality of the lemma for the case in which $\theta \neq 1$. The previous inequality together with the relation $\|s_1 + s_2\|^2 \le (1+t)\|s_1\|^2 + (1+1/t)\|s_2\|^2$ which holds for every $s_1, s_2 \in \mathbb{R}^m$ and $t > 0$ yield

$$\begin{aligned}\Theta_\lambda^k &\le \left(\frac{1}{\beta\theta\sigma^+_{A_p}} + \frac{c_1}{2}\right)\left[(1+t)(\theta-1)^2\|A_p^*\Delta\lambda_{k-1}\|^2 + \left(1+\frac{1}{t}\right)\theta^2\|u_k\|^2\right] - \frac{c_1}{2}\|A_p^*\Delta\lambda_{k-1}\|^2\\ &= \left[\left(\frac{1}{\beta\theta\sigma^+_{A_p}} + \frac{c_1}{2}\right)(1+t)(\theta-1)^2 - \frac{c_1}{2}\right]\|A_p^*\Delta\lambda_{k-1}\|^2 + \left(\frac{1}{\beta\theta\sigma^+_{A_p}} + \frac{c_1}{2}\right)\left(1+\frac{1}{t}\right)\theta^2\|u_k\|^2\\ &= \left\{\frac{(1+t)(\theta-1)^2}{\beta\theta\sigma^+_{A_p}} - \left[1-(1+t)(\theta-1)^2\right]\frac{c_1}{2}\right\}\|A_p^*\Delta\lambda_{k-1}\|^2 + \left(\frac{1}{\beta\theta\sigma^+_{A_p}} + \frac{c_1}{2}\right)\left(1+\frac{1}{t}\right)\theta^2\|u_k\|^2.\end{aligned}$$

Using the above expression with $t = -1 + 1/|\theta-1|$ and noting that $t > 0$ in view of the assumption that $\theta \in (0,2)$, we conclude that

$$\begin{aligned}\Theta_\lambda^k &\le \left[\frac{1}{\beta\theta\sigma^+_{A_p}}|\theta-1| - (1-|\theta-1|)\frac{c_1}{2}\right]\|A_p^*\Delta\lambda_{k-1}\|^2 + \left(\frac{1}{\beta\theta\sigma^+_{A_p}} + \frac{c_1}{2}\right)\frac{\theta^2}{1-|\theta-1|}\|u_k\|^2\\ &= \frac{1}{\beta\theta\sigma^+_{A_p}}\left(1 + \frac{|\theta-1|}{1-|\theta-1|}\right)\frac{\theta^2}{1-|\theta-1|}\|u_k\|^2\end{aligned}$$

where the last equality is due to (29). Hence, in view of (12), the first inequality of the lemma is proved.

We now prove the second inequality in (38). Due to $R_p^0 = M_p\Delta x_p^0$, $w_p^k \in \mathcal{D}_{\mathbb{R}^{n_p}}(m_p, M_p)$, assumption **(A2)**, and relation (22), we obtain

$$\begin{aligned}\|u^k\|^2 &= \left\|\Delta f_p^k + \Delta R_p^k\right\|^2\\ &\le \left[L_p\|\Delta x_p^k\| + \sum_{j=1}^{p-1}\beta\|A_p^*A_j\|\left(\|\Delta x_j^k\| + \|\Delta x_j^{k-1}\|\right) + M_p\left(\|\Delta x_p^k\| + \|\Delta x_p^{k-1}\|\right)\right]^2\\ &\le (p+1)\left[L_p^2\|\Delta x_p^k\|^2 + \sum_{j=1}^{p-1}\beta^2\|A_p^*A_j\|^2\left(\|\Delta x_j^k\| + \|\Delta x_j^{k-1}\|\right)^2 + M_p^2\left(\|\Delta x_p^k\| + \|\Delta x_p^{k-1}\|\right)^2\right]\end{aligned}$$

where the inequalities follow from the triangle inequality for norms, definition of ΔR_p^k in (22), and the relation $\left(\sum_{i=1}^l s_i\right)^2 \le l\left(\sum_{i=1}^l s_i^2\right)$ for $s_i \in \mathbb{R}$, $i = 1,\ldots,l$. Hence the proof of Lemma 2.8 follows. □

B Proof of Lemma 2.9(b)

Note that due to (a), we just need to prove the statement of (b) for $k \ge 1$. Hence, assume by contradiction that there exists an index $k_0 \ge 0$ such that $\hat{\mathcal{L}}_{k_0+1} < v(\beta)$. Since by (a), $\{\hat{\mathcal{L}}_k\}$ is decreasing, we obtain

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

which implies that

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

On the other hand, it follows from (6), (14), (26) and **(A3)** that

$$\begin{aligned}
\hat{\mathcal{L}}_k &= \mathcal{L}_\beta(x^k,\lambda^k) + \eta_k \geq \mathcal{L}_\beta(x^k,\lambda^k) \\
&= \sum_{i=1}^{p} f_i(x_i^k) + \frac{\beta}{2}\left\|\sum_{i=1}^{p} A_i x_i^k - b\right\|^2 + \frac{1}{\beta\theta}\langle\lambda^k, \lambda^k - \lambda^{k-1}\rangle \\
&\geq v(\beta) + \frac{1}{2\beta\theta}\left(\|\lambda^k\|^2 - \|\lambda^{k-1}\|^2 + \|\lambda^k - \lambda^{k-1}\|^2\right) \geq v(\beta) + \frac{1}{2\beta\theta}\left(\|\lambda^k\|^2 - \|\lambda^{k-1}\|^2\right)
\end{aligned}$$

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] 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.

[3] 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.

[4] Wei Deng, Ming-Jun Lai, Zhimin Peng, and Wotao Yin. Parallel multi-block admm with o(1/k) convergence. *Journal of Scientific Computing*, 71(2):712–736, 2017.

[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, H-K. Xu, and X. Yuan. On the proximal jacobian decomposition of alm for multiple-block separable convex minimization problems and its relationship to admm. *Journal of Scientific Computing*, 66(3):1204–1217, 2016.

[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] M. Li and X. Yuan. The augmented lagrangian method with full jacobian decomposition and logarithmic-quadratic proximal regularization for multiple-block separable convex programming. *Preprint*, 2015.

[24] 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.

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

[26] 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.

[27] 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.

[28] 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.

[29] Q. Liu, X. Shen, and Y. Gu. Linearized admm for non-convex non-smooth optimization with convergence analysis. *Arxiv Preprint: https://arxiv.org/abs/1705.02502v1*, 2017.

[30] J. G. Melo and R. D. C. Monteiro. Iteration-complexity of a linearized proximal multiblock admm class for linearly constrained nonconvex optimization problems. *Available on: http://www.optimization-online.org*, 2017.

[31] 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.

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

[33] 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.

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

[35] 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.

[36] 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.

[37] H. Wang, A. Banerjee, and Z-Q. Luo. Parallel direction method of multipliers. *Arxiv Preprint: https://arxiv.org/abs/1406.4064*, 2014.

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

[39] 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.

[40] 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.

[41] 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.

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