The Direct Extension of ADMM for Multi-block Convex Minimization Problems is Not Necessarily Convergent

Caihua Chen[1] **Bingsheng He**[2] **Yinyu Ye**[3] **Xiaoming Yuan**[4]

First Version: September 30, 2013; Second Version: January 22, 2014

Abstract. The alternating direction method of multipliers (ADMM) is now widely used in many fields, and its convergence was proved when two blocks of variables are alternatively updated. It is strongly desirable and practically valuable to extend the ADMM directly to the case of a multi-block convex minimization problem where its objective function is the sum of more than two separable convex functions. However, the convergence of this extension has been missing for a long time — neither an affirmative convergence proof nor an example showing its non-convergence is known in the literature. In this paper we give a negative answer to this long-standing open question: The direct extension of ADMM is not necessarily convergent. We present a sufficient condition to ensure the convergence of the direct extension of ADMM, and give an example to show its divergence.

Keywords. Alternating direction method of multipliers, Convergence analysis, Convex programming, Splitting methods

1 Introduction

We consider the convex minimization model with linear constraints and an objective function which is the sum of three functions without coupled variables:

$$\begin{array}{rl} \min & \theta_1(x_1) + \theta_2(x_2) + \theta_3(x_3) \\ \text{s.t.} & A_1x_1 + A_2x_2 + A_3x_3 = b, \\ & x_1 \in \mathcal{X}_1,\, x_2 \in \mathcal{X}_2,\, x_3 \in \mathcal{X}_3, \end{array} \tag{1.1}$$

where $A_i \in \Re^{p\times n_i}$ $(i = 1, 2, 3)$, $b \in \Re^p$, $\mathcal{X}_i \subset \Re^{n_i}$ $(i = 1, 2, 3)$ are closed convex sets; and $\theta_i : \Re^{n_i} \to \Re$ $(i = 1, 2, 3)$ are closed convex but not necessarily smooth functions. The solution set of (1.1) is assumed to be nonempty. The abstract model (1.1) captures many applications in diversifying areas — e.g. see the image alignment problem in [22], the robust principal component analysis model with noisy and incomplete data in [25], the latent variable Gaussian graphical model selection in [5, 21] and the quadratic discriminant analysis model in [20]. Our discussion is inspired by the

[1]International Center of Management Science and Engineering, School of Management and Engineering, Nanjing University, China. This author was supported in part by the Natural Science Foundation of Jiangsu Province under project grant No. BK20130550 and the NSFC Grant 11371192. Email: chchen@nju.edu.cn.

[2]International Centre of Management Science and Engineering, School of Management and Engineering, and Department of Mathematics, Nanjing University, China. This author was supported by the NSFC Grant 91130007 and the MOEC fund 20110091110004. Email: hebma@nju.edu.cn.

[3]Department of Management Science and Engineering, School of Engineering, Stanford University, USA; and International Center of Management Science and Engineering, School of Management and Engineering, Nanjing University, China. This author was supported by AFOSR Grant FA9550-12-1-0396. Email: yyye@stanford.edu.

[4]Department of Mathematics, Hong Kong Baptist University, Hong Kong, China. This author was supported partially by the General Research Fund from Hong Kong Research Grants Council: 203613. Email: xmyuan@hkbu.edu.hk.

scenario where each function θ_i may have some specific properties and it deserves to explore them in algorithmic design. This is often encountered in some sparse and low-rank optimization models, such as the just-mentioned applications of (1.1). We thus do not consider the generic treatment that the sum of three functions is regarded as one general function and possible advantageous properties of each individual θ_i are ignored or not fully used.

The alternating direction method of multipliers (ADMM) was originally proposed in [12] (see also [4, 9]), and it is now a benchmark for the following convex minimization model analogous to (1.1) but with only two blocks of functions and variables:

$$\begin{array}{ll} \min & \theta_1(x_1) + \theta_2(x_2) \\ \text{s.t.} & A_1x_1 + A_2x_2 = b, \\ & x_1 \in \mathcal{X}_1, x_2 \in \mathcal{X}_2 \,. \end{array} \tag{1.2}$$

Let

$$\mathcal{L}_{\mathcal{A}}(x_1, x_2, \lambda) = \theta_1(x_1) + \theta_2(x_2) - \lambda^T\big(A_1x_1 + A_2x_2 - b\big) + \frac{\beta}{2}\big\|A_1x_1 + A_2x_2 - b\big\|^2 \tag{1.3}$$

be the augmented Lagrangian function of (1.2) with the Lagrange multiplier $\lambda \in \Re^p$ and $\beta > 0$ be a penalty parameter. Then, the iterative scheme of ADMM for (1.2) is

$$\text{(ADMM)} \quad \left\{ \begin{array}{ll} x_1^{k+1} = \text{Argmin}\{\mathcal{L}_{\mathcal{A}}(x_1, x_2^k, \lambda^k) \,|\, x_1 \in \mathcal{X}_1\}, & \text{(1.4a)} \\ x_2^{k+1} = \text{Argmin}\{\mathcal{L}_{\mathcal{A}}(x_1^{k+1}, x_2, \lambda^k) \,|\, x_2 \in \mathcal{X}_2\}, & \text{(1.4b)} \\ \lambda^{k+1} = \lambda^k - \beta(A_1x_1^{k+1} + A_2x_2^{k+1} - b). & \text{(1.4c)} \end{array} \right.$$

The iterative scheme of ADMM embeds a Gaussian-Seidel decomposition into each iteration of the augmented Lagrangian method (ALM) in [18, 23]; thus the functions θ_1 and θ_2 are treated individually and so easier subproblems could be generated. This feature is very advantageous for a broad spectrum of application such as partial differential equations, mechanics, image processing, statistical learning, computer vision, and so on. In fact, the ADMM has recently witnessed a "renaissance" in many application domains after a long period without too much attention. We refer to [3, 6, 11] for some review papers on the ADMM.

With the same philosophy as the ADMM to take advantage of each θ_i's properties individually, it is natural to extend the original ADMM (1.4) for (1.2) directly to (1.1) and obtain the scheme

$$\text{(Extended ADMM)} \quad \left\{ \begin{array}{ll} x_1^{k+1} = \text{Argmin}\{\mathcal{L}_{\mathcal{A}}(x_1, x_2^k, x_3^k, \lambda^k) \,|\, x_1 \in \mathcal{X}_1\}, & \text{(1.5a)} \\ x_2^{k+1} = \text{Argmin}\{\mathcal{L}_{\mathcal{A}}(x_1^{k+1}, x_2, x_3^k, \lambda^k) \,|\, x_2 \in \mathcal{X}_2\}, & \text{(1.5b)} \\ x_3^{k+1} = \text{Argmin}\{\mathcal{L}_{\mathcal{A}}(x_1^{k+1}, x_2^{k+1}, x_3, \lambda^k) \,|\, x_3 \in \mathcal{X}_3\}, & \text{(1.5c)} \\ \lambda^{k+1} = \lambda^k - \beta(A_1x_1^{k+1} + A_2x_2^{k+1} + A_3x_3^{k+1} - b), & \text{(1.5d)} \end{array} \right.$$

where

$$\mathcal{L}_{\mathcal{A}}(x_1, x_2, x_3, \lambda) = \sum_{i=1}^{3} \theta_i(x_i) - \lambda^T\big(A_1x_1 + A_2x_2 + A_3x_3 - b\big) + \frac{\beta}{2}\big\|A_1x_1 + A_2x_2 + A_3x_3 - b\big\|^2 \tag{1.6}$$

is the augmented Lagrangian function of (1.1). This direct extension of ADMM is strongly desired and practically used by many users, see e.g. [22, 25]. The convergence of (1.5), however, has been ambiguous for a long time — there is neither an affirmative convergence proof nor an example showing its divergence in the literature. This convergence ambiguity has inspired an active research topic of developing such algorithms that are somehow slightly twisted versions of (1.5) but with provable convergence and competitive numerical efficiency and iteration simplicity, see e.g. [15, 16, 19]. Since the direct extension of ADMM (1.5) does work well for some applications (e.g. [22, 25]), users have the inclination to imagine that this scheme seems convergent even though they are perplexed by the rigorous proof. In the literature, there was even very little hint for the difficulty in the convergence proof for (1.5), see [6] for an insightful explanation.

The main result of this paper is to answer this long-standing open question negatively: The direct extension of ADMM (1.5) is not necessarily convergent. We organize the rest of this paper as follows. In Section 2, we present a sufficient condition to ensure the convergence of (1.5). Then, based on the analysis in Section 2, we construct an example to demonstrate the divergence of the direct extension of ADMM (1.5) in Section 3. Some extensions of the paper's main result are discussed in Section 4. Finally, some concluding remarks are given in Section 5.

2 A Sufficient Condition Ensuring the Convergence of (1.5)

We first study a condition that can ensure the convergence for the direct extension of ADMM (1.5). Our methodology of constructing a counter example to show the divergence of (1.5) is also clear via this study.

Our claim is that the convergence of (1.5) is guaranteed when any two coefficient matrices in (1.1) are orthogonal. We thus will discuss the cases: $A_1^TA_2 = 0$, $A_2^TA_3 = 0$ and $A_1^TA_3 = 0$. This new condition does not impose any strong convexity on the objective function in (1.1), and it simply requires to check the orthogonality of the coefficient matrices. So, it is more checkable than some conditions in the literature such as those in [14, 19] (to be delineated in Section 4).

2.1 Case 1: $A_1^TA_2 = 0$ or $A_2^TA_3 = 0$

We remark that if two coefficient matrices of (1.1) in consecutive order are orthogonal, i.e., $A_1^TA_2 = 0$ or $A_2^TA_3 = 0$, then the direct extension of ADMM (1.5) reduces to a special case of the original ADMM (1.4). Thus the convergence of (1.5) under this condition is implied by well known results in ADMM literature.

To see this, let us first assume $A_1^TA_2 = 0$. According to the first-order optimality conditions of the minimization problems in (1.5), we have $x_i^{k+1} \in \mathcal{X}_i$ $(i = 1, 2, 3)$ and

$$\left\{\begin{array}{ll}\theta_1(x_1) - \theta_1(x_1^{k+1}) + (x_1 - x_1^{k+1})^T\{-A_1^T[\lambda^k - \beta(A_1x_1^{k+1} + A_2x_2^k + A_3x_3^k - b)]\} \ge 0, & \forall x_1 \in \mathcal{X}_1, \quad (2.1a)\\ \theta_2(x_2) - \theta_2(x_2^{k+1}) + (x_2 - x_2^{k+1})^T\{-A_2^T[\lambda^k - \beta(A_1x_1^{k+1} + A_2x_2^{k+1} + A_3x_3^k - b)]\} \ge 0, & \forall x_2 \in \mathcal{X}_2, \quad (2.1b)\\ \theta_3(z) - \theta_3(x_3^{k+1}) + (x_3 - x_3^{k+1})^T\{-A_3^T[\lambda^k - \beta(A_1x_1^{k+1} + A_2x_2^{k+1} + A_3x_3^{k+1} - b)]\} \ge 0, & \forall x_3 \in \mathcal{X}_3. \quad (2.1c)\end{array}\right.$$

Then, because of $A_1^TA_2 = 0$, it follows from (2.1) that

$$\left\{\begin{array}{ll}\theta_1(x_1) - \theta_1(x_1^{k+1}) + (x_1 - x_1^{k+1})^T\{-A_1^T[\lambda^k - \beta(A_1x_1^{k+1} + A_3x_3^k - b)]\} \ge 0, \ \forall x_1 \in \mathcal{X}_1, & (2.2a)\\ \theta_2(x_2) - \theta_2(x_2^{k+1}) + (x_2 - x_2^{k+1})^T\{-A_2^T[\lambda^k - \beta(A_2x_2^{k+1} + A_3x_3^k - b)]\} \ge 0, \ \forall x_2 \in \mathcal{X}_2, & (2.2b)\\ \theta_3(x_3) - \theta_3(x_3^{k+1}) + (x_3 - x_3^{k+1})^T\{-A_3^T[\lambda^k - \beta(A_1x_1^{k+1} + A_2x^{k+1} + A_3x_2^k - b)]\} \ge 0, \ \forall x_3 \in \mathcal{X}_3, & (2.2c)\end{array}\right.$$

which is also the first-order optimality condition of the scheme

$$\begin{cases} (x_1^{k+1}, x_2^{k+1}) = \operatorname{Argmin}\left\{ \begin{array}{c} \theta_1(x_1) + \theta_2(x_2) - (\lambda^k)^T(A_1x_1 + A_2x_2) \\ +\frac{\beta}{2}\|A_1x_1 + A_2x_2 + A_3x_3^k - b\|^2 \end{array} \middle| \begin{array}{c} x_1 \in \mathcal{X}_1, \\ x_2 \in \mathcal{X}_2 \end{array} \right\}, & \text{(2.3a)} \\ x_3^{k+1} = \operatorname{Argmin}\{\theta_3(x_3) - (\lambda^k)^T A_3x_3 + \frac{\beta}{2}\|A_1x_1^{k+1} + A_2x_2^{k+1} + A_3x_3 - b\|^2 | x_3 \in \mathcal{X}_3, \}, & \text{(2.3b)} \\ \lambda^{k+1} = \lambda^k - \beta(A_1x_1^{k+1} + A_2x_2^{k+1} + A_3x_3^{k+1} - b). & \text{(2.3c)} \end{cases}$$

Clearly, (2.3) is a specific application of the original ADMM (1.4) to (1.1) by regarding (x_1, x_2) as one variable, $[A_1, A_2]$ as one matrix and $\theta_1(x_1) + \theta_2(x_2)$ as one function. Note that both x_1^k and x_2^k are not required to generate the $(k+1)$-th iteration under the orthogonality condition $A_1^T A_2 = 0$ in (2.3). Existing convergence results for the original ADMM such as those in [8, 17] thus hold for the special case of (1.5) with the orthogonality condition $A_1^T A_2 = 0$.

Similar discussion can be carried out under the orthogonality condition $A_2^T A_3 = 0$.

2.2 Case 2: $A_1^T A_3 = 0$

In the last subsection, we have discussed the cases where two consecutive coefficient matrices are orthogonal. Now, we pay attention to the case where $A_1^T A_3 = 0$ and show that it can also ensure the convergence of (1.5).

To prepare for the proof, we need to make something clear. First, note that the update order of (1.5) at each iteration is $x_1 \to x_2 \to x_3 \to \lambda$ and then it repeats cyclically. Equivalently, we can update the variables via the order $x_2 \to x_3 \to \lambda \to x_1$ and thus have the following iterative form:

$$\begin{cases} x_2^{k+1} = \operatorname{Argmin}\{\mathcal{L}_\mathcal{A}(x_1^k, x_2, x_3^k, \lambda^k) \,|\, x_2 \in \mathcal{X}_2\}, & \text{(2.4a)} \\ x_3^{k+1} = \operatorname{Argmin}\{\mathcal{L}_\mathcal{A}(x_1^k, x_2^{k+1}, x_3, \lambda^k) \,|\, x_3 \in \mathcal{X}_3\}, & \text{(2.4b)} \\ \lambda^{k+1} = \lambda^k - \beta(A_1x_1^k + A_2x_2^{k+1} + A_3x_3^{k+1} - b), & \text{(2.4c)} \\ x_1^{k+1} = \operatorname{Argmin}\{\mathcal{L}_\mathcal{A}(x_1, x_2^{k+1}, x_3^{k+1}, \lambda^{k+1}) \,|\, x_1 \in \mathcal{X}_1\}. & \text{(2.4d)} \end{cases}$$

According to (2.4), there is a update for the variable λ between the updates for x_3 and x_1. Thus, the case $A_1^T A_3 = 0$ requires discussion different from that in the last subsection. Moreover, when x_1^k is taken as x_1^{k+1} and x_1^{k+1} as x_1^{k+2}, the scheme (2.4) reduces exactly to the direct extension of ADMM (1.5). Therefore, the convergence analysis for the scheme (1.5) is equivalent to that for (2.4). For notational simplicity, we will focus on the representation of (2.4) within this subsection.

Second, it worths to mention that the variable x_2 is not involved in the iteration of (2.4), meaning the scheme (2.4) generating a new iterate only based on $(x_1^k, x_3^k, \lambda^k)$. We thus follow the terminology in [3] to call x_2 an intermediate variable; and correspondingly call (x_1, x_3, λ) essential variables because they are really necessary to execute the iteration of (2.4). Accordingly, we use the notations $w^k = (x_1^k, x_2^k, x_3^k, \lambda^k)$, $u^k = w^k \setminus \lambda^k = (x_1^k, x_2^k, x_3^k)$, $v^k = w^k \setminus x_2^k = (x_1^k, x_3^k, \lambda^k)$, $v = w \setminus x_2 = (x_1, x_3, \lambda)$, $\mathcal{V} = \mathcal{X}_1 \times \mathcal{X}_3 \times \Re^p$ and

$$\mathcal{V}^* := \{v^* = (x_1^*, x_3^*, \lambda^*) \,|\, w^* = (x_1^*, x_2^*, x_3^*, \lambda^*) \in \Omega^*\}.$$

Third, it is useful to characterize the model (1.1) by a variational inequality. More specifically, finding a saddle point of the Lagrange function of (1.1) is equivalent to solving the variational inequality problem: Finding $w^* \in \Omega$ such that

$$\text{VI}(\Omega, F, \theta): \quad \theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \geq 0, \quad \forall\, w \in \Omega, \qquad \text{(2.5a)}$$

where

$$u = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}, \quad w = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \lambda \end{pmatrix}, \quad \theta(u) = \theta_1(x_1) + \theta_2(x_2) + \theta_3(x_3), \tag{2.5b}$$

and

$$F(w) = \begin{pmatrix} -A_1^T\lambda \\ -A_2^T\lambda \\ -A_3^T\lambda \\ A_1x_1 + A_2x_2 + A_3x_3 - b \end{pmatrix}. \tag{2.5c}$$

Obviously, the mapping $F(\cdot)$ defined in (2.5c) is monotone because it is affine with a skew-symmetric matrix.

Last, let us take a deeper look at the output of (2.4) and investigate some of its properties. In fact, deriving the first-order optimality condition of the minimization problems in (2.4) and rewriting (2.4c) appropriately, we obtain

$$\begin{cases} \theta_2(x_2) - \theta_2(x_2^{k+1}) + (x_2 - x_2^{k+1})^T\{-A_2^T[\lambda^k - \beta(A_1x_1^k + A_2x_2^{k+1} + A_3x_3^k - b)]\} \geq 0, \quad \forall x_2 \in \mathcal{X}_2, & (2.6a) \\ \theta_3(x_3) - \theta_3(x_3^{k+1}) + (x_3 - x_3^{k+1})^T\{-A_3^T[\lambda^k - \beta(A_1x_1^k + A_2x_2^{k+1} + A_3x_3^{k+1} - b)]\} \geq 0, \ \forall x_3 \in \mathcal{X}_3, & (2.6b) \\ (A_1x_1^k + A_2x_2^{k+1} + A_3x_3^{k+1} - b) + \frac{1}{\beta}(\lambda^{k+1} - \lambda^k) = 0, & (2.6c) \\ \theta_1(x_1) - \theta_1(x_1^k) + (x_1 - x_1^{k+1})^T\{-A_1^T[\lambda^{k+1} - \beta(A_1x_1^{k+1} + A_2x_2^{k+1} + A_3x_3^{k+1} - b)]\} \geq 0, \ \forall x_1 \in \mathcal{X}_1. & (2.6d) \end{cases}$$

Then, substituting (2.6c) into (2.6a), (2.6b) and (2.6d); and using $A_1^TA_3 = 0$, we get

$$\begin{cases} \theta_2(x_2) - \theta_2(x_2^{k+1}) + (x_2 - x_2^{k+1})^T\{-A_2^T\lambda^{k+1} + \beta A_2^TA_3(x_3^k - x_3^{k+1})\} \geq 0, \quad \forall x_2 \in \mathcal{X}_2, & (2.7a) \\ \theta_3(x_3) - \theta_3(x_3^{k+1}) + (x_3 - x_3^{k+1})^T\{-A_3^T\lambda^{k+1}\} \geq 0, \ \forall x_3 \in \mathcal{X}_3, & (2.7b) \\ (A_1x_1^{k+1} + A_2x_2^{k+1} + A_3x_3^{k+1} - b) + A_1(x_1^k - x_1^{k+1}) - \frac{1}{\beta}(\lambda^k - \lambda^{k+1}) = 0, & (2.7c) \\ \theta_1(x_1) - \theta_1(x_1^{k+1}) + (x_1 - x_1^{k+1})^T\{-A_1^T\lambda^{k+1} - \beta A_1^TA_1(x_1^k - x_1^{k+1}) + A_1^T(\lambda^k - \lambda^{k+1})\} \geq 0, \ \forall x_1 \in \mathcal{X}_1. & (2.7d) \end{cases}$$

With the definitions of θ, F, Ω, u^k and v^k, we can rewrite (2.7) as a compact form. We summarize it in the next lemma and omit its proof as it is just a compact reformulation of (2.7).

Lemma 2.1. *Let $w^{k+1} = (x_1^{k+1}, x_2^{k+1}, x_3^{k+1}, \lambda^{k+1})$ be generated by* (2.4) *from given $v^k = (x_1^k, x_3^k, \lambda^k)$. Then we have*

$$w^{k+1} \in \Omega, \quad \theta(u) - \theta(u^{k+1}) + (w - w^{k+1})^T\{F(w^{k+1}) + Q(v^k - v^{k+1})\} \geq 0, \quad \forall\, w \in \Omega, \tag{2.8}$$

where

$$Q = \begin{pmatrix} -\beta A_1^TA_1 & 0 & A_1^T \\ 0 & \beta A_2^TA_3 & 0 \\ 0 & 0 & 0 \\ A_1 & 0 & -\frac{1}{\beta}I \end{pmatrix}. \tag{2.9}$$

Note that the assertion (2.8) is useful for quantifying the accuracy of w^{k+1} to a solution point of $\text{VI}(\Omega, F, \theta)$, because of the variational inequality reformulation (2.5) of (1.1).

Now, we are ready to prove the convergence for the direct extension of ADMM under the condition $A_1^TA_3 = 0$. We first refine the assertion (2.8) under this additional condition.

Lemma 2.2. *Let* $w^{k+1} = (x_1^k, x_2^{k+1}, x_3^{k+1}, \lambda^{k+1})$ *be generated by* (2.4) *from given* $v^k = (x_1^k, x_3^k, \lambda^k)$. *If* $A_1^T A_3 = 0$, *then we have*

$$\begin{aligned} w^{k+1} \in \Omega, \quad & \theta(u) - \theta(u^{k+1}) + (w - w^{k+1})^T \{F(w^{k+1}) + \beta P A_3 (x_3^k - x_3^{k+1})\} \\ & \geq (v - v^{k+1})^T H (v^k - v^{k+1}), \quad \forall\, w \in \Omega, \end{aligned} \tag{2.10}$$

where

$$P = \begin{pmatrix} A_1^T \\ A_2^T \\ A_3^T \\ 0 \end{pmatrix}, \quad v = \begin{pmatrix} x_1 \\ x_3 \\ \lambda \end{pmatrix} \quad \text{and} \quad H = \begin{pmatrix} \beta A_1^T A_1 & 0 & -A_1^T \\ 0 & \beta A_3^T A_3 & 0 \\ -A_1 & 0 & \frac{1}{\beta} I \end{pmatrix}. \tag{2.11}$$

Proof. Since $A_1^T A_3 = 0$, the following is an identity:

$$\begin{aligned} & \begin{pmatrix} x_1 - x_1^{k+1} \\ x_2 - x_2^{k+1} \\ x_3 - x_3^{k+1} \\ \lambda - \lambda^{k+1} \end{pmatrix}^T \begin{pmatrix} \beta A_1^T A_1 & \beta A_1^T A_3 & -A_1^T \\ 0 & 0 & 0 \\ 0 & \beta A_3^T A_3 & 0 \\ -A_1 & 0 & \frac{1}{\beta} I \end{pmatrix} \begin{pmatrix} x_1^k - x_1^{k+1} \\ x_3^k - x_3^{k+1} \\ \lambda^k - \lambda^{k+1} \end{pmatrix} \\ = & \begin{pmatrix} x_1 - x_1^{k+1} \\ x_3 - x_3^{k+1} \\ \lambda - \lambda^{k+1} \end{pmatrix}^T \begin{pmatrix} \beta A_1^T A_1 & 0 & -A_1^T \\ 0 & \beta A_3^T A_3 & 0 \\ -A_1 & 0 & \frac{1}{\beta} I \end{pmatrix} \begin{pmatrix} x_1^k - x_1^{k+1} \\ x_3^k - x_3^{k+1} \\ \lambda^k - \lambda^{k+1} \end{pmatrix}. \end{aligned}$$

Adding the above identity to the both sides of (2.8) and using the notations of v and H, we obtain

$$\begin{aligned} w^{k+1} \in \Omega, \quad & \theta(u) - \theta(u^{k+1}) + (w - w^{k+1})^T \{F(w^{k+1}) + Q_0 (v^k - v^{k+1})\} \\ & \geq (v - v^{k+1})^T H (v^k - v^{k+1}), \quad \forall w \in \Omega, \end{aligned} \tag{2.12}$$

where (see Q in (2.9))

$$Q_0 = Q + \begin{pmatrix} \beta A_1^T A_1 & \beta A_1^T A_3 & -A_1^T \\ 0 & 0 & 0 \\ 0 & \beta A_3^T A_3 & 0 \\ -A_1 & 0 & \frac{1}{\beta} I \end{pmatrix} = \begin{pmatrix} 0 & \beta A_1^T A_3 & 0 \\ 0 & \beta A_2^T A_3 & 0 \\ 0 & \beta A_3^T A_3 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

By using the structures of the matrices Q_0 and P (see (2.11)), and the vector v, we have

$$(w - w^{k+1})^T Q_0 (v^k - v^{k+1}) = (w - w^{k+1})^T \beta P A_3 (x_3^k - x_3^{k+1}).$$

The assertion (2.10) is proved. □

Let us define two auxiliary sequences which will only serve for simplifying our notation in convergence analysis:

$$\tilde{w}^k = \begin{pmatrix} \tilde{x}_1^k \\ \tilde{x}_2^k \\ \tilde{x}_3^k \\ \tilde{\lambda}^k \end{pmatrix} = \begin{pmatrix} x_1^{k+1} \\ x_2^{k+1} \\ x_3^{k+1} \\ \lambda^{k+1} - \beta A_3 (x_3^k - x_3^{k+1}) \end{pmatrix} \quad \text{and} \quad \tilde{u}^k = \begin{pmatrix} \tilde{x}_1^k \\ \tilde{x}_2^k \\ \tilde{x}_3^k \end{pmatrix}, \tag{2.13}$$

where $\{x_1^{k+1}, x_2^{k+1}, x_3^{k+1}, \lambda^{k+1}\}$ is generated by (2.4).

In the next lemma, we establish an important inequality based on the assertion in Lemma 2.2, which will play a vital role in convergence analysis.

Lemma 2.3. *Let $w^{k+1}=(x_1^{k+1},x_2^{k+1},x_3^{k+1},\lambda^{k+1})$ be generated by (2.4) from given $v^k=(x_1^k,x_3^k,\lambda^k)$. If $A_1^TA_3=0$, we have $\tilde{w}^k\in\Omega$ and*

$$\theta(u)-\theta(\tilde{u}^k)+(w-\tilde{w}^k)^TF(\tilde{w}^k)\geq\frac{1}{2}\big(\|v-v^{k+1}\|_H^2-\|v-v^k\|_H^2\big)+\frac{1}{2}\|v^k-v^{k+1}\|_H^2,\ \forall\, w\in\Omega, \quad (2.14)$$

where $\tilde{w}^k$ and $\tilde{u}^k$ are defined in (2.13).

Proof. According to the definition of $\tilde{w}^k$ and $F(w)$ (see (2.13) and (2.5c), respectively), (2.10) can be rewritten as

$$\tilde{w}^k\in\Omega,\quad \theta(u)-\theta(\tilde{u}^k)+(w-w^{k+1})^TF(\tilde{w}^k)\geq(v-v^{k+1})^TH(v^k-v^{k+1}),\quad \forall w\in\Omega. \quad (2.15)$$

Note that $w^{k+1}-\tilde{w}^k=\beta\begin{pmatrix}0\\0\\0\\A_3(x_3^k-x_3^{k+1})\end{pmatrix}$, we further obtain that $\tilde{w}^k\in\Omega$, and

$$\begin{aligned}&\theta(u)-\theta(\tilde{u}^k)+(w-\tilde{w}^k)^TF(\tilde{w}^k)\\ &=\ \theta(u)-\theta(\tilde{u}^k)+(w-w^{k+1})^TF(\tilde{w}^k)+(w^{k+1}-\tilde{w}^k)^TF(\tilde{w}^k)\\ &\geq\ (v-v^{k+1})^TH(v^k-v^{k+1})\\ &\quad+\big(A_3(x_3^k-x_3^{k+1})\big)^T\big(\beta(A_1x_1^{k+1}+A_2x_2^{k+1}+A_3x_3^{k+1}-b)\big),\ \forall w\in\Omega.\end{aligned} \quad (2.16)$$

Setting $x_3=x_3^k$ in (2.7b), we obtain

$$\theta_3(x_3^k)-\theta_3(x_3^{k+1})+(x_3^k-x_3^{k+1})^T\{-A_3^T\lambda^{k+1}\}\geq 0. \quad (2.17)$$

Note that (2.7b) is also true for the $(k-1)$th iteration. Thus, it holds that

$$\theta_3(x_3)-\theta_3(x_3^k)+(x_3-x_3^k)^T\{-A_3^T\lambda^k\}\geq 0.$$

Setting $x_3=x_3^{k+1}$ in the last inequality, we obtain

$$\theta_3(x_3^{k+1})-\theta_3(x_3^k)+(x_3^{k+1}-x_3^k)^T\{-A_3^T\lambda^k\}\geq 0, \quad (2.18)$$

which together with (2.17) yields that

$$(\lambda^k-\lambda^{k+1})^TA_3(x_3^k-x_3^{k+1})\geq 0,\quad \forall k\geq 0. \quad (2.19)$$

By using the fact $\lambda^k-\lambda^{k+1}=\beta(A_1x_1^k+A_2x_2^{k+1}+A_3x_3^{k+1}-b)$ (see (2.6c)) and the assumption $A_1^TA_3=0$, we get immediately that

$$\beta(A_1x_1^{k+1}+A_2x_2^{k+1}+A_3x_3^{k+1}-b)^TA_3(x_3^k-x_3^{k+1})\geq 0, \quad (2.20)$$

and hence

$$\tilde{w}^k\in\Omega,\quad \theta(u)-\theta(\tilde{u}^k)+(w-\tilde{w}^k)^TF(\tilde{w}^k)\geq(v-v^{k+1})^TH(v^k-v^{k+1}),\quad \forall\, w\in\Omega. \quad (2.21)$$

By substituting the identity

$$(v-v^{k+1})^TH(v^k-v^{k+1})=\frac{1}{2}\big(\|v-v^{k+1}\|_H^2-\|v-v^k\|_H^2\big)+\frac{1}{2}\|v^k-v^{k+1}\|_H^2$$

into the right-hand side of (2.21), we obtain (2.14). □

Now, we are able to establish the contraction property with respect to the solution set of VI(Ω,F,θ) for the sequence $\{v^k\}$ generated by (2.4), from which the convergence of (2.4) can be easily established.

Theorem 2.4. *Assume* $A_1^T A_3 = 0$ *for the model (1.1). Let* $\{x_1^k, x_2^k, x_3^k, \lambda^k\}$ *be the sequence generated by the direct extension of ADMM (2.4). Then, we have:*

(i) *The sequence* $\{v^k := (x_1^k, x_3^k, \lambda^k)\}$ *is contractive with respective to the solution of VI*(Ω, F, θ)*, i.e.,*

$$\|v^{k+1} - v^*\|_H^2 \leq \|v^k - v^*\|_H^2 - \|v^k - v^{k+1}\|_H^2. \tag{2.22}$$

(ii) *If the matrices* $[A_1,\ A_2]$ *and* A_3 *are assumed to be full column rank, then the sequence* $\{w^k\}$ *converges to a KKT point of the model* (1.1)*.*

Proof. (i) The first assertion is straightforward based on (2.14). Setting $w = w^*$ in (2.14), we get

$$\frac{1}{2}\big(\|v^k - v^*\|_H^2 - \|v^{k+1} - v^*\|_H^2\big) - \frac{1}{2}\|v^k - v^{k+1}\|_H^2 \geq \theta(\tilde{u}^k) - \theta(u^*) + (\tilde{w}^k - w^*)^T F(\tilde{w}^k).$$

From the monotonicity of F and (2.5), it follows that

$$\theta(\tilde{u}^k) - \theta(u^*) + (\tilde{w}^k - w^*)^T F(\tilde{w}^k) \geq \theta(\tilde{u}^k) - \theta(u^*) + (\tilde{w}^k - w^*)^T F(w^*) \geq 0,$$

and thus (2.22) is proved. Clearly, (2.22) indicates that the sequence $\{v^k\}$ is contractive with respect to the solution set of VI(Ω, F, θ), see e.g. [2].

(ii) To prove (ii), by the inequality (2.22) and (see the definitions of v and H in (2.11))

$$\|v^k - v^{k+1}\|_H^2 = \beta\|A_1(x_1^k - x_1^{k+1}) - \frac{1}{\beta}(\lambda^k - \lambda^{k+1})\|^2 + \beta\|A_3(x_3^k - x_3^{k+1})\|^2,$$

it follows that the sequences $\{A_1 x_1^k - \frac{1}{\beta}\lambda^k\}$ and $\{A_3 x_3^k\}$ are both bounded. Since A_3 has full column rank, we deduce that $\{x_3^k\}$ is bounded. Note that

$$A_1 x_1^k + A_2 x_2^k = A_1 x_1^k - \frac{1}{\beta}\lambda^k - (A_1 x_1^{k-1} - \frac{1}{\beta}\lambda^{k-1}) - A_3 x_3^k + b. \tag{2.23}$$

Hence, $\{A_1 x_1^k + A_2 x_2^k\}$ is bounded. Together with the assumption that $[A_1,\ A_2]$ has full column rank, we conclude that the sequences $\{x_1^k\}, \{x_2^k\}$ and $\{\lambda^k\}$ are all bounded. Therefore, there exists a subsequence $\{x_1^{n_k+1}, x_2^{n_k+1}, x_3^{n_k+1}, \lambda^{n_k+1}\}$ that converges to a limit point, say $(x_1^\infty, x_2^\infty, x_3^\infty, \lambda^\infty)$. Moreover, from (2.22), we see immediately that

$$\sum_{k=1}^{\infty} \|v^k - v^{k+1}\|_H^2 < +\infty, \tag{2.24}$$

which shows

$$\lim_{k\to\infty} H(v^k - v^{k+1}) = 0, \tag{2.25}$$

and thus

$$\lim_{k\to\infty} Q(v^k - v^{k+1}) = 0. \tag{2.26}$$

Then, by taking the limits on the both sides of (2.8), using (2.26), one can immediately write

$$w^\infty \in \Omega, \quad \theta(u) - \theta(u^\infty) + \ (w - w^\infty)^T F(w^\infty) \geq 0, \ \ \forall\, w \in \Omega, \tag{2.27}$$

which means $w^\infty = (x_1^\infty, x_2^\infty, x_3^\infty, \lambda^\infty)$ is a KKT point of (1.1). Hence, the inequality (2.22) is also valid if $(x_1^*, x_2^*, x_3^*, \lambda^*)$ is replaced by $(x_1^\infty, x_2^\infty, x_3^\infty, \lambda^\infty)$. Then it holds that

$$\|v^{k+1} - v^\infty\|_H^2 \leq \|v^k - v^\infty\|_H^2, \tag{2.28}$$

which implies that

$$\lim_{k\to\infty} A_1(x_1^k - x_1^\infty) - \frac{1}{\beta}(\lambda^k - \lambda^\infty) = 0, \quad \lim_{k\to\infty} A_3(x_3^k - x_3^\infty) = 0. \tag{2.29}$$

By taking limits to (2.23), using (2.29) and the assumptions, we know

$$\lim_{k\to\infty} x_1^k = x_1^\infty, \quad \lim_{k\to\infty} x_2^k = x_2^\infty, \quad \lim_{k\to\infty} x_3^k = x_3^\infty, \quad \lim_{k\to\infty} \lambda^k = \lambda^\infty. \tag{2.30}$$

which completes the proof of this theorem. □

Inspired by [17], we can also establish a worst-case convergence rate measured by the iteration complexity in the ergodic sense for the direct extension of ADMM (2.4). This is summarized in the following theorem.

Theorem 2.5. *Assume $A_1^T A_3 = 0$ for the model (1.1). Let $\{(x_1^k, x_2^k, x_3^k, \lambda^k)\}$ be the sequence generated by the direct extension of ADMM (2.4) and $\tilde{w}^k$ be defined in* (2.13). *After t iterations of (2.4), we take*

$$\tilde{w}_t = \frac{1}{t+1}\sum_{k=0}^{t} \tilde{w}^k. \tag{2.31}$$

Then, $\tilde{w} \in \mathcal{W}$ and it satisfies

$$\theta(\tilde{u}_t) - \theta(u) + (\tilde{w}_t - w)^T F(w) \le \frac{1}{2(t+1)}\|v - v^0\|_H^2, \quad \forall w \in \Omega. \tag{2.32}$$

Proof. By the monotonicity of F and (2.14), it follows that

$$\tilde{w}^k \in \Omega, \;\; \theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(w) + \frac{1}{2}\|v - v^k\|_H^2 \ge \frac{1}{2}\|v - v^{k+1}\|_H^2, \; \forall w \in \Omega. \tag{2.33}$$

Together with the convexity of $\mathcal{X}_1$, $\mathcal{X}_2$ and $\mathcal{X}_3$, (2.31) implies that $\tilde{w}_t \in \Omega$. Summing the inequality (2.33) over $k = 0, 1, \ldots, t$, we obtain

$$(t+1)\theta(u) - \sum_{k=0}^{t}\theta(\tilde{u}^k) + \Big((t+1)w - \sum_{k=0}^{t}\tilde{w}^k\Big)^T F(w) + \frac{1}{2}\|v - v^0\|_H^2 \ge 0, \quad \forall w \in \Omega.$$

Use the notation of $\tilde{w}_t$, it can be written as

$$\frac{1}{t+1}\sum_{k=0}^{t}\theta(\tilde{u}^k) - \theta(u) + (\tilde{w}_t - w)^T F(w) \le \frac{1}{2(t+1)}\|v - v^0\|_H^2, \quad \forall w \in \Omega. \tag{2.34}$$

Since $\theta(u)$ is convex and

$$\tilde{u}_t = \frac{1}{t+1}\sum_{k=0}^{t}\tilde{u}^k,$$

we have that

$$\theta(\tilde{u}_t) \le \frac{1}{t+1}\sum_{k=0}^{t}\theta(\tilde{u}^k).$$

Substituting it into (2.34), the assertion of this theorem follows directly. □

Remark 2.6. For an arbitrarily given compact set $D \subset \Omega$, let $d = \sup\{\|v - v^0\|_H^2\} \mid v = w\backslash x_2, w \in D\}$, where $v^0 = (x_1^0, x_3^0, \lambda^0)$. Then, after t iterations of the extended ADMM (2.4) , the point $\tilde{w}_t$ defined in (2.31) satisfies

$$\sup\{\theta(\tilde{u}_t) - \theta(u) + (\tilde{w}_t - w)^T F(w)\} \le \frac{d}{2(t+1)}, \tag{2.35}$$

which, according to the definition (2.5), means $\tilde{w}_t$ is an approximate solution of $\mathrm{VI}(\Omega, F, \theta)$ with an accuracy of $O(1/t)$. Thus a worst-case $O(1/t)$ convergence rate in the ergodic sense is established for the direct extension of ADMM (2.4).

3 An Example Showing the Divergence of (1.5)

In the last section, we have shown that if it is additionally assumed that any two coefficient matrices in (1.1) be orthogonal, then the direct extension of ADMM (1.5) is convergent. In this section, we give an example to show the divergence of (1.5) when such an orthogonality condition is missing. The analyses below also present a strategy for constructing more such non-convergent examples.

More specifically, we consider the following linear homogeneous equation with three variables:

$$A_1x_1 + A_2x_2 + A_3x_3 = 0, \tag{3.1}$$

where $A_i \in \Re^3$ $(i = 1, 2, 3)$ are all column vectors and the matrix $[A_1, A_2, A_3]$ is assumed to be nonsingular; and $x_i \in \Re$ $(i = 1, 2, 3)$. The unique solution of (3.1) is thus $x_1 = x_2 = x_3 = 0$. Clearly, (3.1) is a special case of (1.1) where the objective function is null, b is the all-zero vector in $\Re^3$, and $\mathcal{X}_i = \Re$ for $i = 1, 2, 3$. The direct extension of ADMM (1.5) is thus applicable to (3.1), and the corresponding optimal Lagrange multipliers are all 0.

One will see next that the convergence of the direct extension of ADMM (1.5) applied to solving the linear equations with a null objective is independent of the selection of the penalty parameter β. That is, if the direct extension of ADMM (1.5) is convergent for one selected $\beta > 0$, then the it is convergent for every $\beta > 0$. On the other hand, if (1.5) is not convergent for one selected $\beta > 0$, then it is not convergent for all $\beta > 0$. Hence, in our specific example developed below, one can think $\beta = 1$ without loss of generality.

3.1 The Iterative Scheme of (1.5) for (3.1)

Now, we elucidate the iterative scheme when the direct extension of ADMM (1.5) is applied to solve the linear equation (3.1). In fact, as we will show, it can be represented as a matrix recursion.

Specifying the scheme (1.5) with any given $\beta > 0$ by the particular setting in (3.1), we obtain

$$\begin{cases} -A_1^T\lambda^k + \beta A_1^T(A_1x_1^{k+1} + A_2x_2^k + A_3x_3^k) = 0, & (3.2a) \\ -A_2^T\lambda^k + \beta A_2^T(A_1x_1^{k+1} + A_2x_2^{k+1} + A_3x_3^k) = 0, & (3.2b) \\ -A_3^T\lambda^k + \beta A_3^T(A_1x_1^{k+1} + A_2x_2^{k+1} + A_3x_3^{k+1}) = 0, & (3.2c) \\ \beta(A_1x_1^{k+1} + A_2x_2^{k+1} + A_3x_3^{k+1}) + \lambda^{k+1} - \lambda^k = 0. & (3.2d) \end{cases}$$

By introducing a new variable $\mu^k := \lambda^k/\beta$, we can recast the scheme(3.2) as

$$\begin{cases} -A_1^T\mu^k + A_1^T(A_1x_1^{k+1} + A_2x_2^k \quad + A_3x_3^k\ ) = 0, & (3.3a)\\ -A_2^T\mu^k + A_2^T(A_1x_1^{k+1} + A_2x_2^{k+1} + A_3x_3^k\ ) = 0, & (3.3b)\\ -A_3^T\mu^k + A_3^T(A_1x_1^{k+1} + A_2x_2^{k+1} + A_3x_3^{k+1}\ ) = 0, & (3.3c)\\ (A_1x_1^{k+1} + A_2x_2^{k+1} + A_3x_3^{k+1}) + \mu^{k+1} - \mu^k = 0. & (3.3d) \end{cases}$$

It follows from the first equation in (3.3) that

$$x_1^{k+1} = \frac{1}{A_1^TA_1}(-A_1^TA_2x_2^k - A_1^TA_3x_3^k + A_1^T\mu^k). \tag{3.4}$$

Substituting (3.4) into (3.3b), (3.3c) and (3.3d), we obtain a reformulation of (3.3)

$$\begin{pmatrix} A_2^TA_2 & 0 & 0_{1\times 3} \\ A_3^TA_2 & A_3^TA_3 & 0_{1\times 3} \\ A_2 & A_3 & I_{3\times 3} \end{pmatrix}\begin{pmatrix} x_2^{k+1} \\ x_3^{k+1} \\ \mu^{k+1} \end{pmatrix}$$

$$= \left[\begin{pmatrix} 0 & -A_2^TA_3 & A_2^T \\ 0 & 0 & A_3^T \\ 0_{3\times 1} & 0_{3\times 1} & I_{3\times 3} \end{pmatrix} - \frac{1}{A_1^TA_1}\begin{pmatrix} A_2^TA_1 \\ A_3^TA_1 \\ A_1 \end{pmatrix}\left(-A_1^TA_2, -A_1^TA_3, A_1^T\right)\right]\begin{pmatrix} x_2^k \\ x_3^k \\ \mu^k \end{pmatrix}. \tag{3.5}$$

Let

$$L = \begin{pmatrix} A_2^TA_2 & 0 & 0_{1\times 3} \\ A_3^TA_2 & A_3^TA_3 & 0_{1\times 3} \\ A_2 & A_3 & I_{3\times 3} \end{pmatrix} \tag{3.6}$$

and

$$R = \begin{pmatrix} 0 & -A_2^TA_3 & A_2^T \\ 0 & 0 & A_3^T \\ 0_{3\times 1} & 0_{3\times 1} & I_{3\times 3} \end{pmatrix} - \frac{1}{A_1^TA_1}\begin{pmatrix} A_2^TA_1 \\ A_3^TA_1 \\ A_1 \end{pmatrix}\left(-A_1^TA_2, -A_1^TA_3, A_1^T\right). \tag{3.7}$$

Then the iterative formula (3.5) can be rewritten in the following fixed matrix mappings:

$$\begin{pmatrix} x_2^{k+1} \\ x_3^{k+1} \\ \mu^{k+1} \end{pmatrix} = M\begin{pmatrix} x_2^k \\ x_3^k \\ \mu^k \end{pmatrix} = \cdots = M^{k+1}\begin{pmatrix} x_2^0 \\ x_3^0 \\ \mu^0 \end{pmatrix} \tag{3.8}$$

with

$$M = L^{-1}R. \tag{3.9}$$

Therefore, the direct extension of ADMM (1.5) is convergent if the matrix mapping is a contraction, or equivalently, the spectral radius of M, denote by $\rho(M)$, is strictly less than 1. Thus, to construct a divergent example, we would look for a A such that $\rho(M) > 1$.

3.2 A Concrete Example Showing the Divergence of (1.5)

Now we are ready to construct a concrete example to show the divergence of the direct extension of ADMM (1.5) for all $\beta > 0$ when it is applied to solve the model (3.1).

Our previous analysis in Section 2 has shown that the scheme (1.5) is convergent whenever any two coefficient matrices are orthogonal. Thus, to show the divergence of (1.5) for (3.1), the columns A_1, A_2 and A_3 in (3.1) should be chosen such that any two of them are non-orthogonal. Specifically, we thus construct the matrix A as follows:

$$A = (A_1, A_2, A_3) = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 2 \\ 1 & 2 & 2 \end{pmatrix}. \tag{3.10}$$

Given this matrix A, the system of linear equations (3.5) can be specified as

$$\begin{pmatrix} 6 & 0 & 0 & 0 & 0 \\ 7 & 9 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 2 & 0 & 1 & 0 \\ 2 & 2 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_2^{k+1} \\ x_3^{k+1} \\ \mu_1^{k+1} \\ \mu_2^{k+1} \\ \mu_3^{k+1} \end{pmatrix}$$
$$= \left[\begin{pmatrix} 0 & -7 & 1 & 1 & 2 \\ 0 & 0 & 1 & 2 & 2 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} - \frac{1}{3} \begin{pmatrix} 4 \\ 5 \\ 1 \\ 1 \\ 1 \end{pmatrix} \begin{pmatrix} -4, -5, 1, 1, 1 \end{pmatrix} \right] \begin{pmatrix} x_2^k \\ x_3^k \\ \mu_1^k \\ \mu_2^k \\ \mu_3^k \end{pmatrix}.$$

Note with the specification in (3.10), the matrices L in (3.6) and R in (3.7) reduce to

$$L = \begin{pmatrix} 6 & 0 & 0 & 0 & 0 \\ 7 & 9 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 2 & 0 & 1 & 0 \\ 2 & 2 & 0 & 0 & 1 \end{pmatrix} \quad \text{and} \quad R = \frac{1}{3} \begin{pmatrix} 16 & -1 & -1 & -1 & 2 \\ 20 & 25 & -2 & 1 & 1 \\ 4 & 5 & 2 & -1 & -1 \\ 4 & 5 & -1 & 2 & -1 \\ 4 & 5 & -1 & -1 & 2 \end{pmatrix}.$$

Thus we have

$$M = L^{-1}R = \frac{1}{162} \begin{pmatrix} 144 & -9 & -9 & -9 & 18 \\ 8 & 157 & -5 & 13 & -8 \\ 64 & 122 & 122 & -58 & -64 \\ 56 & -35 & -35 & 91 & -56 \\ -88 & -26 & -26 & -62 & 88 \end{pmatrix}.$$

From direct computation, M admits the following eigenvalue decomposition

$$M = V\mathrm{Diag}(\mathrm{d})V^{-1}, \tag{3.11}$$

where

$$d = \begin{pmatrix} 0.9836 + 0.2984i \\ 0.9836 - 0.2984i \\ 0.8744 + 0.2310i \\ 0.8744 - 0.2310i \\ 0 \end{pmatrix}$$

and

$$V = \begin{pmatrix} 0.1314 + 0.2661i & 0.1314 - 0.2661i & 0.1314 - 0.2661i & 0.1314 + 0.2661i & 0 \\ 0.0664 - 0.2718i & 0.0664 + 0.2718i & 0.0664 + 0.2718i & 0.0664 - 0.2718i & 0 \\ -0.2847 - 0.4437i & -0.2847 + 0.4437i & 0.2847 - 0.4437i & 0.2847 + 0.4437i & 0.5774 \\ 0.5694 & 0.5694 & -0.5694 & -0.5694 & 0.5774 \\ -0.4270 + 0.2218i & -0.4270 - 0.2218i & 0.4270 + 0.2218i & 0.4270 - 0.2218i & 0.5774 \end{pmatrix}.$$

An important fact regarding d defined above is that

$$\rho(M) = |d_1| = |d_2| > 1,$$

from which we can construct a divergent sequence $\{x_2^k, x_3^k, \lambda_1^k, \lambda_2^k, \lambda_3^k\}$ starting from certain initial points. The questions are: Can we find real-valued non-convergent starting points? Does the set of non-convergent starting points form a continuously dense set, that is, are they not isolated? We give affirmative answers below.

Indeed, for any initial $(x_2^0, x_3^0, \mu_1^0, \mu_2^0, \mu_3^0)$, let

$$\begin{pmatrix} l_1 \\ l_2 \\ l_3 \\ l_4 \\ l_5 \end{pmatrix} = V^{-1} \begin{pmatrix} x_2^0 \\ x_3^0 \\ \mu_1^0 \\ \mu_2^0 \\ \mu_3^0 \end{pmatrix}. \tag{3.12}$$

From (3.8) and (3.11), we know that

$$\begin{aligned} \begin{pmatrix} x_2^k \\ x_3^k \\ \mu_1^k \\ \mu_2^k \\ \mu_3^k \end{pmatrix} &= V\mathrm{Diag}(d^k)V^{-1} \begin{pmatrix} x_2^0 \\ x_3^0 \\ \mu_1^0 \\ \mu_2^0 \\ \mu_3^0 \end{pmatrix} \\ &= V\mathrm{Diag}(d^k) \begin{pmatrix} l_1 \\ l_2 \\ l_3 \\ l_4 \\ l_5 \end{pmatrix} \\ &= V \begin{pmatrix} l_1(0.9836 + 0.2984i)^k \\ l_2(0.9836 - 0.2984i)^k \\ l_3(0.8744 + 0.2310i)^k \\ l_4(0.8744 - 0.2310i)^k \\ 0 \end{pmatrix}, \end{aligned}$$

Thus, as long as $l_1 l_2 \neq 0$, the sequence would be divergent and there is no way for it to converge to a solution point of (3.1).

There are many choices of the starting point $(x_2^0, x_3^0, \mu_1^0, \mu_2^0, \mu_3^0)$ such that $l_1 l_2 \neq 0$. For example,

$$\begin{pmatrix} x_2^0 \\ x_3^0 \\ \mu_1^0 \\ \mu_2^0 \\ \mu_3^0 \end{pmatrix} = V \begin{pmatrix} \alpha_1 \\ \alpha_1 \\ \alpha_2 \\ \alpha_2 \\ \alpha_3 \end{pmatrix}, \tag{3.13}$$

where α_i are any *real* numbers and $\alpha_1 \neq 0$ (which implies that $l_1 = l_2 = \alpha_1 \neq 0$). Furthermore, it is clear that the pair of $V(1)$ and $V(2)$ are two complex conjugate vectors, so are the pair of $V(3)$ and $V(4)$, where $V(i)$ denotes the i-th column of V. Thus the starting point of (3.13) is *real*-valued.

Since $(\alpha_1 > 0, \alpha_2, \alpha_3) \in \Re^3$ form a continuously dense half space, the non-convergent starting points given by (3.13) with $\alpha_1 > 0$ also form a continuously dense half space. Thus, we conclude the main result of this paper as follows.

Theorem 3.1. *For the three-block convex minimization problem (1.1), there is an example where the direct extension of ADMM* (1.5) *is divergent for any penalty parameter* $\beta > 0$ *and for any starting-point in a continuously dense half space of dimension* 3.

4 Extensions

In this section, we extend our previous analysis to some relevant work in the literature.

4.1 Strongly Convex Case of (1.1)

When all functions θ_i's in (1.1) are further assumed to be strongly convex and the penalty parameter β is restricted into a specific range determined by all the strong convex modulus of these functions, the direct extension of ADMM (1.5) is convergent as proved in [14].

Then, it is interesting to ask whether the scheme (1.5) for a strongly convex minimization model is still convergent when the restriction on β in [14] is removed. In other words, does the strong convexity of the objective function help the convergence of the direct extension of ADMM for the three block convex minimization problem (1.1)? A by-product of this paper is a negative answer to the question.

Theorem 4.1. *For the model (1.1) with the strong convex assumption on its objective function, the direct extension of ADMM (1.5) is not necessarily convergent for all* $\beta > 0$.

Recall that the requirement $\rho(M) > 1$ yields the divergence of the direct extension of ADMM (1.5) when it is applied to solve (3.1). Consider the following strongly convex minimization problem with three variables:

$$\begin{array}{ll} \min & 0.05x_1^2 + 0.05x_2^2 + 0.05x_3^2 \\ \text{s.t.} & \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 2 \\ 1 & 2 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = 0. \end{array} \tag{4.1}$$

One can verify that each iteration of the direct extension of ADMM (1.5) applied to the problem remains a fixed matrix mapping. Based on a simple calculation, it is seen that for (4.1), the spectral radius of the matrix involved in (1.5) with $\beta = 1$ is 1.0087. Thus, by a similar discussion to that in Section 3.2, one can find a proper starting point such that the direct extension of ADMM (1.5) with $\beta = 1$ is divergent. The detail is omitted for succinctness.

4.2 Application to the ADMM Variant with a Small Step-size in [19]

To tackle the convergence ambiguity of the direct extension of ADMM (1.5), it was recently proposed in [19][5] to attach a shrinkage step-size factor to the Lagrange-multiplier updating step in (1.5). An

[5] A more general model with m block of functions and variables was considered in [19]. But here, for the convenience of notation, we only focus on the model (1.1) with $m = 3$ and the analysis can be trivially extended to the general case with a generic m.

interesting ADMM variant was thus proposed whose iterative scheme differs from (1.5) only in the step of updating the Lagrange multiplier:

$$\lambda^{k+1} = \lambda^k - \gamma\beta(A_1x_1^k + A_2x_2^k + A_3x_3^k - b), \tag{4.2}$$

where the "step-size" factor γ is required to be sufficiently small to ensure that a certain error-bound condition is satisfied. With some additional assumptions on θ_i's and A_i's, this ADMM variant would be linearly convergent; as it was rigorously proved in [19].

The sufficiently small requirement on γ indeed plays a significant theoretical role in ensuring the linear convergence of the ADMM variant in [19], and the requirement depends on the objective function and problem data. It would be valuable to investigate the possibility of a pre-determined range for the step-size factor γ, say depending only on the number of blocks in the model. This possibility is not unreasonable. When there is only one block of variable and function in the model, the ADMM reduces to the standard ALM [18, 23], and it is convergent for all $\gamma \in (0, 2)$ as demonstrated in [1, 13, 7, 24]. When there are two blocks of variable and function in the model, the original ADMM (1.4) is convergent for all $\gamma \in (0, \frac{\sqrt{5}+1}{2})$ as shown in [10]. Thus, in these two cases, the convergence is guaranteed even for selecting $\gamma > 1$ in the Lagrange-multiplier update, which typically results in a numerical acceleration of the convergence.

Similarly, we ask whether or not a problem-data-independent positive step-size interval for γ exists such that the ADMM variant in [19] is guaranteed to be convergent. In the next, we construct an example to give a negative answer numerically. In particular, we again consider the linear equation example (3.1) but the matrix A is given by

$$A = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1+\gamma \\ 1 & 1+\gamma & 1+\gamma \end{pmatrix} \tag{4.3}$$

where the positive scalar $\gamma > 0$ is the same step-size factor in (4.2). (Note that the matrix in (3.10) used to show the divergence of the direct extension of ADMM (1.5) is a special case of (4.3) with $\gamma = 1$.)

Let $M(\gamma)$ be the mapping matrix when the ADMM variant (4.2) is applied to the problem (3.1) with the new matrix (4.3). Recall that the matrix mapping is divergent from a certain initial point if the spectral radius of $M(\gamma)$, denoted by $\rho(M(\gamma))$, is greater than 1. In Table 1, we report the values of $\rho(M(\gamma))$ for several choices of γ ranging from 1e-8 to 1. It was found that the ADMM variant remains divergent even if γ is as small as 10^{-8}. It is thus numerically demonstrated that the ADMM variant (4.2) is still divergent even if the step-size factor γ is very small. In fact, not similar as the ALM and the original ADMM, we conjecture that there does not exist a problem-data-independent interval in which any value of the step-size factor γ can ensure the convergence of the ADMM variant (4.2).

Table 1: Spectral radius of the small step-size variant of ADMM ($\beta = 1$)

γ	1	0.1	1e-2	1e-3	1e-4	1e-5	1e-6	1e-7	1e-8
$\rho(M(\gamma))$	1.027839	1.002637	1.000105	1.000004	>1	>1	>1	>1	>1

5 Conclusions

We have shown by an example that the direct extension of the alternating direction method of multiplier (ADMM) is not necessarily convergent for solving a convex minimization model with linear constraints and an objective function in sum of three separable convex functions; which solves a long-standing open problem. Based on the strategies to construct the divergent example, we give answers to some other questions related to the direct extension of ADMM.

The negative answer to the open question thus justifies the rationale of algorithmic design in recent work such as [15, 16], where it was suggested to combine several correction steps with the output of the direct extension of ADMM in order to produce a splitting algorithm with provable convergence under mild assumptions for multi-block convex minimization models.

We also studied a condition that can guarantee the convergence of the direct extension of ADMM. This new sufficient condition is significantly different from those in the literature which often require strong convexity on the objective functions and/or restrictive choices for the penalty parameter. Instead, the new condition simply depends on the orthogonality of the given coefficient matrices in the model and poses no restriction on how to choose the penalty parameter β in algorithmic implementation.

Although our results have focused on the model (1.1) where there are three blocks of variable and function, our analyses can be easily extended to the application of the direct extension of ADMM to a more general convex minimization model where the block number of variable and function is greater than 3.

References

[1] D.P. Bertsekas, Constrained Optimization and Lagrange Multiplier Methods, Academic Press, 1982.

[2] E. Blum and W. Oettli, Mathematische Optimierung. Grundlagen und Verfahren. Ökonometrie und Unternehmensforschung, Springer-Verlag, Berlin-Heidelberg-New York, 1975.

[3] S. Boyd, N. Parikh, E. Chu, B. Peleato and J. Eckstein, *Distributed optimization and statistical learning via the alternating direction method of multipliers*, Found. Trends Mach. Learning, 3 (2010), pp. 1-122.

[4] T. F. Chan and R. Glowinski, *Finite element approximation and iterative solution of a class of mildly non-linear elliptic equations*, technical report, Stanford University, 1978.

[5] V. Chandrasekaran, P. A. Parrilo, and A. S. Willsky, *Latent variable graphical model selection via convex optimization*, Ann. Stat., 40 (2012), pp. 1935-1967.

[6] J. Eckstein, *Augmented Lagrangian and alternating direction methods for convex optimization: A tutorial and some illustrative computational results*, manuscript, 2012.

[7] J. Eckstein and D.P. Bertsekas, *On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators*, Math. Program., 55 (1992), pp. 293-318.

[8] M. Fortin and R. Glowinski, *On Decomposition-Coordination Methods Using an Augmented Lagrangian.* In M. Fortin and R. Glowinski, eds., Augmented Lagrangian Methods: Applications to the solution of Boundary Problems. North- Holland, Amsterdam, 1983.

[9] D. Gabay and B. Mercier, *A dual algorithm for the solution of nonlinear variational problems via finite element approximations*, Comput. Math. Appli., 2 (1976), pp. 17-40.

[10] R. Glowinski, *Numerical methods for nonlinear variational problems*, Springer, 1984.

[11] R. Glowinski, *On alternating directon methods of multipliers: a historical perspective*, Springer Proceedings of a Conference Dedicated to J. Periaux, to appear.

[12] R. Glowinski and A. Marrocco, *Approximation par éléments finis d'ordre un et résolution par pénalisation-dualité d'une classe de problèmes non linéaires*, R.A.I.R.O., R2 (1975), pp. 41-76.

[13] E. G. Gol'shtein and N. V. Tret'yakov, *Modified Lagrangian in convex programming and their generalizations*, Math. Program. Studies, 10 (1979), pp. 86-97.

[14] D. R. Han and X.M. Yuan, *A note on the alternating direction method of multipliers*, J. Optim. Theory Appli., 155 (2012), pp. 227–238.

[15] B. S. He, M. Tao and X. M. Yuan, *Alternating direction method with Gaussian back substitution for separable convex programming*, SIAM J. Optim., 22 (2012), pp. 313-340.

[16] B. S. He, M. Tao and X. M. Yuan, *Convergence rate and iteration complexity on the alternating direction method of multipliers with a substitution procedure for separable convex programming*, Math. Oper. Res., under revision.

[17] B. S. He and X.M. Yuan. On the $O(1/n)$ convergence rate of the Douglas-Rachford alternating direction method, SIAM J. Num. Anal., 50 (2012), pp. 700–709.

[18] M. R. Hestenes, *Multiplier and gradient methods*, J. Optim. Theory Appli., 4 (1969), pp. 303-320.

[19] M. Hong and Z. Q. Luo, *On the linear convergence of the alternating direction method of multipliers*, manuscript, August 2012.

[20] G. J. McLachlan, Discriminant analysis and statistical pattern recognition, volume 544. Wiley-Interscience, 2004.

[21] K. Mohan, P. London, M. Fazel, D. Witten, and S. Lee. Node-based learning of multiple gaussian graphical models. *avialable at: arXiv:/1303.5145*, 2013.

[22] Y. G. Peng, A. Ganesh, J. Wright, W. L. Xu and Y. Ma, *Robust alignment by sparse and low-rank decomposition for linearly correlated images*, IEEE Tran. Pattern Anal. Mach. Intel. 34 (2012), pp. 2233-2246.

[23] M. J. D. Powell, *A method for nonlinear constraints in minimization problems*, In Optimization edited by R. Fletcher, pp. 283-298, Academic Press, New York, 1969.

[24] R.T. Rockafellar, *Augmented Lagrangians and applications of the proximal point algorithm in convex programming*, Math. Oper. Res., 1 (1976), pp. 97-116.

[25] M. Tao and X. M. Yuan, *Recovering low-rank and sparse components of matrices from incomplete and noisy observations*, SIAM J. Optim., 21 (2011), pp. 57-81.