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]

December 1, 2013

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 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 affirmatively proved convergence nor counter example showing its failure of convergence is known in the literature. In this paper we answer this long-standing open question: the direct extension of ADMM is not necessarily convergent. We present an example showing its failure of convergence, and a sufficient condition ensuring its convergence.

Keywords. Alternating direction method of multipliers, Convergence analysis, Convex programming, Sparse optimization, Low-rank Optimization, Image processing, Statistical learning, Computer vision

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:

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

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 [18], the robust principal component analysis model with noisy and incomplete data in [20], the latent variable Gaussian graphical model selection in

[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. 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`.

[4, 17] and the quadratic discriminant analysis model in [16]. Our discussion is inspired by the 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 advantageous properties of each individual θ_i are ignored or not fully used.

The alternating direction method of multipliers (ADMM) was originally proposed in [9] (see also [3, 7]), 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 $\lambda\in\Re^p$ the Lagrange multiplier and $\beta>0$ a penalty parameter. Then, the iterative scheme of ADMM for (1.2) is

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

The iterative scheme of ADMM embeds a Gaussian-Seidel decomposition into each iteration of the augmented Lagrangian method in [14, 19]; 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, ADMM has recently witnessed a "renaissance" in many application domains after a long period without too much attention. We refer to [2, 5, 8] for some review papers on ADMM.

With the same philosophy as 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

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

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. [18, 20]. The convergence of (1.5), however, has been ambiguous for a long time — there is neither affirmative convergence proof nor counter example

showing its failure of convergence in the literature. This convergence ambiguity has inspired an active research topic in developing such algorithms that are some slightly twisted versions of (1.5) but with provable convergence and competitive numerical efficiency and iteration simplicity, see e.g. [11, 12, 15]. Since the extended ADMM scheme (1.5) does work well for some applications (e.g. [18, 20]), users have the inclination to imagine that this scheme seems convergent even though they are perplexed by the rigorous proof. In the literature, there is even very little hint for the difficulty in the convergence proof for (1.5), see [5] for an insightful explanation.

The main purpose of this paper is to answer this long-standing open question negatively: The extended ADMM scheme (1.5) is not necessarily convergent. We will give an example (and a strategy for constructing such an example) to demonstrate its failure of convergence, and show that the convergence of (1.5) can be guaranteed when any two coefficient matrices in (1.1) are orthogonal.

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 failure of convergence for (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 conditions in the literature such as the one in [10] which requires strong convexity on all functions in the objective and restricts the choice of the penalty parameter β into a specific range; and the one in [15] which requires to attach a sufficiently small shrinkage factor to the Lagrange-multiplier updating step (1.5d) such that a certain error-bound condition is satisfied.

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

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

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

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

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

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

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 in (2.3), both x_1^k and x_2^k are not required to generate the $(k+1)$-th iteration under the orthogonality condition $A_1^TA_2 = 0$. Existing convergence results for the original ADMM such as those in [6, 13] thus hold for the special case of (1.5) with the orthogonality condition $A_1^TA_2 = 0$.

Similar discussion can be carried out under the orthogonality condition $A_2^TA_3 = 0$.

2.2 Case 2: $A_1^TA_3 = 0$

In the last subsection, we have discussed the cases where two consecutive coefficient matrices are orthogonal. Now, we pay more attention to the case where $A_1^TA_3 = 0$ 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 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. We thus can rewrite (1.5) in the form

$$\begin{cases} x_2^{k+1} = \text{Argmin}\{\mathcal{L}_{\mathcal{A}}(x_1^{k+1}, x_2, x_3^k, \lambda^k) \,|\, x_2 \in \mathcal{X}_2\}, & (2.4a) \\ x_3^{k+1} = \text{Argmin}\{\mathcal{L}_{\mathcal{A}}(x_1^{k+1}, x_2^{k+1}, x_3, \lambda^k) \,|\, x_3 \in \mathcal{X}_3\}, & (2.4b) \\ \lambda^{k+1} = \lambda^k - \beta(A_1x_1^{k+1} + A_2x_2^{k+1} + A_3x_3^{k+1} - b), & (2.4c) \\ x_1^{k+2} = \text{Argmin}\{\mathcal{L}_{\mathcal{A}}(x_1, x_2^{k+1}, x_3^{k+1}, \lambda^{k+1}) \,|\, x_1 \in \mathcal{X}_1\}. & (2.4d) \end{cases}$$

According to (2.4), there is a update for the variable λ between the updates for x_3 and x_1. Thus, the case $A_1^TA_3 = 0$ requires discussion different from that in the last subsection. We will focus on this representation of (1.5) 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 $(x_1^{k+1}, x_3^k, \lambda^k)$. We thus follow the terminology in [2] 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 $w^k = (x_1^{k+1}, x_2^k, x_3^k, \lambda^k)$, $u^k = w^k \setminus \lambda^k = (x_1^{k+1}, x_2^k, x_3^k)$, $v^k = w^k \setminus x_2^k = (x_1^{k+1}, x_3^k, \lambda^k)$, $v = w \setminus x_2 = (x_1, x_3, \lambda)$, $\mathcal{V} = \mathcal{X}_1 \times \mathcal{X}_3 \times \Re^p$ and

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

Note the first element in w^k, u^k or v^k is x_1^{k+1} rather than x_1^k.

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

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

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

$$\begin{cases} \theta_2(x_2) - \theta_2(x_2^{k+1}) + (x_2 - x_2^{k+1})^T\{-A_2^T[\lambda^k - \beta(A_1x_1^{k+1} + A_2x_2^{k+1} + A_3x_3^k - b)]\} \geq 0, \quad \forall x_2 \in \mathcal{X}_2, & (2.6a) \\ \theta_3(x_3) - \theta_3(x_3^{k+1}) + (x_3 - x_3^{k+1})^T\{-A_3^T[\lambda^k - \beta(A_1x_1^{k+1} + A_2x_2^{k+1} + A_3x_3^{k+1} - b)]\} \geq 0, \ \forall x_3 \in \mathcal{X}_3, & (2.6b) \\ (A_1x_1^{k+1} + A_2x_2^{k+1} + A_3x_3^{k+1} - b) + \dfrac{1}{\beta}(\lambda^{k+1} - \lambda^k) = 0, & (2.6c) \\ \theta_1(x_1) - \theta_1(x_1^{k+2}) + (x_1 - x_1^{k+2})^T\{-A_1^T[\lambda^{k+1} - \beta(A_1x_1^{k+2} + A_2x_2^{k+1} + A_3x_3^{k+1} - b)]\} \geq 0, \ \forall x_1 \in \mathcal{X}_1. & (2.6d) \end{cases}$$

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

$$\begin{cases} \theta_2(x_2) - \theta_2(x_2^{k+1}) + (x_2 - x_2^{k+1})^T\{-A_2^T\lambda^{k+1} + \beta A_2^TA_3(x_3^k - x_3^{k+1})\} \geq 0, \quad \forall x_2 \in \mathcal{X}_2, & (2.7a) \\ \theta_3(x_3) - \theta_3(x_3^{k+1}) + (x_3 - x_3^{k+1})^T\{-A_3^T\lambda^{k+1}\} \geq 0, \ \forall x_3 \in \mathcal{X}_3, & (2.7b) \\ (A_1x_1^{k+2} + A_2x_2^{k+1} + A_3x_3^{k+1} - b) + A_1(x_1^{k+1} - x_1^{k+2}) - \dfrac{1}{\beta}(\lambda^k - \lambda^{k+1}) = 0, & (2.7c) \\ \theta_1(x_1) - \theta_1(x_1^{k+2}) + (x_1 - x_1^{k+2})^T\{-A_1^T\lambda^{k+1} - \beta A_1^TA_1(x_1^{k+1} - x_1^{k+2}) + A_1^T(\lambda^k - \lambda^{k+1})\} \geq 0, \ \forall x_1 \in \mathcal{X}_1. & (2.7d) \end{cases}$$

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 $w^{k+1} = (x_1^{k+2}, x_2^{k+1}, x_3^{k+1}, \lambda^{k+1})$ be generated by (2.4) from given $v^k = (x_1^{k+1}, x_3^k, \lambda^k)$. 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

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

Note the assertion (2.8) is useful for quantifying the accuracy of w^{k+1} to a solution point of $\mathrm{VI}(\Omega, F, \theta)$, 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 $w^{k+1} = (x_1^{k+2}, x_2^{k+1}, x_3^{k+1}, \lambda^{k+1})$ be generated by (2.4) from given $v^k = (x_1^{k+1}, x_3^k, \lambda^k)$. 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:

$$\begin{aligned} &\begin{pmatrix} x_1 - x_1^{k+2} \\ x_2 - x_2^{k+1} \\ x_3 - x_3^{k+1} \\ \lambda - \lambda^{k+1} \end{pmatrix}^T \begin{pmatrix} \beta A_1^T A_1 & \beta A_1^T A_3 & -A_1^T \\ 0 & 0 & 0 \\ 0 & \beta A_3^T A_3 & 0 \\ -A_1 & 0 & \frac{1}{\beta} I \end{pmatrix} \begin{pmatrix} x_1^{k+1} - x_1^{k+2} \\ x_3^k - x_3^{k+1} \\ \lambda - \lambda^{k+1} \end{pmatrix} \\ &= \begin{pmatrix} x_1 - x_1^{k+2} \\ x_3 - x_3^{k+1} \\ \lambda - \lambda^{k+1} \end{pmatrix}^T \begin{pmatrix} \beta A_1^T A_1 & 0 & -A_1^T \\ 0 & \beta A_3^T A_3 & 0 \\ -A_1 & 0 & \frac{1}{\beta} I \end{pmatrix} \begin{pmatrix} x_1^{k+1} - x_1^{k+2} \\ x_3^k - x_3^{k+1} \\ \lambda^k - \lambda^{k+1} \end{pmatrix}. \end{aligned}$$

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:

$$\tilde{w}^k = \begin{pmatrix} \tilde{x}_1^k \\ \tilde{x}_2^k \\ \tilde{x}_3^k \\ \tilde{\lambda}^k \end{pmatrix} = \begin{pmatrix} x_1^{k+2} \\ x_2^{k+1} \\ x_3^{k+1} \\ \lambda^{k+1} - \beta A_3 (x_3^k - x_3^{k+1}) \end{pmatrix} \quad \text{and} \quad \tilde{u}^k = \begin{pmatrix} \tilde{x}_1^k \\ \tilde{x}_2^k \\ \tilde{x}_3^k \end{pmatrix}, \tag{2.13}$$

where $\{x_1^{k+2}, x_2^{k+1}, x_3^{k+1}, \lambda^{k+1}\}$ 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 $w^{k+1} = (x_1^{k+2}, x_2^{k+1}, x_3^{k+1}, \lambda^{k+1})$ be generated by (2.4) from given $v^k = (x_1^{k+1}, x_3^k, \lambda^k)$. If $A_1^T A_3 = 0$, we have $\tilde{w}^k \in \Omega$ and*

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

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

Proof. According to the definition of $\tilde{w}^k$, (2.10) can be rewritten as

$$\begin{aligned}\tilde{w}^k \in \Omega, \quad \theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq \beta(A_1 x_1^{k+2} + A_2 x_2^{k+1} + A_3 x_3^{k+1} - b)^T A_3(x_3^k - x_3^{k+1}) \\ +(v - v^{k+1})^T H(v^k - v^{k+1}), \quad \forall w \in \Omega,\end{aligned} \tag{2.15}$$

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

$$\theta_3(x_3^k) - \theta_3(x_3^{k+1}) + (x_3^k - x_3^{k+1})^T\{-A_3^T \lambda^{k+1}\} \geq 0. \tag{2.16}$$

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

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

which together with (2.16) yields that

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

By using the fact $\lambda^k - \lambda^{k+1} = \beta(A_1 x_1^{k+1} + A_2 x_2^{k+1} + A_3 x_3^{k+1} - b)$ and the assumption $A_1^T A_3 = 0$, we get immdeiately that

$$\beta(A_1 x_1^{k+2} + A_2 x_2^{k+1} + A_3 x_3^{k+1} - b)^T A_3(x_3^k - x_3^{k+1}) \geq 0, \tag{2.19}$$

and hence

$$\tilde{w}^k \in \Omega, \quad \theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (v - v^{k+1})^T H(v^k - v^{k+1}), \quad \forall\, w \in \Omega. \tag{2.20}$$

By substituting the identity

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

into the right-hand side of (2.20), we obtain (2.14). □

Now, we are able to establish the contraction property with respect to the solution set of $\mathrm{VI}(\Omega, F, \theta)$ 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 $\{x_1^{k+1}, x_2^k, x_3^k, \lambda^k\}$ be the sequence generated by the direct extension of ADMM (2.4). Then, we have:*

(i) *The sequence* $\{v^k := (x_1^{k+1}, x_3^k, \lambda^k)\}$ *is contractive with respective to the solution of VI*(Ω, F, θ)*, i.e.,*

$$\|v^{k+1} - v^*\|_H^2 \leq \|v^k - v^*\|_H^2 - \|v^k - v^{k+1}\|_H^2. \tag{2.21}$$

(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 (2.21) is proved. Clearly, (2.21) indicates that the sequence $\{v^k\}$ is contractive with respect to the solution set of VI(Ω, F, θ), see e.g. [1].

(ii) To prove (ii), by the inequality (2.21), it follows that the sequences $\{A_1 x_1^{k+1} - \frac{1}{\beta}\lambda^k\}$ 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

$$A_1 x_1^{k+1} + A_2 x_2^k = A_1 x_1^{k+1} - \frac{1}{\beta}\lambda^k - (A_1 x_1^k - \frac{1}{\beta}\lambda^{k-1}) - A_3 x_3^k + b. \tag{2.22}$$

Hence, $\{A_1 x_1^{k+1} + A_2 x_2^k\}$ is bounded. Together with the assumption that $[A_1,\, A_2]$ has full column rank, we conclude that the sequences $\{x_1^{k+1}\}, \{x_2^k\}$ and $\{\lambda^k\}$ are all bounded. Therefore, there exists a subsequence $\{x_1^{n_k+2}, x_2^{n_k+1}, x_3^{n_k+1}, \lambda^{n_k+1}\}$ that converges to a limit point, say $(x_1^\infty, x_2^\infty, x_3^\infty, \lambda^\infty)$. Moreover, from (2.21), we see immediately that

$$\sum_{k=1}^{\infty} \|v^k - v^{k+1}\|_H^2 < +\infty, \tag{2.23}$$

which shows

$$\lim_{k\to\infty} H(v^k - v^{k+1}) = 0, \tag{2.24}$$

and thus

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

Then, by taking the limits on the both sides of (2.8), using (2.25), one can immediately write

$$w^\infty \in \Omega, \quad \theta(u) - \theta(u^\infty) + \ (w - w^\infty)^T F(w^\infty) \geq 0, \ \ \forall\, w \in \Omega, \tag{2.26}$$

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 (2.21) 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

$$\|v^{k+1} - v^\infty\|_H^2 \leq \|v^k - v^\infty\|_H^2, \tag{2.27}$$

which implies that

$$\lim_{k\to\infty} A_1(x_1^{k+1} - x_1^\infty) - \frac{1}{\beta}(\lambda^k - \lambda^\infty) = 0, \quad \lim_{k\to\infty} A_3(x_3^k - x_3^\infty) = 0. \tag{2.28}$$

By taking limits to (2.22), using (2.28) and the assumptions, we know

$$\lim_{k\to\infty} x_1^{k+1} = x_1^\infty, \quad \lim_{k\to\infty} x_2^k = x_2^\infty, \quad \lim_{k\to\infty} x_3^k = x_3^\infty, \quad \lim_{k\to\infty} \lambda^k = \lambda^\infty. \tag{2.29}$$

which completes the proof of this theorem. □

Inspired by [13], we can also establish a worst-case convergence rate measured by the iteration complexity in 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*

$$\tilde{w}_t = \frac{1}{t+1}\sum_{k=0}^{t} \tilde{w}^k. \tag{2.30}$$

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

$$\theta(\tilde{u}_t) - \theta(u) + (\tilde{w}_t - w)^T F(w) \le \frac{1}{2(t+1)}\|v - v^0\|_H^2, \quad \forall w \in \Omega. \tag{2.31}$$

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

$$\tilde{w}^k \in \Omega, \;\; \theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(w) + \frac{1}{2}\|v - v^k\|_H^2 \ge \frac{1}{2}\|v - v^{k+1}\|_H^2, \; \forall w \in \Omega. \tag{2.32}$$

Together with the convexity of $\mathcal{X}_1$, $\mathcal{X}_2$ and $\mathcal{X}_3$, (2.30) implies that $\tilde{w}_t \in \Omega$. Summing the inequality (2.32) 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

$$\frac{1}{t+1}\sum_{k=0}^{t}\theta(\tilde{u}^k) - \theta(u) + (\tilde{w}_t - w)^T F(w) \le \frac{1}{2(t+1)}\|v - v^0\|_H^2, \quad \forall w \in \Omega. \tag{2.33}$$

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 (2.33), the assertion of this theorem follows directly. □

Remark 2.6. For an arbitrarily given compact set $D \subset \Omega$, let $d = \sup\{\|v - v^0\|_H^2\} \,|\, v = w\backslash x_2, w \in D\}$, where $v^0 = (x_1^1, x_3^0, \lambda^0)$. Then, after t iterations of the extended ADMM (2.4) , the point $\tilde{w}_t$ defined in (2.30) satisfies

$$\sup\{\theta(\tilde{u}_t) - \theta(u) + (\tilde{w}_t - w)^T F(w)\} \le \frac{d}{2(t+1)}, \tag{2.34}$$

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 ergodic sense is established for the direct extension of ADMM (2.4).

3 An Example Showing the Failure of Convergence 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 is convergent in any order. In this section, we give an example to show the failure of convergence when (1.5) is applied to solve (1.1) without such an orthogonality assumption.

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

$$x_1A_1 + x_2A_2 + x_3A_3 = 0, \tag{3.1}$$

where $A_i \in \Re^3$ $(i = 1, 2, 3)$ are all column vectors and $[A_1,\ A_2,\ A_3]$ 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 zero, b is the 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, and the corresponding optimal Lagrange multiplier is 0.

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) in general setting with $\beta = 1$ by the particular setting in (3.1), we obtain

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

It follows from the first equation in (3.2) that

$$x_1^{k+1} = \frac{1}{A_1^TA_1}\big(-A_1^TA_2x_2^k - A_1^TA_3x_3^k + A_1^T\lambda^k\big). \tag{3.3}$$

Substituting (3.3) into (3.2b), (3.2c) and (3.2d), we obtain a reformulation of (3.2)

$$\begin{aligned} &\begin{pmatrix} A_2^TA_2 & 0 & 0_{1\times3} \\ A_3^TA_2 & A_3^TA_3 & 0_{1\times3} \\ A_2 & A_3 & I_{3\times3} \end{pmatrix} \begin{pmatrix} x_2^{k+1} \\ x_3^{k+1} \\ \lambda^{k+1} \end{pmatrix} \\ &= \left[\begin{pmatrix} 0 & -A_2^TA_3 & A_2^T \\ 0 & 0 & A_3^T \\ 0_{3\times1} & 0_{3\times1} & I_{3\times3} \end{pmatrix} - \frac{1}{A_1^TA_1} \begin{pmatrix} A_2^TA_1 \\ A_3^TA_1 \\ A_1 \end{pmatrix} \big(-A_1^TA_2, -A_1^TA_3, A_1^T\big) \right] \begin{pmatrix} x_2^k \\ x_3^k \\ \lambda^k \end{pmatrix}. \end{aligned} \tag{3.4}$$

Let

$$L = \begin{pmatrix} A_2^TA_2 & 0 & 0_{1\times3} \\ A_3^TA_2 & A_3^TA_3 & 0_{1\times3} \\ A_2 & A_3 & I_{3\times3} \end{pmatrix} \tag{3.5}$$

and

$$R = \begin{pmatrix} 0 & -A_2^T A_3 & A_2^T \\ 0 & 0 & A_3^T \\ 0_{3\times 1} & 0_{3\times 1} & I_{3\times 3} \end{pmatrix} - \frac{1}{A_1^T A_1} \begin{pmatrix} A_2^T A_1 \\ A_3^T A_1 \\ A_1 \end{pmatrix} \left(-A_1^T A_2, -A_1^T A_3, A_1^T\right). \tag{3.6}$$

Then the iterative formula (3.4) can be rewritten in the following matrix recursion context:

$$\begin{pmatrix} x_2^{k+1} \\ x_3^{k+1} \\ \lambda^{k+1} \end{pmatrix} = M \begin{pmatrix} x_2^{k} \\ x_3^{k} \\ \lambda^{k} \end{pmatrix} = \cdots = M^{k+1} \begin{pmatrix} x_2^{0} \\ x_3^{0} \\ \lambda^{0} \end{pmatrix} \tag{3.7}$$

with

$$M = L^{-1} R. \tag{3.8}$$

Note for R defined in (3.6), we have

$$R \begin{pmatrix} 0 \\ 0 \\ A_1 \end{pmatrix} = 0.$$

Thus, as will be seen in the next subsection, 0 is an eigenvalue of the matrix M.

3.2 A Concrete Example Showing the Divergence of (1.5)

Now we are ready to construct a concrete example to show the divergence when the direct extension of ADMM (1.5) 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 failure of convergence 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 not orthogonal. Moreover, notice that the sequence generated by (3.7) is divergent if $\rho(M) > 1$. We thus construct the matrix A as follows:

$$A = (A_1, A_2, A_3) = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 2 \\ 1 & 2 & 2 \end{pmatrix}. \tag{3.9}$$

Given this matrix A, the system of linear equations (3.4) can be specified as

$$\begin{aligned} &\begin{pmatrix} 6 & 0 & 0 & 0 & 0 \\ 7 & 9 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 2 & 0 & 1 & 0 \\ 2 & 2 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_2^{k+1} \\ x_3^{k+1} \\ \lambda_1^{k+1} \\ \lambda_2^{k+1} \\ \lambda_3^{k+1} \end{pmatrix} \\ &= \left[\begin{pmatrix} 0 & -7 & 1 & 1 & 2 \\ 0 & 0 & 1 & 2 & 2 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} - \frac{1}{3} \begin{pmatrix} 4 \\ 5 \\ 1 \\ 1 \\ 1 \end{pmatrix} \left(-4, -5, 1, 1, 1\right) \right] \begin{pmatrix} x_2^{k} \\ x_3^{k} \\ \lambda_1^{k} \\ \lambda_2^{k} \\ \lambda_3^{k} \end{pmatrix}. \end{aligned}$$

Note with the specification in (3.9), the matrices L in (3.5) and R in (3.6) 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

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

By direct computation, we know that M admits the following eigenvalue decomposition

$$M = V\mathrm{Diag(d)}V^{-1}, \tag{3.10}$$

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

$$|d_1| = |d_2| > 1,$$

which offers us the opportunity to construct a divergent sequence $\{x_2^k, x_3^k, \lambda_1^k, \lambda_2^k, \lambda_3^k\}$. In fact, let us choose the initial point $(x_2^0, x_3^0, \lambda_1^0, \lambda_2^0, \lambda_3^0)^T$ as $V(:,1) + V(:,2)$. It is clear that the vector $V(:,2)$ is the complex conjugate of the vector $V(:,1)$; thus the starting point is real. Then, since

$$V\begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} x_2^0 \\ x_3^0 \\ \lambda_1^0 \\ \lambda_2^0 \\ \lambda_3^0 \end{pmatrix},$$

we know from (3.7) and (3.10) that

$$\begin{aligned}
\begin{pmatrix} x_2^{k+1} \\ x_3^{k+1} \\ \lambda_1^{k+1} \\ \lambda_2^{k+1} \\ \lambda_3^{k+1} \end{pmatrix} &= V\mathrm{Diag}(d^{k+1})V^{-1} \begin{pmatrix} x_2^0 \\ x_3^0 \\ \lambda_1^0 \\ \lambda_2^0 \\ \lambda_3^0 \end{pmatrix} \\
&= V\mathrm{Diag}(d^{k+1}) \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \\
&= V \begin{pmatrix} (\,0.9836 + 0.2984i\,)^{k+1} \\ (\,0.9836 - 0.2984i\,)^{k+1} \\ 0 \\ 0 \\ 0 \end{pmatrix},
\end{aligned}$$

which is definitely divergent and there is no way to converge to the solution point of (3.1).

4 Conclusions

We showed by an example that the direct extension of the alternating direction method of multiplier (ADMM) is not necessarily convergent for solving a multi-block convex minimization model with linear constraints and an objective function which is the sum of more than two convex functions without coupled variables; a long-standing open problem is thus answered. Based on our strategy for constructing this example, it is easy to find more such examples. The negative answer to this open problem thus verifies the rationale of algorithmic design in our recent work such as [11, 12], where it was suggested to combine some correction steps with the output of the direct extension of ADMM in order to produce a splitting algorithm with rigorous 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 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 in algorithmic implementation.

Our analysis focused on the model (1.1) where there are three convex functions in its objective, because this case represents a wide range of applications and it is easier to demonstrate our main idea with this special case. The analysis can be easily extended to the more general multi-block case of convex minimization model where there are $m > 3$ convex functions in its objective. Moreover, based on our strategy of constructing the example showing the divergence of (1.5), it is easy to show that the direct extension of ADMM with some given value of the penalty parameter is still divergent even if the functions θ_i in (1.1) are all further assumed to be strongly convex. We omit the detail of these extensions for succinctness.

References

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

[2] 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.

[3] 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.

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

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

[6] 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.

[7] 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.

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

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[13] 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.

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

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

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

[17] 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.

[18] 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.

[19] 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.

[20] 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.