Sun et al. *Journal of Inequalities and Applications* (2016) 2016:227
DOI 10.1186/s13660-016-1173-2

Journal of Inequalities and Applications
a SpringerOpen Journal

RESEARCH Open Access

On the global and linear convergence of direct extension of ADMM for 3-block separable convex minimization models

Huijie Sun[1,2,3], Jinjiang Wang[3] and Tingquan Deng[1,2*]

*Correspondence: ws_hrbnu@163.com
[1]College of Computer Science and Technology, Harbin Engineering University, Harbin, 150001, China
[2]College of Science, Harbin Engineering University, Harbin, 150001, China
Full list of author information is available at the end of the article

Abstract
In this paper, we show that when the alternating direction method of multipliers (ADMM) is extended directly to the 3-block separable convex minimization problems, it is convergent if one block in the objective possesses sub-strong monotonicity which is weaker than strong convexity. In particular, we estimate the globally linear convergence rate of the direct extension of ADMM measured by the iteration complexity under some additional conditions.

Keywords: linear convergence rate; separable convex optimization; alternating direction method of multiplier; Karush-Kuhn-Tucher (KKT) system; strong monotonicity

1 Introduction

Because there is still a gap between the empirical efficiency of the direct extension of ADMM for a variety of applications and the lack of theoretical conditions that can both ensure the convergence of the direct extension of ADMM and be satisfied by applications, the main attention of this paper is paid to the study of the convergence of the direct extension of ADMM for the 3-block separable convex optimization problems.

We consider the following separable convex minimization problem whose objective function is the sum of three functions without coupled variables:

$$\begin{aligned} &\min \theta_1(x_1) + \theta_2(x_2) + \theta_3(x_3) \\ &\text{s.t. } A_1x_1 + A_2x_2 + A_3x_3 = b, \end{aligned} \tag{1}$$

where $A_i \in \mathcal{R}^{l\times n_i}$ $(i = 1, 2, 3)$, $b \in \mathcal{R}^l$, and $\theta_i : \mathcal{R}^{n_i} \to (-\infty, +\infty]$ $(i = 1, 2, 3)$ are closed proper convex (not necessarily smooth) functions. This model has a lot of applications in practice. For example, the latent variable Gaussian graphical model selection in [1], the quadratic discriminant analysis model in [2] and the robust principal component analysis model with noisy and incomplete data in [3, 4], and so on. The augmented Lagrangian function of (1) is defined as

$$\mathcal{L}_\beta(x_1, x_2, x_3, \lambda) := \sum_{i=1}^{3} \theta_i(x_i) - \left\langle \lambda, \sum_{i=1}^{3} A_ix_i - b \right\rangle + \frac{\beta}{2} \left\| \sum_{i=1}^{3} A_ix_i - b \right\|^2, \tag{2}$$

where $\lambda \in \mathcal{R}^l$ and $\beta > 0$.

The classical alternating direction method of multipliers (ADMM) for solving the 2-block separable convex minimization problems was first introduced by Gabay and Mercier [5] and Glowinski and Marrocco [6], respectively, and its iterative scheme can be described by

$$x_1^{k+1} = \arg\min_{x_1} \left\{ \theta_1(x_1) + \langle \lambda^k, A_1 x_1 \rangle + \frac{\beta}{2} \| A_1 x_1 + A_2 x_2^k - b \|^2 \right\}, \tag{3a}$$

$$x_2^{k+1} = \arg\min_{x_2} \left\{ \theta_2(x_2) + \langle \lambda^k, A_2 x_2 \rangle + \frac{\beta}{2} \| A_1 x_1^{k+1} + A_2 x_2 - b \|^2 \right\}, \tag{3b}$$

$$\lambda^{k+1} = \lambda^k - \alpha_0 \beta \left(A_1 x_1^{k+1} + A_2 x_2^{k+1} - b \right), \tag{3c}$$

where $\alpha > 0$ is called step-length. The convergence of ADMM has been well established in the literature (see [5, 7, 8]). For more details of the ADMM, the reader can also refer to [5, 9–13].

Due to the classical ADMM extreme simplicity and efficiency in numerous applications such as mathematical imaging science, signal processing, and so on, it is natural to extend the classical ADMM (3a)-(3c) directly to (1). The direct extension of the ADMM for solving problem (1) consists of the following iterations:

$$x_1^{k+1} = \arg\min_{x_1} \mathcal{L}_\beta \left(x_1, x_2^k, x_3^k, \lambda^k \right), \tag{4a}$$

$$x_2^{k+1} = \arg\min_{x_2} \mathcal{L}_\beta \left(x_1^{k+1}, x_2, x_3^k, \lambda^k \right), \tag{4b}$$

$$x_3^{k+1} = \arg\min_{x_3} \mathcal{L}_\beta \left(x_1^{k+1}, x_2^{k+1}, x_3, \lambda^k \right), \tag{4c}$$

$$\lambda^{k+1} = \lambda^k - \alpha \beta \left(A_1 x_1^{k+1} + A_2 x_2^{k+1} + A_3 x_3^{k+1} - b \right). \tag{4d}$$

Despite the scheme working very well for many concrete applications of (1) (see *e.g.* [1, 2, 4, 14]), Chen *et al.* [15] showed by a counter example that the convergence of (4a)-(4d) fails. The absence of the convergence of (8) has inspired some improved algorithms. These algorithms are mainly used the following two ways: One way is to correct the output of (4a)-(4d). For example, the authors of [16, 17] added an additional Gaussian back substitution correction step in each iteration after all the block variables are updated. Although, numerically, these algorithms perform slightly slower than the scheme (4a)-(4d), they possess global convergence. The other way is to employ a simple proximal term to solve inexactly the x_i-subproblem in (4a)-(4d), which can make the subproblems of (4a)-(4d) become much easier to carry out and the entire algorithm runs in less time. The readers can refer to [4, 18–25].

On the other hand, several researchers have also studied the convergence of the direct extension of the ADMM (4a)-(4d) by introducing some strong conditions. Han and Yuan [26] have showed that the scheme (4a)-(4d) with $\alpha = 1$ is convergent if the functions θ_i $(i = 1, 2, 3)$ are all strongly convex and the penalty parameter β chosen in a certain interval. Subsequently, these conditions were weakened in [27, 28], and the authors showed that the condition that the two functions are strongly convex can ensure the convergence of (4a)-(4d) with $\alpha = 1$. Recently, these conditions were further weakened, Cai *et al.* [29] had proved that the scheme (4a)-(4d) with $\alpha = 1$ was convergent if one function in the objective is strongly convex. Very recently, Li *et al.* [30] showed that the directly extended 3-block ADMM with $\alpha \in (0, (1+\sqrt{5})/2)$ is convergent, if β is smaller than a certain threshold and the first and third linear operators in the linear equation constraint are full column rank, and the second function in the objective is strongly convex. However, many applications that can be efficiently solved by the scheme (4a)-(4d) will be excluded because of the strong convexity. Thus, these conditions are of only theoretical interests and they seem to be too strict to be satisfied by many mentioned applications.

In the cyclic sense, the scheme (4a)-(4d) can be rewritten as

$$x_1^{k+1} = \arg\min_{x_1} \mathcal{L}_\beta\left(x_1, x_2^k, x_3^k, \lambda^k\right), \tag{5a}$$

$$x_2^{k+1} = \arg\min_{x_2} \mathcal{L}_\beta\left(x_1^{k+1}, x_2, x_3^k, \lambda^k\right), \tag{5b}$$

$$\lambda^{k+1} = \lambda^k - \beta\left(A_1 x_1^{k+1} + A_2 x_2^{k+1} + A_3 x_3^k - b\right), \tag{5c}$$

$$x_3^{k+1} = \arg\min_{x_3} \mathcal{L}_\beta\left(x_1^{k+1}, x_2^{k+1}, x_3, \lambda^{k+1}\right). \tag{5d}$$

In this manuscript, we show that (5a)-(5d) is convergent if one function in the objective of (1) is sub-strongly monotone together with some minor restrictions on the coefficient matrices A_1, A_2, A_3, and the penalty parameter β, which explains why the direct extension of ADMM (4a)-(4d) works well for some applications, even though there are not strong convex functions in such applications. Furthermore, we establish a globally linear convergence rate for the direct extension of ADMM (5a)-(5d) under some additional conditions.

After presenting in Section 2 the needed preliminary material, we devote Section 3 to a proof of the global and linear convergence of the scheme (5a)-(5d) under some assumptions. In Section 4, we construct an example which satisfies the convergence conditions given in Section 3 but do not satisfy the condition that one of the functions in the objective is strongly convex.

2 Preliminaries

In this section we summarize some of notations and the fundamental tools of variational analysis.

We use $\langle \cdot, \cdot \rangle$ to denote the inner product of $\mathcal{R}^n$, and denote by $\|\cdot\|$ its induced norm. $\mathbf{B}_r(x)$ stands for the closed ball of radius r centered at x. Throughout the paper we let all vectors be column vectors. Let A be a symmetric matrix, $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ to denote the smallest eigenvalue and the largest eigenvalue of A, respectively. A real symmetric matrix $A \in \mathcal{R}^{n\times n}$ is called *positive definite* (or *positive semi-definite*) if for all $x \neq 0$, $x^T Ax > 0$ (or $x^T Ax \geq 0$). We denote this as $A \succ 0$ (or $A \succeq 0$). For any real symmetric matrices $A, B \in \mathcal{R}^{n\times n}$, we use $A \succ B$ (or $A \succeq B$) to mean $A - B \succ 0$ (or $A - B \succeq 0$). We denote by $\|x\|_M := \sqrt{x^T Mx}$ the M-norm of the vector x if the matrix M is symmetric and positive definite. For a given matrix A, we use

$$\|A\| := \sup_{x\neq 0} \frac{\|Ax\|}{\|x\|}$$

to denote its norm.

Given a nonempty subset C in $\mathcal{R}^n$, its indicator function is defined as

$$\delta(x;C) := \begin{cases} 0, & x \in C, \\ +\infty, & \text{otherwise.} \end{cases}$$

A function $f : \mathcal{R}^n \to \mathcal{R}$ is convex if

$$f\bigl(\alpha x + (1-\alpha)y\bigr) \le \alpha f(x) + (1-\alpha)f(y), \quad \forall x, y \in \mathcal{R}^n, \forall \alpha \in [0,1],$$

and it is strongly convex with modulus $\mu > 0$ if

$$f\bigl(\alpha x + (1-\alpha)y\bigr) \le \alpha f(x) + (1-\alpha)f(y) - \frac{\mu}{2}\alpha(1-\alpha)\|x-y\|^2, \quad \forall x, y \in \mathcal{R}^n, \forall \alpha \in [0,1].$$

A multifunction $F : \mathcal{R}^n \rightrightarrows \mathcal{R}^n$ (see [31]) is monotone if

$$\langle y_1 - y_2, x_1 - x_2 \rangle \ge 0, \quad \forall y_1 \in F(x_1), \forall y_2 \in F(x_2),$$

and strongly monotone with modulus $\mu > 0$ if

$$\langle y_1 - y_2, x_1 - x_2 \rangle \ge \mu\|x_1 - x_2\|^2, \quad \forall y_1 \in F(x_1), \forall y_2 \in F(x_2).$$

It is well known that a function f is convex if and only if ∂f, the subdifferential of f, is monotone; and f is strongly convex if and only if ∂f is strongly monotone (see, *e.g.*, [31]).

For a differentiable function f, the gradient ∇f is called *Lipschitz continuous* with constant $L_f > 0$ if

$$\bigl\|\nabla f(x) - \nabla f(y)\bigr\| \le L_f\|x-y\|, \quad \forall x, y \in \mathcal{R}^n.$$

For any two vectors x and y with the same dimension, we have

$$2\langle x, y\rangle \le t\|x\|^2 + \frac{1}{t}\|y\|^2, \quad \forall t > 0. \tag{6}$$

Throughout this paper, we make the following standard assumption.

Assumption 2.1 There is a point $(\hat{x}_1, \hat{x}_2, \hat{x}_3) \in \mathrm{ri}(\mathrm{dom}\,\theta_1 \times \mathrm{dom}\,\theta_2 \times \mathrm{dom}\,\theta_3)$ such that $A_1\hat{x}_1 + A_2\hat{x}_2 + A_3\hat{x}_3 = b$.

Suppose that the constraint qualification (CQ) holds, then we know from Corollary 28.2.2 of [31] and Corollary 28.3.1 of [31] that $(x_1^*, x_2^*, x_3^*) \in \mathrm{ri}(\mathrm{dom}\,\theta_1 \times \mathrm{dom}\,\theta_2 \times \mathrm{dom}\,\theta_3)$ is an optimal solution to problem (1) if and only if there exists a Lagrange multiplier $\lambda^* \in \mathcal{R}^l$ such that $(x_1^*, x_2^*, x_3^*, \lambda^*)$ is a solution to the following Karush-Kuhn-Tucher (KKT) system:

$$0 \in \partial\theta_1\bigl(x_1^*\bigr) - A_1^T\lambda^*, \tag{7a}$$

$$0 \in \partial\theta_2\bigl(x_2^*\bigr) - A_2^T\lambda^*, \tag{7b}$$

$$0 \in \partial\theta_3(x_3^*) - A_3^T\lambda^*, \tag{7c}$$

$$0 = A_1x_1^* + A_2x_2^* + A_3x_3^* - b. \tag{7d}$$

We denote by $\mathcal{W}^*$ the set of the solutions of (7a)-(7d).

3 Convergence

In this section, we prove that the iterative sequence $\{(x_1^k, x_2^k, x_3^k, \lambda^k)\}$ generated by the direct extension of ADMM (5a)-(5d) converges to a point $(x_1^*, x_2^*, x_3^*, \lambda^*)$ which is a solution of the KKT system (7a)-(7d) under the following assumption. In the following, the matrices A_1, A_2, and A_3 are assumed to be full column rank. We define the notations

$$G := \begin{pmatrix} \beta A_2^TA_2 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \beta A_3^TA_3 & -A_3^T \\ \mathbf{0} & -A_3 & \frac{1}{\beta}I \end{pmatrix}, \qquad G_1 := \begin{pmatrix} \beta A_2^TA_2 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \beta(1+\frac{3}{\rho})A_3^TA_3 & -A_3^T \\ \mathbf{0} & -A_3 & \frac{1}{\beta}I \end{pmatrix},$$

and

$$w := (x_1, x_2, x_3, \lambda)^T, \qquad v := (x_2, x_3, \lambda)^T,$$

where $\rho > 0$ and β is the penalty parameter in the direct extension of ADMM (5a)-(5d). Then the matrices G and G_1 are symmetric.

3.1 Global convergence

Assumption 3.1 (Sub-strong monotonicity) There exist $(\tilde{x}_1^*, \tilde{x}_2^*, \tilde{x}_3^*, \tilde{\lambda}^*) \in \mathcal{W}^*$ and a real number $\mu_3 > 0$ such that

$$\langle y_3 - A_3^T\tilde{\lambda}^*, x_3 - \tilde{x}_3^*\rangle \geq \mu_3\|x_3 - \tilde{x}_3^*\|^2, \quad \text{for all } x_3 \in \mathcal{R}^{n_3} \text{ and } y_3 \in \partial\theta_3(x_3). \tag{8}$$

Now, we start proving the convergence of the iterative scheme (5a)-(5d) under Assumption 3.1. First, we give several lemmas.

Lemma 3.1 *Suppose Assumption* 3.1 *holds. For the iterative sequence* $\{(x_1^k, x_2^k, x_3^k, \lambda^k)\}$ *generated by the direct extension of ADMM* (5a)-(5d), *then we have*

$$\begin{aligned}(v^{k+1} - \tilde{v}^*)^T G(v^k - v^{k+1}) \geq{}& \langle A_2(x_2^k - x_2^{k+1}), \lambda^k - \lambda^{k+1}\rangle \\ &- \beta\langle A_2(x_2^k - x_2^{k+1}), A_3(x_3^k - \tilde{x}_3^*)\rangle + \mu_3\|x_3^{k+1} - \tilde{x}_3^*\|^2,\end{aligned} \tag{9}$$

where $\tilde{v}^* = (\tilde{x}_2^*, \tilde{x}_3^*, \tilde{\lambda}^*)$ *in* $(\tilde{x}_1^*, \tilde{x}_2^*, \tilde{x}_3^*, \tilde{\lambda}^*)$ *introduced in Assumption* 3.1.

Proof Indeed, the optimality condition of subproblems in (5a)-(5d) can be written as

$$0 \in \partial\theta_1(x_1^{k+1}) - A_1^T\lambda^k + \beta A_1^T(A_1x_1^{k+1} + A_2x_2^k + A_3x_3^k - b), \tag{10a}$$

$$0 \in \partial\theta_2(x_2^{k+1}) - A_2^T\lambda^k + \beta A_2^T(A_1x_1^{k+1} + A_2x_2^{k+1} + A_3x_3^k - b), \tag{10b}$$

$$0 \in \partial\theta_3(x_3^{k+1}) - A_3^T\lambda^{k+1} + \beta A_3^T(A_1x_1^{k+1} + A_2x_2^{k+1} + A_3x_3^{k+1} - b). \tag{10c}$$

Using (5c), (10a)-(10c) can be rewritten as

$$0 \in \partial\theta_1\left(x_1^{k+1}\right) - A_1^T\lambda^{k+1} + \beta A_1^T A_2\left(x_2^k - x_2^{k+1}\right), \tag{11a}$$

$$0 \in \partial\theta_2\left(x_2^{k+1}\right) - A_2^T\lambda^{k+1}, \tag{11b}$$

$$0 \in \partial\theta_3\left(x_3^{k+1}\right) - A_3^T\lambda^{k+1} + A_3^T\left(\lambda^k - \lambda^{k+1}\right) + \beta A_3^T A_3\left(x_3^{k+1} - x_3^k\right). \tag{11c}$$

Using the monotonicity of the subdifferential and Assumption 3.1, it follows from (7a)-(7d) and (11a)-(11c) that we have

$$\left\langle A_1^T\left(\lambda^{k+1} - \tilde{\lambda}^*\right) + \beta A_1^T A_2\left(x_2^{k+1} - x_2^k\right), x_1^{k+1} - \tilde{x}_1^*\right\rangle \geq 0, \tag{12a}$$

$$\left\langle A_2^T\left(\lambda^{k+1} - \tilde{\lambda}^*\right), x_2^{k+1} - \tilde{x}_2^*\right\rangle \geq 0, \tag{12b}$$

$$\begin{aligned}&\left\langle A_3^T\left(\lambda^{k+1} - \tilde{\lambda}^*\right) + A_3^T\left(\lambda^{k+1} - \lambda^k\right) - \beta A_3^T A_3\left(x_3^{k+1} - x_3^k\right), x_3^{k+1} - \tilde{x}_3^*\right\rangle \\ &\quad \geq \mu_3\left\|x_3^{k+1} - \tilde{x}_3^*\right\|^2,\end{aligned} \tag{12c}$$

where $\mu_3 > 0$. Adding up these three inequalities in (12a)-(12c) and using (7d), we obtain

$$\begin{aligned}&\mu_3\left\|x_3^{k+1} - \tilde{x}_3^*\right\|^2 \\ &\quad \leq \left\langle \lambda^{k+1} - \lambda^*, A_1\left(x_1^{k+1} - \tilde{x}_1^*\right) + A_2\left(x_2^{k+1} - \tilde{x}_2^*\right) + A_3\left(x_3^{k+1} - \tilde{x}_3^*\right)\right\rangle \\ &\qquad + \left\langle x_1^{k+1} - \tilde{x}_1^*, \beta A_1^T A_2\left(x_2^{k+1} - x_2^k\right)\right\rangle + \left\langle x_3^{k+1} - \tilde{x}_3^*, A_3^T\left(\lambda^{k+1} - \lambda^k\right)\right\rangle \\ &\qquad + \left\langle x_3^{k+1} - \tilde{x}_3^*, \beta A_3^T A_3\left(x_3^k - x_3^{k+1}\right)\right\rangle \\ &\quad = \left\langle \lambda^{k+1} - \tilde{\lambda}^*, \frac{1}{\beta}\left(\lambda^k - \lambda^{k+1}\right) - A_3\left(x_3^k - x_3^{k+1}\right)\right\rangle \\ &\qquad + \left\langle x_3^{k+1} - \tilde{x}_3^*, \beta A_3^T A_3\left(x_3^k - x_3^{k+1}\right) - A_3^T\left(\lambda^k - \lambda^{k+1}\right)\right\rangle \\ &\qquad + \left\langle x_1^{k+1} - \tilde{x}_1^*, \beta A_1^T A_2\left(x_2^{k+1} - x_2^k\right)\right\rangle.\end{aligned}$$

Using the notations G and v, we further obtain

$$\begin{aligned}&\left(v^{k+1} - \tilde{v}^*\right)^T G\left(v^k - v^{k+1}\right) \\ &\quad \geq \beta\left\langle A_2\left(x_2^k - x_2^{k+1}\right), A_1\left(x_1^{k+1} - \tilde{x}_1^*\right)\right\rangle + \mu_3\left\|x_3^{k+1} - \tilde{x}_3^*\right\|^2 \\ &\qquad + \beta\left\langle A_2\left(x_2^k - x_2^{k+1}\right), A_2\left(x_2^{k+1} - \tilde{x}_2^*\right)\right\rangle \\ &\quad = \beta\left\langle A_2\left(x_2^k - x_2^{k+1}\right), A_1\left(x_1^{k+1} - \tilde{x}_1^*\right) + A_2\left(x_2^{k+1} - \tilde{x}_2^*\right)\right\rangle + \mu_3\left\|x_3^{k+1} - \tilde{x}_3^*\right\|^2 \\ &\quad = \beta\left\langle A_2\left(x_2^k - x_2^{k+1}\right), \frac{1}{\beta}\left(\lambda^k - \lambda^{k+1}\right) - A_3\left(x_3^k - \tilde{x}_3^*\right)\right\rangle + \mu_3\left\|x_3^{k+1} - \tilde{x}_3^*\right\|^2,\end{aligned} \tag{13}$$

which implies (9) and thus completes the proof. □

Lemma 3.2 *There exists a real number $\rho \in (0,1)$ such that the matrix G_1 is symmetric and positive definite.*

Proof Let

$$P_1 := \begin{pmatrix} \beta(1 + \frac{3}{\rho})A_3^T A_3 & -A_3^T \\ -A_3 & \frac{1}{\beta}I \end{pmatrix}.$$

In order to justify the matrix G_1 is symmetric and positive definite, we only need to show the matrix P_1 is positive definite. Since

$$\begin{aligned}\frac{1}{\beta}I - A_3\left[\beta\left(1+\frac{3}{\rho}\right)A_3^TA_3\right]^{-1}A_3^T &= \frac{1}{\beta}I - \frac{1}{\beta}\cdot\frac{\rho}{3+\rho}A_3\left(A_3^TA_3\right)^{-1}A_3^T \\ &\succeq \frac{1}{\beta}I - \frac{1}{\beta}\cdot\frac{\rho}{3+\rho}\cdot\frac{\lambda_{\max}(A_3A_3^T)}{\lambda_{\min}(A_3^TA_3)}I \\ &= \left[1-\frac{\rho}{3+\rho}\cdot\frac{\lambda_{\max}(A_3A_3^T)}{\lambda_{\min}(A_3^TA_3)}\right]\frac{1}{\beta}I.\end{aligned}$$

If $\lambda_{\max}(A_3A_3^T) \leq \lambda_{\min}(A_3^TA_3)$, then for any $\rho \in (0,1)$, we have

$$\frac{1}{\beta}I - A_3\left[\beta\left(1+\frac{3}{\rho}\right)A_3^TA_3\right]^{-1}A_3^T \succ 0.$$

Otherwise, for any

$$\rho \in \left(0, \frac{3\lambda_{\min}(A_3^TA_3)}{\lambda_{\max}(A_3A_3^T) - \lambda_{\min}(A_3^TA_3)}\right),$$

we have

$$\frac{1}{\beta}I - A_3\left[\beta\left(1+\frac{3}{\rho}\right)A_3^TA_3\right]^{-1}A_3^T \succ 0.$$

Thus, it follows from the Schur complement [32], Section A.5.5, that there exists a real number $\rho \in (0,1)$ such that the matrix P_1 is symmetric and positive definite, and so is G_1. □

Lemma 3.3 *Let the iterative sequence* $\{(x_1^k, x_2^k, x_3^k, \lambda^k)\}$ *be generated by the direct extension of ADMM* (5a)-(5d) *with* $\beta \in (0, 2\rho\mu_3/(5\|A_3^TA_3\|))$ *and* $\rho \in (0,1)$ *defined in Lemma* 3.2. *Suppose Assumption* 3.1 *holds. Then there is a real number* $\eta > 0$ *such that*

$$\left\|v^{k+1} - \tilde{v}^*\right\|_{G_1}^2 \leq \left\|v^k - \tilde{v}^*\right\|_{G_1}^2 - \eta\left\|v^k - v^{k+1}\right\|_{G_1}^2. \tag{14}$$

Proof Note that (11b) is also true for $k := k-1$, *i.e.*,

$$0 \in \partial\theta_2\left(x_2^k\right) - A_2^T\lambda^k.$$

Using the monotonicity of the subdifferential $\partial\theta_2$, we have

$$\begin{aligned}0 &\leq \left\langle A_2^T\lambda^k - A_2^T\lambda^{k+1}, x_2^k - x_2^{k+1}\right\rangle \\ &= \left\langle A_2\left(x_2^k - x_2^{k+1}\right), \lambda^k - \lambda^{k+1}\right\rangle.\end{aligned} \tag{15}$$

It follows from (6) that

$$-2\left\langle A_2\left(x_2^k - x_2^{k+1}\right), A_3\left(x_3^k - \tilde{x}_3^*\right)\right\rangle \geq -\rho\left\|A_2\left(x_2^k - x_2^{k+1}\right)\right\|^2 - \frac{1}{\rho}\left\|A_3\left(x_3^k - \tilde{x}_3^*\right)\right\|^2 \tag{16}$$

and

$$\begin{aligned} -\|A_3(x_3^k - x_3^{k+1})\|^2 &= -\|A_3(x_3^k - \tilde{x}_3^*) - A_3(x_3^{k+1} - \tilde{x}_3^*)\|^2 \\ &= -\|A_3(x_3^k - \tilde{x}_3^*)\|^2 + 2\langle A_3(x_3^k - \tilde{x}_3^*), A_3(x_3^{k+1} - \tilde{x}_3^*)\rangle \\ &\quad - \|A_3(x_3^{k+1} - \tilde{x}_3^*)\|^2 \\ &\geq -2\|A_3(x_3^k - \tilde{x}_3^*)\|^2 - 2\|A_3(x_3^{k+1} - \tilde{x}_3^*)\|^2. \end{aligned} \tag{17}$$

Let

$$P := \begin{pmatrix} \beta(1 + \frac{1}{\rho})A_3^T A_3 & -A_3^T \\ -A_3 & \frac{1}{\beta} I \end{pmatrix}.$$

It follows from (9) that

$$\begin{aligned} &(v^k - \tilde{v}^*)^T G(v^k - \tilde{v}^*) \\ &\quad = [(v^k - v^{k+1}) + (v^{k+1} - \tilde{v}^*)]^T G[(v^k - v^{k+1}) + (v^{k+1} - \tilde{v}^*)] \\ &\quad = (v^k - v^{k+1})^T G(v^k - v^{k+1}) + 2(v^{k+1} - \tilde{v}^*)^T G(v^k - v^{k+1}) \\ &\qquad + (v^{k+1} - \tilde{v}^*)^T G(v^{k+1} - \tilde{v}^*) \\ &\quad \geq (v^{k+1} - \tilde{v}^*)^T G(v^{k+1} - \tilde{v}^*) + (v^k - v^{k+1})^T G(v^k - v^{k+1}) \\ &\qquad + 2\mu_3\|x_3^{k+1} - \tilde{x}_3^*\|^2 + 2\langle A_2(x_2^k - x_2^{k+1}), \lambda^k - \lambda^{k+1}\rangle \\ &\qquad - 2\beta\langle A_2(x_2^k - x_2^{k+1}), A_3(x_3^k - \tilde{x}_3^*)\rangle, \end{aligned}$$

which together with (15), (16), and (17) gives

$$\begin{aligned} &(v^k - \tilde{v}^*)^T G(v^k - \tilde{v}^*) \\ &\quad \geq (v^{k+1} - \tilde{v}^*)^T G(v^{k+1} - \tilde{v}^*) + (v^k - v^{k+1})^T G(v^k - v^{k+1}) \\ &\qquad + 2\mu_3\|x_3^{k+1} - \tilde{x}_3^*\|^2 - \beta\rho\|A_2(x_2^k - x_2^{k+1})\|^2 - \frac{\beta}{\rho}\|A_3(x_3^k - \tilde{x}_3^*)\|^2 \\ &\quad = (v^{k+1} - \tilde{v}^*)^T G(v^{k+1} - \tilde{v}^*) + \beta(1-\rho)\|A_2(x_2^k - x_2^{k+1})\|^2 \\ &\qquad + (x_3^k - x_3^{k+1}, \lambda^k - \lambda^{k+1}) P \begin{pmatrix} x_3^k - x_3^{k+1} \\ \lambda^k - \lambda^{k+1} \end{pmatrix} - \frac{\beta}{\rho}\|A_3(x_3^k - x_3^{k+1})\|^2 \\ &\qquad + 2\mu_3\|x_3^{k+1} - \tilde{x}_3^*\|^2 - \frac{\beta}{\rho}\|A_3(x_3^k - \tilde{x}_3^*)\|^2 \\ &\quad \geq (v^{k+1} - \tilde{v}^*)^T G(v^{k+1} - \tilde{v}^*) + \beta(1-\rho)\|A_2(x_2^k - x_2^{k+1})\|^2 \\ &\qquad + (x_3^k - x_3^{k+1}, \lambda^k - \lambda^{k+1}) P \begin{pmatrix} x_3^k - x_3^{k+1} \\ \lambda^k - \lambda^{k+1} \end{pmatrix} + 2\mu_3\|x_3^{k+1} - \tilde{x}_3^*\|^2 \\ &\qquad - \frac{2\beta}{\rho}\|A_3(x_3^{k+1} - \tilde{x}_3^*)\|^2 - \frac{2\beta}{\rho}\|A_3(x_3^k - \tilde{x}_3^*)\|^2 - \frac{\beta}{\rho}\|A_3(x_3^k - \tilde{x}_3^*)\|^2 \\ &\quad \geq (v^{k+1} - \tilde{v}^*)^T G(v^{k+1} - \tilde{v}^*) + \frac{\beta}{\rho}\|A_3(x_3^{k+1} - \tilde{x}_3^*)\|^2 + \frac{2\beta}{\rho}\|A_3(x_3^{k+1} - \tilde{x}_3^*)\|^2 \end{aligned}$$

$$+\beta(1-\rho)\left\|A_2\left(x_2^k-x_2^{k+1}\right)\right\|^2+\left(x_3^k-x_3^{k+1},\lambda^k-\lambda^{k+1}\right)P\begin{pmatrix}x_3^k-x_3^{k+1}\\ \lambda^k-\lambda^{k+1}\end{pmatrix}$$

$$-\frac{\beta}{\rho}\left\|A_3\left(x_3^k-\tilde{x}_3^*\right)\right\|^2-\frac{2\beta}{\rho}\left\|A_3\left(x_3^k-\tilde{x}_3^*\right)\right\|^2+\left(2\mu_3-\frac{5\beta\|A_3^TA_3\|}{\rho}\right)\left\|x_3^{k+1}-\tilde{x}_3^*\right\|^2,$$

which implies that

$$\begin{aligned}&\left(v^{k+1}-\tilde{v}^*\right)^TG\left(v^{k+1}-\tilde{v}^*\right)+\frac{3\beta}{\rho}\left\|A_3\left(x_3^{k+1}-\tilde{x}_3^*\right)\right\|^2\\&\quad\le\left(v^k-\tilde{v}^*\right)^TG\left(v^k-\tilde{v}^*\right)+\frac{3\beta}{\rho}\left\|A_3\left(x_3^k-\tilde{x}_3^*\right)\right\|^2-\left(2\mu_3-\frac{5\beta\|A_3^TA_3\|}{\rho}\right)\left\|x_3^{k+1}-\tilde{x}_3^*\right\|^2\\&\qquad-\beta(1-\rho)\left\|A_2\left(x_2^k-x_2^{k+1}\right)\right\|^2-\left(x_3^k-x_3^{k+1},\lambda^k-\lambda^{k+1}\right)P\begin{pmatrix}x_3^k-x_3^{k+1}\\ \lambda^k-\lambda^{k+1}\end{pmatrix}.\end{aligned}\tag{18}$$

Using the notation G_1 and (18), we have

$$\begin{aligned}\left\|v^{k+1}-\tilde{v}^*\right\|_{G_1}^2\le{}&\left\|v^k-\tilde{v}^*\right\|_{G_1}^2-\left(x_3^k-x_3^{k+1},\lambda^k-\lambda^{k+1}\right)P\begin{pmatrix}x_3^k-x_3^{k+1}\\ \lambda^k-\lambda^{k+1}\end{pmatrix}\\&-\left(2\mu_3-\frac{5\beta\|A_3^TA_3\|}{\rho}\right)\left\|x_3^{k+1}-\tilde{x}_3^*\right\|^2-\beta(1-\rho)\left\|A_2\left(x_2^k-x_2^{k+1}\right)\right\|^2.\end{aligned}\tag{19}$$

To prove such $\eta>0$ exists for (14), we only need $2\mu_3-\frac{5\beta\|A_3^TA_3\|}{\rho}>0$, which holds if $\beta<2\rho\mu_3/(5\|A_3^TA_3\|)$. □

Now, we are ready to prove the convergence of the sequence $\{(x_1^k,x_2^k,x_3^k,\lambda^k)\}$ generated by the direct extension of ADMM (5a)-(5d) under Assumption 3.1. The result is summarized in the following theorem.

Theorem 3.1 *Let the iterative sequence* $\{(x_1^k,x_2^k,x_3^k,\lambda^k)\}$ *be generated by the direct extension of ADMM* (5a)-(5d) *with* $\beta\in(0,2\rho\mu_3/(5\|A_3^TA_3\|))$ *and* $\rho\in(0,1)$ *defined in Lemma* 3.2. *Suppose Assumption* 3.1 *holds. Then the sequence* $\{(x_1^k,x_2^k,x_3^k,\lambda^k)\}$ *converges to a KKT point in* $\mathcal{W}^*$.

Proof It follows from (14) that

$$\lim_{k\to+\infty}\left\|v^k-v^{k+1}\right\|_{G_1}=0\tag{20}$$

and the sequence $\{v^k\}$ is bounded. Equation (5c) then further implies that $\{x_1^k\}$ is also bounded and hence the sequence $\{(x_1^k,x_2^k,x_3^k,\lambda^k)\}$ generated by (5a)-(5d) is bounded. The boundedness of the sequence $\{(x_1^k,x_2^k,x_3^k,\lambda^k)\}$ indicates that there is at least one cluster point of $\{(x_1^k,x_2^k,x_3^k,\lambda^k)\}$. Let $\bar{w}:=(\bar{x}_1,\bar{x}_2,\bar{x}_3,\bar{\lambda})$ be an arbitrary cluster point of $\{(x_1^k,x_2^k,x_3^k,\lambda^k)\}$ and $\{(x_1^{k_j},x_2^{k_j},x_3^{k_j},\lambda^{k_j})\}$ be the subsequence converging to $\bar{w}$. By the inequality (19), we have $\bar{x}_3=\tilde{x}_3^*$. It follows from (5c) and (11a)-(11c) that

$$0\in\partial\theta_1\left(x_1^{k_j}\right)-A_1^T\lambda^{k_j}+\beta A_1^TA_2\left(x_2^{k_j-1}-x_2^{k_j}\right),\tag{21a}$$

$$0\in\partial\theta_2\left(x_2^{k_j}\right)-A_2^T\lambda^{k_j},\tag{21b}$$

$$0 \in \partial\theta_3\big(x_3^{k_j}\big) - A_3^T\lambda^{k_j} + A_3^T\big(\lambda^{k_j-1} - \lambda^{k_j}\big) + P\big(x_3^{k_j} - x_3^{k_j-1}\big), \tag{21c}$$

$$0 = \frac{1}{\beta}\big(\lambda^{k_j} - \lambda^{k_j-1}\big) + A_1x_1^{k_j} + A_2x_2^{k_j} + A_3x_3^{k_j} - b + A_3\big(x_3^{k_j-1} - x_3^{k_j}\big). \tag{21d}$$

Taking the limit in (21a)-(21d) and using (20), we obtain

$$0 \in \partial\theta_1(\bar{x}_1) - A_1^T\bar{\lambda}, \tag{22a}$$

$$0 \in \partial\theta_2(\bar{x}_2) - A_2^T\bar{\lambda}, \tag{22b}$$

$$0 \in \partial\theta_3(\bar{x}_3) - A_3^T\bar{\lambda}, \tag{22c}$$

$$0 = A_1\bar{x}_1 + A_2\bar{x}_2 + A_3\bar{x}_3 - b, \tag{22d}$$

which implies that $(\bar{x}_1, \bar{x}_2, \bar{x}_3, \bar{\lambda})$ is a KKT point in $\mathcal{W}^*$. It follows from (14) and (5c) that the iterative sequence $\{(x_1^k, x_2^k, x_3^k, \lambda^k)\}$ generated by the direct extension of ADMM (5a)-(5d) converges to a KKT point in $\mathcal{W}^*$. The proof is completed. □

Remark 3.1 If the sequence $\{x_3^k\}$ is bounded, then Assumption 3.1 can be substituted by the following.

Assumption 3.2 There exist $(\tilde{x}_1^*, \tilde{x}_2^*, \tilde{x}_3^*, \tilde{\lambda}^*) \in \mathcal{W}^*$ and a real number $\mu_3 > 0$ such that

$$\big\langle y_3 - A_3^T\tilde{\lambda}^*, x_3 - \tilde{x}_3^*\big\rangle \geq \mu_3\big\|x_3 - \tilde{x}_3^*\big\|^2, \quad \text{for all } x_3 \in \mathbf{B}_a\big(\tilde{x}_3^*\big) \text{ and } y_3 \in \partial\theta_3(x_3), \tag{23}$$

where $a = \max_k\{\|x_3^k - \tilde{x}_3^*\|\}$.

3.2 Global linear convergence

Cai *et al.* [29] show that the globally linear convergence of the direct extension of ADMM (4a)-(4d) can be ensured if θ_2 and θ_3 are strongly convex. In this subsection, we will show that the globally linear convergence rate of the direct extension of ADMM (5a)-(5d) can be ensured under weaker conditions. More precisely, we establish the globally linear convergence result for the iterative scheme (5a)-(5d) by showing that there exist $\sigma \in (0,1)$ and $\eta_1 > 0$ such that

$$\big\|v^k - v^{k+1}\big\|_{G_1} \leq \sigma^k \frac{\|v^0 - \hat{v}^*\|_{G_1}}{\sqrt{\eta_1}}. \tag{24}$$

Assumption 3.3 A_2^T is full column rank. For any $(x_1^*, x_2^*, x_3^*, \lambda^*) \in \mathcal{W}^*$, there exists a real number $\mu_2 > 0$ such that

$$\big\langle y_2 - A_2^T\lambda^*, x_2 - x_2^*\big\rangle \geq \mu_2\big\|x_2 - x_2^*\big\|^2, \quad \text{for all } x_2 \in \mathcal{R}^{n_2} \text{ and } y_2 \in \partial\theta_2(x_2). \tag{25}$$

Theorem 3.2 *Let the iterative sequence* $\{(x_1^k, x_2^k, x_3^k, \lambda^k)\}$ *be generated by the direct extension of ADMM* (5a)-(5d) *with* $\beta \in (0, 2\rho\mu_3/(5\|A_3^TA_3\|))$ *and* $\rho \in (0,1)$ *defined in Lemma* 3.2. *Suppose Assumption* 3.1 *and Assumption* 3.3 *hold. If the function* θ_2 *is differentiable, and its gradient* $\nabla\theta_2$ *is Lipschitz continuous with positive constant* L_2, *then there exists* $\delta > 0$ *such that* (24) *holds.*

Proof Since Assumption 3.1 and Assumption 3.3 hold, following the same discussions as of Lemma 3.1 and Lemma 3.3, we have

$$\begin{aligned}\|v^{k+1}-\tilde{v}^*\|_{G_1}^2 &\le \|v^k-\tilde{v}^*\|_{G_1}^2 \\ &\quad - (x_3^k-x_3^{k+1},\lambda^k-\lambda^{k+1})P\begin{pmatrix}x_3^k-x_3^{k+1}\\ \lambda^k-\lambda^{k+1}\end{pmatrix} - \beta(1-\rho)\|A_2(x_2^k-x_2^{k+1})\|^2 \\ &\quad - 2\mu_2\|x_2^{k+1}-\tilde{x}_2^*\|^2 - \left(2\mu_3-\frac{5\beta\|A_3^TA_3\|}{\rho}\right)\|x_3^{k+1}-\tilde{x}_3^*\|^2,\end{aligned}$$

where $\mu_2>0$ and $\mu_3>0$. Thus, there is a real number $\eta_1>0$ such that

$$\begin{aligned}\|v^k-\tilde{v}^*\|_{G_1}^2-\|v^{k+1}-\tilde{v}^*\|_{G_1}^2 &\ge \eta_1\|v^k-v^{k+1}\|_{G_1}^2+2\mu_2\|x_2^{k+1}-\tilde{x}_2^*\|^2 \\ &\quad + \left(2\mu_3-\frac{5\beta\|A_3^TA_3\|}{\rho}\right)\|x_3^{k+1}-\tilde{x}_3^*\|^2. \end{aligned} \tag{26}$$

Since θ_2 is differentiable and $\nabla\theta_2$ is Lipschitz continuous with positive constant L_2, it follows from (7b) and (11b) that

$$\|x_2^{k+1}-\tilde{x}_2^*\| \ge \frac{1}{L_2}\|A_2^T(\lambda^{k+1}-\tilde{\lambda}^*)\|,$$

which together with (26) yields

$$\begin{aligned}\|v^k-\tilde{v}^*\|_{G_1}^2-\|v^{k+1}-\tilde{v}^*\|_{G_1}^2 &\ge \eta_1\|v^k-v^{k+1}\|_{G_1}^2+\left(2\mu_3-\frac{\beta\|A_3^TA_3\|}{\rho}\right)\|x_3^{k+1}-\tilde{x}_3^*\|^2 \\ &\quad + \mu_2\|x_2^{k+1}-\tilde{x}_2^*\|^2+\mu_2\frac{1}{L_2^2}\|A_2^T(\lambda^{k+1}-\tilde{\lambda}^*)\|^2. \end{aligned} \tag{27}$$

Since the matrix A_2^T is full column rank, the inequality (27) implies that there exists $\delta>0$ such that

$$\|v^k-\tilde{v}^*\|_{G_1}^2 \ge (1+\delta)\|v^{k+1}-\tilde{v}^*\|_{G_1}^2. \tag{28}$$

Using (27) again, we obtain

$$\begin{aligned}\eta_1\|v^k-v^{k+1}\|_{G_1}^2 &\le \|v^k-\tilde{v}^*\|_{G_1}^2-\|v^{k+1}-\tilde{v}^*\|_{G_1}^2 \\ &\le \|v^k-\tilde{v}^*\|_{G_1}^2 \le \frac{1}{(1+\delta)^k}\|v^0-\tilde{v}^*\|_{G_1}^2. \end{aligned} \tag{29}$$

Let $\sigma=\frac{1}{\sqrt{1+\delta}}$, we see that (24) holds. □

Notice that if $\|v^k-v^{k+1}\|_G^2=0$, it follows from (5c) and (11a)-(11c) that

$$\begin{aligned}&0\in\partial\theta_1(x_1^{k+1})-A_1^T\lambda^{k+1},\\ &0\in\partial\theta_2(x_2^{k+1})-A_2^T\lambda^{k+1},\\ &0\in\partial\theta_3(x_3^{k+1})-A_3^T\lambda^{k+1},\\ &0=A_1x_1^{k+1}+A_2x_2^{k+1}+A_3x_3^{k+1}-b,\end{aligned}$$

which shows that $(x_1^{k+1}, x_2^{k+1}, x_3^{k+1}, \lambda^{k+1})$ is a solution of (7a)-(7d). Thus, Theorem 3.2 establishes a globally linear convergence rate for the direct extension of ADMM (5a)-(5d), and it inspires an easily implementable stopping criterion for implementing ADMM (5a)-(5d):

$$\max\left\{\frac{\|x_2^k - x_2^{k+1}\|}{1+\|x_2^k\|}, \frac{\|x_3^k - x_3^{k+1}\|}{1+\|x_3^k\|}, \frac{\|\lambda^k - \lambda^{k+1}\|}{1+\|\lambda^k\|}\right\} < \epsilon. \tag{30}$$

4 Example

Chen *et al.* [15] constructed the following example of solving a 3-dimensional linear system:

$$\begin{aligned} &\min 0 \times x_1 + 0 \times x_2 + 0 \times x_3 \\ &\text{s.t.} \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 2 \\ 1 & 2 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \end{aligned} \tag{31}$$

to show that the directly extended alternating direction method of multipliers applied to the above 3-block (treating each variable as one block) optimization problem will diverge.

We replace the item $0 \times x_3$ by $\mu\|A_3x_3\|_1 + \delta(x_3; \mathbf{B}_r(0))$ with $\mu > 0$ and arbitrary $r > 0$ in (31), and obtain

$$\begin{aligned} &\min 0 \times x_1 + 0 \times x_2 + \mu\|A_3x_3\|_1 + \delta\bigl(x_3; \mathbf{B}_r(0)\bigr) \\ &\text{s.t.} \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 2 \\ 1 & 2 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}. \end{aligned} \tag{32}$$

The example (32) can be rewritten as the 3-block optimization problem (1) with the following specifications:

- $\theta_1(x_1) := 0 \times x_1$, $\theta_2(x_2) := 0 \times x_2$, $\theta_3(x_3) := \mu\|A_3x_3\|_1 + \delta(x_3; \mathbf{B}_r(0))$;
- The coefficients A_i ($i = 1, 2, 3$) and the vector b are given by

$$A_1 := \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \qquad A_2 := \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix}, \qquad A_3 := \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix}, \qquad b := \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

Obviously, $\theta_3(x_3)$ is not strongly convex. For problem (32), $(x_1, x_2, x_3, \lambda) = (0, 0, 0, 0)$ is a KKT point. If the iterative sequence $\{(x_1^k, x_2^k, x_3^k, \lambda^k)\}$ is generated by the direct extension of ADMM (5a)-(5d), then we have $x_3^k \in \mathbf{B}_r(0)$, $\forall k$. To justify the convergence of the direct extension of ADMM (5a)-(5d) applied to (32), one just needs to show that Assumption 3.2 holds at $(\tilde{x}_1^*, \tilde{x}_2^*, \tilde{x}_3^*, \tilde{\lambda}^*) = (0, 0, 0, 0)$.

For any $x_3 \in \mathbf{B}_r(\tilde{x}_3^*)$, we have

$$\mu\|A_3x_3\|_1 = 5\mu|x_3| \geq \frac{5\mu}{r}x_3^2.$$

Thus,

$$\theta_3(x_3) \geq \theta_3\bigl(\tilde{x}_3^*\bigr) + \bigl\langle A_3^T\tilde{\lambda}^*, x_3 - \tilde{x}_3^*\bigr\rangle + \alpha^r\bigl\|x_3 - \tilde{x}_3^*\bigr\|^2, \quad \forall x_3 \in \mathbf{B}_r\bigl(\tilde{x}_3^*\bigr), \tag{33}$$

where $\alpha^r = 5\mu/r$. On the other hand, since the function θ_3 is convex, we have

$$\theta_3(\tilde{x}_3^*) \geq \theta_3(x_3) + \langle y_3, \tilde{x}_3^* - x_3 \rangle, \quad \forall x_3 \in \mathbf{B}_r(\tilde{x}_3^*) \text{ and } \forall y_3 \in \partial\theta_3(x_3). \tag{34}$$

Adding up (33) and (34), we obtain

$$\langle y_3 - A_3^T \tilde{\lambda}^*, x_3 - \tilde{x}_3^* \rangle \geq \alpha^r \| x_3 - \tilde{x}_3^* \|^2, \quad \forall x_3 \in \mathbf{B}_r(\tilde{x}_3^*) \text{ and } \forall y_3 \in \partial\theta_3(x_3).$$

Thus Assumption 3.2 holds for (32) at $(\tilde{x}_1^*, \tilde{x}_2^*, \tilde{x}_3^*, \tilde{\lambda}^*) = (0,0,0,0)$ with $\mu_3 = 5\mu/r$, it follows from Theorem 3.1 that the direct extension of ADMM (5a)-(5d) applied to (32) is convergent.

Remark 4.1 If the function θ_3 is strongly convex, then Assumption 3.1 or Assumption 3.2 holds trivially. The example (32) shows that the direct extension of ADMM (5a)-(5d) applied to (32) is convergent, although θ_3 is not strongly convex. This explains why the original scheme of the direct extension of ADMM works well for some applications even though there is not a strong convex function in the objective.

Competing interests
The authors declare that they have no competing interests.

Authors' contributions
All authors contributed equally to the writing of this paper. All authors read and approved the final manuscript.

Author details
[1]College of Computer Science and Technology, Harbin Engineering University, Harbin, 150001, China. [2]College of Science, Harbin Engineering University, Harbin, 150001, China. [3]College of Computer Science and Information Engineering, Harbin Normal University, Harbin, 150025, China.

Acknowledgements
This work was supported by the National Natural Sciences Grants (No. 11371116, No. 41071262, and No. 41101243).

Received: 28 April 2016 Accepted: 12 September 2016 Published online: 17 September 2016

References
1. Chandrasekaran, V, Parrilo, PA, Willsky, AS: Latent variable graphical model selection via convex optimization. Ann. Stat. **40**, 1935-1967 (2012)
2. McLachlan, GJ: Discriminant Analysis and Statistical Pattern Recognition. Wiley-Interscience, New York (2004)
3. Candès, EJ, Li, X, Ma, Y, Wright, J: Robust principal component analysis. J. ACM **58**(3), Article 11 (2011)
4. Tao, M, Yuan, XM: Recovering low-rank and sparse components of matrices from incomplete and noisy observations. SIAM J. Optim. **21**(1), 57-81 (2011)
5. Gabay, D, Mercier, B: A dual algorithm for the solution of nonlinear variational problems via finite-element approximations. Comput. Math. Appl. **2**, 17-40 (1976)
6. Glowinski, R, Marrocco, A: Sur l'approximation, par éléments fins d'ordren, et la résolution, par pénalisation-dualité, d'une classe de problèmes de dirichlet nonlinéares. Rev. Fr. Autom. Inform. Rech. Opér., Anal. Numér. **9**, 41-76 (1975)
7. Fortin, M, Glowinski, R: Augmented Lagrangian Methods. North-Holland, Amsterdam (1983)
8. Glowinski, R: Lectures on Numerical Methods for Nonlinear Variational Problems. Springer, Berlin (1980)
9. Boyd, S, Parikh, N, Chu, E, Peleato, B, Eckstein, J: Distributed optimization and statistical learning via the alternating direction method of multipliers. Found. Trends Mach. Learn. **3**, 1-122 (2010)
10. Chan, TF, Glowinski, R: Finite element approximation and iterative solution of a class of mildly non-linear elliptic equations. Technical report, Stanford University (1978)
11. Eckstein, J: Augmented Lagrangian and alternating direction methods for convex optimization: a tutorial and some illustrative computational results. RUTCOR research report RRR 32-2012, Rutgers University (2012)
12. He, BS, Liu, H, Wang, ZR, Yuan, XM: A strictly Peaceman-Rachford splitting method for convex programming. SIAM J. Optim. **24**, 1011-1040 (2014)
13. Lin, Z, Chen, M, Wu, L, Ma, Y: The augmented Lagrange multiplier method for exact recovery of corrupted low-rank matrices. Math. Program. **9**(22), 15-26 (2010)
14. Peng, YG, Ganesh, A, Wright, J, Xu, WL, Ma, Y: Robust alignment by sparse and low rank decomposition for linearly correlated images. IEEE Trans. Pattern Anal. Mach. Intell. **34**, 2233-2246 (2012)
15. Chen, CH, He, BS, Ye, YY, Yuan, XM: The direct extension of ADMM for multi-block convex minimization problems is not necessarily convergent. Math. Program. **155**(1-2), 57-79 (2016)

16. He, BS, Tao, M, Yuan, XM: Alternating direction method with Gaussian back substitution for separable convex programming. SIAM J. Optim. **22**, 313-340 (2012)
17. He, BS, Tao, M, Yuan, XM: Convergence rate and iteration complexity on the alternating direction method of multipliers with a substitution procedure for separable convex programming. Manuscript
18. Chen, G, Teboulle, M: A proximal-based decomposition method for convex minimization problems. Math. Program. **64**, 81-101 (1994)
19. Deng, W, Yin, WT: On the global and linear convergence of the generalized alternating direction method of multipliers. J. Sci. Comput. **66**(3), 889-916 (2016)
20. Eckstein, J: Some saddle-function splitting methods for convex programming. Optim. Methods Softw. **4**, 75-83 (1994)
21. Fazel, M, Pong, TK, Sun, DF, Tseng, P: Hankel matrix rank minimization with applications to system identification and realization. SIAM J. Matrix Anal. Appl. **34**, 946-977 (2013)
22. He, BS, Liao, LZ, Han, D, Yang, H: A new inexact alternating direction method for monotone variational inequalities. Math. Program. **92**, 103-118 (2002)
23. Hong, MY, Luo, ZQ: On the linear convergence of the alternating direction method of multipliers. Math. Program. (2016). doi:10.1007/s10107-016-1034-2
24. Xu, MH, Wu, T: A class of linearized proximal alternating direction methods. J. Optim. Theory Appl. **155**, 321-337 (2011)
25. Yang, J, Zhang, Y: Alternating direction algorithms for l_1 problems in compressive sensing. SIAM J. Sci. Comput. **33**, 250-278 (2011)
26. Han, DR, Yuan, XM: A note on the alternating direction method of multipliers. J. Optim. Theory Appl. **155**(1), 227-238 (2013)
27. Chen, CH, Shen, Y, You, YF: On the convergence analysis of the alternating direction method of multipliers with three blocks. Abstr. Appl. Anal. **2013**, Article ID 183961 (2013)
28. Lin, TY, Ma, SQ, Zhang, SZ: On the convergence rate of multi-block ADMM (2014). arXiv:1408.4265
29. Cai, XJ, Han, DR, Yuan, XM: The direct extension of ADMM for three-block separable convex minimization models is convergent when one function is strongly convex. Manuscript (2014)
30. Li, M, Sun, DF, Toh, KC: A convergent 3-block semi-proximal ADMM for convex minimization problems with one strongly convex block. Asia-Pac. J. Oper. Res. **32**(4), Article ID 1550024 (2015)
31. Rockafellar, RT: Convex Analysis. Princeton University Press, Princeton (1970)
32. Boyd, S, Vandenberghe, L: Convex Optimization. Cambridge University Press, Cambridge (2004)