Commun. Math. Stat. (2017) 5:59–82
DOI 10.1007/s40304-017-0100-2

Alternating Direction Method for Separable Variables Under Pair-Wise Constraints

Jiaojiao Yang[1] · Yusheng Li[2] · Xinchang Xie[3] · Zhouwang Yang[1]

Received: 23 May 2016 / Revised: 3 January 2017 / Accepted: 11 January 2017 /
Published online: 28 February 2017

Abstract While the convergence of alternating direction method (ADM) for two separable variables has been established for years, the validity of its direct generalization to more than two blocks has been studying now. In this paper, we propose an additional requirement on the constraints, i.e., the pair-wise linear constraints and establish the convergence of ADM for more than two blocks. Then we apply our approach to two kinds of optimization problems. We also show several numerical experiments to verify the rationality of proposed algorithm.

Keywords Alternating direction method · Convergence analysis · Pair-wise constraints · ℓ_1-Analysis problem

Mathematics Subject Classification 90C25 · 65K05 · 80M50 · 68W40

✉ Zhouwang Yang
yangzw@ustc.edu.cn

Jiaojiao Yang
jiao904@mail.ustc.edu.cn

Yusheng Li
lysh@mail.ustc.edu.cn

Xinchang Xie
xinchang.xie@duke.edu

1 University of Science and Technology of China, Hefei, China

2 Luoyang Electronic Equipment Test Center, Luoyang, China

3 Duke University, Durham, NC, USA

1 Introduction

The original alternating direction method (ADM) could date back to 1970s in the research of PDE, with roots in the 1950s, and is equivalent or closely related to many other algorithms, such as dual decomposition, the method of multipliers and Douglas–Rachford splitting. It is now a benchmark for the following convex minimization model:

$$\begin{cases} F(x) + G(y), \\ \text{s.t. } Ax + By - b = 0, \\ x \in \mathcal{X}, y \in \mathcal{Y}. \end{cases} \tag{1.1}$$

Here, F, G are closed convex but not necessarily smooth functions, and $\mathcal{X}, \mathcal{Y}$ are closed convex sets.

Let

$$L(x, y, \lambda) := F(x) + G(y) + \langle \lambda, Ax + By - b \rangle + \frac{\rho}{2} \|Ax + By - b\|_2^2 \tag{1.2}$$

be the augmented Lagrangian function of (1.1). λ is the Lagrange multiplier and $\rho > 0$ is the penalty parameter. Then, the iterative scheme of ADM of (1.1) is

$$\begin{cases} x^{(k+1)} := \arg\min_x L(x, y^{(k)}, \lambda^{(k)}), \\ y^{(k+1)} := \arg\min_y L(x^{(k+1)}, y, \lambda^{(k)}), \\ \lambda^{(k+1)} := \lambda^{(k)} + \rho(Ax^{(k+1)} + By^{(k+1)} - b). \end{cases} \tag{1.3}$$

The convergence of ADM for two separable variables has been established for years. For more details, see [1,10] and [4]. The direct extension of ADM to m variables is strongly desired and practically used by many users, see e.g., [12,13]; however, its convergence result is still an open problem until this paper [2] gives us an example that the extended ADM scheme is not necessarily convergent. They proved the direct generalization of ADM is convergent while the coefficient matrices in the constraints are mutually orthogonal.

Recently, several results of modified ADM to general case (with more than three separable variables) have been established, see [2,6,7]. [14] introduced different ways of ADM and proved its equivalence. He and Yuan et al. had done lots of work in this topic, see [8,9]. Deng and Yin et al. had proved that the extending ADMM algorithm converges globally at a rate see [3] and [5].

In this paper, we focus on the condition of the linear constraints which can guarantee the convergence of the direct generalization of ADM. We propose an additional requirement of the constraints, i.e., the pair-wise linear constraints, and establish the convergence of ADM for more than three variables.

The remaining is organized as follows. In Sect. 2, we establish the convergence theorem of ADM under pair-wise linear constraint of three separable variables case and the m separable variables case successively. In Sect. 3, we propose the application of pair-wise constraint ADM in two kinds of optimization problems and some numerical experiments are given to show the rationality and utility. The proofs of convergence for the theorems are given in the "Appendix part".

2 Alternating Direction Method and Its Convergence Under Pair-Wise Constraints

Although the convergence of general ADM with common constraints is an open problem, we can deduce the convergence theorem under a certain kind of constraints, which we call "pair-wise constraints", i.e., each linear constraint contains two separable variables.

2.1 Notations

We denote $\langle a, b\rangle$ as the inner product of two vectors a and b, i.e., $\langle a, b\rangle = a^T b$, $\|a\|_2$ the ℓ_2-norm of vector a, i.e., $\|a\|_2^2 = a^T a$. A, B and C denote matrices or linear mappings.

We say the variables have pair-wise linear constraints means:

$$\begin{cases} A_1^{(1,2)} x_1 + A_2^{(2,1)} x_2 - b_{(1,2)} = 0, \\ A_1^{(1,3)} x_1 + A_3^{(3,1)} x_3 - b_{(1,3)} = 0, \\ \cdot \\ \cdot \\ \cdot \\ A_{m-1}^{(m-1,m)} x_{m-1} + A_m^{(m,m-1)} x_m - b_{(m-1,m)} = 0, \end{cases} \tag{2.1}$$

where $A_i^{(i,j)}$ means the matrix is related to the variable x_i and occurs in the constraint of x_i and x_j, i.e., the constraints of x_i and x_j is: $A_i^{(i,j)} x_i + A_j^{(j,i)} x_j - b_{(i,j)} = 0$ (suppose $i < j$). More specifically, we can write the constraints in a compact form:

$$\begin{pmatrix} A_1^{(1,2)} & A_2^{(2,1)} & 0 & 0 & \cdots & 0 & 0 \\ A_1^{(1,3)} & 0 & A_3^{(3,1)} & 0 & \cdots & 0 & 0 \\ A_1^{(1,4)} & 0 & 0 & A_4^{(4,1)} & \cdots & 0 & 0 \\ \cdot & & & & & & \\ \cdot & & & & & & \\ \cdot & & & & & & \\ 0 & 0 & 0 & 0 & \cdots & A_{m-1}^{(m-1,m)} & A_m^{(m,m-1)} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \cdot \\ \cdot \\ \cdot \\ x_m \end{pmatrix} = \begin{pmatrix} b_{(1,2)} \\ b_{(1,3)} \\ b_{(1,4)} \\ \cdot \\ \cdot \\ \cdot \\ b_{(m-1,m)} \end{pmatrix}. \tag{2.2}$$

2.2 Three Separable Variables Case

In this section, we consider the ADM with three separable variables of pair-wise linear constraints optimization problem,

$$\begin{cases} \min F(x) + G(y) + H(z), \\ \text{s.t. } A_1 x + B_1 y - b_1 = 0, \\ A_2 x + C_2 z - b_2 = 0, \\ B_3 y + C_3 z - b_3 = 0, \end{cases} \tag{2.3}$$

where F, G, H are convex functions and $A_i,\ B_i,\ C_i\ (i = 1, 2, 3)$ are matrices. Then the augmented Lagrangian function of (2.3) is:

$$\begin{aligned} L(x, y, z, \lambda_1, \lambda_2, \lambda_3, \rho) = {} & F(x) + G(y) + H(z) + \lambda_1^T (A_1 x + B_1 y - b_1) \\ & + \lambda_2^T (A_2 x + C_2 z - b_2) + \lambda_3^T (B_3 y + C_3 z - b_3) \\ & + \frac{\rho}{2} (\|A_1 x + B_1 y - b_1\|_2^2 + \|A_2 x + C_2 z - b_2\|_2^2 \\ & + \|B_3 y + C_3 z - b_3\|_2^2). \end{aligned} \tag{2.4}$$

Then the general ADM will lead us to three subproblems, in which we fix two separable variables and find the other one that minimize the augmented Lagrangian function:

$$\begin{aligned} & \min_x \left\{ F(x) + \lambda_1^T (A_1 x + B_1 y^{(k)} - b_1) + \lambda_2^T (A_2 x + C_2 z^{(k)} - b_2) \right. \\ & \quad \left. + \frac{\rho}{2} \left(\|A_1 x + B_1 y^{(k)} - b_1\|_2^2 + \|A_2 x + C_2 z^{(k)} - b_2\|_2^2 \right) \right\}, \end{aligned} \tag{2.5}$$

$$\begin{aligned} & \min_y \left\{ G(y) + \lambda_1^T (A_1 x^{(k+1)} + B_1 y - b_1) + \lambda_3^T (B_3 y + C_3 z^{(k)} - b_3) \right. \\ & \quad \left. + \frac{\rho}{2} \left(\|A_1 x^{(k+1)} + B_1 y - b_1\|_2^2 + + \|B_3 y + C_3 z^{(k)} - b_3\|_2^2 \right) \right\}, \end{aligned} \tag{2.6}$$

$$\begin{aligned} & \min_z \left\{ H(z) + \lambda_2^T (A_2 x^{(k+1)} + C_2 z - b_2) + \lambda_3^T (B_3 y^{(k+1)} + C_3 z - b_3) \right. \\ & \quad \left. + \frac{\rho}{2} \left(\|A_2 x^{(k+1)} + C_2 z - b_2\|_2^2 + \|B_3 y^{(k+1)} + C_3 z - b_3\|_2^2 \right) \right\}. \end{aligned} \tag{2.7}$$

After solving three subproblems successively, we update the Lagrangian multipliers as follows:

$$\begin{cases} \lambda_1^{(k+1)} = \lambda_1^{(k)} + \rho (A_1 x^{(k+1)} + B_1 y^{(k+1)} - b_1), \\ \lambda_2^{(k+1)} = \lambda_2^{(k)} + \rho (A_2 x^{(k+1)} + C_2 z^{(k+1)} - b_2), \\ \lambda_3^{(k+1)} = \lambda_3^{(k)} + \rho (B_3 y^{(k+1)} + C_3 z^{(k+1)} - b_3). \end{cases} \tag{2.8}$$

The convergence result of Algorithm 1 is:

Algorithm 1 ADM for three variables case

Require: starting point $(x^{(0)}, y^{(0)}, z^{(0)}, \lambda_1^{(0)}, \lambda_2^{(0)}, \lambda_3^{(0)})$ and parameters $\rho > 0$;
for k = 0 to M **do**
$x^{(k+1)} = \arg\min_x L(x, y^{(k)}, z^{(k)}, \lambda_1^{(k)}, \lambda_2^{(k)}, \lambda_3^{(k)}, \rho)$
$y^{(k+1)} = \arg\min_y L(x^{(k+1)}, y, z^{(k)}, \lambda_1^{(k)}, \lambda_2^{(k)}, \lambda_3^{(k)}, \rho)$
$z^{(k+1)} = \arg\min_z L(x^{(k+1)}, y^{(k+1)}, z, \lambda_1^{(k)}, \lambda_2^{(k)}, \lambda_3^{(k)}, \rho)$

Update Lagrangian multipliers
$\lambda_1^{(k+1)} = \lambda_1^{(k)} + \rho(A_1 x^{(k+1)} + B_1 y^{(k+1)} - b_1)$
$\lambda_2^{(k+1)} = \lambda_2^{(k)} + \rho(A_2 x^{(k+1)} + C_2 z^{(k+1)} - b_2)$
$\lambda_3^{(k+1)} = \lambda_3^{(k)} + \rho(B_3 y^{(k+1)} + C_3 z^{(k+1)} - b_3)$
end for

Theorem 2.1 *Given $\rho > 0$, and matrices $A_i,\ B_i,\ C_i\ (i = 1, 2, 3)$ are full column-rank, then the sequences $\{(x^{(k)}, y^{(k)}, z^{(k)}, \lambda_1^{(k)}, \lambda_2^{(k)}, \lambda_3^{(k)})\}$ generated by Algorithm 1 from any initial point $(x^{(0)}, y^{(0)}, z^{(0)}, \lambda_1^{(0)}, \lambda_2^{(0)}, \lambda_3^{(0)})$ converge to the KKT point. The corresponding objective function value also converges to $F(\hat{x})+G(\hat{y})+H(\hat{z})$, where $(\hat{x}, \hat{y}, \hat{z})$ is a saddle point of the augmented Lagrangian function.*

The proof of Theorem 2.1 is in the "Appendix".

2.3 General *m*-Variable Case

Suppose we have an optimization problem with the form:

$$\begin{cases} \min \sum_{i=1}^{m} F_i(x_i), \\ \text{s.t.} \sum_{i=1}^{m} A_i x_i - b = 0, \end{cases} \tag{2.9}$$

where F_i are convex functions and the linear constraints are pair-wise constraints of separable variables x_i. A_i is the ith column of the matrix in (2.2).

We introduce Lagrangian multiplier λ and penalty factor ρ to compose the augmented Lagrangian function:

$$L(x_1, \ldots, x_m, \lambda, \rho) = \sum_{i=1}^{m} F_i(x_i) + \langle \lambda, \sum_{i=1}^{m} A_i x_i - b \rangle + \frac{\rho}{2} \left\| \sum_{i=1}^{m} A_i x_i - b \right\|_2^2. \tag{2.10}$$

The convergence of Algorithm 2 for the general m-variable case is as follows:

Algorithm 2 ADM for m variables case

Require: starting point $(x_1^{(0)}, x_2^{(0)}, \ldots, x_m^{(0)}, \lambda^{(0)})$ and parameters $\rho > 0$;
for $k = 0$ to M **do**
$x_1^{(k+1)} = \arg\min_{x_1} L(x_1, x_2^{(k)}, x_3^{(k)}, \ldots, x_m^{(k)}, \lambda^{(k)}, \rho)$
$x_2^{(k+1)} = \arg\min_{x_2} L(x_1^{(k+1)}, x_2, x_3^{(k)}, \ldots, x_m^{(k)}, \lambda^{(k)}, \rho)$
...
...
...
$x_m^{(k+1)} = \arg\min_{x_m} L(x_1^{(k+1)}, x_2^{(k+1)}, x_3^{(k+1)}, \ldots, x_{(m-1)}^{(k+1)}, x_m, \lambda^{(k)}, \rho)$
Update the Lagrangian multipliers
$\lambda^{(k+1)} = \lambda^{(k)} + \rho(\sum_{i=1}^m A_i x_i^{(k+1)} - b)$
end for

Theorem 2.2 *Given $\rho > 0$, and each matrix A_i $(i = 1, 2, \ldots, m)$ is full column-rank, then the sequences $\{(x_1^{(k)} x_2^{(k)}, x_3^{(k)}, \ldots, x_m^{(k)}, \lambda^{(k)})\}$ generated by Algorithm 2 from any initial point $(x_1^{(0)}, x_2^{(0)}, x_3^{(0)}, \ldots, x_m^{(0)}, \lambda^{(0)})$ converge to the KKT point. The corresponding objective function value also converges to $\sum_{i=1}^m F_i(\bar{x}_i)(i = 1, 2, \ldots, m)$ where $(\bar{x}_1, \bar{x}_2, \bar{x}_3, \ldots, \bar{x}_m)$ is a saddle point of the augmented Lagrangian function.*

The proof of Theorem 2.2 is also in the "Appendix".

3 Application in Different Problems and Numerical Experiments

3.1 ℓ_1-Analysis Model

In compressed sensing (CS) research community, the first and widely considered model is the Basic Pursuit problem:

$$\begin{cases} \min \|x\|_1, \\ \text{s.t. } Ax = b. \end{cases} \tag{3.1}$$

It is well known that problem (3.1) can be solved by a linear programming. However, the more general case in practice is that the signal is not itself sparse but has few nonzero elements under some transformation. Besides, the sensing procedure may be influenced by some noise. So, the variational problem (3.1) has the following formulation:

$$\begin{cases} \min \|Tx\|_1, \\ \text{s.t. } \|Ax - b\|_2 \leq \delta, \end{cases} \tag{3.2}$$

where T is some nonorthogonal linear transform which can transform x to a more spare representation. If T is an orthogonal linear transform. Introduce variable $y = Tx$, we have $x = T^T y$ and the problem becomes

$$\begin{cases} \min \|y\|_1, \\ \text{s.t. } \|AT^T y - b\|_2 \le \delta. \end{cases} \tag{3.3}$$

It is the normal ℓ_1 problem.

How can we solve problem (3.2) efficiently? Modern convex optimization can solve it by interior point method with high accuracy and acceptable iterations; however, it may fail when dealing with large-scale problems. Here we propose an augmented Lagrangian method based on alternating direction method to solve this model.

Firstly, we introduce two auxiliary variables y and r, and then the model can be reformulated as follows, in which the variables have been separated.

$$\begin{cases} \min \|y\|_1, \\ \text{s.t. } Tx = y, \\ Ax + r - b = 0, \\ \|r\|_2 \le \delta. \end{cases} \tag{3.4}$$

It is a special case of problem (2.9), so the convergence result is obvious.

Secondly, we introduce the augmented Lagrangian function:

$$\begin{aligned} L(x, y, r, \lambda_1, \lambda_2, \rho) = \|y\|_1 + \lambda_1^T(Tx - y) + \lambda_2^T(Ax + r - b) \\ + \frac{\rho}{2}\left(\|Tx - y\|_2^2 + \|Ax + r - b\|_2^2\right). \end{aligned} \tag{3.5}$$

At last, we solve three subproblems successively and update the multipliers at the end of each iteration.

Subproblem (1):

$$y^{(k+1)} = \arg\min_y \left\{\|y\|_1 + (\lambda_1^{(k)})^T (Tx^{(k)} - y) + \frac{\rho}{2}\|Tx^{(k)} - y\|_2^2\right\}. \tag{3.6}$$

It is easy to see that:

$$y^{(k+1)} = \arg\min_y \left\{\|y\|_1 + \frac{\rho}{2}\|Tx^{(k)} - y + \frac{\lambda_1^{(k)}}{2\rho}\|_2^2\right\} \tag{3.7}$$

which has the closed-form solution:

$$y^{(k+1)} = \max\left\{|Tx^{(k)} + \frac{\lambda_1^{(k)}}{\rho}| - \frac{1}{\rho}, 0\right\} .* sign\left(Tx^{(k)} + \frac{\lambda_1^{(k)}}{\rho}\right), \tag{3.8}$$

where $.*$ is the Hadamard product of two vectors, i.e., element-wise product.

Subproblem (2):

$$x^{(k+1)} = \arg\min_x (\lambda_1^{(k)})^T (Tx - y^{(k+1)}) + (\lambda_2^{(k)})^T (Ax + r^{(k)} - b)$$

$$+\frac{\rho}{2}\left(\|Tx-y^{(k+1)}\|_2^2+\|Ax+r^{(k)}-b\|_2^2\right) \tag{3.9}$$

which leads to the linear system of $x^{(k+1)}$:

$$\begin{aligned}(A^TA+T^TT)x^{(k+1)} &= -\frac{1}{\rho}\left(T^T\lambda_1^{(k)}+A^T\lambda_2^{(k)}\right)+T^Ty^{(k+1)}\\ &\quad +A^T(b-r^{(k)}).\end{aligned} \tag{3.10}$$

Noticed that the coefficient matrix (A^TA+T^TT) keeps unchanged in the iterations, we can make a pre-decomposition of it(e.g., Cholesky factorization).

Subproblem (3):

$$\begin{aligned}r^{(k+1)} &= \arg\min_z\left\{(\lambda_2^{(k)})^T(Ax^{(k+1)}+r-b)+\frac{\rho}{2}\|Ax^{(k+1)} s\right.\\ &\quad \left.+r-b\|_2^2 \mid \|r\|_2^2\le\delta\right\}.\end{aligned} \tag{3.11}$$

The solution of subproblem (3) is:

$$r^{(k+1)}=P_{B_\delta}\left\{b-\frac{\lambda_2^{(k)}}{\rho}-Ax^{(k+1)}\right\}, \tag{3.12}$$

where $B_\delta=\{r \mid \|r\|_2\le\delta\}$ and P_{B_δ} is the projection operator to the ball B_δ.

We update the multipliers after solving the three subproblems as:

$$\lambda_1^{(k+1)}=\lambda_1^{(k)}+\rho(Tx^{(k+1)}-y^{(k+1)}), \tag{3.13}$$
$$\lambda_2^{(k+1)}=\lambda_2^{(k)}+\rho(Ax^{(k+1)}+r^{(k+1)}-b). \tag{3.14}$$

Next, we will show some numerical experiments of ADM method on the ℓ_1-analysis signal recovery problems. We take the 1D signal to make the process clear. More specifically, we let the truth x be the pair-wise linear function in which has the following Fig. 1:

We take L be the discrete Laplacian matrix in 1D and Lx is obviously sparse. For the sensing part, we take an $m\times n$ Gauss random matrix ϕ where m is one-tenth of n. Then problem (3.2) has the following form:

$$\begin{cases}\min\|Lx\|_1,\\ \text{s.t. } \|\phi x-b\|_2\le\delta.\end{cases} \tag{3.15}$$

In the test, we add noise of different intensity and try to recover x, that is to say, $b=\phi x+n$ where n denotes the noise vector and $\delta=\|n\|_2$. We can get the recovery of x:

It is easy to see that the algorithm works well in medium noise and noiseless case, while the heavy noise will affects the quality of the recovery signal (Fig. 2).

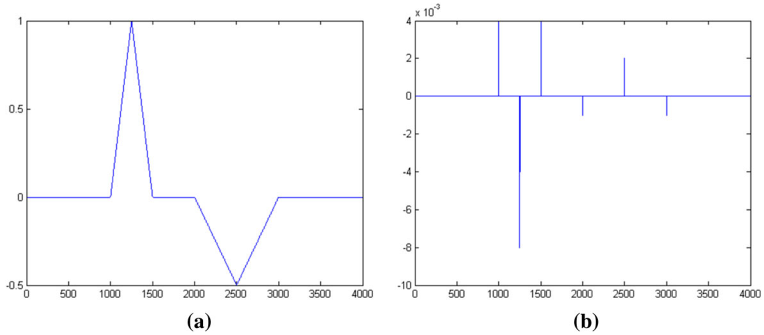

Fig. 1 True signal x (**a**), and true Lx (**b**)

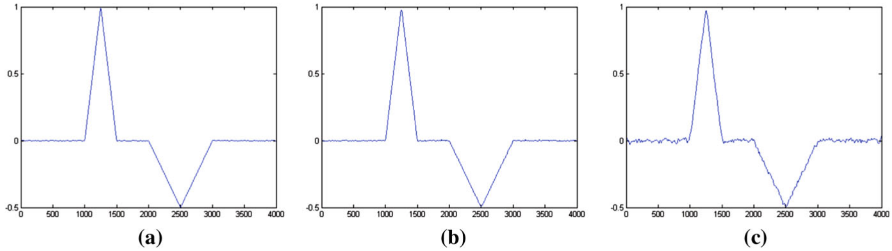

Fig. 2 Noiseless recovery (**a**), medium noise recovery with $\delta = 0.1$ (**b**), and heavy noise recovery with $\delta = 1$, (**c**)

3.2 Earthquake Source Localization Problem

Another useful model in many applications is:

$$\min \ \|Wx\|_p + \mu\|Ax - b\|_q. \tag{3.16}$$

Here W is some weight matrix and p, q are some kinds of convex norms. For an example, if p is $\|\cdot\|_1$ and q is $\|\cdot\|_2^2$, it is the lasso problem. If W is the unit matrix, p is the group $\|\cdot\|_2$ and q is the $\|\cdot\|_2^2$, it becomes the group lasso problem [16], i.e.,

$$\min \sum_{g=1}^{G} \|x_{\mathcal{I}_g}\|_2 + \mu\|Ax - b\|_2^2, \tag{3.17}$$

where $\mathcal{I}_g$ is the index set belonging to the gth group of variables, $g = 1, \ldots, G$.

To solve this kind of model efficiently, firstly we introduce two auxiliary variables y and z and then the model can be reformulated as follows, in which the variables have been separated.

$$\begin{cases} \min \|y\|_p + \mu \|z\|_q, \\ \text{s.t. } Wx - y = 0, \\ Ax - z - b = 0. \end{cases} \tag{3.18}$$

It is also a special case of problem (2.9).

We also introduce the augmented Lagrangian function:

$$\begin{aligned} L(x, y, z, \lambda_1, \lambda_2, \rho) = \|y\|_p + \mu \|z\|_q + \lambda_1^T (Wx - y) + \lambda_2^T (Ax - z - b) \\ + \frac{\rho}{2} \left(\|Wx - y\|_2^2 + \|Ax - z - b\|_2^2 \right). \end{aligned} \tag{3.19}$$

In the end, we solve three subproblems and update the multipliers at the end of each iteration.

Subproblem (1):

$$\begin{aligned} x^{(k+1)} = \arg\min_x (\lambda_1^{(k)})^T (Wx - y^{(k)}) + (\lambda_2^{(k)})^T (Ax - z^{(k)} - b) \\ + \frac{\rho}{2} \left(\|Wx - y^{(k)}\|_2^2 + \|Ax - z^{(k)} - b\|_2^2 \right). \end{aligned} \tag{3.20}$$

It is a quadratic problem of x and can be solved by many algorithms, such as CG (conjugate gradient method) and Newton Method.

Subproblem (2):

$$y^{(k+1)} = \arg\min_y \left\{ \|y\|_p + (\lambda_1^{(k)})^T (Wx^{(k+1)} - y) + \frac{\rho}{2} \|Wx^{(k+1)} - y\|_2^2 \right\}. \tag{3.21}$$

It has the closed-form solution for some special norms, such as $\|\cdot\|_1$, $\|\cdot\|_2$, group norm, and other convex $\|\cdot\|_p$ norm.

Subproblem (3):

$$\begin{aligned} z^{(k+1)} = \arg\min_z \Big\{ \|z\|_q + (\lambda_2^{(k)})^T (Ax^{(k+1)} - z - b) \\ + \frac{\rho}{2} \|Ax^{(k+1)} - z - b\|_2^2 \Big\}. \end{aligned} \tag{3.22}$$

It also has the closed-form solution for the mentioned norms.

In the end, we update the multipliers after solving the three subproblems as:

$$\lambda_1^{(k+1)} = \lambda_1^{(k)} + \rho (Wx^{(k+1)} - y^{(k+1)}), \tag{3.23}$$

$$\lambda_2^{(k+1)} = \lambda_2^{(k)} + \rho (Ax^{(k+1)} - z^{(k+1)} - b). \tag{3.24}$$

Below, we apply model (3.16) to solve earthquake source localization problem [11,15]:

$$\min \ \mu \sum_{j=1}^{n} \|x^j\|_2 + \|vec(AX - B)\|_1. \tag{3.25}$$

Here $A \in \mathbb{C}^{m\times n}$, $B \in \mathbb{C}^{m\times T}$, $B = [b_1, \ldots, b_T]$, $X \in \mathbb{C}^{n\times T}$, $X = [x_1, \ldots, x_T]$, and $x^j = [x_1^j, \ldots, x_T^j]^T$, where x_i^j is the jth number of x_i, i.e., the ith row and jth column of X. $vec(AX - B)$ is $[Ax_1 - b_1, \ldots, Ax_T - b_T]^T$. A is the so-called array manifold matrix, whose i, jth element contains the delay and gain information from the jth source to the ith sensor. B is the observed waveform.

Next, we will show two numerical experiments. In the test, we let $m = 200$, $n = 400$, $T = 5$, $\mu = 0.8$, and we add noise of different intensity and try to recover X, that is to say, $B = AX + N$ where N denotes the noise matrix and $\delta = \|N\|_2$.

Figure 3 is the signal recovery with different noise and Fig. 4 is the corresponding objective value. In this example, our algorithm works well in not so heavy noise, too. In fact, ADM algorithm converges slowly when the iteration points are near the optimum point, especially when the noise is heavy in our problem. We can see in Fig. 4. And this is also a difficulty to be overcome in our following research.

Acknowledgements We would like to thank the anonymous reviewers for their comments and suggestions. The work is supported by the NSF of China (No. 11626253), and the Fundamental Research Funds for the Central Universities.

4 Appendix

4.1 Proof of Theorem 2.1

Proof of Theorem 2.1 The three pair-wise constraints can be rewritten as compact form: $Ax + By + Cz - b = 0$, where $A = \begin{pmatrix} A_1 \\ A_2 \\ 0 \end{pmatrix}$, $B = \begin{pmatrix} B_1 \\ 0 \\ B_3 \end{pmatrix}$, and $C = \begin{pmatrix} 0 \\ C_2 \\ C_3 \end{pmatrix}$.

If we let $\lambda = \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{pmatrix}$, then the augmented Lagrangian function of (2.3) is:

$$\begin{aligned} L(x, y, z, \lambda, \rho) = F(x) + G(y) + H(z) + \lambda^T (Ax + By + Cz - b) \\ + \frac{\rho}{2}\|Ax + By + Cz - b\|_2^2. \end{aligned} \tag{4.1}$$

Since the augmented Lagrangian function always has a saddle point in the finite dimension case, there exists $(\hat{x}, \hat{y}, \hat{z}, \hat{\lambda})$ such that: $L(\hat{x}, \hat{y}, \hat{z}, \lambda, \rho) \le L(\hat{x}, \hat{y}, \hat{z}, \hat{\lambda}, \rho) \le L(x, y, z, \hat{\lambda}, \rho)$ for any (x, y, z, λ).

The left inequality shows that: $A\hat{x} + B\hat{y} + C\hat{z} - b = 0$, which combines with the right inequality shows that:

$$\begin{aligned} F(\hat{x}) + G(\hat{y}) + H(\hat{z}) \le F(x) + G(y) + H(z) + \langle \hat{\lambda}, Ax + By + Cz - b\rangle \\ + \frac{\rho}{2}\|Ax + By + Cz - b\|_2^2. \end{aligned} \tag{4.2}$$

Let $x = \hat{x} + t(w - \hat{x})$, $0 < t < 1$, we have $F(x) - F(\hat{x}) \ge t[F(w) - F(\hat{x})]$ because of the convexity of F. Since the arbitrariness of (x, y, z), let $(x, y, z) = (x, \hat{y}, \hat{z})$ then we get:

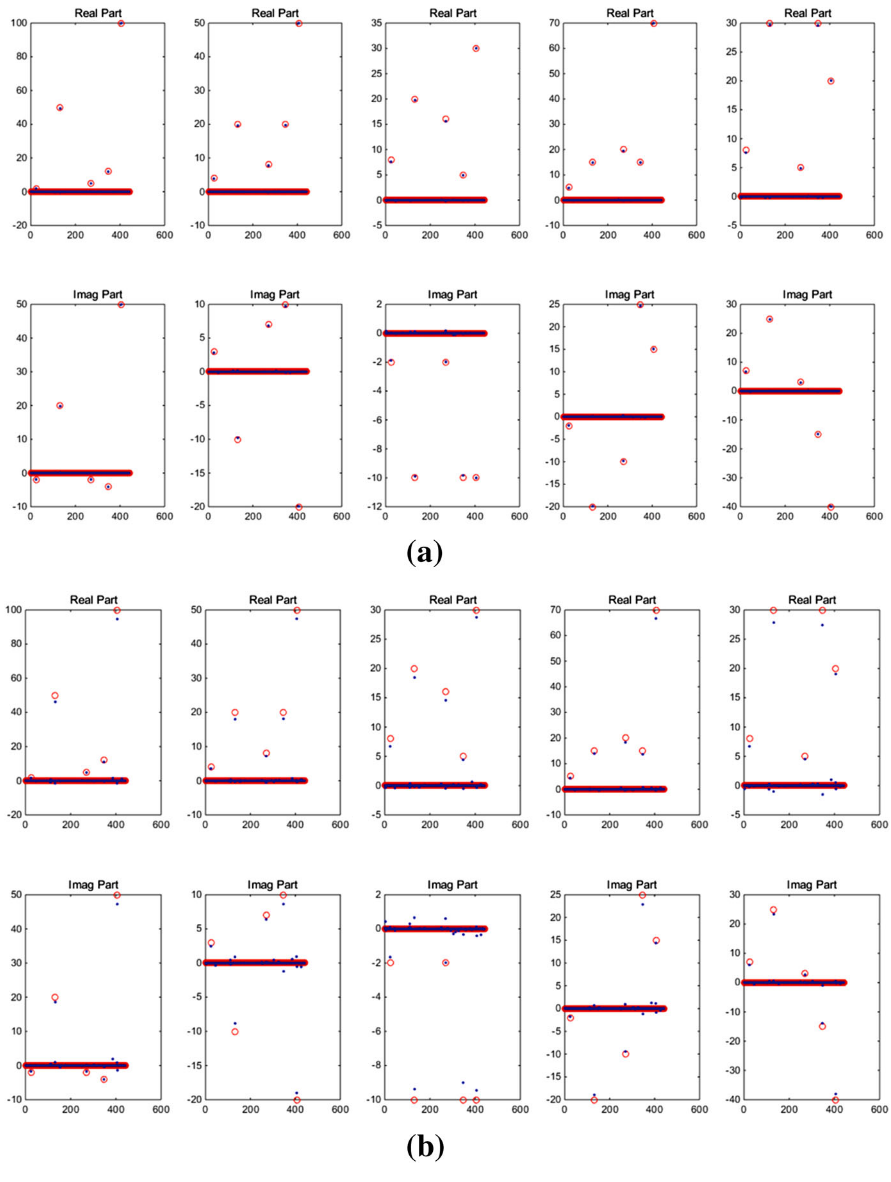

Fig. 3 Signal recovery with different noise. Here, the *red hollow point* is the true signal and the *blue solid point* is the recovery signal. Colour figure online

$$t[F(w) - F(\hat{x})] + t\langle\hat{\lambda}, A(w - \hat{x})\rangle + \frac{t^2\rho}{2}\|A(w - \hat{x})\|_2^2 \geq 0. \tag{4.3}$$

Divide by t on both side of the inequality and let $t \to 0$ we will get:

$$F(w) - F(\hat{x}) + \langle\hat{\lambda}, A(w - \hat{x})\rangle \geq 0. \tag{4.4}$$

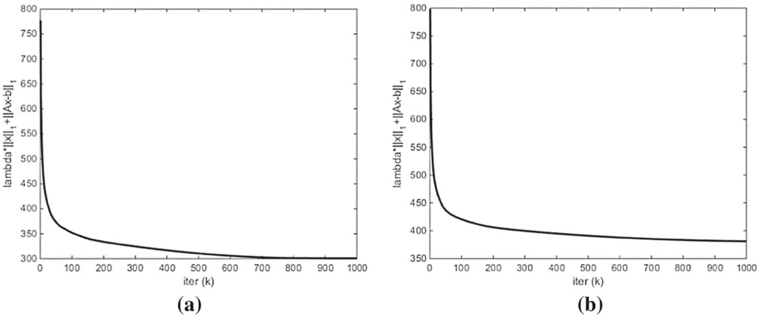

Fig. 4 Signal recovery with different noise. Here is the objective value of these two numerical experiments

Let $w = x^{(n)}$,

$$F(x^{(n)}) - F(\hat{x}) + \langle \hat{\lambda}, A(x^{(n)} - \hat{x}) \rangle, \geq 0, \tag{4.5}$$

Similarly, we can get:

$$G(y^{(n)}) - G(\hat{y}) + \langle \hat{\lambda}, B(y^{(n)} - \hat{y}) \rangle \geq 0. \tag{4.6}$$

$$H(z^{(n)}) - H(\hat{z}) + \langle \hat{\lambda}, C(z^{(n)} - \hat{z}) \rangle \geq 0, \tag{4.7}$$

Add the three inequalities above and we will get:

$$\begin{aligned} &F(x^{(n)}) + G(y^{(n)}) + H(z^{(n)}) - F(\hat{x}) - G(\hat{y}) - H(\hat{z}) \\ &\quad + \langle \hat{\lambda}, A(x^{(n)} - \hat{x}) + B(y^{(n)} - \hat{y}) + C(z^{(n)} - \hat{z}) \rangle \geq 0. \end{aligned} \tag{4.8}$$

Noted the update of x in Algorithm 1, we can get:

$$\begin{aligned} &F(x) + \frac{\rho}{2} \|Ax + By^{(n-1)} + Cz^{(n-1)-b}\|_2^2 \\ &\quad + \langle \lambda^{(n)}, Ax + By^{(n-1)} + Cz^{(n-1)} - b \rangle \\ &\geq F(x^{(n)}) + \frac{\rho}{2} \|Ax^{(n)} + By^{(n-1)} + Cz^{(n-1)-b}\|_2^2 \\ &\quad + \langle \lambda^{(n)}, Ax^{(n)} + By^{(n-1)} + Cz^{(n-1)} - b \rangle \end{aligned} \tag{4.9}$$

for any x. Simple derivations of the inequality show that:

$$\begin{aligned} &F(x) - F(x^{(n)}) + \langle A(x - x^{(n)}), \lambda^{(n)} + \rho(Ax^{(n)} + By^{(n-1)} + Cz^{(n-1)} - b) \rangle \\ &\quad + \frac{\rho}{2} \|A(x - x^{(n)})\|_2^2 \geq 0. \end{aligned} \tag{4.10}$$

Here we use the similar skill, i.e., let $x = x^{(n)} + t(w - x^{(n)}), 0 < t < 1$ and let $t \to 0$:

$$\begin{aligned} &F(w) - F(x^{(n)}) + \langle A(w - x^{(n)}), \lambda^{(n)} \\ &\quad + \rho(Ax^{(n)} + By^{(n-1)} + Cz^{(n-1)} - b) \rangle \geq 0. \end{aligned} \tag{4.11}$$

Letting $w = \hat{x}$, then we can obtain that:

$$F(\hat{x}) - F(x^{(n)}) + \langle A(\hat{x} - x^{(n)}), \lambda^{(n)} + \rho(Ax^{(n)} + By^{(n-1)} + Cz^{(n-1)} - b)\rangle \geq 0. \tag{4.12}$$

Similarly, following the order of subproblems in Algorithm 1 we can get:

$$G(\hat{y}) - G(y^{(n)}) + \langle B(\hat{y} - y^{(n)}), \lambda^{(n)} + \rho(Ax^{(n)} + By^{(n)} + Cz^{(n-1)} - b)\rangle \geq 0, \tag{4.13}$$

$$H(\hat{z}) - H(z^{(n)}) + \langle C(\hat{z} - z^{(n)}), \lambda^{(n)} + \rho(Ax^{(n)} + By^{(n)} + Cz^{(n)} - b)\rangle \geq 0. \tag{4.14}$$

Add the three inequalities and use the notations $\bar{x}^{(n)} = x^{(n)} - \hat{x}$, $\bar{y}^{(n)} = y^{(n)} - \hat{y}$, $\bar{z}^{(n)} = z^{(n)} - \hat{z}$ we can get:

$$\begin{aligned}
&F(\hat{x}) + G(\hat{y}) + H(\hat{z}) - F(x^{(n)}) - G(y^{(n)}) - H(z^{(n)}) \\
&\quad - \langle \lambda^{(n)}, A\bar{x}^{(n)} + B\bar{y}^{(n)} + C\bar{z}^{(n)}\rangle - \rho\|A\bar{x}^{(n)} + B\bar{y}^{(n)} + C\bar{z}^{(n)}\|_2^2 \\
&\quad + \rho\langle B\bar{y}^{(n)} - B\bar{y}^{(n-1)} + C\bar{z}^{(n)} - C\bar{z}^{(n-1)}, A\bar{x}^{(n)}\rangle \\
&\quad + \rho\langle C\bar{z}^{(n)} - C\bar{z}^{(n-1)}, B\bar{y}^{(n)}\rangle \geq 0.
\end{aligned} \tag{4.15}$$

Add (4.15) to (4.8) and denote $\bar{\lambda}^{(n)} = \lambda^{(n)} - \lambda$:

$$\begin{aligned}
&-\langle \bar{\lambda}^{(n)}, A\bar{x}^{(n)} + B\bar{y}^{(n)} + C\bar{z}^{(n)}\rangle - \rho\|A\bar{x}^{(n)} + B\bar{y}^{(n)} + C\bar{z}^{(n)}\|_2^2 \\
&+ \rho\langle B\bar{y}^{(n)} - B\bar{y}^{(n-1)} + C\bar{z}^{(n)} - C\bar{z}^{(n-1)}, A\bar{x}^{(n)}\rangle + \rho\langle C\bar{z}^{(n)} - C\bar{z}^{(n-1)}, B\bar{y}^{(n)}\rangle \geq 0.
\end{aligned} \tag{4.16}$$

According to the update of λ in Algorithm 1, we know that:

$$\begin{aligned}
\bar{\lambda}^{(n+1)} - \bar{\lambda}^{(n)} &= \lambda^{(n+1)} - \lambda^{(n)} \\
&= \rho(Ax^{(n)} + By^{(n)} + C^{(n)} - b) \\
&= \rho(A\bar{x}^{(n)} + B\bar{y}^{(n)} + C\bar{z}^{(n)}).
\end{aligned} \tag{4.17}$$

Therefore,

$$\begin{aligned}
|\bar{\lambda}^{(n)}|^2 - |\bar{\lambda}^{(n+1)}|^2 &= \langle \bar{\lambda}^{(n)} - \bar{\lambda}^{(n+1)}, \bar{\lambda}^{(n)} + \bar{\lambda}^{(n+1)}\rangle \\
&= -\rho\langle A\bar{x}^{(n)} + B\bar{y}^{(n)} + C\bar{z}^{(n)}, 2\bar{\lambda}^{(n)} + \rho(A\bar{x}^{(n)} + B\bar{y}^{(n)} + C\bar{z}^{(n)})\rangle \\
&= -\rho^2\|A\bar{x}^{(n)} + B\bar{y}^{(n)} + C\bar{z}^{(n)}\|_2^2 \\
&\quad - 2\rho\langle \bar{\lambda}^{(n)}, A\bar{x}^{(n)} + B\bar{y}^{(n)} + C\bar{z}^{(n)}\rangle \\
&\geq \rho^2\|A\bar{x}^{(n)} + B\bar{y}^{(n)} + C\bar{z}^{(n)}\|_2^2 \\
&\quad - 2\rho^2\langle B\bar{y}^{(n)} - B\bar{y}^{(n-1)} + C\bar{z}^{(n)} - C\bar{z}^{(n-1)}, A\bar{x}^{(n)}\rangle \\
&\quad - 2\rho^2\langle C\bar{z}^{(n)} - C\bar{z}^{(n-1)}, B\bar{y}^{(n)}\rangle,
\end{aligned} \tag{4.18}$$

where the last inequality follows from (4.16).

From the definition of (4.14) we can replace $\hat{z}$ with $z^{(n-1)}$:

$$H(z^{(n-1)}) - H(z^{(n)}) + \langle C(z^{(n-1)} - z^{(n)}), \lambda^{(n)} + \rho(Ax^{(n)} + By^{(n)} + Cz^{(n)} - b)\rangle \geq 0. \tag{4.19}$$

On the other hand, take the $(n-1)$th iteration of the inequality (4.14) and let $\hat{z} = z^{(n)}$:

$$\begin{aligned} &H(z^{(n)}) - H(z^{(n-1)}) + \langle C(z^{(n)} - z^{(n-1)}), \lambda^{(n-1)} + \rho(Ax^{(n-1)} \\ &\quad + By^{(n-1)} + Cz^{(n-1)} - b)\rangle \geq 0. \end{aligned} \tag{4.20}$$

Add the above two inequalities and make some derivations of the inequality then we can see:

$$\begin{aligned} &-\rho\langle C\bar{z}^{(n)} - C\bar{z}^{(n-1)}, A\bar{x}^{(n)} + B\bar{y}^{(n)}\rangle \\ &\geq \rho\|C\bar{z}^{(n)} - C\bar{z}^{(n-1)}\|_2^2 + \rho\langle C\bar{z}^{(n)} - C\bar{z}^{(n-1)}, C\bar{z}^{(n-1)}\rangle. \end{aligned} \tag{4.21}$$

Take (4.21) into (4.18):

$$\begin{aligned} |\bar{\lambda}^{(n)}|^2 - |\bar{\lambda}^{(n+1)}|^2 \geq{}& \rho^2\|A\bar{x}^{(n)} + B\bar{y}^{(n)} + C\bar{z}^{(n)}\|_2^2 - 2\rho^2\langle B\bar{y}^{(n)} - B\bar{y}^{(n-1)}, A\bar{x}^{(n)}\rangle \\ &+ 2\rho^2\|C\bar{z}^{(n)} - C\bar{z}^{(n-1)}\|_2^2 + 2\rho^2\langle C\bar{z}^{(n)} - C\bar{z}^{(n-1)}, C\bar{z}^{(n-1)}\rangle. \end{aligned} \tag{4.22}$$

Since

$$\begin{aligned} &2\|C\bar{z}^{(n)} - C\bar{z}^{(n-1)}\|_2^2 + 2\langle C\bar{z}^{(n)} - C\bar{z}^{(n-1)}, C\bar{z}^{(n-1)}\rangle = \\ &\|C\bar{z}^{(n)}\|_2^2 - \|C\bar{z}^{(n-1)}\|_2^2 + \|C\bar{z}^{(n)} - C\bar{z}^{(n-1)}\|_2^2 \end{aligned} \tag{4.23}$$

yields:

$$\begin{aligned} |\bar{\lambda}^{(n)}|^2 - |\bar{\lambda}^{(n+1)}|^2 \geq{}& \rho^2\|A\bar{x}^{(n)} + B\bar{y}^{(n)} + C\bar{z}^{(n)}\|_2^2 \\ &+ \rho^2(\|C\bar{z}^{(n)}\|_2^2 - \|C\bar{z}^{(n-1)}\|_2^2) \\ &+ \rho^2\|C\bar{z}^{(n)} - C\bar{z}^{(n-1)}\|_2^2 \\ &- 2\rho^2\langle B\bar{y}^{(n)} - B\bar{y}^{(n-1)}, A\bar{x}^{(n)}\rangle. \end{aligned} \tag{4.24}$$

Then

$$\begin{aligned} &(|\bar{\lambda}^{(n)}|^2 + \rho^2\|C\bar{z}^{(n-1)}\|_2^2) - (|\bar{\lambda}^{(n+1)}|^2 + \rho^2\|C\bar{z}^{(n)}\|_2^2) \\ &\geq \rho^2\|C\bar{z}^{(n)} - C\bar{z}^{(n-1)}\|_2^2 + \rho^2\|A\bar{x}^{(n)} + B\bar{y}^{(n)} + C\bar{z}^{(n)}\|_2^2 \\ &\quad - 2\rho^2\langle B\bar{y}^{(n)} - B\bar{y}^{(n-1)}, A\bar{x}^{(n)}\rangle. \end{aligned} \tag{4.25}$$

It is easy to see that:

$$\|A\bar{x}^{(n)} + B\bar{y}^{(n)} + C\bar{z}^{(n)}\|_2^2 - 2(B\bar{y}^{(n)} - B\bar{y}^{(n-1)}, A\bar{x}^{(n)})$$

$$\geq \|A_1\bar{x}^{(n)} + B_1\bar{y}^{(n)}\|_2^2 - 2\langle B_1\bar{y}^{(n)} - B_1\bar{y}^{(n-1)}, A_1\bar{x}^{(n)}\rangle$$
$$\geq \|A_1\bar{x}^{(n)} + B_1\bar{y}^{(n-1)}\|_2^2 + \|B_1\bar{y}^{(n)}\|_2^2 - \|B_1\bar{y}^{(n-1)}\|_2^2. \tag{4.26}$$

Combine (4.26) with (4.25) we can get:

$$\begin{aligned}
&\left(|\bar{\lambda}^{(n)}|^2 + \rho^2\|C\bar{z}^{(n-1)}\|_2^2 + \rho^2\|B_1\bar{y}^{(n-1)}\|_2^2\right) \\
&\quad - \left(|\bar{\lambda}^{(n+1)}|^2 + \rho^2\|C\bar{z}^{(n)}\|_2^2 + \rho^2\|B_1\bar{y}^{(n)}\|_2^2\right) \\
&\geq \rho^2\|C\bar{z}^{(n)} - C\bar{z}^{(n-1)}\|_2^2 + \rho^2\|A_1\bar{x}^{(n)} + B_1\bar{y}^{(n-1)}\|_2^2.
\end{aligned} \tag{4.27}$$

Then the right-hand side is nonnegative, that is to say, the nonnegative sequences $\left\{(|\bar{\lambda}^{(n)}|^2 + \rho^2\|C\bar{z}^{(n-1)}\|_2^2 + \rho^2\|B_1\bar{y}^{(n-1)}\|_2^2)\right\}$ are decreasing. Therefore, it converges while the nonnegative sequences $\{\|C\bar{z}^{(n)} - C\bar{z}^{(n-1)}\|_2^2\}$ and $\{\|A_1\bar{x}^{(n)} + B_1\bar{y}^{(n-1)}\|_2^2\}$ converge to zero when n goes to infinity. Therefore, we have:

$$\lim_{n\to\infty} C\bar{z}^{(n)} - C\bar{z}^{(n-1)} = \lim_{n\to\infty} A_1\bar{x}^{(n)} + B_1\bar{y}^{(n-1)}. \tag{4.28}$$

The last two limits yield that:

$$\lim_{n\to\infty} B_1\bar{y}^{(n)} - B_1\bar{y}^{(n-1)} = \lim_{n\to\infty} A_1\bar{x}^{(n)} + B_1\bar{y}^{(n-1)} = 0 \tag{4.29}$$

which, combines with the pair-wise structure of the constraints, shows that:

$$\lim_{n\to\infty}\langle B\bar{y}^{(n)} - B\bar{y}^{(n-1)}, A\bar{x}^{(n)}\rangle = \lim_{n\to\infty}\langle B_1\bar{y}^{(n)} - B_1\bar{y}^{(n-1)}, A_1\bar{x}^{(n)}\rangle = 0. \tag{4.30}$$

Bring (4.28) and (4.30) into (4.8) and (4.15) and take the limit we can get:

$$\begin{aligned}
&\limsup_{n\to\infty} F(x^{(n)}) + G(y^{(n)}) + H(z^{(n)}) \leq F(\hat{x}) + G(\hat{y}) + H(\hat{z}) \\
&\leq \liminf_{n\to\infty} F(x^{(n)}) + G(y^{(n)}) + H(z^{(n)})
\end{aligned} \tag{4.31}$$

that is,

$$\lim_{n\to\infty} F(x^{(n)}) + G(y^{(n)}) + H(z^{(n)}) = F(\hat{x}) + G(\hat{y}) + H(\hat{z}). \tag{4.32}$$

Then $\{(x^{(n)}, y^{(n)}, z^{(n)})\}$ is a minimizing sequence of objective function, and it is convergent. Next, we need to prove the sequence $\{(x^{(n)}, y^{(n)}, z^{(n)}, \lambda^{(n)})\}$ converge to the KKT point.

From the above equation, the sequence $\{(x^{(n)}, y^{(n)}, z^{(n)}, \lambda^{(n)})\}$ has a convergent subsequence $\{(x^{(n_k)}, y^{(n_k)}, z^{(n_k)}, \lambda^{(n_k)})\}$. Letting

$$\lim_{k\to\infty}\{(x^{(n_k)}, y^{(n_k)}, z^{(n_k)}, \lambda^{(n_k)})\} = (\tilde{x}, \tilde{y}, \tilde{z}, \tilde{\lambda}). \tag{4.33}$$

Then $(\tilde{x}, \tilde{y}, \tilde{z}, \tilde{\lambda})$ is an optimal solution of the objective function and $\tilde{x}$, $\tilde{y}$ and $\tilde{z}$ satisfy the constraint of problem (2.3):

$$A\tilde{x} + B\tilde{y} + C\tilde{z} = b. \tag{4.34}$$

Considering the following x-subproblem:

$$x^{(n_k+1)} := \arg\min \left\{ F(x) + \frac{\rho}{2}\|Ax + By^{(n_k)} + Cz^{(n_k)} - b - \frac{\lambda^{(n_k)}}{\rho}\|_2^2 \right\}. \tag{4.35}$$

Its optimality condition is given by

$$\begin{aligned} &A^T(\lambda^{(n_k)} - \rho(Ax^{(n_k+1)} + By^{(n_k+1)} + Cz^{(n_k+1)} - b)) \\ &\quad + \rho A^T(B(y^{(n_k+1)} - y^{(n_k)}) + C(z^{(n_k+1)} - z^{(n_k)})) \in \partial F(x^{(n_k+1)}). \end{aligned} \tag{4.36}$$

Because of $\lambda^{(n_k+1)} = \lambda^{(n_k)} - \rho(Ax^{(n_k+1)} + By^{(n_k+1)} + Cz^{(n_k+1)} - b)$. Then (4.36) can be rewritten as

$$A^T\lambda^{(n_k+1)} + \rho(A_1^T B_1(y^{(n_k+1)} - y^{(n_k)}) + A_2^T C_2(z^{(n_k+1)} - z^{(n_k)})) \in \partial F(x^{(n_k+1)}). \tag{4.37}$$

By(4.28) and (4.29), and taking the limit for both sides that means $k \to \infty$, then we can obtain

$$A^T\tilde{\lambda} \in \partial F(\tilde{x}). \tag{4.38}$$

The same as the y-subproblem and z-subproblem and we can get:

$$B^T\lambda^{(n_k+1)} + \rho B_3^T C_3(z^{(n_k+1)} - z^{(n_k)}) \in \partial G(y^{(n_k+1)}), \tag{4.39}$$

$$C^T\lambda^{(n_k+1)} \in \partial H(z^{(n_k+1)}). \tag{4.40}$$

Using the property of limit, we can know that $\tilde{x}$, $\tilde{y}$, $\tilde{z}$ and $\tilde{\lambda}$ satisfy the KKT conditions of original problem (2.3):

$$\begin{cases} A^T\tilde{\lambda} \in \partial G(\tilde{x}), \\ B^T\tilde{\lambda} \in \partial G(\tilde{y}), \\ C^T\tilde{\lambda} \in \partial G(\tilde{z}), \\ A\tilde{x} + B\tilde{y} + C\tilde{z} = b. \end{cases} \tag{4.41}$$

Therefore, the sequence generated by Algorithm 1 converges to the KKT point $(\tilde{x}, \tilde{y}, \tilde{z}, \tilde{\lambda})$. The proof is done. □

4.2 Lemma 1

Lemma 1 *The pair-wise linear constraints in* (2.1) *or* (2.2) *have the following property*:

$$\begin{aligned}&\left\|\sum_{i=1}^{m} A_i\bar{x}_i^{(n)}\right\|_2^2 - 2\sum_{i=1}^{m-2}\langle A_i x_i^{(n)}, \sum_{j=i+1}^{m-1}(A_j\bar{x}_j^{(n)} - A_j\bar{x}_j^{(n-1)})\rangle \\ &\geq \sum_{i=1}^{m-2}\sum_{j=i+1}^{m-1} \|A_i^{(i,j)}\bar{x}_i^{(n)} + A_j^{(i,j)}\bar{x}_j^{(n-1)}\|_2^2 + \|A_j^{(i,j)}\bar{x}_j^{(n)}\|_2^2 - \|A_j^{(i,j)}\bar{x}_j^{(n-1)}\|_2^2.\end{aligned} \tag{4.42}$$

Proof According to the pair-wise structure of the constraints, the $L2$-norm can be separated to match the inner product terms:

$$\begin{aligned}&\|\sum_{i=1}^{m} A_i\bar{x}_i^{(n)}\|_2^2 = \sum_{i=1}^{m-1}\sum_{j=i+1}^{m} \|A_i^{(i,j)}\bar{x}_i^{(n)} + A_j^{(i,j)}\bar{x}_j^{(n)}\|_2^2 \\ &\geq \sum_{i=1}^{m-2}\sum_{j=i+1}^{m-1} \|A_i^{(i,j)}\bar{x}_i^{(n)} + A_j^{(i,j)}\bar{x}_j^{(n)}\|_2^2 \\ &\langle A_i x_i^{(n)}, (A_j\bar{x}_j^{(n)} - A_j\bar{x}_j^{(n-1)})\rangle = \langle A_i^{(i,j)} x_i^{(n)}, (A_j\bar{x}_j^{(n)} - A_j^{(i,j)}\bar{x}_j^{(n-1)})\rangle.\end{aligned} \tag{4.43}$$

For each (i, j), we have:

$$\begin{aligned}&\|A_i^{(i,j)}\bar{x}_i^{(n)} + A_j^{(i,j)}\bar{x}_j^{(n)}\|_2^2 - 2\langle A_i^{(i,j)} x_i^{(n)}, (A_j\bar{x}_j^{(n)} - A_j^{(i,j)}\bar{x}_j^{(n-1)})\rangle \\ &= \|A_i^{(i,j)}\bar{x}_i^{(n)} + A_j^{(i,j)}\bar{x}_j^{(n-1)}\|_2^2 + \|A_i^{(i,j)} x_i^{(n)}\|_2^2 - \|A_i^{(i,j)} x_i^{(n-1)}\|_2^2.\end{aligned} \tag{4.44}$$

Combining the last two equations will lead to the inequality we want. □

4.3 Proof of Theorem 2.2

Proof of Theorem 2.2 Similar to three-variable case, the augmented Lagrangian function of problem (2.9) always has a saddle point $(\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_m, \bar{\lambda})$ such that for any $(x_1, x_2, \ldots, x_m, \lambda)$

$$L(\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_m, \lambda) \leq L(\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_m, \bar{\lambda}) \leq L(x_1, x_2, \ldots, x_m, \bar{\lambda}). \tag{4.45}$$

The same derivation in the proof of Theorem 2.1 shows that:

$$\sum_{i=1}^{m} A_i\bar{x}_i - b = 0, \tag{4.46}$$

$$\sum_{i=1}^{m} F_i(x_i^{(n)}) - \sum_{i=1}^{m} F_i(\bar{x}_i) + \langle \bar{\lambda}, \sum_{i=1}^{m} A_i \bar{x}_i^{(n)} \rangle \geq 0, \tag{4.47}$$

where $\bar{x}_i^{(n)} = x_i^{(n)} - \bar{x}_i, i = 1, \ldots, m.$

On the other hand, according to the order of subproblems, we have: (for any x_i)

$$\begin{aligned} &F_i(x_i) - F_i(x_i^{(n)}) - \langle A_i \bar{x}_i^{(n)}, \lambda^{(n)} + \rho(A_1 x_1^{(n)} + \\ &\cdots + A_i x_i^{(n)} + A_{i+1} x_{i+1}^{(n-1)} + \cdots + A_m x_m^{(n-1)} - b) \rangle \geq 0 \;\; i = 1, \ldots, m. \end{aligned} \tag{4.48}$$

Add all m inequalities, let $x_i = \bar{x}_i$ and make some simplifications we will have:

$$\begin{aligned} &\sum_{i=1}^{m} F_i(\bar{x}_i) - \sum_{i=1}^{m} F_i(x_i^{(n)}) - \langle \lambda^{(n)}, \sum_{i=1}^{m} A_i \bar{x}_i^{(n)} \rangle - \rho \left\| \sum_{i=1}^{m} A_i \bar{x}_i^{(n)} \right\|_2^2 \\ &\quad + \rho \sum_{i=1}^{m-1} \langle A_i x_i^{(n)}, \sum_{j=i}^{m} (A_j \bar{x}_j^{(n)} - A_j \bar{x}_j^{(n-1)}) \rangle \geq 0. \end{aligned} \tag{4.49}$$

Add (4.49) to (4.47) and take the notation $\bar{\lambda}^{(n)} = \lambda^{(n)} - \bar{\lambda}$, we can see:

$$\begin{aligned} &-\langle \bar{\lambda}^{(n)}, \sum_{i=1}^{m} A_i \bar{x}_i^{(n)} \rangle - \rho \left\| \sum_{i=1}^{m} A_i \bar{x}_i^{(n)} \right\|_2^2 \\ &+ \rho \sum_{i=1}^{m-1} \langle A_i x_i^{(n)}, \sum_{j=i}^{m} (A_j \bar{x}_j^{(n)} - A_j \bar{x}_j^{(n-1)}) \rangle \geq 0. \end{aligned} \tag{4.50}$$

According to the update of Lagrangian multiplier, we have:

$$\bar{\lambda}^{(n+1)} - \bar{\lambda}^{(n)} = \lambda^{(n+1)} - \lambda^{(n)} = \rho \left(\sum_{i=1}^{m} A_i x_i^{(n)} - b \right) = \rho \sum_{i=1}^{m} A_i \bar{x}_i^{(n)}. \tag{4.51}$$

Therefore,

$$\begin{aligned} |\bar{\lambda}^{(n)}|^2 - |\bar{\lambda}^{(n+1)}|^2 &= \langle \bar{\lambda}^{(n)} - \bar{\lambda}^{(n+1)}, \bar{\lambda}^{(n)} + \bar{\lambda}^{(n+1)} \rangle \\ &= -\rho \langle \sum_{i=1}^{m} A_i \bar{x}_i^{(n)}, 2\bar{\lambda}^{(n)} + \rho \sum_{i=1}^{m} A_i \bar{x}_i^{(n)} \rangle \\ &= -\rho^2 \left\| \sum_{i=1}^{m} A_i \bar{x}_i^{(n)}) \right\|_2^2 - 2\rho \langle \bar{\lambda}^{(n)}, \sum_{i=1}^{m} A_i \bar{x}_i^{(n)} \rangle \\ &\geq \rho^2 \left\| \sum_{i=1}^{m} A_i \bar{x}_i^{(n)}) \right\|_2^2 \end{aligned}$$

$$-2\rho^2 \sum_{i=1}^{m-1} \langle A_i x_i^{(n)}, \sum_{j=i+1}^{m} (A_j \bar{x}_j^{(n)} - A_j \bar{x}_j^{(n-1)}) \rangle. \quad (4.52)$$

Take $x_i = \bar{x}_m$ in (4.48) we have:

$$F_m(\bar{x}_m) - F_m(x_m^{(n)}) - \langle A_m \bar{x}_m^{(n)}, \lambda^{(n)} + \rho \sum_{i=1}^{m} A_i \bar{x}_i^{(n)} \rangle \geq 0. \quad (4.53)$$

Let $\bar{x}_m = x_m^{(n-1)}$ in (4.53) and take the $(n-1)$th iteration. We will have:

$$F_m(x_m^{(n-1)}) - F_m(x_m^{(n)}) - \langle A_m \bar{x}_m^{(n)} - A_m \bar{x}_m^{(n-1)}, \lambda^{(n)} + \rho \sum_{i=1}^{m} A_i \bar{x}_i^{(n)} \rangle \geq 0, \quad (4.54)$$

$$F_m(x_m^{(n)}) - F_m(x_m^{(n-1)}) - \langle A_m \bar{x}_m^{(n-1)} - A_m \bar{x}_m^{(n)}, \lambda^{(n-1)} + \rho \sum_{i=1}^{m} A_i \bar{x}_i^{(n-1)} \rangle \geq 0. \quad (4.55)$$

Add the two equations together and we will see:

$$\begin{aligned} &-\langle \lambda^{(n)} - \lambda^{(n-1)}, A_m \bar{x}_m^{(n)} - A_m \bar{x}_m^{(n-1)} \rangle \\ &- \rho \langle A_m \bar{x}_m^{(n)} - A_m \bar{x}_m^{(n-1)}, \sum_{i=1}^{m} A_i \bar{x}_i^{(n)} - \sum_{i=1}^{m} A_i \bar{x}_i^{(n-1)} \rangle \geq 0. \end{aligned} \quad (4.56)$$

Combining with the update of λ yields that:

$$\begin{aligned} &- \rho \langle \sum_{i=1}^{m} A_i \bar{x}_i^{(n-1)}, A_m \bar{x}_m^{(n)} - A_m \bar{x}_m^{(n-1)} \rangle \\ &- \rho \langle A_m \bar{x}_m^{(n)} - A_m \bar{x}_m^{(n-1)}, \sum_{i=1}^{m} A_i \bar{x}_i^{(n)} - \sum_{i=1}^{m} A_i \bar{x}_i^{(n-1)} \rangle \geq 0. \end{aligned} \quad (4.57)$$

Then we can see:

$$\begin{aligned} |\bar{\lambda}^{(n)}|^2 - |\bar{\lambda}^{(n+1)}|^2 \geq \rho^2 \| \sum_{i=1}^{m} A_i \bar{x}_i^{(n)}) \|_2^2 + 2\rho^2 \langle A_m \bar{x}_m^{(n)} - A_m \bar{x}_m^{(n-1)}, A_m \bar{x}_m^{(n)} \rangle \\ - 2\rho^2 \sum_{i=1}^{m-2} \langle A_i x_i^{(n)}, \sum_{j=i+1}^{m-1} (A_j \bar{x}_j^{(n)} - A_j \bar{x}_j^{(n-1)}) \rangle. \end{aligned} \quad (4.58)$$

Take the identity

$$2\langle A_m \bar{x}_m^{(n)} - A_m \bar{x}_m^{(n-1)}, A_m \bar{x}_m^{(n)} \rangle$$

$$= \|A_m\bar{x}_m^{(n)}\|_2^2 - \|A_m\bar{x}_m^{(n-1)}\|_2^2 + \|A_m\bar{x}_m^{(n)} - A_m\bar{x}_m^{(n-1)}\|_2^2 \tag{4.59}$$

into (4.58):

$$\begin{aligned}
&(|\bar{\lambda}^{(n)}|^2 + \rho^2\|A_m\bar{x}_m^{(n-1)}\|_2^2) - (|\bar{\lambda}^{(n+1)}|^2 + \rho^2\|A_m\bar{x}_m^{(n)}\|_2^2) \\
&\geq \rho^2\|A_m\bar{x}_m^{(n)} - A_m\bar{x}_m^{(n-1)}\|_2^2 + \rho^2\|\sum_{i=1}^{m} A_i\bar{x}_i^{(n)}\|_2^2 \\
&\quad - 2\rho^2\sum_{i=1}^{m-2}\langle A_i x_i^{(n)}, \sum_{j=i+1}^{m-1}(A_j\bar{x}_j^{(n)} - A_j\bar{x}_j^{(n-1)})\rangle.
\end{aligned} \tag{4.60}$$

According to Lemma 1, we can get:

$$\begin{aligned}
&(|\bar{\lambda}^{(n)}|^2 + \rho^2\|A_m\bar{x}_m^{(n-1)}\|_2^2 + \rho^2\sum_{i=1}^{m-2}\sum_{j=i}^{m-1}\|A_j^{(i,j)}\bar{x}_j^{(n-1)}\|_2^2) \\
&\quad - (|\bar{\lambda}^{(n+1)}|^2 + \rho^2\|A_m\bar{x}_m^{(n)}\|_2^2 + \rho^2\sum_{i=1}^{m-2}\sum_{j=i}^{m-1}\|A_j^{(i,j)}\bar{x}_j^{(n)}\|_2^2) \\
&\geq \rho^2\|A_m\bar{x}_m^{(n)} - A_m\bar{x}_m^{(n-1)}\|_2^2 + \rho^2\sum_{i=1}^{m-2}\sum_{j=i+1}^{m-1}\|A_i^{(i,j)}\bar{x}_i^{(n)} + A_j^{(i,j)}\bar{x}_j^{(n-1)}\|_2^2.
\end{aligned} \tag{4.61}$$

Then the nonnegative sequence

$$\left(|\bar{\lambda}^{(n)}|^2 + \rho^2\|A_m\bar{x}_m^{(n-1)}\|_2^2 + \rho^2\sum_{i=1}^{m-2}\sum_{j=i}^{m-1}\|A_j^{(i,j)}\bar{x}_j^{(n-1)}\|_2^2\right) \tag{4.62}$$

is decreasing and has a lower bound, which means it converges. Therefore, the non-negative sequences of right-hand side in (4.61) have the limits 0, which yields:

$$\begin{aligned}
&\lim_{n\to\infty} A_m\bar{x}_m^{(n)} - A_m\bar{x}_m^{(n-1)} = \lim_{n\to\infty}\sum_{i=1}^{m} A_i\bar{x}_i^{(n)} = 0, \\
&\lim_{n\to\infty} A_i^{(i,j)}\bar{x}_i^{(n)} + A_j^{(i,j)}\bar{x}_j^{(n-1)} = 0\ (1 \leq i < j \leq m-1).
\end{aligned} \tag{4.63}$$

Based on the structure of pair-wise constraints and the last two limits, we can get:

$$\lim_{n\to\infty} A_j^{(i,j)}\bar{x}_j^{(n-1)} - A_j^{(i,j)}\bar{x}_j^{(n-1)} = 0,\ 1 \leq i < j \leq m-1 \tag{4.64}$$

which means:

$$\lim_{n\to\infty}\sum_{i=1}^{m-2}\langle A_i^{(i,j)}x_i^{(n)},\sum_{j=i}^{m-1}(A_j^{(i,j)}\bar{x}_j^{(n)}-A_j^{(i,j)}\bar{x}_j^{(n-1)})\rangle=0. \tag{4.65}$$

Take (4.65) and the first two limits in (4.63) to (4.49) and take the superior limit:

$$\sum_{i=1}^{m}F_i(\bar{x}_i)\geq\limsup_{n\to\infty}\sum_{i=1}^{m}F_i(x_i^{(n)}). \tag{4.66}$$

On the other hand, take the second limit in (4.63) to (4.47) and take the inferior limit:

$$\sum_{i=1}^{m}F_i(\bar{x}_i)\leq\liminf_{n\to\infty}\sum_{i=1}^{m}F_i(x_i^{(n)}). \tag{4.67}$$

Now we can say that Algorithm 2 converges to the minimum point:

$$\lim_{n\to\infty}\sum_{i=1}^{m}F_i(x_i^{(n)})=\sum_{i=1}^{m}F_i(\bar{x}_i). \tag{4.68}$$

Then $\{(x_1^{(n)},x_2^{(n)},\ldots,x_m^{(n)})\}$ is a minimizing sequence of objective function, and it is convergent. Next, we need to prove the sequence $\{(x_1^{(n)},x_2^{(n)},\ldots,x_m^{(n)},\lambda^{(n)})\}$ converge to the KKT point.

From the above equation, the sequence $\{(x_1^{(n)},x_2^{(n)},\ldots,x_m^{(n)},\lambda^{(n)})\}$ has a convergent subsequence $\{(x_1^{(n_k)},x_2^{(n_k)},\ldots,x_m^{(n_k)},\lambda^{(n_k)})\}$. Letting

$$\lim_{k\to\infty}\{(x_1^{(n_k)},x_2^{(n_k)},\ldots,x_m^{(n_k)},\lambda^{(n_k)})\}=(\tilde{x}_1,\tilde{x}_2,\ldots,\tilde{x}_m,\tilde{\lambda}). \tag{4.69}$$

Then $(\tilde{x}_1,\tilde{x}_2,\ldots,\tilde{x}_m,\tilde{\lambda})$ is an optimal solution of the objective function and $x_i(i=1,2,\ldots,m)$ satisfy the constraint of problem (2.3):

$$\sum_{i=1}^{m}A_i\tilde{x}_i=b. \tag{4.70}$$

Considering the following x_i-subproblem:

$$x_i^{(n_k+1)}:=\arg\min_{x_i}\{F_i(x_i)+\frac{\rho}{2}\|A_ix_i+\sum_{j=1,j\neq i}^{m}A_jx_j^{(n_k)}-b-\frac{\lambda^{(n_k)}}{\rho}\|_2^2\}. \tag{4.71}$$

Its optimality condition is given by

$$A_i^T\left(\lambda^{(n_k)}-\rho(\sum_{i=1}^{m}A_i x_i^{(n_k+1)}-b)\right)$$
$$+\rho A_i^T\left(\sum_{j=1,j\neq i}^{m}A_j(x_j^{(n_k+1)}-x_j^{(n_k)})\right)\in\partial F(x^{(n_k+1)}). \tag{4.72}$$

Because of $\lambda^{(n_k+1)}=\lambda^{(n_k)}-\rho(\sum_{i=1}^{m}A_i x_i^{(n_k+1)}-b)$. Then (4.72) can be rewritten as

$$A_i^T\lambda^{(n_k+1)}+\rho A_i^T\left(\sum_{j=1,j\neq i}^{m}A_j\left(x_j^{(n_k+1)}-x_j^{(n_k)}\right)\right)\in\partial F_i(x_i^{(n_k+1)}). \tag{4.73}$$

By(4.63) and (4.64), and taking the limit for both sides that means $k\to\infty$, then we can obtain

$$A_i^T\tilde{\lambda}\in\partial F_i(\tilde{x}_i). \tag{4.74}$$

Using the same method, we can get

$$A_j^T\tilde{\lambda}\in\partial F_j(\tilde{x}_j),\ (j=1,2,\ldots,m). \tag{4.75}$$

Then we can know that $\tilde{x}_i(i=1,\ldots,m)$ and $\tilde{\lambda}$ satisfy the KKT conditions of original problem (2.9):

$$\begin{cases}A_i^T\tilde{\lambda}\in\partial F_j(\tilde{x}_i),\ (i=1,2,\ldots,m),\\ \sum_{i=1}^{m}A_i\tilde{x}_i=b.\end{cases} \tag{4.76}$$

Then $(\tilde{x}_1,\tilde{x}_2,\ldots,\tilde{x}_m,\tilde{\lambda})$ is the KKT point, which completes the proof. □

References

1. 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), 1–122 (2011)
2. Chen, C., He, B., Ye, Y., Yuan, X.: The direct extension of ADMM for multi-block convex minimization problems is not necessarily convergent. Math. Program. **155**(1–2), 57–79 (2016)
3. Deng, W., Lai, M., Peng, Z., Yin, W.: Parallel multi-block ADMM with o (1/k) convergence. arXiv preprint arXiv:1312.3040 (2013)
4. Deng, W., Yin, W.: On the global and linear convergence of the generalized alternating direction method of multipliers. Rice University CAAM Technical Report, (TR12-14) (2012)
5. Deng, W., Yin, W.: On the global and linear convergence of the generalized alternating direction method of multipliers. J. Sci. Comput. **66**(3), 889–916 (2016)
6. He, B., Tao, M., Yuan, X.: A splitting method for separate convex programming with linking linear constraints. Optimization Online (2010)
7. He, B., Tao, M., Yuan, X.: Alternating direction method with gaussian back substitution for separable convex programming. SIAM J. Optim. **22**(2), 313–340 (2012)
8. He, B., Tao, M., Yuan, X.: A splitting method for separable convex programming. IMA J. Numer. Anal. **35**(1), 394–426 (2015)

9. He, B., Yuan, X.: Block-wise alternating direction method of multipliers for multiple-block convex programming and beyond. SMAI J. Comput. Math. **1**, 145–175 (2015)
10. Hong, M., Luo, Z.-Q.: On the linear convergence of the alternating direction method of multipliers. arXiv preprint arXiv:1208.3922 (2012)
11. Malioutov, D., Çetin, M., Willsky, A.S.: A sparse signal reconstruction perspective for source localization with sensor arrays. IEEE Trans. Signal Process. **53**(8), 3010–3022 (2005)
12. Peng, Y., Ganesh, A., Wright, J., Wenli, X., Ma, Y.: Rasl: Robust alignment by sparse and low-rank decomposition for linearly correlated images. IEEE Trans. Pattern Anal. Mach. Intell. **34**(11), 2233–2246 (2012)
13. Tao, M., Yuan, X.: Recovering low-rank and sparse components of matrices from incomplete and noisy observations. SIAM J. Optim. **21**(1), 57–81 (2011)
14. Yan, M., Yin, W.: Self equivalence of the alternating direction method of multipliers. arXiv preprint arXiv:1407.7400 (2014)
15. Yao, H., Gerstoft, P., Shearer, P.M., Mecklenbräuker, C.: Compressive sensing of the Tohoku-Oki Mw 9.0 earthquake: Frequency-dependent rupture modes. Geophys. Res. Lett. **38**(20), L20310 (2011)
16. Yuan, M., Lin, Y.: Model selection and estimation in regression with grouped variables. J. R. Stat. Soc. Ser. B **68**(1), 49–67 (2006)