ANZIAM J. 60(2018), 95–117
doi:10.1017/S1446181118000184

ON THE $O(1/K)$ CONVERGENCE RATE OF THE ALTERNATING DIRECTION METHOD OF MULTIPLIERS IN A COMPLEX DOMAIN

L. LI ✉ [1], G. Q. WANG [1] and J. L. ZHANG [1]

(Received 25 March, 2017; accepted 4 February, 2018; first published online 30 August 2018)

Abstract

We focus on the convergence rate of the alternating direction method of multipliers (ADMM) in a complex domain. First, the complex form of variational inequality (VI) is established by using the Wirtinger calculus technique. Second, the $O(1/K)$ convergence rate of the ADMM in a complex domain is provided. Third, the ADMM in a complex domain is applied to the least absolute shrinkage and selectionator operator (LASSO). Finally, numerical simulations are provided to show that ADMM in a complex domain has the $O(1/K)$ convergence rate and that it has certain advantages compared with the ADMM in a real domain.

2010 *Mathematics subject classification*: primary 90C25; secondary 65K10, 93C83.

Keywords and phrases: the alternating direction method of multipliers, convergence rate, Wirtinger calculus, least absolute shrinkage and selectionator operator.

1. Introduction

The augmented Lagrangian methods (ALMs) are a certain class of algorithms for solving constrained optimization problems. They were originally known as the method of multipliers (Hestenes [17]). In particular, a variant of the standard ALMs that uses partial updates, also known as the alternating direction method of multipliers (ADMM), has gained some attention. This was originally proposed by Gabay and Mercier [11] in the 1970s, and its convergence has been explored by many authors, including Gabay [10] and Eckstein [8]. Furthermore, the ADMM has been extensively explored in recent years due to its broad applications and empirical performance in a wide variety of problems such as image processing [27], machine learning and statistics [2, 29], sparse optimizations [25], signal processing [18] and many other relevant fields [1, 19–21].

[1] School of Mathematics, Physics and Statistics, Shanghai University of Engineering Science, Shanghai, China; e-mail: lilu@sues.edu.cn, guoq_wang@hotmail.com, xzhzhangjuli@163.com.

The convergence rate of the ADMM has been widely discussed. The $O(1/K)$ convergence rate of the ADMM with a separable two-block structure was shown by Monteiro and Svaiter [26] with some additional assumptions (for example, matrix A being full rank), and both two-block structures of ADMM were solved exactly. This was the first time that the ergodic iteration-complexity of the classical ADMM for a class of linearly constrained convex programming problems was established, with proper closed convex objective functions. He and Yuan [15] provided a unified proof of the $O(1/K)$ convergence rate for both the original ADMM and its linearized variant based on a variational inequality (VI) approach. A worst-case $O(1/K)$ convergence rate was proposed for the ADMM in a nonergodic sense by He and Yuan [16]. Lin and Ma [23] showed that, under some easily verifiable and reasonable conditions, the global linear convergence of the ADMM when $N \geq 3$ can still be assured. This is important, since the ADMM is a popular method for solving large-scale multi-block optimization models, and it is known to perform very well in practice even when $N \geq 3$. Cai and Han [3] showed that for the three-block case in the ADMM, when one of them is strongly convex, the direct extension of the ADMM is convergent and that the worst-case convergence rate is also estimated in both the ergodic and nonergodic senses for the direct extension of the ADMM.

Many nonlinear optimization problems in complex variables are commonly encountered in the domain of applied mathematics and engineering applications, for example signal processing and control theory. The usual method of analysing a complex-valued optimization problem is to separate it into the real and imaginary parts, and then to recast it into a equivalent real-valued optimization problem by doubling the size of the constraint conditions (see [24, 31, 33] and the references therein). Often, the classical optimization problem is dealt with by separating it into real and imaginary parts, but, in this way, it may lose unknown coupling relationship between the signals themselves [22, 30].

Recently, the ADMM in a complex domain (also denoted as the complex ADMM) was studied by Li et al. [22]. Based on the theory of Wirtinger calculus, the Lagrange function and the augmented Lagrangian function in a complex domain were studied, and the convergence of the complex ADMM was obtained. In what follows, we briefly recall the main ideas and the results considered in the earlier paper [22].

For convenience, we first explain some notation used. Let $\mathrm{Re}\{z\}$ and $\mathrm{Im}\{z\}$ denote the real and imaginary part of z, respectively. The superscripts $(\cdot)^T$, $\bar{(\cdot)}$, $(\cdot)^H$ and $(\cdot)^{-1}$ are used for the transpose, complex conjugate, complex conjugate transpose and inverse of a matrix, respectively. The one-norm and two-norm are denoted by $\|\centerdot\|_1$ and $\|\centerdot\|_2$, respectively.

Consider the separable convex optimization problem of a real-valued function in a complex domain with linear equality constraints

$$\underset{x,y}{\text{minimize}}\{f(x) + g(y) \mid Ax + By = b, x \in C^n, y \in C^m\}, \tag{1.1}$$

where $f : C^n \to R \cup \{+\infty\}$ and $g : C^m \to R \cup \{+\infty\}$ are proper, closed and convex functions (see, for example, [22]); $A \in C^{p\times n}$ and $B \in C^{p\times m}$ are given matrices and $b \in C^p$ is a given vector.

https://doi.org/10.1017/S1446181118000184 Published online by Cambridge University Press

The Lagrangian function $L_0(x, y, \lambda)$ and the augmented Lagrangian function $L_\rho(x, y, \lambda)$ of (1.1) are [22]

$$L_0(x, y, \lambda) = f(x) + g(y) + 2\,\mathrm{Re}\{\lambda^H(Ax + By - b)\} \tag{1.2}$$

and

$$L_\rho(x, y, \lambda) = f(x) + g(y) + 2\,\mathrm{Re}\{\lambda^H(Ax + By - b)\} + \rho\|Ax + By - b\|_2^2, \tag{1.3}$$

respectively, where $\lambda \in C^p$ is the dual variable and $\rho > 0$ is the penalty parameter.

The iterative scheme of the complex ADMM for (1.1) is

$$\begin{cases} x^{k+1} = \arg\min_x L_\rho(x, y^k, \lambda^k) & \text{(1.4a)} \\ y^{k+1} = \arg\min_y L_\rho(x^{k+1}, y, \lambda^k) & \text{(1.4b)} \\ \lambda^{k+1} = \lambda^k + \rho(Ax^{k+1} + By^{k+1} - b). & \text{(1.4c)} \end{cases}$$

Without loss of generality, we make the following two assumptions [22].

Assumption 1.1. The Lagrangian function $L_0(x, y, \lambda)$ given by (1.2) has a saddle point (x^*, y^*, λ^*).

Assumption 1.2. The (extended-real-valued) functions $f : C^n \to R \cup \{+\infty\}$ and $g : C^m \to R \cup \{+\infty\}$ are proper, closed and convex.

Theorem 1.3 [22]. *Under Assumptions 1.1 and 1.2, the complex ADMM iterations (1.4a)–(1.4c) have the following conclusions:*

(1) *residual convergence, that is, $r^k = Ax^k + By^k - b \to 0$ as $k \to \infty$;*
(2) *objective convergence, that is, $f(x^k) + g(y^k) \to f(x^*) + g(y^*)$ as $k \to \infty$; and*
(3) *dual variable convergence, that is, $\lambda^k \to \lambda^*$ as $k \to \infty$.*

Although the convergence of the complex ADMM has been achieved and the linear convergence rate in a real domain was shown by He and Yuan [15], the convergence rate for the complex ADMM was not obtained. An interesting question is whether we can generalize the $O(1/K)$ convergence rate for the ADMM in a real domain to the complex ADMM. As we will see later, this extension is by no means obvious or expected.

The purpose of this paper is to establish the $O(1/K)$ convergence rate for the complex ADMM. By means of the Wirtinger calculus technique, we give the form of VI in a complex domain, and then obtain the $O(1/K)$ convergence rate for the complex ADMM. Furthermore, the complex ADMM is applied to the standard and generalized least absolute shrinkage and selectionator operator (LASSO) models. Some numerical simulation results are reported to show that the complex ADMM has the $O(1/K)$ convergence rate and is indeed more efficient.

The outline of the paper is as follows. In Section 2, the form of VI in a complex domain is given by using the Wirtinger calculus technique. In Section 3, we establish the $O(1/K)$ convergence rate for the complex ADMM, and, in Section 4, the complex ADMM is applied to the LASSO model. In Section 5, some numerical simulations are provided. Finally, we make some concluding remarks in Section 6.

https://doi.org/10.1017/S1446181118000184 Published online by Cambridge University Press

2. VI in a complex domain

VI is an inequality involving a function, which has to be solved for all possible values of a given variable, belonging to a convex set [6]. The mathematical theory of VI was initially developed to deal with equilibrium problems; the Signorini problem [9], to be precise. The applicability of the theory has since been expanded to include problems from economics, finance, optimization and game theory [5, 12, 34]. In this section, the form of VI in a complex domain will be presented.

2.1. VI of convex optimization with linear equality constraints in a complex domain We consider the following convex optimization problem for a real-valued function in complex domain with linear equality constraints, namely,

$$\min_x \{f(x) \mid Ax = b, x \in C^n\}, \tag{2.1}$$

where f is a real-valued convex function with complex variables $x \in C^n$, $A \in C^{p\times n}$ is a given matrix and $b \in C^p$ is a given vector. An equivalent form of (2.1) (see for example, [22]) is

$$\min_x \{f(x) \mid A^c x^c = b^c\}, \tag{2.2}$$

where

$$A^c = \begin{pmatrix} A & 0 \\ 0 & \bar{A} \end{pmatrix} \in C^{2p\times 2n}, \quad x^c = \begin{pmatrix} x \\ \bar{x} \end{pmatrix} \in C^{2n} \quad \text{and} \quad b^c = \begin{pmatrix} b \\ \bar{b} \end{pmatrix} \in C^{2p}.$$

The Lagrangian function $L_0(x, \lambda)$ of the optimization problem (2.2) is

$$L_0(x, \lambda) = f(x) + (\lambda^c)^H (A^c x^c - b^c) = f(x) + 2\,\mathrm{Re}\{\lambda^H (Ax - b)\}. \tag{2.3}$$

Similarly to Assumption 1.1, we make the following assumption.

ASSUMPTION 2.1. The Lagrangian function $L_0(x, \lambda)$ given by (2.3) has a saddle point, that is, there exists (x^*, λ^*) for which

$$L_0(x^*, \lambda) \le L_0(x^*, \lambda^*) \le L_0(x, \lambda^*)$$

holds for all x and λ.

This can be expressed as

$$\begin{cases} L_0(x, \lambda^*) - L_0(x^*, \lambda^*) \ge 0 \\ L_0(x^*, \lambda^*) - L_0(x^*, \lambda) \ge 0. \end{cases} \tag{2.4}$$

An equivalent expression of (2.4) is the following VI, namely,

$$\begin{cases} x^* \in C^n, & f(x) - f(x^*) + 2\,\mathrm{Re}\{(x - x^*)^H (A^H \lambda^*)\} \ge 0 \quad \text{for all } x \in C^n, \\ \lambda^* \in C^m, & 2\,\mathrm{Re}\{(\lambda - \lambda^*)^H [-(Ax^* - b)]\} \ge 0 \quad \text{for all } \lambda \in C^m. \end{cases} \tag{2.5}$$

An optimal condition of (2.5) can be characterized as finding a μ^* that satisfies

$$f(x) - f(x^*) + 2\,\mathrm{Re}\{(\mu - \mu^*)^H \Psi(\mu^*)\} \ge 0 \quad \text{for all } \mu \in \Omega^2,$$

where

$$\mu = \begin{pmatrix} x \\ \lambda \end{pmatrix}, \quad \Psi(\mu) = \begin{pmatrix} A^H \lambda \\ -(Ax - b) \end{pmatrix} \quad \text{and} \quad \Omega^2 = C^n \times C^p.$$

2.2. VI of separable convex optimization with linear equality constraints in a complex domain By Assumption 1.1, the Lagrangian function $L_0(x, y, \lambda)$ given by (1.2) has a saddle point, that is, there exists (x^*, y^*, λ^*), for which

$$L_0(x^*, y^*, \lambda) \le L_0(x^*, y^*, \lambda^*) \le L_0(x, y, \lambda^*) \tag{2.6}$$

holds for all x, y, λ.

An equivalent expression of (2.6) is the following VI, namely,

$$\begin{cases} L_0(x, y^*, \lambda^*) - L_0(x^*, y^*, \lambda^*) \ge 0 & \text{for all } x \in C^n \\ L_0(x^*, y, \lambda^*) - L_0(x^*, y^*, \lambda^*) \ge 0 & \text{for all } y \in C^m \\ L_0(x^*, y^*, \lambda^*) - L_0(x^*, y^*, \lambda) \ge 0 & \text{for all } \lambda \in C^p. \end{cases} \tag{2.7}$$

To simplify (2.7),

$$\begin{cases} f(x) - f(x^*) + 2\,\mathrm{Re}\{(x - x^*)^H (A^H \lambda^*)\} \ge 0 & \text{for all } x \in C^n \\ g(y) - g(y^*) + 2\,\mathrm{Re}\{(y - y^*)^H (B^H \lambda^*)\} \ge 0 & \text{for all } y \in C^m \\ 2\,\mathrm{Re}\{(\lambda - \lambda^*)^H [-(Ax^* + By^* - b)]\} \ge 0 & \text{for all } \lambda \in C^p. \end{cases} \tag{2.8}$$

This implies that the VI reformulation of (2.8) is to find a $w^* = (x^*, y^*, \lambda^*) \in \Omega^3$ such that

$$\Phi(\mu) - \Phi(\mu^*) + 2\,\mathrm{Re}\{(w - w^*)^H \Psi(w^*)\} \ge 0 \quad \text{for all } w \in \Omega^3, \tag{2.9}$$

where

$$\mu = \begin{pmatrix} x \\ y \end{pmatrix}, \quad \omega = \begin{pmatrix} x \\ y \\ \lambda \end{pmatrix}, \quad \Psi(\omega) = \begin{pmatrix} A^H \lambda \\ B^H \lambda \\ -(Ax + By - b) \end{pmatrix},$$
$$\Phi(\mu) = f(x) + g(y), \quad \Omega^3 = C^n \times C^m \times C^p.$$

3. $O(1/K)$ convergence rate of the complex ADMM

3.1. Applying VI to the complex ADMM As noted by Boyed et al. [2], the variable x^{k+1} is an intermediate variable during the complex ADMM iterations (2.7), since it essentially requires only $(y^k, \lambda^k)^T$ to generate the x^{k+1}. Suppose $v^k = (y^k, \lambda^k)^T$ and $V^* = \{v^* = (y^*, \lambda^*)^T \mid \omega^* = (x^*, y^*, \lambda^*)^T \in \Omega^*\}$. Here, we use Ω^* to denote the solution set of (2.9).

LEMMA 3.1. *Let $\omega^{k+1} = (x^{k+1}, y^{k+1}, \lambda^{k+1})^T$ given by (1.4a)–(1.4c). Then, for all $\omega \in \Omega$,*

$$\Phi(\mu) - \Phi(\mu^{k+1}) + 2\,\mathrm{Re}\{(\omega - \omega^{k+1})^H [\Psi(\omega^{k+1}) + \Gamma(y^k, y^{k+1}) + P_0(v^{k+1} - v^k)]\} \ge 0, \tag{3.1}$$

where

$$\Gamma(y^k, y^{k+1}) = \rho \begin{pmatrix} A^H \\ B^H \\ 0 \end{pmatrix} B(y^k - y^{k+1}), \quad P_0 = \begin{pmatrix} 0 & 0 \\ \rho B^H B & 0 \\ 0 & I_p/\rho \end{pmatrix},$$

with ρ as in equation (1.3).

PROOF. From (1.4a) and (1.4b), we can conclude that x^{k+1} and y^{k+1} satisfy

$$f(x) - f(x^{k+1}) + 2\rho\,\mathrm{Re}\Big\{(x - x^{k+1})^H A^H\Big(Ax^{k+1} + By^k - b + \frac{\lambda^k}{\rho}\Big)\Big\} \geq 0 \tag{3.2}$$

and

$$g(y) - g(y^{k+1}) + 2\rho\,\mathrm{Re}\Big\{(y - y^{k+1})^H B^H\Big(Ax^{k+1} + By^{k+1} - b + \frac{\lambda^k}{\rho}\Big)\Big\} \geq 0, \tag{3.3}$$

respectively.

Substituting

$$\lambda^{k+1} = \lambda^k + \rho(Ax^{k+1} + By^{k+1} - b)$$

into (3.2) and (3.3) gives

$$f(x) - f(x^{k+1}) + 2\,\mathrm{Re}\{(x - x^{k+1})^H[A^H\lambda^{k+1} + \rho A^H B(y^k - y^{k+1})]\} \geq 0 \tag{3.4}$$

and

$$g(y) - g(y^{k+1}) + 2\,\mathrm{Re}\{(y - y^{k+1})^H B^H\lambda^{k+1}\} \geq 0, \tag{3.5}$$

respectively.

Reforming (3.4) and (3.5) yields

$$\Phi(\mu) - \Phi(\mu)^{k+1} + 2\,\mathrm{Re}\bigg\{\begin{pmatrix} x - x^{k+1} \\ y - y^{k+1}\end{pmatrix}^H\bigg[\begin{pmatrix} A^H\lambda^{k+1} \\ B^H\lambda^{k+1}\end{pmatrix} + \rho\begin{pmatrix} A^H \\ B^H\end{pmatrix}B(y^k - y^{k+1}) + \begin{pmatrix} 0 & 0 \\ 0 & \rho B^H B\end{pmatrix}\begin{pmatrix} x^{k+1} - x^k \\ y^{k+1} - y^k\end{pmatrix}\bigg]\bigg\} \geq 0.$$

Combining this with (1.4c) gives

$$\Phi(\mu) - \Phi(\mu)^{k+1} + 2\,\mathrm{Re}\left\{\begin{pmatrix} x - x^{k+1} \\ y - y^{k+1} \\ \lambda - \lambda^{k+1}\end{pmatrix}^H\left[\begin{pmatrix} A^H\lambda^{k+1} \\ B^H\lambda^{k+1} \\ -(Ax^{k+1} + By^{k+1} - b)\end{pmatrix} + \rho\begin{pmatrix} A^H \\ B^H \\ 0\end{pmatrix}B(y^k - y^{k+1}) + \begin{pmatrix} 0 & 0 \\ \rho B^H B & 0 \\ 0 & I_p/\rho\end{pmatrix}\begin{pmatrix} y^{k+1} - y^k \\ \lambda^{k+1} - \lambda^k\end{pmatrix}\right]\right\} \geq 0.$$

This completes the proof. □

LEMMA 3.2. *Let* $\omega^{k+1} = (x^{k+1}, y^{k+1}, \lambda^{k+1})^T$ *be given by* (1.4a)–(1.4c). *Then*

$$\mathrm{Re}\{(v^{k+1} - v^*)^H P(v^k - v^{k+1})\} \geq \mathrm{Re}\{(\omega^{k+1} - \omega^*)^H\Gamma(y^k, y^{k+1})\}. \tag{3.6}$$

PROOF. Substituting $\omega^* \in \Omega^*$ into (3.1) yields

$$\Phi(\mu^*) - \Phi(\mu^{k+1}) + 2\,\mathrm{Re}\{(\omega^* - \omega^{k+1})^H[\Psi(\omega^{k+1}) + \Gamma(y^k, y^{k+1}) + P_0(v^{k+1} - v^k)]\} \geq 0.$$

Regrouping it gives

$$2\,\mathrm{Re}\{(v^{k+1} - v^*)^H P(v^k - v^{k+1})\} \geq \Phi(\mu^{k+1}) - \Phi(\mu^*) + 2\,\mathrm{Re}\{(\omega^{k+1} - \omega^*)^H\Gamma(y^k, y^{k+1}) + (\omega^{k+1} - \omega^*)^H\Psi(\omega^{k+1})\}. \tag{3.7}$$

https://doi.org/10.1017/S1446181118000184 Published online by Cambridge University Press

Since Ψ is a monotone mapping, it can be written as

$$\begin{aligned}\Phi(\mu^{k+1}) - \Phi(\mu^*) + 2\,\mathrm{Re}\{\omega^{k+1} - \omega^*)^H\Psi(\omega^{k+1})\} &\geq \Phi(\mu^{k+1}) - \Phi(\mu^*) \\ &\quad + 2\,\mathrm{Re}\{\omega^{k+1} - \omega^*)^H\Psi(\omega^*)\} \\ &\geq 0. \end{aligned} \tag{3.8}$$

Substituting (3.8) into (3.7), we get the conclusion. This completes the proof. □

Lemma 3.3. *Let $\omega^{k+1} = (x^{k+1}, y^{k+1}, \lambda^{k+1})^T$ be given by (1.4a)–(1.4c). Then*

$$\mathrm{Re}\{(\omega^{k+1} - \omega^*)^H\Gamma(y^k, y^{k+1})\} = \mathrm{Re}\{(\lambda^{k+1} - \lambda^k)^H B(y^k - y^{k+1})\} \tag{3.9}$$

and

$$\mathrm{Re}\{(v^{k+1} - v^*)^H P(v^k - v^{k+1})\} \geq 0 \quad \textit{for all } v^* \in V^*,$$

where $P = \begin{pmatrix} \rho B^H B & 0 \\ 0 & I_p/\rho \end{pmatrix}$.

Proof. Since $Ax^* + By^* = b$ and $\lambda^{k+1} = \lambda^k + \rho(Ax^{k+1} + By^{k+1} - b)$,

$$\begin{aligned}\mathrm{Re}\{(\omega^{k+1} - \omega^*)^H\Gamma(y^k, y^{k+1})\} &= \mathrm{Re}\{[(\omega^{k+1} - \omega^*)^H\Gamma(y^k, y^{k+1})]^H\} \\ &= \mathrm{Re}\{[B(y^k - y^{k+1})]^H\rho[(Ax^{k+1} + By^{k+1}) - (Ax^* + By^*)]\} \\ &= \mathrm{Re}\{(\lambda^{k+1} - \lambda^k)^H B(y^k - y^{k+1})\}.\end{aligned}$$

It follows from (3.5) that

$$g(y) - g(y^{k+1}) + 2\,\mathrm{Re}\{(y - y^{k+1})B^H\lambda^{k+1}\} \geq 0 \tag{3.10}$$

and

$$g(y) - g(y^k) + 2\,\mathrm{Re}\{(y - y^k)B^H\lambda^k\} \geq 0. \tag{3.11}$$

Setting y for y^k in (3.10) and y^{k+1} in (3.11), and then adding the two resulting inequalities gives

$$\mathrm{Re}\{(\lambda^{k+1} - \lambda^k)^H B(y^k - y^{k+1})\} \geq 0. \tag{3.12}$$

Combining (3.6), (3.9) and (3.12) gives

$$\mathrm{Re}\{(v^{k+1} - v^*)^H P(v^k - v^{k+1})\} \geq 0 \quad \text{for all } v^* \in V^*. \tag{3.13}$$

This completes the proof. □

Because matrix P is positive semidefinite, we can define P-norm of vectors as

$$\|v - v^*\|_P^2 = (v - v^*)^H P(v - v^*) = \rho\|B(y - y^*)\|^2 + \frac{1}{\rho}\|\lambda - \lambda^*\|^2.$$

Theorem 3.4. *Let $\omega^{k+1} = (x^{k+1}, y^{k+1}, \lambda^{k+1})^T$ be given by (1.4a)–(1.4c), and $v^k = (y^k, \lambda^k)$. Then*

$$\|v^{k+1} - v^*\|_P^2 \leq \|v^k - v^*\|_P^2 - \|v^k - v^{k+1}\|_P^2 \quad \textit{for all } v^* \in V^*.$$

https://doi.org/10.1017/S1446181118000184 Published online by Cambridge University Press

PROOF. From (3.13),

$$\begin{aligned}\|v^k - v^*\|_P^2 &= \|(v^{k+1} - v^*) + (v^k - v^{k+1})\|_P^2 \\ &= \|v^{k+1} - v^*\|_P^2 + 2\,\mathrm{Re}\{(v^{k+1} - v^*)^H P(v^k - v^{k+1})\} + \|v^k - v^{k+1}\|_P^2 \\ &\geq \|v^{k+1} - v^*\|_P^2 + \|v^k - v^{k+1}\|_P^2.\end{aligned}$$

By rewriting this, the conclusion follows directly. This completes the proof. □

Theorem 3.4 implies the contractive property of the complex ADMM; for example, the residual $\|v^{k+1} - v^*\|_P^2$ decreases in each iteration.

3.2. Construction and depiction of instrumental variables $\widetilde{\omega}$

Let

$$\widetilde{\omega}^k = \begin{pmatrix} \widetilde{x}^k \\ \widetilde{y}^k \\ \widetilde{\lambda}^k \end{pmatrix} = \begin{pmatrix} x^{k+1} \\ y^{k+1} \\ \lambda^k + \rho(Ax^{k+1} + By^k - b) \end{pmatrix}. \tag{3.14}$$

By comparing [ ] ω^{k+1} and $\widetilde{\omega}^k$, we find that only the dual variable λ^{k+1} is different from $\widetilde{\lambda}^k$. From (3.14),

$$\begin{pmatrix} y^{k+1} \\ \lambda^{k+1} \end{pmatrix} = \begin{pmatrix} y^k \\ \lambda^k \end{pmatrix} - \begin{pmatrix} I_m & 0 \\ \rho B & I_p \end{pmatrix} \begin{pmatrix} y^k - \widetilde{y}^k \\ \lambda^k - \widetilde{\lambda}^k \end{pmatrix},$$

which can be represented as

$$v^{k+1} = v^k - M(v^k - \widetilde{v}^k), \tag{3.15}$$

where $M = \begin{pmatrix} I_m & 0 \\ \rho B & I_p \end{pmatrix}$.

LEMMA 3.5. *Let $\omega^{k+1} = (x^{k+1}, y^{k+1}, \lambda^{k+1})^T$ and $\widetilde{\omega}^k$ be given by (1.4a)–(1.4c) and (3.14), respectively. Then*

$$\Phi(\mu) - \Phi(\widetilde{\mu}^k) + 2\,\mathrm{Re}\{(\omega - \widetilde{\omega}^k)^H(\Psi(\widetilde{\omega}^k) + Q_0(\widetilde{v}^k - v^k))\} \geq 0 \quad \text{for all } \omega \in \Omega, \tag{3.16}$$

where $Q_0 = \begin{pmatrix} 0 & 0 \\ \rho B^H B & 0 \\ B & I_p/\rho \end{pmatrix}$.

PROOF. By (3.14),

$$x^{k+1} = \widetilde{x}^k, \quad y^{k+1} = \widetilde{y}^k \quad \text{and} \quad \lambda^{k+1} = \widetilde{\lambda}^k - \rho B(y^k - y^{k+1}),$$

$$\Psi(\omega^{k+1}) + \Gamma(y^k, y^{k+1}) = \begin{pmatrix} A^H \widetilde{\lambda}^k \\ B^H \widetilde{\lambda}^k \\ -(A^H \widetilde{x}^k + B^H \widetilde{y}^k - b) \end{pmatrix} = \Psi(\widetilde{\omega}^k) \tag{3.17}$$

and

$$\begin{aligned} P_0(v^{k+1} - v^k) &= \begin{pmatrix} 0 & 0 \\ \rho B^H B & 0 \\ 0 & I_p/\rho \end{pmatrix} \begin{pmatrix} y^{k+1} - y^k \\ \lambda^{k+1} - \lambda^k \end{pmatrix} \\ &= \begin{pmatrix} 0 & 0 \\ \rho B^H B & 0 \\ B & I_p/\rho \end{pmatrix} (\widetilde{v}^k - v^k). \end{aligned} \tag{3.18}$$

https://doi.org/10.1017/S1446181118000184 Published online by Cambridge University Press

It follows from (3.17), (3.18) and Lemma 3.1 that the conclusion (3.16) can be obtained. This completes the proof. □

LEMMA 3.6. *The solution set of VI given by* (2.9) *is convex and it can be characterized as*

$$\Omega^* = \bigcap_{\omega\in\Omega}\{\widetilde{\omega} \in \Omega \mid \Phi(\mu) - \Phi(\widetilde{\mu}) + 2\,\mathrm{Re}\{(\omega - \widetilde{\omega})^H\Psi(\omega) \geq 0)\}\}.$$

PROOF. Let $\widetilde{\omega} \in \Omega^*$. Then

$$\Phi(\mu) - \Phi(\widetilde{\mu}) + 2\,\mathrm{Re}\{(\omega - \widetilde{\omega})^H\Psi(\widetilde{\omega})\} \geq 0 \quad \text{for all } \omega \in \Omega.$$

Since the mapping Ψ is monotonous on Ω,

$$\Phi(\mu) - \Phi(\widetilde{\mu}) + 2\,\mathrm{Re}\{(\omega - \widetilde{\omega})^H\Psi(\omega)\} \geq 0 \quad \text{for all } \omega \in \Omega.$$

This implies that $\widetilde{\omega}$ belongs to the right-hand set in (3.6).

Let $\omega \in \Omega$. Then the vector

$$\bar{\omega} = \alpha\widetilde{\omega} + (1 - \alpha)\omega \quad \text{for all } \alpha \in (0, 1)$$

belongs to Ω. Furthermore, we can conclude that

$$\Phi(\bar{\mu}) - \Phi(\widetilde{\mu}) + 2\,\mathrm{Re}\{(\bar{\omega} - \widetilde{\omega})^H\Psi(\bar{\omega})\} \geq 0.$$

On the other hand, since the function Φ is convex,

$$\Phi(\bar{\mu}) \leq \alpha\Phi(\widetilde{\mu}) + (1 - \alpha)\Phi(\mu, \text{for all } \alpha \in (0, 1)).$$

Therefore,

$$\Phi(\mu) - \Phi(\widetilde{\mu}) + 2\mathrm{Re}\{(\omega - \widetilde{\omega})^H\Psi(\alpha\widetilde{\omega} + (1 - \alpha)\omega)\} \geq 0.$$

Let $\alpha \to 1$. Then

$$\Phi(\mu) - \Phi(\widetilde{\mu}) + 2\,\mathrm{Re}\{(\omega - \widetilde{\omega})^H\Psi(\widetilde{\omega})\} \geq 0,$$

which implies that $\widetilde{\omega} \in \Omega^*$.

For all $\omega \in \Omega$, the set

$$\{\widetilde{\omega} \in \Omega \mid \Phi(\widetilde{\mu}) + 2\,\mathrm{Re}\{\widetilde{\omega}^H\Psi(\omega)\} \leq \Phi(\mu) + 2\,\mathrm{Re}\{\omega^H\Psi(\omega)\}\}$$

and its equivalent form

$$\{\widetilde{\omega} \in \Omega \mid \Phi(\mu) - \Phi(\widetilde{\mu}) + 2\,\mathrm{Re}\{(\omega - \widetilde{\omega})^H\Psi(\omega)\} \geq 0, \text{for all } \omega \in \Omega\}$$

is convex. Since the intersection of any number of convex sets is convex, it follows that the solution set of Ω^* is convex. This completes the proof. □

Lemma 3.6 implies that $\widetilde{\omega} \in \Omega$ is an approximate solution of Ω^* with the accuracy $\varepsilon > 0$, namely,

$$\Phi(\mu) - \Phi(\widetilde{\mu}) + 2\,\mathrm{Re}\{(\omega - \widetilde{\omega})^H\Psi(\omega)\} \geq -\varepsilon \quad \text{for all } \omega \in \Omega.$$

Theorem 3.7. *Let* $\omega^{k+1} = (x^{k+1}, y^{k+1}, \lambda^{k+1})$ *and* $\widetilde{\omega}^k$ *be given by* (1.4a)–(1.4c) *and* (3.14), *respectively. Then*

$$\mathrm{Re}\{(v - \widetilde{v}^k)^H PM(v^k - \widetilde{v}^k)\} \ge \tfrac{1}{2}(\|v - v^{k+1}\|_P^2 - \|v - v^k\|_P^2) \quad \textit{for all } v \in V.$$

Proof. It follows from (3.15) that $M(v^k - \widetilde{v}^k) = (v^k - v^{k+1})$. Then

$$\mathrm{Re}\{(v - \widetilde{v}^k)^H PM(v^k - \widetilde{v}^k)\} = \mathrm{Re}\{(v - \widetilde{v}^k)^H P(v^k - v^{k+1})\}. \tag{3.19}$$

To get (3.7), we only need to prove that

$$\mathrm{Re}\{(v - \widetilde{v}^k)^H P(v^k - v^{k+1})\} \ge \tfrac{1}{2}(\|v - v^{k+1}\|_P^2 - \|v - v^k\|_P^2) \quad \text{for all } v \in V^*.$$

Let $x_1, x_2, x_3, x_4 \in C^{m+p}$. Then

$$\begin{aligned}\mathrm{Re}\{(x_1 - x_2)^H P(x_3 - x_4)\} &= \tfrac{1}{2}(\|x_1 - x_4\|_P^2 - \|x_1 - x_3\|_P^2) \\ &\quad + \tfrac{1}{2}(\|x_3 - x_2\|_P^2 - \|x_4 - x_2\|_P^2).\end{aligned} \tag{3.20}$$

Substituting $\mathrm{Re}\{(v - \widetilde{v}^k)^H P(v^k - v^{k+1})\}$ into (3.20) gives

$$\begin{aligned}\mathrm{Re}\{(v - \widetilde{v}^k)^H P(v^k - v^{k+1})\} &= \tfrac{1}{2}(\|v - v^{k+1}\|_P^2 - \|v - v^k\|_P^2) \\ &\quad + \tfrac{1}{2}(\|v^k - \widetilde{v}^k\|_P^2 - \|v^{k+1} - \widetilde{v}^k\|_P^2).\end{aligned} \tag{3.21}$$

Since

$$\|v^k - \widetilde{v}^k\|_P^2 = \rho\|B(y^k - \widetilde{y}^k)\|^2 + \frac{1}{\rho}\|\lambda^k - \widetilde{\lambda}^k)\|^2$$

and

$$\|v^{k+1} - \widetilde{v}^k\|_P^2 = \frac{1}{\rho}\|\lambda^{k+1} - \widetilde{\lambda}^k\|^2 = \frac{1}{\rho}\|\rho B(y^k - \widetilde{y}^k)\|^2 = \rho\|B(y^k - \widetilde{y}^k)\|^2,$$

we have

$$\|v^k - \widetilde{v}^k\|_P^2 - \|v^{k+1} - \widetilde{v}^k\|_P^2 = \frac{1}{\rho}\|\lambda^k - \widetilde{\lambda}^k\|^2 \ge 0.$$

It follows from (3.21) that

$$\mathrm{Re}\{(v - \widetilde{v}^k)^H P(v^k - v^{k+1})\} \ge \tfrac{1}{2}(\|v - v^{k+1}\|_P^2 - \|v - v^k\|_P^2). \tag{3.22}$$

The conclusion follows directly by combining (3.19) and (3.22). This completes the proof. □

3.3. $O(1/K)$ Convergence rate of the complex ADMM

Theorem 3.8. *Let* w^k *and* $\widetilde{\omega}^k$ *be the sequences generated by* (1.4a)–(1.4c) *and* (3.14) *respectively. For any integer* $K > 0$, *let*

$$\widetilde{\omega}_K = \frac{1}{K+1}\sum_{k=0}^{K}\widetilde{\omega}^k. \tag{3.23}$$

Then $\widetilde{\omega}_K \in \Omega$ *and*

$$\Phi(\widetilde{\mu}_K) - \Phi(\mu) + 2\,\mathrm{Re}\{(\widetilde{\omega}_K - \omega)^H \Psi(\omega)\} \le \frac{1}{2(K+1)}\|v - v^0\|_P^2 \quad \textit{for all } \omega \in \Omega. \tag{3.24}$$

https://doi.org/10.1017/S1446181118000184 Published online by Cambridge University Press

Proof. It follows from (3.14) and $w^k \in \Omega$ that $\widetilde{\omega}^k \in \Omega$ for all $k \geq 0$. Then, by (3.23),

$$\widetilde{\omega}_K \in \Omega.$$

On the other hand, the inequalities (3.16) and (3.7) imply that

$$\Phi(\mu) - \Phi(\widetilde{\mu}^k) + 2\,\mathrm{Re}\{(\omega - \widetilde{\omega}^k)^H \Psi(\omega)\} + \tfrac{1}{2}\|\nu - \nu^k\|_P^2 \geq \tfrac{1}{2}\|\nu - \nu^{k+1}\|_P^2 \quad \text{for all } \omega \in \Omega. \tag{3.25}$$

Summing the inequality (3.25) over $k = 0, 1, \ldots, K$ gives

$$(K+1)\Phi(\mu) - \sum_{k=0}^{K} \Phi(\widetilde{\mu}^k) + 2\,\mathrm{Re}\Big\{\Big[(K+1)\omega - \sum_{k=0}^{K} \widetilde{\omega}^k\Big]^H \Psi(\omega)\Big\} + \tfrac{1}{2}\|\omega - \omega^0\|_P^2 \geq 0$$

for all $\omega \in \Omega$. It follows from (3.23) that

$$\frac{1}{K+1}\sum_{k=0}^{K} \Phi(\widetilde{\mu}^k) - \Phi(\mu) + 2\,\mathrm{Re}\{(\widetilde{\omega}_K - \omega)^H \Psi(\omega)\} \leq \frac{1}{2(K+1)}\|\nu - \nu^0\|_P^2 \quad \text{for all } \omega \in \Omega. \tag{3.26}$$

Since Ψ is convex and

$$\widetilde{\mu}_K = \frac{1}{K+1}\sum_{k=0}^{K} \widetilde{\mu}^k, \tag{3.27}$$

we have

$$\Phi(\widetilde{\mu}_K) = \frac{1}{K+1}\sum_{k=0}^{K} \Phi(\widetilde{\mu}^k).$$

Combining (3.26) and (3.27), the result (3.24) follows directly. This completes the proof. □

For a given compact set $\mathcal{D} \subset \Omega$, let $d = \sup\{\|\omega - \omega^0\|_P \mid \omega \in \mathcal{D}\}$, where $\omega^0 = (x^0, y^0, \lambda^0)$ is the initial iterate. Then, after K iterations of the complex ADMM (1.4a)–(1.4c), the point $\widetilde{w}_K \in \Omega$ defined in (3.23) satisfies

$$\sup_{\omega \in \mathcal{D}}\{\Phi(\widetilde{\mu}_K) - \Phi(\mu) + 2\,\mathrm{Re}\{(\widetilde{\omega}_K - \omega)^H \Psi(\omega)\}\} \leq \frac{d^2}{2(K+1)},$$

which means that $\widetilde{\omega}_K$ is an approximate solution of $\mathrm{VI}(\Omega, F, \theta)$ with the accuracy $O(1/K)$. That is, the convergence rate $O(1/K)$ of the complex ADMM (1.4) is established.

4. LASSO with the complex ADMM

In order to minimize deviation, due to the lack of independent variables, the model will usually choose the independent variables as much as possible. The modelling process, however, needs to find the explanatory power of the independent variables on the dependent variables which are collected most often, and a good choice of the independent variables is required to improve the explanatory and predictive precision of the model. Index selection is extremely important in the process of statistical modelling problems. The LASSO is an effective estimation method to realize index set to streamline. In typical applications, there are many more features than training examples, and the goal is to find a parsimonious model for the data. (For general background on the LASSO, see [32]). The LASSO has been widely applied, particularly in the analysis of biological data, where only a small fraction of a huge number of possible factors are actually predictive of some outcome of interest (see [14] for a representative case study).

4.1. Standard LASSO with the complex ADMM An important special case of l_1 regularized loss minimization in a complex domain is regularized linear regression, also called the LASSO [32]. This involves solving

$$\underset{x}{\text{minimize}}\{\|Ax-b\|_2^2+\delta\|x\|_1 \mid x\in C^n\}, \tag{4.1}$$

where $A\in C^{p\times n}$ is a given matrix, $b\in C^p$ is a given vector and $\delta>0$ is a scalar regularization parameter.

In the complex ADMM form, the LASSO problem (4.1) can be written as

$$\underset{x,y}{\text{minimize}}\{\|Ax-b\|_2^2+\delta\|y\|_1 \mid x=y, x,y\in C^n\}. \tag{4.2}$$

From (1.4a)–(1.4c), the iterations of the complex ADMM for (4.2) are

$$\begin{cases} x^{k+1}=\arg\min\limits_{x}\{\|Ax-b\|_2^2+\rho\|x-y^k+\tau^k\|_2^2\}, & (4.3a)\\ y^{k+1}=\arg\min\limits_{y}\Big\{\|y\|_1+\dfrac{\rho}{\delta}\|x^{k+1}-y+\tau^k\|_2^2\Big\}, & (4.3b)\\ \tau^{k+1}=\tau^k+x^{k+1}-y^{k+1}, & (4.3c)\end{cases}$$

where $\rho>0$ is the penalty parameter and $\tau=\lambda/\rho$ is a scaled dual variable.

THEOREM 4.1. *The analytical solution of* (4.3a) *is*

$$x^{k+1}=(A^HA+\rho I)^{-1}\{A^Hb+\rho(y^k-\tau^k)\}.$$

PROOF. It follows from the first-order optimality condition of (4.3a) that

$$2A^H(Ax-b)+2\rho(x-y^k+\tau^k)=0.$$

Simplifying it gives

$$x^{k+1}=(A^HA+\rho I)^{-1}\{A^Hb+\rho(y^k-\tau^k)\}.$$

This completes the proof. □

If the optimization problem (4.3a) is in real domain, then

$$x^{k+1}=(A^TA+\rho I)^{-1}(A^Tb+\rho(y^k-\tau^k)),$$

which is the same as in [2, Section 6.4].

https://doi.org/10.1017/S1446181118000184 Published online by Cambridge University Press

The optimization problem (4.3b) can be solved by the soft thresholding operator in a complex domain, denoted by [22]

$$y^{k+1} = S_{\delta/(2\rho)}(x^{k+1} + \tau^k).$$

From what has been discussed above, the iterations of the LASSO are

$$\begin{cases} x^{k+1} = (A^H A + \rho I)^{-1}(A^H b + \rho(y^k - \tau^k)) \\ y^{k+1} = S_{\delta/(2\rho)}(x^{k+1} + \tau^k) \\ \tau^{k+1} = \tau^k + x^{k+1} - y^{k+1}. \end{cases} \tag{4.4}$$

For (4.4), $r^k = x^k - y^k$ can be viewed as a residual for the primal feasibility condition and $s^k = -\rho(y^k - y^{k-1})$ can be viewed as a residual for the dual feasibility condition. When the two residuals are small, the error must be small [2]. Thus an appropriate termination criterion is that the primal residuals r^k and dual residuals s^k are small simultaneously, that is, $\|r^k\|_2 \le \varepsilon^{\text{pri}}$ and $\|s^k\|_2 \le \varepsilon^{\text{dual}}$, where ε^{pri} and $\varepsilon^{\text{dual}}$ are the tolerances for the primal and dual feasibility, respectively.

These tolerances can be chosen using an absolute and relative criterion such as $\varepsilon^{\text{pri}} = \sqrt{n}\varepsilon^{\text{abs}} + \varepsilon^{\text{rel}} \max\{\|x^k\|_2, \|y^k\|_2\}$, $\varepsilon^{\text{dual}} = \sqrt{n}\varepsilon^{\text{abs}} + \varepsilon^{\text{rel}}\|\rho\tau\|_2$, where $\varepsilon^{\text{abs}} > 0$ is an absolute tolerance and $\varepsilon^{\text{rel}} > 0$ is a relative tolerance [2].

4.2. Generalized LASSO with the complex ADMM The LASSO problem in a complex domain can be generalized as

$$\underset{x}{\text{minimize}}\{\|Ax - b\|_2^2 + \delta\|Fx - c\|_1 \mid x \in C^n\}, \tag{4.5}$$

where $A \in C^{p\times n}$ is a given matrix, $b \in C^p$ is a given vector, $\delta > 0$ is a scalar regularization parameter and $F \in C^{m\times n}$ is an arbitrary linear transformation [2]. An important special case is when $F \in R^{(n-1)\times n}$ is the difference matrix, that is

$$F_{i,j} = \begin{cases} 1 & \text{if } j = i+1, \\ -1 & \text{if } j = i, \\ 0 & \text{otherwise,} \end{cases}$$

and $A = I$, in which case the generalization reduces to

$$\underset{x}{\text{minimize}}\left\{\|x - b\|_2^2 + \delta \sum_{k=1}^{n-1} |x_{i+1} - x_i| \; x \in C^n\right\}.$$

Since the second term is the total variation of x, the problem is often called total variation denoising, and its applications in signal processing can be referred to [28].

In the complex ADMM form, the problem (4.5) can be written as

$$\underset{x,y}{\text{minimize}}\{\|Ax - b\|_2^2 + \delta\|y\|_1 \mid Fx - c = y, x \in C^n, y \in C^m\}. \tag{4.6}$$

Then the iterations of the complex ADMM for (4.6) are

$$\begin{aligned} x^{k+1} &= \arg\min_x (\|Ax - b\|_2^2 + \rho\|Fx - c - y^k + \tau^k\|_2^2) \\ y^{k+1} &= \arg\min_y \left(\|y\|_1 + \frac{\rho}{\delta}\|Fx^{k+1} - c - y + \tau^k\|_2^2\right) \\ \tau^{k+1} &= \tau^k + Fx^{k+1} - c - y^{k+1}. \end{aligned}$$

https://doi.org/10.1017/S1446181118000184 Published online by Cambridge University Press

Furthermore, the corresponding iterations of the generalized LASSO are

$$\begin{cases} x^{k+1} = (A^H A + \rho F^H F)^{-1}(A^H b + \rho F^H (y^k + c - \tau^k)) \\ y^{k+1} = S_{\delta/(2\rho)}(F x^{k+1} - c + \tau^k) \\ \tau^{k+1} = \tau^k + F x^{k+1} - c - y^{k+1}. \end{cases} \tag{4.7}$$

5. Numerical simulation

Here, we report some numerical simulation on the standard LASSO and the generalized LASSO to illustrate the performance of the complex ADMM proposed in this paper. All our numerical experiments are carried out on a PC with Intel(R) Core(TM) i5-4200U CPU at 2.30 GHz and 8 GB of physical memory. The PC runs MATLAB Version: R2017a on Window 7 Enterprise 64-bit operating system.

5.1. Numerical simulation of the complex ADMM for standard LASSO

Assume that $x_o \in C^n$ is a discrete complex signal generated by random $N(0, 1)$ with the length $n = 2000$. x_o is r-sparse, which contains (at most) $r = 200$ nonzero entries with $r \ll n$. The sparse ratio is 10%. Select $p = 800$ $(p < n)$ measurements uniformly at random matrix $A_{p\times n}$ via $A_{p\times n}x = b$. Hence reconstructing signal x from measurement b is generally an ill-posed problem, which is an undetermined system of linear equations [4, 7]. This problem can be formulated as the standard LASSO in a complex domain (4.1), namely,

$$\underset{x}{\text{minimize}}\{\|Ax - b\|_2^2 + \delta\|x\|_1 \mid x \in C^n\},$$

and its sparsest solution can be obtained from the LASSO iterations (4.4).

5.1.1 *Influence of the penalty parameter* ρ. It is well known that the penalty parameter ρ influences the convergence rate of the ADMM in a real domain. In this section, we consider the role of the penalty parameter ρ in the complex ADMM. For this purpose, we repeated the same experiments on a set of 100 randomly generated problems with the penalty parameter ρ from 1 to 100 with the interval step 1 by fixing the other parameters as follows:

$$A = \text{randn}(800, 2000) + i \cdot \text{randn}(800, 2000), \quad b = Ax_o, \delta = 1,$$

with the scaled dual variable $\tau = \lambda/\rho$. In the algorithm, we chose the absolute tolerance $\varepsilon^{\text{abs}} = 10^{-5}$, the relative tolerance $\varepsilon^{\text{rel}} = 10^{-5}$, the initial values of x, y and τ to be all zeros, and the maximum number of iterations $N_{c\text{ADMM}} = 5000$. The number of iterations and the time consumption with different values of the parameter ρ from 1 to 100 with interval step 1 are plotted in Figure 1, which shows that, when $20 \le \rho \le 40$, the performance of the complex ADMM is optimum.

In addition, we took $\varepsilon^{\text{abs}} = 10^{-3}, \varepsilon^{\text{rel}} = 10^{-3}$, and we repeated the same experiments on a set of 100 randomly generated problems with the penalty parameter ρ from 0.1 to 10 and the interval step 0.1 by fixing the other parameters, as above. The number of the iterations and the time consumption with the different values of parameter ρ

https://doi.org/10.1017/S1446181118000184 Published online by Cambridge University Press

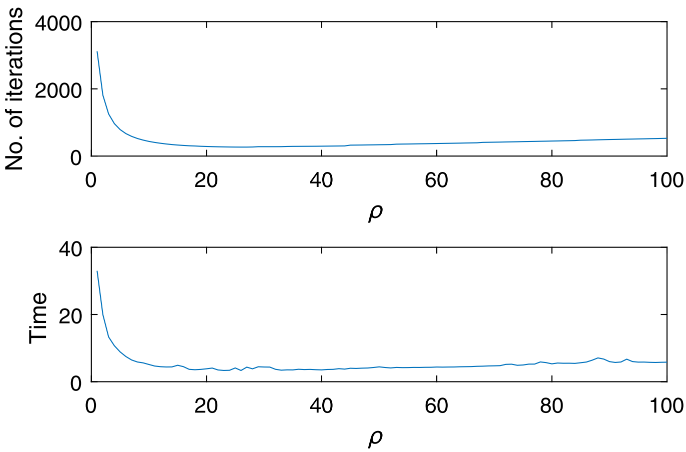


FIGURE 1. The convergence rate with different ρ for small tolerance.

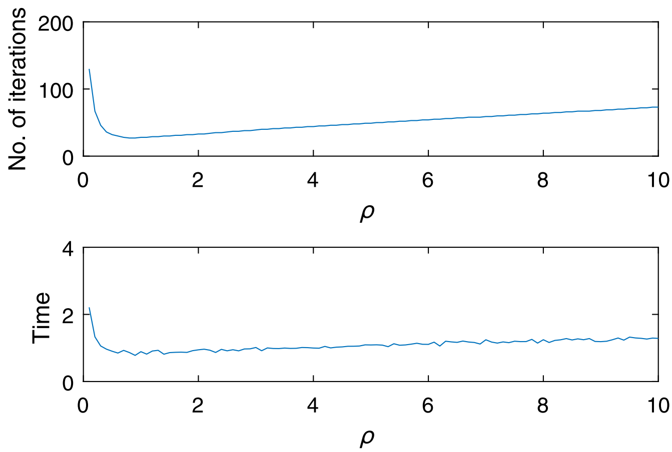


FIGURE 2. The convergence rate with different ρ for large tolerance.

are plotted in Figure 2, which shows that when the penalty parameter $\rho \approx 0.8$, the performance of the complex ADMM is optimum. However, it is hard to reach a definite conclusion about the role of the penalty parameter ρ in the complex ADMM from these experiments. The above observations are indicative. For smaller values of the absolute tolerance $\varepsilon^{\mathrm{abs}}$ and the relative tolerance $\varepsilon^{\mathrm{rel}}$ (for example, $\varepsilon^{\mathrm{abs}} = 10^{-5}, \varepsilon^{\mathrm{rel}=10^{-5}}$), the penalty parameter ρ needs to be be larger (for $\rho = 20$). Otherwise, smaller values of the penalty parameter ρ give better results.

5.1.2 *Numerical simulation of the complex ADMM with repeated experiments.* We repeated the random experiment with the penalty parameter $\rho = 20$ by fixing other parameters.

https://doi.org/10.1017/S1446181118000184 Published online by Cambridge University Press

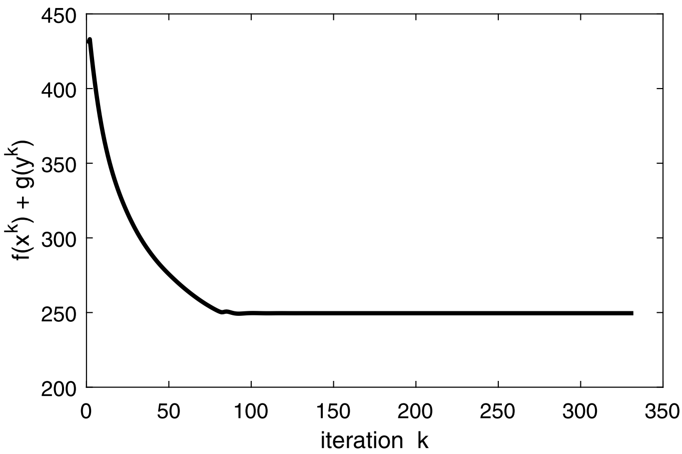
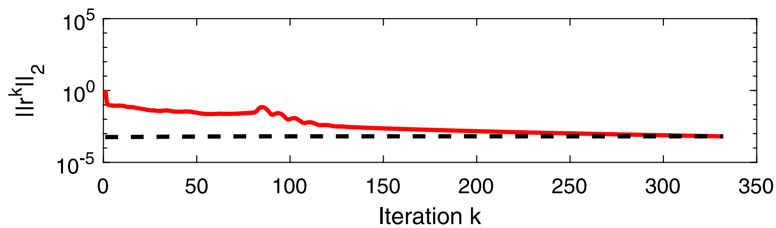
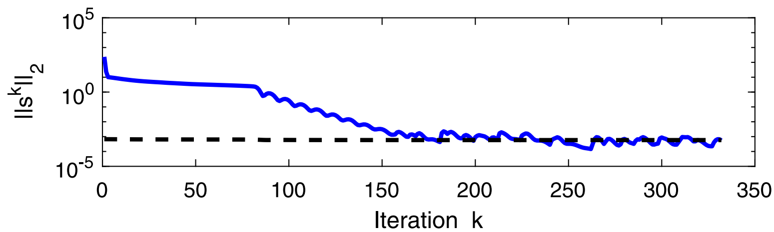


FIGURE 3. The objective function value by iteration k.



(a)



(b)

FIGURE 4. The primal residuals and the dual residuals by iteration k.

The performance of the objective function values in the process of the iteration is shown in Figure 3, which implies that the objective function values decrease through the iterations. In Figure 4(a), the full line describes the changes of the primal residuals r^k. In Figure 4(b), the full line describes the changes of the dual residuals s^k. The dotted lines in Figures 4(a) and 4(b) represent the absolute residual tolerance ε^{pri} and dual relative residuals tolerance $\varepsilon^{\text{dual}}$, respectively. From Figure 4, we can see that the two residuals descend linearly. In Figure 5, the original signal, the reconstructed signal and their comparison are displayed. The real part and the imaginary part of the

https://doi.org/10.1017/S1446181118000184 Published online by Cambridge University Press

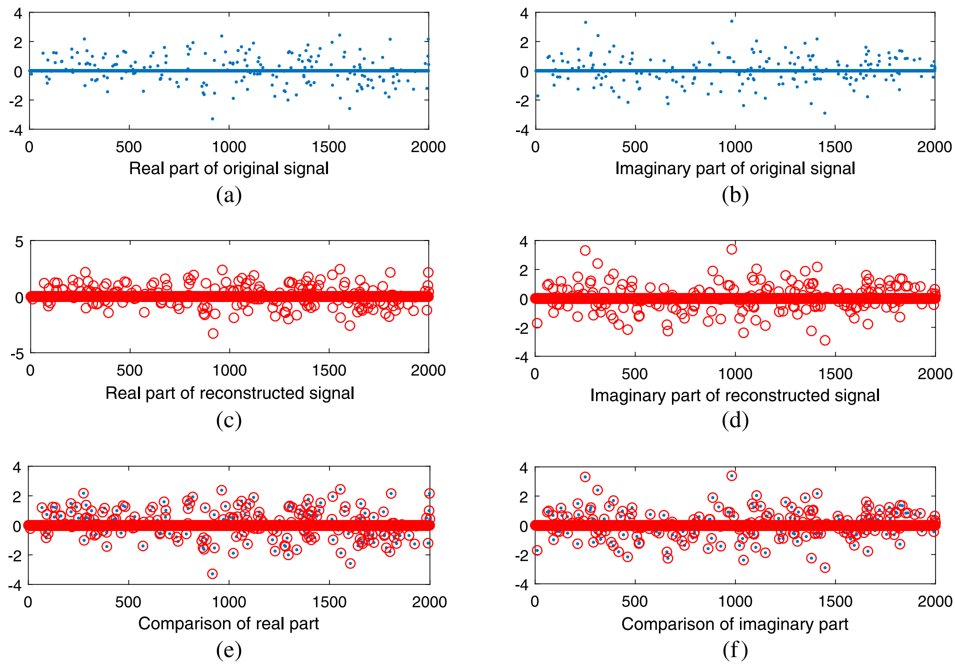


FIGURE 5. Comparison of the original signal and the reconstructed signal.

original signal (blue dots online) are shown in Figure 5(a) and 5(b), respectively. The real part and the imaginary part of the reconstructed signal (red circles online) are shown in Figure 5(c) and 5(d), respectively. A comparison of the real and imaginary parts of the original signal and the reconstructed signal is given in Figure 5(e) and 5(f), respectively. This shows that the reconstructed signal has restored the original signal very well.

Keeping the parameters fixed, we repeated the same experiments on a set of 100 randomly generated problems with different random signals, and we recorded the primal residual, the dual residual, the iteration number, the time consumption and the relative errors $\text{Error}(x_{\text{re}})$ and $\text{Error}(x_{\text{im}})$ with maximum, minimum and average values. The results are summarized in Table 1, which shows that the complex ADMM is efficient and robust. In Table 1,

$$\text{Error}(x_{\text{re}}) = \frac{\|\text{re}(x - x_o)\|_2}{\|\text{re}(x_o)\|_2} \quad \text{and} \quad \text{Error}(x_{\text{im}}) = \frac{\|\text{im}(x - x_o)\|_2}{\|\text{im}(x_o)\|_2}$$

indicate the relative error of the real and the imaginary parts of the reconstructed signal x.

5.1.3 *Comparison of the complex ADMM and the ADMM in a real domain.* To compare the efficiencies of the complex ADMM and the ADMM in a real domain, we separated the standard LASSO in a complex domain into the real part and the

https://doi.org/10.1017/S1446181118000184 Published online by Cambridge University Press

TABLE 1. The 100 experimental results of the complex ADMM.

Parameter	Maximum	Minimum	Average
Primal residual	6.5621×10^{-4}	6.1572×10^{-4}	6.4525×10^{-4}
Dual residual	5.7315×10^{-4}	1.0444×10^{-4}	3.3526×10^{-4}
Iteration number	345	321	331.4
Time consumption	9.0461	7.0766	7.5976
Error(x_{re})	4.9238×10^{-4}	3.1661×10^{-4}	3.9086×10^{-4}
Error(x_{im})	4.9245×10^{-4}	3.1790×10^{-4}	3.8777×10^{-4}

imaginary part, and then we recast it into an equivalent real-valued optimization problem by doubling the size of the constraint conditions (see [24, 31, 33]).

Assume that random signal x_o is same as in Section 5.1. Let $x_o^R = ({x_{o\text{re}}}^T, {x_{o\text{im}}}^T)^T \in R^{2n}$ denote the real composite $2n$-dimensional vector, where $x_{o\text{re}}$ and $x_{o\text{im}}$ are the real part and the imaginary part of the complex vector x_o, respectively. Then the standard LASSO in a complex domain (4.1) can be written as the standard LASSO in a real domain, namely,

$$\underset{x^R}{\text{minimize}}\{\|A^R x^R - b^R\|_2^2 + \delta\|x^R\|_1 \mid x^R \in R^{2n}\},$$

where

$$A^R = \begin{pmatrix} A_{\text{re}} & -A_{\text{im}} \\ A_{\text{im}} & A_{\text{re}} \end{pmatrix} \in R^{2p\times 2n}, \quad b^R = A^R x_o^R \in R^{2p}, \quad \delta = 1,$$

and A_{re} and A_{im} are the real part and the imaginary part of the complex matrix A, the penalty parameter $\rho = 20$ and the scaled dual variable $\tau = \lambda/\rho$. In the algorithm, we chose the absolute tolerance $\varepsilon^{\text{abs}} = 10^{-5}$, the relative tolerance $\varepsilon^{\text{rel}} = 10^{-5}$, the initial values of x, y and τ to be all zeros, and the maximum number of iterations $N_{\text{ADMM}} = 5000$.

The real part and imaginary part of the reconstructed complex signal x are the first half and second half of x^R, respectively, for example,

$$x = x^R(1 : n) + i \cdot x^R(n + 1 : 2n).$$

We repeated the same experiments on a set of 100 randomly generated problems and recorded the primal residual, the dual residual, the iteration number, the time consumption and the relative errors Error(x_{re}) and Error(x_{im}) with maximum, minimum and average values. The results are summarized in Table 2.

It follows from Tables 1 and 2 that the relative errors Error(x_{re}) and Error(x_{im}) of the complex ADMM are smaller than for the ADMM in a real domain, and other results are similar.

5.1.4 *Comparison of the complex ADMM and CVX.* To compare the efficiencies of the complex ADMM and the existing methods, we solved the optimization problem

https://doi.org/10.1017/S1446181118000184 Published online by Cambridge University Press

TABLE 2. The 100 experimental results of the ADMM in a real domain.

Parameter	Maximum	Minimum	Average
Primal residual	8.4144×10^{-4}	8.1988×10^{-4}	8.2504×10^{-4}
Dual residual	9.7520×10^{-4}	1.9255×10^{-4}	5.9882×10^{-4}
Iteration number	408	384	390.82
Time consumption	8.3651	7.0385	7.4040
Error(x_{re})	1.9039×10^{-3}	1.2302×10^{-3}	1.4758×10^{-3}
Error(x_{im})	1.9559×10^{-3}	1.1571×10^{-3}	1.4913×10^{-3}

TABLE 3. Comparison of the complex ADMM and CVX.

	Complex ADMM	cvx
Error(x_{re})	3.9086×10^{-4}	7.4573×10^{-4}
Error(x_{im})	3.8777×10^{-4}	7.4595×10^{-4}
Time consumption	7.5976	360.4572

(4.2) 100 times with the same scale data as in Section 5.1.2 by CVX, a package for specifying and solving convex programs [13].

The average values of Error(x_{re}), Error(x_{im}) and the time consumption of the complex ADMM and CVX are summarized in Table 3. It follows from Table 3 that the relative errors of the complex ADMM are smaller than for CVX, and the time consumed by the former is much less than by the latter.

5.2. Numerical simulation of the complex ADMM for generalized LASSO

Signal x_o is a limited variation signal with length $n = 2000$, that is, signal x_o is constant in a random $r = 50$ continuous interval, and such a signal can be seen as the image signal, as shown in Figure 8(a) and 8(b), respectively. Select $p = 200$ ($p = 4r, p < n$) measurements uniformly at random matrix $A_{p\times n}$ via $A_{p\times n}x_o = b$. Hence, a reconstruction of signal x_o from measurement b can be obtained by solving the generalized LASSO model (4.6)

$$\underset{x,y}{\text{minimize}}\{\|Ax-b\|_2^2 + \delta\|y\|_1 \mid Fx - c = y, x \in C^n, y \in C^p\},$$

and its sparsest solution can be obtained from the LASSO iterations (4.7). We consider the following random experiment:

$$A = \mathrm{randn}(200, 2000) + i * \mathrm{randn}(200, 2000),$$
$$b = Ax_o, \quad F_{1999\times 2000} = \begin{cases} 1 & \text{if } j = i + 1, \\ -1 & \text{if } j = i, \\ 0 & \text{otherwise}, \end{cases}$$

with the penalty parameter $\rho = 20, \delta = 1$ and the scaled dual variable $\tau = \rho/\lambda$. In the algorithm, we chose the absolute tolerance $\varepsilon^{\mathrm{abs}} = 10^{-5}$, the relative tolerance

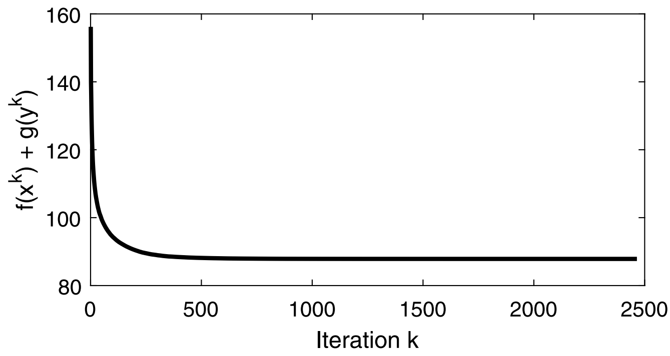
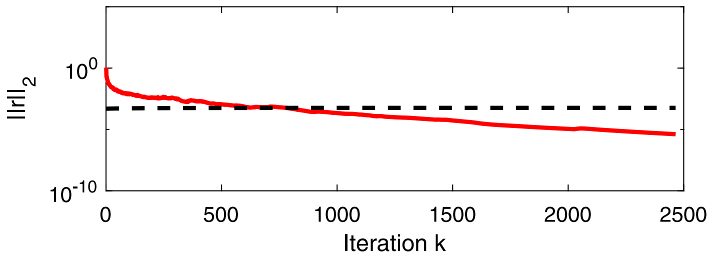
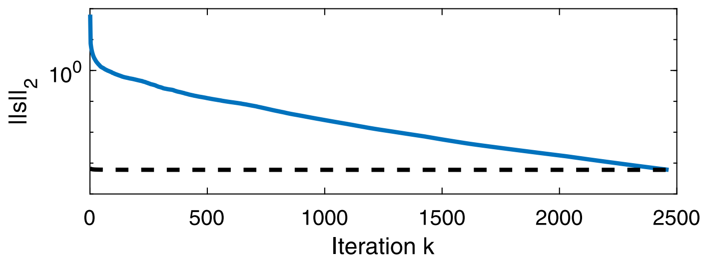


FIGURE 6. The objective function value by iteration k.



(a)



(b)

FIGURE 7. The primal residual and dual residual by iteration k.

$\varepsilon^{\text{rel}} = 10^{-5}$, the initial values of x, y and τ to be all zeros, and the maximum number of iterations $N_{c\text{ADMM}} = 5000$.

The objective function values in the process of iteration are plotted in Figure 6, which shows that the objective function values decrease monotonously through the iterations.

In Figure 7(a), the full line describes the changes in the primal residuals $\|r^k\|_2$ and descends linearly. In Figure 7(b), the full line describes the changes in the dual residuals $\|s^k\|_2$ and descends linearly too. The dotted lines in Figures 7(a)

https://doi.org/10.1017/S1446181118000184 Published online by Cambridge University Press

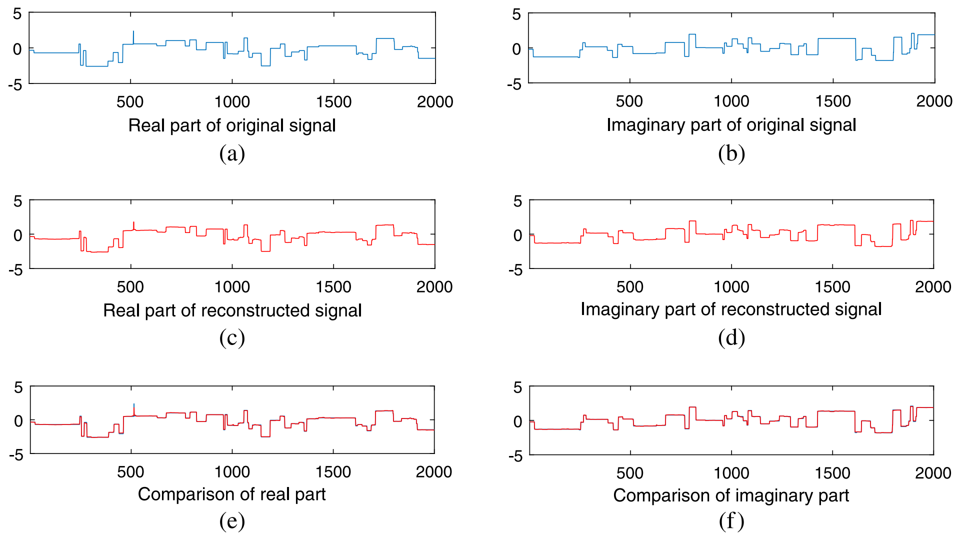


FIGURE 8. Comparison of the original signal and the reconstructed signal.

TABLE 4. The 100 experimental results of the ADMM in a real domain.

Parameter	Maximum	Minimum	Average
Primal residual	6.0874×10^{-4}	1.8385×10^{-4}	4.0680×10^{-4}
Dual residual	6.0569×10^{-4}	0.86990×10^{-4}	5.0717×10^{-4}
Iteration number	1743	443	652.6
Time(s)	130.3388	39.9726	62.0056
Error(x_{re})	0.1425	1.4711×10^{-4}	0.0425
Error(x_{im})	0.1443	2.4059×10^{-3}	0.0415

and 7(b) represent the primal residual tolerance ε^{pri} and the dual residual tolerance ε^{dul}, respectively.

The original signals (blue lines online), the reconstructed signals (red lines online) and their comparison are shown in Figure 8. This shows that the reconstructed signals have restored the original signals very well. Keeping the parameters fixed, we repeated the experiment 100 times with different random signals, and we recorded the average relative error and the time. The results are given in Table 4, which shows that the proposed complex ADMM is efficient.

6. Conclusions

In this paper, we explore the problem of convergence speed of the complex ADMM. First, we present the VI of the separable convex optimization of real-valued functions in a complex domain with linear equality constraints, based on the theory

https://doi.org/10.1017/S1446181118000184 Published online by Cambridge University Press

of complex analysis. Next, with the help of the contraction and the auxiliary variable sequences of the complex ADMM, its $O(1/K)$ convergence rate is established. Some preliminary numerical simulations on the standard LASSO model and the generalized LASSO model are also reported.

Acknowledgement

This research was supported by the National Natural Science Foundation of China (No. 11471211).

References

[1] Y. Q. Bai and K. J. Shen, "Alternating direction method of multipliers for ℓ_1–ℓ_2-regularized logistic regression model", *J. Oper. Res. Soc. China* **4** (2016) 243–253; doi:10.1007/s40305-015-0090-2.

[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. Learn.* **3** (2011) 1–122; doi:10.1561/2200000016.

[3] X. J. Cai, D. R. Han and X. M. Yuan, "On the convergence of the direct extension of ADMM for three-block separable convex minimization models with one strongly convex function", *Comput. Optim. Appl.* **66** (2017) 39–73; doi:10.1007/s10589-016-9860-y.

[4] E. J. Candès, J. Romberg and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information", *IEEE Trans. Inform. Theory* **52** (2006) 489–509; doi:10.1109/TIT.2005.862083.

[5] R. W. Cottle, F. Giannessi and J. L. Lions, *Variational inequalities and complementarity problems: theory and applications* (Wiley, Chichester, 1980); doi:10.1002/zamm.19810610316.

[6] S. C. Dafermosa, "An iterative scheme for variational inequalities", *Math. Program.* **26** (1983) 40–47; doi:10.1007/BF02591891.

[7] D. L. Donoho, "Compressed sensing", *IEEE Trans. Inform. Theory* **52** (2006) 1289–1306; doi:10.1109/TIT.2006.871582.

[8] J. Eckstein and D. P. Bertsekas, "On the Douglas–Rachford splitting method and the proximal point algorithm for maximal monotone operators", *Math. Program.* **55** (1992) 293–318; doi:10.1007/BF01581204.

[9] G. Fichera, "Abstract unilateral problems", in: *Conference on the theory of ordinary and partial differential equations* (2006) 51–64; doi:10.1007/BFb0066918.

[10] D. Gabay, "Applications of the method of multipliers to variational inequalities", *Stud. Math. Appl.* **15** (1983) 299–331; doi:10.1016/S0168-2024(08)70034-1.

[11] D. Gabay and B. Mercier, "A dual algorithm for the solution of nonlinear variational problems via finite element approximation", *Comput. Math. Appl.* **2** (1976) 17–40; doi:10.1016/0898-1221(76)90003-1.

[12] R. Glowinski, J. L. Lions and R. Tremoliérès, *Numerical analysis of variational inequalities* (North-Holland, Amsterdam, 1981); ISBN: 9780080875293.

[13] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.1", (2014); http://cvxr.com/cvx.

[14] T. Hastie, R. Tibshirani and J. Friedman, *The elements of statistical learning: data mining, inference and prediction* (Springer, New York, 2009); doi:10.1007/978-0-387-84858-7.

[15] B. S. He and X. M. Yuan, "On the $O(1/n)$ convergence rate of the Douglas–Rachford alternating direction method", *SIAM J. Numer. Anal.* **50** (2012) 700–709; doi:10.1137/110836936.

[16] B. S. He and X. M. Yuan, "On non-ergodic convergence rate of Douglas–Rachford alternating direction method of multipliers", *Numer. Math.* **130** (2015) 567–577; doi:10.1007/s00211-014-0673-6.

[17] M. R. Hestenes, "Multiplier and gradient methods", *J. Optim. Theory Appl.* **4** (1969) 303–320; doi:10.1007/BF00927673.
[18] T. Hinamoto, A. Doi and W. S. Lu, "Realization of 3-D separable-denominator digital filters with low l_2-sensitivity", *IEEE Trans. Signal Process.* **60** (2012) 6282–6293; doi:10.1109/TSP.2012.2215027.
[19] C. J. Li, C. Liu, K. Deng, X. H. Yu and T. W. Huang, "Data-driven charging strategy of PEVs under transformer aging risk", *IEEE Trans. Control Syst. Technol.* **Vol PP** (2017) 1–14; doi:10.1109/TCST.2017.2713321.
[20] J. Y. Li, C. J. Li, Z. Y. Wu, X. Y. Wang, K. L. Teo and C. Z. Wu, "Sparsity-promoting distributed charging control for plug-in electric vehicles over distribution networks", *Appl. Math. Model.* **58** (2018) 111–127; doi:10.1016/j.apm.2017.10.034.
[21] J. Y. Li, C. J. Li, Y. Xu, Z. Dong, K. Wong and T. W. Huang, "Noncooperative game-based distributed charging control for plug-in electric vehicles in distribution networks", *IEEE Trans. Ind. Inform.* **14** (2018) 301–310; doi:10.1109/TII.2016.2632761.
[22] L. Li, X. Y. Wang and G. Q. Wang, "Alternating direction method of multipliers for separable convex optimization of real functions in complex variables", *Math. Probl. Eng.* **2015** (2015) 1–14; doi:10.1155/2015/104531.
[23] T. Y. Lin, S. Q. Ma and S. Z. Zhang, "On the sublinear convergence rate of multi-block ADMM", *J. Oper. Res. Soc. China* **3** (2015) 251–274; doi:10.1007/s40305-015-0092-0.
[24] J. Lofberg, "YALMIP: a toolbox for modeling and optimization in MATLAB", *Optim.* **2004** (2004) 284–289; doi:10.1109/CACSD.2004.1393890.
[25] W. S. Lu and T. Hinamoto, "Two-dimensional digital filters with sparse coefficients", *Multidimens. Syst. Signal Process.* **22** (2011) 173–189; doi:10.1007/s11045-010-0129-9.
[26] R. D. C. Monteiro and B. F. Svaiter, "Iteration-complexity of block-decomposition algorithms and the alternating direction method of multipliers", *SIAM J. Optim.* **23** (2010) 475–507; doi:10.1137/110849468.
[27] M. Ng, P. Weiss and X. M. Yuan, "Solving constrained total-variation image restoration and reconstruction problems via alternating direction methods", *SIAM J. Sci. Comput.* **32** (2010) 2710–2736; doi:10.1137/090774823.
[28] L. I. Rudin, S. Osher and E. Fatemi, "Nonlinear total variation based noise removal algorithms", in: *Experimental mathematics, computational issues in nonlinear science: Proc. of the Eleventh Annual Int. Conf. of the Center for Nonlinear Studies, Los Alamos, NM, USA, 20–24 May 1991*. Volume 60 (Elsevier Science Publishers B.V., Amsterdam, 1992) 259–268; doi:10.1016/0167-2789(92)90242-F.
[29] K. Scheinberg, S. Q. Ma and D. Goldfarb, "Sparse inverse covariance selection via alternating linearization methods", in: *Twenty-fourth annual conference on neural information processing systems (NIPS) Vancouver, Canada, 6–9 December, 2010*, Volume 23 *Advances in Neural Information Processing Systems* (Curran Associates Inc., Red Hook, New York, 2010) 2101–2109; ISBN: 978-1-61782-380-0.
[30] L. Sorber, M. V. Barel and L. D. Lathauwer, "Unconstrained optimization of real functions in complex variables", *SIAM J. Optim.* **22** (2012) 8791–8898; doi:10.1137/110832124.
[31] J. F. Sturm, "Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cons", *Optim. Methods Softw.* **11** (1999) 625–653; doi:10.1080/10556789908805766.
[32] R. Tibshirani, "Regression shrinkage selection via the LASSO", *J. R. Stat. Soc. Ser. B Stat. Methodol.* **58** (1996) 267–288; URL: http://www.jstor.org/stable/2346178.
[33] K. C. Toh, M. J. Todd and R. H. Tütüncü, "SDPT3 – a Matlab software package for semidefinite programming, Version 1.3", *Optim. Methods Softw.* **11** (1999) 545–581; doi:10.1080/10556789908805762.
[34] L. Zhao and S. Dafermos, "General economic equilibrium and variational inequalities", *Oper. Res. Lett.* **10** (1991) 369–376; doi:10.1016/0167-6377(91)90037-P.