An Adaptive Sieving Strategy for the Complex Lasso Problem

Xiaoning Bai
School of Mathematics and Statistics
Beijing Institute of Technology
Beijing, P. R. China
bxn_2922@163.com

Yuge Ye
School of Mathematics and Statistics
Beijing Institute of Technology
Beijing, P. R. China
3120215720@bit.edu.cn

Qingna Li*
Beijing Key Laboratory on MCAACI
Beijing Institute of Technology
Beijing, P. R. China
qnl@bit.edu.cn

***Abstract*—The reconstruction of complex sparse signals is widely used in magnetic resonance imaging, radar and other fields. Most of the current algorithms convert problems with complex variables into real variables, which may cause the special structure of complex numbers to be ignored. In this paper, we propose an adaptive sieving (AS) strategy to effectively reduce the size of complex domain problems. The sieving strategy makes full use of the sparsity of the solution, which can effectively and safely select the nonzero features in the problem and reduce the CPU time. We apply the fast iterative shrinkage-thresholding algorithm (FISTA) to solve the small-scale subproblems after sieving. Numerical results show that the adaptive sieving strategy on FISTA (AS-FISTA) can effectively identify the nonzero features with low computational cost and fast speed.**

***Keywords*-Adaptive sieving strategy; Fast iterative shrinkage-thresholding algorithm; Complex lasso.**

I. Introduction

In this paper, we consider the recovery of complex sparse signals in linear models via the complex lasso (c-lasso) model [1], also known as the complex group lasso model [2]. It can be written as

$$\min_{x\in C^n} \frac{1}{2}\|Ax-b\|^2+\lambda\|x\|_1, \tag{c-lasso}$$

where $A: \ \mathcal{X}\to\mathcal{Y}$ is a linear map, and $\lambda>0$ is a given data. Here $\mathcal{X}$ and $\mathcal{Y}$ are two finite complex dimensional Euclidean spaces.

Restoring sparse signals from linear models has been studied extensively. Since Tibshirani proposed the lasso model [3] which can effectively generate sparse solutions, it has been applied in many fields. The existing works can be summarized into two lines. One is to transfer it into a problem in real variables and the other line is to develop methods directly dealing the complex variables. We briefly review the both lines below.

For the first line, based on the extensive research conducted on the lasso problem with real variables, a natural way is to reconstruct the c-lasso problem as an equivalent real variables optimization problem [4]. If it is transformed into a problem of real vectors, it is actually equivalent to solving the group lasso model, that is

$$\min_{\tilde{x}\in\mathbb{R}^{2n}} \frac{1}{2}\|\tilde{A}\tilde{x}-\tilde{b}\|^2+\lambda\sum_{i=1}^{n}\sqrt{\tilde{x}_i^2+\tilde{x}_{i+n}^2}, \tag{1}$$

where $\tilde{A} = \begin{bmatrix} A^{re} & -A^{im} \\ A^{im} & A^{re}\end{bmatrix}$, $\tilde{b} = \begin{bmatrix} b^{re} \\ b^{im}\end{bmatrix}$, $\tilde{x} = \begin{bmatrix} x^{re} \\ x^{im}\end{bmatrix}$ and a^{re}, a^{im} represent the real and imaginary parts of a respectively, a can be a complex vector or matrix. There are many first-order algorithms that can solve the c-lasso problem, including spectral projected-gradient algorithm [5], methods based on coordinate descent [6, 7], trust-region algorithm [8], active-set algorithm [9] and so on. Zhang et al. [10] proposed a semismooth Newton augmented Lagrangian (Ssnal) method which exploits fully the underlying second order sparsity for the sparse group lasso problem. This can also be used to solve (1).

For the second line, research has been focused on solving the c-lasso problem directly in the complex field using first-order methods. Hu [11] extended the block-coordinate descent (BCD) algorithms for the group lasso model [7] and obtained the BCD method for complex variables. The alternating direction method of multipliers (ADMM) has been also extended to separable convex optimization of real functions in complex variables [12].

On the other hand, to deal with large scale problems, in statistics and optimization, there is research on developing sieving methods to reduce the size of problem for many sparse models such as lasso model [13–17], group lasso model [14, 15], sparse group lasso model [18, 19], fused lasso model [20] etc. In addition, the adaptive sieving strategy in [21] starts with a few nonzero components and gradually adds more nonzero components by checking the KKT condition of the exclusive lasso problem. Yuan et al. generalized the results of [21] to apply this adaptive sieving strategy to a more complicated case. See [22] for more details.

To conclude, in many applications, such as magnetic resonance imaging and radar, variables are more easily represented in complex domains [23–26]. For example, many communication signals can be compressed on a Fourier

*Corresponding author. This author's research is supported by the National Science Foundation of China (NSFC) 12071032.

basis, while discrete cosine and wavelet basis are often suitable for compressing natural images [27]. That is, in many recovery problems of complex signals, the matrix A has a special structure, such as Fourier transform, circulant matrix and so on. For a complex variable $x \in C^n$, one can directly and quickly calculate the value of Ax without explicitly writing out the structure of matrix A. Therefore, to deal with the c-lasso problem in high-dimensional case, a natural question is whether one can develop a sieving strategy for the c-lasso problem, while maintaining the complex structure of the problem. This motivates the work in this paper.

The contributions of this paper are as follows. Inspired by [21], we propose an adaptive sieving (AS) strategy to reduce the size of the c-lasso problem based on checking the KKT condition of the problem. We also demonstrate the convergence of the AS algorithm. For the subproblem, we apply the fast iterative shrinkage-thresholding algorithm (FISTA) [28]. Numerical results show that the proposed AS strategy is very effective in reducing the size of the c-lasso problem and AS-FISTA significantly reduces the CPU time.

This paper is organized as follows. In Section II, we derive the KKT condition, give the definition of proximal residual function and the subgradient required in the sieving strategy. In Section III, we propose our AS strategy for the c-lasso problem and prove its convergence. Then we introduce how FISTA is applied to solve the subproblem. In Section IV, numerical results on different datasets verify the efficiency of our proposed AS-FISTA for the c-lasso problem. We conclude this paper in Section V.

II. Preliminaries

In this section, We derive the KKT condition and its equivalent form of the c-lasso problem.

Although the Cauchy-Riemann condition is not satisfied for the c-lasso problem, it is a real-valued function, the two conditions for a stationary point from the Wirtinger derivatives [29] of the c-lasso problem can been reduced down to a single condition [30]. For the case of $x = 0$, we try to define the subgradient of $\|x\|_1$. That is, a vector $v \in C$ is said to be a subgradient of $\|x\|_1$ at a point $x = 0$ if

$$\|y\|_1 \geq \|x\|_1 + \mathrm{Re}\left(v^H\left(y-x\right)\right),\ \forall\ y \in C.$$

Then $\partial\|x\|_1 = \{v \in C \mid \|v\|_1 \leq 1\}$ at $x = 0$, the KKT condition of the c-lasso problem is

$$0 \in A^H(Ax+b) + \partial\|x\|_1, \tag{2}$$

where

$$(\partial\|x\|_1)_i = \begin{cases} \frac{x_i}{\|x_i\|_1}, & x_i \neq 0, \\ \{v \in C \mid \|v\|_1 \leq 1\}, & x_i = 0, \end{cases} \quad i = 1, \ldots, n.$$

For the convenience of calculation, we next derive the equivalent form of the KKT condition. We can notice that the L_1 norm of a complex vector x can be regarded as a convex function of a real vector $\tilde{x} := \begin{bmatrix} x^{re} \\ x^{im} \end{bmatrix}$, i.e., $\|x\|_1 = \sum_{i=1}^n \sqrt{(x_i^{im})^2 + (x_i^{im})^2} := p(\tilde{x})$. For $x \in C^n,\ t > 0$, the shrinkage operator $\mathcal{T}$ of the L_1 norm in the complex field is

$$\mathrm{Prox}_{t\|\cdot\|_1}(x) = \underset{y\in C^n}{\arg\min}\, \frac{1}{2}\|y-x\|^2 + t\|y\|_1.$$

For any $x \in C^n$, $y \in C^n$,

$$\begin{aligned} &\min_{y\in C^n} \frac{1}{2}\|y-x\|^2 + t\|y\|_1 \\ &= \min_{y\in C^n} \sum_{i=1}^n \left(\frac{1}{2}\|y_i - x_i\|^2 + t\sqrt{(y_i^{re})^2 + (y_i^{im})^2}\right). \end{aligned}$$

so we can get $y_i^* = \mathrm{sgn}(x_i)\max(|x_i| - t, 0)$, where

$$\mathrm{sgn}(x_i) = \begin{cases} \frac{x_i}{|x_i|}, & x_i \neq 0, \\ 0, & x_i = 0. \end{cases}$$

For a real vector $\tilde{x}$, the Moreau envelope of p can be defined by

$$M_{tp}(\tilde{x}) := \min_{\tilde{y}\in\mathbb{R}^{2n}} \left\{\frac{1}{2}\|\tilde{y}-\tilde{x}\|^2 + tp(\tilde{y})\right\},$$

where $\tilde{y} = \begin{bmatrix} y^{re} \\ y^{im} \end{bmatrix}$, and $\mathrm{Prox}_{tp}(\tilde{x})$ is the unique optimal solution of the above problem. It is known that $\nabla M_{tp}(\tilde{x}) = \tilde{x} - \mathrm{Prox}_{tp}(\tilde{x})$, and $\mathrm{Prox}_{tp}(\cdot)$ is Lipschitz continuous with modulus 1 [31]. In this way, for a complex vector $x \in C^n$, $\mathrm{Prox}_{\mathrm{t}\|\cdot\|_1}(x) = \tilde{y}_{1:n} + \tilde{y}_{n+1:2n}i$, where $\tilde{y} = \mathrm{Prox}_{tp}(\tilde{x})$.

The proximal residual function $\mathrm{PR}(x)\ :\ C^n \to C^n$ associated with the c-lasso problem is

$$\mathrm{PR}\left(x\right) := x - \mathrm{Prox}_{\lambda\|\cdot\|_1}\left(x - A^H\left(Ax-b\right)\right),\ \forall\ x \in C^n,$$

and $\bar{x}$ satisfies the KKT condition if and only if $\mathrm{PR}(\bar{x}) = 0$.

III. Adaptive Sieving Strategy (AS) for the c-lasso Problem

Motivated by [21], we consider designing an AS strategy for the c-lasso problem by checking the KKT condition and discuss the convergence of AS. Next, we consider to apply FISTA to solve the subproblem.

A. AS Strategy for the c-lasso problem

The idea of our AS strategy is as follows. In each iteration l, let I^l be the set of indices corresponding to the nonzero elements in x^l. To estimate the set of indices I^{l+1} in next iteration, we check whether the KKT condition of the c-lasso problem are satisfied approximately. That is,

$$\mathrm{PR}\left(x^l\right) < \epsilon, \tag{3}$$

where $\epsilon > 0$ is a prescribed parameter. If (3) fails, we select the following indices based on the violation of KKT condition:

$$\begin{aligned} J^{l+1} := \Big\{ j \in \bar{I}^l \ \Big|\ &-\left(A^H\left(Ax-b\right)\right)_j \notin \\ &\left(\lambda\partial\left\|x^l\right\|_1 + q\mathbb{B}_\infty\right)_j \Big\}. \end{aligned}$$

Here, $\bar{I}^l = \{1, \ldots, p\} \backslash I^l$. The set of nonzero elements in next iteration is given by $I^{l+1} = I^l \bigcup J^{l+1}$.

The details of our AS strategy are summarized in Algorithm 1.

Algorithm 1 AS Algorithm for the c-lasso problem

Input: $A \in C^{n \times n},\ b \in C^n,\ \lambda > 0$, tolerance $\epsilon > \epsilon_1 > 0$.

Initialization: Construct $I^0 \subseteq \{1, \ldots, n\}$, where I^0 is the set of indices assuming x^* may not be zero. Let $l = 1$.

S0: Solve

$$\min_{x \in C^n} \left\{ \frac{1}{2} \|Ax - b\|^2 + \lambda \|x\|_1 - \langle \delta^0, x \rangle \mid x_{\bar{I}^0} = 0 \right\}$$

to find a solution x^0. Let $I := \{1, \ldots, n\},\ \bar{I}^0 := I \backslash I^0$. $\delta^0 \in C^n$ is an error vector such that $\|\delta^0\| \le \epsilon_1,\ (\delta^0)_{\bar{I}^0} = 0$.

S1: Create J^{l+1} by

$$J^{l+1} := \left\{ j \in \bar{I}^l \;\middle|\; -\left(A^H \left(Ax^l - b\right)\right)_j \notin \left(\lambda \partial \left\|x^l\right\|_1 + q\mathbb{B}_\infty\right)_j \right\},$$

where $q \le \frac{\epsilon - \epsilon_1}{\sqrt{|\bar{I}^l|}}$, $\mathbb{B}_\infty := \{v \mid \|v\|_\infty \le 1\}$. Let $I^{l+1} = I^l \cup J^{l+1}$.

S2: Solve the following problem:

$$\min_{x \in C^n} \left\{ \frac{1}{2} \|Ax - b\|^2 + \lambda \|x\|_1 - \langle \delta^{l+1}, x \rangle \mid x_{\bar{I}^{l+1}} = 0 \right\}$$

to find a solution x^{l+1}, $\delta^{l+1} \in C^n$ is an error vector such that $\|\delta^{l+1}\| \le \epsilon_1,\ (\delta^{l+1})_{\bar{I}^{l+1}} = 0$.

S3: Let $\bar{I}^{l+1} = I \backslash I^{l+1}$.

S4: If $\mathrm{PR}\left(x^{l+1}\right) \le \epsilon$, stop; otherwise, $l := l + 1$, go to **S1**.

B. Convergence of AS strategy

In Algorithm 1, we check the violation of the KKT condition by J^{l+1}, and gradually add the elements in $\bar{I}^l$ to I^{l+1}. As long as the required accuracy is not achieved, the set J^{l+1} is not empty. This is demonstrated in the following theorem.

Theorem 1 *If the proximal residual function* $\mathrm{PR}\left(x^l\right)$ *does not satisfy* (3)*, the set* J^{l+1} *is not empty.*

Proof: For contradiction, assume that $J^{l+1} = \emptyset$. The following holds

$$-\left(A^H \left(Ax^l - b\right)\right)_j \in \left(\lambda \partial \left\|x^l\right\|_1 + q\mathbb{B}_\infty\right)_j,\ \forall\ j \in \bar{I}^l.$$

Then there exists $\tilde{\delta} \in C^{|\bar{I}^l|}$, such that $\|\tilde{\delta}\|_\infty \le q$ and

$$-(A^H \left(Ax^l - b\right))_{\bar{I}^l} + \tilde{\delta} \in \lambda \partial \left\|\left(x^l\right)_{\bar{I}^l}\right\|_1. \qquad (4)$$

On the other hand, note that x^l is the solution of the following problem

$$\min_{x \in C^n} \left\{ \frac{1}{2} \|Ax - b\|^2 + \lambda \|x\|_1 - \langle \delta^l, x \rangle \;\middle|\; x^l_{\bar{I}^l} = 0 \right\},$$

where $\|\delta^l\| \le \epsilon_1,\ (\delta^l)_{\bar{I}^l} = 0$. Consequently, we know that there exists $y \in C^{|\bar{I}^l|}$ such that

$$\begin{cases} 0 \in A^H_{I^l} \left(Ax^l - b\right) - (\delta^l)_{I^l} + \lambda \partial \| \left(x^l\right)_{I^l} \|_1, \\ 0 \in A^H_{\bar{I}^l} \left(Ax^l - b\right) - (\delta^l)_{\bar{I}^l} + \lambda \partial \| \left(x^l\right)_{\bar{I}^l} \|_1 - y, \\ 0 = \left(x^l\right)_{\bar{I}^l}. \end{cases}$$

Therefore, together with (4), we have

$$-(A^H \left(Ax^l - b\right)) + \hat{\delta}^l \in \lambda \partial \left\|x^l\right\|_1.$$

where $\hat{\delta}^l \in C^n$ is defined as $\left(\hat{\delta}^l\right)_{\bar{I}^l} = \tilde{\delta},\ \left(\hat{\delta}^l\right)_{I^l} = (\delta^l)_{I^l}$, which means

$$x^l = \mathrm{Prox}_{\lambda \|\cdot\|_1}(x^l - A^H \left(Ax^l - b\right) + \hat{\delta}^l).$$

As a result, it holds that

$$\begin{aligned} &\left\|x^l - \mathrm{Prox}_{\lambda\|\cdot\|_1}(x^l - A^H \left(Ax^l - b\right))\right\| \\ =&\left\|\mathrm{Prox}_{\lambda\|\cdot\|_1}(x^l - A^H \left(Ax^l - b\right) + \hat{\delta}^l) \right. \\ &\left. - \mathrm{Prox}_{\lambda\|\cdot\|_1}(x^l - A^H \left(Ax^l - b\right))\right\| \\ \le&\|\tilde{\delta}\| + \|\delta^l\| \\ \le&\epsilon. \end{aligned}$$

It implies that $\mathrm{PR}\left(x^l\right) < \epsilon$. It contradicts with the assumption that x^l violates (3). Therefore if the condition (3) is not satisfied, the set J^{l+1} must be nonempty. ■

Remark 1 *Note that in Algorithm 1, the error vectors* $\delta^0,\ \delta^{l+1}$ *imply that the corresponding minimization problems could be solved inexactly. Actually, the vectors are not given in advance but they are the errors that occur when the original problems are solved inexactly.*

C. FISTA for the c-lasso problem

Considering that FISTA is an efficient algorithm that is widely used and easy to promote. In this part we use FISTA to solve the subproblem. As deduced above, we have obtained the shrinkage operator $\mathcal{T}$ of the L_1 norm in the complex field. Now we can apply FISTA to the c-lasso model as follows.

IV. NUMERCIAL RESULTS

In this section, we conduct numerical experiments to evaluate the performance of AS-FISTA for solving the c-lasso problem. It contains two parts. In the first part, we investigate the performance of Algorithm 1 based on test problems with different sizes. In the second part, we compare Algorithm 1 with Ssnal for different test problems. All of our results are obtained by running Matlab R2020a on

Algorithm 2 FISTA for the c-lasso problem

Input: $A \in C^{n\times n},\ b \in C^n,\ \lambda \in \mathbb{R}^+$, tolerance $\epsilon > 0$.

Initialization: $z^0 = x^0 \in C^n,\ t \in \left(0, \frac{1}{\|A^H A\|}\right),\ \alpha^0 = 1,\ k = 0$.

S0: Calculate $y^{k+1} = z^k - tA^H(Az^k - b)$.

S1: Update $x^{k+1} = \mathcal{T}_{t\lambda}(y^{k+1})$.

S2: Calculate $\alpha^{k+1} = \frac{1+\sqrt{1+4(\alpha^k)^2}}{2}$, $z^{k+1} = x^{k+1} + \frac{\alpha^k - 1}{\alpha^{k+1}}(x^{k+1} - x^k)$.

S3: If $\mathrm{PR}(x^{k+1}) \le \epsilon$, stop; otherwise, $k := k+1$, go to **S0**.

a windows workstation (Intel Core i5-7200U @ 2.50GHz, 12G RAM).

Test Problems We test the following examples.

- E1 We simulate random data from the regression model $Y_j = X_j^\top x^* + \epsilon_j, j = 1, \ldots, n$, where X is a circular matrix generated by a complex vector $c \in C^n$, which real part and imaginary part are generated from $N(0, 0.1)$ respectively and is independent of ϵ_j, $x^* = ((1+i)\mathbf{1}_{10}, (0.5+0.5i)\mathbf{1}_{10}, (0.25+0.25i)\mathbf{1}_{10}, \mathbf{0}_{n-30})^\top$, where $\mathbf{1}_{10}$ is a 10-dimensional vector of ones, and $\mathbf{0}_{p-30}$ is a $(p-30)$-dimensional vector of zeros. ϵ_j is generated by normal distribution with mean 0 and variance 1, where ϵ_j^{re} and ϵ_j^{im} are n-dimensional errors generated by $N(0, 1)$ respectively.
- E2 Similar as in E1, we take the error as $\epsilon \sim \sqrt{2}t(4)$, where $t(4)$ denotes the t distribution with 4 degree of freedom.
- E3 [1] We consider $X \in C^{N\times n}$, where $N = \lfloor 0.25n \rfloor$, the real and imaginary parts of every element of X is generated from $N\left(0, \frac{1}{N}\right)$. There are $\lfloor 0.1N \rfloor$ nonzero elements in x^*, and the amplitude of these elements are all 1, constructed with uniform phase.
- E4 [1] Similar as in E3, elements of X with real and imaginary parts distributed according to Rademacher distribution, i.e., $P\left(X_{i,j} = -\sqrt{\frac{1}{2N}}\right) = \frac{1}{2},\ P\left(X_{i,j} = \sqrt{\frac{1}{2N}}\right) = \frac{1}{2}$.
- E4 We consider $X \in C^{N\times n}$ to be a Hankel matrix whose first column is $c \in C^N$ generated by N-dimensional normal distribution and last row is $r \in C^n$ generated by n-dimensional normal distribution. The rest is the same as E1.
- E6 Similar as in E5, we take the error as $\epsilon \sim \sqrt{2}t(4)$, where $t(4)$ denotes the t distribution with 4 degree of freedom.

The parameters are taken as $\epsilon = 1.0 \times 10^{-5},\ \epsilon_1 = 0.9 \times 10^{-5}$. The initial value for $x_{J^{l+1}}^{l+1}$ is set to 0. We adopt the method in [32] to construct the initial I^0.

Taking into account of numerical rounding errors, we define the number of nonzero elements of a vector $x \in C^n$ and the index set of nonzero components of x as follows (see [33])

$$\hat{k} := \min\left\{k \;\middle|\; \sum_{i=1}^{k} |\tilde{x}_i| \ge 0.9999\,\|x\|_1\right\},$$
$$I(x) = \left\{i \;\middle|\; |x_i| \ge |\tilde{x}_{\hat{k}}|\ ,\ i = 1, \ldots, n\right\},$$

where $\tilde{x}$ is sorted, i.e., $|\tilde{x}_1| \ge |\tilde{x}_2| \ge \cdots \ge |\tilde{x}_n|$. Let $\bar{I}(x) = \{1, \ldots, n\} \setminus I(x)$.

We report the following information: the relative residual KKT of the solution obtained by AS-FISTA $\left(\eta_{as} = \frac{\|\mathrm{PR}(x^l)\|}{1+\|x^l\|}\right)$; the relative residual KKT of the solution obtained by Ssnal (η_{Ssn}); the running time of solving the original problem by Ssnal in second (T_{Ssn}); the running time of solving the original problem by FISTA in second (T_f); the running time of solving the subproblem by AS-FISTA in second (T_{as}); as in [34], the ratio of speed up (Rsp) is defined as T_f/T_{as}; the root mean square error (RMSE) between the solutions obtained by FISTA and that obtained by AS-FISTA.

A. Comparison with FISTA

In Fig 1, we show the variation of the relative residual KKT as the number of FISTA iterations increases. We can see that the relative residual KKT of AS-FISTA presents a stepwise change. This is because every change of the index set I will cause a significant drop in the relative residual KKT. When solving subproblems, FISTA will reduce the relative residual KKT at a relatively stable speed. From Fig 1, we can see that after AS-FISTA finds the correct index set, it achieves the accuracy required for the solution with higher efficiency.

In order to visualize the effect of our sieving strategy, we draw the following diagram on E1 with $n = 2000$. In this way, we can observe whether the added components gradually cover $I(\hat{x})$ after each iteration of AS. In Fig 2, the white circles represent the nonzero components of the true solution $\hat{x}$, and the black circles represent the components selected in each sieving. We can find that as the number of iterations increases, the white circles are gradually covered. This indicates that AS gradually selects the real nonzero components. It verifies that our AS strategy can effectively select the components that are nonzero after adding new components each time.

Table I shows the mean values after 20 simulations for different datasets. It can be seen that AS-FISTA successfully reduces the computation time, and the maximum speedup reaches 60, which indicates that our sieving rules reduce the computation time by more than 98%. AS-FISTA and FISTA return almost the same solutions due to the small values in RMSE. By comparing the results of the datasets of $n = 700$ and $n = 7000$, we can find that the larger and sparser the problems are, the greater the speedup of our AS strategy is.

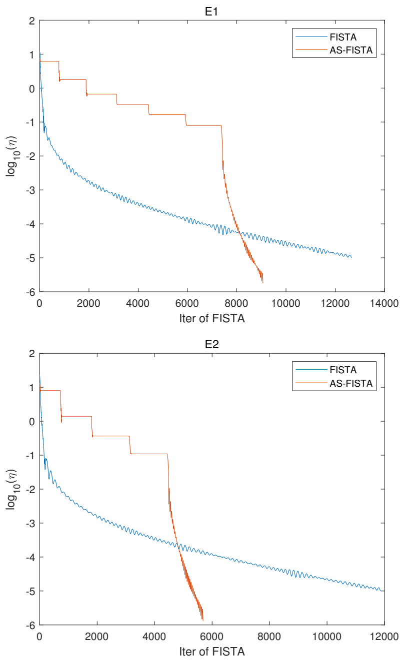

Figure 1. The relative residual KKT on E1 and E2 (n=700).

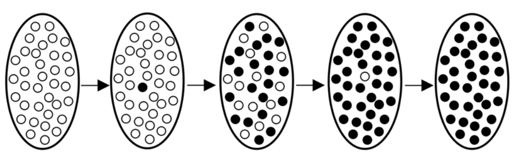

Figure 2. Schematic diagram of sieving effect in AS-FISTA.

Table I
RESULTS ON E1 AND E2.

n	Datasets	T_f	T_{as}	Rsp	RMSE
700	E1	1.79	0.52	3	5.27e-07
	E2	1.76	0.30	6	4.53e-07
7000	E1	190.88	3.85	50	1.44e-08
	E2	183.70	3.05	60	1.44e-08

B. Comparison with Ssnal

In this part, we compare Algorithm 1 with Ssnal[1] for the c-lasso problem. Table II summarizes the results for all test problems.

From Table II, we can see the quality of solutions provided by AS-FISTA is better on E1 and E2 datasets, while Ssnal has better solutions on E3 and E4. However, AS-FISTA is much faster than Ssnal. Even for E2, the dataset with the smallest gap, the time consumed by AS-FISTA is about a fifth of Ssnal. For the result on E5, where the gap is the largest, the time required for AS-FISTA is only less than 2% of Ssnal.

[1] https://github.com/YangjingZhang/SparseGroupLasso

Table II
RESULTS WITH $n = 7000$.

Datasets	$T_{as} \mid T_{Ssn}$	$\eta_{as} \mid \eta_{Ssn}$
E1	4.238 | 24.874	2.21e-06 | 6.46e-04
E2	3.112 | 15.631	2.31e-06 | 8.65e-04
E3	0.098 | 1.322	6.75e-06 | 9.53e-07
E4	0.220 | 8.111	6.82e-06 | 1.40e-07
E5	0.122 | 8.079	8.95e-06 | 5.57e-06
E6	1.322 | 7.167	8.85e-06 | 4.79e-06

V. CONCLUSION

In this paper, we design an AS strategy for the c-lasso problem. When solving subproblem with the same form as the original model, considering the c-lasso in many backgrounds, it is more beneficial to keep the complex form for solving to take advantage of the special structure of complex vectors. We extend FISTA to the c-lasso model, and give the optimality condition. AS-FISTA takes full advantage of the sparsity of knowledge, and it can effectively identify the nonzero component features. Experimental results show that our sieving rules are safe and effective, reducing computation time by up to 99%.

REFERENCES

[1] A. Maleki, L. Anitori, Z. Yang, and R. G. Baraniuk, "Asymptotic analysis of complex lasso via complex approximate message passing (camp)," *IEEE Transactions on Information Theory*, vol. 59, no. 7, pp. 4290–4308, 2013.

[2] Q. Wu, Y. D. Zhang, M. G. Amin, and B. Himed, "Complex multitask bayesian compressive sensing," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2014, pp. 3375–3379.

[3] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 58, no. 1, pp. 267–288, 1996.

[4] L. Sorber, M. V. Barel, and L. D. Lathauwer, "Unconstrained optimization of real functions in complex variables," *SIAM Journal on Optimization*, vol. 22, no. 3, pp. 879–898, 2012.

[5] E. Van Den Berg and M. P. Friedlander, "Probing the pareto frontier for basis pursuit solutions," *Siam journal on scientific computing*, vol. 31, no. 2, pp. 890–912, 2009.

[6] Y. Yang and H. Zou, "A fast unified algorithm for solving group-lasso penalize learning problems," *Statistics and Computing*, vol. 25, pp. 1129–1141, 2015.

[7] Z. Qin, K. Scheinberg, and D. Goldfarb, "Efficient block-coordinate descent algorithms for the

group lasso," *Mathematical Programming Computation*, vol. 5, no. 2, pp. 143–169, 2013.

[8] D. Kim, S. Sra, and I. Dhillon, "A scalable trust-region algorithm with application to mixed-norm regression," in *27th International Conference on Machine Learning (ICML 2010)*. Omnipress, 2010, pp. 519–526.

[9] V. Roth and B. Fischer, "The group-lasso for generalized linear models: uniqueness of solutions and efficient algorithms," in *Proceedings of the 25th international conference on Machine learning*, 2008, pp. 848–855.

[10] Y. Zhang, N. Zhang, D. Sun, and K.-C. Toh, "An efficient hessian based algorithm for solving large-scale sparse group lasso problems," *Mathematical Programming*, vol. 179, pp. 223–263, 2020.

[11] Y. Hu, "Complex-valued group lasso for tensor autoregressive models," 2016.

[12] L. Li, X. Wang, and G. Wang, "Alternating direction method of multipliers for separable convex optimization of real functions in complex variables," *Mathematical Problems in Engineering*, vol. 2015, 2015.

[13] L. El Ghaoui, V. Viallon, and T. Rabbani, "Safe feature elimination in sparse supervised learning technical report no," *Technical report, UCB/EECS-2010–126, EECS Department, University of California, Berkeley*, 2010.

[14] J. Wang, J. Zhou, P. Wonka, and J. Ye, "Lasso screening rules via dual polytope projection," vol. 26, 2013.

[15] R. Tibshirani, J. Bien, J. Friedman, T. Hastie, N. Simon, J. Taylor, and R. J. Tibshirani, "Strong rules for discarding predictors in lasso-type problems," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 74, no. 2, pp. 245–266, 2012.

[16] A. Bonnefoy, V. Emiya, L. Ralaivola, and R. Gribonval, "A dynamic screening principle for the lasso," in *2014 22nd European Signal Processing Conference (EUSIPCO)*. IEEE, 2014, pp. 6–10.

[17] O. Fercoq, A. Gramfort, and J. Salmon, "Mind the duality gap: safer rules for the lasso," in *International Conference on Machine Learning*. PMLR, 2015, pp. 333–342.

[18] J. Wang and J. Ye, "Two-layer feature reduction for sparse-group lasso via decomposition of convex sets," *Advances in Neural Information Processing Systems*, vol. 27, 2014.

[19] E. Ndiaye, O. Fercoq, A. Gramfort, and J. Salmon, "Gap safe screening rules for sparse-group lasso," in *Proceedings of the 30th International Conference on Neural Information Processing Systems*, ser. NIPS'16. Red Hook, NY, USA: Curran Associates Inc., 2016, p. 388–396.

[20] J. Wang, W. Fan, and J. Ye, "Fused lasso screening rules via the monotonicity of subdifferentials," *IEEE transactions on pattern analysis and machine intelligence*, vol. 37, no. 9, pp. 1806–1820, 2015.

[21] M. Lin, Y. Yuan, D. Sun, and K.-C. Toh, "Adaptive sieving with ppdna: Generating solution paths of exclusive lasso models," *arXiv preprint arXiv:2009.08719*, 2020.

[22] Y. Yuan, T.-H. Chang, D. Sun, and K.-C. Toh, "A dimension reduction technique for large-scale structured sparse optimization problems with application to convex clustering," *SIAM Journal on Optimization*, vol. 32, no. 3, pp. 2294–2318, 2022.

[23] M. Lustig, D. Donoho, and J. M. Pauly, "Sparse mri: The application of compressed sensing for rapid mr imaging," *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, vol. 58, no. 6, pp. 1182–1195, 2007.

[24] L. Anitori, M. Otten, W. Van Rossum, A. Maleki, and R. Baraniuk, "Compressive cfar radar detection," in *2012 IEEE Radar Conference*. IEEE, 2012, pp. 0320–0325.

[25] R. Baraniuk and P. Steeghs, "Compressive radar imaging," in *2007 IEEE radar conference*. IEEE, 2007, pp. 128–133.

[26] M. A. Herman and T. Strohmer, "High-resolution radar via compressed sensing," *IEEE transactions on signal processing*, vol. 57, no. 6, pp. 2275–2284, 2009.

[27] M. Stephane, "A wavelet tour of signal processing," 1999.

[28] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM journal on imaging sciences*, vol. 2, no. 1, pp. 183–202, 2009.

[29] R. Remmert, *Theory of complex functions*. Springer Science & Business Media, 1991, vol. 122.

[30] D. Messerschmitt *et al.*, "Stationary points of a real-valued function of a complex variable," *EECS Department, University of California, Berkeley, Tech. Rep. UCB/EECS-2006-93*, 2006.

[31] J.-J. Moreau, "Proximité et dualité dans un espace hilbertien," *Bulletin de la Société mathématique de France*, vol. 93, pp. 273–299, 1965.

[32] J. Fan and J. Lv, "Sure independence screening for ultrahigh dimensional feature space," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 70, no. 5, pp. 849–911, 2008.

[33] X. Li, D. Sun, and K.-C. Toh, "A highly efficient semismooth newton augmented lagrangian method for solving lasso problems," *SIAM Journal on Optimization*, vol. 28, no. 1, pp. 433–458, 2018.

[34] S. Ren, S. Huang, J. Ye, and X. Qian, "Safe feature screening for generalized lasso," *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 12, pp. 2992–3006, 2017.