



Short communication

# Robust sparse recovery via weakly convex optimization in impulsive noise

Qi Liu<sup>a,\*</sup>, Chengzhu Yang<sup>b</sup>, Yuantao Gu<sup>b</sup>, Hing Cheung So<sup>a,1</sup><sup>a</sup> Department of Electronic Engineering, City University of Hong Kong, Hong Kong, China<sup>b</sup> Department of Electronic Engineering and Tsinghua National Laboratory for Information Science and Technology (TNList), Tsinghua University, Beijing, China

## ARTICLE INFO

## Article history:

Received 11 February 2018

Revised 6 May 2018

Accepted 22 May 2018

Available online 23 May 2018

## Keywords:

Weakly convex optimization

Robust sparse recovery

ADMM

Global optimum

## ABSTRACT

We propose a robust sparse recovery formulation in impulsive noise, where  $\ell_1$  norm as the metric for the residual error and a class of weakly convex functions for inducing sparsity are employed. To solve the corresponding nonconvex and nonsmooth minimization, a slack variable is introduced to guarantee the convexity of the equivalent optimization problem in each block of variables. An efficient algorithm is developed for minimizing the surrogate Lagrangian based on the alternating direction method of multipliers. Model analysis guarantees that this novel robust sparse recovery formulation guarantees to attain the global optimum. Compared with several state-of-the-art algorithms, our method attains better recovery performance in the presence of outliers.

© 2018 Elsevier B.V. All rights reserved.

## 1. Introduction

Sparse recovery (SR) is of great interest in recent years [1], which is a paradigm to acquire sparse or compressible signals at a rate lower than that of the Nyquist sampling. Mathematically, given a known matrix  $\mathbf{A} \in \mathbb{R}^{M \times N}$  with  $M < N$ , the underlying task is to recover a target signal  $\mathbf{x} \in \mathbb{R}^N$  from its undersampled set of noisy observations  $\mathbf{y} \in \mathbb{R}^M$ :

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n} \quad (1)$$

where  $\mathbf{n} = [n_1 \ n_2 \ \dots \ n_M]^T$  is the additive disturbance vector. In sparse signal recovery, the desired sparsity structure can be enforced by either  $\ell_0$  or  $\ell_1$  norm penalties, where  $\ell_2$  norm data fitting model is employed as the metric for the residual error [1]. However, it is well known that least squares-based estimators are highly sensitive to outliers present in the measurement vector, leading to poor recovery. In practical applications, the measurement noise may be of different kinds or combinations. Impulsive noise is a typical representative which can model large errors in observations and has been widely studied in robust statistics [2]. Hence, the  $\ell_2$  norm data fitting model may be inefficient.

Based on the Huber penalty function as the metric for the residual error, a robust sparse recovery formulation is proposed to mitigate the effect of the impulsive noise in the compressed measurements [3]. However, the robust sparse recovery framework is not necessarily restricted to the Huber loss function and indeed many loss functions [4,5] can be used to cater for different noise types. One particular interest is the  $\ell_1$  norm loss function [5–7], which is optimal when the impulsive noise is modeled as a Laplace distribution. Efficient solvers have been presented in [3,5–8] based on the fast iterative shrinkage algorithm (FISTA) [9] and alternating direction method of multipliers (ADMM) [10].

In this work, we adopt weakly convex sparseness measure to constitute the sparsity-inducing penalty, and obtain the following robust SR formulation:

$$\min_{\mathbf{x}} \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_1 + \lambda J(\mathbf{x}) \quad (2)$$

where  $J(\mathbf{x})$  is a sparsity-inducing penalty. Weakly convex penalty functions are also known as semi-convex functions [11], and most commonly used nonconvex penalties are formed by weakly convex sparseness [12–15]. Our work is motivated by the fact that  $\ell_1$  norm loss function has been widely used in designing robust methods due to its simultaneous convexity and robustness. Compared to  $\ell_1$  norm sparsity-inducing, improved results are obtained by replacing  $\ell_1$  norm by a suitably chosen nonconvex regularization [13,16–19], which is advantageous with fewer measurements, faster convergence and better robustness against noise. Since the weakly convex (nonconvex) regularization and  $\ell_1$  norm loss func-

\* Corresponding author.

E-mail addresses: [qiliu47-c@my.cityu.edu.hk](mailto:qiliu47-c@my.cityu.edu.hk) (Q. Liu), [ycz16@mails.tsinghua.edu.cn](mailto:ycz16@mails.tsinghua.edu.cn) (C. Yang), [gyt@tsinghua.edu.cn](mailto:gyt@tsinghua.edu.cn) (Y. Gu), [hcs0@ee.cityu.edu.hk](mailto:hcs0@ee.cityu.edu.hk) (H.C. So).<sup>1</sup> EURASIP member.

tion are adopted, the resultant minimization problem in (2) is non-convex and nonsmooth, which is difficult to solve.

We briefly summarize the contributions of this work as follows: i) By combining the concept of weak convexity with  $\ell_1$  norm loss function, a robust sparse recovery framework for impulsive noise is proposed in (2) and analyzed. Our theoretical results state that if the extended measurement matrix satisfies the restricted isometry property (RIP) with a mild constant, the proposed framework can robustly reconstruct the original signal. ii) A solution based on the ADMM is derived for (2). We show that robust recovery can be implemented readily by decomposing the augmented Lagrangian separately into a series of simpler problems whose solutions approach that of the main problem (2).

The  $\|\mathbf{x}\|_2$ ,  $\|\mathbf{x}\|_1$  and  $\|\mathbf{x}\|_0$  are the  $\ell_2$  norm,  $\ell_1$  norm and  $\ell_0$  norm of a vector  $\mathbf{x}$ , respectively. In particular,  $\|\mathbf{x}\|_0 = \#\{i : x_i \neq 0\}$  counts the nonzero elements of  $\mathbf{x}$ .  $(\cdot)^T$  and  $(\cdot)^{-1}$  stand for the transpose and inverse operators, respectively.  $\text{sign}(\cdot)$  denotes the sign of a quantity with  $\text{sign}(0) = 0$ .  $\mathbf{I}$  represents an identity matrix of appropriate dimensions.

## 2. Algorithm development

We now tackle the problem of robust recovery minimization in (2), regularized by a weakly convex (nonconvex) function  $J(\mathbf{x})$ . The weakly convex penalty  $J(\mathbf{x})$  in (2) is defined as:

$$J(\mathbf{x}) := \sum_{i=1}^N F(x_i) \quad (3)$$

where  $F(x)$  is a weakly convex sparseness measure. Note that  $F(x)$  is weakly convex if and only if there exists a convex function  $H(x) = F(x) - \xi x^2$  when  $\xi < 0$ . A class of sparsity-inducing penalties has been introduced in [13], satisfying the following properties.

- Definition 1.** (a)  $F(0) = 0$ ,  $F(\cdot)$  is even and not identically zero;  
 (b)  $F(\cdot)$  is nondecreasing on  $[0, +\infty)$ ;  
 (c) The function  $x \rightarrow F(x)/x$  is nonincreasing on  $[0, +\infty)$ ;  
 (d)  $F(\cdot)$  is weakly convex on  $[0, +\infty)$ .

From Lemma 1.1 in [13],  $F(x)/x \rightarrow \alpha$  as  $x \rightarrow 0^+$  for  $\alpha > 0$ . Hence, according to Definition 1, we let  $\beta \triangleq -\xi/\alpha$  to characterize the non-convexity of  $F(x)$  and  $J(x)$ , where  $-\xi$  divided by  $\alpha$  is to remove the scaling effect on the penalty. For example, the weakly convex sparseness function in (3) may be chosen as [20]:

$$F(x) = (|x| - \beta x^2) \mathbf{1}_{|x| \leq \frac{1}{2\beta}}(x) + \frac{1}{4\beta} \mathbf{1}_{|x| > \frac{1}{2\beta}}(x) \quad (4)$$

where  $\mathbf{1}_P(x)$  is the indicator function with value 1 when the argument satisfying  $P$ , and 0 otherwise.  $F(x)$  is a continuous piecewise quadratic function, which is easily verified to satisfy Definition 1 when  $\alpha = 1$  and  $\beta = -\xi$ . The weakly convex sparseness function in (4) is also known as the minimax-concave penalty function [21]. Note that when  $\xi = 0$ , the robust recovery formulation in (2) becomes convex.

Next, we solve (2) regularized by any weakly convex function  $J(\mathbf{x})$  satisfying Definition 1. It is worth noting that the main method to be proposed is similar to [22], but there are significant differences between them. The main challenge is that the variables are coupled through  $\mathbf{A}$ . This makes it rather difficult when the extra constraint with nonsmooth  $\ell_1$  loss function is introduced. With the use of operator splitting [23], we separate the nonsmooth weakly convex term from the  $\ell_1$  norm loss term. Hence, problem (2) can be rewritten equivalently by introducing an auxiliary variable vector  $\mathbf{z} \in \mathbb{R}^M$ , which is tied to the original variable via an affine constraint:

$$\min_{\mathbf{x}} \lambda J(\mathbf{x}) + \|\mathbf{z}\|_1 \quad \text{s.t. } \mathbf{A}\mathbf{x} - \mathbf{z} = \mathbf{y}. \quad (5)$$

Nevertheless, the objective function is nonconvex with respect to  $\mathbf{x}$ . Since  $\lambda > 0$ ,  $J$  is nonconvex, and it may be nondifferentiable, which indicates that an optimization problem with  $J$  in the objective function can be hard to solve. In our study, a slack variable  $\mathbf{w}$  is introduced to solve (5). Then, (5) is rewritten as:

$$\min_{\mathbf{x}} \lambda J(\mathbf{x}) + \|\mathbf{z}\|_1 \quad \text{s.t. } [\mathbf{A} \quad -\mathbf{I}]\mathbf{w} = \mathbf{y}, [\mathbf{x}^T \quad \mathbf{z}^T]^T = \mathbf{w}. \quad (6)$$

For this type of regularized objective function (6), ADMM [10] considers the following augmented Lagrangian, given by

$$\begin{aligned} L(\mathbf{x}, \mathbf{z}, \mathbf{w}, \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2) &= \lambda J(\mathbf{x}) + \|\mathbf{z}\|_1 \\ &+ \boldsymbol{\gamma}_1^T ([\mathbf{A} \quad -\mathbf{I}]\mathbf{w} - \mathbf{y}) + \frac{\rho}{2} \|\mathbf{A} - \mathbf{I}\mathbf{w} - \mathbf{y}\|_2^2 \\ &+ \boldsymbol{\gamma}_2^T ([\mathbf{x}^T \quad \mathbf{z}^T]^T - \mathbf{w}) + \frac{\rho}{2} \|\mathbf{x}^T \quad \mathbf{z}^T]^T - \mathbf{w}\|_2^2 \end{aligned} \quad (7)$$

where  $\boldsymbol{\gamma}_1 \in \mathbb{R}^M$  and  $\boldsymbol{\gamma}_2 \in \mathbb{R}^{M+N}$  are dual variable vectors and  $\rho > 0$  is the penalty parameter. The fact that by adding a quadratic term  $\frac{\rho}{2} \|\mathbf{x}^T \quad \mathbf{z}^T]^T - \mathbf{w}\|_2^2$ , the function  $J$  becomes convex, which leads to some favorable properties. For example, the separable function defined in (3) can be expressed as the Moreau envelope [21,24] of the weakly convex function, i.e.,

$$J(\mathbf{x}) = \min_{\mathbf{v} \in \mathbb{R}^N} \left\{ J(\mathbf{v}) + \frac{1}{2} \|\mathbf{x} - \mathbf{v}\|_2^2 \right\}. \quad (8)$$

If  $\beta$  is small enough such that  $\beta\xi < \frac{1}{2}$ , then the objective function in (8) is strongly convex, and the minimizer is unique. The proximal operators of some weakly convex functions are well defined and have closed-form expressions which are relatively easy to compute [13]. Denote  $\mathbf{w}^T = [\mathbf{w}_1^T \quad \mathbf{w}_2^T]$  and  $\boldsymbol{\gamma}_2^T = [\boldsymbol{\gamma}_{21}^T \quad \boldsymbol{\gamma}_{22}^T]$ . The strategy for minimizing the augmented Lagrangian is iteratively updating of the primal and dual variables. That is, ADMM applied to (7) consists of the following iterative steps:

$$\begin{aligned} \mathbf{x}^{t+1} &= \arg \min_{\mathbf{x}} \left\{ \lambda J(\mathbf{x}) + \frac{\rho}{2} \|\mathbf{x} - \mathbf{w}_1^t + \frac{\boldsymbol{\gamma}_{21}^t}{\rho}\|_2^2 \right\} \\ &= \text{prox}_{\frac{\lambda}{\rho} J(\cdot)} \left( \mathbf{w}_1^t - \frac{\boldsymbol{\gamma}_{21}^t}{\rho} \right) \end{aligned} \quad (9)$$

$$\begin{aligned} \mathbf{z}^{t+1} &= \arg \min_{\mathbf{z}} \left\{ \|\mathbf{z}\|_1 + \frac{\rho}{2} \|\mathbf{z} - \mathbf{w}_2^t + \frac{\boldsymbol{\gamma}_{22}^t}{\rho}\|_2^2 \right\} \\ &= \text{soft} \left( \mathbf{w}_2^t - \frac{\boldsymbol{\gamma}_{22}^t}{\rho}, \frac{1}{\rho} \right) \end{aligned} \quad (10)$$

$$\begin{aligned} \mathbf{w}^{t+1} &= \arg \min_{\mathbf{w}} \left\{ \frac{\rho}{2} \|\mathbf{w} - [(\mathbf{x}^{t+1})^T \quad (\mathbf{z}^{t+1})^T]^T\|_2^2 \right. \\ &\quad \left. + \mathbf{w}^T ([\mathbf{A} \quad -\mathbf{I}]^T (\boldsymbol{\gamma}_1^t - \rho \mathbf{y}) - \boldsymbol{\gamma}_2^t) + \frac{\rho}{2} \|\mathbf{A} - \mathbf{I}\mathbf{w}\|_2^2 \right\} \end{aligned} \quad (11)$$

$$\boldsymbol{\gamma}_1^{t+1} = \boldsymbol{\gamma}_1^t + \rho([\mathbf{A} \quad -\mathbf{I}]\mathbf{w}^{t+1} - \mathbf{y}) \quad (12)$$

$$\boldsymbol{\gamma}_2^{t+1} = \boldsymbol{\gamma}_2^t + \rho([\mathbf{x}^{t+1})^T \quad (\mathbf{z}^{t+1})^T]^T - \mathbf{w}^{t+1}). \quad (13)$$

in which  $\text{prox}(\cdot)$  and  $\text{soft}(\cdot)$  are proximal and soft-thresholding shrinkage operators, respectively.

In the process of the  $\mathbf{x}$ -minimization (9), for  $F(x)$  in (4), when  $\beta < 1/(2\zeta)$ , its proximal operator is [13]

$$\text{prox}_{\zeta F}(v) = \frac{v - \zeta \text{sign}(v)}{1 - 2\zeta\beta} \mathbf{1}_{\zeta \leq |v| \leq \frac{1}{2\beta}}(v) + v \mathbf{1}_{|v| > \frac{1}{2\beta}}(v). \quad (14)$$

The  $\mathbf{w}$ -minimization (11) is a convex formulation and has a closed-form solution, but it is not computationally efficient for large-scale data due to the inverse computation of the multiplication of the sensing matrix, i.e.,  $(\mathbf{I} + \rho[\mathbf{A} \quad -\mathbf{I}]^T[\mathbf{A} \quad -\mathbf{I}])^{-1}$ , which can

be approximated by a design matrix  $\mathbf{B}$ . Let  $\tilde{\mathbf{A}} := (\mathbf{I} + \rho[\mathbf{A} - \mathbf{I}]^T[\mathbf{A} - \mathbf{I}])$ . A well-known iterative method [13] to compute  $\mathbf{B}$  is

$$\mathbf{B}_0 = \frac{1}{2\delta} \tilde{\mathbf{A}}^T, \quad (15)$$

$$\mathbf{B}_k = \mathbf{B}_{k-1}(2\mathbf{I} - \tilde{\mathbf{A}}\mathbf{B}_{k-1}), \quad (16)$$

where  $0 < \delta < 2/\|\tilde{\mathbf{A}}\tilde{\mathbf{A}}^T\|_1$ . When  $\mathbf{B}_k$  satisfies  $\|\mathbf{A}^{-1} - \mathbf{B}_k\|_F^2 \leq \varepsilon$  where  $\varepsilon$  is a small positive value, a nice approximate analytic solution of (11) can be easily obtained without evaluating the matrix inverse:

$$\mathbf{w}^{t+1} = \mathbf{B} \left( [(\mathbf{x}^{t+1})^T \ (\mathbf{z}^{t+1})^T]^T - \frac{1}{\rho}([\mathbf{A} - \mathbf{I}](\boldsymbol{\gamma}^t - \rho\boldsymbol{\gamma}) - \boldsymbol{\gamma}^t) \right). \quad (17)$$

### 3. Model analysis

In this section, we prove that when a class of sparsity-inducing penalties satisfies the Definition 1 in (2), global minimum of the model can be found. Recovery guarantees based on the RIP of the extended matrix  $[\mathbf{A} \ -\mathbf{I}]$  have been reported in [25].

For any matrix  $\hat{\mathbf{A}} = [\mathbf{A}\mathbf{O} \ -\mathbf{I}]$ , define the RIP-constant  $\delta_{k_1, k_2}$  by the infimum value of  $\delta$  such that

$$\begin{aligned} (1 - \delta)(\|\mathbf{x}\|_2^2 + \|\mathbf{z}\|_2^2) &\leq \|\hat{\mathbf{A}}[\mathbf{x}^T \ \mathbf{z}^T]^T\|_2^2 \\ &\leq (1 + \delta)(\|\mathbf{x}\|_2^2 + \|\mathbf{z}\|_2^2) \end{aligned} \quad (18)$$

holds for any  $\mathbf{x}$  with  $|\text{supp}(\mathbf{x})| \leq k_1$  and  $\mathbf{z}$  with  $|\text{supp}(\mathbf{z})| \leq k_2$ .

**Lemma 1.** For any  $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^N$  and  $\mathbf{z}_1, \mathbf{z}_2 \in \mathbb{R}^M$  such that  $\text{supp}(\mathbf{x}_1) \cap \text{supp}(\mathbf{x}_2) = \emptyset$ ,  $\text{supp}(\mathbf{z}_1) \cap \text{supp}(\mathbf{z}_2) = \emptyset$ , and  $|\text{supp}(\mathbf{x}_1)| + |\text{supp}(\mathbf{x}_2)| \leq k_1$ ,  $|\text{supp}(\mathbf{z}_1)| + |\text{supp}(\mathbf{z}_2)| \leq k_2$ , there exists  $|t| \leq \delta_{k_1, k_2}$ , such that

$$\|\hat{\mathbf{A}}[\mathbf{x}_1^T \ \mathbf{z}_1^T]^T\|_2^2 = (1 + t)\|[\mathbf{x}_1^T \ \mathbf{z}_1^T]^T\|_2^2, \quad (19)$$

and

$$\begin{aligned} &| \langle \hat{\mathbf{A}}[\mathbf{x}_1^T \ \mathbf{z}_1^T]^T, \hat{\mathbf{A}}[\mathbf{x}_2^T \ \mathbf{z}_2^T]^T \rangle | \\ &\leq \sqrt{\delta_{2k_1, 2k_2}^2 - t^2} \sqrt{\|\mathbf{x}_1\|_2^2 + \|\mathbf{z}_1\|_2^2} \sqrt{\|\mathbf{x}_2\|_2^2 + \|\mathbf{z}_2\|_2^2}. \end{aligned} \quad (20)$$

**Proof.** It is similar to the proof of Theorem 6.14 of [26] and omitted here to save space.  $\square$

**Theorem 1.** Suppose that  $\frac{1}{\gamma} \sqrt{\frac{k_2}{k_1}} \leq \lambda \leq \gamma \sqrt{\frac{k_2}{k_1}}$  ( $\gamma \geq 1$ ) and the extended sensing matrix  $\hat{\mathbf{A}}$  satisfies the RIP of order  $\{2k_1, 2k_2\}$  with

$$\delta_{2k_1, 2k_2} \leq \frac{1}{\sqrt{1 + 2(\gamma^2 + \frac{1}{4})^2}}. \quad (21)$$

Then for  $\mathbf{y} = \mathbf{A}\mathbf{x} - \mathbf{z}$ , the solution of (5) denoted by  $\{\mathbf{x}^*, \mathbf{z}^*\}$ , obeys:

$$J(\mathbf{x}^* - \mathbf{x}) \leq C_1 J(\mathbf{x}_{-k_1}) + C_2 \|\mathbf{z}_{-k_2}\|_1, \quad (22)$$

where the constants  $C_1$  and  $C_2$  depend on  $\delta_{2k_1, 2k_2}$ , and  $\mathbf{x}_{-k_1}$  ( $\mathbf{z}_{-k_2}$ ) denotes the vector setting the  $k_1$  ( $k_2$ ) largest absolute entries of  $\mathbf{x}$  ( $\mathbf{z}$ ) to be 0 and keeping others.

**Proof.** Let  $\mathbf{u} = \mathbf{x} - \mathbf{x}^*$  and  $\mathbf{v} = \mathbf{z} - \mathbf{z}^*$ . By the definitions of  $\mathbf{x}^*$  and  $\mathbf{z}^*$ , it is obvious that

$$\hat{\mathbf{A}}[\mathbf{u}; \mathbf{v}] = \mathbf{0}. \quad (23)$$

For notational simplicity, stacking of two vectors  $\mathbf{u}$  and  $\mathbf{v}$ ,  $[\mathbf{u}^T \ \mathbf{v}^T]^T$ , is denoted by  $[\mathbf{u}; \mathbf{v}]$ . Consider an index set  $S_0$  of  $k_1$  largest absolute entries of  $\mathbf{u}$ . At first, denote  $\bar{S}_0 = S_1 \cup S_2 \cup \dots$ , where

- $S_1$ : index set of  $k_1$  largest absolute entries of  $\mathbf{u}$  in  $\bar{S}_0$ ,
- $S_2$ : index set of  $k_1$  largest absolute entries of  $\mathbf{u}$  in  $\bar{S}_0 \cup \bar{S}_1$ , etc.

In a similar manner, index sets  $T_0, T_1, T_2, \dots$  are defined for  $\mathbf{v}$ . We observe that

$$\begin{aligned} \|\hat{\mathbf{A}}[\mathbf{u}_{S_0}; \mathbf{v}_{T_0}]\|_2^2 &= \langle \hat{\mathbf{A}}[\mathbf{u}_{S_0}; \mathbf{v}_{T_0}], \hat{\mathbf{A}}[\mathbf{u}; \mathbf{v}] \rangle \\ &\quad - \langle \hat{\mathbf{A}}[\mathbf{u}_{S_0}; \mathbf{v}_{T_0}], \sum_{p \geq 1} \mathbf{A}\mathbf{u}_{S_p} - \sum_{q \geq 1} \mathbf{v}_{T_q} \rangle \\ &= \langle \hat{\mathbf{A}}[\mathbf{u}_{S_0}; \mathbf{v}_{T_0}], \sum_{p \geq 1} \mathbf{A}\mathbf{u}_{S_p} - \sum_{q \geq 1} \mathbf{v}_{T_q} \rangle. \end{aligned} \quad (24)$$

On the one hand, according to Lemma 1, the left hand side satisfies,

$$\|\hat{\mathbf{A}}[\mathbf{u}_{S_0}; \mathbf{v}_{T_0}]\|_2^2 \geq (1 + t)\|[\mathbf{u}_{S_0}; \mathbf{v}_{T_0}]\|_2^2. \quad (25)$$

Recalling Lemma 1, we have

$$\langle \hat{\mathbf{A}}[\mathbf{u}_{S_0}; \mathbf{v}_{T_0}], \mathbf{A}\mathbf{u}_{S_p} \rangle \leq \sqrt{\delta_{2k_1, k_2}^2 - t^2} \|[\mathbf{u}_{S_0}; \mathbf{v}_{T_0}]\|_2 \|\mathbf{u}_{S_p}\|_2, \quad (26)$$

$$\langle \hat{\mathbf{A}}[\mathbf{u}_{S_0}; \mathbf{v}_{T_0}], \mathbf{v}_{T_q} \rangle \leq \sqrt{\delta_{k_1, 2k_2}^2 - t^2} \|[\mathbf{u}_{S_0}; \mathbf{v}_{T_0}]\|_2 \|\mathbf{v}_{T_q}\|_2, \quad (27)$$

Inserting (26) and (27) into the left hand side of (24), we have

$$\begin{aligned} &\langle \hat{\mathbf{A}}[\mathbf{u}_{S_0}; \mathbf{v}_{T_0}], \sum_{p \geq 1} \mathbf{A}\mathbf{u}_{S_p} - \sum_{q \geq 1} \mathbf{v}_{T_q} \rangle \\ &\leq \sqrt{\delta_{2k_1, 2k_2}^2 - t^2} \|[\mathbf{u}_{S_0}; \mathbf{v}_{T_0}]\|_2 \left( \sum_{p \geq 1} \|\mathbf{u}_{S_p}\|_2 + \sum_{q \geq 1} \|\mathbf{v}_{T_q}\|_2 \right). \end{aligned} \quad (28)$$

Combining (25) and (28) into (24), and applying Lemma 1 again, we get:

$$(1 + t)\|[\mathbf{u}_{S_0}; \mathbf{v}_{T_0}]\|_2 \leq \sqrt{\delta_{2k_1, 2k_2}^2 - t^2} \left( \sum_{p \geq 1} \|\mathbf{u}_{S_p}\|_2 + \sum_{q \geq 1} \|\mathbf{v}_{T_q}\|_2 \right). \quad (29)$$

For each  $p \geq 1$ , the smallest and largest absolute entries of  $\mathbf{u}_{S_p}$  are denoted by  $u_p^-$  and  $u_p^+$ , respectively. Similarly,  $v_q^+$  and  $v_q^-$  are defined for  $\mathbf{v}_{T_q}$  with  $q \geq 1$ . Using Lemma 6.14 of [26], we obtain:

$$\begin{aligned} \sum_{p \geq 1} \|\mathbf{u}_{S_p}\|_2 &\leq \sum_{p \geq 1} \left( \frac{1}{\sqrt{k_1}} \|\mathbf{u}_{S_p}\|_1 + \frac{\sqrt{k_1}}{4} (u_p^+ - u_p^-) \right) \\ &\leq \frac{1}{\sqrt{k_1}} \|\mathbf{u}_{S_0}\|_1 + \frac{1}{4\sqrt{k_1}} \|\mathbf{u}_{S_0}\|_1, \end{aligned} \quad (30)$$

$$\begin{aligned} \sum_{q \geq 1} \|\mathbf{v}_{T_q}\|_2 &\leq \sum_{q \geq 1} \left( \frac{1}{\sqrt{k_2}} \|\mathbf{v}_{T_q}\|_1 + \frac{\sqrt{k_2}}{4} (v_q^+ - v_q^-) \right) \\ &\leq \frac{1}{\sqrt{k_2}} \|\mathbf{v}_{T_0}\|_1 + \frac{1}{4\sqrt{k_2}} \|\mathbf{v}_{T_0}\|_1. \end{aligned} \quad (31)$$

Inserting (30) and (31) into (29), and noticing

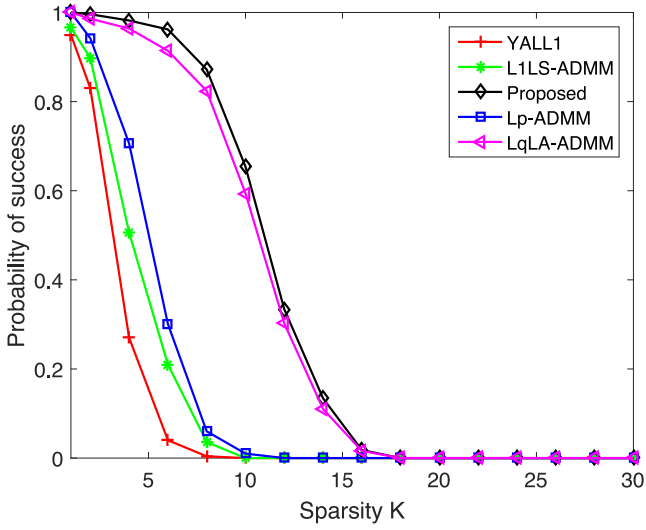
$$\|[\mathbf{u}_{S_0}; \mathbf{v}_{T_0}]\|_2 \geq \frac{1}{\sqrt{2k_1}} \|\mathbf{u}_{S_0}\|_1 + \frac{1}{\sqrt{2k_2}} \|\mathbf{v}_{T_0}\|_1, \quad (32)$$

$$\frac{\delta_{2k_1, 2k_2}}{\sqrt{1 - \delta_{2k_1, 2k_2}^2}} = \max_t \frac{\sqrt{\delta_{2k_1, 2k_2}^2 - t^2}}{1 + t}, \quad \text{s.t. } |t| \leq \delta_{2k_1, 2k_2}.$$

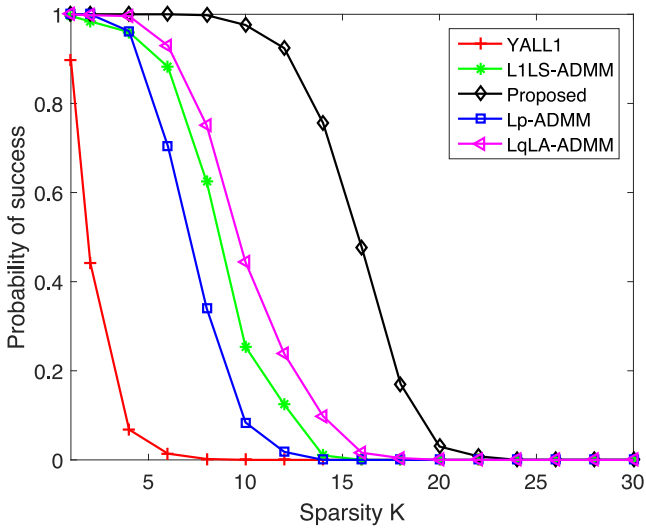
$$\delta_{2k_1, k_2} \leq \delta_{2k_1, 2k_2}, \quad \delta_{k_1, 2k_2} \leq \delta_{2k_1, 2k_2}, \quad (33)$$

we have:

$$\begin{aligned} \frac{1}{\sqrt{2k_1}} \|\mathbf{u}_{S_0}\|_1 + \frac{1}{\sqrt{2k_2}} \|\mathbf{v}_{T_0}\|_1 &\leq \frac{\sqrt{\delta_{2k_1, 2k_2}^2 - t^2}}{1 + t} \left( \frac{1}{\sqrt{k_1}} \|\mathbf{u}_{S_0}\|_1 \right. \\ &\quad \left. + \frac{1}{4\sqrt{k_1}} \|\mathbf{u}_{S_0}\|_1 + \frac{1}{\sqrt{k_2}} \|\mathbf{v}_{T_0}\|_1 + \frac{1}{4\sqrt{k_2}} \|\mathbf{v}_{T_0}\|_1 \right) \end{aligned}$$



(a) GMM noise



(b)  $S\alpha S$  noise

**Fig. 1.** Probability of success versus sparsity.

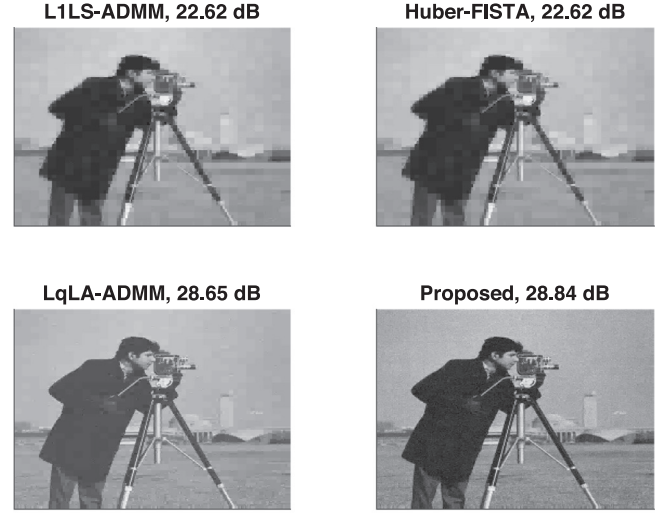
$$\begin{aligned} &\Rightarrow \frac{\sqrt{k_2}}{\sqrt{k_1}} \|\mathbf{u}_{S_0}\|_1 + \|\mathbf{v}_{T_0}\|_1 \\ &\leq \frac{\delta_{2k_1, 2k_2}}{\sqrt{(1 - \delta_{2k_1, 2k_2}^2)/2 - \delta_{2k_1, 2k_2}/4}} \left( \frac{\sqrt{k_2}}{\sqrt{k_1}} \|\mathbf{u}_{\bar{S}_0}\|_1 + \|\mathbf{v}_{\bar{T}_0}\|_1 \right) \end{aligned} \quad (34)$$

where  $\bar{T}_0$  and  $\bar{S}_0$  denote the complementary sets of  $T_0$  and  $S_0$ , respectively. By the assumption of  $\lambda$ ,

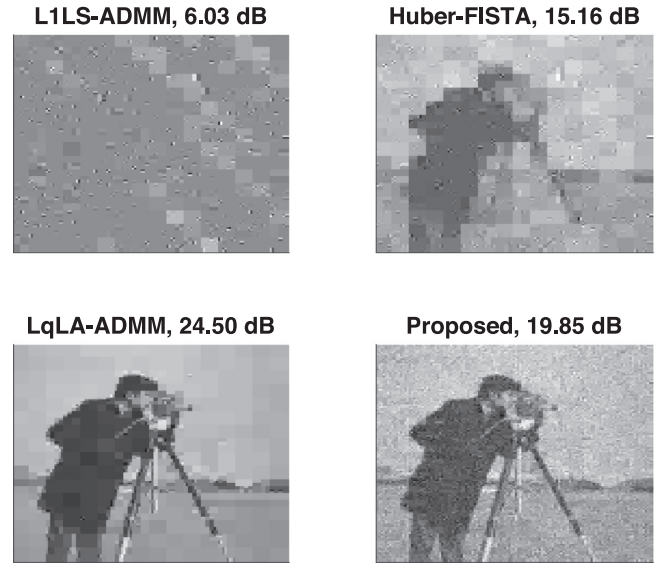
$$\begin{aligned} \frac{\sqrt{k_2}}{\sqrt{k_1}} \|\mathbf{u}_{S_0}\|_1 + \|\mathbf{v}_{T_0}\|_1 &\geq \frac{1}{\gamma} (\lambda \|\mathbf{u}_{S_0}\|_1 + \|\mathbf{v}_{T_0}\|_1), \\ \frac{\sqrt{k_2}}{\sqrt{k_1}} \|\mathbf{u}_{\bar{S}_0}\|_1 + \|\mathbf{v}_{\bar{T}_0}\|_1 &\leq \gamma (\lambda \|\mathbf{u}_{\bar{S}_0}\|_1 + \|\mathbf{v}_{\bar{T}_0}\|_1). \end{aligned} \quad (35)$$

Inserting (35) into (34) yields:

$$\lambda \|\mathbf{u}_{S_0}\|_1 + \|\mathbf{v}_{T_0}\|_1$$



(a) GMM noise



(b)  $S\alpha S$  noise

**Fig. 2.** Recovery of Cameraman image.

$$\leq \frac{\gamma^2 \delta_{2k_1, 2k_2}}{\sqrt{(1 - \delta_{2k_1, 2k_2}^2)/2 - \delta_{2k_1, 2k_2}/4}} (\lambda \|\mathbf{u}_{\bar{S}_0}\|_1 + \|\mathbf{v}_{\bar{T}_0}\|_1). \quad (36)$$

Let  $\rho := \frac{\gamma^2 \delta_{2k_1, 2k_2}}{\sqrt{(1 - \delta_{2k_1, 2k_2}^2)/2 - \delta_{2k_1, 2k_2}/4}}$ . Following the last inequality and observing that  $\frac{\lambda J(\mathbf{u}_{S_0}) + \|\mathbf{v}_{T_0}\|_1}{\lambda J(\mathbf{u}_{\bar{S}_0}) + \|\mathbf{v}_{\bar{T}_0}\|_1}$  is a nondecreasing function of  $\beta$ , we have

$$\frac{\lambda J(\mathbf{u}_{S_0}) + \|\mathbf{v}_{T_0}\|_1}{\lambda J(\mathbf{u}_{\bar{S}_0}) + \|\mathbf{v}_{\bar{T}_0}\|_1} \leq \lim_{\beta \rightarrow 0} \frac{\lambda J(\mathbf{u}_{S_0}) + \|\mathbf{v}_{T_0}\|_1}{\lambda J(\mathbf{u}_{\bar{S}_0}) + \|\mathbf{v}_{\bar{T}_0}\|_1}$$

$$= \frac{\lambda \|\mathbf{u}_{S_0}\|_1 + \|\mathbf{v}_{T_0}\|_1}{\lambda \|\mathbf{u}_{S_0^-}\| + \|\mathbf{v}_{T_0^-}\|_1} \leq \rho,$$

$$\Rightarrow \lambda J(\mathbf{u}_{S_0}) + \|\mathbf{v}_{T_0}\|_1 \leq \rho (\lambda J(\mathbf{u}_{S_0^-}) + \|\mathbf{v}_{T_0^-}\|_1), \quad (37)$$

$$\Leftrightarrow \lambda J(\mathbf{u}) + \|\mathbf{v}\|_1 \leq (1 + \rho) (\lambda J(\mathbf{u}_{S_0^-}) + \|\mathbf{v}_{T_0^-}\|_1). \quad (38)$$

Let  $S$  denote the index set of  $k_1$  largest absolute entries of  $\mathbf{x}$  and  $T$  denote the index set of  $k_2$  largest absolute entries of  $\mathbf{z}$ . Because of the definitions of  $S_0$ ,  $T_0$  and  $S$ ,  $T$ , (37) and (38) still hold with  $S$  in place of  $S_0$  and  $T$  in place of  $T_0$ , respectively. By Lemma 4.15 of [26], it is easy to prove

$$J(\mathbf{u}_{S^-}) \leq J(\mathbf{x}^*) - J(\mathbf{x}) + J(\mathbf{u}_S)_1 + 2J(\mathbf{x}_S), \quad (39)$$

$$\|\mathbf{v}_T\|_1 \leq \|\mathbf{z}^*\|_1 - \|\mathbf{z}\|_1 + \|\mathbf{v}_T\|_1 + 2\|\mathbf{z}_T\|_1. \quad (40)$$

Substituting (39) and (40) into (38) and employing (37), we obtain:

$$\lambda J(\mathbf{u}) + \|\mathbf{v}\|_1 \leq \frac{1 + \rho}{1 - \rho} ((\lambda J(\mathbf{x}^*) + \|\mathbf{z}^*\|_1) - (\lambda J(\mathbf{x}) + \|\mathbf{z}\|_1) + 2J(\mathbf{x}_S) + 2\|\mathbf{z}_T\|_1) \quad (41)$$

where  $\rho \leq 1$ , which requires

$$\delta_{2k_1, 2k_2} \leq \frac{1}{\sqrt{1 + 2(\gamma^2 + \frac{1}{4})^2}}. \quad (42)$$

By the definitions of  $\mathbf{x}^*$  and  $\mathbf{z}^*$ , (41) gives

$$\lambda J(\mathbf{u}) + \|\mathbf{v}\|_1 \leq 2 \frac{1 + \rho}{1 - \rho} (J(\mathbf{x}_S) + \|\mathbf{z}_T\|_1),$$

$$\Rightarrow J(\mathbf{u}) \leq \frac{2(1 + \rho)}{\lambda(1 - \rho)} (J(\mathbf{x}_S) + \|\mathbf{z}_T\|_1), \quad (43)$$

which concludes the proof.  $\square$

**Remark 1.** If  $\gamma = 1$ ,  $\delta_{2k_1, 2k_2} \leq \sqrt{\frac{8}{33}} \approx 0.492$ ; and if  $\gamma = 2$ ,  $\delta_{2k_1, 2k_2} \leq \sqrt{\frac{8}{293}} \approx 0.165$ , which is better than the bound of [27], which claims  $\delta_{2k_1, 2k_2} \leq \frac{1}{18} \approx 0.056$  for  $\gamma = 2$ .

**Remark 2.** If the assumptions of Theorem 1 hold,  $\mathbf{x}$  and  $\mathbf{z}$  are  $k_1$ -sparse and  $k_2$ -sparse, respectively, then for any  $y \leq 1$ , it is clear that  $\mathbf{x}^* = \mathbf{x}$  holds because  $J(\mathbf{u}) = 0 \Leftrightarrow \mathbf{u} = \mathbf{0}$  and  $J(\mathbf{x}_{-k_1}) = \|\mathbf{z}_{-k_2}\|_1 = 0$ , namely, the sparse solution  $\{\mathbf{x}, \mathbf{z}\}$  is the global minimum of (5).

**Remark 3.** From Theorem 1,  $J(\mathbf{x}^* - \mathbf{x})$ , a function of reconstruction error  $\mathbf{x}^* - \mathbf{x}$ , is bounded by the tails of the signal and noise. Actually, if  $\|\mathbf{x}^* - \mathbf{x}\|_2 \leq \mathbf{r}$ , it is easy to deduce that

$$\|\mathbf{x}^* - \mathbf{x}\|_2 \leq \frac{\mathbf{r}}{\mathbf{F}(\mathbf{r})} (\mathbf{C}_1 J(\mathbf{x}_{-k_1}) + \mathbf{C}_2 \|\mathbf{z}_{-k_2}\|_1), \quad (44)$$

which means the reconstruction error is bounded if the distance between the reconstructed and true signals can be roughly estimated.

**Remark 4.** Since the convergence analysis of ADMM with three blocks is very hard and few references are helpful in our case, which remains an open problem for future research. Although the convergence of the proposed method is not proved theoretically, extensive simulation results indicate convergence in empirical sense.

## 4. Numerical examples

In this section, two widely used probability density functions for impulsive noise, i.e., symmetric  $\alpha$ -stable ( $S\alpha S$ ) ( $\varphi(\omega) = \exp(-\gamma^\alpha |\omega|^\alpha)$ ,  $\alpha = 1$ ,  $\gamma = 10^{-4}$ ) and Gaussian mixture model (GMM) ( $p_n(n) = \sum_{i=1}^2 \frac{c_i}{\pi \sigma_i^2} \exp\left(-\frac{|n|^2}{\sigma_i^2}\right)$ ,  $\sigma_2^2 = 1000\sigma_1^2$ ,  $c_2 = 0.1$  and signal-to-noise ratio (SNR) = 30 dB), are taken to model the additive noise. Simulations are implemented to evaluate the robustness of the proposed method for sparse recovery.

In Fig. 1, compared with the YALL1 [5], L1LS-ADMM [28], Lp-ADMM [6] and LqLA-ADMM [7], we evaluate the robustness of the proposed method using simulated sparse signal in impulsive noise. The simulated  $K$ -sparse signal is constructed as follows: the positions of  $K$  nonzeros are uniformly and randomly chosen and the amplitude of each nonzero entry is generated by Gaussian distribution.  $\mathbf{A}$  is of dimensions  $40 \times 200$  and randomly generated with entries drawn from the standard Gaussian distribution, and each entry follows  $\mathcal{N}(0, 1/N)$ . If the relative reconstruction error is less than  $10^{-2}$ , the recovery is regarded as a success.  $\rho = 4$ ,  $\beta = 10^{-0.3}$  and  $\lambda$  in each method is chosen such that the best performance in terms of relative recovery error [7,29] is attained. As can be seen, the proposed method guarantees successful recovery for more sparse signal than the other algorithms.

In Fig. 2, performance of the proposed method is compared with the L1LS-ADMM [28], Huber-FISTA [3] and LqLA-ADMM [7] in image recovery experiments. We consider  $\mathbf{A}$  to be a partial discrete cosine transform (DCT) matrix via randomly selecting  $M$  out of  $N$  rows of the full counterpart, where  $M = 0.5N$ . It has  $256 \times 256 = 65536$  pixels. Accordingly, the ground-truth sparse  $\mathbf{x}$  contains the partial DCT coefficients, which is obtained from the 2096 DCT coefficients with largest magnitudes using  $\mathbf{A}$  and the real-world image, i.e., Cameraman. And then we recover the coefficients from the observations corrupted by impulsive noise, and compare the recovered image with the original one. Note that the coefficients of a real-world image are not strictly sparse but rather approximately follow an exponential decay, which is referred to as compressible [30]. The proposed method significantly outperforms the other algorithms (except LqLA-ADMM in  $S\alpha S$  noise) with higher peak SNR in recovering the real-world image.

## 5. Conclusion

By combining the concept of weak convexity with  $\ell_1$  norm loss function, a robust sparse recovery framework for impulsive noise is proposed in (2) and theoretically analyzed. In particular, if the extended measurement matrix satisfies the RIP with a mild constant, our devised framework can robustly reconstruct the original signal.

## Acknowledgments

This work was supported by a grant from the NSFC/RGC Joint Research Scheme sponsored by the National Natural Science Foundation of China and the Research Grants Council of Hong Kong (61531166005, N\_CityU 104/15).

## References

- [1] E.J. Candès, J.K. Romberg, T. Tao, Stable signal recovery from incomplete and inaccurate measurements, *Commun. Pure Appl. Math.* 59 (8) (2006) 1207–1223.
- [2] A.M. Zoubir, V. Koivunen, Y. Chakhchoukh, M. Muma, Robust estimation in signal processing: a tutorial-style treatment of fundamental concepts, *IEEE Signal Process. Mag.* 29 (4) (2012) 61–80.
- [3] D.S. Pham, S. Venkatesh, Efficient algorithms for robust recovery of images from compressed data, *IEEE Trans. Image Process.* 22 (12) (2013) 4724–4737.
- [4] W.-J. Zeng, H.C. So, X. Jiang, Outlier-robust greedy pursuit algorithms in  $\ell_p$ -space for sparse approximation, *IEEE Trans. Signal Process.* 64 (1) (2016) 60–75.



- [5] J. Yang, Y. Zhang, Alternating direction algorithms for  $\ell_1$ -problems in compressive sensing, *SIAM J. Scient. Comput.* 33 (1) (2011) 250–278.
- [6] F. Wen, P. Liu, Y. Liu, R.C. Qiu, W. Yu, Robust sparse recovery in impulsive noise via  $\ell_p$ - $\ell_1$  optimization, *IEEE Trans. Signal Process.* 65 (1) (2017) 105–118.
- [7] F. Wen, L. Pei, Y. Yang, W. Yu, P. Liu, Efficient and robust recovery of sparse signal and image using generalized nonconvex regularization, *IEEE Trans. Comput. Imaging* 3 (4) (2017) 566–579.
- [8] D.-S. Pham, S. Venkatesh, Improved image recovery from compressed data contaminated with impulsive noise, *IEEE Trans. Image Process.* 21 (2012) 397–405.
- [9] A. Beck, M. Teboulle, A fast iterative shrinkage-thresholding algorithm for linear inverse problems, *SIAM J. Imag. Sci.* 2 (1) (2009) 183–202.
- [10] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein, Distributed optimization and statistical learning via the alternating direction method of multipliers, *Found. Trends Mach. Learn.* 3 (1) (2011) 1–122.
- [11] A. Colesanti, D. Hug, Hessian measures of semi-convex functions and applications to support measures of convex bodies, *Manuscripta Mathematica* 101 (2) (2000) 209–238.
- [12] E.J. Candès, M.B. Wakin, S.P. Boyd, Enhancing sparsity by reweighted  $\ell_1$  minimization, *J. Fourier Anal. Appl.* 14 (5) (2008) 877–905.
- [13] L. Chen, Y. Gu, The convergence guarantees of a non-convex approach for sparse recovery, *IEEE Trans. Signal Process.* 62 (15) (2014) 3754–3767.
- [14] J. Trzasko, A. Manduca, Highly undersampled magnetic resonance image reconstruction via homotopic  $\ell_0$ -minimization, *IEEE Trans. Med. Imaging* 28 (1) (2009) 106–121.
- [15] Q. Liu, H.C. So, Y. Gu, Off-grid DOA estimation with nonconvex regularization via joint sparse representation, *Signal Process.* 140 (2017) 171–176.
- [16] A. Antoniadis, J. Fan, Regularization of wavelet approximations, *J. Am. Stat. Assoc.* 96 (455) (2001) 939–955.
- [17] P. Yin, Y. Lou, Q. He, J. Xin, Minimization of  $\ell_{1-2}$  for compressed sensing, *SIAM J. Scient. Comput.* 37 (1) (2015) A536–A563.
- [18] R. Chartrand, Shrinkage mappings and their induced penalty functions, in: 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Florence, Italy, 2014, pp. 1026–1029.
- [19] A. Parekh, I.W. Selesnick, Convex denoising using non-convex tight frame regularization, *IEEE Signal Process. Lett.* 22 (10) (2015) 1786–1790.
- [20] C.-H. Zhang, Nearly unbiased variable selection under minimax concave penalty, *Ann. Stat.* 38 (2) (2010) 894–942.
- [21] A. Lanza, S. Morigi, I. Selesnick, F. Sgallari, Nonconvex nonsmooth optimization via convex–nonconvex majorization–minimization, *Numerische Mathematik* 136 (2) (2017) 343–381.
- [22] X. Shen, L. Chen, Y. Gu, H.C. So, Square-root lasso with nonconvex regularization: an ADMM approach, *IEEE Signal Process. Lett.* 23 (7) (2016) 934–938.
- [23] J. Eckstein, D.P. Bertsekas, On the Douglas–Rachford splitting method and the proximal point algorithm for maximal monotone operators, *Math. Program.* 55 (1) (1992) 293–318.
- [24] N. Parikh, S. Boyd, Proximal algorithms, *Found. Trends Optim.* 1 (3) (2014) 127–239.
- [25] J.N. Laska, M.A. Davenport, R.G. Baraniuk, Exact signal recovery from sparsely corrupted measurements through the pursuit of justice, in: 2009 Conference Record of the Forty-Third Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA, USA, 2009, pp. 1556–1560.
- [26] S. Foucart, H. Rauhut, *A Mathematical Introduction to Compressive Sensing*, Birkhäuser Basel, 2013.
- [27] X. Li, Compressed sensing and matrix completion with constant proportion of corruptions, *Construct. Approx.* 37 (1) (2011) 73–99.
- [28] G. Li, T.K. Pong, Global convergence of splitting method for nonconvex composite optimization, *SIAM J. Optim.* 25 (4) (2015) 2434–2460.
- [29] R.T. Trevor Hastie, J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Second, Springer-Verlag New York, 2009.
- [30] L.B. Montefusco, D. Lazzaro, S. Papi, Fast sparse image reconstruction using adaptive nonlinear filtering, *IEEE Trans. Image Process.* 20 (2) (2011) 534–544.