

## Research Article

# Sparse Signal Inversion with Impulsive Noise by Dual Spectral Projected Gradient Method

Liang Ding,<sup>1</sup> Hui Zhao,<sup>1</sup> and Yixin Dou<sup>2</sup>

<sup>1</sup>Department of Mathematics, Northeast Forestry University, No. 26 Hexing Street, Xiangfang District, Harbin, China

<sup>2</sup>School of Economics and Finance, Harbin University of Commerce, No. 1 Xuehai Street, Songbei District, Harbin, China

Correspondence should be addressed to Liang Ding; dl@nefu.edu.cn

Received 20 April 2017; Accepted 24 July 2017; Published 14 September 2017

Academic Editor: Haranath Kar

Copyright © 2017 Liang Ding et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We consider sparse signal inversion with impulsive noise. There are three major ingredients. The first is regularizing properties; we discuss convergence rate of regularized solutions. The second is devoted to the numerical solutions. It is challenging due to the fact that both fidelity and regularization term lack differentiability. Moreover, for ill-conditioned problems, sparsity regularization is often unstable. We propose a novel dual spectral projected gradient (DSPG) method which combines the dual problem of multiparameter regularization with spectral projection gradient method to solve the nonsmooth  $\ell^1 + \ell^1$  optimization functional. We show that one can overcome the nondifferentiability and instability by adding a smooth  $\ell^2$  regularization term to the original optimization functional. The advantage of the proposed functional is that its convex duality reduced to a constraint smooth functional. Moreover, it is stable even for ill-conditioned problems. Spectral projected gradient algorithm is used to compute the minimizers and we prove the convergence. The third is numerical simulation. Some experiments are performed, using compressed sensing and image inpainting, to demonstrate the efficiency of the proposed approach.

## 1. Introduction

In the present manuscript we are concerned with ill-posed linear operator equation:

$$Ax = y, \quad (1)$$

where  $x$  is sparse with respect to an orthonormal basis and  $A : D(A) \subset X \rightarrow Y$  is a bounded linear operator. In practice, exact data  $y$  are not known precisely, but only an approximation  $y^\delta$  with

$$\|y - y^\delta\| \leq \delta \quad (2)$$

is available. We call  $y^\delta$  the noisy data and  $\delta$  the noise level. It is well known that the conventional method for solving (1) is sparsity regularization, which provides an efficient way to extract the essential features of sparse solutions compared with oversmoothed classical Tikhonov regularization.

In the past ten years, sparsity regularization has certainly become an important concept in inverse problems. The theory of sparse recovery has largely been driven by the needs of applications in compressed sensing [1, 2],

bioluminescence tomography [3], seismic tomography [4], parameter identification [5], and so forth. For accounts of the regularizing properties and computational techniques in sparsity regularization we refer the reader to [5–7] and the references given there. In general, sparsity regularization is given by

$$\min_x \|Ax - y^\delta\|_{\ell^2}^2 + \alpha \|x\|_{w,p}^p, \quad (3)$$

where  $\|x\|_{w,p}^p = \sum_\gamma \omega_\gamma |\langle \varphi_\gamma, x \rangle|^p$  ( $1 \leq p \leq 2$ ),  $\alpha$  is the regularization parameter balancing the fidelity  $\|Ax - y^\delta\|_{\ell^2}^2$  and regularization term  $\|x\|_{w,p}^p$ . The functional in (3) is not convex if  $p < 1$ , it is challenging to investigate the regularizing properties and numerical computing method of minimizers. Limited work has been done for  $p < 1$ ; we refer the reader to [8–11] for a recent account of the theory. In this paper, we will focus our main attention on the situation of  $p = 1$ .

The aim of this paper is to consider a regularization functional of the form

$$\min_x \|Ax - y^\delta\|_{\ell^1} + \alpha \sum_\gamma \omega_\gamma |\langle \varphi_\gamma, x \rangle|. \quad (4)$$

We call (4)  $\ell^1 + \ell^1$  problem. A main motivation to investigate the  $\ell^1 + \ell^1$  problem is that noisy data  $y^\delta$  often contain impulsive noise. For Gaussian noise,  $\ell^2$  fidelity is a natural choice. However, a typical nondifferentiable fidelity used in application involving impulsive noise is the  $\ell^1$  fidelity, which is more robust than  $\ell^2$  fidelity [12].

$\ell^1$  fidelity technique was motivated by [13] for signal processing and later had attracted considerable attention in image processing, especially in image denoising. In image processing,  $\ell^1$  fidelity typically combines with TV regularization term. For details of  $\ell^1 + \text{TV}$  problems, we refer the reader to [14–17]. Nowadays  $\ell^1$  fidelity has received growing interest in the inverse problems where solutions are sparse with respect to an orthonormal basis. Minimizers of cost functions involving  $\ell^1$  fidelity combined with sparsity regularization have been studied. We refer the reader to [18–23] and the references given there.

For  $\ell^1$  fidelity, regularizing properties must be handled specifically. Burger and Osher [24] proved convergence rate when the regularization parameter is chosen arbitrarily fixed but sufficiently small; the authors call this phenomenon exact penalization. The most curious situation is that, in [25], the optimal convergence rate requires that regularization parameter must be equal to a fixed value. In [26], Grasmair et al. proved that source condition together with finite basis injectivity (FBI) condition is weaker than the restricted isometry property and obtained linear convergence  $O(\delta)$ . Flemming and Hegland discussed convergence rates in  $\ell^1$ -regularization when the basis is not smooth enough [27]. König et al. obtained high order polynomial rates of convergence in the special corrupted domain, even though the underlying problem is exponentially ill-posed in the classical sense [28].

Though  $\ell^1$  fidelity is robust, more researchers prefer to use  $\ell^2$  fidelity because of its differentiability. Hence a key issue for the  $\ell^1$  fidelity is the numerical computing methods. In the past few years, numerous algorithms have been systematically proposed for the  $\ell^1 + \text{TV}$  problems. On the other hand, in spite of growing interests in the  $\ell^1$  fidelity, we can indicate limited work has been done for numerical methods of  $\ell^1 + \ell^1$  problems. For sparsity regularization, the popular algorithms, for example, homotopy (LARS) method [29], iteratively reweighted least squares (IRLS) method [30], and iterative thresholding algorithm [31, 32], cannot be directly applied to  $\ell^1 + \ell^1$  problem due to the fact that both fidelity and regularization term lack differentiability. There are only a few papers, in which numerical algorithms for  $\ell^1 + \ell^1$  problems have been discussed systematically. References [20, 22] proposed the new model by statistics method. Reference [21] discussed the prior sparse representation and the data-fidelity term and proposed the two-phase approach. In [23], authors propose a robust bisparsity model (RBSM) to effectively exploit the prior knowledge about the similarities and the distinctions of signals. However, the researchers often devoted them to compressive sensing problem, where random matrices are well-conditioned. For ill-conditioned problems, these methods are often unstable [33] [Chap. 5].

Moreover, the researchers assume that the solution is sparse itself, which is different from the general assumption that the solution is sparse with respect to an orthonormal basis. In [19], Yang and Zhang proposed a Primal Dual-Interior Point Methods (PD-IPM) for EIT problem, which is efficient at dealing with the nondifferentiability. However, they did not give the convergence proof. Yang and Zhang reformulated the  $\ell^1 + \ell^1$  problem into the basis pursuit model which can be solved effectively by ADM method [18]. It is a competitive method compared with other algorithms for compressive sensing. In [34], Xiao et al. applied ADM method to  $\ell^1 + \ell^1$  problem directly. Numerical results illustrated that the proposed algorithm performs better than Yall1 [18].

In this paper, we investigate regularizing properties of  $\ell^1 + \ell^1$  problems. The convergence rates of  $\|x_\alpha^\delta - x^\dagger\|_{\ell^2}$  can be shown to be  $O(\delta^{1-\epsilon})$  and  $O(\delta)$  by a priori and a posteriori choice with the additional source condition and FBI condition. As mentioned above, dual is a conventional technique to solve the Tikhonov regularization with  $\ell^1$  fidelity. However, there are some limitations to this approach to  $\ell^1 + \ell^1$  problems due to the fact that it is difficult to obtain the dual formulation of  $\ell^1 + \ell^1$  problems. Inspired by [14] and multiregularization theory [35–37], a smooth  $\ell^2$  term is added to original functional of regularization. The dual problem of this new cost function is reduced to a constraint smooth functional. Moreover, the smooth  $\ell^2$  regularization term can improve the stability. The conventional method to solve the constraint smooth functional is projected gradient algorithm. We use spectral projected gradient (SPG) method to seek for the minimizers of the constraint functional and prove the convergence.

An outline of this paper is as follows. We devote Section 2 to a discussion of regularizing properties, including well-posedness and convergence rate. The convergence rate for a priori and a posteriori choice is established. In Section 3, inspired by the theory of multiregularization, we construct a new functional which is equal to a constraint smooth functional according to Frechel duality. In Section 4, the spectral projected gradient method is applied to compute the minimizers. Section 5 provides a detailed exposition of multiparameter choice rule based on the balancing principle. Numerical experiments involving compressed sensing and image inpainting are presented in Section 6, showing that our proposed approaches are robust and efficient.

## 2. Regularization Properties

*2.1. Notation and Assumptions.* For the approximate solutions of  $Ax = y$ , we consider the minimization of the regularization functional

$$\begin{aligned} \mathcal{F}_{\alpha,\delta}(x) &= \begin{cases} \|Ax - y^\delta\|_{\ell^1} + \alpha R(x), & x \in \text{dom}(A) \cap \text{dom}(R), \\ +\infty, & x \notin \text{dom}(A) \cap \text{dom}(R), \end{cases} \quad (5) \end{aligned}$$

where  $R(x) := \sum_\gamma \omega_\gamma |\langle \varphi_\gamma, x \rangle|$  and the subdifferential of  $R(x)$  at  $x$  is denoted by  $\partial R(x) \subset X$ . All along this paper,  $X$  denotes a Hilbert space which is a subspace of  $\ell^2$  space and  $\langle \cdot, \cdot \rangle$

denotes the inner product;  $Y$  denotes a Banach space which is a subspace of  $\ell^1$  space.  $A : \text{dom}(A) \subseteq X \rightarrow Y$  is a bounded linear operator and  $\text{dom}(A) \cap \text{dom}(R) \neq \emptyset$ .  $(\varphi_\gamma)_{\gamma \in \Lambda} \subset X$  is an orthonormal basis where  $\Lambda$  is some countable index set. From now, we denote

$$x_\gamma = \langle x, \varphi_\gamma \rangle, \quad (6)$$

$$\|x\|_{\ell^2} = \left( \sum_\gamma |\langle \varphi_\gamma, x \rangle|^2 \right)^{1/2} = \left( \sum_\gamma |x_\gamma|^2 \right)^{1/2}.$$

In order to prove convergence rate results we denote by  $x_\alpha^\delta$  the minimizer of the regularization functional  $J_{\alpha,\delta}(x)$  for every  $\alpha > 0$  and use the following definition of  $R(x)$ -minimum norm solution.

*Definition 1.* An element  $x^\dagger$  is called a  $R(x)$ -minimum norm solution of linear problem  $Ax = y$  if

$$Ax^\dagger = y, \quad (7)$$

$$R(x^\dagger) = \min \{R(x) \mid Kx = y\}.$$

We define the sparsity as follows.

*Definition 2.*  $x \in X$  is sparse with respect to  $\{\varphi_\gamma\}_{\gamma \in \Lambda}$  in the sense that  $\text{supp}(x) := \{\gamma \in \Lambda : \langle \varphi_\gamma, x \rangle \neq 0\}$  is finite. If  $\|\text{supp}(x)\|_0 = s$  for some  $s \in \mathbb{N}$ , the  $x \in X$  is called  $s$ -sparse.

For a subset of indices  $J \subset \Lambda$ , we denote by

$$J = \{\gamma \in \Lambda : |\langle \varphi_\gamma, \xi \rangle| \geq \omega_{\min} > 0\}, \quad (8)$$

$$W = \text{span} \{\varphi_\gamma : \gamma \in J\},$$

where  $\xi \in \partial R(x^\dagger)$ . We denote by  $J^c$  the complement of  $J$ . In addition, let  $P_J x := \sum_{\gamma \in J} \langle \varphi_\gamma, x \rangle \varphi_\gamma$  and  $P_{J^c} x := \sum_{\gamma \notin J} \langle \varphi_\gamma, x \rangle \varphi_\gamma$ .

Source condition or variational inequality are necessary for analysis of convergence rate. With these conditions one can obtain convergence rate  $o(\delta^{1/2})$ , even saturation convergence rate  $o(\delta^{2/3})$ . But for sparsity regularization, in order to obtain convergence rate  $o(\delta^{1-\epsilon})$  ( $0 < \epsilon < 1$ ), even linear convergence  $o(\delta)$ , one needs the following assumptions.

*Assumption 3.* Assume that the following hold:

- (i) Source condition: there exists  $\eta \in Y^*$  and some  $\xi \in \partial R(x^\dagger)$  satisfying  $\xi = A^* \eta$ .
- (ii) Finite basis injectivity (FBI) condition: for every finite set  $\Lambda \subseteq \mathbb{N}$ , the restriction of  $A$  to  $\{\varphi_\gamma : \gamma \in \Lambda\}$  is injective.

The FBI condition was originally introduced in [8]. In [26], Grasmair et al. used a slightly weaker condition based on FBI condition and showed that Assumption 3 is in some sense the weakest possible condition that guarantees the linear convergence rate. Actually, if  $x^\dagger$  is the unique  $R(x)$ -minimizing solution, Assumption 3 is satisfied [26]. For  $0 \leq p < 1$ ,  $x^\dagger$  being the unique  $R(x)$ -minimizing solution

can not guarantee that Assumption 3 is satisfied. Even if Assumption 3 is satisfied for  $0 \leq p < 1$ , we do not know if linear convergence can be obtained due to fact that traditional proof for linear convergence needs the convexity of  $R(X)$ .

*2.2. Well-Posedness and Convergence Rate.* In this subsection, well-posedness and convergence rate of the regularization method are given.

**Proposition 4** (well-posedness). *For every  $\alpha > 0$  and  $y^\delta \in Y$ , minimizing  $J_{\alpha,\delta}(x)$  is well-defined, stable, and convergent.*

*Proof.* In  $J_{\alpha,\delta}(x)$ ,  $A$  is a linear bounded operator,  $X$  is Hilbert space, and  $R(x)$  is sequentially lower semicontinuous. It is easy to verify Assumption 3.13 [38]; then proof is along the lines of the proofs of Theorem 3.22 (existence), 3.23 (stability), and 3.26 (convergence) in [38].  $\square$

We now turn to convergence rate. For  $\ell^2 + \ell^1$  problem, Lorenz proved convergence rate  $O(\delta^{1/2})$  with source condition and FBI condition [7]. Linear convergence rate  $O(\delta)$  was improved by Grasmair et al. for nonlinear equation with additional two nonlinear conditions [8]. In [26], Grasmair et al. proved that source condition together with FBI condition is weaker than the restricted isometry property and obtained linear convergence  $O(\delta)$ . Flemming discussed convergence rates in  $\ell^1$ -regularization when the basis is not smooth enough [27].

*Remark 5.* We note that Bregman distance cannot be used as an error measure in this section due to the fact that  $R(x) := \sum_\gamma \omega_\gamma |\langle \varphi_\gamma, x \rangle|$  fails to be strictly convex. We refer reader to [8, 39] for details. In this section we use  $\ell^2$  norm as error measure.

The next results show a convergence rate for a priori parameter choice rule.

**Lemma 6.** *Assume that Assumption 3 holds. Then there exists a constant  $C > 0$  such that*

$$\|x - x^\dagger\|_{\ell^2} \leq C \|Ax - Ax^\dagger\|_{\ell^1} + (1 + C \|A\|) \|P_{J^c} x\|_{\ell^2} \quad (9)$$

for all  $x \in X$ .

*Proof.* Since  $\xi_\gamma = \langle \varphi_\gamma, \xi \rangle \in X$ , we see that

$$\lim_{\gamma \rightarrow \infty} \xi_\gamma = 0. \quad (10)$$

Together with the definition of  $J$ , (10) implies that  $J$  is a finite set. From Assumption 3, the restriction  $A$  to every  $x \in W$  is injective. Then for some constant  $C > 0$

$$\|x\|_{\ell^2} = \|A^{-1} Ax\|_{\ell^2} \leq \|A^{-1}\| \|Ax\|_{\ell^1} \leq C \|Ax\|_{\ell^1} \quad (11)$$

for all  $x \in W$ . By the definition of  $J$  and  $x^\dagger \in X$  we have that  $\langle \varphi_\gamma, x^\dagger \rangle = 0$  for every  $\gamma \notin J$  and  $x^\dagger \in W$ . From (11), it is easy to show that

$$\|P_J x - x^\dagger\|_{\ell^2} \leq C \|AP_J x - Ax^\dagger\|_{\ell^1}. \quad (12)$$

Then we conclude that

$$\begin{aligned}
\|x - x^\dagger\|_{\ell^2} &= \|P_J x - x^\dagger + P_{J^c} x\|_{\ell^2} \\
&\leq C \|AP_J x - Ax^\dagger\|_{\ell^1} + \|P_{J^c} x\|_{\ell^2} \\
&= C \|A(x - x^\dagger) - AP_{J^c} x\|_{\ell^1} + \|P_{J^c} x\|_{\ell^2} \\
&\leq C \|A(x - x^\dagger)\|_{\ell^1} + C \|A\| \|P_{J^c}(x)\|_{\ell^2} \\
&\quad + \|P_{J^c} x\|_{\ell^2} \\
&\leq C \|Ax - Ax^\dagger\|_{\ell^1} \\
&\quad + (C \|A\| + 1) \|P_{J^c}(x)\|_{\ell^2},
\end{aligned} \tag{13}$$

and the proof is completed.  $\square$

**Lemma 7.** Assume there exist  $\beta_1, \beta_2 > 0$  such that

$$R(x) - R(x^\dagger) \geq \beta_1 \|x - x^\dagger\|_{\ell^2} - \beta_2 \|Ax - Ax^\dagger\|_{\ell^1} \tag{14}$$

for all  $x \in \text{dom}(A)$ . If  $\alpha\beta_2 < 1$ , then

$$\|x_\alpha^\delta - x^\dagger\|_{\ell^2} \leq \frac{(1 + \alpha\beta_2)\delta}{\alpha\beta_1}. \tag{15}$$

*Proof.* Since  $x_\alpha^\delta$  minimize  $J_{\alpha,\delta}(x)$ , then

$$\|Ax_\alpha^\delta - y^\delta\|_{\ell^1} + \alpha R(x_\alpha^\delta) \leq \|Ax^\dagger - y^\delta\|_{\ell^1} + \alpha R(x^\dagger). \tag{16}$$

Together with condition (14) this implies the inequality

$$\begin{aligned}
\delta &\geq \|Ax^\dagger - y^\delta\|_{\ell^1} \\
&\geq \|Ax_\alpha^\delta - y^\delta\|_{\ell^1} + \alpha (R(x_\alpha^\delta) - R(x^\dagger)) \\
&\geq \|Ax_\alpha^\delta - y^\delta\|_{\ell^1} + \alpha\beta_1 \|x_\alpha^\delta - x^\dagger\|_{\ell^2} \\
&\quad - \alpha\beta_2 \|Ax_\alpha^\delta - Ax^\dagger\|_{\ell^1} \\
&\geq \|Ax_\alpha^\delta - y^\delta\|_{\ell^1} + \alpha\beta_1 \|x_\alpha^\delta - x^\dagger\|_{\ell^2} \\
&\quad - \alpha\beta_2 \|Ax_\alpha^\delta - y^\delta\|_{\ell^1} - \alpha\beta_2 \delta \\
&\geq \alpha\beta_1 \|x_\alpha^\delta - x^\dagger\|_{\ell^2} - \alpha\beta_2 \delta,
\end{aligned} \tag{17}$$

which proves Lemma 7.  $\square$

**Theorem 8.** Assume that Assumption 3 holds. Then for the choice  $\alpha = O(\delta^\epsilon)$  ( $0 < \epsilon < 1$ ), we deduce that

$$\|x_\alpha^\delta - x^\dagger\|_{\ell^2} = O(\delta^{1-\epsilon}). \tag{18}$$

*Proof.* Let  $m := \max\{|\langle \varphi_\gamma, \xi \rangle| \mid \gamma \notin J\}$ ; we can estimate

$$\begin{aligned}
\|P_{J^c}(x)\|_{\ell^2} &= \left( \sum_{\gamma \notin J} |\langle \varphi_\gamma, x \rangle|^2 \right)^{1/2} \leq \sum_{\gamma \notin J} |\langle \varphi_\gamma, x \rangle| \\
&\leq \frac{1}{\omega_{\min} - m} \sum_{\gamma \notin J} (\omega_\gamma - m) |\langle \varphi_\gamma, x \rangle| \leq \frac{1}{\omega_{\min} - m} \\
&\quad \cdot \sum_{\gamma \notin J} (\omega_\gamma |\langle \varphi_\gamma, x \rangle| - \langle \varphi_\gamma, \xi \rangle \langle \varphi_\gamma, x \rangle) \leq \frac{1}{\omega_{\min} - m} \\
&\quad \cdot \sum_{\gamma \notin J} (\omega_\gamma |\langle \varphi_\gamma, x \rangle| - \omega_\gamma |\langle \varphi_\gamma, x^\dagger \rangle| \\
&\quad - \langle \varphi_\gamma, \xi \rangle \langle \varphi_\gamma, x - x^\dagger \rangle) = \frac{1}{\omega_{\min} - m} (R(x) \\
&\quad - R(x^\dagger) - \langle \xi, x - x^\dagger \rangle) = \frac{1}{\omega_{\min} - m} (R(x) \\
&\quad - R(x^\dagger) + \langle \eta, Ax^\dagger - Ax \rangle) = \frac{1}{\omega_{\min} - m} (R(x) \\
&\quad - R(x^\dagger) + \|\eta\|_{\ell^\infty} \|Ax - Ax^\dagger\|_{\ell^1}).
\end{aligned} \tag{19}$$

From Lemma 6, we obtain

$$\begin{aligned}
\|x - x^\dagger\|_{\ell^2} &\leq C \|Ax - Ax^\dagger\|_{\ell^1} + (C \|A\| + 1) \\
&\quad \cdot \left( \frac{1}{\omega_{\min} - m} (R(x) - R(x^\dagger)) \right. \\
&\quad \left. + \|\eta\|_{\ell^\infty} \|Ax - Ax^\dagger\|_{\ell^1} \right);
\end{aligned} \tag{20}$$

that is,

$$\begin{aligned}
R(x) - R(x^\dagger) &\geq \frac{\omega_{\min} - m}{C \|A\| + 1} \|x - x^\dagger\|_{\ell^2} \\
&\quad - \left( \|\eta\|_{\ell^\infty} + \frac{C(\omega_{\min} - m)}{C \|A\| + 1} \right) \|Ax - Ax^\dagger\|_{\ell^1}.
\end{aligned} \tag{21}$$

Let

$$\begin{aligned}
\beta_1 &= \frac{\omega_{\min} - m}{C \|A\| + 1}, \\
\beta_2 &= \|\eta\|_{\ell^\infty} + \frac{C(\omega_{\min} - m)}{C \|A\| + 1}, \\
\alpha\beta_2 &< 1.
\end{aligned} \tag{22}$$

According to Lemma 7, it follows that

$$\|x_\alpha^\delta - x^\dagger\|_{\ell^2} \leq \frac{(1 + \alpha\beta_2)\delta}{\alpha\beta_1}. \tag{23}$$

The assertion follows from  $\alpha = O(\delta^\epsilon)$ .  $\square$

Next we turn our attention to a posteriori parameter choice which is so-called discrepancy principle due to Morozov [40]. The regularization parameter defined via the discrepancy principle is

$$\alpha := \sup \{ \alpha > 0 \mid \|Ax_\alpha^\delta - y^\delta\|_{\ell^1} = \tau\delta \} \quad (24)$$

for some  $\tau \geq 1$ . Morozov discrepancy principle is a widely used and easy numerical implementation rule. Furthermore, for  $\tau = 1$ , (4) is equivalent to a constrained minimization problem [26]

$$\begin{aligned} R(x^\dagger) &\longrightarrow \min \\ \text{subject to } &\|Ax_\alpha^\delta - y^\delta\|_{\ell^1} \leq \tau\delta, \end{aligned} \quad (25)$$

which is widely used in compressive sensing. The next results show a convergence rate by the Morozov discrepancy principle.

**Theorem 9.** *Assume that Assumption 3 holds and that the regularization parameter is determined by (24). Then*

$$\|x_\alpha^\delta - x^\dagger\|_{\ell^2} = O(\delta). \quad (26)$$

*Proof.* Since  $x_\alpha^\delta$  is a minimizer of  $J_{\alpha,\delta}(x)$ , the inequality

$$\begin{aligned} \|Ax_\alpha^\delta - y^\delta\|_{\ell^1} + \alpha R(x_\alpha^\delta) &\leq \|Ax^\dagger - y^\delta\|_{\ell^1} + \alpha R(x^\dagger) \\ &\leq \delta + \alpha R(x^\dagger) \end{aligned} \quad (27)$$

holds. This together with  $\|Ax_\alpha^\delta - y^\delta\|_{\ell^1} = \tau\delta$  ( $\tau \geq 1$ ) implies that

$$R(x_\alpha^\delta) \leq R(x^\dagger). \quad (28)$$

Proceeding as in the proof of Theorem 8, we obtain that

$$\begin{aligned} \|P_{J^c}(x_\alpha^\delta)\|_{\ell^2} &\leq \frac{1}{\omega_{\min} - m} (R(x_\alpha^\delta) - R(x^\dagger)) \\ &+ \|\eta\|_{\ell^\infty} \|Ax_\alpha^\delta - Ax^\dagger\|_{\ell^1} \leq \frac{\|\eta\|_{\ell^\infty} \|Ax_\alpha^\delta - Ax^\dagger\|_{\ell^1}}{\omega_{\min} - m}. \end{aligned} \quad (29)$$

From Lemma 6, it follows that

$$\begin{aligned} \|x_\alpha^\delta - x^\dagger\|_{\ell^2} &\leq C \|Ax_\alpha^\delta - Ax^\dagger\|_{\ell^1} + (C\|A\| + 1) \\ &\cdot \|P_{J^c}(x_\alpha^\delta)\|_{\ell^2} \leq C \|Ax_\alpha^\delta - Ax^\dagger\|_{\ell^1} + (C\|A\| + 1) \\ &\cdot \frac{\|\eta\|_{\ell^\infty} \|Ax_\alpha^\delta - Ax^\dagger\|_{\ell^1}}{\omega_{\min} - m} \\ &\leq \left( C + \frac{\|\eta\|_{\ell^\infty} (C\|A\| + 1)}{\omega_{\min} - m} \right) \|Ax_\alpha^\delta - Ax^\dagger\|_{\ell^1} \\ &\leq \left( C + \frac{\|\eta\|_{\ell^\infty} (C\|A\| + 1)}{\omega_{\min} - m} \right) \\ &\cdot (\|Ax_\alpha^\delta - y^\delta\|_{\ell^1} + \|Ax^\dagger - y^\delta\|_{\ell^1}) \\ &\leq \left( C + \frac{\|\eta\|_{\ell^\infty} (C\|A\| + 1)}{\omega_{\min} - m} \right) (\tau + 1) \delta. \end{aligned} \quad (30)$$

The proof is completed.  $\square$

*Remark 10.* Linear convergence rate  $O(\delta)$  can also be obtained when the a priori choice and the Morozov discrepancy principle are applied to  $\|Kx_\alpha^\delta - y^\delta\|_{\ell^p}^p$  ( $1 < p \leq 2$ ) fidelity. However, one must let  $\alpha = O(\delta)$  in the a priori parameter choice rule.

### 3. Dual Problem

In this section, let  $\omega_\gamma$  in (4) take the same value; that is,  $\omega_\gamma = \mu > 0$  for all  $\gamma \in \Gamma$ . It is reasonable because convergence can be obtained when  $\delta/\alpha \rightarrow 0$  [6]. Let  $\alpha := \alpha\mu = \alpha\omega_\gamma$ ; then (4) is equivalent to

$$\min_x \|Ax - y^\delta\|_{\ell^1} + \alpha \sum_\gamma |\langle \varphi_\gamma, x \rangle|. \quad (31)$$

Let  $u = (x_1, x_2, \dots, x_\gamma, \dots) \in \ell^2$ , where  $x_\gamma = \langle \varphi_\gamma, x \rangle$ . In addition, we denote by  $D: \ell^2 \rightarrow \ell^2$  a dictionary satisfied with  $u = Dx$  and  $u^\dagger = Dx^\dagger$ . For example, in the field of wavelet transform,  $D$  is a wavelet decomposition operator and  $D^T$  is a wavelet reconstruction operator [41, 42]. Let  $K = A \circ D^T$ ; then (4) is equivalent to

$$\mathcal{P}: \min_u \{ \mathcal{F}_\alpha(u) = \|Ku - y^\delta\|_{\ell^1} + \alpha \|u\|_{\ell^1} \}. \quad (32)$$

Dual is a conventional method for solving Tikhonov regularization with  $\ell^1$  fidelity. However, there are some limitations to this approach to solve (32). The main difficulty is that both the  $\ell^1$  fidelity and the  $\ell^1$  regularization term are nondifferentiable. Moreover, for ill-conditioned problems, sparsity regularization is often unstable. We add the smooth penalty  $(\beta/2)\|u\|_{\ell^2}^2$  to (32) to construct the following functional:

$$\mathcal{P}_\beta: \min_u \{ \mathcal{F}_{\alpha,\beta}(u) = \|Ku - y^\delta\|_{\ell^1} + \alpha \|u\|_{\ell^1} + \frac{\beta}{2} \|u\|_{\ell^2}^2 \}. \quad (33)$$

The advantage of problem (33) in place of (32) is that the dual problem of (33) is a constraint smooth functional and projected gradient algorithm can be used to compute minimizers. Moreover, the regularization effect of  $\ell^1$  penalty is weak and  $\ell^2$  penalty can improve the stability of (32).

Next we will investigate the convergence of the minimizers to the functional  $\mathcal{P}_\beta$  as  $\beta$  tending to zero.

**Theorem 11.** *Let  $\alpha$  be fixed and  $\{\beta_n\}$  be a sequence tending to zero. Then the minimizers  $u_{\alpha,\beta_n}^{\beta_n}$  to the functional  $J_{\alpha,\beta_n}(u)$  have a subsequence converging to  $u^*$  being a minimizer of the functional  $J_\alpha(u)$ .*

*Proof.* By the definition of  $u_{\alpha,\beta_n}^{\beta_n}$ , we have

$$\begin{aligned} \mathcal{F}_{\alpha,\beta_n}(u_{\alpha,\beta_n}^{\beta_n}) &= \|Ku_{\alpha,\beta_n}^{\beta_n} - y^\delta\|_{\ell^1} + \alpha \|u_{\alpha,\beta_n}^{\beta_n}\|_{\ell^1} + \frac{\beta_n}{2} \|u_{\alpha,\beta_n}^{\beta_n}\|_{\ell^2}^2 \\ &\leq \|Ku^\dagger - y^\delta\|_{\ell^1} + \alpha \|u^\dagger\|_{\ell^1} + \frac{\beta_n}{2} \|u^\dagger\|_{\ell^2}^2 \\ &\leq \delta + \alpha \|u^\dagger\|_{\ell^1} + \frac{\beta_n}{2} \|u^\dagger\|_{\ell^2}^2, \end{aligned} \quad (34)$$

and hence

$$\|u_\alpha^{\beta_n}\|_{\ell^1} \leq \frac{\delta}{\alpha} + \|u^\dagger\|_{\ell^1} + \frac{\beta_n}{2\alpha} \|u^\dagger\|_{\ell^2}^2. \quad (35)$$

Since  $\alpha$  is a fixed value and  $\beta_n \rightarrow 0$ , there exists a constant  $C > 0$  such that

$$\|u_\alpha^{\beta_n}\|_{\ell^2} \leq \|u_\alpha^{\beta_n}\|_{\ell^1} \leq \frac{\delta}{\alpha} + \|u^\dagger\|_{\ell^1} + \frac{\beta_n}{2\alpha} \|u^\dagger\|_{\ell^2}^2 < C. \quad (36)$$

It follows that  $\{u_\alpha^{\beta_n}\}$  is uniformly bounded. Therefore, a subsequence  $\{u_\alpha^{\beta_m}\}$  of  $\{u_\alpha^{\beta_n}\}$  and  $u^*$  exist such that

$$u_\alpha^{\beta_m} \rightarrow u^*. \quad (37)$$

By the weak lower semicontinuity of norm, we obtain that

$$\begin{aligned} \|Ku^* - y^\delta\|_{\ell^1} &\leq \liminf \|Ku_\alpha^{\beta_m} - y^\delta\|_{\ell^1}, \\ \|u^*\|_{\ell^1} &\leq \liminf \|u_\alpha^{\beta_m}\|_{\ell^1}, \end{aligned} \quad (38)$$

and hence that

$$\begin{aligned} \|Ku^* - y^\delta\|_{\ell^1} + \alpha \|u^*\|_{\ell^1} &\leq \liminf \left( \|Ku_\alpha^{\beta_m} - y^\delta\|_{\ell^1} \right. \\ &\quad \left. + \alpha \|u_\alpha^{\beta_m}\|_{\ell^1} + \frac{\beta_m}{2} \|u_\alpha^{\beta_m}\|_{\ell^2}^2 \right) \\ &\leq \limsup \left( \|Ku_\alpha^{\beta_m} - y^\delta\|_{\ell^1} + \alpha \|u_\alpha^{\beta_m}\|_{\ell^1} \right. \\ &\quad \left. + \frac{\beta_m}{2} \|u_\alpha^{\beta_m}\|_{\ell^2}^2 \right) \leq \limsup \left( \|Ku - y^\delta\|_{\ell^1} + \alpha \|u\|_{\ell^1} \right. \\ &\quad \left. + \frac{\beta_m}{2} \|u\|_{\ell^2}^2 \right) = \|Ku - y^\delta\|_{\ell^1} + \alpha \|u\|_{\ell^1} \end{aligned} \quad (39)$$

for all  $u \in D(K)$ . Therefore,  $u^*$  is a minimizer of the functional  $J_\alpha(u)$  and the proof is completed.  $\square$

Next we consider the dual problem of  $\mathcal{P}_\beta$ . We will show that the constraint smooth minimization problems  $\mathcal{P}_\beta^*$

$$\begin{aligned} \mathcal{P}_\beta^*: \min \quad & \frac{1}{2\beta} \|K^* p\|_{\ell^2}^2 - \langle p, y^\delta \rangle, \\ \text{s.t.} \quad & \|p\|_{\ell^\infty} \leq \min\{\alpha, 1\}, \end{aligned} \quad (40)$$

are the dual problem of  $\mathcal{P}_\beta$ .

**Theorem 12.**  $\mathcal{P}_\beta^*$  is the dual problem of  $\mathcal{P}_\beta$ . The solutions  $u_\alpha^\beta$  of  $\mathcal{P}_\beta$  and  $p_\alpha^\beta$  of  $\mathcal{P}_\beta^*$  have the following relation:

$$(u_\alpha^\beta)_i = S_{\alpha/\beta} \left( \frac{(K^* p_\alpha^\beta)_i}{\beta} \right), \quad i = 1, 2, \dots, \quad (41)$$

$$\langle Ku_\alpha^\beta - y^\delta, p - p_\alpha^\beta \rangle \geq 0$$

for all  $p \in \ell^2$  with  $\|p\|_{\ell^\infty} \leq \min\{\alpha, 1\}$ , where

$$S_{\alpha/\beta}(x) = \begin{cases} x - \frac{\alpha}{\beta}, & \text{if } x \geq \frac{\alpha}{\beta}, \\ 0, & \text{if } |x| < \frac{\alpha}{\beta}, \\ x + \frac{\alpha}{\beta}, & \text{if } x \leq -\frac{\alpha}{\beta} \end{cases} \quad (42)$$

is the soft threshold function.

*Proof.* Let

$$\begin{aligned} F(u) &= \|u - y^\delta\|_{\ell^1}; \\ R(u) &= \frac{\beta}{2} \|u\|_{\ell^2}^2 + \alpha \|u\|_{\ell^1}, \end{aligned} \quad (43)$$

then problem  $\mathcal{P}_\beta$  can be rewritten as

$$\inf_{u \in \ell^2} F(Ku) + R(u). \quad (44)$$

Let us denote by  $F^*$  and  $R^*$  the conjugate function of  $F$  and  $R$ . By the Fenchel duality [43] [Chap 3, Chap 10], it follows that

$$\begin{aligned} R^*(p) &= \begin{cases} \frac{1}{2\beta} \|p\|_{\ell^2}^2, & \text{if } \|p\|_{\ell^\infty} \leq \alpha, \\ \infty, & \text{else,} \end{cases} \\ F^*(p) &= \begin{cases} \langle p, y^\delta \rangle_{\ell^2}, & \text{if } \|p\|_{\ell^\infty} \leq 1, \\ \infty, & \text{else.} \end{cases} \end{aligned} \quad (45)$$

In  $J_{\alpha,\delta}(x)$ , the functionals  $F$  and  $R$  are proper, convex lower semicontinuous. By the Fenchel duality theorem [43] [Chap 3, Prop. 2.4, Prop. 4.1, Rem. 4.2], it is easy to show that

$$\inf_{u \in \ell^2} F(Ku) + R(u) = \sup_{p \in \ell^2} -F^*(-p) - R^*(K^*p); \quad (46)$$

that is,

$$\inf_{u \in \ell^2} F(Ku) + R(u) = -\inf_{p \in \ell^2} F^*(-p) + R^*(K^*p). \quad (47)$$

Since  $u_\alpha^\beta$  and  $p_\alpha^\beta$  are minimizers of  $\mathcal{P}_\beta$  and  $\mathcal{P}_\beta^*$ , we have that

$$F(Ku_\alpha^\beta) + R(u_\alpha^\beta) + F^*(-p_\alpha^\beta) + R^*(K^*p_\alpha^\beta) = 0. \quad (48)$$

Moreover, the extremality condition (48) is equivalent to the Kuhn-Tucker conditions

$$\begin{aligned} -p_\alpha^\beta &\in \partial F(Ku_\alpha^\beta), \\ K^* p_\alpha^\beta &\in \partial R(u_\alpha^\beta). \end{aligned} \quad (49)$$

By the definition of the subgradient and  $-p_\alpha^\beta \in \partial F(Ku_\alpha^\beta)$  in (49), it follows that

$$\begin{aligned} Ku_\alpha^\beta &\in \partial F^*(-p_\alpha^\beta), \\ Ku_\alpha^\beta - y^\delta &\in \partial I_{\{\| \cdot \|_{\ell^\infty} \leq 1\}}. \end{aligned} \quad (50)$$

```

(1) Set  $\gamma, \epsilon, \sigma_1, \sigma_2, \alpha_{\min}, \alpha_{\max} > 0, M \geq 0, k = 1$ ; choose  $p_1 \in \ell^2, \alpha_1$ 
(2)  $d_k = P_c(p_k - \alpha_k \nabla f(p_k)) - p_k$ 
(3) Select  $\lambda_k$  using line research Algorithm 2
(4) compute  $p_{k+1} = p_k + \lambda_k d_k, s_k = p_{k+1} - p_k, q_k = \nabla f(p_{k+1}) - \nabla f(p_k), \beta_k = \langle s_k, q_k \rangle$ 
(5)
if  $\beta_k \leq 0$  then
    Set  $\alpha_{k+1} = \alpha_{\max}$ 
    Else
    Set  $\alpha_{k+1} = \min \{ \alpha_{\max}, \max \{ \alpha_{\min}, \langle s_k, s_k \rangle / \beta_k \} \}$ 
end if
(6)  $k = k + 1$ 
Until  $\|P_c(x_k + \nabla f(x_k)) - x_k\|_2 \leq \epsilon$  or  $k = k_{\max}$ 
    
```

 ALGORITHM 1: Spectral projected gradient method for  $\mathcal{F}_\beta^*$ .

Then we obtain

$$\langle Ku_\alpha^\beta - y^\delta, p - p_\alpha^\beta \rangle \geq 0. \quad (51)$$

By the definition of the subgradient and  $K^* p_\alpha^\beta \in \partial R(u_\alpha^\beta)$  in (49), we conclude that

$$(K^* p_\alpha^\beta)_i = \beta (u_\alpha^\beta)_i + \alpha \text{sign}((u_\alpha^\beta)_i), \quad i = 1, 2, \dots \quad (52)$$

For  $(u_\alpha^\beta)_i > 0$ , this leads to  $(u_\alpha^\beta)_i = (K^*(p_\alpha^\beta))_i / \beta - \alpha / \beta$ ; we must impose in this case that  $(K^*(p_\alpha^\beta))_i / \beta > \alpha / \beta$ . For  $(u_\alpha^\beta)_i < 0$ , this leads to  $(u_\alpha^\beta)_i = (K^*(p_\alpha^\beta))_i / \beta + \alpha / \beta$ , valid only when  $(K^*(p_\alpha^\beta))_i / \beta < -\alpha / \beta$ . When  $|(K^*(p_\alpha^\beta))_i / \beta| < \alpha / \beta$ , we put  $(u_\alpha^\beta)_i = 0$ . Summarizing

$$(u_\alpha^\beta)_i = S_{\alpha/\beta} \left( \frac{(K^*(p_\alpha^\beta))_i}{\beta} \right), \quad (53)$$

where  $S_{\alpha/\beta}$  is soft threshold function which is defined by

$$S_{\alpha/\beta}(x) = \begin{cases} x - \frac{\alpha}{\beta}, & \text{if } x \geq \frac{\alpha}{\beta}, \\ 0, & \text{if } |x| < \frac{\alpha}{\beta}, \\ x + \frac{\alpha}{\beta}, & \text{if } x \leq -\frac{\alpha}{\beta}, \end{cases} \quad (54)$$

the proof is completed.  $\square$

*Remark 13.* In [39] (chap 1), numerical experiments showed that the choice of  $p$  slightly larger than 1 gives even better results than  $p = 1$ . It is necessary to discuss the functional

$$\min_u \left\{ \mathcal{J}_\alpha(u) = \|Ku - y^\delta\|_{\ell^1} + \frac{\alpha}{p} \|u\|_{\ell^p}^p \right\}, \quad (55)$$

where  $1 < p < 2$ ; we call (55)  $\ell^1 + \ell^p$  problem. We do not have to add additional  $\ell^2$  penalty to (55); its dual problem can be obtained by Fenchel dual theory directly.

## 4. Computation of Minimizers

As mentioned above, the functional equation (33) can be transformed using Fenchel duality into a smooth functional with a box constraint, which can be solved effectively using projected gradient-type methods. The spectral projected gradient (SPG) method [44] has been proved to be effective for the minimization of differentiable functional with box constraints. The function  $f(p)$  is defined by

$$f(p) = \frac{1}{2\beta} \|K^* p\|_{\ell^2}^2 - \langle p, y^\delta \rangle; \quad (56)$$

then

$$\nabla f(p) = \frac{1}{\beta} KK^* p - y^\delta. \quad (57)$$

By setting

$$B(c) := \{p \in U^* \mid \|p\|_{\ell^\infty} \leq c\}, \quad (58)$$

where  $c = \min\{\alpha, 1\}$ ,

we denote by  $P_{B(c)}(z)$  the orthogonal projection on the ball  $B(c)$

$$P_{B(c)}(x)(i, j) = c \frac{x(i, j)}{\max(c, |x(i, j)|)}. \quad (59)$$

Given initial value  $x_0$  and a constraint on step length  $\alpha_k$ , that is,  $0 < \alpha_{\min} \leq \alpha_k \leq \alpha_{\max}$ , let parameter  $\gamma \in (0, 1)$ , and  $0 < \sigma_1 < \sigma_2 < 1, M > 1$ . We use the convergence criteria given by

$$\|P_c(p_k + \nabla f(p_k)) - p_k\|_{\ell^2} \leq \epsilon \quad (60)$$

(see Algorithm 1). We have the following convergence results.

**Proposition 14.** For  $\gamma \in (0, 1)$  and  $0 < \sigma_1 < \sigma_2 < 1$ , algorithm SPG is well-defined; the sequence  $p_k$  generated by algorithm SPG converges to a stationary point of the functional  $P_\alpha^\beta$ .

*Proof.* Since  $f(p)$  is convex and has continuous derivatives, the convergence follows from standard results (refer to Theorem 2.1 in [44]).  $\square$

```

(1) Set  $\lambda = 1$ 
(2) Compute  $\bar{p} = p_k + \lambda d_k$ 
(3)
if
     $f(\bar{p}) \leq \max_{0 \leq j \leq \min\{k, M-1\}} f(p_{k-j}) + \gamma \lambda \langle d_k, \nabla f(p_k) \rangle$ 
then
    Let  $\lambda_k = \lambda$ 
    Else
    Let  $\lambda \in [\sigma_1 \lambda, \sigma_2 \lambda]$ , and go to (2)
end if

```

ALGORITHM 2: Line research.

```

(1) Choose  $\alpha^0$  and  $\beta^0$ , and let  $k = 0$ 
(2)  $u_{k+1} = \arg \min_x \left( \|Ku - y^\delta\|_{\ell^1} + \alpha \|u\|_{\ell^1} + \frac{\beta}{2} \|u\|_{\ell^2}^2 \right)$ 
(3)  $\alpha_{k+1} = \frac{1}{1 + \gamma} \frac{\|Ku_{k+1} - y^\delta\|_{\ell^1} + \beta \|u_{k+1}\|_{\ell^2}^2}{\|u_{k+1}\|_{\ell^1}}$ 
     $\beta_{k+1} = \frac{1}{1 + \gamma} \frac{\|Ku_{k+1} - y^\delta\|_{\ell^1} + \beta \|u_{k+1}\|_{\ell^1}}{\|u_{k+1}\|_{\ell^2}^2}$ 
Until a stopping criterion satisfied.

```

ALGORITHM 3: Fixed point algorithm for  $\alpha$  and  $\beta$ .

## 5. Choice of Parameter $\alpha$ and $\beta$

The solution  $u_\alpha^\beta$  of  $\mathcal{P}_\beta$  converges theoretically to the solution  $u_\alpha^\delta$  of  $\mathcal{P}$  as  $\beta \rightarrow 0$ . Obviously smaller  $\beta$  is better. However,  $(1/2\beta)\|K^*p\|_{\ell^2}^2$  tends to infinite as  $\beta \rightarrow 0$ . Moreover, the smaller  $\beta$  weakens the regularization effect of the  $\ell^2$  penalty, which leads to instability. If the solution is sparse, we can say that the nonzero coefficients are impulsive parts and the zero coefficients are smooth parts. Multiparameter regularization

$$P_\alpha^\beta: \min_{u \in \ell^2} \left\{ \mathcal{F}_{\alpha, \beta}(u) = \|Ku - y^\delta\|_{\ell^1} + \alpha \|u\|_{\ell^1} + \frac{\beta}{2} \|u\|_{\ell^2}^2 \right\} \quad (61)$$

is a conventional method for ill-posed problems if the solutions have a number of different structures. In [35, 37], numerical experiments show that multiparameter regularization can effectively recover the different part of solutions. If the solutions contain only a single structure, multiparameter regularization also has better performance. Hence it does not necessarily require the parameter  $\beta$  tending to zero when we choose the parameters  $\alpha$  and  $\beta$ . A conventional method for the choice of parameters  $\alpha$  and  $\beta$  is multiparameter regularization choice principle, for example, discrepancy principle [35] and balance principle [37]. In this paper, we use multiparameter balance principle for the choice of regularization parameters  $\alpha$  and  $\beta$ .

Balance principle is to compute minimizers of the function

$$\Phi_\gamma(\alpha, \beta) = c_\gamma \frac{(\inf \|Ku - y^\delta\|_{\ell^1} + \alpha \|u\|_{\ell^1} + (\beta/2) \|u\|_{\ell^2}^2)^{\gamma+2}}{\alpha\beta}, \quad (62)$$

where  $c_\gamma$  is a fixed constant, and (62) is equivalent to [37]

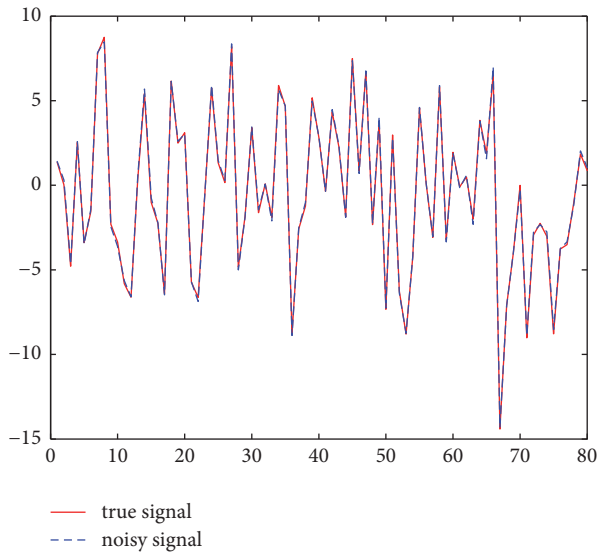
$$\gamma\alpha \|u_{\alpha, \beta}^\delta\|_{\ell^1} = \gamma\beta \|u_{\alpha, \beta}^\delta\|_{\ell^2}^2 = \|Ku_{\alpha, \beta}^\delta - y^\delta\|_{\ell^1}. \quad (63)$$

A fixed point algorithm for  $\alpha$  and  $\beta$  according to (63) is as in Algorithm 3.

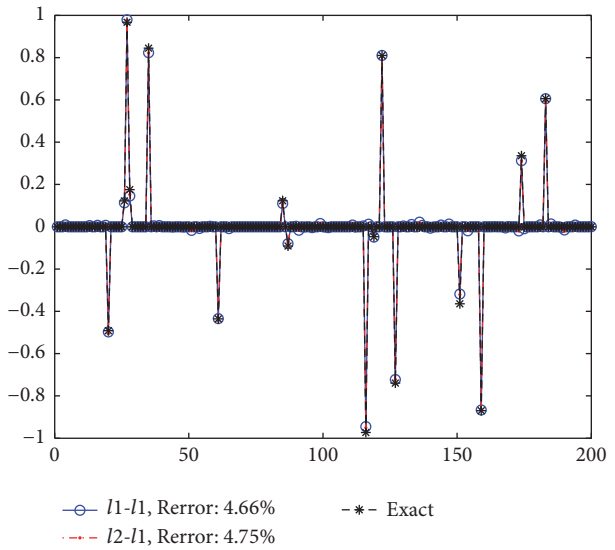
## 6. Numerical Simulations

In this section, we present some numerical experiments to illustrate the efficiency of the proposed method. In Section 6.1, numerical experiments involve compressive sensing. We aim to demonstrate that  $\ell^1$  fidelity is more stable than the  $\ell^2$  fidelity and is capable of handling both impulsive and Gaussian noises. In Section 6.2, we first compare the performance of the DSPG method with the alternating direction method (ADM) and TNIP method by well-conditioned compressive sensing problems. In the second example, we discuss an ill-posed problem where the condition number of linear operator  $A$  is 255; we aim to demonstrate that

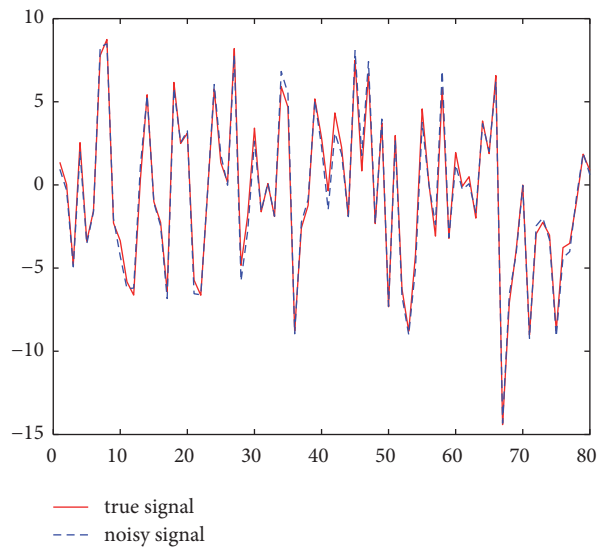




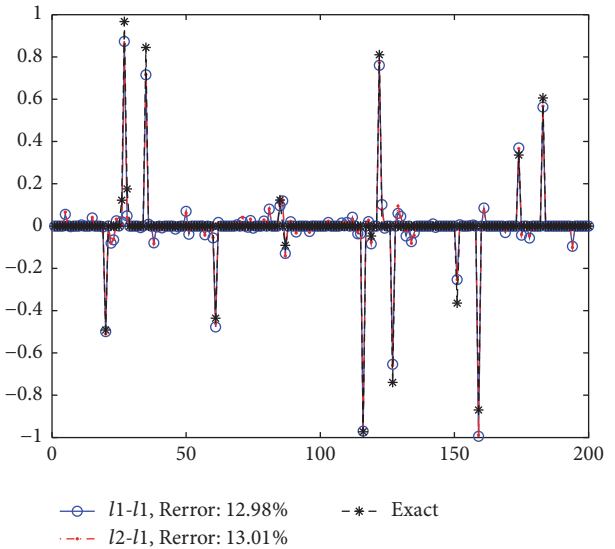
(a) 30 dB Gaussian noise



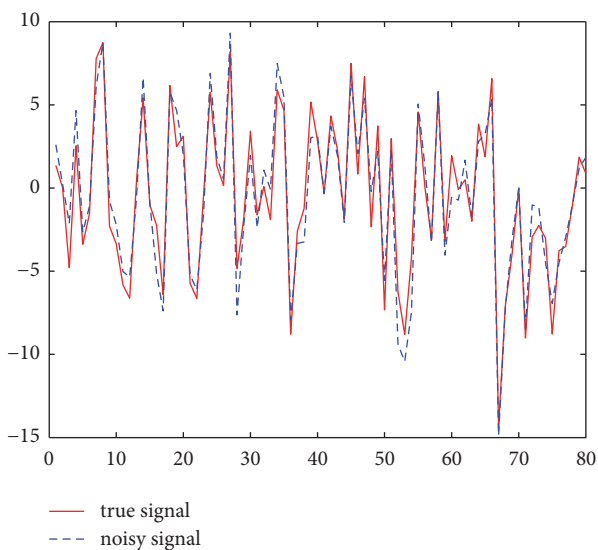
(b) Restorations with 30 dB Gaussian noise



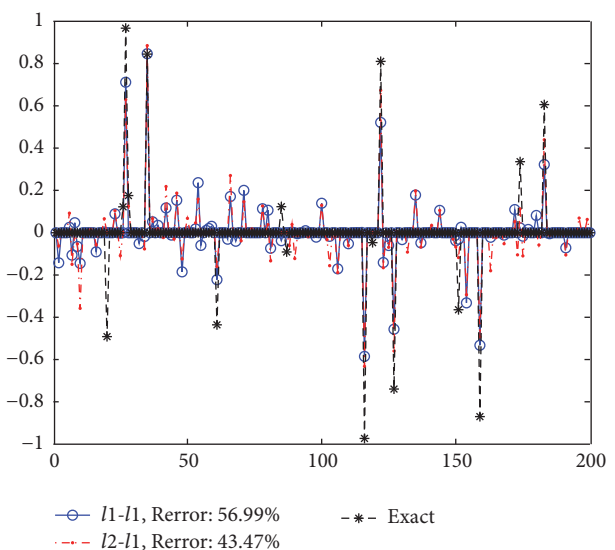
(c) 20 dB Gaussian noise



(d) Restorations with 20 dB Gaussian noise



(e) 10 dB Gaussian noise



(f) Restorations with 10 dB Gaussian noise

FIGURE 1: Continued.

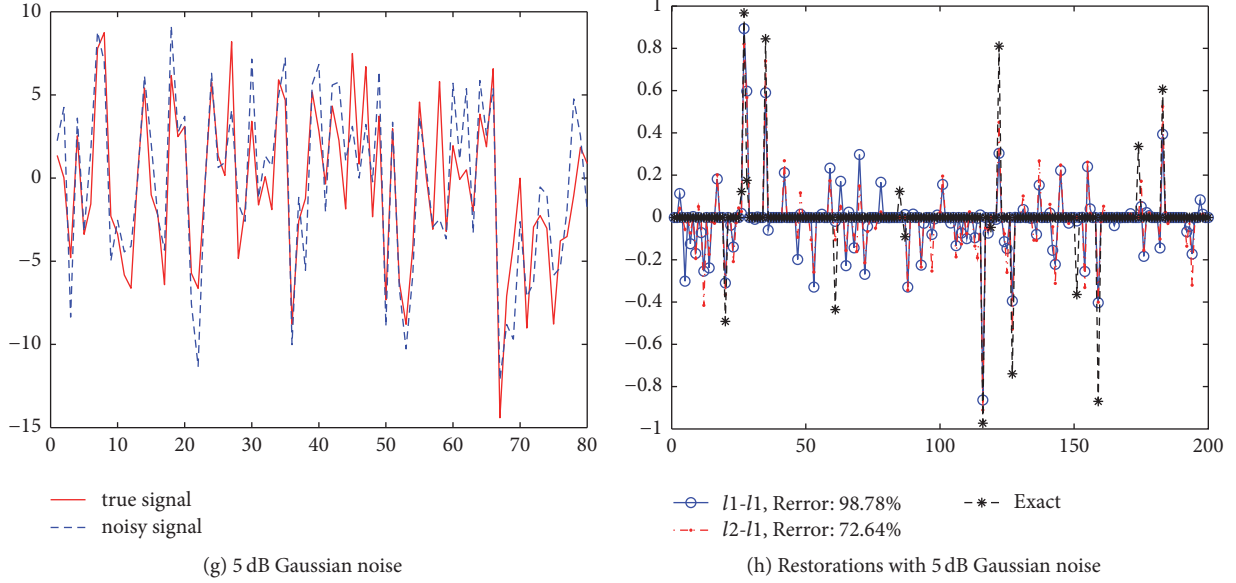


FIGURE 1: Restorations of  $\ell^1$  and  $\ell^2$  fidelities with different Gaussian noise levels.

the proposed method is stable even with large condition numbers. In Section 6.3, we discuss the image inpainting where images are sparse with respect to the Daubechies wavelets. For image inpainting, the linear operator  $A$  is moderate ill-condition and the condition number is around 4000. In order to compare the restoration results, the quality of the computed solution  $x$  is measured by relative error Rerr and PSNR which are, respectively, defined by

$$\begin{aligned} \text{Rerr}(x) &= \frac{\|x - x^\dagger\|}{\|x^\dagger\|} \times 100\%, \\ \text{PSNR}(y^\delta) &= -20 \log_{10} \left( \frac{\|x - x^\dagger\|}{n} \right). \end{aligned} \quad (64)$$

All experiments were performed under Windows 7 and Matlab R2010a on HP ProBook 4431s with Intel Core i5 2410M CPU 2.30 GHz and 4 GB of memory.

**6.1. Comparison with  $\ell^1$  and  $\ell^2$  Fidelity.** This example involves compressive sensing problem  $Ax = y^\delta$ , where matrix  $A_{80 \times 200}$  is random Gaussian and  $y^\delta = Ax^\dagger + \delta$  is the observed data containing white noise or impulsive noise. The true solution  $x^\dagger$  is 16-sparse with respect to natural basis of  $\ell^2$  space which is defined by

$$\varphi_\gamma = e_\gamma = \left( \underbrace{0, 0, \dots, 0}_\gamma, 1, 0, \dots \right). \quad (65)$$

White noise is generated such that data  $y^\delta$  attains a desired SNR, which is defined by

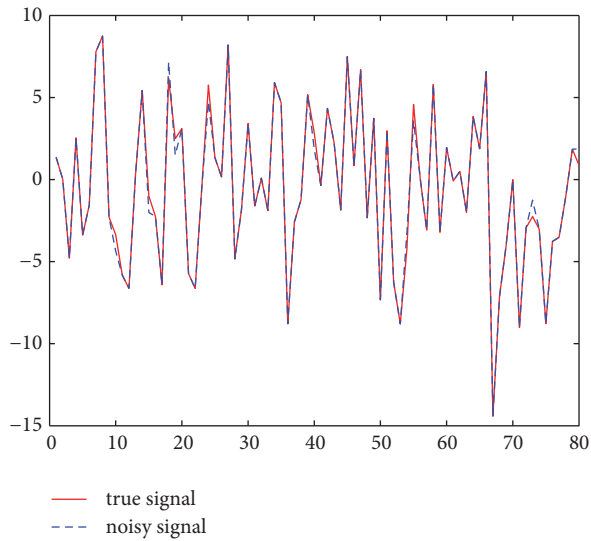
$$\text{SNR}(y^\delta) = 20 \log_{10} \left( \frac{\|y^\delta - E(y^\delta)\|}{\|\delta\|} \right). \quad (66)$$

The impulsive noise is measured by relative error, which is defined by

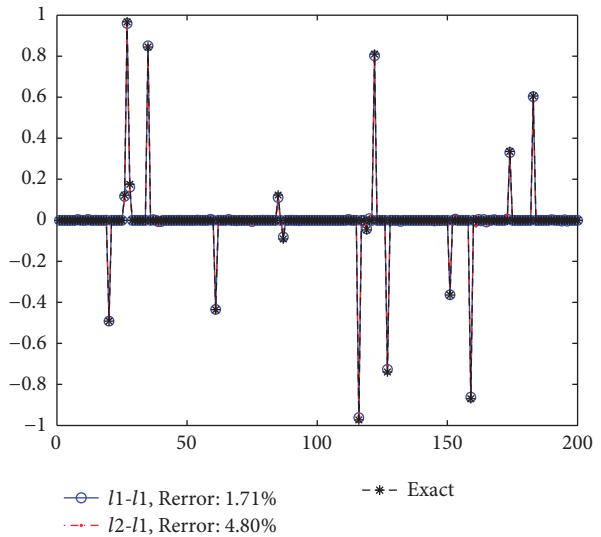
$$\text{Rerr}(\delta) = \frac{\|y - y^\delta\|}{\|y\|} \times 100\%. \quad (67)$$

Figure 1 presents the restoration results by  $\ell^1$  and  $\ell^2$  fidelities. The left column describes the data with different white noise levels. The SNR of data are 30 dB, 20 dB, 10 dB, and 5 dB. The right column is restoration results; the regularization parameters  $(\alpha, \beta)$  are  $(2.37e-3, 1.26e-3)$ ,  $(6.32e-2, 2.63e-2)$ ,  $(4.35e-1, 1.68e-1)$ , and  $(9.05e-1, 7.45e-1)$ . As can be seen from Figure 1, the quality of restoration by  $\ell^1$  and  $\ell^2$  fidelities is similar if the data are contaminated with low noise levels. If data contain high noise levels, the performance of  $\ell^2$  fidelity is better. However, we can not recover high-quality solutions with high noise levels. In Figure 2, the left column describes the data which are contaminated by different impulsive noise levels. Rerr( $\delta$ ) of noise level are 3%, 7%, 15%, and 22%. The value of impulsive noise is  $\pm 1$  at random positions and 0 at other positions. The right column contains restoration results according to different noise levels. The regularization parameters  $(\alpha, \beta)$  are  $(2.43e-2, 1.58e-2)$ ,  $(8.29e-2, 4.63e-2)$ ,  $(2.12e-1, 1.06e-1)$ , and  $(4.25e-1, 2.46e-1)$ . As can be seen from Figure 2, the  $\ell^1$  fidelity is more stable for impulsive noise and always offers high-quality restoration even with poor data. In contrast to the  $\ell^2$  fidelity, the quality of restoration results by the  $\ell^2$  fidelity is always poor.

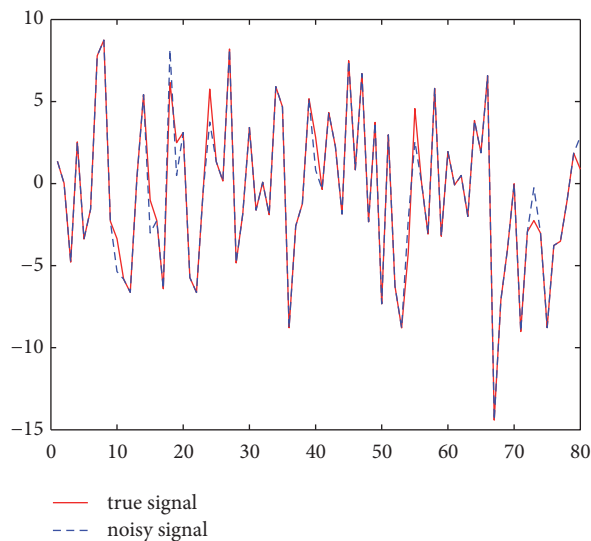
**6.2. Comparison of DSPG with ADM- $\ell^1$ , ADM- $\ell^2$ , and TNIP.** In this subsection, we present the comparison results of DSPG with ADM- $\ell^1$ , ADM- $\ell^2$ , and TNIP. ADM- $\ell^1$  [18] is an efficient alternating direction method for  $\ell^1 + \ell^1$  problem. ADM- $\ell^2$  [18] and TNIP [45] (truncated Newton interior point) are for  $\ell^2 + \ell^1$  problem. In the first experiment, we



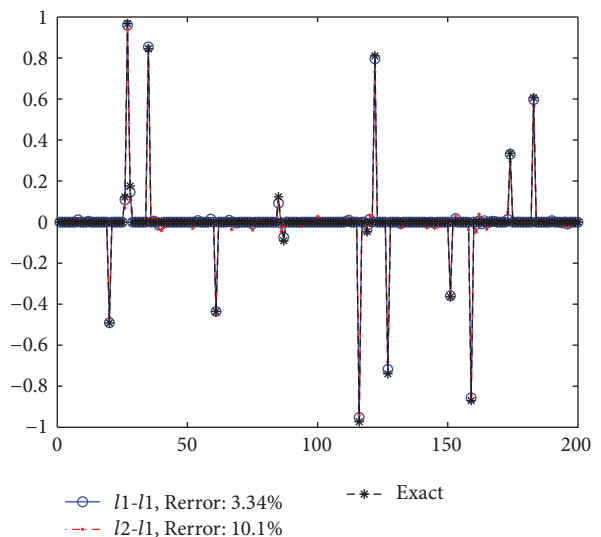
(a) 3% Impulsive noise



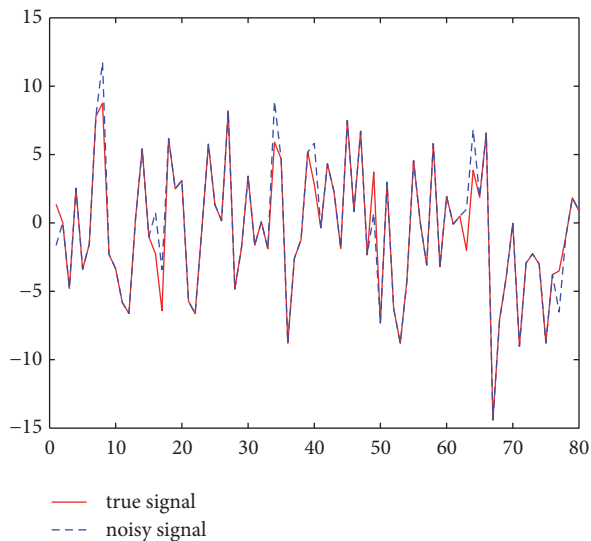
(b) Restorations with 3% Impulsive noise



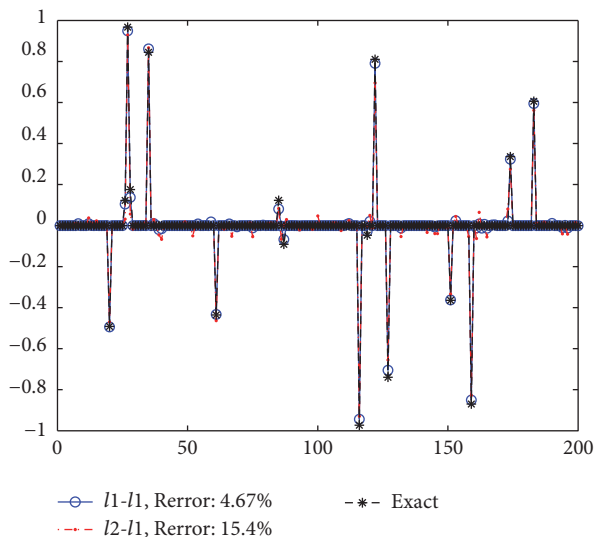
(c) 7% Impulsive noise



(d) Restorations with 7% Impulsive noise



(e) 10% Impulsive noise



(f) Restorations with 10% Impulsive noise

FIGURE 2: Continued.

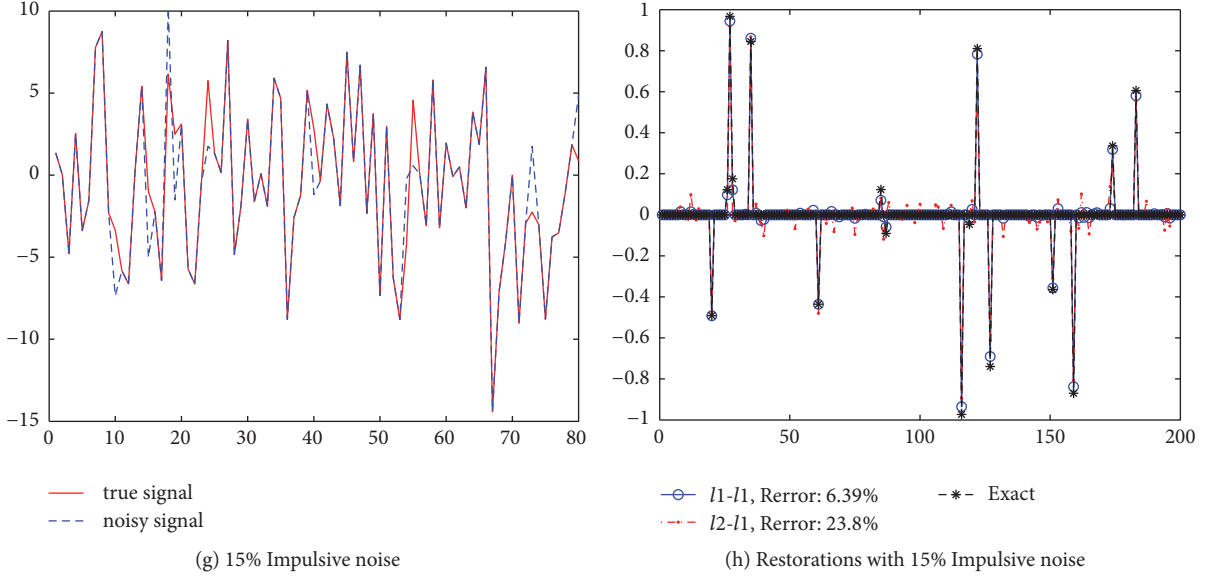
FIGURE 2: Restorations of  $\ell^1$  and  $\ell^2$  fidelities with different impulsive noise levels.

TABLE 1: Comparisons of different algorithms.

Impulsive noise $\delta$	TNIP		AMD- $\ell^2$		AMD- $\ell^1$		DSPG	
	Error	PSNR	Error	PSNR	Error	PSNR	Error	PSNR
0.1%	0.610%	54.49	0.399%	57.97	0.907%	50.84	0.387%	57.28
0.3%	1.412%	47.01	0.559%	55.04	1.813%	44.83	0.548%	55.36
0.5%	2.295%	42.78	1.078%	49.34	2.719%	41.31	0.956%	50.14
1.0%	3.187%	39.92	3.178%	39.93	3.619%	38.82	2.305%	42.65
3.0%	4.993%	36.03	4.956%	36.12	5.542%	35.31	3.063%	40.15
5.0%	8.608%	31.30	8.602%	31.30	9.101%	30.87	5.217%	36.25
10%	17.66%	25.06	17.66%	25.08	18.11%	24.84	12.64%	29.21

use random Gaussian matrix  $A_{m \times n}$ , where sampling length is  $m$  and signal length  $n = 200$ . The condition number of random Gaussian matrix  $A_{m \times n}$  is around 5. The signal is  $p$ -sparsity. We add 1% impulsive noise to true signal. For each fixed pair  $(m, p)$ , we take 100 iterations, where  $m/n = 0.5, 0.4, 0.3, 0.2$  and  $0.1$ ,  $p/m = 0.1$  and  $0.2$ . Figure 3 presents comparison results for four different iterative algorithms. The left column describes convergence rates of  $\text{Rerr}(x)$  for  $p/m = 0.1$ . The right column describes convergence rates of  $\text{Rerr}(x)$  for  $p/m = 0.2$ .  $m/n$  increase from top to bottom row. As can be seen from Figure 3, ADM- $\ell^1$  converges faster than DSPG and the accuracy is better than DSPG when  $m/n = 0.4$ . With  $m/n$  decreasing, DSPG method performs more competitively. The accuracy of DSPG method is even better than ADM- $\ell^1$  when  $m/n = 0.1$ . Though the optimal relative error of ADM- $\ell^1$  is better than DSPG method, the corresponding optimal iteration number or stopping tolerance of ADM- $\ell^1$  is difficult to estimate in practice. The other  $\ell^2$  fidelity algorithms, especially TNIP method, converge obviously faster than ADM- $\ell^1$  and DSPG method. However, the accuracy of the

two algorithms is worse than ADM- $\ell^1$  and DSPG method no matter what value the  $m/n$  is.

Next, in order to test the stability of the DSPG method for ill-conditioned problems, we use matrix  $A_{n \times n}$  ( $n = 200$ ) whose condition number is 255. This problem was discussed by Lorenz in [8] where the ill-conditioned matrix is generated by Matlab code:  $A = \text{tril}(\text{ones}(200))$ . The signal is  $p$ -sparsity where  $p/n = 0.1$  and  $0.2$ . We add 1% impulsive noise to data. As can be seen from Figure 4, DSPG method converges obviously faster than the ADM- $\ell^1$  method. The relative error of DSPG method is also better than ADM- $\ell^1$  method. It is shown that DSPG method is stable even for large condition number matrices. In Table 1, data contain impulsive noise with corruption  $\text{Rerr}(\delta) = 0.1\%, 0.3\%, 0.5\%, 1\%, 3\%, 5\%, 10\%$ . As can be seen from Table 1, the quality of restoration improves as noise level  $\delta$  decreases. Theoretically, ADM- $\ell^1$  and DSPG methods are adept to process impulsive noise. However, ADM- $\ell^1$  method is sensitive to noise when the operators are ill-conditioned. In this case, ADM- $\ell^1$  cannot obtain reasonable restoration. ADM- $\ell^2$  and DSPG methods

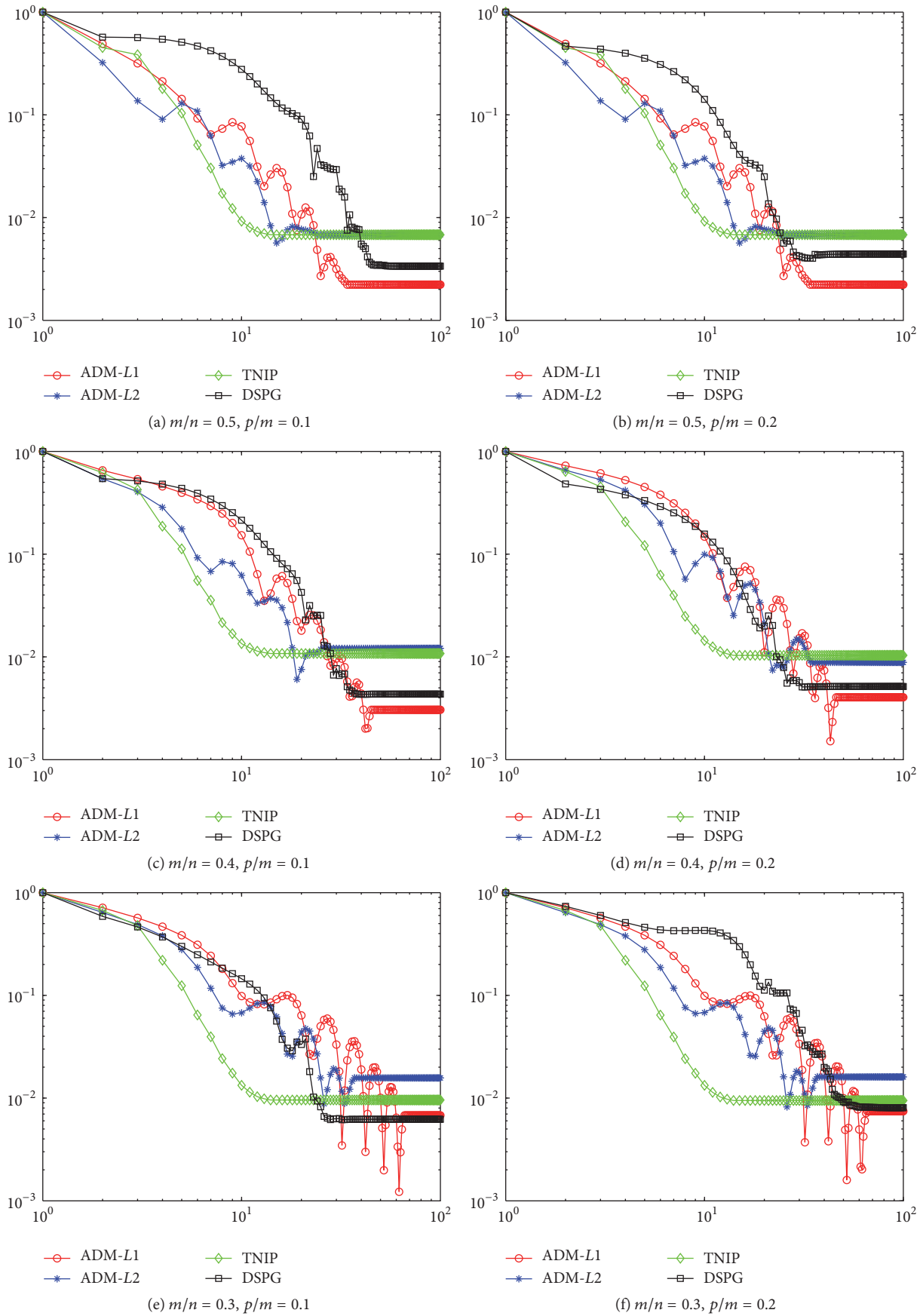


FIGURE 3: Continued.

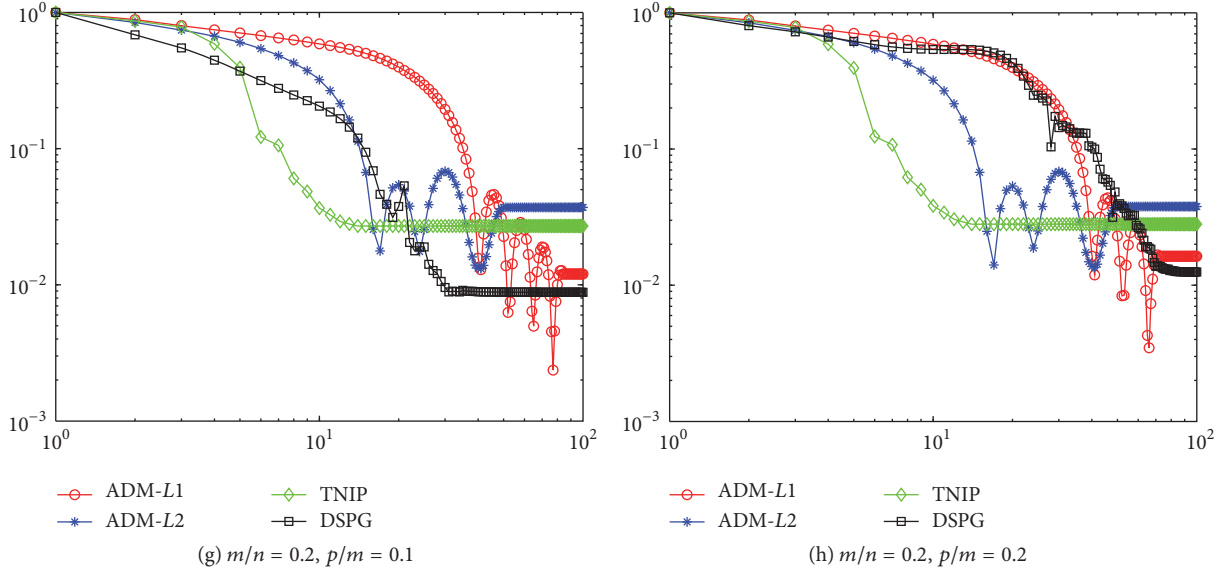
FIGURE 3: Comparisons of DSPG with  $\text{ADM-}\ell^1$ ,  $\text{ADM-}\ell^2$ , and TNIP.

TABLE 2: Restoration of four images with different salt-and-pepper noise levels.

Noise level	Lena	Boat	Barbara	Goldhill	Camerman	Peppers	Mandrill	Pirate
0.05	26.21	25.16	22.16	25.58	25.12	26.25	26.68	25.09
0.08	24.87	23.88	21.52	25.17	25.03	26.11	26.16	25.01
0.10	22.82	21.56	20.35	24.07	24.23	25.01	25.32	24.58
0.15	20.42	19.55	19.23	22.74	22.42	23.16	23.85	22.75

are more stable to noise level  $\delta$  even if matrix  $A$  has large condition numbers. For low noise levels,  $\text{ADM-}\ell^2$  and DSPG methods have advantage over the other two methods. For high noise levels, the restoration results of  $\text{ADM-}\ell^1$  method are very close to  $\text{ADM-}\ell^2$  and TNIP methods. Restoration results of the DSPG method are obviously better than the other three methods.

**6.3. Image Inpainting.** We consider 2D image inpainting problems. The image is Barbara ( $n = 512$ ; cf. Figure 5). We randomly remove 40% pixels of Barbara to create an incomplete image. In this case, the image inpainting is a moderate ill-posed problem. The condition number of the matrix is around 4000. For our purpose, we make use of Daubechies 4 wavelet basis as a dictionary. We use four scales, for a total of  $8192 \times 512$  wavelet and scaling coefficients (cf. Figure 6). As seen from Figure 6, the representation of the image with respect to Daubechies 4 basis is sparse. We add salt-and-pepper noise by Matlab code: `imnoise(image, 'salt & pepper', d)`. In this example,  $d = 0.05$ . The restoration results are shown in Figure 5(d). Restoration results of four images with different noise levels are given in Table 2. Restoration results show that if images have a sparse representation with respect to an orthogonal basis, DSPG method can obtain

satisfactory results even if the image inpainting are moderate ill-posed problems.

## 7. Conclusion

With source and FBI conditions, we have proved that  $\ell^1 + \ell^1$  regularization method yields convergence rates of order  $\delta^{1-\epsilon}$  and  $\delta$ . For numerical solutions, we have proposed a novel DSPG approach for sparsity regularization with  $\ell^1$  fidelity. Numerical results indicate that the proposed DSPG algorithm performs competitively with several state-of-art algorithms such as ADM method. On various classes of test problems with different condition numbers, the proposed DSPG method has exhibited the following advantages: (i) compared with other algorithms, the dual problem of our methods is more simple which is a box-constraint smooth functional and can be solved effectively by projected methods; (ii) for ill-conditioned problems, DSPG method is more stable with respect to noise level. We remark that for well-conditioned problems, for example, compressive sensing, optimal accuracy of  $\text{ADM-}\ell^1$  is better than DSPG method. However, the optimal accuracy of  $\text{ADM-}\ell^1$  method is strongly dependant on stopping tolerance values which can be difficult to estimate in practice. To the best of the author's

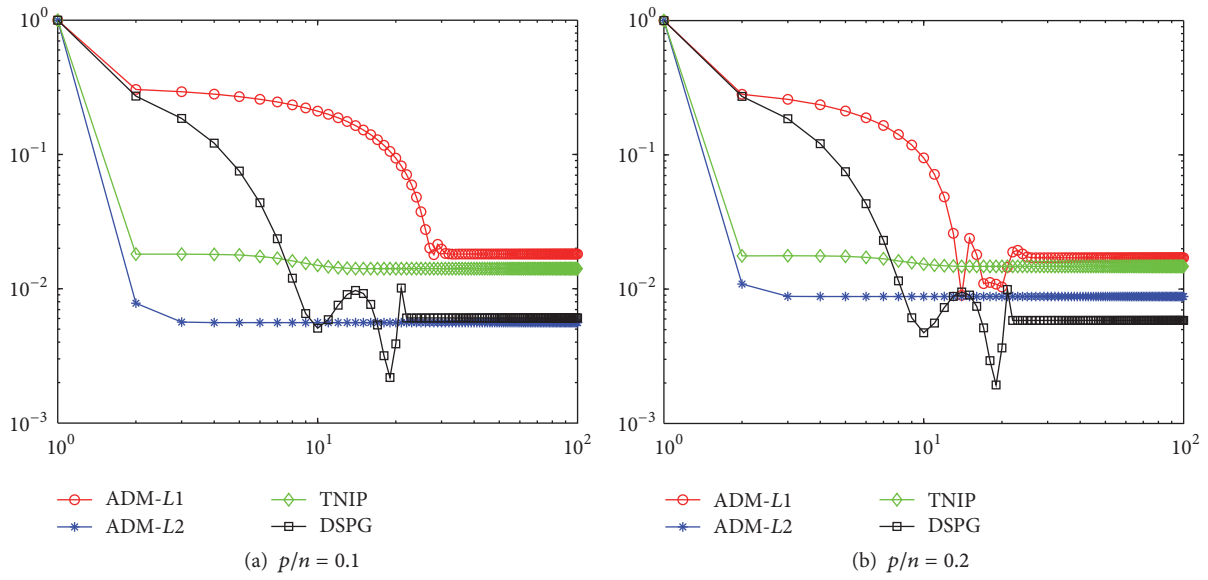


FIGURE 4: Comparisons of DSPG with  $ADM-\ell^1$ ,  $ADM-\ell^2$ , and TNIP.

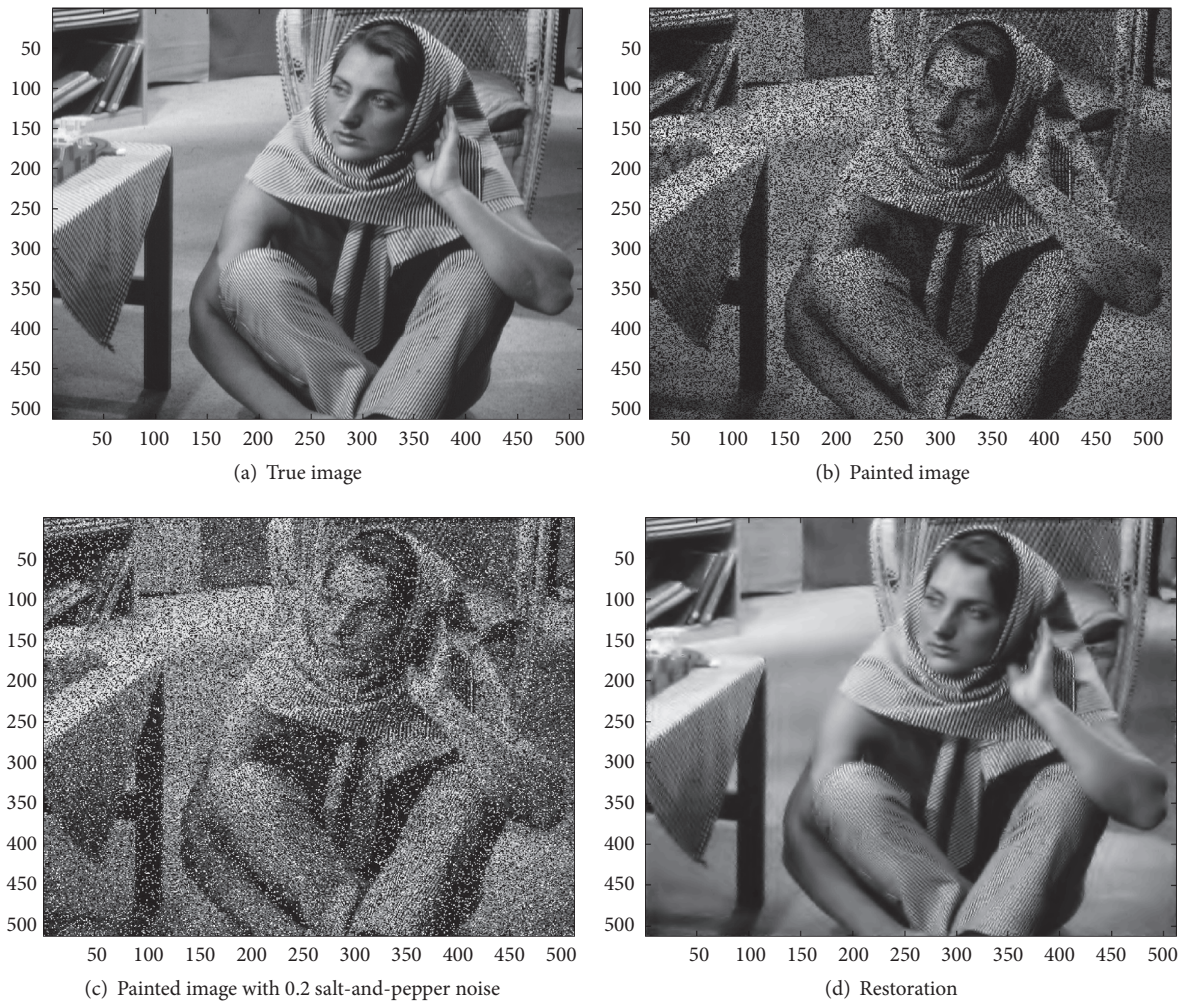


FIGURE 5: Restoration of Barbara with salt-and-pepper noise,  $d = 0.2$ .

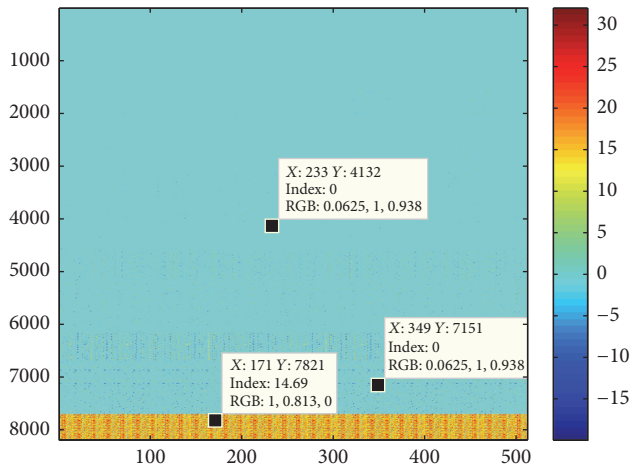


FIGURE 6: Wavelet and scale coefficients.

knowledge, multiparameter regularization is first used to obtain the dual formulation of  $\ell^1 + \ell^1$  problems. One can also try to use multiparameter regularization strategy to solve dual problem of the other nonsmooth functionals, for example,  $\ell^1 + \text{TV}$  problems.

### Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

### Acknowledgments

This work was funded by the National Natural Science Foundation of China (Grants nos. 41304093 and 11301119) and the Fundamental Research Funds for the Central Universities (Grant no. 2572015CB19) and Heilongjiang Postdoctoral Research Develop-Mental Fund (no. LBH-Q16008). The authors are grateful to Professor Wen Song of Harbin Normal University for many useful discussions and suggestions.

### References

- [1] E. J. Candès, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on Pure and Applied Mathematics*, vol. 59, no. 8, pp. 1207–1223, 2006.
- [2] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [3] H. Gao and H. Zhao, "Multilevel bioluminescence tomography based on radiative transfer equation part 1:  $\ell^1$  regularization," *Optics Express*, vol. 18, no. 3, pp. 1854–1871, 2010.
- [4] I. Loris, G. Nolet, I. Daubechies, and F. A. Dahlen, "Tomographic inversion using  $\ell^1$ -norm regularization of wavelet coefficients," *Geophysical Journal International*, vol. 170, no. 1, pp. 359–370, 2007.
- [5] B. Jin and P. Maass, "Sparsity regularization for parameter identification problems," *Inverse Problems*, vol. 28, no. 12, Article ID 123001, 2012.
- [6] I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Communications on Pure and Applied Mathematics*, vol. 57, no. 11, pp. 1413–1457, 2004.
- [7] M. Fornasier, *Theoretical Foundations and Numerical Methods for Sparse Recovery*, De Gruyter, 2010.
- [8] M. Grasmair, "Non-convex sparse regularisation," *Journal of Mathematical Analysis and Applications*, vol. 365, no. 1, pp. 19–28, 2010.
- [9] M. Nikolova, "Description of the minimizers of least squares regularized with  $l(0)$ -norm. Uniqueness of the global minimizer," *SIAM Journal on Imaging Sciences*, vol. 6, no. 2, pp. 904–937, 2013.
- [10] C. A. Zarzer, "On Tikhonov regularization with non-convex sparsity constraints," *Inverse Problems*, vol. 25, no. 2, Article ID 025006, 2009.
- [11] W. Wang, S. Lu, H. Mao, and J. Cheng, "Multi-parameter Tikhonov regularization with the  $l(0)$  sparsity constraint," *Inverse Problems*, vol. 29, no. 6, Article ID 065018, 2013.
- [12] P. J. Rousseeuw and A. M. Leroy, *Robust Regression and Outlier Detection*, John Wiley & Sons, New York, NY, USA, 1987.
- [13] S. Alliney, "Digital filters as absolute norm regularizers," *IEEE Transactions on Signal Processing*, vol. 40, no. 6, pp. 1548–1562, 1992.
- [14] C. Clason, B. Jin, and K. Kunisch, "A duality-based splitting method for  $l_1$ -TV image restoration with automatic regularization parameter choice," *SIAM Journal on Scientific Computing*, vol. 32, no. 3, pp. 1484–1505, 2010.
- [15] M. Nikolova, "Minimizers of cost-functions involving nonsmooth data-fidelity terms. Application to the processing of outliers," *SIAM Journal on Numerical Analysis*, vol. 40, no. 3, pp. 965–994, 2002.
- [16] Y. Dong, M. Hintermüller, and M. Neri, "An efficient primal-dual method for  $l_1$ -TV image restoration," *SIAM Journal on Imaging Sciences*, vol. 2, no. 4, pp. 1168–1189, 2009.
- [17] W. Yin, D. Goldfarb, and S. Osher, "The total variation regularized  $l_1$  model for multiscale decomposition," *Multiscale Modeling and Simulation*, vol. 6, no. 1, pp. 190–211, 2007.
- [18] J. Yang and Y. Zhang, "Alternating direction algorithms for  $l_1$ -problems in compressive sensing," *SIAM Journal on Scientific Computing*, vol. 33, no. 1, pp. 250–278, 2011.
- [19] J. Yang and Y. Zhang, "A primal-dual interior-point framework for using the  $l_1$  or  $l_2$  norm on the data and regularization terms of inverse problems," *SIAM Journal on Scientific Computing*, vol. 33, no. 1, pp. 250–278, 2011.
- [20] S. Stanković, I. Orović, and M. Amin, "L-statistics based modification of reconstruction algorithms for compressive sensing in the presence of impulse noise," *Signal Processing*, vol. 93, no. 11, pp. 2927–2931, 2013.
- [21] L. Ma, J. Yu, and T. Zeng, "Sparse representation prior and total variation-based image deblurring under impulse noise," *SIAM Journal on Imaging Sciences*, vol. 6, no. 4, pp. 2258–2284, 2013.
- [22] J. Shang, Z. Wang, and Q. Huang, "A robust algorithm for joint sparse recovery in presence of impulsive noise," *IEEE Signal Processing Letters*, vol. 22, no. 8, pp. 1166–1170, 2015.
- [23] L. Chen, L. Liu, and C. L. Philip Chen, "A robust bi-sparsity model with non-local regularization for mixed noise reduction," *Information Sciences*, vol. 354, pp. 101–111, 2016.
- [24] M. Burger and S. Osher, "Convergence rates of convex variational regularization," *Inverse Problems*, vol. 20, no. 5, pp. 1411–1421, 2004.



- [25] A. Neubauer, T. Hein, B. Hofmann, S. Kindermann, and U. Tautenhahn, "Improved and extended results for enhanced convergence rates of Tikhonov regularization in Banach spaces," *Applicable Analysis*, vol. 89, no. 11, pp. 1729–1743, 2010.
- [26] M. Grasmair, M. Haltmeier, and O. Scherzer, "Necessary and sufficient conditions for linear convergence of  $l_1$ -regularization," *Communications on Pure and Applied Mathematics*, vol. 64, no. 2, pp. 161–182, 2011.
- [27] J. Flemming and M. Hegland, "Convergence rates in  $l_1$ -regularization when the basis is not smooth enough," *Applicable Analysis*, vol. 94, no. 3, pp. 464–476, 2015.
- [28] C. König, F. Werner, and T. Hohage, "Convergence rates for exponentially ill-posed inverse problems with impulsive noise," *SIAM Journal on Numerical Analysis*, vol. 54, no. 1, pp. 341–360, 2016.
- [29] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, "Least angle regression," *The Annals of Statistics*, vol. 32, no. 2, pp. 407–499, 2004.
- [30] I. Daubechies, R. DeVore, M. Fornasier, and C. S. Güntürk, "Iteratively reweighted least squares minimization for sparse recovery," *Communications on Pure and Applied Mathematics*, vol. 63, no. 1, pp. 1–38, 2010.
- [31] T. Blumensath and M. E. Davies, "Iterative thresholding for sparse approximations," *The Journal of Fourier Analysis and Applications*, vol. 14, no. 5-6, pp. 629–654, 2008.
- [32] T. Blumensath and M. E. Davies, "Iterative hard thresholding for compressed sensing," *Applied and Computational Harmonic Analysis*, vol. 27, no. 3, pp. 265–274, 2009.
- [33] M. Cullen, M. A. Freitag, and S. Kindermann, *Large Scale Inverse Problems: Computational Methods and Applications in the Earth Sciences*, De Gruyter, Austria, 2013.
- [34] Y. Xiao, H. Zhu, and S.-Y. Wu, "Primal and dual alternating direction algorithms for  $l_1$ - $l_1$  norm minimization problems in compressive sensing," *Computational Optimization and Applications*, vol. 54, no. 2, pp. 441–459, 2013.
- [35] S. Lu and S. V. Pereverzev, "Multi-parameter regularization and its numerical realization," *Numerische Mathematik*, vol. 118, no. 1, pp. 1–31, 2011.
- [36] S. Lu, S. V. Pereverzev, Y. Shao, and U. Tautenhahn, "Discrepancy curves for multi-parameter regularization," *Journal of Inverse and Ill-Posed Problems*, vol. 18, no. 6, pp. 655–676, 2010.
- [37] K. Ito, B. Jin, and T. Takeuchi, "Multi-parameter Tikhonov regularization," *Methods & Applications of Analysis*, vol. 18, no. 1, pp. 31–46, 2011.
- [38] O. Scherzer, M. Grasmair, H. Grossauer, M. Haltmeier, and F. Lenzen, *Variational Methods in Imaging*, Springer, 2011.
- [39] T. Schuster, B. Kaltenbacher, B. Hofmann, and K. S. Kazimierski, *Regularization Methods in Banach Spaces*, De Gruyter, 2012.
- [40] V. A. Morozov, "On the solution of functional equations by the method of regularization," *Soviet Mathematics. Doklady*, vol. 7, pp. 510–512, 1966.
- [41] S. G. Mallet, *A Wavelet Tour of Signal Processing*, Elsevier, Oxford, UK, 2010.
- [42] I. Loris, G. Nolet, I. Daubechies, and F. A. Dahlen, "Tomographic inversion using  $l_1$ -norm regularization of wavelet coefficients," *Geophysical Journal International*, vol. 170, no. 1, pp. 359–370, 2007.
- [43] I. Ekeland and R. Man, *Convex Analysis and Variational Problems*, SIAM, Philadelphia, Pa, USA, 1999.
- [44] E. G. Birgin, J. M. Martinez, and M. Raydan, "Nonmonotone spectral projected gradient methods on convex sets," *SIAM Journal on Optimization*, vol. 10, no. 4, pp. 1196–1211, 2000.
- [45] S. J. Kim, K. Koh, and M. Lustig, "A method for large-scale  $l_1$ -regularized least squares," *IEEE Journal on Selected Topics in Signal Processing*, vol. 1, no. 4, pp. 606–617, 2007.



# Hindawi

Submit your manuscripts at  
<https://www.hindawi.com>

