Dual Splitting Method For Sparsity Signal Restoration With Impulsive Noise

Liang Ding and Lei Xu

Department of Mathematics Northeast Forestry University 26 Hexing Road, Harbin, China dl@nefu.edu.cn; mathding@163.com

Received December, 2016; revised May, 2017

ABSTRACT. We consider sparsity signal restoration with impulsive noise by sparsity regularization. It is challenging due to the fact that both fidelity and regularization term lack of differentiability. Moreover, for ill-conditioned problems, sparsity regularization is often unstable. We propose a novel dual splitting method and show that one can overcome the non-differentiability and instability by adding a smooth splitting ℓ^2 regularization term to the original optimization functional. The functional is split into an $\ell^1 + \ell^2$ part and an $\ell^2 + \ell^1$ part. The advantage of the proposed method is that the convex duality reduced to a constraint smooth functional which can be solved easily by projected gradient method. Moreover, it is stable even for ill-conditioned problems. Some experiments are performed, using compressed sensing and image inpainting, to demonstrate the efficiency of the proposed approach.

Keywords: Sparsity signal, Impulsive noise, Inversion, Duality, Splitting

1. Introduction. In the present manuscript we are concerned with ill-posed linear operator equation

$$Ax = y, (1)$$

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

$$\|y - y^{\delta}\| \le \delta \tag{2}$$

is available. We call y^{δ} the noisy data and δ the noise level. It is well known that the conventional method for solving eq.(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], etc. For accounts of the regularizing properties and computational techniques in sparsity regularization we refer the reader to [5, 6, 7, 8, 9, 10] and the references given there. In general, sparsity regularization is given by

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

where $\|x\|_{w,p}^{p} = \sum_{\gamma} \omega_{\gamma} |\langle \varphi_{\gamma}, x \rangle|^{p} (1 \le p \le 2), \alpha$ is the regularization parameter balancing the fidelity $\|Ax - y^{\delta}\|_{\ell^{2}}^{2}$ and regularization term $\|x\|_{w,p}^{p}$. The functional in eq.(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 references[11, 12, 13, 14] 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

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

We call eq.(4) $\ell^1 + \ell^1$ problem. A main motivation to investigate the $\ell^1 + \ell^1$ problem is that noisy data y^{δ} often contain impulsive noise. For Gaussian noise, ℓ^2 fidelity is a natural choice. However, a typical non-differentiable fidelity used in application involving impulsive noise is the ℓ^1 fidelity, which is more robust than ℓ^2 fidelity[15].

Nowadays ℓ^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 ℓ^1 fidelity combined with sparsity regularization have been studied. We refer the reader to [17] and the references given there.

Though ℓ^1 fidelity is robust, more researchers prefer to use ℓ^2 fidelity because of its differentiability. Hence a key issue for the ℓ^1 fidelity is the numerical computing methods. In the past few years, numerous algorithms have been systematically proposed for the $\ell^1 + TV$ problems. On the other hand, in spite of growing interests in the ℓ^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, e.g. homotopy (LARS) method^[19], iteratively reweighted least squares (IRLS) method [20] and iterative thresholding algorithm [21, 22] cannot be directly applied to $\ell^1 + \ell^1$ problem due to the fact that both fidelity and regularization term lack of differentiability. Only a few papers, in which numerical algorithms for $\ell^1 + \ell^1$ problems have been discussed systematically. For ill-conditioned problems, these methods are often unstable^[23][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[18], Borsic and Adler proposed a Primal Dual-Interior Point Methods (PD-IPM) for EIT problem, which is efficient at dealing with the non-differentiability. However, they didn't give the convergence proof. Yang, Zhang and Yin reformulated the $\ell^1 + \ell^1$ problem into the basis pursuit model which can be solved effectively by ADM method[17]. It is a competitive method compared with other algorithms for compressive sensing. In [29], Xiao, Zhu and Wu applied ADM method to $\ell^1 + \ell^1$ problem directly, Numerical results illustrated that the proposed algorithm performs better than Yall1[17].

In this paper, we investigate numerical method for $\ell^1 + \ell^1$ problems. As above mentioned, dual is a conventional technique to solve the Tikhonov regularization with ℓ^1 fidelity. However, there are some limitations to this approach to $\ell^1 + \ell^1$ problems due to the fact that it's difficult to obtain the dual formulation of $\ell^1 + \ell^1$ problems. Inspired by [16], a smooth spliting ℓ^2 term is added to original functional of regularization. The functional is split into an $\ell^1 + \ell^2$ part and an $\ell^2 + \ell^1$ part. The dual problem of this new cost functional is reduced to a constraint smooth functional. Moreover, the smooth splitting ℓ^2 regularization term can improve the stability. We use projected gradient method to seek for the minimizers of the constraint functional.

An outline of this paper is as follows. We devote Section 2 to introduce some notations and preliminaries. In Section 3, inspired by the theory of duality, we construct a new

946

functional and investigate the convergence of the minimizers to the functional. In Section 4, we show that the Frechel duality of the functional is equal to a constraint smooth functional and the projected gradient method can be used to compute the minimizers. Numerical experiments involving compressed sensing and image inpainting are presented in Section 5, showing that our proposed approaches are robust and efficient.

2. Notation and Preliminaries. The specific notations will be introduced and recalled in the following sections. For the approximate solutions of Ax = y, we consider the minimization of the regularization functional

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

where $R(x) := \sum_{\gamma} \omega_{\gamma} |\langle \varphi_{\gamma}, x \rangle|$, the subdifferential of R(x) at x is denoted by $\partial R(x) \subset X$.

All along this paper, X and Y denote Hilbert space which is a subspace of ℓ^2 space and $\langle \cdot, \cdot \rangle$ denotes the inner product. $A : \operatorname{dom}(A) \subseteq X \to Y$ is a bounded linear operator and $\operatorname{dom}(A) \cap \operatorname{dom}(R) \neq \emptyset$. We denote with A^* the adjoint of the operator A. We will always work on real vector spaces, hence, in finite dimensions, A^* usually coincides with the transposed matrix of A. $(\varphi_{\gamma})_{\gamma \in \Lambda} \subset X$ is an orthonormal basis where Λ is some countable index set. From now, we denote

$$x_{\gamma} = \langle x, \varphi_{\gamma} \rangle,$$

$$\|x\|_{\ell^{p}} = \left(\sum_{\gamma} |\langle \varphi_{\gamma}, x \rangle|^{p}\right)^{\frac{1}{p}} = \left(\sum_{\gamma} |x_{\gamma}|^{p}\right)^{\frac{1}{p}}, 0$$

and

$$||x||_{\ell^{\infty}} = \operatorname{Max}_{\gamma}|x_{\gamma}|, p = +\infty$$

Associated to these norms we denote their unit balls by

$$B_{\ell^{\infty}}(1) = \{ p \in X : \|p\|_{\ell^{\infty}} \le 1 \}$$

and the balls of radius α by

$$B_{\ell^{\infty}}(\alpha) = \{q \in Y : \|q\|_{\ell^{\infty}} \le \alpha\}$$

We denote by x_{α}^{δ} 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 2.1. An element x^{\dagger} is called a R(x)-minimum norm solution of linear problem Ax = y if

$$Ax^{\dagger} = y \text{ and } R(x^{\dagger}) = \min\{R(x) | Kx = y\}.$$

We define the sparsity as follows:

Definition 2.2. $x \in X$ is sparse with respect to $\{\varphi_{\gamma}\}_{\gamma \in \Lambda}$ in the sense that $\operatorname{supp}(x) := \{\gamma \in \Lambda : \langle \varphi_{\gamma}, x \rangle \neq 0\}$ is finite. If $\|\operatorname{supp}(x)\|_0 = s$ for some $s \in \mathbb{N}$, the $x \in X$ is called *s*-sparse.

3. **Primal Problem.** We consider the splitting of $\ell^1 + \ell^1$ problem. let ω_{γ} in (4) take the same value, i.e. $\omega_{\gamma} = \mu > 0$ for all $\gamma \in \Lambda$. It is reasonable because convergence can be obtained when $\frac{\delta}{\alpha} \to 0[6]$. Let $\alpha := \alpha \mu = \alpha \omega_{\gamma}$, then (4) is equivalent to

$$\min_{x} \| Ax - y^{\delta} \|_{\ell^{1}} + \alpha \sum_{\gamma} | \langle \varphi_{\gamma}, x \rangle |.$$
(6)

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 $D : \ell^2 \to \ell^2$ a dictionary which 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[24, 25]. Let $K = A \circ D^T$, then (4) is equivalent to

$$\mathcal{P}: \min_{u} \{ J_{\alpha}(u) = \| Ku - y^{\delta} \|_{\ell^{1}} + \alpha \| u \|_{\ell^{1}} \}.$$
(7)

Dual is a popular technique to solve Tikhonov regularization with ℓ^1 fidelity. However, there are some limitations to this approach to solve (7). The main difficulty is that both the ℓ^1 fidelity and the ℓ^1 regularization term are non-differentiable. Moreover, for ill-conditioned problems, sparsity regularization is often unstable. We add the smooth splitting penalty $\frac{1}{2\beta} \parallel u - v \parallel^2_{\ell^2}$ to (7) to construct the following functional

$$\mathcal{P}_{\beta} : \min_{u} \{ J_{\alpha,\beta}(u) = \| Ku - y^{\delta} \|_{\ell^{1}} + \alpha \| v \|_{\ell^{1}} + \frac{1}{2\beta} \| u - v \|_{\ell^{2}}^{2} \}.$$
(8)

The advantage of problem (8) in place of (7) is that the dual problem of (8) is a constraint smooth functional and projected gradient algorithm can be used to compute minimizers. Moreover, the regularization effect of ℓ^1 penalty is weak, ℓ^2 penalty can improve the stability of (7). Next we will investigate the convergence of the minimizers to the functional \mathcal{P}_{β} as β tends to zero.

Theorem 3.1. Let α be fixed and $\{\beta_k\}$ be a sequence converge to zero. Then the minimizers $\{(u_{\beta_k}, v_{\beta_k})\}$ of the problem (8) has a subsequence converging to (u^*, v^*) with $u^* = v^*$ being a minimizer of the problem (7).

Proof: Let u^{\dagger} be a global minimizer of the problem (7). By the definition of $(u_{\beta_k}, v_{\beta_k})$, we have

$$J_{\alpha,\beta_{k}}(u_{\beta_{k}},v_{\beta_{k}}) = \| Ku_{\beta_{k}} - y^{\delta} \|_{\ell^{1}} + \alpha \| u_{\beta_{k}} \|_{\ell^{1}} + \frac{1}{2\beta_{k}} \| u_{\beta_{k}} - v_{\beta_{k}} \|_{\ell^{2}}^{2}$$

$$\leq \| Ku^{\dagger} - y^{\delta} \|_{\ell^{1}} + \alpha \| u^{\dagger} \|_{\ell^{1}} < \delta + \alpha \| u^{\dagger} \|_{\ell^{1}},$$
(9)

hence that

$$\parallel u_{\beta_k} \parallel_{\ell^1} < \frac{\delta}{\alpha} + \parallel u^{\dagger} \parallel_{\ell^1}$$
(10)

and

$$\| v_{\beta_k} \|_{\ell^2}^2 < 2\beta_k \frac{\delta}{\alpha} + 2\beta_k \| u^{\dagger} \|_{\ell^1}$$

$$\tag{11}$$

Since α is a fixed value and $\beta_k \to 0$, there exists a constant C > 0 such that

$$\| v_{\beta_k} \|_{\ell^2}^2 < 2\beta_k \frac{\delta}{\alpha} + 2\beta_k \| u^{\dagger} \|_{\ell^1} < C$$

$$\tag{12}$$

It follows that $\{(u_{\beta_k}, v_{\beta_k})\}$ is uniformly bounded. Therefore, a subsequence $(u_{\beta_n}, v_{\beta_n})$ of $(u_{\beta_k}, v_{\beta_k})$ and (u^*, v^*) exist such that

$$(u_{\beta_n}, v_{\beta_n}) \longrightarrow (u^*, v^*).$$
(13)

Splitting Method For Sparsity regularization

From (9), we have

$$\| u_{\beta_{k}} - v_{\beta_{k}} \|_{\ell^{2}}^{2} \leq 2\beta_{k} [\| Ku^{\dagger} - y^{\delta} \|_{\ell^{1}} + \alpha \| u^{\dagger} \|_{\ell^{1}} - (\| Ku_{\beta_{k}} - y^{\delta} \|_{\ell^{1}} + \alpha \| u_{\beta_{k}} \|_{\ell^{1}})].$$

$$(14)$$

By the weak lower semicontinuity of the norm, we obtain that

$$\| u^{*} - v^{*} \|_{\ell^{2}}^{2} \leq \liminf_{u} \| u_{\beta_{k}} - v_{\beta_{k}} \|_{\ell^{2}}^{2} \leq \lim_{k \to \infty} 2\beta_{k} [\| Ku^{\dagger} - y^{\delta} \|_{\ell^{1}} + \alpha \| u^{\dagger} \|_{\ell^{1}} - (\| Ku_{\beta_{k}} - y^{\delta} \|_{\ell^{1}} + \alpha \| u_{\beta_{k}} \|_{\ell^{1}})] = 0.$$

$$(15)$$

Therefore, we deduce that $u^* = v^*$. From (9), we obtain

$$\| Ku^{*} - y^{\delta} \|_{\ell^{1}} + \alpha \| u^{*} \|_{\ell^{1}} \leq \lim_{k \to \infty} \| Ku_{\beta_{k}} - y^{\delta} \|_{\ell^{1}}$$

+
$$\lim_{k \to \infty} \alpha \| u_{\beta_{k}} \|_{\ell^{1}} + \lim_{k \to \infty} \frac{1}{2\beta_{k}} \| u_{\beta_{k}} - v_{\beta_{k}} \|_{\ell^{2}}^{2} \leq \| Ku^{\dagger} - y^{\delta} \|_{\ell^{1}} + \alpha \| u^{\dagger} \|_{\ell^{1}}$$
 (16)

Therefore $u^* = v^*$ being a minimizer of the problem (7) and the proof is complete.

4. Dual Problem and Computation of Minimizers. It is challenging to solve (8) directly due to the fact that two ℓ^1 term lack of differentiability. Next we consider the dual problem of \mathcal{P}_{β} . We will show that the constraint smooth minimization problems \mathcal{P}^*_{β}

$$\mathcal{P}_{\beta}^{*}: \begin{cases} \min\frac{\beta}{4} \parallel K^{*}p \parallel_{\ell^{2}}^{2} + \frac{\beta}{4} \parallel q \parallel_{\ell^{2}}^{2} - \langle p, y^{\delta} \rangle, \\ s.t. \parallel p \parallel_{\ell^{\infty}} \leq 1, \parallel q \parallel_{\ell^{\infty}} \leq \alpha, K^{*}p - q = 0. \end{cases}$$
(17)

is the dual problem of \mathcal{P}_{β} . The duality is a constraint smooth functional which could be solved easily by projected gradient method.

Theorem 4.1. \mathcal{P}^*_{β} is the dual problem of \mathcal{P}_{β} . The solutions (u_{β}, v_{β}) of \mathcal{P}_{β} and (p_{β}, q_{β}) of \mathcal{P}^*_{β} have the following relation

$$\begin{cases} \beta K^* p_{\beta} = u_{\beta} - v_{\beta}, \\ -\beta q_{\beta} = -u_{\beta} + v_{\beta}, \\ \langle K u_{\beta} - y^{\delta}, p - p_{\beta} \rangle \ge 0, \\ \langle v_{\beta}, q - q_{\beta} \rangle \ge 0 \end{cases}$$
(18)

for all $\parallel p \parallel_{\ell^{\infty}} \leq 1$, $\parallel q \parallel_{\ell^{\infty}} \leq \alpha$.

Proof: Let

$$F(u,v) = \frac{1}{2\beta} \| u - v \|_{\ell^2}^2, R(u) = \| u - y^{\delta} \|_{\ell^1} + \alpha \| v \|_{\ell^1}, \Lambda(u,v) = (Ku,v).$$
(19)

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

$$\inf_{u \in \ell^2} F(u, v) + R(\Lambda(u, v)).$$
(20)

Let us denote by F^* and R^* the conjugate function of F and R. By the Fenchel duality[26][Chap 3,Chap 10], it follows that

$$F^{*}(p,q) = \begin{cases} \frac{\beta}{4} \parallel p \parallel_{\ell^{2}}^{2} + \frac{\beta}{4} \parallel q \parallel_{\ell^{2}}^{2}, p+q=0, \\ \infty, \ else \end{cases}$$
(21)

949

L. Ding, and L. Xu

and

950

$$R^*(p) = \begin{cases} \langle p, y^{\delta} \rangle_{\ell^2}, if \parallel p \parallel_{\ell^{\infty}} \leq 1, \parallel q \parallel_{\ell^{\infty}} \leq \alpha \\ \infty, \ else. \end{cases}$$
(22)

In $J_{\alpha,\delta}(x)$, the functional F and R are proper, convex lower semicontinuous and continuous at 0. By the Fenchel duality theorem[26](Chap 3, Prop.2.4, Prop.4.1, Rem.4.2), it is easy to show that

$$\inf_{(u,v)\in\ell^2} F(u,v) + R(\Lambda(u,v)) = \sup_{(p,q)\in\ell^2} -F^*(\Lambda^*(p,q)) - R^*(-p,-q).$$
(23)

i.e.

$$\inf_{(u,v)\in\ell^2} F(u,v) + R(\Lambda(u,v)) = -\inf_{(p,q)\in\ell^2} F^*(\Lambda^*(p,q)) + R^*(-p,-q).$$
(24)

And the right side of (24) has at least one solution (p_{β}, q_{β}) . Since (u_{β}, v_{β}) and (p_{β}, q_{β}) are minimizers of \mathcal{P}_{β} and \mathcal{P}_{β}^{*} , we have that

$$F(u_{\beta}, v_{\beta}) + R(\Lambda(u, v)) + F^*(\Lambda^*(p_{\beta}, q_{\beta})) + R^*(-p_{\beta}, -q_{\beta}) = 0.$$
(25)

Moreover, the extremality condition (25) is equivalent to the Kuhn-Tucker conditions

$$(-p_{\beta}, -q_{\beta}) \in \partial R(\Lambda(u_{\beta}, v_{\beta})) \quad and \quad \Lambda^*(p_{\beta}, q_{\beta}) \in \partial F(u_{\beta}, v_{\beta}).$$
 (26)

By the definition of the subgradient and $(-p_{\beta}, -q_{\beta}) \in \partial R(\Lambda(u_{\beta}, v_{\beta}))$ and $\Lambda^*(p_{\beta}, q_{\beta}) \in \partial F(u_{\beta}, v_{\beta})$, it follows that

$$Ku_{\beta} - y^{\delta} \in \partial I_{\{\|-p_{\beta}\|_{\ell^{\infty}} \le 1\}}$$

$$\tag{27}$$

and

$$v_{\beta} \in \partial I_{\{\|-q_{\beta}\|_{\ell^{\infty}} \le \alpha\}}.$$
(28)

Then we have

$$\begin{cases} \langle Ku_{\beta} - y^{\delta}, p - p_{\beta} \rangle \ge 0, \\ \langle v_{\beta}, q - q_{\beta} \rangle \ge 0 \end{cases}$$
(29)

By the definition of the subgradient and $\Lambda^*(p_\beta, q_\beta) \in \partial F(u_\beta, v_\beta)$, it follows that

$$\begin{cases} \beta K^* p_\beta = u_\beta - v_\beta, \\ -\beta q_\beta = -u_\beta + v_\beta \end{cases}$$
(30)

and the proof is completed.

As above mentioned, the functional (8) can be transformed using Fenchel duality into a smooth functional with a box constraint, which is easy to compute. For linear equality constraint $K^*p - q = 0$, we consider an augmented Lagrangian formulation of this constraint, which can be solved efficiently using classical Arrow-Hurwicz method proposed by[27, 28]. The augmented Lagrangian functional $f(p, q, \lambda)$ is defined by

$$f(p,q,\lambda) = \frac{\beta}{4} \parallel K^*p \parallel_{\ell^2}^2 + \frac{\beta}{4} \parallel q \parallel_{\ell^2}^2 - \langle p, y^\delta \rangle + \langle \lambda, K^*p - q \rangle + \frac{c}{2} \parallel K^*p - q \parallel_{\ell^2}^2, \quad (31)$$

then the frechét derivative of f with respect to p, q and λ are

$$f_p(p,q,\lambda) = \frac{\beta}{2}KK^*p - y^{\delta} + \langle \lambda, K^* \rangle + cK(K^*p - q), \qquad (32)$$

$$f_q(p,q,\lambda) = \frac{\beta}{2}q - y^{\delta} - \lambda - c(K^*p - q), \qquad (33)$$

and

$$f_{\lambda}(p,q,\lambda) = K^* p - q. \tag{34}$$

By setting unit balls by

$$B_{\ell^{\infty}}(1) = \{ p \in X : \|p\|_{\ell^{\infty}} \le 1 \}$$

and the balls of radius α by

$$B_{\ell^{\infty}}(\alpha) = \{q \in Y : \|q\|_{\ell^{\infty}} \le \alpha\}$$

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

$$P_{B(c)}(x)(i,j) = c \frac{x(i,j)}{\max(c, |x(i,j)|)}.$$
(35)

Given initial value p_0 , q_0 and λ_0 and a constraint on step length $\gamma_p \gamma_q$ and γ_{λ} , i.e.

$$0 < \gamma_{\min} \le \gamma_p, \gamma_q, \gamma_\lambda \le \gamma_{\max}.$$
(36)

Let step length $\sigma \in (0, 1)$. We use the convergence criteria given by

$$\| P_{B(1)}(p_k + \gamma_p d_p^k) - p_k \|_{\ell^2} \leq \varepsilon$$

$$(37)$$

and

$$\| P_{B(\alpha)}(q_k + \gamma_q d_q^k) - q_k \|_{\ell^2} \leq \varepsilon.$$
(38)

The algorithm is given as follows:

Algorithm 1 Projected Gradient method for \mathcal{P}_{β}^{*} 1: Set $p_{0} = q_{0} = \lambda_{0} = 0$; set step sizes $\gamma_{p}, \gamma_{q}, \gamma_{\lambda}, c$ and σ ; k=1, 2: Compute $d_{p}^{k} = f_{p}(p_{k-1}, q_{k-1}, \lambda_{k-1}),$ $d_{q}^{k} = f_{p}(p_{k-1}, q_{k-1}, \lambda_{k-1}),$ $d_{\lambda}^{k} = f_{p}(p_{k-1}, q_{k-1}, \lambda_{k-1}),$ 3: Compute $p_{k} = P_{B(1)}(p_{k-1} + \gamma_{p}d_{p}^{k}),$ $q_{k} = P_{B(\alpha)}(q_{k-1} + \gamma_{q}d_{q}^{k}),$ $\lambda_{k} = \lambda_{k-1} + \sigma\gamma_{\lambda}d_{\lambda}^{k},$ 4: k = k + 1,Until the convergence criteria obtained or $k = k_{\text{max}}.$

5. Numerical Simulations. In this section, we present some numerical experiments to illustrate the efficiency of the proposed method. In Section 5.1, numerical experiments involve compressive sensing. We aim to demonstrate that ℓ^1 fidelity is more stable than the ℓ^2 fidelity and is capable of handing impulsive noises. In Section5.2, we compare the performance of the splitting method with the alternating direction method (ADM) and TNIP method. We discuss an ill-posed problem where the condition number of linear operator A is 255, we aim to demonstrate that the proposed method is stable. In Section 5.3, we discuss the image inpairing 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

$$\operatorname{Rerr}(x) = \frac{\|x - x^{\dagger}\|}{\|x^{\dagger}\|} \times 100\%$$
(39)

and

$$PSNR(y^{\delta}) = -20\log_{10}(\frac{\|x - x^{\dagger}\|}{n}).$$
(40)

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

5.1. Comparison with ℓ^1 and ℓ^2 fidelity. This example involves compressive sensing problem

$$Ax = y^{\delta} \tag{41}$$

where matrix $A_{80\times 200}$ is random Gaussian, $y^{\delta} = Ax^{\dagger} + \delta$ is the observed data containing white noise or impulsive noise. The true solution x^{\dagger} is 16-sparse with respect to natural basis of ℓ^2 space which is defined by

$$\varphi_{\gamma} = e_{\gamma} = (\underbrace{0, 0, \cdots, 0, 1}_{\gamma}, 0, \cdots).$$
(42)

White noise is generated such that data y^{δ} attains a desired SNR, which is defined by

$$\operatorname{SNR}(y^{\delta}) = 20 \log_{10}\left(\frac{\|y^{\delta} - E(y^{\delta})\|}{\|\delta\|}\right)$$
(43)

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

$$\operatorname{Rerr}(\delta) = \frac{\|y - y^{\delta}\|}{\|y\|} \times 100\%$$
(44)

Figure 1 shows the performance of the ℓ^1 and ℓ^2 fidelities with different impulsive noise levels. The left column describes the data which are contaminated by different impulsive noise levels. The Rerr(δ) of noise level are 3%, 7%, 15% and 22%. The value of impulsive noise is ±1 at random positions and 0 at other positions. The right column contains restoration results according to different noise levels. As can be seen from Figure 1 the ℓ^1 fidelity is more stable for impulsive noise and always offer high quality restoration even with poor data. In contrast to the ℓ^2 fidelity, the quality of restoration results by the ℓ^2 fidelity is always poor.

5.2. Comparison of Splitting with ADM- ℓ^2 and TNIP. in order to test the stability of the splitting method for ill-conditioned problems, we use matrix $A_{n\times n}$ (n=200) whose condition number is 255. This problem was discussed by Lorentz in [10] where the illconditioned matrix is generated by Matlab code: "A=tril(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 Fig.2, splitting converge obviously faster than the ADM- ℓ^1 method. The relative error of splitting method is also better than ADM- ℓ^1 method. It is shown that splitting method is stable even for large condition number matrices. Theoretically, ADM- ℓ^1 and splitting method are adept to process impulsive noise. However, ADM- ℓ^1 method is sensitive to noise when the operators are ill-conditioned. In this case, ADM- ℓ^1 cannot obtain reasonable restoration. Splitting methods are more stable to noise level δ even if matrix K has large condition numbers. Restoration results of the splitting method are obviously better than the other two methods.

5.3. Image inpainting. We present the comparison results of splitting algorithm with ADM algorithm by 2D image inpainting problems. The image is Lena(n=128; cf. Fig.3). We remove eight vertical grids and eight horizontal grids pixels of Lena to create an incomplete image. In this case, the image inpainting is a moderate ill-conditioned 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 8192×512 wavelet and scaling coefficients(cf.Fig.3). As seen from Fig.3, the representation of the image with respect to Daubechies 4 basis is sparse. We add impulsive noise by Matlab code "imnoise(image, 'salt & pepper', d)". In the example, d = 0.02, the restoration results are shown in Fig.3. In this case, the operator K is moderate ill-conditioned,

performance of splitting is obviously better than ADM. Restoration results(PSNR) of ten classical images in image processing by ADM and splitting methods are given in Table.1. Restoration results show that if images have a sparse representation with respect to an orthogonal basis, splitting method are competitive, which can obtain satisfactory results even if the image inpainting are moderate ill-posed problems.

noise	Lena	Babara	boat	goldhill	cameraman
level	ADM Splitting	ADM Splitting	ADM Splitting	ADM Splitting	ADM Splitting
0.01	23.15 25.45	21.28 22.52	23.84 25.38	23.14 25.26	23.74 25.17
0.02	22.45 24.23	$20.52\ 21.73$	22.48 23.98	22.23 24.46	22.34 23.86
0.03	20.23 22.17	$17.46 \ 19.89$	20.23 21.49	20.69 22.58	20.47 21.72
noise	peppers	mandrill	pirate	jetplane	lake
level	ADM Splitting	ADM Splitting	ADM Splitting	ADM Splitting	ADM Splitting
0.01	24.05 26.28	24.18 26.20	23.74 25.06	$23.25 \ 25.09$	23.14 25.87
0.02	23.48 25.37	23.42 25.11	22.38 23.97	22.32 23.79	22.46 24.93
0.03	21.53 23.70	21.60 23.01	20.33 21.79	20.58 22.00	20.49 22.82

TABLE 1. Restoration results (PSNR) of ten images by ADM and splitting methods.

6. Conclusions. For $\ell^1 + \ell^1$ problems, we have proposed a novel dual splitting method method for sparsity regularization. This method transform $\ell^1 + \ell^1$ problems to box constraint smooth functional which can be easily solved. the Numerical results indicate that the proposed splitting algorithm performs competitively with several state-of-art algorithms such as ADM method. We remark that for well-conditioned problems e.g. compressive sensing, ℓ^1 fidelity is more accurate than ℓ^2 fidelity with impulsive noise. On various classes of test problems with different condition numbers, the proposed splitting method is more stable with respect to noise levels compared with ADM- ℓ^1 algorithm.

Acknowledgment. This paper is supported by Project supported by the National Nature Science Foundation of China (no.41304093), the Fundamental Research Funds for the Central Universities (no.2572015CB19), Heilongjiang Postdoctoral Research Developmental Fund(no.LBH-Q16008).

REFERENCES

- E. J. Candes, J. 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, 2007.
- [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: 11 regularization, Optics Express, vol.18, no.3, pp.1854–1871, 2010.
- [4] I. Loris, G. Nolet, and I. Daubechies, Tomographic inversion using l1-norm regularization of wavelet coefficients, *Geophysical Journal International*, vol.170, no.1, pp.359–370,2007.
- [5] B. T. Jin, and P. Maass, Sparsity regularization for parameter identification problems, *Inverse Problems*, vol.28, no.12, pp.123001, 2012.
- [6] I. Daubechies, M. Defrise, and C. D. 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, Austria, 2010.

- [8] D. A. Lorenz, Convergence rates and source conditions for Tikhonov regularization with sparsity constraints, *Journal of Inverse and Ill-posed Problems*, vol.16, no.5, pp.463–478, 2008.
- [9] M. Grasmair, M. Haltmeier, and O. Scherzer, Sparse regularization with ℓ^q penalty term, *Inverse Problems*, vol.24, no.5, pp.055020, 2008.
- [10] R. Griesse, and D. A. Lorenz, A semismooth Newton method for Tikhonov functionals with sparsity constraints, *Inverse Problems*, vol.24, no.3, pp.035007, 2008. 24(3)035007.
- [11] M. Grasmair, Non-convex sparse regularisation, Journal of Mathematical Analysis and Applications, vol.365, no.1, pp.19–28, 2010.
- [12] 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.
- [13] C. A. Zarzer, On Tikhonov regularization with non-convex sparsity constraints, *Inverse Problems*, vol.25, no.2, pp.025006, 2009.
- [14] 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, pp.065018, 2013.
- [15] P. J. Rousseeuw, and A. M. Leroy, Robust regression and outlier detection, John Wiley Sons, New York, USA, 1987.
- [16] C. Clason, B. Jin, and K. Kunisch, A Duality-Based Splitting Method for ℓ¹-TV Image Restoration with Automatic Regularization Parameter Choice, SIAM Journal on Scientific Computing, vol.32, no.3, pp.1484–1505, 2010.
- [17] J. Yang, and Y. Zhang, Alternating Direction Algorithms for l₁-Problems in Compressive Sensing, SIAM Journal on Scientific Computing, vol.33, no.1, pp.250–278, 2011.
- [18] A. Borsic, and A. Adler, A primalCdual interior-point framework for using the L1 or L2 norm on the data and regularization terms of inverse problems, *Inverse Problems*, vol.28, no.9, pp.095011, 2012.
- [19] B. Efron, T. Hastie, and I. Johnstone, [Least Angle Regression]: Rejoinder, Annals of Statistics, vol.32, no.2, pp.494–499, 2004.
- [20] I. Daubechies, R. Devore, and M. Fornasier, Iteratively reweighted least squares minimization for sparse recovery, *Communications on Pure & Applied Mathematics*, vol.63, no.1, pp.1–38, 2010.
- [21] T. Blumensath, and M. E. Davies, Iterative Thresholding for Sparse Approximations, Journal of Fourier Analysis & Applications, vol.14, no.5, pp.629–654, 2008.
- [22] T. Blumensath, and M. E. Davies, Iterative hard thresholding for compressed sensing, Applied & Computational Harmonic Analysis, vol.27, no.3, pp.265–274, 2008.
- [23] M. Cullen, M. A. Freitag, and S. Kindermann, Large Scale Inverse Problems: Computational Methods and Applications in the Earth Sciences, De Gruyter, Austria, 2013.
- [24] S. G. Mallet, A wavelet tour of signal processing, Elsevier, Oxford, UK, 2010.
- [25] I. Loris, G. Nolet, and I. Daubechies, Tomographic inversion using l1 -norm regularization of wavelet coefficients, *Geophysical Journal International*, vol.170, no.1, pp.359-C370, 2007.
- [26] I. Ekeland, and R. Man, Convex analysis and variational problems, SIAM, US, 1999.
- [27] 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.
- [28] E. G. Birgin, M. Raydan, SPG: Software for Convex-Constrained Optimization, Competent International Journal for Computation & Mathematics in Electrical & Electronic Engineering, vol.29, no.5, pp.1331–1344, 2001.
- [29] Y. Xiao, H. Zhu and S. Y. Wu, Primal and dual alternating direction algorithms for 11-11 norm minimization problems in compressive sensing, *Computational Optimization and Applications*, vol.54, no.2, pp.441–459, 2013.







(b) solution with 3% Impulsive noise



(d) solution with 7% Impulsive noise



(f) solution with 10% Impulsive noise



(h) solution with 15% Impulsive noise

FIGURE 1. Restorations of ℓ^1 and ℓ^2 fidelities with Impulsive noises



FIGURE 2. Comparisons of Splitting method with $ADM-\ell^1$ and TNIP



FIGURE 3. Restoration of Lena with impulsive noise d=0.02