Journal of Information Hiding and Multimedia Signal Processing ©2018 ISSN 2073-4212
Ubiquitous International Volume **9**, Number **6**, November **2018**

A Stable Multi-parameter Thresholding Method for Ill-conditioned Sparse Signal Restoration and Optimal Parameter Choices

Liang Ding[1] and Guan Wang[1,2]

1. Department of Mathematics
Northeast Forestry University
26 Hexing Road, Harbin, China
2. School of Mathematics
Tianjin University
135 Yaguan Road, Tianjin, China
dl@nefu.edu.cn

Received March, 2017; Revised May, 2018

Abstract. *In this paper, we will solve sparse regularization with ill-conditioned problems. Typically, sparse regularization method is appropriate for compressive sensing problem, where random matrices are well-conditioned. For ill-conditioned problems, for example image inpainting, image deblurring, sparsity regularization is often unstable. The motivation of this paper is that the traditional sparse regularization method cannot effectively recovery the approximate solution with ill-conditioned problems because the stability of ℓ^1 regularization is weaker than ℓ^2 regularization in statistics. To improve the stability of sparsity regularization, a smooth ℓ^2 term is added to original sparsity regularization. This method admits sparsity inverse problems are ill-conditioned. Contributions of this paper are as follows. Convergence of the minimizer and its stability are studied. A stable multi-parameter thresholding algorithm for ill-conditioned problems are proposed and numerical results are presented to illustrate the features of the functional and algorithms.*
Keywords: Sparse signal recovery, regularization, Multi-parameter thresholding, Ill-conditioned problems, balancing principle

---

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

$$Kx = y \tag{1}$$

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

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

is available. We call y^δ the noisy data and δ the noise level. It is well known that the conventional method for solving (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, 11, 12, 13, 14] and the references given there.

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

$$\mathcal{J}(x) = \min_{x} \parallel Kx - y^{\delta} \parallel_{\ell^2}^2 + \alpha \sum_{i} \omega_i |\langle \phi_i, x \rangle| \tag{3}$$

where K is bounded linear ill-conditioned operator, we call (3) ℓ^2+ℓ^1 problem. In the past few years, numerous algorithms have been systematically proposed for the ℓ^2+ℓ^1 problems. For sparsity regularization, the popular algorithms, e.g. homotopy (LARS) method [16, 17] and iteratively reweighted least squares method (IRLS) [18] cannot be directly applied to ℓ^2+ℓ^1 problem if A is ill-conditioned and regularization term lack of stability. For ill-conditioned problems, these methods are often unstable[20][Chaper 5].

Ill-conditioned sparse linear systems arise in a wide variety of science and engineering applications, ranging from geomechanical problems to computational number theory. On the other hand, in spite of growing interests in the ill-conditioned sparse linear systems, we can indicate limited work has been done for numerical methods of sparsity regularization with ill-conditioned problems. In [21] Herrholz and Teschke concerned with compressive sampling strategies and sparse recovery principles for linear inverse and ill-posed problems. They provide compressed measurement models for ill-posed problems and recovery accuracy estimates for sparse approximations of the solution of the underlying inverse problem. Gholami and Siahkoohi[22] adopt sparsity regularization to solve geophysical ill-posed problems and developed a new regularization scheme for a high resolution solution of linear and non-linear inverse problems which benefits from the advantages of two different sparsifying operators in representation of the desired model. Haber, Magnant etc.[23] proposed a new numerical methods for A-optimal designs with a sparsity constraint for ill-posed inverse problems. In[24], Golmohammadi, Khaninezhad and Jafarpour proposed a group-sparsity regularization for ill-posed subsurface flow inverse problems. Carson and Higham[25] gave a new analysis of iterative refinement and its application to accurate solution of ill-conditioned sparse linear systems.

The motivation of this paper is that the traditional sparse regularization method cannot effectively recovery the approximate solution with ill-conditioned problems because the stability of l1 regularization is weaker than l2 regularization in statistics[26]. To improve the stability of sparsity regularization, inspired by multi-regularization theory[31, 32, 33], a smooth l2 term is added to original sparsity regularization to construct multi-parameter regularization.

The advantage of problem multi-parameter regularization in place of (3) is that the regularization effect of ℓ^1 penalty is weak, ℓ^2 penalty can improve the stability of (3). Moreover, such functional is appropriate for the sparsity recovery due to the fact that sparse signal typically contain smooth and impulsive features simultaneously. We investigate regularizing properties of multi-parameter regularization with ill-conditioned problems, a stable multi-parameter iterative threshold algorithm is applied to multi-parameter regularization for numerical solution.

An outline of this paper is as follows. We devote Section 2 to a discussion of regularizing properties, including well-posedness and convergence rate. In Section 3, multi-parameter iterative threshold algorithm is applied to compute the minimizers. Section 4 provides a detailed exposition of multi-parameter choice rule based on the balancing principle. Finally, Numerical experiments involving compressed sensing and image inpainting are presented in Section 5, showing that our proposed approaches are robust and efficient.

2. **Regularizating properties.** In general, for the approximate solutions of $Kx = y$, sparsity regularization is given by

$$\min_x \| Kx - y^\delta \|_{\ell^2}^2 + \alpha \| x \|_{w,p}^p \tag{4}$$

where $\| x \|_{w,p}^p = \sum_\gamma \omega_\gamma |\langle \varphi_\gamma, x\rangle|^p (1 \le p \le 2)$, α is the regularization parameter balancing the fidelity $\| Kx - y^\delta \|_{\ell^2}^2$ and regularization term $\| x \|_{w,p}^p$. The problem (4) 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 [27, 28, 29, 30] for a recent account of the theory.

In this paper, We add the smooth penalty $\beta \| u \|_{\ell^2}^2$ to traditional sparsity regularization (3) to consider the minimization of the following multi-parameter regularization functional

$$\Gamma_{\alpha,\beta}(x) = \|Kx - y^\delta\|_{\ell^2}^2 + \alpha \sum_{i\in\mathbb{N}} \omega_i |\langle \phi_i, x\rangle| + \beta \|x\|_{\ell^2}^2 \tag{5}$$

where $R(x) = \alpha\|x\|_{\omega,1}^1 + \beta\|x\|_{\ell^2}^2$, $\|x\|_{\omega,1}^1 = \sum_{i\in\mathbb{N}} \omega_i |\langle \phi_i, x\rangle|$, 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. $K : \mathrm{dom}(K) \subseteq X \to Y$ is a bounded linear operator and $\mathrm{dom}(K) \cap \mathrm{dom}(R) \neq \emptyset$. $(\phi_i)_{i\in\Lambda} \subset X$ is an orthonormal basis where Λ is some countable index set. From now, we denote

$$x_i = \langle x, \phi_i\rangle,$$

and

$$\|x\|_{\ell^2} = (\sum_i |\langle \phi_i, x\rangle|^2)^{\frac{1}{2}} = (\sum_i |x_i|^2)^{\frac{1}{2}}.$$

To prove convergence rate results we denote by $x_{\alpha,\beta}^\delta$ the minimizer of the regularization functional $\Gamma_{\alpha,\delta}(x)$ 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*

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

We define the sparsity as follows:

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

In this subsection, well-posedness and convergence rate of the regularization method are given.

Proposition 2.1. *(stability)Let $\alpha > 0$, $\beta > 0$. $\{y_k\}$ and $\{x_k\}$ are two sequences. If $y_k \to y^\delta$, for every $\{x_k\}$, there exists*

$$x_k \in \arg\min\{\|Kx - y_k\|_{\ell^2}^2 + \alpha\|x\|_{\omega,1}^1 + \beta\|x\|_{\ell^2}^2\}.$$

Then there exists a convergent subsequence of $\{x_k\}$. And the subsequence convergence to the minimizer of (5).

Proof: By definition of $\{x_k\}$, we have

$$\begin{aligned} &\|Kx_k - y_k\|_{\ell^2}^2 + \alpha\|x_k\|_{\omega,1}^1 + \beta\|x_k\|_{\ell^2}^2 \\ &\le \|Kx - y_k\|_{\ell^2}^2 + \alpha\|x\|_{\omega,1}^1 + \beta\|x\|_{\ell^2}^2 \end{aligned} \tag{6}$$

for $\forall x \in dom(K) \cap dom(R)$. As $\|x_k\|_{\omega,1}^1$ and $\|x_k\|_{\ell^2}^2$ are bounded, there exist a subsequence $\{x_m\}$ of $\{x_k\}$ and a $\bar{x}$, such that

$$x_m \rightharpoonup \bar{x}$$

and

$$Kx_m \rightharpoonup \bar{x},$$

where we denote by $\rightharpoonup$ weak convergence. By weak lower semi-continuity of the norm, we have

$$\begin{aligned} &\|\bar{x}\|_{\omega,1}^1 \le \liminf_{m\to\infty} \|x_m\|_{\omega,1}^1, \\ &\|\bar{x}\|_{\ell^2}^2 \le \liminf_{m\to\infty} \|x_m\|_{\ell^2}^2 \end{aligned} \tag{7}$$

and

$$\|K\bar{x} - y^\delta\|_{\ell^2}^2 \le \liminf_{m\to\infty} \|Kx_m - y_m\|_{\ell^2}^2. \tag{8}$$

Hence that

$$\begin{aligned} &\|K\bar{x} - y^\delta\|_{\ell^2}^2 + \alpha\|\bar{x}\|_{\omega,1}^1 + \beta\|\bar{x}\|_{\ell^2}^2 \\ &\le \liminf_{m\to\infty}(\|Kx_m - y_m\|_{\ell^2}^2 + \alpha\|x_m\|_{\omega,1}^1 + \beta\|x_m\|_{\ell^2}^2) \\ &\le \limsup_{m\to\infty}(\|Kx_m - y_m\|_{\ell^2}^2 + \alpha\|x_m\|_{\omega,1}^1 + \beta\|x_m\|_{\ell^2}^2) \\ &\le \lim_{m\to\infty}(\|Kx - y_m\|_{\ell^2}^2 + \alpha\|x\|_{\omega,1}^1 + \beta\|x\|_{\ell^2}^2) \\ &= \|Kx - y^\delta\|_{\ell^2}^2 + \alpha\|x\|_{\omega,1}^1 + \beta\|x\|_{\ell^2}^2). \end{aligned} \tag{9}$$

This implies that $\bar{x}$ is a minimizer of (5), and that

$$\begin{aligned} &\lim_{m\to\infty}(\|Kx_m - y_m\|_{\ell^2}^2 + \alpha\|x_m\|_{\omega,1}^1 + \beta\|x_m\|_{\ell^2}^2) \\ &= \|K\bar{x} - y^\delta\|_{\ell^2}^2 + \alpha\|\bar{x}\|_{\omega,1}^1 + \beta\|\bar{x}\|_{\ell^2}^2. \end{aligned} \tag{10}$$

Now, assume that

$$x_m \to \bar{x}$$

is false. Then

$$c := \limsup_{m\to\infty}(\|x_m\|_{\ell^2}^2 + \|x_m\|_{\omega,1}^1 > \|\bar{x}\|_{\ell^2}^2 + \|\bar{x}\|_{\omega,1}^1)$$

and there exists a subsequence $\{x_n\}$ of $\{x_m\}$ such that

$$x_n \to \bar{x},$$

,

$$Kx_n \to \ K\bar{x}$$

and

$$\|x_n\|_{\ell^2}^2 + \|x_n\|_{\omega,1}^1 \to c.$$

Then we obtain

$$\begin{aligned} &\lim_{n\to\infty}(\|Kx_n - y_n\|_{\ell^2}^2) \\ &= \|K\bar{x} - y^\delta\|_{\ell^2}^2 + (\alpha\|\bar{x}\|_{\omega,1}^1 + \beta\|\bar{x}\|_{\ell^2}^2 - c) \\ &< \|K\bar{x} - y^\delta\|_{\ell^2}^2 \end{aligned} \tag{11}$$

in contradict to (8). This argument shows that $x_m \to \bar{x}$.

Assumption 2.3. *The functional $\|Kx - y^\delta\|_{\ell^2}^2$ vanishes if and only if $Kx = y^\delta$, and satisfies*

$$\|Kx - y^\delta\| \le c(\|Kx' - (y^\delta)'\| + \|Kx - (y^\delta)'\|)$$

for constant c and x' with $Kx' = y^\delta$.

According to discrepancy principle, we have a appropriate regularization parameter (α,β) by

$$\|Kx^{\delta}_{\alpha,\beta}-y^{\delta}\|^2_{\ell^2}=c_m\delta^2 \tag{12}$$

for some constant $c_m\geq 1$.

Let $\omega_i=\beta$ below.

Theorem 2.1. *(convergence)Let Assumption 2.3 be satisfied and the operator K be injective. Then for any $(\alpha,\beta)\equiv(\alpha(\delta),\beta(\delta))$ satisfying $\|Kx^{\delta}_{\alpha,\beta}-y^{\delta}\|^2_{\ell^2}=c_m\delta^2$ and $c_0\leq \dfrac{\alpha(\delta)}{\beta(\delta)}\leq c_1$ for some $c_0,c_1>0$, there holds $\lim\limits_{\delta\to 0}x^{\delta}_{\alpha,\beta}=x^{\dagger}$ in X.*

Proof: Because $x^{\delta}_{\alpha,\beta}$ is minimizer of the regularization functional $\Gamma_{\alpha,\delta}(x)$, which implies that

$$\begin{aligned}
&\|Kx^{\delta}_{\alpha,\beta}-y^{\delta}\|^2_{\ell^2}+(\alpha,\beta)\cdot(\sum_{i\in\mathbb{N}}\omega_i|\langle\phi_i,x^{\delta}_{\alpha,\beta}\rangle|+\|x^{\delta}_{\alpha,\beta}\|^2_{\ell^2})\\
&\leq \|Kx^{\dagger}-y^{\delta}\|^2_{\ell^2}+(\alpha,\beta)\cdot(\sum_{i\in\mathbb{N}}\omega_i|\langle\phi_i,x^{\dagger}\rangle|+\|x^{\dagger}\|^2_{\ell^2})\\
&\leq \delta^2+(\alpha,\beta)\cdot(\sum_{i\in\mathbb{N}}\omega_i|\langle\phi_i,x^{\dagger}\rangle|+\|x^{\dagger}\|^2_{\ell^2}).
\end{aligned} \tag{13}$$

Combination of (13) and discrepancy principle imply that

$$\begin{aligned}
&(\alpha,\beta)\cdot(\sum_{i\in\mathbb{N}}\omega_i|\langle\phi_i,x^{\delta}_{\alpha,\beta}\rangle|+\|x^{\delta}_{\alpha,\beta}\|^2_{\ell^2})\\
&\leq (\alpha,\beta)\cdot(\sum_{i\in\mathbb{N}}\omega_i|\langle\phi_i,x^{\dagger}\rangle|+\|x^{\dagger}\|^2_{\ell^2}).
\end{aligned} \tag{14}$$

Therefore, either

$$\sum_{i\in\mathbb{N}}\omega_i|\langle\phi_i,x^{\delta}_{\alpha,\beta}\rangle|\leq\sum_{i\in\mathbb{N}}\omega_i|\langle\phi_i,x^{\dagger}\rangle|$$

or $\|x^{\delta}_{\alpha,\beta}\|^2_{\ell^2}\leq\|x^{\dagger}\|^2_{\ell^2}$ holds. The assumption

$$c_0\leq\frac{\alpha(\delta)}{\beta(\delta)}\leq c_1$$

implies that the penalties are bounded. From the coercivity of $\|x\|^2_{\ell^2}$ and $\|x\|^1_{\omega,1}$, we have the sequence $\{x^{\delta}_{\alpha,\beta}\}$ is uniformly bounded. Thus there exists a subsequence, also denoted by $\{x^{\delta}_{\alpha,\beta}\}$, and some x^*, such that $x^{\delta}_{\alpha,\beta}\to x^*$ in X. The lower semi-continuity of the functional $\|Kx-y^{\delta}\|^2_{\ell^2}$ and Assumption 2.3 yields

$$\begin{aligned}
0\leq\|Kx^*-y^{\dagger}\|^2_{\ell^2}&\leq c\liminf_{\delta\to 0}(\|Kx^{\dagger}-y^{\delta}\|^2_{\ell^2}+\|Kx^{\delta}_{\alpha,\beta}-y^{\delta}\|^2_{\ell^2})\\
&\leq\liminf_{\delta\to 0}c(1+c_m)\delta^2=0.
\end{aligned} \tag{15}$$

In particular, $\|Kx^*-y^{\dagger}\|^2_{\ell^2}=0$, i.e. $Kx^*=y^{\dagger}$ and K be injective, which imply that $x^*=x^{\dagger}$.

Next we use Bregman distance to measure the convergence rate. The definition of the Bregman distance in this work is

$$d_{\xi_1}(x,x^{\dagger})=\|x\|^1_{\omega,1}-\|x^{\dagger}\|^1_{\omega,1}-\langle\xi_1,x-x^{\dagger}\rangle$$

or

$$d_{\xi_2}(x,x^{\dagger})=\|x\|^2_{\ell^2}-\|x^{\dagger}\|^2_{\ell^2}-\langle\xi_2,x-x^{\dagger}\rangle.$$

Theorem 2.2. [33] *(convergence rate)If the exact solution $x^\dagger$ satisfies the source condition:*

$$range(K^*) \cap \partial(\|x^\dagger\|_{\ell^2}^2) \cap \partial(\|x^\dagger\|_{\omega,1}^1) \neq \emptyset.$$

Then for any α, β solving (12), there exists $\xi_2 \in \partial(\|x^\dagger\|_{\ell^2}^2)$(or $\xi_1 \in \partial(\|x\|_{\omega,1}^1)$) such that

$$d_{\xi_2}(x_\alpha^\delta, x^\dagger) \leq C\delta$$

or

$$d_{\xi_1}(x_\alpha^\delta, x^\dagger) \leq C\delta.$$

Proof: Because $x_{\alpha,\beta}^\delta$ is minimizer, it implies that

$$\begin{aligned} &\|Kx_\alpha^\delta - y^\delta\|_{\ell^2}^2 + (\alpha, \beta) \cdot (\sum_{i\in\mathbb{N}} \omega_i |\langle \phi_i, x_\alpha^\delta \rangle| + \|x_\alpha^\delta\|_{\ell^2}^2) \\ &\leq \|Kx^\dagger - y^\delta\|_{\ell^2}^2 + (\alpha, \beta) \cdot (\sum_{i\in\mathbb{N}} \omega_i |\langle \phi_i, x^\dagger \rangle| + \|x^\dagger\|_{\ell^2}^2) \\ &\leq \delta^2 + (\alpha, \beta) \cdot (\sum_{i\in\mathbb{N}} \omega_i |\langle \phi_i, x^\dagger \rangle| + \|x^\dagger\|_{\ell^2}^2). \end{aligned} \tag{16}$$

According to discrepancy principle

$$\begin{aligned} &(\alpha, \beta) \cdot (\sum_{i\in\mathbb{N}} \omega_i |\langle \phi_i, x_\alpha^\delta \rangle| + \|x_\alpha^\delta\|_{\ell^2}^2) \\ &\leq (\alpha, \beta) \cdot (\sum_{i\in\mathbb{N}} \omega_i |\langle \phi_i, x^\dagger \rangle| + \|x^\dagger\|_{\ell^2}^2). \end{aligned} \tag{17}$$

Consequently, we have that there holds

$$\|x_\alpha^\delta\|_{\ell^2}^2 \leq \|x^\dagger\|_{\ell^2}^2 \tag{18}$$

or

$$\sum_{i\in\mathbb{N}} \omega_i |\langle \phi_i, x_\alpha^\delta \rangle| \leq \sum_{i\in\mathbb{N}} \omega_i |\langle \phi_i, x^\dagger \rangle|. \tag{19}$$

Therefore, by the source condition, and Cauchy-Schwarz inequality, we have

$$\begin{aligned} &d_\xi(x_\alpha^\delta, x^\dagger) = \|x_\alpha^\delta\|_{\ell^2}^2 - \|x^\dagger\|_{\ell^2}^2 - \langle \xi, x_\alpha^\delta - x^\dagger \rangle \\ &\leq -\langle \xi, x_\alpha^\delta - x^\dagger \rangle = -\langle K^*\omega_i, x_\alpha^\delta - x^\dagger \rangle \\ &= -\langle \omega_i, K(x_\alpha^\delta - x^\dagger) \rangle \leq \|\omega_i\|_{\ell^2}^2 \|K(x_\alpha^\delta - x^\dagger)\|_{\ell^2}^2 \\ &\leq \|\omega_i\|_{\ell^2}^2 (\|Kx_\alpha^\delta - y^\delta\|_{\ell^2}^2 + \|y^\delta - Kx^\dagger\|_{\ell^2}^2) \\ &\leq (1 + c_m)\|\omega_i\|_{\ell^2}^2 \delta \end{aligned} \tag{20}$$

where ω_i is source representer. The proof is similar if

$$\sum_{i\in\mathbb{N}} \omega_i |\langle \phi_i, x_\alpha^\delta \rangle| \leq \sum_{i\in\mathbb{N}} \omega_i |\langle \phi_i, x^\dagger \rangle|$$

3. The Multi-parameter Iterative Threshold Algorithm.

The multi-paremeter iterative threshold algorithm plays an important role of solving regularization. In this section, the iterative algorithm will be introduced. Define functional

$$\Theta(x; a) = \|x - a\|^2 - \|Kx - Ka\|^2 \tag{21}$$

and

$$\Phi_{\alpha,\beta}(x, a) = \Gamma_{\alpha,\beta}(x) + \Theta(x; a). \tag{22}$$

Denote $\omega_i := \alpha\omega_i$, then

$$\begin{aligned}\Phi_{\alpha,\beta}(x,a) &= \Gamma_{\alpha,\beta}(x) - \|Kx-Ka\|^2 + \|x-a\|^2 \\ &= \|Kx-y^\delta\|^2 + \sum_i \omega_i|\langle x,\phi_i\rangle| + \beta\|x\|^2 \\ &-\|Kx-Ka\|^2 + \|x-a\|^2 \\ &= \|x\|^2 - 2\langle x, a+K^*y^\delta - K^*Ka\rangle \\ &+\sum_i \omega_i|\langle x,\phi_i\rangle| + \|y^\delta\|^2 + \|a\|^2 - \|Ka\|^2 + \beta\|x\|^2 \\ &= \sum_i [x_i^2 - 2x_i(a+K^*y^\delta - K^*Ka)_i \\ &+\omega_i|x_i|] + \|y^\delta\|^2 + \|a\|^2 - \|Ka\|^2 + \beta\|x\|^2.\end{aligned} \tag{23}$$

The Euler equation of (23) is

$$2(1+\beta)x_i + \omega_i sign(x_i) = 2(a_i + [K^*(y^\delta - Ka)]_i).$$

For $x_i > 0$, we have

$$x_i = \frac{1}{1+\beta}(a_i + [K^*(y^\delta - Ka)]_i) - \frac{\omega_i}{2(1+\beta)},$$

that is

$$\frac{1}{1+\beta}(a_i + [K^*(y^\delta - Ka)]_i) > \frac{\omega_i}{2(1+\beta)}.$$

For $x_i < 0$, we will see that

$$x_i = \frac{1}{1+\beta}(a_i + [K^*(y^\delta - Ka)]_i) + \frac{\omega_i}{2(1+\beta)},$$

that is

$$\frac{1}{1+\beta}(a_i + [K^*(y^\delta - Ka)]_i) < -\frac{\omega_i}{2(1+\beta)}.$$

Let $x_i = 0$, if

$$|\frac{a_i + [K^*(y^\delta - Ka)]_i}{1+\beta}| \le \frac{\omega_i}{2(1+\beta)}.$$

To summarize

$$x_i = S_{\omega_i,1}(\frac{a_i + [K^*(y^\delta - Ka)]_i}{1+\beta}), \tag{24}$$

where

$$\begin{cases}\langle Ku_\beta - y^\delta, p - p_\beta\rangle \ge 0, \\ \langle v_\beta, q - q_\beta\rangle \ge 0\end{cases} \tag{25}$$

$$S^\beta_{\omega,1}(u) = \begin{cases} u - \frac{\omega}{2(1+\beta)}, u \ge \frac{\omega}{2(1+\beta)} \\ 0, |u| < \frac{\omega}{2(1+\beta)} \\ u + \frac{\omega}{2(1+\beta)}, |u| \le -\frac{\omega}{2(1+\beta)} \end{cases} \tag{26}$$

The Algorithm 1 is given as follows: Multi-parameter iterative threshold method for $\Phi_{\alpha,\beta}$,
1:Choose a, set ω_0, α_0, β_0, $x^0 = a$, $n = 1$
2:$x^n = S^{\beta_{n-1}}_{\omega_{n-1},1}(x^{n-1} + K^*(y^\delta - Kx^{n-1}))$
3:Compute α_n and β_n using fixed point Algorithm 2
4:Compute $\omega_n = \alpha_n \cdot \omega_{n-1}$
5:$n = n + 1$
Until a stopping criterion satisfied.

4. **Choice of parameter α and β.** The choice of regularization parameter of (5) is also important for the performance of algorithm from the numerical perspective. Morozov discrepancy principle is a good choice from theoretical perspective. However, it is difficult to compute because we usually cant obtain accurate estimate of noise. So in this paper we adopt balance principle. The solution $u^\delta_{\alpha,\beta}$ of $\Gamma_{\alpha,\beta}(x)$ in (3) converges theoretically to the solution u^δ_α of $\mathcal{J}(x)$ in (5) as $\beta \to 0$. Obviously smaller β is better. However, the smaller β weaken the regularization effect of the ℓ^2 penalty, which leads to instability. If the solution is sparse, we can say that the non-zero coefficients are impulsive parts and the zero coefficients are smooth parts. Multi-parameter optimization functional

$$\Gamma_{\alpha,\beta}(x) = \|Kx - y^\delta\|^2_{\ell^2} + \alpha \sum_{i\in\mathbb{N}} \omega_i |\langle \phi_i, x\rangle| + \beta\|x\|^2_{\ell^2} \tag{27}$$

is a conventional method for ill-posed problems if the solutions have a number of different structures. In [31, 33], numerical experiments show that multi-parameter regularization can recover the different part of solutions efficiently. If the solutions contain only a single structure, multi-parameter regularization also have better performance. So when we choose the parameters α and β, it is not necessarily require the parameter β tending to zero. A conventional method for the choice of parameters α and β is multi-parameter regularization choice principle, e.g., discrepancy principle[31], balance principle[33]. In this paper, we use multi-parameter balance principle for the choice of regularization parameters α and β.

Balance principle is to compute minimizers of the function

$$\Phi_\gamma(\alpha, \beta) = c_\gamma \frac{(\inf \parallel Ku - y^\delta \parallel_{\ell^1} + \alpha \parallel u \parallel_{\ell^1} + \beta \parallel u \parallel^2_{\ell^2})^{\gamma+2}}{\alpha\beta}, \tag{28}$$

where c_γ is a fixed constant, (28) is equivalent to [33]

$$\gamma\alpha \parallel u^\delta_{\alpha,\beta} \parallel_{\ell^1} = \gamma\beta \parallel u^\delta_{\alpha,\beta} \parallel^2_{\ell^2} = \parallel Ku^\delta_{\alpha,\beta} - y^\delta \parallel_{\ell^1} . \tag{29}$$

The Algorithm 2 is given as follows:
Fixed point algorithm for α and β according to eq.(29),
1:Choose α^0 and β^0, and let k=0
2:$u_{k+1} = \arg\min_x (\parallel Ku - y^\delta \parallel_{\ell^1} + \alpha \parallel u \parallel_{\ell^1} + \frac{\beta}{2} \parallel u \parallel^2_{\ell^2})$
3:$\alpha_{k+1} = \frac{1}{1+\gamma} \frac{\parallel Ku_{k+1} - y^\delta \parallel_{\ell^1} + \beta \parallel u_{k+1} \parallel^2_{\ell^2}}{\parallel u_{k+1} \parallel_{\ell^1}}$
$\beta_{k+1} = \frac{1}{1+\gamma} \frac{\parallel Ku_{k+1} - y^\delta \parallel_{\ell^1} + \beta \parallel u_{k+1} \parallel_{\ell^1}}{\parallel u_{k+1} \parallel^2_{\ell^2}}$
Until a stopping criterion satisfied.

5. **Numerical implementation.** In this section, we present practical application framework and some numerical experiments to illustrate the efficiency of the proposed method. In Section 5.1, we first discuss the practical application framework that how to apply our theory in practical applications. In Section5.2, we give three examples including well-posed and ill-posed problems to support our theory.

5.1. **Practical application framework.** To apply our theory in practical applications, the first step i.e. how to choose orthogonal basis or dictionary is critical. If the true image or signal is sparse itself, then it is not necessary to choose orthogonal basis. On the contrary, if the true image or signal is not sparse, we should choose suitable basis where image is sparse under the basis. Traditional selection for basis is wavelets. Wavelets provide orthogonal bases of $L^2(R^d)$ with localization in space and in scale. This makes them more suitable than, e.g., Fourier expansions for an efficient representation of functions that have space-varying smoothness properties. However, wavelets are not suitable for the images which have lots of curve. On the contrary, in order to obtain more efficient (sparser) expansions of image which have lots of curve, other expansions have to be used, using e.g. ridgelets, curvelets or shearlet.

A strict sparsity assumption, as introduced above, is that the orthogonal basis has to be perfectly adapted to the original image or signal, in the sense that image is only allowed to have finitely many non-zero expansion coefficients with respect to basis. While in some application the basis arises naturally, in other applications this might not be the case. For example if there is a natural image e.g. of a human face, which basis should be chosen to ensure sparsity? One remedy to this problem might be to choose a more general basis than an orthogonal basis, say a frame or a dictionary. However, the heart of the problem remains unchanged, even in these cases: the function is still allowed to have only finitely many non-zero expansion coefficients. Furthermore, one must be very cautious when dealing with function systems which are non-orthogonal, since a badly chosen system might worsen the stability of the reconstruction and hence render it meaningless. Fortunately, in most cases one has an idea which system should be chosen, in order to ensure almost sparsity. By this we mean that only a small number of expansion coefficients carry almost all information about image. Returning to our example of face images, it is well-known that the Fourier-basis or the Daubechies-basis are good choices, since images are almost sparse with respect to these basis. Image compression algorithms like the JPEG algorithm or the JPEG2000 algorithm employ this idea in order to store images in an efficient way.

If a strict sparsity assumption is hold, then at the next step, the linear equation $Kx = y^\delta$ should be discretized to a matrix equation. For example, if it is a image denoising problem, then linear operator K is discretized to the identity matrix. If the linear equation is inverse integration (or differentiation) equation, then linear operator K is discretized to the lower triangular matrix. Then the linear equation $Kx = y^\delta$ is discretized to a discrete system $Kx = y^\delta$ where K is $K : R^N \to R^N$.

With the discrete system, one can construct optimization functional

$$\Gamma_{\alpha,\beta}(x) = \|Kx - y^\delta\|_{\ell^2}^2 + \alpha \sum_{i\in\mathbb{N}} \omega_i |\langle \phi_i, x\rangle| + \beta \|x\|_{\ell^2}^2 \tag{30}$$

Given initial value u_0,α_0, β_0, ω and noise level δ, one can use Algorithm 1 to solve u_1. Then one can use Algorithm 2 to solve α_1, β_1. Then with the update u_1,α_1, β_1, one can solve the new iterative image u. Until the stop principle is fulfilled, one can obtain the numerical solution. With some evaluation criteria, e.g. PSNR, one can evaluate the performance of our method with other algorithm.

5.2. **numerical experiments.** In Section5.2.1, we first compare the performance of the multi-parameter iterative threshold(MPIT) method with the alternating direction mehtod(ADM) and truncated Newton interior-point(TNIP) method by well-conditioned compressive sensing problems. In the second example, we discuss an ill-conditioned problem where the condition number of linear operator K is 255, we aim to demonstrate that the proposed method is stable. In Section5.2.2, we discuss the image inpainting where images are sparse with respect to the Daubechies 4 wavelets. For image inpainting, the linear operator K is moderate ill-conditioned and the condition number is around 4000. 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

$$\mathrm{Rerr}(x) = \frac{\|x - x^{\dagger}\|}{\|x^{\dagger}\|} \times 100\% \tag{31}$$

and

$$\mathrm{PSNR}(y^{\delta}) = -20\log_{10}(\frac{\|x - x^{\dagger}\|}{n}). \tag{32}$$

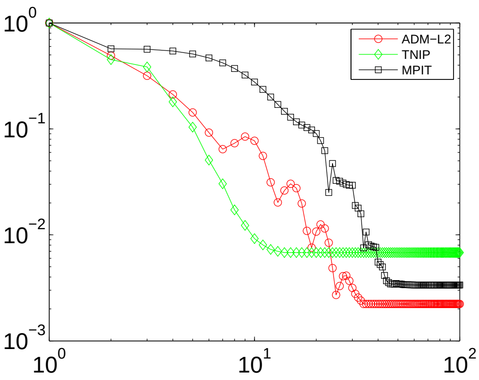
where x and $x^{\dagger}$ are iterative solution and true solution respectively. All experiments were performed under Windows 7 and Matlab R2010a on Thinkpad P50s with Intel Core i7 6500U CPU 2.50GHz 2.59GHz and 8GB RAM.

5.2.1. *Comparison of MPIT with ADM-ℓ^2 and TNIP.* This example involves compressive sensing problem

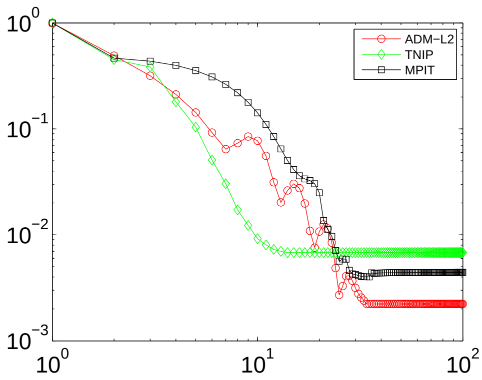
$$Kx = y^{\delta} \tag{33}$$

where matrix K is random Gaussian, $y^{\delta} = Kx^{\dagger} + \delta$ is the observed data containing white noise. We present the comparison results of MPIT with ADM-ℓ^2 and TNIP. ADM-ℓ^2[15] is an efficient alternating direction method for ℓ^2+ℓ^1 problem. TNIP[34] method uses truncated Newton interior-point for ℓ^2+ℓ^1 problem. In the first experiment, we use random Gaussian matrix $A_{m\times n}$, where sampling length is m, 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% Gaussian noise to exact 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. Fig.1 presents comparison results for three different iterative algorithms. The left column describes convergence rates of Rerr(x) for $p/m = 0.1$. The right column describes convergence rates of Rerr(x) for $p/m = 0.2$. m/n increase from top to bottom row. As can be seen from Fig.1, ADM-ℓ^2 converges faster than DSPG and the accuracy is better than MPIT when $m/n = 0.4$. With m/n decreasing, MPIT method performs more competitively. The accuracy of MPIT method is even better than ADM-ℓ^2 when $m/n = 0.1$. Though the optimal relative error of ADM-ℓ^2 is better than MPIT method, the corresponding optimal iteration number or stopping tolerance is difficult to estimate in practice. The TNIP method converges obviously faster than ADM-ℓ^2 and MPIT method. However, the accuracy of the two algorithms is worse compared with ADM-ℓ^2 and MPIT method no matter what value the m/n is.

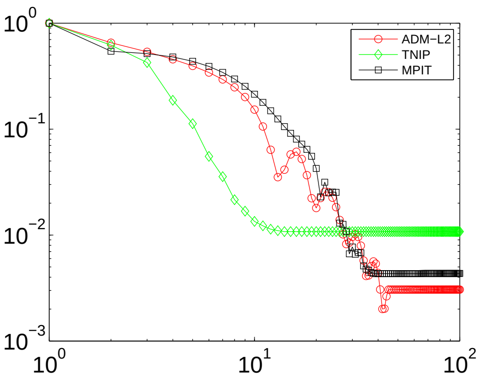
Next, in order to test the stability of the MPIT for ill-conditioned problems, we use matrix $A_{n\times n}$(n=200) whose condition number is 255. This problem was discussed by Lorentz in [10] where the ill-conditioned matrix is generated by Matlab code: "A=tril(ones(200))". The signal is p-sparsity where p/n=0.1 and 0.2. We add 1% Gaussian noise to data. As can be seen from Fig.2, MPIT converge obviously faster than the ADM-ℓ^2 method. The relative error of MPIT method is also better than ADM-ℓ^2 method. It is shown that MPIT method is stable even for large condition number matrices. In Table.1, data contain Gaussian noise with corruption Rerr(δ)=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 δ decreasing. Theoretically, ADM-ℓ^2 and MPIT method are adept to process Gaussian noise. However, ADM-ℓ^2 method is sensitive to noise when the operators are ill-conditioned. In this case, ADM-ℓ^2 cannot obtain reasonable restoration. MPIT methods are more stable to noise level δ even if matrix K has large condition numbers. For low noise levels, MPIT methods have advantage over the other two methods. For high noise levels, the restoration results of ADM-ℓ^2 method are similar with TNIP methods. Restoration results of the MPIT method are obviously better than the other two methods.



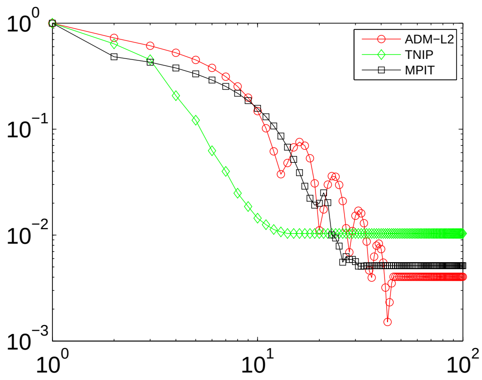
(a) m/n=0.5,p/m=0.1



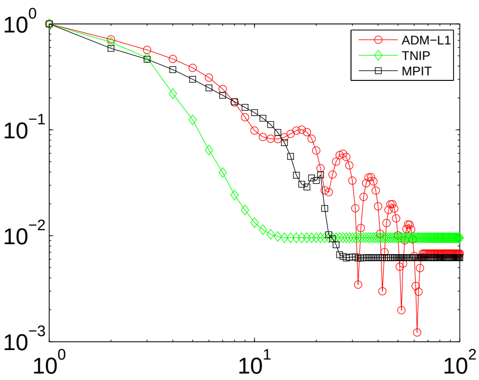
(b) m/n=0.5,p/m=0.2



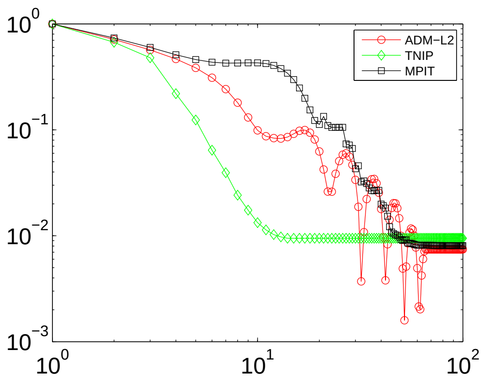
(c) m/n=0.4,p/m=0.1



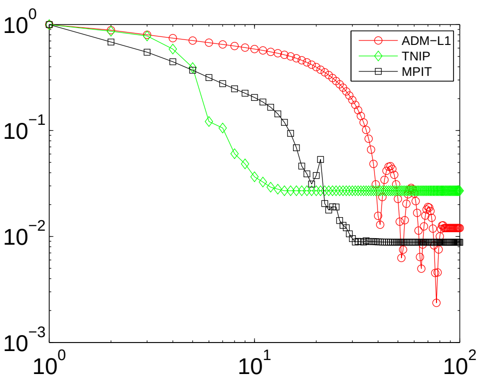
(d) m/n=0.4,p/m=0.2



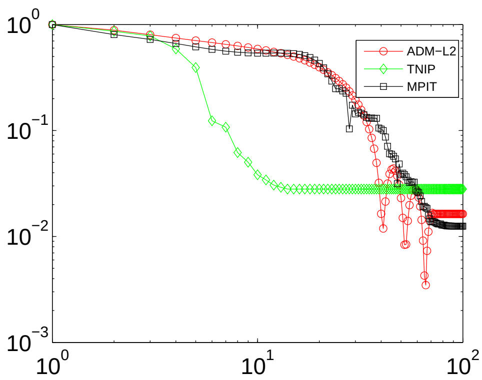
(e) m/n=0.3,p/m=0.1



(f) m/n=0.3,p/m=0.2

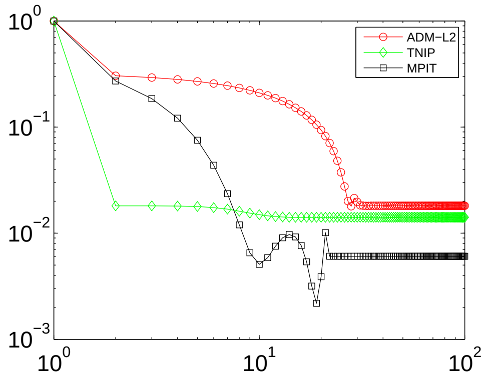


(g) m/n=0.2,p/m=0.1

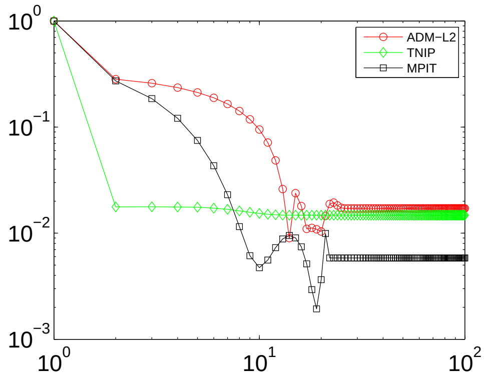


(h) m/n=0.2,p/m=0.2

FIGURE 1. Comparisons of MPIT with ADM-ℓ^2 and TNIP



(a) p/n=0.1



(b) p/n=0.2

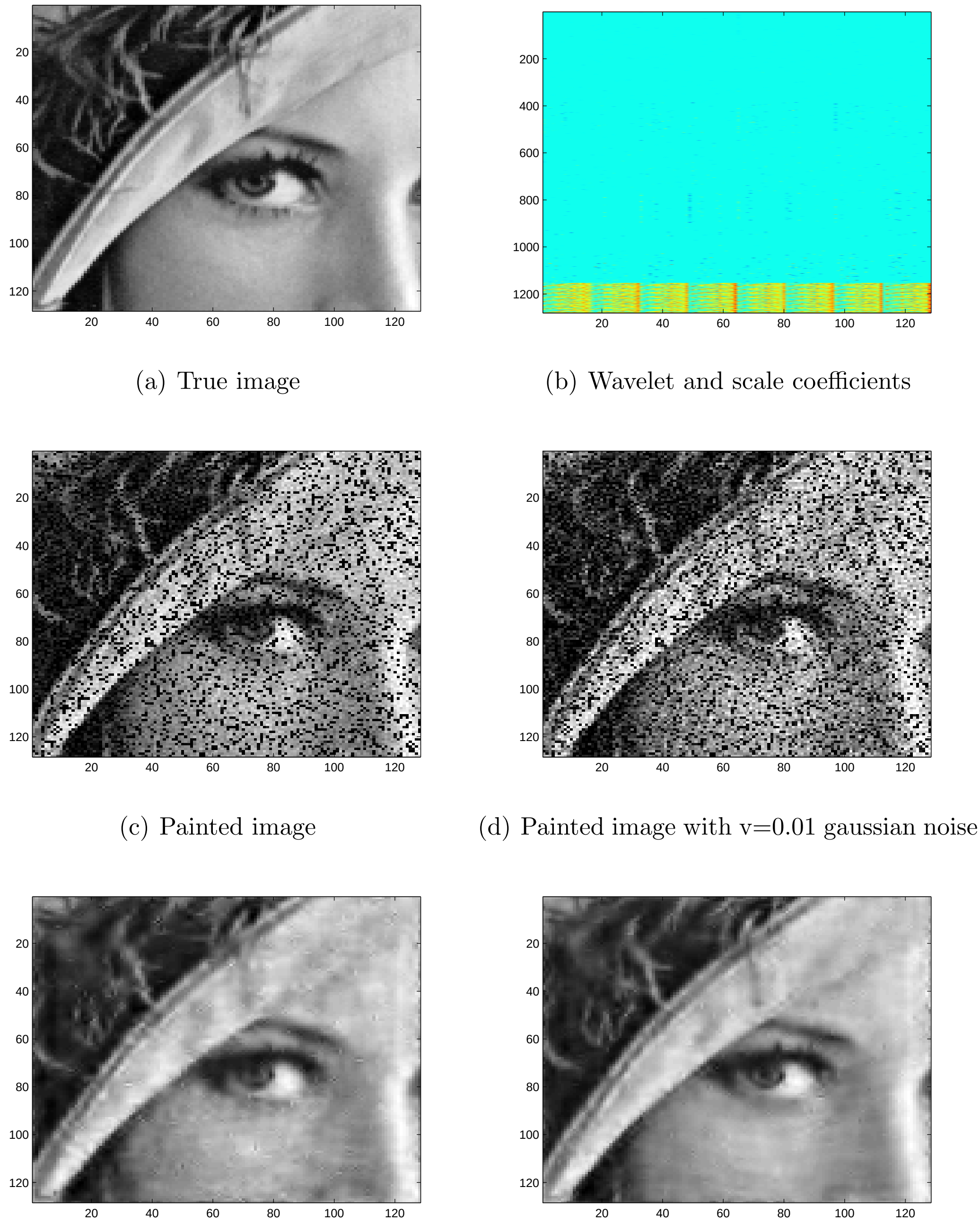
FIGURE 2. Comparisons of MPIT with ADM-ℓ^2 and TNIP

Gaussian Noise δ	TNIP Rerror	TNIP PSNR	AMD-ℓ^2 Rerror	AMD-ℓ^2 PSNR	MPIT Rerror	MPIT PSNR
0.1%	0.590%	52.87	0.707%	51.85	0.374%	55.65
0.3%	1.562%	46.98	1.607%	46.13	0.648%	52.26
0.5%	2.275%	43.09	2.556%	42.63	0.956%	49.26
1.0%	3.208%	38.08	3.932%	37.02	2.955%	41.45
3.0%	4.609%	36.12	5.156%	34.65	4.973%	38.12
5.0%	8.598%	30.45	9.963%	29.96	6.325%	35.25
10%	18.80%	24.76	18.98%	24.32	13.68%	30.28

TABLE 1. Comparison of different algorithms

5.2.2. *Image inpainting.* We present the comparison results of multi-parameter iterative threshold(MPIT) algorithm with soft iterative threshold(SIT) algorithm by 2D image inpainting problems. The image is Lena(n=128; cf. Fig.3). We randomly remove 40% 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.Figure 3). As seen from Fig.3, the representation of the image with respect to Daubechies 4 basis is sparse. We add Gaussian noise by Matlab code "imnoise(image, 'gaussian', m, v)" . In the first example, $m = 0$ and $v = 0.05$, the restoration results are shown in Fig.3. In this case, the operator K is moderate ill-conditioned, performance of MPIT is obviously better than SIT due to fact that regularization $\beta\|x\|_{\ell^2}^2$ improve the stability. Restoration results of four images with different noise levels v are given in Table.2. Restoration results show that if images have a sparse representation with respect to an orthogonal basis, MPIT method are competitive, which can obtain satisfactory results even if the image inpainting are moderate ill-posed problems.

6. **Conclusions.** To extend sparsity regularization method to moderate ill-conditioned problems, we applied multi-parameter optimization functional method to sparse inverse problems. Well-posedness and convergence rate are given. For numerical solutions, we have proposed a novel multi-parameter iterative threshold (MPIT) method for sparsity regularization. Numerical results indicate that the proposed MPIT algorithm performs



(a) True image (b) Wavelet and scale coefficients

(c) Painted image (d) Painted image with v=0.01 gaussian noise

(e) STI,PSNR=24.76 (f) MPIT,PSNR=26.85

FIGURE 3. Restoration of Lena with Gaussian noise, m=0,v=0.01

competitively with several state-of-art algorithms such as ADM method. We remark that for well-conditioned problems e.g. compressive sensing, optimal accuracy of ADM-ℓ^2 is better than DSPG method. However, the optimal accuracy of ADM-ℓ^2 method is strongly depended on stopping tolerance values which can be difficult to estimate in practice. On various classes of test problems with different condition numbers, the proposed MPIT method is more stable with respect to noise levels compared with ADM-ℓ^2 and soft threshold algorithm.

Gaussian Noise v	Boat SIT MPIT	Babara SIT MPIT	Goldhill SIT MPIT
0.005	23.26 25.54	22.01 24.32	24.65 26.96
0.01	21.13 23.23	19.84 21.85	23.56 25.31
0.02	19.06 21.45	18.86 20.25	21.95 24.32
0.03	16.85 19.89	16.19 19.08	19.52 22.46

TABLE 2. Comparison of MPIT and SIT

Acknowledgment. This paper is supported by Project supported by the Fundamental Research Funds for the Central Universities (no.2572018BC02), the National Nature Science Foundation of China (no.41304093), Heilongjiang Postdoctoral Research Developmental Fund(no.LBH-Q16008).

REFERENCES

[1] 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: l1 regularization,*Optics Express*, vol.18, no.3, pp.1854–1871, 2010.

[4] I. Loris, G. Nolet, and I. Daubechies, Tomographic inversion using $\ell 1$-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] E. Ozkan, V. Vishnevsky, and O. Goksel, Inverse Problem of Ultrasound Beamforming With Sparsity Constraints and Regularization, *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol.65, no.3, pp.356–365, 2018.

[12] Y. H. Jia, S. Kwong, J. H. Hou, Semi-Supervised Spectral Clustering With Structured Sparsity Regularization, *IEEE Signal Processing Letters*, vol.25, no.3, pp.403–407, 2018.

[13] Y. Wang, G. L. Zhou, X. Zhang, The Non-convex Sparse Problem with Nonnegative Constraint for Signal Reconstruction, *Journal of Optimization Theory and Applications*, vol.170, no.3, pp.1009–1025, 2016.

[14] Z. Y. Zha, X. G. Zhang, Q. Wang, Group-based sparse representation for image compressive sensing reconstruction with non-convex regularization, *Neurocomputing*, vol.296, pp.1–11, 2018.

[15] 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.

[16] B. Efron, T. Hastie, and I. Johnstone, [Least Angle Regression]: Rejoinder, *Annals of Statistics*, vol.32, no.2, pp.494–499, 2004.

[17] M. R. Osborne, B. Presnell and B. A. Turlach, On the LASSO and Its Dual, *Journal of Computational & Graphical Statistics*, vol.9, no.2, pp.319–337, 1999.

[18] 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.

[19] T. Blumensath, and M. E. Davies, Iterative Thresholding for Sparse Approximations, *Journal of Fourier Analysis & Applications*, vol.14, no.5, pp.629–654, 2008.

[20] M. Cullen, M. A. Freitag, and S. Kindermann, *Large Scale Inverse Problems:Computational Methods and Applications in the Earth Sciences*, De Gruyter, Austria, 2013.

[21] H. Evelyn and T. GerdM, *Compressive sensing principles and iterative sparse recovery for inverse and ill-posed problems*, vol.26, no.12, pp.125012, 2010.

[22] A. Gholami and H. R. Siahkoohi, *Regularization of linear and non-linear geophysical ill-posed problems with joint sparsity constraints*, vol.2010, no.180, pp.871–882, 2010.

[23] E.Haber, Z. Magnant, C. Lucero, L. Tenorio, *Numerical methods for A-optimal designs with a sparsity constraint for ill-posed inverse problems*, vol.2012, no.52, pp.293–314, 2012.

[24] A. Golmohammadi, M. M. Khaninezhad and B. Jafarpour, *Group-sparsity regularization for ill-posed subsurface flow inverse problems*, vol.2015, no.51, pp.8607–8626, 2015.

[25] E. Carson and N. J. Higham, *A New analysis of iterative refinement and its application to accurate solution of ill-conditioned sparse linear systems*, vol.39, no.6, pp.A2834–2856, 2017.

[26] B. Jin, D. A. Lorenz and S. Schiffler, *Elastic-net regularization: error estimates and active set methods*, vol.25, no.11, pp.115022, 2009.

[27] M. Grasmair, Non-convex sparse regularisation, *Journal of Mathematical Analysis and Applications*, vol.365, no.1, pp.19–28, 2010.

[28] 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.

[29] C. A. Zarzer, On Tikhonov regularization with non-convex sparsity constraints, *Inverse Problems*, vol.25, no.2, pp.025006, 2009.

[30] 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.

[31] S. Lu, S. V. Pereverzev, Multi-parameter regularization and its numerical realization, *Numerische Mathematik*, vol.118, no.1, pp.1–31, 2011.

[32] S. Lu, S. V. Pereverzev and Y. Shao, Discrepancy curves for multi-parameter regularization, *Journal of Inverse and Ill-posed Problems*, vol.18, no.6, pp.655–676, 2010.

[33] K. Ito, B. Jin and T. Takeuchi, Multi-Parameter Tikhonov Regularization, *Methods & Applications of Analysis*, vol.18, no.1, pp.31–46, 2011.

[34] S. J. Kim, K Koh and M. Lustig, A method for large-scale l1-regularized least squares, *IEEE Journal on Selected Topics in Signal Processing*, vol.1, no.4, pp.606–617, 2007.