Inverse Problems and Imaging
Vol. 17, No. 1, February 2023, pp. 157–179
doi:10.3934/ipi.2022035

MOROZOV'S DISCREPANCY PRINCIPLE FOR $\alpha\ell_1 - \beta\ell_2$ SPARSITY REGULARIZATION

Liang Ding*

Department of Mathematics, Northeast Forestry University
Harbin 150040, China

Weimin Han

Department of Mathematics, University of Iowa
Iowa City, IA 52242, USA

(Communicated by Gang Bao)

Abstract. In this paper, Morozov's discrepancy principle is considered for the non-convex $\alpha\ell_1 - \beta\ell_2$ sparsity regularization ($\alpha > \beta > 0$). It is shown that if $\tau > 1$ satisfies some conditions, there exists a regularization parameter α such that $\delta \le \|A(x^\delta_{\alpha,\beta}) - y^\delta\|_Y \le \tau\delta$ holds. Furthermore, it is shown that α converges to 0 as $\delta \to 0$. In addition, well-posedness and convergence rate results are presented for the regularized solution under Morozov's discrepancy principle. Numerical simulation results are reported to illustrate the efficiency of the proposed approach.

1. **Introduction.** In this paper, we consider solving an ill-posed operator equation of the form

$$A(x) = y, \tag{1}$$

where $x \in \ell_2$ is sparse, $A : \ell_2 \to Y$ is an operator from the ℓ_2 space to a Banach space Y, not necessarily linear. Norms in ℓ_2 and Y are denoted by $\|\cdot\|_{\ell_2}$ and $\|\cdot\|_Y$, respectively. In practice, the right-hand side y is known only approximately with an error up to a level $\delta \ge 0$. Therefore, we assume that we know $y^\delta \in Y$ with $\|y^\delta - y\|_Y \le \delta$ for a given $\delta \ge 0$. The most commonly adopted technique to solve problem (1) is the ℓ_p-norm sparsity regularization with $1 \le p < 2$, cf. the monographs [20, 41] and the special issues [5, 14, 29, 30] for many developments on regularization properties and minimization schemes. For $0 \le p < 1$, the ℓ_p-norm sparsity regularization has been studied in [7, 8, 10, 21, 45] and other references. We refer the reader to [27, 33, 38] for alternatives to the ℓ_0-norm sparsity regularization.

A non-convex regularization term of the form $\alpha\|\cdot\|_{\ell_1} - \beta\|\cdot\|_{\ell_2}$ ($\alpha > \beta > 0$) has attracted attention in the area of sparse recovery over the last five years, see [15, 16, 34, 35, 47, 49] and references therein. In [15, 16], we investigate the well-posedness and convergence rate of the non-convex sparsity regularization problem

$$\min\left\{\mathcal{J}^\delta_{\alpha,\beta}(x) = \frac{1}{2}\|A(x) - y^\delta\|^2_Y + \mathcal{R}_{\alpha,\beta}(x)\right\}, \tag{2}$$

2020 *Mathematics Subject Classification.* Primary: 49J27, 49J52; Secondary: 65J15, 65J20.

Key words and phrases. Ill-posed problems, Morozov's discrepancy principle, $\alpha\ell_1 - \beta\ell_2$ regularization, non-convex regularizer, sparsity.

* Corresponding author: Liang Ding.

where $x \in \ell_2$ and

$$\mathcal{R}_{\alpha,\beta}(x) := \alpha\|x\|_{\ell_1} - \beta\|x\|_{\ell_2}, \quad \alpha > \beta > 0.$$

Denoting $\eta = \beta/\alpha$, we can equivalently express the objective function $\mathcal{J}^\delta_{\alpha,\beta}(x)$ in (2) as

$$\frac{1}{2}\|A(x) - y^\delta\|_Y^2 + \alpha\mathcal{R}_\eta(x), \tag{3}$$

where

$$\mathcal{R}_\eta(x) := \|x\|_{\ell_1} - \eta\|x\|_{\ell_2}, \quad \alpha > 0,\ 1 > \eta > 0.$$

The $\ell_1 - \ell_2$ regularizer was first addressed in [19] for nonnegative least squares problems. Then it was extended to $\alpha\ell_1 - \beta\ell_2$ regularization ([15]). In [15], we present the analysis of well-posedness and a numerical algorithm. The $\alpha\ell_1 - \beta\ell_2$ regularizer behaves more and more like the ℓ_0-norm as $\beta/\alpha \to 1$. Meanwhile, $\mathcal{R}_{\alpha,\beta}(x)$ behaves like to a constant multiple of the ℓ_1-norm as $\beta/\alpha \to 0$. For the case $\beta/\alpha = 1$, $\mathcal{R}_{\alpha,\beta}(x)$ is a good approximation of a constant multiple of $\|x\|_{\ell_0}$. In addition, some analogous penalties have been proposed, e.g. $\ell_1^2 - \ell_2^2$, ℓ_1/ℓ_2 etc. Nevertheless, there lacks the theoretical analysis on the well-posedness of the regularization. These will be done in forthcoming papers.

In practice, it is crucial to choose an appropriate regularization parameter α for problem (2) where the parameter η is fixed. Generally speaking, there are two kinds of rules to determine α, one is of a priori kind and the other is of a posterior kind. For an a priori rule, α is chosen by $\alpha = O(\delta)$, e.g. $\alpha = c\,\delta$ with a fixed constant $c > 0$. However, it is challenging to determine an optimal value of c. Among a posteriori rules, Morozov's discrepancy principle (MDP) is the most commonly adopted technique to determine the regularization parameter $\alpha = \alpha(\delta, y^\delta) > 0$ such that

$$\delta \le \|A(x^\delta_{\alpha,\beta}) - y^\delta\|_Y \le \tau\delta \tag{4}$$

holds, where $\tau > 1$ and $x^\delta_{\alpha,\beta}$ is a minimizer of $\mathcal{J}^\delta_{\alpha,\beta}(x)$ in (2). If the minimizer of $\mathcal{J}^\delta_{\alpha,\beta}(x)$ is unique, it can be shown that there exists a regularization parameter α such that (4) holds ([1, 2, 43]). However, due to the non-convexity of $\mathcal{J}^\delta_{\alpha,\beta}(x)$, there may exist multiple minimizers of $\mathcal{J}^\delta_{\alpha,\beta}(x)$. One can not ensure the existence of α such that (4) holds ([40, 43]). In this paper, we show that if τ is large enough, then the existence of α satisfying (4) can be guaranteed. Furthermore, we discuss the well-posedness and convergence rate of the $\alpha\ell_1 - \beta\ell_2$ regularization under MDP. In addition, we extend the technique discussed above to a general convex penalty term.

1.1. **Related works.** The first theoretical analysis on MDP for Tikhonov regularization dates back to 1966 ([39]). Several numerical algorithms have been proposed to compute the regularization parameter α for the classical quadratic Tikhonov regularization ([22, 36, 37]). Subsequently, MDP is extended to general convex regularizations. In [9], MDP is applied to the Tikhonov regularization with convex penalty terms more general than the classical quadratic one. In [31], two iterative parameter choice methods by MDP are proposed for non-smooth Tikhonov regularization with general convex penalty terms.

We note that the results cited in the previous paragraph are limited to linear ill-posed problems. Due to the existence of multiple minimizers, the techniques for linear ill-posed problems can not be extended to nonlinear ill-posed operator equations directly. Indeed, a drawback of MDP lies in the fact that a regularization parameter with (4) might not exist for nonlinear operator equations. Special regularization techniques are needed to analyze the existence of the regularization parameter α which is determined by MDP. In [40], for the classical quadratic Tikhonov regularization, it is shown that minor restrictions to the nonlinear operator F and the solution $x^\dagger$ of the equation $F(x) = y$ can guarantee the existence of a regularization parameter α such that $\delta \leq \|F(x_\alpha^\delta) - y^\delta\| \leq \tau\delta$ ($\tau > 1$), and a convergence rate result is proved. In [2], for a general convex penalty term, the existence of α is shown under certain conditions. It is illustrated that for this parameter choice rule, $\alpha \to 0$ and $\delta^q/\alpha \to 0$ ($q \geq 1$) as the noise level δ goes to 0. In addition, convergence rate analysis is given with respect to the generalized Bregman distance. In [3], convergence rates are investigated using variational inequalities, where the regularization parameter is determined by MDP. A relation between uniqueness of minimizers in Tikhonov-type regularization and Morozov-like discrepancy principles is shown in [1].

When the penalty term is non-convex, fewer results are available. Convergence rate results can be found in [46] for the $\|\cdot\|_{\ell_0} + \|\cdot\|_{\ell_2}^2$ penalty term and in [15] for the $\|\cdot\|_{\ell_1} - \eta\|\cdot\|_{\ell_2}$ penalty term. In these references, one needs additional source condition and the assumption that there exists α such that MDP holds. To the best of our knowledge, no result is available in the literature on the validity of MDP when the penalty term is non-convex.

1.2. **Contribution and organization.** In [15], we introduced the non-convex $\alpha\ell_1 - \beta\ell_2$ sparsity regularization method, with the primary interest in the regularization properties. However, there lacks a systematical and theoretical analysis for MDP, especially on the existence of the regularization parameter α. The aim of this paper is to extend MDP to the non-convex $\alpha\ell_1 - \beta\ell_2$ sparsity regularization.

Throughout this paper, we assume the operator A has the following property: there exists $\gamma > 0$ such that for arbitrary $x_1, x_2 \in \ell_2$,

$$\|A(x_2) - A(x_1) - A'(x_1)(x_2 - x_1)\|_Y \leq \gamma\|A(x_2) - A(x_1)\|_Y.$$

We prove there exits at least one α such that

$$\delta \leq \|A(x_{\alpha,\beta}^\delta) - y^\delta\|_Y \leq \left(\max\left\{\tau^2\delta^2, (3+2\gamma)\delta^2 + \delta(2+2\gamma)\|y^\delta - A(0)\|_Y\right\}\right)^{\frac{1}{2}}. \quad (5)$$

Under (5), we investigate the well-posedness and convergence rate of the regularized solution of problem (2). We show that with the parameter choice rule (5), $\alpha \to 0$ as the noise level δ goes to 0. When A is linear and bounded, we prove there exits at least one α such that

$$\delta \leq \|Ax_{\alpha,\beta}^\delta - y^\delta\|_Y \leq \left(\max\left\{\tau^2\delta^2, 3\delta^2 + 2\delta\|y^\delta\|_Y\right\}\right)^{\frac{1}{2}}.$$

For traditional MDP with linear ill-posed problems, the existence of α can be guaranteed. Nevertheless, for nonlinear ill-posed problems, the discrepancy is not continuous with respect to α. So it is challenge to choose an appropriate τ to ensure the existence of α. The upper bound $c(\delta)$ in (5) seems loose. Nevertheless, the modified MDP (5) can ensure the existence α. Once the existence of α is guaranteed, we can determine an α by Algorithm 1 such that the modified MDP (5) holds. So the upper bound $c(\delta)$ is natural for the development of the algorithm. We present a numerical experiment to show the upper bound $c(\delta)$ is actually needed, cf. Fig. 7.

Note that we can apply the same technique to nonlinear ill-posed operator equations with a general convex penalty term $\Phi(x)$. From [2], for a general convex penalty term, the existence of α can be guaranteed under a rigid condition, i.e., there is no $\alpha > 0$ with minimizers $x^{\delta,1}_{\alpha,\beta}, x^{\delta,2}_{\alpha,\beta}$ such that

$$\|A(x^{\delta,1}_{\alpha,\beta}) - y^\delta\|_Y < \delta < \tau\delta < \|A(x^{\delta,2}_{\alpha,\beta}) - y^\delta\|_Y.$$

This condition is difficult to verify. Indeed, if the general convex penalty term fulfills some conditions, there exits at least one α such that (5) holds. These will be done in forthcoming papers.

An outline of the rest of this paper is as follows. In the next section we introduce the notation and review results of the $\alpha\ell_1 - \beta\ell_2$ sparsity regularization. In Section 3, we investigate the existence of α determined by MDP. In addition, we give analogous results for linear ill-posed problems. In Section 4, we present the well-posedness and convergence rate of the regularized solution under MDP. Furthermore, we show that MDP can guarantee $\alpha(\delta, y^\delta) \to 0$ as $\delta \to 0$. Finally, numerical experiments about linear and nonlinear ill-posed problems are presented in Section 5.

2. **Preliminaries.**

Before discussing MDP, we briefly introduce some notation and results of the $\alpha\ell_1 - \beta\ell_2$ sparsity regularization.

Denote by

$$x^\delta_{\alpha,\beta} \in \arg\min_{x\in\ell_2} \left\{ \frac{1}{2}\|A(x) - y^\delta\|_Y^2 + \alpha\mathcal{R}_\eta(x) \right\} \tag{6}$$

a minimizer of the regularization function $\mathcal{J}^\delta_{\alpha,\beta}(x)$ in (2). Let $\mathcal{L}^\delta_{\alpha,\beta}$ be the set of all such minimizers $x^\delta_{\alpha,\beta}$. An element $x^\dagger \in \ell_2$ is called an $\mathcal{R}_\eta$-minimum solution of the equation (1) if

$$A(x^\dagger) = y \textit{ and } \mathcal{R}_\eta(x^\dagger) = \min_{x\in\ell_2}\{\mathcal{R}_\eta(x) \mid A(x) = y\}.$$

For any $x = (x_1, x_2, \cdots) \in \ell_2$, we define the index set

$$I(x) = \{i \in \mathbb{N} \mid x_i \neq 0\}.$$

Throughout this paper, we will assume the operator A and the data y^δ have the following properties.

Condition 2.1. (i) $A : \ell_2 \to Y$ is Fréchet differentiable.

(ii) $A : \ell_2 \to Y$ is weakly sequentially closed, i.e. $x_n \rightharpoonup x$ in ℓ_2 and $A(x_n) \rightharpoonup y$ in Y implies that $A(x) = y$.

(iii) There exists a constant $\gamma > 0$ such that

$$\|A(x_2) - A(x_1) - A'(x_1)(x_2 - x_1)\|_Y \leq \gamma\|A(x_2) - A(x_1)\|_Y \quad \forall\, x_1, x_2 \in \ell_2. \tag{7}$$

(iv) There exist $\delta > 0$ and $\tau > 1$ such that

$$\|y - y^\delta\|_Y \leq \delta < \tau\delta \leq \|A(0) - y^\delta\|_Y. \tag{8}$$

The condition (7) was used in [18, pp. 278–279], [32, pp. 6], [42, pp. 69–70]. It follows from (7) that

$$\|A'(x_1)(x_2 - x_1)\|_Y \leq (1 + \gamma)\|A(x_2) - A(x_1)\|_Y$$

which has been adopted by several researchers.

In (8), we require $\|A(0) - y^\delta\|_Y \geq \tau\delta$, which is a reasonable assumption. Indeed, if $\delta \leq \|A(0) - y^\delta\|_Y < \tau\delta$, then 0 can be viewed as a good approximation to the regularized solution $x^\delta_{\alpha,\beta}$. Moreover, in applications, it is almost impossible to recover a solution from observed data of a size in the same order as the noise.

Next, we recall some properties of $\mathcal{R}_\eta$ $(0 < \eta < 1)$ which are crucial tools in analyzing the existence of the regularization parameter α and the well-posedness of the regularization, cf. [15] for their proofs.

Lemma 2.2. *Let $0 < \eta < 1$. The function $\mathcal{R}_\eta(x)$ in (3) has the following properties:*
(i) (Coercivity) $\|x\|_{\ell_2} \to \infty$ *implies* $\mathcal{R}_\eta(x) \to \infty$.
(ii) (Weak lower semi-continuity) $x_n \rightharpoonup x$ *in* ℓ_2 *implies* $\liminf_n \mathcal{R}_\eta(x_n) \geq \mathcal{R}_\eta(x)$.
(iii) (Radon-Riesz property) $x_n \rightharpoonup x$ *in* ℓ_2 *and* $\mathcal{R}_\eta(x_n) \to \mathcal{R}_\eta(x)$ *implies* $x_n \to x$ *in* ℓ_2.

Definition 2.3. (Morozov's discrepancy principle) For fixed $0 < \eta < 1$, given $\tau > 1$, choose $\alpha = \alpha(\delta, y^\delta) > 0$ such that

$$\delta \leq \|A(x^\delta_{\alpha,\beta}) - y^\delta\|_Y \leq \tau\delta \tag{9}$$

holds for an element $x^\delta_{\alpha,\beta} \in \mathcal{L}^\delta_{\alpha,\beta}$.

The following first order necessary condition for problem (2) is standard ([16, Lemma 3. 1]).

Lemma 2.4. *Let*

$$\mathcal{J}(x) := G(x) + \Phi(x), \quad x \in \ell_2 \tag{10}$$

where G is Fréchet differentiable and Φ is convex. If $\hat{x}$ is a minimizer of $\mathcal{J}(x)$, then

$$\langle G^{'}(\hat{x}), z - \hat{x}\rangle \geq \Phi(\hat{x}) - \Phi(z) \text{ for all } z \in \ell_2. \tag{11}$$

Corollary 1. *Let $0 \neq x^\delta_{\alpha,\beta} \in \ell_2$ be a minimizer of $\mathcal{J}^\delta_{\alpha,\beta}(x)$ in (2). Then*

$$\left\langle A'(x^\delta_{\alpha,\beta})^*\left(A(x^\delta_{\alpha,\beta}) - y^\delta\right) - \frac{\beta x^\delta_{\alpha,\beta}}{\|x^\delta_{\alpha,\beta}\|_{\ell_2}}, z - x^\delta_{\alpha,\beta}\right\rangle \geq \alpha\|x^\delta_{\alpha,\beta}\|_{\ell_1} - \alpha\|z\|_{\ell_1} \text{ for all } z \in \ell_2. \tag{12}$$

Proof. Define

$$G(x) := \frac{1}{2}\|A(x) - y^\delta\|_Y^2 - \beta\|x\|_{\ell_2} \quad \text{and} \quad \Phi(x) := \alpha\|x\|_{\ell_1}.$$

Then

$$G'(x) = A'(x)^*\left(A(x) - y^\delta\right) - \frac{\beta x}{\|x\|_{\ell_2}}.$$

By Lemma 2.4, we have

$$\langle G'(x^\delta_{\alpha,\beta}), z - x^\delta_{\alpha,\beta}\rangle \geq \Phi(x^\delta_{\alpha,\beta}) - \Phi(z) \text{ for all } z \in \ell_2. \tag{13}$$

This proves the corollary. □

3. Existence of regularization parameter.

In this section, we discuss the existence of the regularization parameter α which is determined by MDP. In the rest of this paper, we let $0 < \eta < 1$ be fixed, and view α as the regularization parameter. Then, $\beta = \alpha\eta$. For fixed $0 < \eta < 1$ and noise level $\delta > 0$, define functions

$$F(x^\delta_{\alpha,\beta}) := \frac{1}{2}\|A(x^\delta_{\alpha,\beta}) - y^\delta\|_Y^2, \tag{14}$$

$$m(\alpha) := \mathcal{J}^\delta_{\alpha,\beta}(x^\delta_{\alpha,\beta}) = \min \mathcal{J}^\delta_{\alpha,\beta}(x) \tag{15}$$

for $\alpha \in (0, \infty)$.

Next, we recall some properties of $\mathcal{R}_\eta(x^\delta_{\alpha,\beta})$, $F(x^\delta_{\alpha,\beta})$ and $m(\alpha)$, cf. [44] for their proofs.

Lemma 3.1. *$\mathcal{R}_\eta(x^\delta_{\alpha,\beta})$ is non-increasing, and $F(x^\delta_{\alpha,\beta})$ and $m(\alpha)$ are non-decreasing with respect to $\alpha \in (0,\infty)$.*

Lemma 3.2. *Under Condition 2.1, there exist $\alpha_1, \alpha_2 \in \mathbb{R}^+$ such that*
$$\|A(x^\delta_{\alpha_1,\beta_1}) - y^\delta\|_Y < \delta < \tau\delta < \|A(x^\delta_{\alpha_2,\beta_2}) - y^\delta\|_Y,$$
where $\beta_1 = \alpha_1\eta$, $\beta_2 = \alpha_2\eta$.

Proof. Let $\alpha_n \to 0$ as $n \to \infty$ and denote $x_n := x^\delta_{\alpha_n,\beta_n} \in \mathcal{L}^\delta_{\alpha_n,\beta_n}$. Let $x^\dagger$ be an $\mathcal{R}_\eta$-minimum solution of the equation (1). Then,
$$\frac{1}{2}\|A(x_n) - y^\delta\|_Y^2 \le m(\alpha_n) \le \mathcal{J}^\delta_{\alpha_n,\beta_n}(x^\dagger) \le \frac{1}{2}\delta^2 + \alpha_n\mathcal{R}_\eta(x^\dagger) \to \frac{1}{2}\delta^2.$$
This implies that there exists a small enough α_1 such that $\|A(x^\delta_{\alpha_1,\beta_1}) - y^\delta\|_Y < \delta$.

Now let $\alpha_n \to \infty$ as $n \to \infty$. Then
$$0 \le \mathcal{R}_\eta(x_n) \le \frac{1}{\alpha_n}\mathcal{J}^\delta_{\alpha_n,\beta_n}(x_n) \le \frac{1}{\alpha_n}\left(\frac{1}{2}\|A(0) - y^\delta\|_Y^2\right) \to 0 = \mathcal{R}_\eta(0).$$
By Lemma 2.2 (iii), this implies that $x_n \to 0$. Since A is weakly sequentially closed, $A(x_n) \rightharpoonup A(0)$ in Y. Hence,
$$\liminf_{n\to\infty} \|A(x_n) - y^\delta\|_Y \ge \|A(0) - y^\delta\|_Y > \tau\delta.$$
So there exists a sufficiently large α_2 such that $\|A(x^\delta_{\alpha_2,\beta_2}) - y^\delta\|_Y > \tau\delta$. □

Thanks to Condition 2.1 and Lemma 2.2, the following result is standard, cf. [18] for its proof.

Lemma 3.3. *Let $\alpha_n \to \alpha > 0$ as $n \to \infty$. Denote by $x_n := x^\delta_{\alpha_n,\beta_n}$ a minimizer of $\mathcal{J}^\delta_{\alpha_n,\beta_n}(x)$. Then $\{x_n\}$ contains a convergent subsequence $\{x_{n_k}\}$ such that $x_{n_k} \to x^\delta_{\alpha,\beta}$ in ℓ_2, where $x^\delta_{\alpha,\beta}$ is a minimizer of $\mathcal{J}^\delta_{\alpha,\beta}(x)$.*

Proposition 3.4. *The function $m(\alpha)$ is continuous.*

Proof. Suppose $\alpha \to \alpha_0 \neq 0$. Let $\alpha_0/2 < \alpha < 3\alpha_0/2$, $\beta = \alpha\eta$ and $\beta_0 = \alpha_0\eta$. We have
$$\begin{aligned}\lim_{\alpha\to\alpha_0} \left[\mathcal{J}^\delta_{\alpha,\beta}(x^\delta_{\alpha,\beta}) - \mathcal{J}^\delta_{\alpha_0,\beta_0}(x^\delta_{\alpha,\beta})\right] &\le \lim_{\alpha\to\alpha_0} \left[\mathcal{J}^\delta_{\alpha,\beta}(x^\delta_{\alpha,\beta}) - \mathcal{J}^\delta_{\alpha_0,\beta_0}(x^\delta_{\alpha_0,\beta_0})\right] \\ &\le \lim_{\alpha\to\alpha_0} \left[\mathcal{J}^\delta_{\alpha,\beta}(x^\delta_{\alpha_0,\beta_0}) - \mathcal{J}^\delta_{\alpha_0,\beta_0}(x^\delta_{\alpha_0,\beta_0})\right],\end{aligned}$$
i.e.
$$\lim_{\alpha\to\alpha_0} (\alpha-\alpha_0)\mathcal{R}_\eta(x^\delta_{\alpha,\beta}) \le \lim_{\alpha\to\alpha_0} \left[\mathcal{J}^\delta_{\alpha,\beta}(x^\delta_{\alpha,\beta}) - \mathcal{J}^\delta_{\alpha_0,\beta_0}(x^\delta_{\alpha_0,\beta_0})\right] \le \lim_{\alpha\to\alpha_0} (\alpha-\alpha_0)\mathcal{R}_\eta(x^\delta_{\alpha_0,\beta_0}). \tag{16}$$
Since $x^\delta_{\alpha,\beta}$ is a minimizer of $\mathcal{J}^\delta_{\alpha,\beta}(x)$,
$$F(x^\delta_{\alpha,\beta}) + \alpha\mathcal{R}_\eta(x^\delta_{\alpha,\beta}) \le F(0) + \alpha\mathcal{R}_\eta(0) = \frac{1}{2}\|A(0) - y^\delta\|_Y^2.$$
Hence, $\mathcal{R}_\eta(x^\delta_{\alpha,\beta}) \le \|A(0) - y^\delta\|_Y^2/\alpha_0$, which implies $\mathcal{R}_\eta(x^\delta_{\alpha,\beta})$ is bounded. Thus,
$$\lim_{\alpha\to\alpha_0} (\alpha-\alpha_0)\mathcal{R}_\eta(x^\delta_{\alpha,\beta}) = \lim_{\alpha\to\alpha_0} (\alpha-\alpha_0)\mathcal{R}_\eta(x^\delta_{\alpha_0,\beta_0}) = 0. \tag{17}$$
A combination of (16) and (17) implies that
$$\lim_{\alpha\to\alpha_0} \left[\mathcal{J}^\delta_{\alpha,\beta}(x^\delta_{\alpha,\beta}) - \mathcal{J}^\delta_{\alpha_0,\beta_0}(x^\delta_{\alpha_0,\beta_0})\right] = 0.$$
This proves the lemma. □

Even though $m(\alpha)$ is continuous, the functions $F(x^\delta_{\alpha,\beta})$ and $\mathcal{R}_\eta(x^\delta_{\alpha,\beta})$ are not necessarily continuous with respect to α. If (9) does not hold for any parameter α, then $\|A(x^\delta_{\alpha,\beta}) - y^\delta\|_Y$ has a jump at a certain parameter. Actually, we have the following lemma, the proof is similar to that of Theorem 2.3 in [44].

Lemma 3.5. *For each $\alpha > 0$, there exist $x^{\delta,1}_{\alpha,\beta}, x^{\delta,2}_{\alpha,\beta} \in \mathcal{L}^\delta_{\alpha,\beta}$ such that*

$$F(x^{\delta,1}_{\alpha,\beta}) = \inf_{x\in\mathcal{L}^\delta_{\alpha,\beta}} F(x), \quad F(x^{\delta,2}_{\alpha,\beta}) = \sup_{x\in\mathcal{L}^\delta_{\alpha,\beta}} F(x).$$

A minimizer of $\mathcal{J}^\delta_{\alpha,\beta}$ may be non-unique. In general, $F(x^\delta_{\alpha,\beta})$ is set-valued and discontinuous. Hence, the inequality (9) may not hold. More precisely, there exists a regularization parameter α such that

$$\|A(x^{\delta,1}_{\alpha,\beta}) - y^\delta\|_Y < \delta < \tau\delta < \|A(x^{\delta,2}_{\alpha,\beta}) - y^\delta\|_Y. \tag{18}$$

To prove the existence of a regularization parameter α satisfying (9), it is sufficient to ensure that there are no minimizers $x^{\delta,1}_{\alpha,\beta}$, $x^{\delta,2}_{\alpha,\beta}$ of $\mathcal{J}^\delta_{\alpha,\beta}(x)$ for (18) to be valid. Next we show if there is no parameter α such that (18) holds, then there exists α satisfying (9). The proof is similar to that of Theorem 3.10 in [2].

Lemma 3.6. *Assume Condition 2.1. If there is no $\alpha > 0$ with minimizers $x^{\delta,1}_{\alpha,\beta}, x^{\delta,2}_{\alpha,\beta} \in \mathcal{L}^\delta_{\alpha,\beta}$ such that (18) is valid, then there exist $\alpha = \alpha(\delta, y^\delta) > 0$ and $x^\delta_{\alpha,\beta} \in \mathcal{L}^\delta_{\alpha,\beta}$ such that (9) holds.*

Proof. Assume that no α fulfilling (9) exists. Define two sets

$$M := \{\alpha : \exists\, x^\delta_{\alpha,\beta} \in \mathcal{L}^\delta_{\alpha,\beta} \text{ with } \|A(x^\delta_{\alpha,\beta}) - y^\delta\|_Y < \delta\},$$
$$\tilde{M} := \{\alpha : \exists\, x^\delta_{\alpha,\beta} \in \mathcal{L}^\delta_{\alpha,\beta} \text{ with } \|A(x^\delta_{\alpha,\beta}) - y^\delta\|_Y > \tau\delta\}.$$

Since there is no $\alpha > 0$ with minimizers $x^{\delta,1}_{\alpha,\beta}, x^{\delta,2}_{\alpha,\beta} \in \mathcal{L}^\delta_{\alpha,\beta}$ such that

$$\|A(x^{\delta,1}_{\alpha,\beta}) - y^\delta\|_Y < \delta < \tau\delta < \|A(x^{\delta,2}_{\alpha,\beta}) - y^\delta\|_Y,$$

$M \cap \tilde{M} = \emptyset$ and $M \cup \tilde{M} = \mathbb{R}^+$. Denote $\bar{\alpha} := \sup M$. It follows from Lemma 3.2 and the monotonicity of $F(x^\delta_{\alpha,\beta})$ with respect to α that $0 < \bar{\alpha} < \infty$. Next, we consider two cases:

Case 1. $\bar{\alpha} \in M$. Then we choose a sequence $\alpha_n \to \bar{\alpha}+$. Denote $x_n := x^\delta_{\alpha_n,\beta_n} \in \mathcal{L}^\delta_{\alpha,\beta}$. By Lemma 3.3, we can find a subsequence $\{x_{n_k}\}$ of $\{x_n\}$ and a minimizer $x^\delta_{\bar{\alpha},\bar{\beta}}$ of $\mathcal{J}^\delta_{\bar{\alpha},\bar{\beta}}(x)$ such that $x_{n_k} \to x^\delta_{\bar{\alpha},\bar{\beta}}$. Now $\alpha_{n_k} > \bar{\alpha}$, and because no parameter α satisfying (18),

$$\|A(x_{n_k}) - y^\delta\|_Y > \tau\delta \tag{19}$$

for all α_{n_k}. By Condition 2.1 (i), A is continuous. Then

$$\|A(x_{n_k}) - y^\delta\|_Y \to \|A(x^\delta_{\bar{\alpha},\bar{\beta}}) - y^\delta\|_Y < \delta, \tag{20}$$

which contradicts to (19).

Case 2. $\bar{\alpha} \notin M$. Here we choose a sequence $\alpha_n \to \bar{\alpha}-$ and we can find a subsequence $\{x_{n_k}\}$ of $\{x_n\}$ with $x_{n_k} \to x^\delta_{\bar{\alpha},\bar{\beta}}$. Then

$$\|A(x_{n_k}) - y^\delta\|_Y < \delta. \tag{21}$$

Meanwhile,

$$\|A(x_{n_k}) - y^\delta\|_Y \to \|A(x^\delta_{\bar{\alpha},\bar{\beta}}) - y^\delta\|_Y > \tau\delta, \tag{22}$$

for $k \to \infty$, which is a contradiction to (21). $\square$

Note that the inequality (9) holds under the assumption there is no α satisfying (18). This condition is difficult to verify. Next we prove that if $x_{\alpha,\beta}^{\delta,1}$, $x_{\alpha,\beta}^{\delta,2}$ satisfy (18), then we can provide an estimate for $\|A(x_{\alpha,\beta}^{\delta,2}) - y^\delta\|_Y^2$.

Lemma 3.7. *Assume $x_{\alpha,\beta}^{\delta,1}$, $x_{\alpha,\beta}^{\delta,2}$ satisfy (18). Then*

$$\tau^2\delta^2 < \|A(x_{\alpha,\beta}^{\delta,2}) - y^\delta\|_Y^2 \le (3+2\gamma)\delta^2 + \delta(2+2\gamma)\|y^\delta - A(0)\|_Y.$$

Proof. Denote $x_1 := x_{\alpha,\beta}^{\delta,1}$, $x_2 := x_{\alpha,\beta}^{\delta,2}$. If $x_1 = 0$, by (18), $\|A(0) - y^\delta\|_Y < \delta$, contradicting to the assumption $\|A(0) - y^\delta\| > \delta$. Hence, $x_1 \neq 0$. From Corollary 1, we have

$$\left\langle A'(x_1)^*\left(A(x_1) - y^\delta\right) - \frac{\beta x_1}{\|x_1\|_{\ell_2}}, z - x_1 \right\rangle \ge \alpha\|x_1\|_{\ell_1} - \alpha\|z\|_{\ell_1} \text{ for all } z \in \ell_2. \quad (23)$$

Take $z = 0$ in (23),

$$\left\langle A'(x_1)^*\left(A(x_1) - y^\delta\right) - \frac{\beta x_1}{\|x_1\|_{\ell_2}}, 0 - x_1 \right\rangle \ge \alpha\|x_1\|_{\ell_1} - \alpha\|0\|_{\ell_1}.$$

Then,

$$\begin{aligned}\alpha\|x_1\|_{\ell_1} &\le \left\langle A'(x_1)^*(A(x_1) - y^\delta) - \frac{\beta x_1}{\|x_1\|_{\ell_2}}, -x_1 \right\rangle \\ &\le \|A(x_1) - y^\delta\|_Y\|A'(x_1)x_1\|_Y + \beta\|x_1\|_{\ell_2}.\end{aligned}$$

It follows from Condition 2.1 (iii) that

$$\begin{aligned}\alpha\|x_1\|_{\ell_1} - \beta\|x_1\|_{\ell_2} &\le \|A(x_1) - y^\delta\|_Y\|A'(x_1)x_1\|_Y \\ &\le \|A(x_1) - y^\delta\|_Y(1+\gamma)\|A(0) - A(x_1)\|_Y. \quad (24)\end{aligned}$$

In addition,

$$\begin{aligned}\frac{1}{2}\|A(x_2) - y^\delta\|_Y^2 &\le \frac{1}{2}\|A(x_2) - y^\delta\|_Y^2 + \alpha\|x_2\|_1 - \beta\|x_2\|_2 \\ &= \frac{1}{2}\|A(x_1) - y^\delta\|_Y^2 + \alpha\|x_1\|_1 - \beta\|x_1\|_2. \quad (25)\end{aligned}$$

A combination of (24) and (25) implies that

$$\begin{aligned}\frac{1}{2}\|A(x_2) - y^\delta\|_Y^2 &\le \frac{1}{2}\delta^2 + \delta(1+\gamma)\|A(x_1) - y^\delta + y^\delta - A(0)\|_Y \\ &\le \frac{2\gamma+3}{2}\delta^2 + \delta(1+\gamma)\|A(0) - y^\delta\|_Y.\end{aligned}$$

This proves the lemma. ∎

From Lemmas 3.6 and 3.7, we deduce the following result.

Theorem 3.8. *Assume Condition 2.1. Then there exist a regularization parameter α such that*

$$\delta \le \|A(x_{\alpha,\beta}^\delta) - y^\delta\|_Y \le c(\delta), \quad (26)$$

where $c(\delta) := \left(\max\left\{\tau^2\delta^2, (3+2\gamma)\delta^2 + \delta(2+2\gamma)\|y^\delta - A(0)\|_Y\right\}\right)^{1/2}$.

We note that the upper bound $c(\delta)$ from Theorem 3.8 is only $O(\delta^{1/2})$. Nevertheless, in Theorem 3.10, we prove $\alpha \to 0$ as $\delta \to 0$. So if the noise level is small enough, we can obtain a good approximation of the value of α. Furthermore, in Section 4, we show $x^\delta_{\alpha,\beta} \to x^\dagger$ as $\delta \to 0$. It implies that we can obtain a satisfactory inversion results if the noise level is small enough.

For a numerical realization of MDP (26), we can make use of an iterative algorithm described in [40], cf. Algorithm 1. Even though α is determined from the upper bound: $\|A(x^\delta_{\alpha,\beta}) - y^\delta\|_Y = c(\delta)$, we can still obtain the convergence of the regularized solution. Numerical simulation results reported in Section 5 show that we can obtain better inversion results if α is determined by Algorithm 1.

Though the upper bound $c(\delta)$ seems loose, the existence of α can be guaranteed. In some applications of inverse problems, one has to assume the existence of α such that $\delta \le \|y^\delta - A(x^\delta_{\alpha,\beta})\|_Y \le c\,\delta$. Here, the upper bound is $O(\delta)$, but it is challenging to choose an appropriate c. As a result, one can not find α with an algorithm similar to Algorithm 1.

Algorithm 1 Iterative algorithm for α under MDP (5)

```
Choose τ > 0, η = 1, 0 < q < 1, α_0 > 0 with ‖A(x^δ_{α_0,β_0}) − y^δ‖_Y > c(δ)
set j = 0
while ‖A(x^δ_{α_j,β_j}) − y^δ‖_Y > c(δ)
  j = j + 1
  α_j = q α_{j−1}
  compute x^δ_{α_j,β_j}
end
α_j^max = α_{j−1}, α_j^min = α_j
while δ ≤ ‖A(x^δ_{α_j,β_j}) − y^δ‖_Y ≤ c(δ) not true
  j = j + 1
  α_j = (α_{j−1}^min + α_{j−1}^max) / 2
  compute x^δ_{α_j,β_j}
  if ‖A(x^δ_{α_j,β_j}) − y^δ‖_Y > c(δ) then α_j^max = α_j
  if ‖A(x^δ_{α_j,β_j}) − y^δ‖_Y < δ then α_j^min = α_j
end
```

In the following, we will show that for MDP, $\alpha \equiv \alpha(\delta, y^\delta) \to 0$ as the noise level $\delta \to 0$.

Lemma 3.9. *Let $\alpha > 0$ be fixed. If $A(x^*) = y$ and*

$$x^* \in \arg\min_{x \in \ell_2} \left\{ \frac{1}{2}\|A(x) - y\|_Y^2 + \alpha \mathcal{R}_\eta(x) \right\},$$

then $x^* = 0$.

Proof. By the assumption, we have

$$\begin{aligned} \alpha \mathcal{R}_\eta(x^*) &= \frac{1}{2}\|A(x^*) - y\|_Y^2 + \alpha \mathcal{R}_\eta(x^*) \\ &\le \frac{1}{2}\|A((1-t)x^*) - y\|_Y^2 + \alpha \mathcal{R}_\eta((1-t)x^*), \end{aligned} \tag{27}$$

for any $t \in (0,1)$. Since $\mathcal{R}_\eta((1-t)x^*) = (1-t)\mathcal{R}_\eta(x^*)$, we can rewrite (27) as

$$0 \leq \alpha\mathcal{R}_\eta(x^*) \leq \frac{1}{2}\frac{\|A((1-t)x^*) - y\|_Y^2}{t}.$$

Since

$$A((1-t)x^*) - y = A(x^*) + A'(x^*)(-tx^*) + o(\|-tx^*\|_{\ell_2}) - y = A'(x^*)(-tx^*) + o(t),$$

we have

$$0 \leq \alpha\mathcal{R}_\eta(x^*) \leq \frac{t}{2}\|A'(x^*)x^* + o(1)\|_Y^2.$$

Take the limit as $t \to 0^+$ to obtain $\mathcal{R}_\eta(x^*) = 0$. Since

$$(1-\eta)\|x^*\|_{\ell_1} + \eta(\|x^*\|_{\ell_1} - \|x^*\|_{\ell_2}) = \mathcal{R}_\eta(x^*) = 0,$$

we must have $x^* = 0$. $\square$

Theorem 3.10. *Let $\delta_n \to 0$ and $y^{\delta_n} \to y$ as $n \to \infty$. Then $\alpha_n \to 0$ where $\alpha_n := \alpha(\delta_n, y^{\delta_n})$ is the regularization parameter obtained from MDP (9) with δ replaced by δ_n.*

Proof. Suppose α_n does not converge to 0, i.e. $\exists\ \alpha_0 > 0$, $\forall\ N \in \mathbb{N}^+$, $\exists\ n_0 > N$ such that $|\alpha_{n_0} - 0| \geq \alpha_0$. This implies that there exists a subsequence of $\{\alpha_n\}$, still denoted by $\{\alpha_n\}$ such that $\alpha_n \geq \alpha_0$. Denote

$$x_n^{\alpha_0} \in \arg\min_{x\in\ell_2}\left\{\frac{1}{2}\|A(x) - y^{\delta_n}\|_Y^2 + \alpha_0\mathcal{R}_\eta(x)\right\}$$

and

$$x_n := x_{\alpha_n,\beta_n}^{\delta_n} \in \arg\min_{x\in\ell_2}\left\{\frac{1}{2}\|A(x) - y^{\delta_n}\|_Y^2 + \alpha_n\mathcal{R}_\eta(x)\right\}.$$

By Lemma 3.1, $\|A(x_{\alpha,\beta}^\delta) - y^\delta\|_Y^2$ is non-decreasing with respect to α. Hence,

$$\begin{aligned}\frac{1}{2}\|A(x_n^{\alpha_0}) - y^{\delta_n}\|_Y^2 &\leq \frac{1}{2}\|A(x_n) - y^{\delta_n}\|_Y^2 \\ &\leq \frac{1}{2}\max\{\tau^2\delta_n^2, (3+2\gamma)\delta_n^2 + \delta_n(2+2\gamma)\|y^{\delta_n} - A(0)\|_Y\} \to 0.\end{aligned} \tag{28}$$

By the definition of $x_n^{\alpha_0}$, there exist $x^* \in \ell_2$ and a subsequence $\{x_{n_k}^{\alpha_0}\}$ such that $x_{n_k}^{\alpha_0} \rightharpoonup x^*$ in ℓ_2. Since A is weakly sequently closed and $y^{\delta_n} \to y$, $A(x_{n_k}^{\alpha_0}) - y^{\delta_{n_k}} \rightharpoonup A(x^*) - y$, it follows from the weak lower semi-continuity of the norm that

$$\begin{aligned}0 \leq \frac{1}{2}\|A(x^*) - y\|_Y^2 &\leq \liminf_{k\to\infty}\frac{1}{2}\|A(x_{n_k}^{\alpha_0}) - y^{\delta_{n_k}}\|_Y^2 \\ &\leq \frac{1}{2}\max\{\tau^2\delta_{n_k}^2, (3+2\gamma)\delta_{n_k}^2 + \delta_{n_k}(2+2\gamma)\|y^{\delta_{n_k}} - A(0)\|_Y\} \to 0.\end{aligned} \tag{29}$$

Hence, $Ax^* = y$. In addition,

$$\begin{aligned}\frac{1}{2}\|A(x^*) - y\|_Y^2 + \alpha_0\mathcal{R}_\eta(x^*) &\leq \liminf_{k\to\infty}\left\{\frac{1}{2}\|A(x_{n_k}^{\alpha_0}) - y^{\delta_{n_k}}\|_Y^2 + \alpha_0\mathcal{R}_\eta(x_{n_k}^{\alpha_0})\right\} \\ &\leq \liminf_{k\to\infty}\left\{\frac{1}{2}\|A(x) - y^{\delta_{n_k}}\|_Y^2 + \alpha_0\mathcal{R}_\eta(x)\right\} \\ &= \frac{1}{2}\|A(x) - y\|_Y^2 + \alpha_0\mathcal{R}_\eta(x)\end{aligned} \tag{30}$$

for any $x \in \ell_2$. Thus, x^* is a minimizer of $\frac{1}{2}\|A(x)-y\|_Y^2+\alpha_0\mathcal{R}_\eta(x)$. By Lemma 3.9, this implies that $x^*=0$. Then $y=A(0)$. Hence, $\|A(0)-y^\delta\|_Y=\|y-y^\delta\|_Y\le\delta$, contradicting to $\|A(0)-y^\delta\|>\delta$ in Condition 2.1 (iv). This proves the theorem. $\square$

The $\alpha\ell_1-\beta\ell_2$ sparsity regularization has been applied to solve linear ill-posed problems, e.g. compressive sensing. Even if A is linear, $\mathcal{J}^\delta_{\alpha,\beta}$ is in general non-convex. We require the linear operator A to be bounded. Then A satisfies Condition 2.1 (ii). Hence, all lemmas and propositions in Section 2 and 3 hold. Note that if A is linear, Condition 2.1 (iii) holds for any $\gamma\ge 0$. When $\gamma=0$, we can state the following result.

Theorem 3.11. *If A in (1) is linear and bounded, then there exists a parameter α such that*

$$\delta^2 \le \|Ax^\delta_{\alpha,\beta}-y^\delta\|_Y^2 \le \max\left\{\tau^2\delta^2, 3\delta^2+2\delta\|y^\delta\|_Y\right\}. \tag{31}$$

Note that the analysis of well-posedness and convergence rate under (31) is similar to that for non-linear ill-posed problems.

4. Regularization properties.

In this section we consider the well-posedness of the regularization method. We prove that the regularized solution $x^\delta_{\alpha,\beta}$ defined by (6) converges to an $\mathcal{R}_\eta$-minimum solution of the problem $A(x)=y$. In addition, we discuss the convergence rate of $x^\delta_{\alpha,\beta}$. The proofs are along the lines of standard quadratic Tikhonov regularization ([18]) and sparsity regularization ([24, 28, 41, 46]). However, the convergence rate is different from that in [15] since the regularization parameter α is now determined by (26).

Theorem 4.1. (Convergence) *Let $x^{\delta_n}_{\alpha_n,\beta_n}$ be a minimizer of $\mathcal{J}^{\delta_n}_{\alpha_n,\beta_n}(x)$ defined by (2) with the data y^{δ_n} satisfying $\|y-y^{\delta_n}\|\le\delta_n$, where $\delta_n\to 0$ if $n\to\infty$ and y^{δ_n} belongs to the range of A. Let α_n be chosen by the discrepancy principle (26). Then there exists a subsequence of $\{x^{\delta_n}_{\alpha_n,\beta_n}\}$, still denoted by $\{x^{\delta_n}_{\alpha_n,\beta_n}\}$, that converges to an $\mathcal{R}_\eta$-minimizing solution $x^\dagger$ in ℓ_2. In addition, if the $\mathcal{R}_\eta$-minimizing solution $x^\dagger$ is unique, then*

$$\lim_{n\to\infty}\|x^{\delta_n}_{\alpha_n,\beta_n}-x^\dagger\|_{\ell_2}=0.$$

Proof. Denote $y_n:=y^{\delta_n}$, $x_n:=x^{\delta_n}_{\alpha_n,\beta_n}$, $\eta:=\beta_n/\alpha_n$. By the definition of x_n, we obtain

$$\begin{aligned}&\frac{1}{2}\|A(x_n)-y_n\|_Y^2+\alpha_n\|x_n\|_{\ell_1}-\beta_n\|x_n\|_{\ell_2}\\ &\le \frac{1}{2}\|A(x^\dagger)-y_n\|_Y^2+\alpha_n\|x^\dagger\|_{\ell_1}-\beta_n\|x^\dagger\|_{\ell_2}\\ &\le \frac{1}{2}\delta_n^2+\alpha_n\|x^\dagger\|_{\ell_1}-\beta_n\|x^\dagger\|_{\ell_2}.\end{aligned} \tag{32}$$

By (26), this implies that $\mathcal{R}_\eta(x_n)\le\mathcal{R}_\eta(x^\dagger)$. Hence, the sequence $\{\mathcal{R}(x_n)\}$ is bounded. Denote

$$c(\delta_n):=\left(\max\left\{\tau^2\delta_n^2,(3+2\gamma)\delta_n^2+\delta_n(2+2\gamma)\|y^{\delta_n}-A(0)\|_Y\right\}\right)^{\frac{1}{2}}.$$

We have $c(\delta_n)\to 0$ as $\delta_n\to 0$. In addition,

$$\|A(x_n)-y\|_Y\le\|A(x_n)-y_n\|_Y+\|y-y_n\|_Y\le c(\delta_n)+\delta_n. \tag{33}$$

Then

$$\lim_{n\to\infty}A(x_n)=y. \tag{34}$$

On the other hand, it follows from (32) that

$$\limsup_{n\to\infty} (\|x_n\|_{\ell_1} - \eta\|x_n\|_{\ell_2}) \le \|x^\dagger\|_{\ell_1} - \eta\|x^\dagger\|_{\ell_2}. \tag{35}$$

Since $\|x_n\|_{\ell_1} - \eta\|x_n\|_{\ell_2}$ is bounded, there exist an element $x^* \in \ell_2$ and a subsequence of $\{x_n\}$, still denoted by $\{x_n\}$, such that $x_n \rightharpoonup x^*$ in ℓ_2. Together with (34), it follows that

$$\|A(x^*) - y\|_Y \le \liminf_{n\to\infty} \|A(x_n) - y\|_Y = 0.$$

Hence, $Ax^* = y$. Meanwhile, by Lemma 2.2 (ii), we have

$$\begin{aligned} \|x^*\|_{\ell_1} - \eta\|x^*\|_{\ell_2} &\le \liminf_n (\|x_n\|_{\ell_1} - \eta\|x_n\|_{\ell_2}) \\ &\le \|x^\dagger\|_{\ell_1} - \eta\|x^\dagger\|_{\ell_2}. \end{aligned} \tag{36}$$

By the definition of $x^\dagger$, x^* is an $\mathcal{R}_\eta$-minimizing solution. If the $\mathcal{R}_\eta$-minimizing solution is unique, then $x^* = x^\dagger$. A combination of (35) and (36) implies $\|x_n\|_{\ell_1} - \eta\|x_n\|_{\ell_2} \to \|x^\dagger\|_{\ell_1} - \eta\|x^\dagger\|_{\ell_2}$. Thus, $\mathcal{R}_\eta(x_n) \to \mathcal{R}_\eta(x^\dagger)$. Then $\lim_{n\to\infty} \|x_n - x^\dagger\|_{\ell_2} = 0$ by Lemma 2.2 (iii). □

Assumption 4.2. *Let $x^\dagger \neq 0$ be an $\mathcal{R}_\eta$-minimizing solution of the problem $A(x) = y$ that is sparse. There exists an $\omega_i \in Y$ such that*

$$e_i = A'(x^\dagger)^* \omega_i \quad \forall\, i \in I(x^\dagger). \tag{37}$$

In addition, assume there exists a fixed $M > 0$ such that $\|x\|_{\ell_\infty} \le M$. Define $\mathcal{M} := \{x \in \ell_2 : \|x\|_\infty \le M\}$.

Assumption 4.2 and its modified form were introduced in [11, 23]. This assumption can be viewed as a source condition and it implies that the operator A fulfills some kind of "finite basis injectivity condition" which is commonly used in sparsity regularization.

Next, we present an inequality under the source condition. The linear convergence rate $O(\delta)$ (or $O(\delta^{1/2})$) can be derived from this inequality.

Lemma 4.3. *Assume Condition 2.1 and Assumption 4.2. Then there exist constants $c_1 > c_2$, depending on M, such that*

$$(\alpha - \beta)\|x - x^\dagger\|_{\ell_1} \le \mathcal{R}_{\alpha,\beta}(x) - \mathcal{R}_{\alpha,\beta}(x^\dagger) + (c_1\alpha - c_2\beta)\|A(x) - A(x^\dagger)\|_Y \tag{38}$$

for any $x \in \mathcal{M}$.

Proof. By the definition of $I(x^\dagger)$, we have

$$(\alpha - \beta)\|x - x^\dagger\|_{\ell_1} = (\alpha - \beta)\left(\sum_{i\in I(x^\dagger)} |x_i - x_i^\dagger| + \sum_{i\notin I(x^\dagger)} |x_i| \right).$$

Then,

$$\begin{aligned} &(\alpha - \beta)\|x - x^\dagger\|_{\ell_1} - (\mathcal{R}_{\alpha,\beta}(x) - \mathcal{R}_{\alpha,\beta}(x^\dagger)) \\ &= -\alpha \sum_{i\in I(x^\dagger)} \left(|x_i| - |x_i^\dagger|\right) \\ &\quad + (\alpha - \beta) \sum_{i\in I(x^\dagger)} |x_i - x_i^\dagger| + \beta(T_1 - T_2), \end{aligned} \tag{39}$$

where

$$T_1 = \left(\sum_i |x_i|^2 \right)^{1/2} - \left(\sum_{i \notin I(x^\dagger)} |x_i|^2 \right)^{1/2} - \left(\sum_{i \in I(x^\dagger)} |x_i^\dagger|^2 \right)^{1/2},$$

$$T_2 = \sum_{i \notin I(x^\dagger)} |x_i| - \left(\sum_{i \notin I(x^\dagger)} |x_i|^2 \right)^{1/2}.$$

Observe that $T_2 \geq 0$. Since

$$\left(\sum_i |x_i|^2 \right)^{1/2} \leq \left(\sum_{i \in I(x^\dagger)} |x_i|^2 \right)^{1/2} + \left(\sum_{i \notin I(x^\dagger)} |x_i|^2 \right)^{1/2},$$

we see that

$$T_1 \leq T_3 := \left(\sum_{i \in I(x^\dagger)} |x_i|^2 \right)^{1/2} - \left(\sum_{i \in I(x^\dagger)} |x_i^\dagger|^2 \right)^{1/2}. \tag{40}$$

Thus, from (39),

$$\begin{aligned}(\alpha - \beta)\|x - x^\dagger\|_{\ell_1} \leq \mathcal{R}_{\alpha,\beta}(x) - \mathcal{R}_{\alpha,\beta}(x^\dagger) + \alpha \sum_{i \in I(x^\dagger)} |x_i - x_i^\dagger| \\ + (\alpha - \beta) \sum_{i \in I(x^\dagger)} |x_i - x_i^\dagger| + \beta T_3.\end{aligned} \tag{41}$$

By the assumption, we have

$$|x_i| + |x_i^\dagger| \leq M + \|x^\dagger\|_{\ell_2},$$

$$0 < \|x^\dagger\|_{\ell_2} \leq \left(\sum_{i \in I(x^\dagger)} |x_i|^2 \right)^{1/2} + \left(\sum_{i \in I(x^\dagger)} |x_i^\dagger|^2 \right)^{1/2}.$$

Consequently,

$$T_3 = \frac{\sum\limits_{i \in I(x^\dagger)} (|x_i| - |x_i^\dagger|)(|x_i| + |x_i^\dagger|)}{\left(\sum\limits_{i \in I(x^\dagger)} |x_i|^2 \right)^{1/2} + \left(\sum\limits_{i \in I(x^\dagger)} |x_i^\dagger|^2 \right)^{1/2}} \leq \frac{M + \|x^\dagger\|_{\ell_2}}{\|x^\dagger\|_{\ell_2}} \sum_{i \in I(x^\dagger)} |x_i - x_i^\dagger|. \tag{42}$$

A combination of (41) and (42) shows that

$$(\alpha - \beta)\|x - x^\dagger\|_{\ell_1} \leq \mathcal{R}_{\alpha,\beta}(x) - \mathcal{R}_{\alpha,\beta}(x^\dagger) + \left(2\alpha + \frac{M}{\|x^\dagger\|_{\ell_2}} \beta \right) \sum_{i \in I(x^\dagger)} |x_i - x_i^\dagger|. \tag{43}$$

In addition, by Assumption 4.2,

$$|x_i - x_i^\dagger| = |\langle e_i, x - x^\dagger \rangle| = |\langle \omega_i, A'(x^\dagger)(x - x^\dagger) \rangle| \leq \max_{i \in I(x^\dagger)} \|\omega_i\|_Y \|A'(x^\dagger)(x - x^\dagger)\|_Y.$$

Hence,

$$\sum_{i \in I(x^\dagger)} |x_i - x_i^\dagger| \leq |I(x^\dagger)| \max_{i \in I(x^\dagger)} \|\omega_i\|_Y \|A'(x^\dagger)(x - x^\dagger)\|_Y.$$

Then, by Condition 2.1 (iii), we have

$$\sum_{i\in I(x^\dagger)} |x_i - x_i^\dagger| \le |I(x^\dagger)| \max_{i\in I(x^\dagger)} \|\omega_i\|_Y (1+\gamma)\|A(x) - A(x^\dagger)\|_Y. \tag{44}$$

A combination of (43) and (44) implies that

$$\begin{aligned}
&(\alpha-\beta)\|x - x^\dagger\|_{\ell_1} \\
&\le \mathcal{R}_{\alpha,\beta}(x) - \mathcal{R}_{\alpha,\beta}(x^\dagger) \\
&\quad + \left(2\alpha + \frac{M_1}{\|x^\dagger\|_{\ell_2}}\beta\right) |I(x^\dagger)| \max_{i\in I(x^\dagger)} \|\omega_i\|_Y (1+\gamma)\|A(x) - A(x^\dagger)\|_Y,
\end{aligned} \tag{45}$$

i.e.

$$(\alpha-\beta)\|x - x^\dagger\|_{\ell_1} \le \mathcal{R}_{\alpha,\beta}(x) - \mathcal{R}_{\alpha,\beta}(x^\dagger) + (c_1\alpha - c_2\beta)\|A(x) - A(x^\dagger)\|_Y,$$

where

$$c_1 = 2|I(x^\dagger)| \max_{i\in I(x^\dagger)} \|\omega_i\|_Y (1+\gamma), \quad c_2 = -\frac{M}{\|x^\dagger\|_{\ell_2}} |I(x^\dagger)| \max_{i\in I(x^\dagger)} \|\omega_i\|_Y (1+\gamma).$$

Obviously, $c_1 > c_2$ and $c_1\alpha - c_2\beta > 0$. The proof is completed. □

Theorem 4.4. (Convergence rate) *Keep the assumptions of Lemma 4.3 and let $x^\delta_{\alpha,\beta}$ be defined by (6). Assume that there exist parameters α and β ($\beta = \eta\alpha$) satisfying*

$$\delta \le \|A(x^\delta_{\alpha,\beta}) - y^\delta\|_Y \le c(\delta),$$

where $c(\delta) := \left(\max\left\{\tau^2\delta^2, (3+2\gamma)\delta^2 + \delta(2+2\gamma)\|y^\delta - A(0)\|_Y\right\}\right)^{1/2}$. *Then*

$$\|x^\delta_{\alpha,\beta} - x^\dagger\|_{\ell_2} \le \frac{(c_1 - c_2\eta)(c(\delta)+\delta)}{1-\eta}.$$

Proof. By the definitions of $x^\delta_{\alpha,\beta}$, α and β, we see that

$$\frac{1}{2}\delta^2 + \mathcal{R}_{\alpha,\beta}(x^\delta_{\alpha,\beta}) \le \frac{1}{2}\|A(x^\delta_{\alpha,\beta}) - y^\delta\|_Y^2 + \mathcal{R}_{\alpha,\beta}(x^\delta_{\alpha,\beta}) \le \frac{1}{2}\|A(x^\dagger) - y^\delta\|_Y^2 + \mathcal{R}_{\alpha,\beta}(x^\dagger). \tag{46}$$

Hence $\mathcal{R}_{\alpha,\beta}(x^\delta_{\alpha,\beta}) \le \mathcal{R}_{\alpha,\beta}(x^\dagger)$. It follows from Lemma 4.3 that

$$\begin{aligned}
0 &\ge \mathcal{R}_{\alpha,\beta}(x^\delta_{\alpha,\beta}) - \mathcal{R}_{\alpha,\beta}(x^\dagger) \\
&\ge (\alpha-\beta)\|x^\delta_{\alpha,\beta} - x^\dagger\|_{\ell_1} - (c_1\alpha - c_2\beta)\|A(x^\delta_{\alpha,\beta}) - A(x^\dagger)\|_Y \\
&\ge (\alpha-\beta)\|x^\delta_{\alpha,\beta} - x^\dagger\|_{\ell_1} - (c_1\alpha - c_2\beta)(c(\delta)+\delta).
\end{aligned} \tag{47}$$

Then

$$\|x^\delta_{\alpha,\beta} - x^\dagger\|_{\ell_2} \le \|x^\delta_{\alpha,\beta} - x^\dagger\|_{\ell_1} \le \frac{(c_1\alpha - c_2\beta)(c(\delta)+\delta)}{\alpha-\beta}.$$

The theorem is proven with $\beta = \eta\alpha$. □

5. **Numerical experiments.** In this section, we present results from three numerical experiments to demonstrate the efficiency of MDP for the $\alpha\ell_1 - \beta\ell_2$ regularization. In these experiments, the ST-($\alpha\ell_1 - \beta\ell_2$) algorithm is used to compute the iterative solution, cf. [15, 17] for details of the algorithm. The relative error (Rerror) is utilized to measure the performance of the reconstruction x^*:

$$\text{Rerror} := \frac{\|x^* - x^\dagger\|_{\ell_2}}{\|x^\dagger\|_{\ell_2}},$$

where $x^\dagger$ is a true solution. Exact data $y^\dagger$ is generated by $y^\dagger = A(x^\dagger)$. White Gaussian noise is added to the exact data $y^\dagger$ by calling $\mathrm{y}^\delta = \mathrm{awgn}(\mathrm{Ax}^\dagger, \sigma)$ in MATLAB, where σ is the added noise, measured in dB, which measures the ratio between the true (noise free) data $y^\dagger$ or $A(x^\dagger)$ and Gaussian noise. The noise level δ is defined by $\delta := \|A(x^\dagger) - y^\delta\|_2$. The first example deals with a linear ill-conditioned image deblurring problem. In the second example, we consider a nonlinear ill-posed Hammerstein equation. The third example deals with a nonlinear compressive sensing problem. All numerical experiments were tested in MATLAB R2010a on an i7-1165G7 2.80GHz workstation with 32Gb RAM.

5.1. **Ill-conditioned image deblurring problem.** In the first example, we test the ill-conditioned image deblurring problem which is the process of removing blurring artifacts from images, such as blur caused by defocus aberration or motion blur. The blur is typically modeled by a Fredholm integral equation of the first kind

$$\int_a^b K(s,t)\, f(t)\, dt = g(s),$$

where $K(s,t)$ is the kernel function, $g(s)$ is the observed image and $f(t)$ is the true image. Note that image deblurring problem is a linear ill-conditioned problem, $\gamma = 0$ and $A(0) = 0$ in (5). We utilize the blur problem from MATLAB Regularization Tools ([26]) by calling $[A, y^\dagger, x^\dagger] = \mathrm{blur}(n, \mathit{band}, \mu)$. We choose $n = 64$, $\mathit{band} = 3$, $\mu = 0.7$, $\tau = 2$. We rescale the matrix A by $A \to 0.05A$. Note that the condition number does not change under the matrix rescaling. The value of $\|y^\dagger\|_2$ is 3.15 and the condition number is around 31.4537.

To show the efficiency of MDP (5), various noise levels δ are added to the exact data $y^\dagger$. Note that, for various noise levels, if $\delta < 2\|y^\delta\|_2/(\tau^2 - 3)$, then $\tau^2\delta^2 < 3\delta^2 + 2\delta\|y^\delta\|_2$. Meanwhile, if $\tau \le \sqrt{3}$, it is obvious that $\tau^2\delta^2 < 3\delta^2 + 2\delta\|y^\delta\|_2$. With $\tau = 2$ and $\delta < \|y^\delta\|$, it is obvious that $\tau^2\delta^2 < 3\delta^2 + 2\delta\|y^\delta\|_2$.

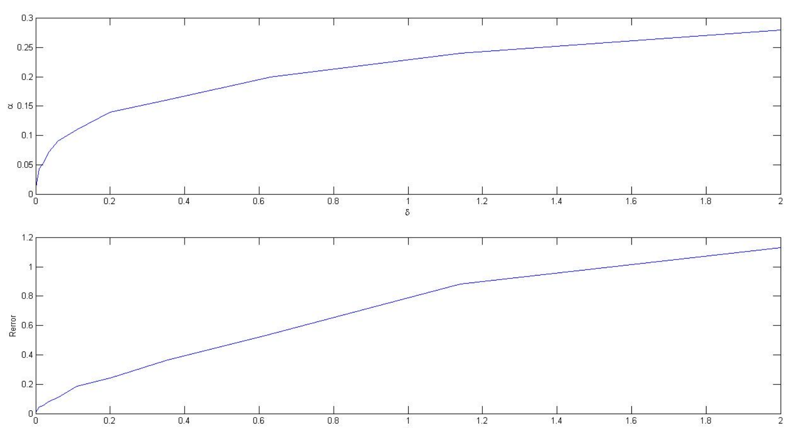

FIGURE 1. (a) α vs. δ where $\alpha : \|A(x^\delta_{\alpha,\beta}) - y^\delta\|_Y = c(\delta)$; (b) Rerror vs. δ.

To test sharpness of the upper bound in (5), we choose α with $\|A(x^\delta_{\alpha,\beta}) - y^\delta\|_Y = c(\delta)$. Note that, from the perspective of computation, $\|A(x^\delta_{\alpha,\beta}) - y^\delta\|_Y = c(\delta)$ is unsolvable. It is infeasible to obtain the true value of α such that $\|A(x^\delta_{\alpha,\beta}) - y^\delta\|_Y =$

$c(\delta)$. In computational practice, this value of α can be found approximately by a sorting from a given set of values $\alpha_1, \alpha_2, \cdots, \alpha_n$. For example, with a good initial guess α_0, we let $\alpha_n = \alpha_0 + 0.001 * n$ (or $\alpha_n = \alpha_0 - 0.001 * n$) to find a satisfactory approximate value of α by $\|A(x^\delta_{\alpha,\beta}) - y^\delta\|_Y = c(\delta)$. In Fig. 1 (a), it is shown that $\alpha \to 0$ as the noise level $\delta \to 0$. Moreover, in Fig. 1 (b), we see that Rerror decreases with respect to σ. So it illustrates the convergence of the regularized solution with respect to the noise level δ. Even though the upper bound in (5) seems loose, if the noise levels are small enough, we can still recover satisfactory results.

Next, we test the convergence of the regularized solutions where α is determined by Algorithm 1. Since the upper bound in (5) is loose, we choose a smaller value of q. In this example, we let $q = 0.1$. In Fig. 2, it is shown that $\alpha \to 0$ and Rerror converges to 0 as $\delta \to 0$.

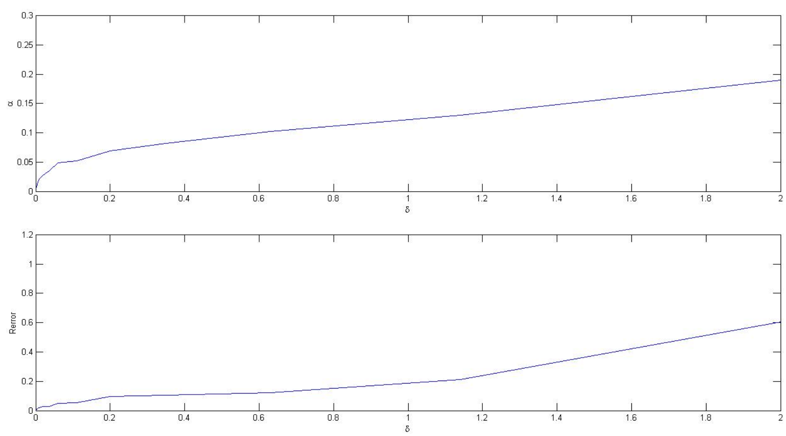

FIGURE 2. (a) α vs. δ where α is determined by Algorithm 1; (b) Rerror vs. δ.

Graphs of the reconstruction x^* corresponding to $\sigma = 70$ ($\delta = 0.0102$ and $c(\delta) = 0.18$) are provided in Fig. 3. In Fig. 3 (c), the reconstruction x^* is computed with $\alpha : \|A(x^\delta_{\alpha_j,\beta_j}) - y^\delta\|_Y = c(\delta)$; Rerror is 5.38%. In Fig. 3 (d), the reconstruction x^* is computed with α determined by Algorithm 1; Rerror is 2.89%. It is shown that we can obtain better results if α determined by Algorithm 1.

5.2. **Nonlinear Hammerstein equation.** In the second example, we test a nonlinear Hammerstein equation $A(x) = y$ with

$$A(x)(t) = \int_0^t x(s)^2\, ds.$$

The Fréchet derivative of A is of the form ([25])

$$(A'(x)h)(t) = \int_0^t 2x(s)h(s)ds.$$

The nonlinearity assumption (7) with $\gamma = 1/2$ on the forward operator A is verified in [25]. The exact solution $x^\dagger$ is s-sparse, i.e. $\|x^\dagger\|_0 = s$. We let $n = 100$ and $s = 50$. The exact data $y^\dagger$ is obtained by $y^\dagger = A(x^\dagger)(t) = \int_0^t x^\dagger(s)^2 ds$. The value

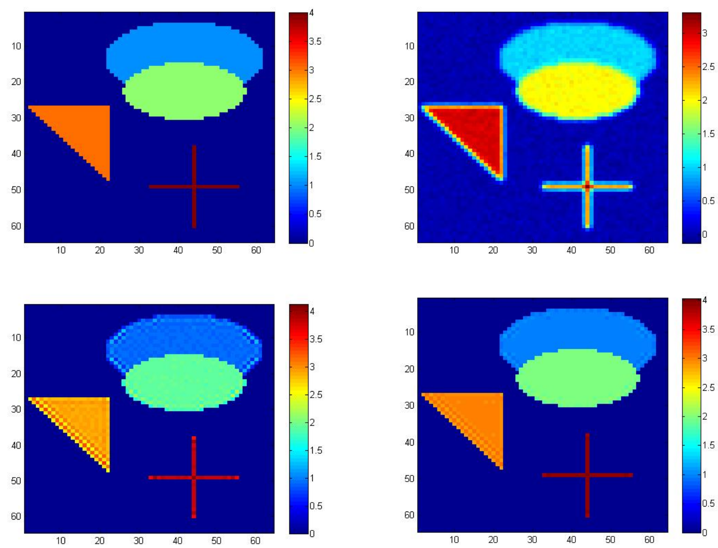

FIGURE 3. (a) True image; (b) Observed image; (c) Recovered image with α : $\|A(x^\delta_{\alpha_j,\beta_j}) - y^\delta\|_Y = c(\delta)$; (d) Recovered image with α is determined by Algorithm 1.

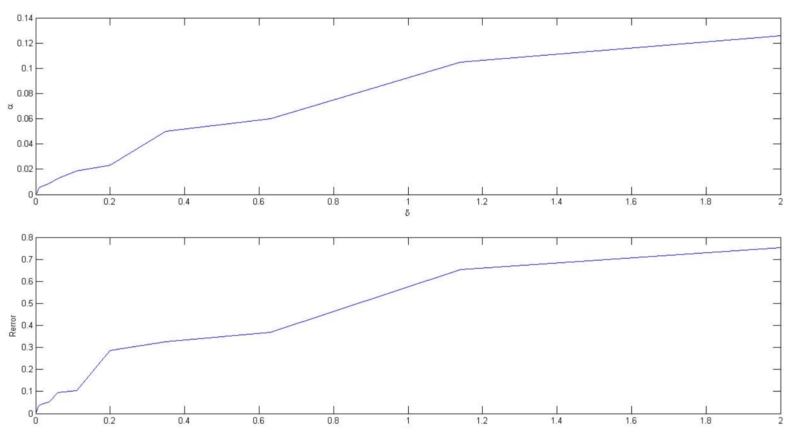

FIGURE 4. (a) α vs. δ where α : $\|A(x^\delta_{\alpha,\beta}) - y^\delta\|_Y = c(\delta)$; (b) Rerror vs. δ.

of $\|A(0) - y^\dagger\|_2$ is 1.3760. With $\tau = 2$ and $\delta < \|A(0) - y^\delta\|$, it is obvious that $\tau^2\delta^2 < (3+\gamma)\delta^2 + (2+\gamma)\delta\|y^\delta\|_2$.

In Fig. 4, we test sharpness of the upper bound in (5). In Fig. 4 (a), it is seen that $\alpha \to 0$ as the noise level $\delta \to 0$, where α is determined with $\|A(x^\delta_{\alpha,\beta}) - y^\delta\|_Y = c(\delta)$. In Fig. 4 (b), we see that Rerror decreases with respect to σ. The numerical results illustrate the convergence of the regularized solution with respect to the noise level δ.

Next, we test the convergence of the regularized solutions where α is determined by Algorithm 1. In Fig. 5, it is seen that $\alpha \to 0$ as the noise level $\delta \to 0$ and Rerror converges to 0 as $\delta \to 0$.

Fig. 6 provides graphs of the reconstruction x^* corresponding to $\sigma = 60$ ($\delta = 0.0114$ and $c(\delta) = 0.2177$). In Fig. 6 (c), the reconstruction x^* is computed with $\alpha : \|A(x^\delta_{\alpha,\beta}) - y^\delta\|_Y = c(\delta)$; Rerror is 8.28%. In Fig. 6 (d), the reconstruction x^* is computed with α determined by Algorithm 1; Rerror is 4.65%. We obtain better results when α is determined by Algorithm 1.

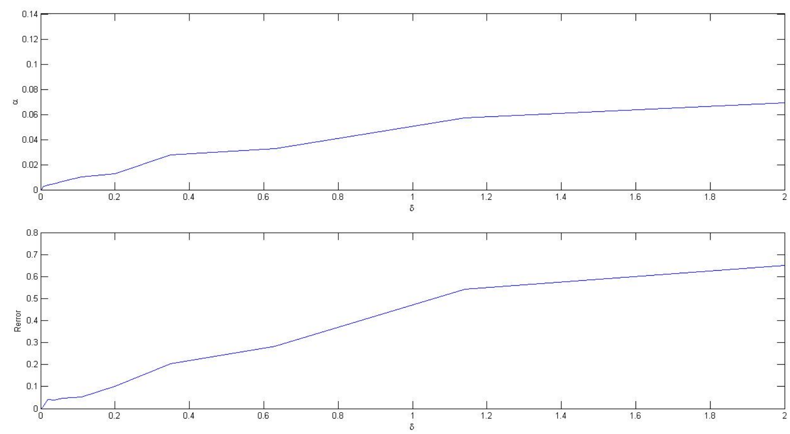

FIGURE 5. (a) α vs. δ where α is determined by Algorithm 1; (b) Rerror vs. δ.

5.3. **Nonlinear compressive sensing.** As the third example, we test the efficiency of MDP in (5) with a nonlinear compressive sensing (CS) problem ([4, 6, 13, 48]) of the form

$$y = A(x) := \hat{a}(\Phi \hat{b}(x)), \tag{48}$$

where Φ is a CS matrix, $\hat{a}$ and $\hat{b}$ are nonlinear operators, respectively. For simplicity, we write $\hat{a}(x) = x + a(x)$ and $\hat{b}(x) = x + b(x)$, where a and b are nonlinear maps. In particular, let $a(x) = x^c$ and $b(x) = x^d$, where $c, d \in \mathbb{N}_+$. An explicit formula of the Fréchet derivative of A is derived in [6]:

$$A'(x) = [I + a'_x(\Phi \hat{b}(x))][\Phi(I + b'_x(x))]$$

and $A'(x)$ is Lipschitz continuous. It is not known whether the forward operator A fulfils the nonlinearity assumption (7) ([6]). Here, we consider a potential application of MDP (5) in nonlinear compressive sensing.

In finite-dimensional spaces, the nonlinear CS problem is of the form $\hat{a}(\Phi_{m\times n}\hat{b}(x_n)) = y_m$, where $\Phi_{m\times n}$ is a Gaussian random measurement matrix. The exact solution $x^\dagger$ is s-sparse. We let $n = 200$, $m = 0.4n$, $s = 0.2m$, $\gamma = 1/2$, $c = 2$ and $d = 3$. The exact data $y^\dagger$ is obtained by $y^\dagger = \hat{a}(\Phi\hat{b}(x^\dagger))$.

Next we show if τ is chosen small, the existence of α can not be guaranteed under the condition (4). We let $\sigma = 30$, then $\delta = 0.2903$ and $c(\delta) = 1.1675$. In Fig. 7, it is shown if $\tau < 2$, then $\tau\delta < 0.5823$. Consequently, there is no α such that the traditional MDP (4) holds. Even with a large value, e.g. $\tau = 3$, it is challenging to find α such that (4) holds. Indeed, for traditional MDP with linear

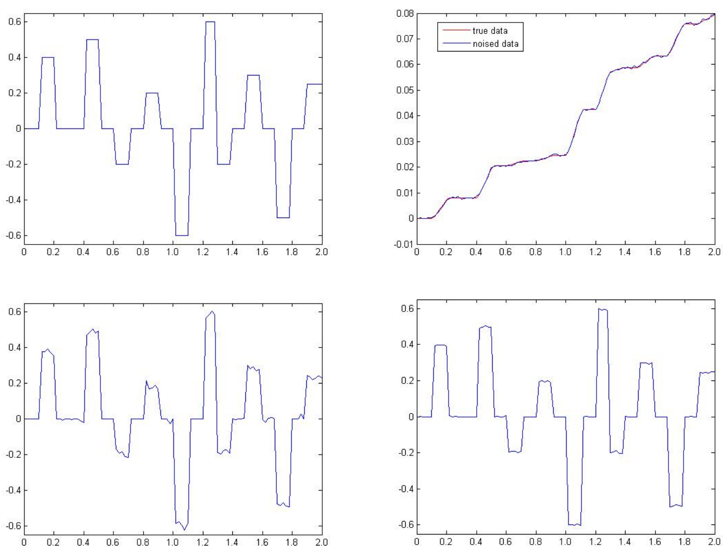

FIGURE 6. (a) True signal; (b) Observed data; (c) Recovered signal with $\alpha : \|A(x^{\delta}_{\alpha_j,\beta_j}) - y^{\delta}\|_Y = c(\delta)$; Rerror= 0.0828; (d) Recovered signal with α is determined by Algorithm 1; Rerror= 0.0465.

ill-posed problems, one commonly tries $\alpha_j = \frac{\alpha}{2^j}$, $j = 1, 2, \cdots$. With j increasing, we calculate $x^{\delta}_{\alpha,\beta}$ until (4) is valid. Nevertheless, for nonlinear ill-posed problems, the discrepancy is not continuous with respect to α. If $\alpha_j = 0.2$, then $\alpha_{j+1} = 0.1$, we can not find an α such that $\delta \le \|A(x^{\delta}_{\alpha,\beta}) - y^{\delta}\|_Y \le 3\delta$ holds (cf. Fig. 7). Certainly, we can try $\alpha_j = \frac{\alpha}{3^j}$ or $\alpha_j = \frac{\alpha}{4^j}$, etc. Nevertheless, it needs more computational time. In addition, one still does not have the theoretical assurance on the existence of α. So it is challenging to choose an appropriate τ to ensure the existence of α. Although the upper bound $c(\delta)$ in (5) seems loose, it guarantees the existence of α. We can determine an α by Algorithm 1 such that the modified MDP (5) holds. So the upper bound $c(\delta)$ appears to be natural.

In Fig. 8 (a), it is shown that $\alpha \to 0$ as the noise level $\delta \to 0$, where α is determined with $\|A(x^{\delta}_{\alpha,\beta}) - y^{\delta}\|_Y = c(\delta)$. In Fig. 8 (b), we see the decreasing tendency of Rerror with respect to σ. In Fig. 9, it is shown that $\alpha \to 0$ as the noise level $\delta \to 0$ and Rerror converge to 0 as $\delta \to 0$, where α is determined by Algorithm 1.

Fig. 10 provides graphs of the reconstruction x^* corresponding to $\sigma = 60$ ($\delta = 0.0095$ and $c(\delta) = 0.3498$). In Fig. 10 (c), the reconstruction x^* is computed with α from $\|A(x^{\delta}_{\alpha,\beta}) - y^{\delta}\|_Y = c(\delta)$; Rerror is 7.34%. In Fig. 10 (d), the reconstruction x^* is computed with α determined by Algorithm 1; Rerror is 1.37%.

6. **Conclusion.** For the non-convex $\alpha\ell_1 - \beta\ell_2$ sparsity regularization, we prove there exist a regularization parameter α such that

$$\delta \le \|A(x^{\delta}_{\alpha,\beta}) - y^{\delta}\|_Y \le c(\delta),$$

where $c(\delta) = \left(\max\left\{\tau^2\delta^2, (3+2\gamma)\delta^2 + \delta(2+2\gamma)\|y^{\delta} - A(0)\|_Y\right\}\right)^{1/2}$. Furthermore, we show that $\alpha \equiv \alpha(\delta, y^{\delta}) \to 0$ as the noise level $\delta \to 0$. It is shown that the $\alpha\ell_1 - \beta\ell_2$

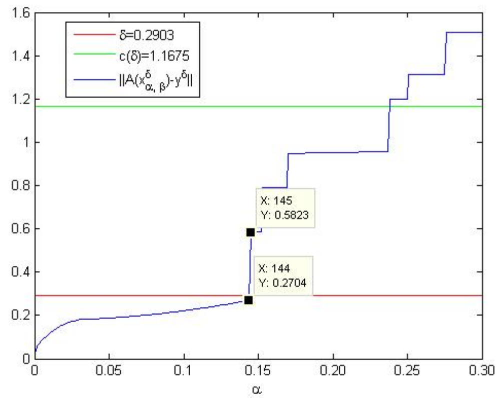

FIGURE 7. δ, the upper bound $c(\delta)$ and $\|A(x^\delta_{\alpha,\beta}) - y^\delta\|$ vs. α.

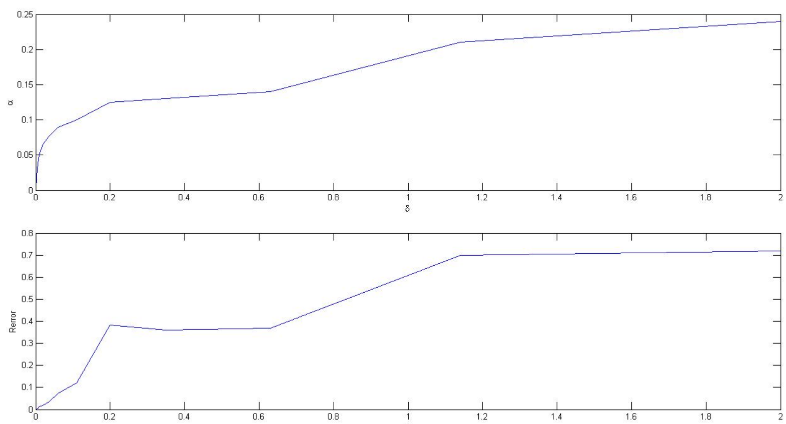

FIGURE 8. (a) α vs. δ where $\alpha : \|A(x^\delta_{\alpha,\beta}) - y^\delta\|_Y = c(\delta)$; (b) Rerror vs. δ.

sparsity regularization is well-posed when α is determined by MDP. Under certain conditions, we prove $\|x^\delta_{\alpha,\beta} - x^\dagger\|_{\ell_2} \leq O(c(\delta) + \delta)$. For linear ill-posed problems, we prove there exists a regularization parameter α such that

$$\delta^2 \leq \|Ax^\delta_{\alpha,\beta} - y^\delta\|_Y^2 \leq \max\left\{\tau^2\delta^2, 3\delta^2 + 2\delta\|y^\delta\|_Y\right\}$$

holds. Numerical experiments indicate that the relative error decreases with respect to δ. Even though the upper bound in (5) seems loose, the existence of α can be guaranteed. In addition, if the noise levels are small enough, we can still recover satisfactory results.

Acknowledgments. The first author was supported by the Fundamental Research Funds for the Central Universities (no. 2572021DJ03), Heilongjiang Postdoctoral Research Developmental Fund (no. LBH-Q16008), the National Nature Science Foundation of China (no. 41304093). The second author was supported by Simons Foundation Collaboration (No. 850737).

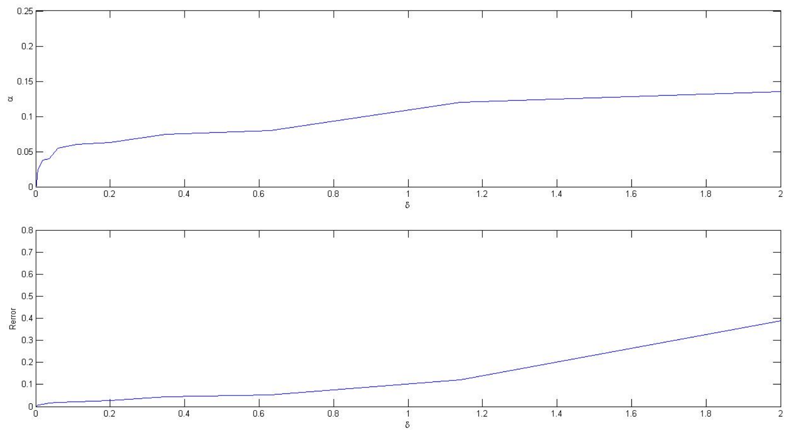

FIGURE 9. (a) α vs. δ where α is determined by Algorithm 1; (b) Rerror vs. δ.

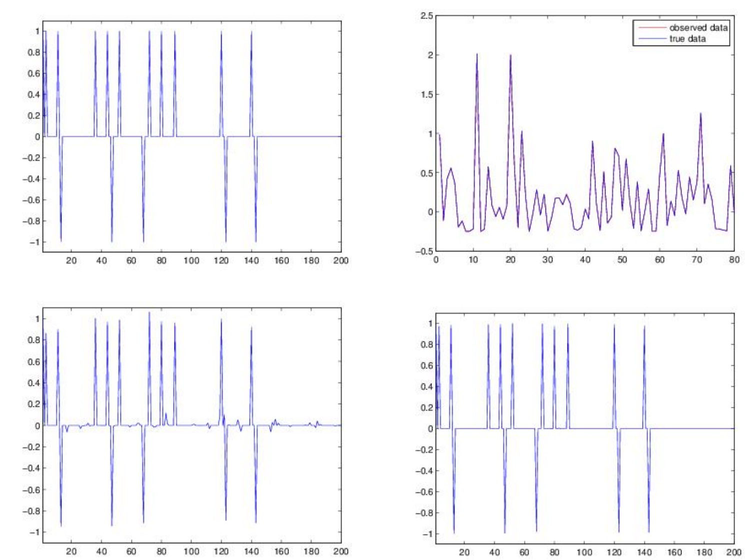

FIGURE 10. (a) True signal; (b) Observed data; (c) Recovered signal with $\alpha : \|A(x^{\delta}_{\alpha,\beta}) - y^{\delta}\|_Y = c(\delta)$ Rerror= 0.0734; (d) Recovered signal with α is determined by Algorithm 1 Rerror= 0.0137.

REFERENCES

[1] V. Albani and A. De Cezaro, A connection between uniqueness of minimizers in Tikhonov-type regularization and Morozov-like discrepancy principles, *Inverse Problems and Imaging*, **13** (2019), 211–229.

[2] S. W. Anzengruber and R. Ramlau, Morozov's discrepancy principle for Tikhonov-type functionals with nonlinear operators, *Inverse Problems*, **26** (2010), 025001, 17 pp.

[3] S. W. Anzengruber and R. Ramlau, Convergence rates for Morozov's discrepancy principle using variational inequalities, *Inverse Problems*, **27** (2011), 105007, 18 pp.

[4] A. Beck and Y. C. Eldar, Sparsity constrained nonlinear optimization: Optimality conditions and algorithms, *SIAM Journal on Optimization*, **23** (2013), 1480–1509.

[5] M. Benning and M. Burger, Modern regularization methods for inverse problems, *Acta Numerica*, **27** (2018), 1–111.

[6] T. Blumensath, Compressed sensing with nonlinear observations and related nonlinear optimization problems, *IEEE Transactions on Information Theory*, **59** (2013), 3466–3474.

[7] T. Blumensath and M. E. Davies, Iterative thresholding for sparse approximations, *J. Fourier Anal. Appl.*, **14** (2008), 629–654.

[8] T. Blumensath and M. E. Davies, Iterative hard thresholding for compressed sensing, *Appl. Comput. Harmon. Anal.*, **27** (2009), 265–274.

[9] T. Bonesky, Morozov's discrepancy principle and Tikhonov-type functionals, *Inverse Problems*, **25** (2009), 015015, 11 pp.

[10] K. Bredies and D. A. Lorenz, Iterated hard shrinkage for minimization problems with sparsity constraints, *SIAM J. Sci. Comput.*, **30** (2008), 657–683.

[11] K. Bredies and D. A. Lorenz, Regularization with non-convex separable constraints, *Inverse Problems*, **25** (2009), 085011, 14 pp.

[12] K. Bredies, D. A. Lorenz and P. Maass, A generalized conditional gradient method and its connection to an iterative shrinkage method, *Comput. Optim. Appl.*, **42** (2009), 173–193.

[13] S. P. Chepuri and G. Leus, Sparsity-promoting sensor selection for non-linear measurement models, *IEEE Trans. Signal Process.*, **63** (2015), 684–698.

[14] I. Daubechies, M. Defrise and C. De Mol, Sparsity-enforcing regularisation and ISTA revisited, *Inverse Problems*, **32** (2016), 104001, 15 pp.

[15] L. Ding and W. Han, $\alpha\ell_1 - \beta\ell_2$ regularization for sparse recovery, *Inverse Problems*, **35** (2019), 125009, 26 pp.

[16] L. Ding and W. Han, A projected gradient method for $\alpha\ell_1 - \beta\ell_2$ sparsity regularization, *Inverse Problems*, **36** (2020), 125012, 30 pp.

[17] L. Ding and W. Han, $\alpha\ell_1 - \beta\ell_2$ sparsity regularization for nonlinear ill-posed problems, preprint, 2020, `arXiv:2007.11377v1`.

[18] H. W. Engl, M. Hanke and A. Neubauer, *Regularization of Inverse Problems*, Mathematics and its Applications vol 375: Dordrecht: Kluwer, 1996.

[19] E. Esser, Y. Lou and J. Xin, A method for finding structured sparse solutions to non-negative least squares problems with applications, *SIAM J. Imaging Sci.*, **6** (2013), 2010–2046.

[20] M. Fornasier, eds, *Theoretical Foundations and Numerical Methods for Sparse Recovery*, De Gruyter, 2010.

[21] M. Fornasier and H. Rauhut, Iterative thresholding algorithms, *Applied and Computational Harmonic Analysis*, **25** (2008), 187–208.

[22] A. Frommer and P. Maass, Fast CG-based methods for Tikhonov-Phillips regularization, *SIAM J. Sci. Comp.*, **20** (1999), 1831–1850.

[23] M. Grasmair, Well-posedness and convergence rates for sparse regularization with sublinear ℓ^q penalty term, *Inverse Problems and Imaging*, **3** (2009), 383–387.

[24] M. Grasmair, M. Haltmeier and O. Scherzer, Sparse regularization with ℓ^q penalty term, *Inverse Problems*, **24** (2008), 055020, 13 pp.

[25] M. Hanke, A. Neubauer and O. Scherzer, A convergence analysis of the Landweber iteration for nonlinear ill-posed problems, *Numer. Math.*, **72** (1995), 21–37.

[26] P. C. Hansen, Regularization Tools version 4.0 for Matlab 7.3, *Numerical Algorithms*, **46** (2007), 189–194.

[27] X.-L. Huang, L. Shi and M. Yan, Nonconvex sorted ℓ_1 minimization for sparse approximation, *J. Oper. Res. Soc. China*, **3** (2015), 207–229.

[28] B. Jin, D. A. Lorenz and S. Schiffer, Elastic-net regularization: Error estimates and active set methods, *Inverse Problems*, **25** (2009), 115022, 26 pp.

[29] B. Jin and P. Maass, Sparsity regularization for parameter identification problems, *Inverse Problems*, **28** (2012), 123001, 70 pp.

[30] B. Jin, P. Maass and O. Scherzer, Sparsity regularization in inverse problems, *Inverse Problems*, **33** (2017), 060301, 4 pp.

[31] B. Jin, Y. Zhao and J. Zou, Iterative parameter choice by discrepancy principle, *IMA J. Numer. Anal.*, **32** (2012), 1714–1732.

[32] B. Kaltenbacher, A. Neubauer and O. Scherzer, *Iterative Regularization Methods for Nonlinear Ill-posed Problems*, De Gruyter, 2012.

[33] D. Lazzaro, E. L. Piccolomini and F. Zama, A nonconvex penalization algorithm with automatic choice of the regularization parameter in sparse imaging, *Inverse Problems*, **35** (2019), 084002, 17 pp.

[34] P. Li, W. Chen, H. Ge and M. K. Ng, ℓ_1-$\alpha\ell_2$ minimization methods for signal and image reconstruction with impulsive noise removal, *Inverse Problems*, **36** (2020), 055009, 30 pp.

[35] Y. Lou and M. Yan, Fast L1-L2 minimization via a proximal operator, *J. Sci. Comput.*, **74** (2018), 767–785.

[36] A. K. Louis, *Inverse und Schlecht Gestellte Probleme*, Teubner: Stuttgart. 1989.

[37] P. Maass, S. V. Pereverzev, R. Ramlau and S. G. Solodky, An adaptive discretization scheme for Tikhonov-regularization with a posteriori parameter selection, *Numerische Mathematik*, **87** (2001), 485–502.

[38] L. B. Montefusco, D. Lazzaro and S. Papi, A fast algorithm for nonconvex approaches to sparse recovery problems, *Signal Proc.*, **93** (2013), 2636–2647.

[39] V. A. Morozov, On the solution of functional equations by the method of regularization, *Soviet Math. Dokl.*, **7** (1966), 414–417.

[40] R. Ramlau, Morozov's discrepancy principle for Tikhonov regularization of nonlinear operators, *Numer. Funct. Anal. Optim.*, **23** (2002), 147–172.

[41] O. Scherzer, M. Grasmair, H. Grossauer, M. Haltmeier and F. Lenzen, *Variational Methods in Imaging*, Applied Mathematical Sciences, Vol.167, Newyork: Springer, 2009.

[42] T. Scherzer, B. Kaltenbacher, B. Hofmann and K. Kazimierski, *Regularization Methods in Banach Spaces*, De Gruyter, 2012.

[43] A. N. Tikhonov and V. Y. Arsenin, *Solutions of Ill-Posed Problems*, Winston Wiley: New York, 1977.

[44] A. N. Tikhonov, A. S. Leonov and A. G. Yagola, *Nonlinear Ill-Posed Problems*, London: Chapman & Hall, 1998.

[45] W. Wang, S. Lu, B. Hofmann and J. Cheng, Tikhonov regularization with ℓ_0-term complementing a convex penalty: ℓ_1-convergence under sparsity constraints, *J. Inverse Ill-Posed Probl.*, **27** (2019), 575–590.

[46] W. Wang, S. Lu, H. Mao and J. Cheng, Multi-parameter Tikhonov regularization with the ℓ_0 sparsity constrain, *Inverse Problems*, **29** (2013), 065018, 18 pp.

[47] L. Yan, Y. Shin and D. Xiu, Sparse approximation using ℓ_1-ℓ_2 minimization and its application to stochastic collocation, *SIAM J. Sci. Comput.*, **39** (2017), A214–A239.

[48] S. Yang, M. Wang, P. Li, L. Jin, B. Wu and L. Jiao, Compressive hyperspectral imaging via sparse tensor and nonlinear compressed sensing, *IEEE Transactions on Geoscience and Remote Sensing*, **53** (2015), 5943–5957.

[49] P. Yin, Y. Lou, Q. He and J. Xin, Minimization of ℓ_{1-2} for compressed sensing, *SIAM J. Sci. Comput.*, **37** (2015), A536–A563.

Received January 2022; revised April 2022; early access June 2022.

E-mail address: dl@nefu.edu.cn
E-mail address: weimin-han@uiowa.edu