Journal of Computational and Applied Mathematics 450 (2024) 115987

Contents lists available at ScienceDirect

Journal of Computational and Applied Mathematics

journal homepage: www.elsevier.com/locate/cam

$\alpha\ell_1 - \beta\ell_2$ sparsity regularization for nonlinear ill-posed problems

Liang Ding [a,1,*], Weimin Han [b]

[a] Department of Mathematics, Northeast Forestry University, Harbin 150040, China
[b] Department of Mathematics, University of Iowa, Iowa City, IA 52242, USA

ARTICLE INFO

Keywords:
Sparsity regularization
Nonlinear inverse problem
$\alpha\ell_1$-$\beta\ell_2$ regularization
Non-convex
Iterative soft thresholding algorithm

ABSTRACT

In this paper, the $\alpha\|\cdot\|_{\ell_1} - \beta\|\cdot\|_{\ell_2}$ sparsity regularization with parameters $\alpha \geq \beta \geq 0$ is studied for nonlinear ill-posed inverse problems. The well-posedness of the regularization is investigated. Compared to the case where $\alpha > \beta \geq 0$, the results for the case $\alpha = \beta > 0$ are weaker due to the lack of coercivity and Radon-Riesz property of the regularization term. Under certain conditions on the nonlinearity of F, sparsity is shown for every minimizer of the $\alpha\|\cdot\|_{\ell_1} - \beta\|\cdot\|_{\ell_2}$ regularized inverse problem. Moreover, for the case $\alpha > \beta \geq 0$, convergence rates $O(\delta^{\frac{1}{2}})$ and $O(\delta)$ are proved for the regularized solution toward a sparse exact solution, under different yet commonly adopted conditions on the nonlinearity of F. The iterative soft thresholding algorithm is shown to be useful to solve the $\alpha\|\cdot\|_{\ell_1} - \beta\|\cdot\|_{\ell_2}$ regularized problem for nonlinear ill-posed equations. Numerical results illustrate the efficiency of the proposed method.

1. Introduction

Investigation of the non-convex $\alpha\|\cdot\|_{\ell_1} - \beta\|\cdot\|_{\ell_2}$ $(\alpha \geq \beta \geq 0)$ regularization has attracted attention in the field of sparse recovery in the recent years, see [1–5] and references therein. As an alternative to the ℓ_p-norm with $0 \leq p < 1$, the advantages of using the functional $\alpha\|\cdot\|_{\ell_1} - \beta\|\cdot\|_{\ell_2}$ $(\alpha \geq \beta \geq 0)$ lie in the fact that it is a good approximation of the ℓ_0-norm and it has a simpler structure than the ℓ_0-norm from the perspective of computation. It is known to be difficult to determine the optimal exponent p for the ℓ_p $(0 \leq p < 1)$ regularization [6]. For the $\alpha\|\cdot\|_{\ell_1} - \beta\|\cdot\|_{\ell_2}$ regularization, it can be shown that $\eta = \beta/\alpha$ plays a role similar to that of p in the ℓ_p regularization, see [1, Fig. 1] for details. In this paper, we investigate the potential of the $\alpha\|\cdot\|_{\ell_1} - \beta\|\cdot\|_{\ell_2}$ regularization method for solving nonlinear ill-posed operator equations with sparse solutions. In addition, we analyze the well-posedness of the regularization for the particular case $\alpha = \beta$.

We are interested in solving an ill-posed operator equation of the form

$$F(x) = y, \tag{1.1}$$

where x is sparse, $F : \ell_2 \to Y$ is a weakly sequentially closed nonlinear operator mapping between the ℓ_2 space and a Hilbert space Y with norms $\|\cdot\|_{\ell_2}$ and $\|\cdot\|_Y$, respectively. Throughout this paper, we let $\langle\cdot,\cdot\rangle$ denote the inner product in the ℓ_2 space and $e_i = (\underbrace{0,\ldots,0,1}_{i},0,\ldots)$, $i \geq 1$. The exact data $y^\dagger$ and the observed data y^δ satisfy $\|y^\delta - y^\dagger\|_Y \leq \delta$ with a noise level $\delta > 0$. The most commonly adopted technique to solve the problem (1.1) is sparsity regularization, see the monographs [7,8] and the special issues [9–12] for many developments on regularizing properties and minimization schemes.

* Corresponding author.
E-mail addresses: dl@nefu.edu.cn (L. Ding), weimin-han@uiowa.edu (W. Han).
1 The work of this author was supported by the Fundamental Research Funds for the Central Universities (no. 2572021DJ03).

https://doi.org/10.1016/j.cam.2024.115987
Received 16 February 2023; Received in revised form 6 August 2023
Available online 10 May 2024

1.1. Related works

The first theoretical analysis on sparsity regularization for ill-posed inverse problems dates back to 2004. In the seminal paper [13], Daubechies et al. proposed an ℓ_p ($1 \le p \le 2$) sparsity regularization for linear ill-posed problems and established the convergence of an iterative soft thresholding algorithm. Inspired by [13], many investigations focused on the regularizing properties and iteration schemes for linear ill-posed inverse problems, see [7–9]. Subsequently, the schemes and their analysis were quickly extended to nonlinear ill-posed inverse problems. Much effort has been devoted to investigating the regularization properties as well as the minimization of the sparsity regularization for nonlinear ill-posed inverse problems, see [11,14–19] and the references therein. We emphasize that in the above cited references only the convex case $p \ge 1$ is investigated. For the non-convex case $0 \le p < 1$, special conditions and techniques are needed to analyze the well-posedness and convergence rate. In [20], a sub-linear ℓ_p regularization is proposed and convergence is proved in the sense of the weak* topology on ℓ_1. A multi-parameter Tikhonov regularization with ℓ_0 constraint is presented in [21,22], where results on regularizing properties and convergence rates are obtained. In [23], with the use of a superposition operator $\mathcal{N}_{p,q}$, the sparsity regularization with $0 \le p < 1$ can be studied within a more classical convex formulation with $1 \le q \le 2$. Then the well-known results on regularizing properties of convex sparsity regularization can be utilized to analyze the original non-convex sparsity regularization.

Concerning the minimization of the ℓ_p sparsity regularization with $0 \le p < 1$, several numerical algorithms were developed for linear ill-posed inverse problems, e.g. alternating direction method of multipliers (ADMM) [24], iteratively reweighted least squares (IRLS) [25], primal–dual active set method [26] and iterative hard thresholding [27]. Unfortunately, these algorithms cannot be extended to nonlinear ill-posed equations directly. Sparsity regularization with non-convex regularized terms for nonlinear ill-posed inverse problems is far from being investigated systematically. Though there is a great potential in the non-convex sparsity regularization for nonlinear ill-posed inverse problems, to the best of our knowledge, only two papers are available in the literature. In [28], the non-convex Tikhonov functional is transformed to a more viable one. Then a surrogate functional approach is applied to the new convex functional straightforwardly. In [29], an iterative algorithm is developed and analyzed, which aims at minimizing non-smooth and non-convex functionals, covering the important special case of Tikhonov functionals for non-linear operators and non-convex penalty terms.

1.2. Contribution and organization

In this paper, we solve the nonlinear ill-posed inverse problem (1.1) by the following regularization method:

$$\min_{x\in\ell_2} \mathcal{J}^{\delta}_{\alpha,\beta}(x), \tag{1.2}$$

with

$$\mathcal{J}^{\delta}_{\alpha,\beta}(x) = \frac{1}{q}\|F(x) - y^{\delta}\|^q_Y + \mathcal{R}_{\alpha,\beta}(x),$$

where $q \ge 1$ and

$$\mathcal{R}_{\alpha,\beta}(x) := \alpha\|x\|_{\ell_1} - \beta\|x\|_{\ell_2}, \quad \alpha \ge \beta \ge 0. \tag{1.3}$$

For $\alpha > 0$, denoting $\eta = \beta/\alpha$, we can equivalently express the functional in (1.3) as

$$\mathcal{R}_{\alpha,\beta}(x) = \alpha\,\mathcal{R}_{\eta}(x),$$

where $\mathcal{R}_{\eta}(x) := \|x\|_{\ell_1} - \eta\|x\|_{\ell_2}$, $1 \ge \eta \ge 0$. We will investigate the well-posedness of the problem (1.2). For the case $\alpha > \beta \ge 0$, we show the existence, stability as well as convergence of regularized solutions under the assumption that the nonlinear operator F is weakly sequentially closed. The numerical results reported in [1] show that we can obtain satisfactory results even when $\alpha = \beta$. Actually, $\mathcal{R}_{\alpha,\beta}$ behaves more and more like a constant multiple of the ℓ_0-norm as $\beta/\alpha \to 1$. So in this paper, we also analyze properties of $\mathcal{R}_{\alpha,\beta}$ when $\alpha = \beta$, even though the well-posedness results of the regularization are weaker than that in the case $\alpha > \beta \ge 0$. For the case $\alpha > \beta \ge 0$, we identify the convergence rate under an appropriate source condition. As is standard in analyzing convergence rates, we need to impose restrictions on the nonlinearity of the operator F. Typically, the restrictions are utilized to bound the crucial term $\langle F'(x^{\dagger})(x - x^{\dagger}), \omega_i\rangle$ in deriving convergence rate results. Under two commonly adopted conditions on the nonlinearity of F, we get convergence rates $O(\delta^{\frac{1}{2}})$ and $O(\delta)$ of the regularized solution in the ℓ_2-norm, respectively.

For the minimization problem (1.2), we propose an iterative soft thresholding algorithm [13,30] based on the generalized conditional gradient method (GCGM). In [31,32], GCGM is applied to solve the minimization problem for sparsity regularization with the convex regularization term $\sum_n w_n|\langle u, \phi_n\rangle|^p$ with $p \ge 1$, where $\{w_n > 0\}$ are the weights, and $\{\phi_n\}$ is an orthonormal basis of a Hilbert space. In this paper, it is shown that this method can be applied to the non-convex $\alpha\ell_1 - \beta\ell_2$ sparsity regularization for nonlinear inverse problems. For the case $q = 2$, we rewrite the functional $\mathcal{J}^{\delta}_{\alpha,\beta}$ in (1.2) as

$$\mathcal{J}^{\delta}_{\alpha,\beta}(x) = G(x) + \Phi(x),$$

where $G(x) = (1/2)\,\|F(x) - y^{\delta}\|^2_Y - \Theta(x)$, $\Phi(x) = \Theta(x) + \alpha\|x\|_{\ell_1} - \beta\|x\|_{\ell_2}$ and $\Theta(x) = (\lambda/2)\,\|x\|^2_{\ell_2} + \beta\|x\|_{\ell_2}$. Here $\lambda > 0$ is a parameter whose effect on the performance of the proposed algorithm is shown in Table 1 of Section 5. We show that if the nonlinear operator F is continuously Fréchet differentiable and F is bounded on bounded sets, then the iterative soft thresholding algorithm is convergent.

The rest of the paper is organized as follows. In Section 2, we analyze the well-posedness of the $\alpha\|\cdot\|_{\ell_1} - \beta\|\cdot\|_{\ell_2}$ $(\alpha \geq \beta \geq 0)$ regularization. In Section 3, we derive the convergence rates in the ℓ_2-norm under an appropriate source condition and two commonly adopted conditions on the nonlinearity of F. In Section 4, we present an iterative soft thresholding algorithm based on GCGM and discuss its convergence. Finally, some numerical experiments are presented in Section 5.

2. Well-posedness of regularization problem

In this section we analyze the well-posedness of the regularization method, i.e., existence, stability as well as convergence of regularized solutions. For the case $\alpha = \beta$, $\mathcal{R}_{\alpha,\beta}$ does not have coercivity nor Radon-Riesz property, and the well-posedness result of the regularization is weaker than that in the case $\alpha > \beta$.

Let us denote a general minimizer of the functional $\mathcal{J}^\delta_{\alpha,\beta}$ by $x^\delta_{\alpha,\beta}$, i.e.

$$x^\delta_{\alpha,\beta} \in \arg\min_x \mathcal{J}^\delta_{\alpha,\beta}(x), \quad \mathcal{J}^\delta_{\alpha,\beta}(x) = \frac{1}{q}\|F(x) - y^\delta\|^q_Y + \mathcal{R}_{\alpha,\beta}(x). \tag{2.1}$$

Definition 2.1. An element $x^\dagger \in \ell_2$ is called an $\mathcal{R}_\eta$-minimum solution to the problem (1.1) if

$$x^\dagger \in \arg\min\left\{\mathcal{R}_\eta(x) \mid x \in \ell_2,\ F(x) = y\right\}.$$

Definition 2.2. $x \in \ell_2$ is called sparse if $\operatorname{supp}(x) := \{i \in \mathbb{N} \mid x_i \neq 0\}$ is finite, where x_i is the ith component of x.

To characterize the sparsity, as in [13], we define the index set

$$I(x^\dagger) = \{i \in \mathbb{N} \mid x^\dagger_i \neq 0\}, \tag{2.2}$$

where $x^\dagger_i$ is the ith component of $x^\dagger$. Next, we present a result on the non-negativity of $\mathcal{R}_{\alpha,\beta}$.

Lemma 2.3. *If $\alpha \geq \beta \geq 0$, then $\mathcal{R}_{\alpha,\beta}(x) \geq 0$ for any $x \in \ell_2$.*

Proof. By the definition of $\mathcal{R}_{\alpha,\beta}$, we have $\mathcal{R}_{\alpha,\beta}(x) = (\alpha-\beta)\|x\|_{\ell_1} + \beta(\|x\|_{\ell_1} - \|x\|_{\ell_2})$. Since $\|x\|_{\ell_2} \leq \|x\|_{\ell_1}$ and $0 \leq \beta \leq \alpha$, this implies $\mathcal{R}_{\alpha,\beta}(x) \geq 0$. ■

2.1. The case $\alpha > \beta \geq 0$

We first recall some properties of $\mathcal{R}_{\alpha,\beta}$ $(\alpha > \beta \geq 0)$ which are crucial tools in analyzing the well-posedness of the regularization, cf. [1] for the proofs.

Lemma 2.4. *The functional $\mathcal{R}_{\alpha,\beta}$ $(\alpha > \beta \geq 0)$ has the following properties:*
(i) *(Coercivity) For $x \in \ell_2$, $\|x\|_{\ell_2} \to \infty$ implies $\mathcal{R}_{\alpha,\beta}(x) \to \infty$.*
(ii) (Weak lower semi-continuity) *If $x_n \rightharpoonup x$ in ℓ_2 and $\{\mathcal{R}_{\alpha,\beta}(x_n)\}$ is bounded, then*

$$\liminf_n \mathcal{R}_{\alpha,\beta}(x_n) \geq \mathcal{R}_{\alpha,\beta}(x).$$

(iii) (Radon-Riesz property) *If $x_n \rightharpoonup x$ in ℓ_2 and $\mathcal{R}_{\alpha,\beta}(x_n) \to \mathcal{R}_{\alpha,\beta}(x)$, then $\|x_n - x\|_{\ell_2} \to 0$.*

Lemma 2.5. *Assume the sequence $\{\|y_n\|_Y\}$ is bounded in Y. For a given $M > 0$, let $x_n \in \ell_2$, $n = 1, 2, \ldots$ and*

$$\frac{1}{q}\|F(x_n) - y_n\|^q_Y + \mathcal{R}_{\alpha,\beta}(x_n) \leq M. \tag{2.3}$$

Then there exist an element $x \in \ell_2$ and a subsequence $\{x_{n_k}\}$ of $\{x_n\}$ such that $x_{n_k} \rightharpoonup x$ and $F(x_{n_k}) \rightharpoonup F(x)$.

Proof. By (2.3), $\{\mathcal{R}_{\alpha,\beta}(x_n)\}$ is bounded. It follows from the coercivity of $\mathcal{R}_{\alpha,\beta}(x)$ that $\{\|x_n\|_{\ell_2}\}$ is bounded. Meanwhile, since $\{\|y_n\|_Y\}$ is bounded, $\{\|F(x_n)\|_Y\}$ is bounded. Hence, there exists a subsequence $\{x_{n_k}\}$ of $\{x_n\}$, $x \in \ell_2$ and $y \in Y$ such that

$$x_{n_k} \rightharpoonup x \text{ in } \ell_2, \quad F(x_{n_k}) \rightharpoonup y \text{ in } Y.$$

Since F is weakly sequentially closed, $F(x) = y$. This proves the lemma. ■

We have the existence, stability as well as convergence of the regularized solution given in the next three results, similar to Theorems 2.11, 2.12 and 2.13 in [1]. Their proofs are based on the properties stated in Lemmas 2.4 and 2.5.

Theorem 2.6 (*Existence*). *For any $y^\delta \in Y$, there exists at least one minimizer to $\mathcal{J}^\delta_{\alpha,\beta}$ in ℓ_2.*

Theorem 2.7 (*Stability*). *Let $\alpha_n > \beta_n \geq 0$, $\alpha_n \to \alpha$ $(\alpha > 0)$, $\beta_n \to \beta$ as $n \to \infty$. Let the sequence $\{y_n\} \subset Y$ be convergent to $y^\delta \in Y$, and let x_n be a minimizer to $\mathcal{J}^{\delta_n}_{\alpha_n,\beta_n}$. Then the sequence $\{x_n\}$ contains a subsequence $\{x_{n_k}\}$ converging to a minimizer of $\mathcal{J}^\delta_{\alpha,\beta}$. Furthermore, if $\mathcal{J}^\delta_{\alpha,\beta}$ has a unique minimizer $x^\delta_{\alpha,\beta}$, then $\lim_{k\to\infty}\|x_{n_k} - x^\delta_{\alpha,\beta}\|_{\ell_2} = 0$.*

Theorem 2.8 (*Convergence*). *Let $\alpha_n := \alpha(\delta_n)$, $\beta_n := \beta(\delta_n)$, $\alpha_n > \beta_n \geq 0$ satisfy*

$$\lim_{n\to\infty} \alpha_n = 0, \quad \lim_{n\to\infty} \beta_n = 0 \quad \text{and} \quad \lim_{n\to\infty} \frac{\delta_n^q}{\alpha_n} = 0.$$

Assume that $\eta = \lim_{n\to\infty} \eta_n \in [0, 1)$ exists, where $\eta_n = \beta_n/\alpha_n$. Let $\delta_n \to 0$ as $n \to +\infty$ and y^{δ_n} satisfy $\|y - y^{\delta_n}\| \leq \delta_n$. Moreover, let

$$x^{\delta_n}_{\alpha_n,\beta_n} \in \arg\min_x \mathcal{J}^{\delta_n}_{\alpha_n,\beta_n}(x).$$

Then $\{x^{\delta_n}_{\alpha_n,\beta_n}\}$ has a subsequence, still denoted by $\{x^{\delta_n}_{\alpha_n,\beta_n}\}$, converging to an $\mathcal{R}_\eta$-minimizing solution $x^\dagger$ in ℓ_2. Furthermore, if the $\mathcal{R}_\eta$-minimizing solution $x^\dagger$ is unique, then the entire sequence $\{x^{\delta_n}_{\alpha_n,\beta_n}\}$ converges to $x^\dagger$ in ℓ_2.

2.2. The case $\alpha = \beta > 0$

We turn to the case $\alpha = \beta > 0$. The functional $\mathcal{R}_{\alpha,\beta}$ remains to be weakly lower semi-continuous, see [1, Lemma 2.8, Remark 2.9] for details. However, coercivity and Radon-Riesz property cannot be extended to the case $\alpha = \beta$, cf. Examples 2.9 and 2.11 below.

Example 2.9 (*Non-coercivity*). Let $x = te_i$ for some i, where $e_i = (\underbrace{0, \ldots, 0, 1}_{i}, 0, \ldots)$. Then, $\|x\|_{\ell_2} \to \infty$ as $t \to \infty$. However, $\mathcal{R}_{\alpha,\alpha}(x) = 0$ for each t. So $\mathcal{R}_{\alpha,\alpha}$ is not coercive.

Note that the standard proof of the well-posedness of Tikhonov regularization is invalid without the coercivity of the regularization term. So to ensure the well-posedness of the problem (1.2) in the case $\alpha = \beta$, we provide a result next where an additional restriction, i.e. coercivity is imposed on the nonlinear operator F; see [26,33] for some examples of the nonlinear (or linear) coercive operator.

Lemma 2.10. *Assume $F(x)$ is coercive with respect to $\|x\|_{\ell_2}$, i.e. $\|x\|_{\ell_2} \to \infty$ implies $\|F(x)\|_Y \to \infty$. Then the functional $\mathcal{J}^\delta_{\alpha,\alpha}$ is coercive.*

Proof. By the definition of $\mathcal{J}^\delta_{\alpha,\beta}$,

$$\mathcal{J}^\delta_{\alpha,\alpha}(x) = \frac{1}{q}\|F(x) - y^\delta\|_Y^q + \alpha\|x\|_{\ell_1} - \alpha\|x\|_{\ell_2} \geq \frac{1}{q}\Big|\|F(x)\|_Y - \|y^\delta\|_Y\Big|^q.$$

Since F is coercive, it is obvious that $\mathcal{J}^\delta_{\alpha,\alpha}(x) \to \infty$ as $\|x\|_{\ell_2} \to \infty$. ■

Note that if $F = A$ is linear, then its coercivity is equivalent to the existence of a positive constant $c_0 > 0$ such that

$$\|Ax\|_Y \geq c_0\|x\|_{\ell_2} \quad \forall x \in \ell_2.$$

So when F is linear and coercive, the problem (1.1) is well-posed and there would be no need to solve it with regularization.

Based on Lemma 2.10, we can demonstrate the existence of the regularized solution; the proof is similar to that in Theorem 2.11 in [1].

Next we give an example to show that x_n does not necessarily converge strongly to x even if $x_n \rightharpoonup x$ in ℓ_2 and $\mathcal{R}_{\alpha,\alpha}(x_n) \to \mathcal{R}_{\alpha,\alpha}(x)$. Thus $\mathcal{R}_{\alpha,\alpha}$ fails to satisfy the Radon-Riesz property.

Example 2.11 (*Non-Radon-Riesz Property*). Let $x_n = (\underbrace{0, \ldots, 0, 1}_{n}, 0, \ldots)$ and $x = 0$. Then $x_n \rightharpoonup x$ in ℓ_2. We have

$$\mathcal{R}_{\alpha,\alpha}(x_n) = \alpha(\|x_n\|_{\ell_1} - \|x_n\|_{\ell_2}) = 0 \quad \text{and} \quad \mathcal{R}_{\alpha,\alpha}(x) = 0.$$

So $\mathcal{R}_{\alpha,\alpha}(x_n) \to \mathcal{R}_{\alpha,\alpha}(x)$. However, $\|x_n - x\|_{\ell_2} = 1$, which implies that x_n does not converge strongly to x.

Since $\mathcal{R}_{\alpha,\alpha}$ fails to satisfy the Radon-Riesz property, we do not have stability and convergence properties similar to the ones stated in Theorems 2.7 and 2.8. Nevertheless, under the additional assumption of the coercivity of F, with the help of Lemma 2.10, stability and convergence properties can be proved, similar to that of Theorems 2.12 and 2.13 in [1].

2.3. Sparsity

Next we turn to a discussion of the sparsity of the regularized solution. When $q = 2$, under a restriction on the nonlinearity of F, it can be shown that every minimizer of $\mathcal{J}^\delta_{\alpha,\beta}$ is sparse whenever $\alpha > \beta$ or $\alpha = \beta$.

Assumption 2.12. Assume that $F : \ell_2 \to Y$ is Fréchet-differentiable and there exists $\gamma > 0$ such that

$$\|F'(y) - F'(x)\|_{L(\ell_2,Y)} \leq \gamma\|y - x\|_{\ell_2} \tag{2.4}$$

for any $y \in B_\delta(x)$, where x is a minimizer of $\mathcal{J}^\delta_{\alpha,\beta}$ and $B_\delta(x) := \{y \mid \|y - x\|_{\ell_2} \leq \delta\}$, $\delta \geq \|x\|_\infty$.

Remark 2.13. By [11, p. 14], (2.4) implies

$$\|F(y)-F(x)-F'(x)(y-x)\|_Y \le \frac{\gamma}{2}\|y-x\|_{\ell_2}^2 \tag{2.5}$$

for any $y \in B_\delta(x)$, $\delta \ge \|x\|_\infty$.

Proposition 2.14 (*Sparsity*). *Let x be a minimizer of $\mathcal{J}_{\alpha,\beta}^{\delta}$ for $\alpha \ge \beta \ge 0$ and $q = 2$. If Assumption 2.12 holds, then x is sparse.*

Proof. For $i \in \mathbb{N}$, consider $\bar{x} := x - x_i e_i$, where x_i is the ith component of x. It is clear that $\bar{x} \in B_\delta(x)$. By the definition of x,

$$\frac{1}{2}\|F(x)-y^\delta\|_Y^2 + \mathcal{R}_{\alpha,\beta}(x) \le \frac{1}{2}\|F(\bar{x})-y^\delta\|_Y^2 + \mathcal{R}_{\alpha,\beta}(\bar{x}). \tag{2.6}$$

If $x = 0$, then x is sparse. Suppose $x \ne 0$. By (2.6), we see that

$$\begin{aligned}\alpha|x_i| - \beta\frac{|x_i|^2}{\|x\|_{\ell_2}+\|\bar{x}\|_{\ell_2}} &= \mathcal{R}_{\alpha,\beta}(x) - \mathcal{R}_{\alpha,\beta}(\bar{x}) \\ &\le \frac{1}{2}\|F(\bar{x})-y^\delta\|_Y^2 - \frac{1}{2}\|F(x)-y^\delta\|_Y^2 \\ &= \frac{1}{2}\|F(\bar{x})-F(x)\|_Y^2 + \langle F(x)-y^\delta, F(\bar{x})-F(x)\rangle. \end{aligned} \tag{2.7}$$

From Remark 2.13, we see that

$$F(\bar{x}) = F(x) + F'(x)(\bar{x}-x) + r_\alpha^\delta \tag{2.8}$$

with

$$\|r_\alpha^\delta\|_Y \le \frac{\gamma}{2}\|\bar{x}-x\|_{\ell_2}^2. \tag{2.9}$$

A combination of (2.8) and (2.9) implies that

$$\begin{aligned}\|F(\bar{x})-F(x)\|_Y^2 &= \|F'(x)(\bar{x}-x)\|_Y^2 + \|r_\alpha^\delta\|_Y^2 + 2\langle F'(x)(\bar{x}-x), r_\alpha^\delta\rangle \\ &\le \|F'(x)\|_{L(\ell_2,Y)}^2\|\bar{x}-x\|_{\ell_2}^2 + \frac{\gamma^2}{4}\|\bar{x}-x\|_{\ell_2}^4 + \gamma\|F'(x)\|_{L(\ell_2,Y)}\|\bar{x}-x\|_{\ell_2}^3 \\ &= |x_i|^2\|F'(x)\|_{L(\ell_2,Y)}^2 + \frac{\gamma^2}{4}|x_i|^4 + \gamma|x_i|^3\|F'(x)\|_{L(\ell_2,Y)}. \end{aligned} \tag{2.10}$$

Moreover,

$$\begin{aligned}\langle F(x)-y^\delta, F(\bar{x})-F(x)\rangle &= \langle F(x)-y^\delta, F'(x)(\bar{x}-x)+r_\alpha^\delta\rangle \\ &\le -x_i\langle F'(x)^*(F(x)-y^\delta), e_i\rangle + \frac{\gamma}{2}|x_i|^2\|F(x)-y^\delta\|_Y. \end{aligned} \tag{2.11}$$

A combination of (2.7), (2.10) and (2.11) implies that

$$\begin{aligned}\alpha|x_i| - \beta\frac{|x_i|^2}{\|x\|_{\ell_2}+\|\bar{x}\|_{\ell_2}} \le{}& \frac{1}{2}|x_i|^2\|F'(x)\|_{L(\ell_2,Y)}^2 + \frac{\gamma^2}{8}|x_i|^4 + \frac{1}{2}\gamma|x_i|^3\|F'(x)\|_{L(\ell_2,Y)} \\ &- x_i\langle F'(x)^*(F(x)-y^\delta), e_i\rangle + \frac{\gamma}{2}|x_i|^2\|F(x)-y^\delta\|_Y \end{aligned} \tag{2.12}$$

for every $i \in \mathbb{N}$. Define $\|x\|_0 := \sum_{i\in\mathbb{N}} \mathrm{sgn}(|x_i|)$, where sgn is the sign function. Now if $\|x\|_0 = 1$, then x is sparse. Otherwise, $\|x\|_0 \ge 2$ and then $\frac{|x_i|}{\|x\|_{\ell_2}+\|\bar{x}\|_{\ell_2}} < 1$. Thus, there exists a constant $c > 0$ such that

$$\frac{c+\eta|x_i|}{\|x\|_{\ell_2}+\|\bar{x}\|_{\ell_2}} \le 1, \quad \text{i.e.} \quad \frac{c}{\|x\|_{\ell_2}+\|\bar{x}\|_{\ell_2}} \le 1 - \frac{\eta|x_i|}{\|x\|_{\ell_2}+\|\bar{x}\|_{\ell_2}}. \tag{2.13}$$

Multiplying (2.13) by $\alpha|x_i|$, we have

$$\alpha c\frac{|x_i|}{\|x\|_{\ell_2}+\|\bar{x}\|_{\ell_2}} \le \alpha|x_i| - \beta\frac{|x_i|^2}{\|x\|_{\ell_2}+\|\bar{x}\|_{\ell_2}}. \tag{2.14}$$

Denote

$$\begin{aligned}K_i :={}& \frac{(\|x\|_{\ell_2}+\|\bar{x}\|_{\ell_2})\left(\frac{1}{2}x_i\|F'(x)\|_{L(\ell_2,Y)}^2 + \frac{\gamma^2}{8}x_i^3 + \frac{1}{2}\gamma x_i|x_i|\|F'(x)\|_{L(\ell_2,Y)}\right)}{c\alpha} \\ &+ \frac{(\|x\|_{\ell_2}+\|\bar{x}\|_{\ell_2})\left(-\langle F'(x)^*(F(x)-y^\delta), e_i\rangle + \frac{\gamma}{2}x_i\|F(x)-y^\delta\|_Y\right)}{c\alpha}. \end{aligned}$$

Then a combination of (2.12) and (2.14) implies that

$$K_i x_i \ge |x_i|, \quad i \in \mathbb{N}.$$

Since x is a minimizer of $\mathcal{J}^{\delta}_{\alpha,\beta}$, $\|F(x)-y^{\delta}\|_Y$ is finite. In addition, since F is Fréchet-differentiable, $\|F'(x)\|_{L(\ell_2,Y)}$ is also finite. Since $x, \bar{x} \in \ell_2$, $F'(x)^*(F(x)-y^{\delta}) \in \ell_2$, we have $K_i \to 0$ as $i \to \infty$, and this implies that $\Lambda := \{i \in \mathbb{N} \mid |K_i| \geq 1\}$ is finite. It is obvious that $x_i = 0$ whenever $i \notin \Lambda$. This proves the proposition. ■

Note that the above result holds only for the case $q = 2$. It is not clear whether each minimizer of $\mathcal{J}^{\delta}_{\alpha,\beta}$ is sparse when $q \neq 2$.

3. Convergence rate of the regularized solutions

We consider the convergence rate for the case $\alpha > \beta \geq 0$ in this section. For this purpose, we need to impose a restriction on the smoothness of $x^{\dagger}$. Meanwhile, we impose two commonly adopted conditions on the nonlinearity of F, and derive two corresponding inequalities. Then we get convergence rates $O(\delta^{\frac{1}{2}})$ and $O(\delta)$ in the ℓ_2-norm based on the two inequalities, respectively.

3.1. Convergence rate $O(\delta^{\frac{1}{2}})$

Assumption 3.1. Let $x^{\dagger} \neq 0$ be an $\mathcal{R}_{\eta}$-minimizing solution of the problem (1.1) that is sparse. Assume that

(i) F is continuously Fréchet differentiable. For every $i \in I(x^{\dagger})$, there exists $\omega_i \in Y$ such that

$$e_i = F'(x^{\dagger})^* \omega_i, \tag{3.1}$$

where $I(x^{\dagger})$ is defined in (2.2).

(ii) There exist $\gamma > 0$, $\rho > 0$, $\mathcal{R}_{\eta}(x^{\dagger}) < \rho$ such that

$$\|F'(x) - F'(x^{\dagger})\|_{L(\ell_2,Y)} \leq \gamma \|x - x^{\dagger}\|_{\ell_2} \tag{3.2}$$

for all $x \in \ell_2$ satisfying $\mathcal{R}_{\eta}(x) \leq \rho$.

Assumption 3.1(i) and other analogous conditions were introduced in [20,34]. Actually, Assumption 3.1(i) is a source condition which imposes the smoothness on the solution $x^{\dagger}$. Assumption 3.1 (ii) is a restriction on F which has two-fold meaning. One is to impose nonlinearity condition on F. Another more crucial effect is to estimate the term $\langle F'(x^{\dagger})(x - x^{\dagger}), \omega_i \rangle$, where ω_i are the same as that in (3.1). Many authors pointed out that the restrictions on the nonlinearity of F coupled with source conditions prove to be a powerful tool to obtain convergence rates in regularization [35–37]. There are several ways to choose the restrictions on the nonlinearity of F. A commonly adopted restriction is (3.2), i.e. F' is Lipschitz continuous [11,38].

Remark 3.2. Note that for all $x \in \ell_2$ satisfying $\mathcal{R}_{\eta}(x) \leq \rho$, (3.2) implies

$$\|F(x) - F(x^{\dagger}) - F'(x^{\dagger})(x - x^{\dagger})\|_Y \leq \frac{\gamma}{2}\|x - x^{\dagger}\|^2_{\ell_2} \tag{3.3}$$

from a Taylor approximation of F. Thus, with the triangle inequality, we obtain

$$\|F'(x^{\dagger})(x - x^{\dagger})\|_Y \leq \frac{\gamma}{2}\|x - x^{\dagger}\|^2_{\ell_2} + \|F(x) - F(x^{\dagger})\|_Y \tag{3.4}$$

for all $x \in \ell_2$ satisfying $\mathcal{R}_{\eta}(x) \leq \rho$, which can be used to give an upper bound of the term $\langle F'(x^{\dagger})(x - x^{\dagger}), \omega_i \rangle$.

Lemma 3.3. *If Assumption 3.1 holds, then there exists $\omega^{\dagger} \in Y$ such that $x^{\dagger} = F'(x^{\dagger})^* \omega^{\dagger}$.*

This result is verified easily by setting $\omega^{\dagger} = \sum_{i \in I(x^{\dagger})} x^{\dagger}_i \omega_i$.

Next we derive an inequality needed in the proof of the convergence rate. By Lemma 2.4(i), for any $M > 0$, there exists $M_1 > 0$ such that $\mathcal{R}_{\alpha,\beta}(x) \leq M$ for $x \in \ell_2$ implies $\|x\|_{\ell_2} \leq M_1$. We further denote

$$c_1 = M_1 + \|x^{\dagger}\|_{\ell_2}, \tag{3.5}$$

$$c_2 = \left(1 + \frac{c_1}{\|x^{\dagger}\|_{\ell_2}}\right) |I(x^{\dagger})| \max_{i \in I(x^{\dagger})} \|\omega_i\|_Y + \frac{2\|\omega^{\dagger}\|_Y}{\|x^{\dagger}\|_{\ell_2}}, \tag{3.6}$$

$$c_3 = \frac{2\|\omega^{\dagger}\|_Y}{\|x^{\dagger}\|_{\ell_2}}. \tag{3.7}$$

Lemma 3.4. *Let $M > 0$ be given and define c_1, c_2 and c_3 by (3.5)–(3.7). Under Assumption 3.1, if $\Gamma := \frac{\gamma c_1 (c_2 \alpha - c_3 \beta)}{2(\alpha - \beta)} < 1$, then*

$$\|x - x^{\dagger}\|^2_{\ell_2} \leq \frac{1}{(1-\Gamma)} \left[\frac{c_1}{\alpha - \beta} (\mathcal{R}_{\alpha,\beta}(x) - \mathcal{R}_{\alpha,\beta}(x^{\dagger})) + \frac{c_1}{\alpha - \beta}(c_2 \alpha - c_3 \beta) \|F(x) - F(x^{\dagger})\|_Y \right]$$

for any $x \in \ell_2$ with $\mathcal{R}_{\alpha,\beta}(x) \leq M$ satisfying $\mathcal{R}_{\eta}(x) \leq \rho$.

Proof. By the definition of $\mathcal{R}_{\alpha,\beta}$ in (1.3), it is clear that

$$\begin{aligned}\mathcal{R}_{\alpha,\beta}(x) &= \alpha\|x\|_{\ell_1} - \beta\|x\|_{\ell_2} = \alpha(\|x\|_{\ell_1} - \|x\|_{\ell_2}) + (\alpha-\beta)\|x\|_{\ell_2} \\ &= \alpha\mathcal{K}(x) + (\alpha-\beta)\|x\|_{\ell_2},\end{aligned}$$

where $\mathcal{K}(x) := \|x\|_{\ell_1} - \|x\|_{\ell_2}$. We see that

$$\begin{aligned}\mathcal{R}_{\alpha,\beta}(x) - \mathcal{R}_{\alpha,\beta}(x^\dagger) &= \alpha[\mathcal{K}(x) - \mathcal{K}(x^\dagger)] + (\alpha-\beta)(\|x\|_{\ell_2} - \|x^\dagger\|_{\ell_2}) \\ &= \alpha[\mathcal{K}(x) - \mathcal{K}(x^\dagger)] + (\alpha-\beta)\frac{\|x\|_{\ell_2}^2 - \|x^\dagger\|_{\ell_2}^2}{\|x\|_{\ell_2} + \|x^\dagger\|_{\ell_2}} \\ &= \alpha[\mathcal{K}(x) - \mathcal{K}(x^\dagger)] + (\alpha-\beta)\frac{\|x-x^\dagger\|_{\ell_2}^2 + 2\langle x^\dagger, x-x^\dagger\rangle}{\|x\|_{\ell_2} + \|x^\dagger\|_{\ell_2}}.\end{aligned} \tag{3.8}$$

From the definition of $\mathcal{K}(x)$, we have

$$\mathcal{K}(x) - \mathcal{K}(x^\dagger) = \|x\|_{\ell_1} - \|x\|_{\ell_2} - \|x^\dagger\|_{\ell_1} + \|x^\dagger\|_{\ell_2}.$$

With the definition of index set $I(x^\dagger)$ in (2.2), we obtain that

$$\begin{aligned}&\|x\|_{\ell_1} = \sum_{i\in I(x^\dagger)} |x_i| + \sum_{i\notin I(x^\dagger)} |x_i|, \\ &-\|x\|_{\ell_2} \geq -\left(\sum_{i\in I(x^\dagger)} |x_i|^2\right)^{\frac{1}{2}} - \left(\sum_{i\notin I(x^\dagger)} |x_i|^2\right)^{\frac{1}{2}} \geq -\left(\sum_{i\in I(x^\dagger)} |x_i|^2\right)^{\frac{1}{2}} - \sum_{i\notin I(x^\dagger)} |x_i|.\end{aligned}$$

Then,

$$\mathcal{K}(x) - \mathcal{K}(x^\dagger) \geq \sum_{i\in I(x^\dagger)} \left(|x_i| - |x_i^\dagger|\right) - \left[\left(\sum_{i\in I(x^\dagger)} |x_i|^2\right)^{\frac{1}{2}} - \left(\sum_{i\in I(x^\dagger)} |x_i^\dagger|^2\right)^{\frac{1}{2}}\right],$$

which is rewritten as

$$\mathcal{K}(x) - \mathcal{K}(x^\dagger) \geq \sum_{i\in I(x^\dagger)} (|x_i| - |x_i^\dagger|) - \frac{\sum_{i\in I(x^\dagger)}(|x_i|+|x_i^\dagger|)(|x_i|-|x_i^\dagger|)}{\left(\sum_{i\in I(x^\dagger)}|x_i|^2\right)^{\frac{1}{2}} + \left(\sum_{i\in I(x^\dagger)}|x_i^\dagger|^2\right)^{\frac{1}{2}}}. \tag{3.9}$$

By the definition of M_1, we have

$$\left(\sum_{i\in I(x^\dagger)} |x_i|^2\right)^{\frac{1}{2}} + \left(\sum_{i\in I(x^\dagger)} |x_i^\dagger|^2\right)^{\frac{1}{2}} \leq M_1 + \|x^\dagger\|_{\ell_2}.$$

Thus,

$$|x_i| + |x_i^\dagger| \leq M_1 + \|x^\dagger\|_{\ell_2}, \quad \forall i \in I(x^\dagger). \tag{3.10}$$

Meanwhile, we have

$$0 < \|x^\dagger\|_{\ell_2} \leq \left(\sum_{i\in I(x^\dagger)} |x_i|^2\right)^{\frac{1}{2}} + \left(\sum_{i\in I(x^\dagger)} |x_i^\dagger|^2\right)^{\frac{1}{2}}. \tag{3.11}$$

A combination of (3.9), (3.10) and (3.11) implies that

$$\mathcal{K}(x) - \mathcal{K}(x^\dagger) \geq -\sum_{i\in I(x^\dagger)} |x_i - x_i^\dagger| - \frac{M_1 + \|x^\dagger\|_{\ell_2}}{\|x^\dagger\|_{\ell_2}} \sum_{i\in I(x^\dagger)} |x_i - x_i^\dagger|,$$

i.e.

$$\mathcal{K}(x) - \mathcal{K}(x^\dagger) \geq -\left(2 + \frac{M_1}{\|x^\dagger\|_{\ell_2}}\right) \sum_{i\in I(x^\dagger)} |x_i - x_i^\dagger|. \tag{3.12}$$

A combination of (3.8) and (3.12) implies that

$$\begin{aligned}(\alpha-\beta)\frac{\|x-x^\dagger\|_{\ell_2}^2}{\|x\|_{\ell_2} + \|x^\dagger\|_{\ell_2}} \leq\ & \mathcal{R}_{\alpha,\beta}(x) - \mathcal{R}_{\alpha,\beta}(x^\dagger) + \alpha\left(2 + \frac{M_1}{\|x^\dagger\|_{\ell_2}}\right)\sum_{i\in I(x^\dagger)} |x_i - x_i^\dagger| \\ & - 2(\alpha-\beta)\frac{\langle x^\dagger, x-x^\dagger\rangle}{\|x\|_{\ell_2} + \|x^\dagger\|_{\ell_2}}.\end{aligned} \tag{3.13}$$

Since $\|x\|_{\ell_2} \le M_1$, by (3.13), we see that

$$\frac{\alpha-\beta}{M_1+\|x^\dagger\|_{\ell_2}}\|x-x^\dagger\|_{\ell_2}^2 \le \mathcal{R}_{\alpha,\beta}(x)-\mathcal{R}_{\alpha,\beta}(x^\dagger)+\alpha\left(2+\frac{M_1}{\|x^\dagger\|_{\ell_2}}\right)\sum_{i\in I(x^\dagger)}|x_i-x_i^\dagger| - 2(\alpha-\beta)\frac{\langle x^\dagger, x-x^\dagger\rangle}{\|x\|_{\ell_2}+\|x^\dagger\|_{\ell_2}}. \tag{3.14}$$

By Assumption 3.1 (i), we have

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

Hence,

$$\sum_{i\in I(x^\dagger)}|x_i-x_i^\dagger| \le |I(x^\dagger)|\max_{i\in I(x^\dagger)}\|\omega_i\|_Y\|F'(x^\dagger)(x-x^\dagger)\|_Y, \tag{3.15}$$

where $|I(x^\dagger)|$ denotes the cardinality of the index set $I(x^\dagger)$. On the other hand, by Lemma 3.3, we see that

$$\frac{|\langle x^\dagger, x-x^\dagger\rangle|}{\|x\|_{\ell_2}+\|x^\dagger\|_{\ell_2}} = \frac{|\langle \omega^\dagger, F'(x^\dagger)(x-x^\dagger)\rangle|}{\|x\|_{\ell_2}+\|x^\dagger\|_{\ell_2}} \le \frac{\|\omega^\dagger\|_Y\|F'(x^\dagger)(x-x^\dagger)\|_Y}{\|x^\dagger\|_{\ell_2}}. \tag{3.16}$$

A combination of (3.14), (3.15) (3.4) and (3.16) implies that

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

i.e.

$$\|x-x^\dagger\|_{\ell_2}^2 \le \frac{1}{\left(1-\frac{\gamma c_1(c_2\alpha-c_3\beta)}{2(\alpha-\beta)}\right)}\left[\frac{c_1}{\alpha-\beta}(\mathcal{R}_{\alpha,\beta}(x)-\mathcal{R}_{\alpha,\beta}(x^\dagger))+\frac{c_1}{\alpha-\beta}(c_2\alpha-c_3\beta)\|F(x)-F(x^\dagger)\|_Y\right],$$

where c_1, c_2 and c_3 are defined by (3.5)–(3.7). ■

Theorem 3.5. *Suppose Assumption 3.1 holds. Let $x_{\alpha,\beta}^\delta$ be defined by (2.1), and let the constants $c_1>0$, $c_2>c_3>0$ be as in Lemma 3.4. Assume $\Gamma := \frac{\gamma c_1(c_2\alpha-c_3\beta)}{2(\alpha-\beta)} < 1$.*

1. If $q=1$ and $c_2\alpha-c_3\beta<1$, then

$$\|x_{\alpha,\beta}^\delta-x^\dagger\|_{\ell_2}^2 \le \frac{c_1[1+(c_2\alpha-c_3\beta)]\delta}{(\alpha-\beta)(1-\Gamma)}, \qquad \|F(x_{\alpha,\beta}^\delta)-y^\delta\|_Y \le \frac{[1+(c_2\alpha-c_3\beta)]\delta}{1-(c_2\alpha-c_3\beta)}. \tag{3.17a}$$

2. If $q>1$, then

$$\begin{aligned}&\|x_{\alpha,\beta}^\delta-x^\dagger\|_{\ell_2}^2 \le \frac{c_1}{(\alpha-\beta)(1-\Gamma)}\left[\frac{\delta^q}{q}+(c_2\alpha-c_3\beta)\delta+\frac{(q-1)2^{\frac{1}{q-1}}(c_2\alpha-c_3\beta)^{\frac{q}{q-1}}}{q}\right],\\ &\|F(x_{\alpha,\beta}^\delta)-y^\delta\|_Y^q \le q\left[\frac{\delta^q}{q}+(c_2\alpha-c_3\beta)\delta+\frac{(q-1)2^{\frac{1}{q-1}}(c_2\alpha-c_3\beta)^{\frac{q}{q-1}}}{q}\right].\end{aligned} \tag{3.17b}$$

Proof. By the definition of $x_{\alpha,\beta}^\delta$, it is clear that

$$\tfrac{1}{q}\|F(x_{\alpha,\beta}^\delta)-y^\delta\|_Y^q+\mathcal{R}(x_{\alpha,\beta}^\delta) \le \tfrac{1}{q}\|F(x^\dagger)-y^\delta\|_Y^q+\mathcal{R}(x^\dagger),$$

i.e.

$$\frac{\delta^q}{q} \ge \mathcal{R}(x_{\alpha,\beta}^\delta)-\mathcal{R}(x^\dagger)+\frac{1}{q}\|F(x_{\alpha,\beta}^\delta)-y^\delta\|_Y^q.$$

Then $\mathcal{R}(x_{\alpha,\beta}^\delta)$ is bounded. Applying Lemma 3.4, we see that

$$\begin{aligned}\frac{\delta^q}{q} &\ge \frac{(\alpha-\beta)(1-\Gamma)}{c_1}\|x_{\alpha,\beta}^\delta-x^\dagger\|_{\ell_2}^2-(c_2\alpha-c_3\beta)\|F(x_{\alpha,\beta}^\delta)-F(x^\dagger)\|_Y+\frac{1}{q}\|F(x_{\alpha,\beta}^\delta)-y^\delta\|_Y^q\\ &\ge \frac{(\alpha-\beta)(1-\Gamma)}{c_1}\|x_{\alpha,\beta}^\delta-x^\dagger\|_{\ell_2}^2-(c_2\alpha-c_3\beta)\|F(x_{\alpha,\beta}^\delta)-y^\delta\|_Y\\ &\quad-(c_2\alpha-c_3\beta)\delta+\frac{1}{q}\|F(x_{\alpha,\beta}^\delta)-y^\delta\|_Y^q.\end{aligned} \tag{3.18}$$

So if $q=1$ and $c_2\alpha-c_3\beta<1$, then (3.17a) holds.

If $q > 1$, we apply Young's inequality $ab \le \frac{a^q}{q} + \frac{b^{q^*}}{q^*}$ for $a, b \ge 0$ and $q^* > 1$ defined by $\frac{1}{q} + \frac{1}{q^*} = 1$. We have

$$(c_2\alpha - c_3\beta)\|F(x^\delta_{\alpha,\beta}) - y^\delta\|_Y = 2^{\frac{1}{q}}(c_2\alpha - c_3\beta)2^{-\frac{1}{q}}\|F(x^\delta_{\alpha,\beta}) - y^\delta\|_Y \le \frac{1}{2q}\|F(x^\delta_{\alpha,\beta}) - y^\delta\|^q_Y + \frac{(q-1)2^{\frac{1}{q-1}}\left(c_2\alpha - c_3\beta\right)^{\frac{q}{q-1}}}{q}. \tag{3.19}$$

A combination with (3.18) and (3.19) implies (3.17b). ■

Note that by the definition of $x^\delta_{\alpha,\beta}$,

$$\frac{1}{q}\|F(x^\delta_{\alpha,\beta}) - y^\delta\|^q_Y + \mathcal{R}(x^\delta_{\alpha,\beta}) \le \frac{1}{q}\|F(0) - y^\delta\|^q_Y + \mathcal{R}(0) = \frac{1}{q}\|F(0) - y^\delta\|^q_Y.$$

Hence, the upper bound of $\mathcal{R}(x^\delta_{\alpha,\beta})$ depends on the noise level δ.

Remark 3.6 (*A-priori Estimation*)**.** Let $\beta = \eta\alpha$ be a fixed constant. For the case $q > 1$, if $\alpha \sim \delta^{q-1}$, then $\|x^\delta_{\alpha,\beta} - x^\dagger\|_{\ell_2} \le c\delta^{\frac{1}{2}}$ for some constant $c > 0$. For the particular case $q = 1$, if $\alpha \sim \delta^{1-\epsilon}$ $(0 < \epsilon < 1)$, then $\|x^\delta_{\alpha,\beta} - x^\dagger\|_{\ell_2} \le c\delta^{\frac{\epsilon}{2}}$ for some constant $c > 0$.

Note that due to the presence of the term $\frac{\gamma}{2}\|x - x^\dagger\|^2_{\ell_2}$ in the estimation (3.4), we need an additional condition to obtain the convergence rate, i.e. $\gamma > 0$ must be small enough such that $\Gamma < 1$. This additional condition is similar to the condition $\gamma\,\|\omega\| < 1$ in the classical quadratic regularization [38].

Theorem 3.7 (*Discrepancy Principle*)**.** *Keep the assumptions of Lemma 3.4 and let $x^\delta_{\alpha,\beta}$ be defined by (2.1). Assume there exist parameters α and β $(\beta = \eta\alpha)$ which are determined by the discrepancy principle*

$$\delta \le \|F(x^\delta_{\alpha,\beta}) - y^\delta\|_Y \le \tau\delta \quad (\tau \ge 1).$$

Then

$$\|x^\delta_{\alpha,\beta} - x^\dagger\|_{\ell_2} \le \sqrt{\frac{c_1(c_2 - c_3\eta)(\tau+1)\delta}{(1-\eta)(1-\Gamma)}}.$$

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

$$\frac{1}{q}\delta^q + \mathcal{R}_{\alpha,\beta}(x^\delta_{\alpha,\beta}) \le \frac{1}{q}\|F(x^\delta_{\alpha,\beta}) - y^\delta\|^q_Y + \mathcal{R}_{\alpha,\beta}(x^\delta_{\alpha,\beta}) \le \frac{1}{q}\|F(x^\dagger) - y^\delta\|^q_Y + \mathcal{R}_{\alpha,\beta}(x^\dagger). \tag{3.20}$$

Hence $\mathcal{R}_{\alpha,\beta}(x^\delta_{\alpha,\beta}) \le \mathcal{R}_{\alpha,\beta}(x^\dagger)$. It follows from Lemma 3.4 that

$$0 \ge \mathcal{R}_{\alpha,\beta}(x^\delta_{\alpha,\beta}) - \mathcal{R}_{\alpha,\beta}(x^\dagger) \ge \frac{(\alpha-\beta)(1-\Gamma)}{c_1}\|x^\delta_{\alpha,\beta} - x^\dagger\|^2_{\ell_2} - (c_2\alpha - c_3\beta)\|F(x^\delta_{\alpha,\beta}) - F(x^\dagger)\|_Y \ge \frac{(\alpha-\beta)(1-\Gamma)}{c_1}\|x^\delta_{\alpha,\beta} - x^\dagger\|^2_{\ell_2} - (c_2\alpha - c_3\beta)(\tau+1)\delta. \tag{3.21}$$

Then

$$\|x^\delta_{\alpha,\beta} - x^\dagger\|^2_{\ell_2} \le \frac{c_1(c_2\alpha - c_3\beta)(\tau+1)\delta}{(\alpha-\beta)(1-\Gamma)}.$$

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

Note that a drawback of the discrepancy principle is that a regularization parameter with

$$\delta \le \|F(x^\delta_{\alpha,\beta}) - y^\delta\|_Y \le \tau\delta \quad (\tau \ge 1)$$

might not exist for general nonlinear operators F. Actually, for nonlinear operators, it is hard to ensure the existence of the regularization parameter α which determined by Morozov's discrepancy principle. We need to impose some conditions on F. In this paper, we are mainly interested in the convergence rate under Morozov's discrepancy principle. Existence of α will be done in forthcoming papers.

3.2. Convergence rate $O(\delta)$

In [39, p. 6], it is pointed out that for ill-posed problems, (3.3) does not carry enough information about the local behavior of F around $x^\dagger$ to draw conclusions about convergence, since the left hand side of (3.3) can be much smaller than the right hand side for certain pairs of points x and $x^\dagger$. Therefore, several researchers adopted an alternative

$$\|F(x) - F(x^\dagger) - F'(x^\dagger)(x - x^\dagger)\|_Y \le \gamma\|F(x) - F(x^\dagger)\|_Y, \quad 0 < \gamma < \frac{1}{2} \tag{3.22}$$

as the condition on the nonlinearity of F; see [38, pp. 278–279], [39, p. 6], [37, pp. 69–70].

Assumption 3.8. Let $x^\dagger \neq 0$ be an $\mathcal{R}_\eta$-minimizing solution of the problem $F(x) = y$ that is sparse. We further assume that

(i) F is Fréchet differentiable at $x^\dagger$. For each $i \in I(x^\dagger)$, there exists an $\omega_i \in Y$ such that

$$e_i = F'(x^\dagger)^*\omega_i \tag{3.23}$$

holds, where $I(x^\dagger)$ is defined in (2.2).

(ii) There exist constants $0 < \gamma < \frac{1}{2}$, $\rho > 0$, $\mathcal{R}_\eta(x^\dagger) < \rho$ such that

$$\|F(x) - F(x^\dagger) - F'(x^\dagger)(x - x^\dagger)\|_Y \leq \gamma\|F(x) - F(x^\dagger)\|_Y \tag{3.24}$$

for any $x \in \ell_2$ with $\mathcal{R}_\eta(x) \leq \rho$.

Remark 3.9. Thanks to the triangle inequality, it follows from (3.24) that

$$\frac{1}{1+\gamma}\|F'(x^\dagger)(x - x^\dagger)\|_Y \leq \|F(x) - F(x^\dagger)\|_Y \leq \frac{1}{1-\gamma}\|F'(x^\dagger)(x - x^\dagger)\|_Y \tag{3.25}$$

which provides two-sided bounds on $\|F'(x^\dagger)(x - x^\dagger)\|_Y$. The condition

$$\|F'(x^\dagger)(x - x^\dagger)\|_Y \leq (1+\gamma)\|F(x) - F(x^\dagger)\|_Y$$

has been adopted by several researchers. This assumption immediately leads to a bound of the critical inner product $\langle F'(x^\dagger)(x - x^\dagger), \omega_i\rangle$.

Next, we derive an inequality from the condition (3.24). The linear convergence rate $O(\delta)$ follows from the inequality directly.

Lemma 3.10. *Let Assumption 3.8 hold and $\mathcal{R}_{\alpha,\beta}(x) \leq M$ for a given $M > 0$. Then there exist constants $c_4 > c_5$ such that*

$$(\alpha - \beta)\|x - x^\dagger\|_{\ell_1} \leq \mathcal{R}_{\alpha,\beta}(x) - \mathcal{R}_{\alpha,\beta}(x^\dagger) + (c_4\alpha - c_5\beta)\|F(x) - F(x^\dagger)\|_Y \tag{3.26}$$

for any $x \in \ell_2$ satisfying $\mathcal{R}_\eta(x) \leq \rho$.

Proof. By Lemma 2.4(i), for $M > 0$, there exists $M_1 > 0$ such that $\mathcal{R}_{\alpha,\beta}(x) \leq M$ for $x \in \ell_2$ implies $\|x\|_{\ell_2} \leq M_1$. Then, in analogy with the proof in [1, Theorem 2.17], we have

$$(\alpha - \beta)\|x - x^\dagger\|_{\ell_1} \leq \mathcal{R}_{\alpha,\beta}(x) - \mathcal{R}_{\alpha,\beta}(x^\dagger) + \left(2\alpha + \frac{M_1}{\|x^\dagger\|_{\ell_2}}\beta\right)\sum_{i\in I(x^\dagger)}|x_i - x_i^\dagger|. \tag{3.27}$$

In addition, by Assumption 3.8,

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

Hence,

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

Then, by (3.24), we have

$$\sum_{i\in I(x^\dagger)}|x_i - x_i^\dagger| \leq |I(x^\dagger)|\max_{i\in I(x^\dagger)}\|\omega_i\|_Y(1+\gamma)\|F(x) - F(x^\dagger)\|_Y. \tag{3.28}$$

A combination of (3.27) and (3.28) implies that

$$\begin{aligned}(\alpha - \beta)\|x - x^\dagger\|_{\ell_1} &\leq \mathcal{R}_{\alpha,\beta}(x) - \mathcal{R}_{\alpha,\beta}(x^\dagger)\\ &\quad + \left(2\alpha + \frac{M_1}{\|x^\dagger\|_{\ell_2}}\beta\right)|I(x^\dagger)|\max_{i\in I(x^\dagger)}\|\omega_i\|_Y(1+\gamma)\|F(x) - F(x^\dagger)\|_Y,\end{aligned} \tag{3.29}$$

i.e.,

$$(\alpha - \beta)\|x - x^\dagger\|_{\ell_1} \leq \mathcal{R}_{\alpha,\beta}(x) - \mathcal{R}_{\alpha,\beta}(x^\dagger) + (c_4\alpha - c_5\beta)\|F(x) - F(x^\dagger)\|_Y,$$

where

$$c_4 = 2|I(x^\dagger)|\max_{i\in I(x^\dagger)}\|\omega_i\|_Y(1+\gamma), \quad c_5 = -\frac{M_1}{\|x^\dagger\|_{\ell_2}}|I(x^\dagger)|\max_{i\in I(x^\dagger)}\|\omega_i\|_Y(1+\gamma).$$

Obviously, $c_4 > c_5$ and $c_4\alpha - c_5\beta > 0$. The proof is completed. ∎

Note that the proof of Lemma 2.17 in [1] can be performed up to (2.32) in [1] and that (2.32) in [1] corresponds to this estimate choosing m_1 as the right-hand side of (3.10). With Lemma 3.10 at our disposal, we can obtain the following result, the proof is almost same as that in [1, Theorem 2.18].

Theorem 3.11. *Suppose Assumption 3.8 holds. Let $x^\delta_{\alpha,\beta}$ be defined by (2.1) and let the constants $c_4 > c_5$ be as in Lemma 3.10.*

1. If $q = 1$ and $1 - (c_4\alpha - c_5\beta) > 0$, then

$$\|x^\delta_{\alpha,\beta} - x^\dagger\|_{\ell_1} \le \frac{1 + (c_4\alpha - c_5\beta)}{(\alpha - \beta)}\delta, \qquad \|F(x^\delta_{\alpha,\beta}) - y^\delta\|_Y \le \frac{1 + (c_4\alpha - c_5\beta)}{1 - (c_4\alpha - c_5\beta)}\delta. \tag{3.30a}$$

2. If $q > 1$, then

$$\|x^\delta_{\alpha,\beta} - x^\dagger\|_{\ell_1} \le \frac{1}{\alpha - \beta}\left[\frac{\delta^q}{q} + (c_4\alpha - c_5\beta)\delta + \frac{(q-1)2^{\frac{1}{q-1}}(c_4\alpha - c_5\beta)^{\frac{q}{q-1}}}{q}\right],$$

$$\|F(x^\delta_{\alpha,\beta}) - y^\delta\|_Y^q \le q\left[\frac{\delta^q}{q} + (c_4\alpha - c_5\beta)\delta + \frac{(q-1)2^{\frac{1}{q-1}}(c_4\alpha - c_5\beta)^{\frac{q}{q-1}}}{q}\right]. \tag{3.30b}$$

Obviously, if $\alpha \sim \delta^{q-1}$ with $q > 1$, then $\|x^\delta_{\alpha,\beta} - x^\dagger\|_{\ell_2} \le c\delta$ for some constant $c > 0$. For the particular case $q = 1$, if $\alpha \sim \delta^{1-\epsilon}$ $(0 < \epsilon < 1)$, then $\|x^\delta_{\alpha,\beta} - x^\dagger\|_{\ell_2} \le c\delta^\epsilon$ for some constant $c > 0$, where $\beta = \eta\alpha$.

Note that we cannot get the inequality (3.26) if the restriction (3.24) is replaced by (3.2). Also, the above results on the convergence rate hold only for the case $\alpha > \beta$. When $\alpha = \beta$, Lemmas 3.4 and 3.10 are no longer meaningful. So the proofs of the convergence rate are invalid if $\alpha = \beta$.

If the regularization parameter α is determined by Morozov's discrepancy principle, we can also obtain the convergence rate $O(\delta)$, cf. [1, Theorem 2.20] for a proof.

Theorem 3.12 (*Discrepancy Principle*). *Keep the assumptions of Lemma 3.10 and let $x^\delta_{\alpha,\beta}$ be defined by (2.1). If there exist α and β $(\beta = \eta\alpha)$ such that*

$$\delta \le \|F(x^\delta_{\alpha,\beta}) - y^\delta\|_Y \le \tau\delta \quad (\tau \ge 1),$$

then

$$\|x^\delta_{\alpha,\beta} - x^\dagger\|_{\ell_2} \le \frac{(c_4 - c_5\eta)(\tau + 1)\delta}{1 - \eta}.$$

4. Computational approach

In this section we introduce and analyze a solution algorithm for the problem (1.2) in the finite dimensional space $\mathbb{R}^n$. We propose an iterative soft thresholding algorithm based on the generalized conditional gradient method (GCGM). We prove the convergence of the algorithm and show that GCGM can be applied to the $\alpha\|x\|_{\ell_1} - \beta\|x\|_{\ell_2}$ $(\alpha \ge \beta \ge 0)$ sparsity regularization for nonlinear inverse problems.

4.1. Generalized conditional gradient method

In [31,32], GCGM was proposed to solve a minimization problem for a functional $G + \Phi$ on a Hilbert space H, where $G : H \to \mathbb{R}$ is continuously Fréchet differentiable and $\Phi : H \to \mathbb{R} \cup \{\infty\}$ is proper, convex, lower semi-continuous and coercive. In addition, GCGM has been applied to solve the classical sparsity regularization by setting $G(x) = \frac{1}{2}\|F(x) - y^\delta\|_Y^2 - \frac{\lambda}{2}\|x\|_{\ell_2}^2$ and $\Phi(x) = \frac{\lambda}{2}\|x\|_{\ell_2}^2 + \alpha\sum_n w_n|\langle x, \phi_n\rangle|^p$ with $p \ge 1$, where $\{w_n > 0\}$ are the weights, $\{\phi_n\}$ is an orthonormal basis of H, and F is a linear (or nonlinear) operator. GCGM from [32] is stated in the form of Algorithm 1.

Algorithm 1 Generalized conditional gradient method

1: Choose $x^0 \in H$ such that $\Phi(x^0) < +\infty$, and set $k = 0$.

2: Determine a solution z^k by solving

$$\min_{z\in H}\langle G'(x^k), z\rangle + \Phi(z).$$

3: Set a step size s^k as a solution of

$$\min_{s\in[0,1]} G(x^k + s(z^k - x^k)) + \Phi(x^k + s(z^k - x^k)).$$

4: Put $x^{k+1} = x^k + s^k(z^k - x^k)$, and $k \leftarrow k + 1$, return to Step 2.

We now consider applying GCGM to solve the problem (1.2) in the finite dimensional space $\mathbb{R}^n$. In this section, we assume that the nonlinear operator $F : \mathbb{R}^n \to \mathbb{R}^m$ is continuously Fréchet differentiable and is bounded on bounded sets. For simplicity, we only consider the case $q = 2$ in (1.2). Since the term $\alpha\|x\|_{\ell_1} - \beta\|x\|_{\ell_2}$ is not convex, a property required by GCGM, we rewrite $\mathcal{J}^{\delta}_{\alpha,\beta}(x)$ in (1.2) in the finite dimensional space $\mathbb{R}^n$ as

$$\mathcal{J}^{\delta}_{\alpha,\beta}(x) = G(x) + \Phi(x), \tag{4.1}$$

where

$$G(x) = \frac{1}{2}\|F(x) - y^{\delta}\|^2_{\ell_2} - \Theta(x), \quad \Phi(x) = \Theta(x) + \alpha\|x\|_{\ell_1} - \beta\|x\|_{\ell_2},$$

and $\Theta(x) = \frac{\lambda}{2}\|x\|^2_{\ell_2} + \beta\|x\|_{\ell_2}$, $\lambda > 0$. Thus, the problem (1.2) can be expressed as

$$\min_x \mathcal{J}^{\delta}_{\alpha,\beta}(x). \tag{4.2}$$

It is clear that $\Phi(x) = \alpha\|x\|_{\ell_1} + \frac{\lambda}{2}\|x\|^2_{\ell_2}$ is proper, convex, lower semi-continuous and coercive in ℓ_2. Unfortunately, since $\Theta'(x) = \lambda x + \frac{\beta x}{\|x\|_{\ell_2}}$, $G(x) = \frac{1}{2}\|F(x) - y^{\delta}\|^2_Y - \left(\frac{\lambda}{2}\|x\|^2_{\ell_2} + \beta\|x\|_{\ell_2}\right)$ is not Fréchet differentiable at $x = 0$. So G fails to fulfill the smoothness condition required by GCGM. We recall the definition of the soft-thresholding (ST) function

$$\mathbb{S}_{\alpha/\lambda}(x) = \sum_i S_{\alpha/\lambda}(x_i)e_i, \tag{4.3}$$

where $S_{\alpha/\lambda} : \mathbb{R} \to \mathbb{R}$ is defined by

$$S_{\alpha/\lambda}(t) = \begin{cases} t - \frac{\alpha}{\lambda} & \text{if} \quad t \geq \frac{\alpha}{\lambda}, \\ 0 & \text{if} \quad |t| < \frac{\alpha}{\lambda}, \\ t + \frac{\alpha}{\lambda} & \text{if} \quad t \leq -\frac{\alpha}{\lambda}. \end{cases} \tag{4.4}$$

In this paper, we propose a numerical algorithm based on the idea that when an iterate is zero, the next iterate is computed by solving the classical ℓ_1 sparsity regularization problem, and otherwise, the next iterate is obtained by solving the minimization problem (4.2) with GCGM. We call it ST-($\alpha\ell_1 - \beta\ell_2$) algorithm which is summarized in Algorithm 2.

Algorithm 2 ST-($\alpha\ell_1 - \beta\ell_2$) algorithm for problem (1.2) in the finite dimensional space $\mathbb{R}^n$

Let $x^0 = 0$ and choose a small number $\epsilon > 0$ for stopping criterion.
Determine a solution x^1 of the problem $\min \frac{1}{2}\|F(x) - y^{\delta}\|^2_Y + \alpha\|x\|_{\ell_1}$
If $x^1 = 0$, then stop and take 0 as the solution;
otherwise, for $k \geq 1$ do the following until $\|x^{k+1} - x^k\|_{\ell_2} < \epsilon$:
Determine a solution z^k of the problem $\min_{z\in H}\langle G'(x^k), z\rangle + \Phi(z)$ by computing

$$z^k = \mathbb{S}_{\alpha/\lambda}\left(\left(\frac{\beta}{\lambda\|x^k\|_{\ell_2}} + 1\right)x^k - \frac{1}{\lambda}F'(x^k)^*(F(x^k) - y^{\delta})\right)$$

Set a step size s^k as a solution of

$$\min_{s\in[0,1]} G(x^k + s(z^k - x^k)) + \Phi(x^k + s(z^k - x^k)).$$

$$x^{k+1} = x^k + s^k(z^k - x^k)$$
$$k \leftarrow k + 1$$

For convenience in presentation, in case $x^1 = 0$, we formally let $x^k = 0$ for $k \geq 2$. Next, we recall a result proved in [32].

Lemma 4.1. *Let $G : \ell_2 \to \mathbb{R}$ denote a Gâteaux-differentiable functional and let $\Phi : \ell_2 \to \mathbb{R}$ be proper, convex, lower semi-continuous and coercive. Then, the first order necessary condition for optimality in (4.2) is*

$$x \in \ell_2 : \quad \langle G'(x), y - x\rangle \geq \Phi(x) - \Phi(y) \quad for\ all \quad y \in \ell_2. \tag{4.5}$$

This condition is equivalent to

$$\langle G'(x), x\rangle + \Phi(x) = \min_{y\in\ell_2}(\langle G'(x), y\rangle + \Phi(y)). \tag{4.6}$$

In the following, we show that $\mathcal{J}^{\delta}_{\alpha,\beta}(x^k)$ decreases with respect to k, where $\{x^k\}$ is generated by Algorithm 2.

Lemma 4.2. *Denote by $\{x^k\}$ the sequence generated by Algorithm 2. Suppose x^k does not fulfill the first order optimality condition (4.5). Then $\mathcal{J}^{\delta}_{\alpha,\beta}(x^{k+1}) \leq \mathcal{J}^{\delta}_{\alpha,\beta}(x^k)$.*

Proof. If $x^k = 0$, by Algorithm 2, we have

$$\mathcal{J}^{\delta}_{\alpha,\beta}(x^{k+1}) = G(x^{k+1}) + \Phi(x^{k+1}) = \frac{1}{2}\|F(x^{k+1}) - y^{\delta}\|^2_{\ell_2} + \alpha\|x^{k+1}\|_{\ell_1} - \beta\|x^{k+1}\|_{\ell_2}$$

$$\leq \frac{1}{2}\|F(0)-y^\delta\|_{\ell_2}^2+\alpha\|0\|_{\ell_1}-\beta\|x^{k+1}\|_{\ell_2} \leq \frac{1}{2}\|F(0)-y^\delta\|_{\ell_2}^2+\alpha\|0\|_{\ell_1}-\beta\|0\|_{\ell_2}$$
$$= \mathcal{J}_{\alpha,\beta}^\delta(x^k). \tag{4.7}$$

If $x^k \neq 0$, G is Fréchet differentiable at x^k and the rest of the proof is similar to that of Lemma 2 in [32]. ■

Remark 4.3. Note that if $x^1 \neq 0$, then the second inequality in (4.7) is strict and so

$$\mathcal{J}_{\alpha,\beta}^\delta(x^1) < \mathcal{J}_{\alpha,\beta}^\delta(0).$$

By Lemma 4.2,

$$\mathcal{J}_{\alpha,\beta}^\delta(x^{k+1}) \leq \mathcal{J}_{\alpha,\beta}^\delta(x^k) \leq \cdots \leq \mathcal{J}_{\alpha,\beta}^\delta(x^1) < \mathcal{J}_{\alpha,\beta}^\delta(0). \tag{4.8}$$

Consequently, $x^k \neq 0$ for any $k \geq 1$. Moreover, the limit x^* of any convergent subsequence of $\{x^k\}$ satisfies

$$x^* \neq 0.$$

This is easily proven from (4.8).

In Algorithm 2, to determine x^1, we need to solve the problem

$$\min \frac{1}{2}\|F(x)-y^\delta\|_Y^2+\alpha\|x\|_{\ell_1}. \tag{4.9}$$

This is the classical ℓ_1-norm sparsity regularization for nonlinear ill-posed problems. There are several algorithms for solving the problem (4.9) [15], e.g., GCGM, surrogate functional approach, the quadratic approximation method and generalized gradient projection method. In this paper, we use GCGM to solve the problem (4.9),

$$x_{l+1}^1 = \mathbb{S}_{\alpha/\lambda}\left(x_l^1 - \frac{1}{\lambda}F'(x_l^1)^*(F(x_l^1)-y^\delta)\right), \quad l = 0, 1, 2, \ldots,$$

cf. [15] for the proof of the convergence. Note that x^1 is not necessary a global minimizer. It could be a stationary point. Nevertheless, it can be ensured that the objective function is decreasing.

4.2. Convergence analysis

Definition 4.4. An element $0 \neq x^* \in \ell_2$ is called a stationary point of $\mathcal{J}_{\alpha,\beta}^\delta$ if it satisfies

$$\langle G'(x^*), y-x^*\rangle \geq \Phi(x^*)-\Phi(y) \quad \forall\, y \in \ell_2,$$

or

$$\langle G'(x^*), x^*\rangle + \Phi(x^*) = \min_{y\in\ell_2}(\langle G'(x^*), y\rangle + \Phi(y)).$$

Note that Algorithm 2 either "produces the sequence $\{0, 0, 0, \ldots\}$ and $x^* = 0$ is the solution" or "produces a sequence x^k with $x^k \neq 0$ for any $k \geq 1$ and any cluster point of the sequence $\{x^k\}$ is not zero". That is why Definition 4.4 is limited to the case $0 \neq x^* \in \ell_2$.

Next, we recall a result proved in [32].

Lemma 4.5. *Let Φ be proper, convex, lower semi-continuous and coercive, and let F be continuously Fréchet differentiable. Then*

$$\Psi(x) := \langle G'(x), x\rangle + \Phi(x) - \min_{y\in\ell_2}(\langle G'(x), y\rangle + \Phi(y)) \tag{4.10}$$

is lower semi-continuous.

We have the following lemma, cf. [32, Theorem 1] for the proof.

Lemma 4.6. *Denote by $\{x^k\}$ the sequence generated by Algorithm 2. Then $\lim_k \Psi(x^k) = 0$.*

Theorem 4.7. *Denote by $\{x^k\}$ the sequence generated by Algorithm 2. $\{x^k\}$ has a subsequence converging to an element x^*. If $x^1 \neq 0$, then $x^* \neq 0$, and x^* is a stationary point of the functional $\mathcal{J}_{\alpha,\beta}^\delta$.*

Proof. Since $\{\|x^k\|_{\ell_2}\}$ is bounded, there exist an element x^* and a convergent subsequence of $\{x^k\}$, still denoted by $\{x^k\}$, such that $x^k \to x^*$. If $x^1 \neq 0$, then $x^* \neq 0$ by Remark 4.3. For Ψ from Lemma 4.5 it holds that $\liminf_{k\to\infty} \Psi(x^k) \geq \Psi(x^*)$. Then, by Lemma 4.6, we have $\Psi(x^*) = 0$, which completes the proof of the theorem. ■

Since determining the optimal step size s^k is expensive, next we show that the step size can be chosen as $s^k = 1$, $\forall\, k \in \mathbb{N}$, if λ is sufficiently large. Note that $s^k = 1$ is not necessary the optimal step size; nevertheless, we can show that $\mathcal{J}_{\alpha,\beta}^\delta$ decreases with constant step size $s^k = 1$. The proof is along the lines of that of Lemma 2.4 in [31].

Assumption 4.8. Define $B_R := \{x \mid \|x-0\|_{\ell_2} \le R\}$. Assume that F is Fréchet differentiable and there exists a constant $L_{F'}(R)$ such that

$$\|F'(x) - F'(y)\|_{L(\mathbb{R}^n,\mathbb{R}^m)} \le L_{F'}(R)\|x-y\|_{\ell_2} \quad \forall\, x, y \in B_R. \tag{4.11}$$

Lemma 4.9. *Under Assumption 4.8,*

$$\|F(x) - F(y)\|_{\ell_2} \le \left(2L_{F'}(R)\,R + \|F'(0)\|_{L(\mathbb{R}^n,\mathbb{R}^m)}\right)\|x-y\|_{\ell_2} \quad \forall\, x, y \in B_R.$$

Proof. Write

$$\begin{aligned} F(x) - F(y) &= \int_0^1 \frac{d}{ds} F(sx + (1-s)y)\,ds \\ &= \int_0^1 F'(sx + (1-s)y)\,(x-y)\,ds \\ &= \int_0^1 \left[F'(sx + (1-s)y) - F'(y)\right](x-y)\,ds + F'(y)\,(x-y). \end{aligned}$$

So

$$\|F(x) - F(y)\|_{\ell_2} \le L_{F'}(R)\int_0^1 s\,\|x-y\|_{\ell_2}^2\,ds + \|F'(y)\|_{L(\mathbb{R}^n,\mathbb{R}^m)}\,\|x-y\|_{\ell_2}.$$

Note that

$$\|F'(y)\|_{L(\mathbb{R}^n,\mathbb{R}^m)} \le \|F'(y) - F'(0)\|_{L(\mathbb{R}^n,\mathbb{R}^m)} + \|F'(0)\|_{L(\mathbb{R}^n,\mathbb{R}^m)} \le L_{F'}(R)\,R + \|F'(0)\|_{L(\mathbb{R}^n,\mathbb{R}^m)}$$

and

$$\|x-y\|_{\ell_2} \le 2\,R.$$

Hence, the stated inequality is proven. ■

So under Assumption 4.8, F' and F are bounded on bounded sets.

Lemma 4.10. *Keep Assumption 4.8 and define* $\phi(x) = \frac{1}{2}\|F(x) - y^\delta\|_{\ell_2}^2$. *Then* ϕ' *is locally Lipschitz continuous, i.e. there exist a constant* $L_{\phi'}(R)$ *that depends on* R *such that*

$$\|\phi'(x) - \phi'(y)\|_{L(\mathbb{R}^n,\mathbb{R}^m)} \le L_{\phi'}(R)\|x-y\|_{\ell_2} \quad \forall\, x, y \in B_R. \tag{4.12}$$

Proof. Let $x, y \in B_R$. Then

$$\begin{aligned} \|\phi'(x) - \phi'(y)\|_{L(\mathbb{R}^n,\mathbb{R}^m)} &= \|F'(x)^*(F(x) - y^\delta) - F'(y)^*(F(y) - y^\delta)\|_{L(\mathbb{R}^n,\mathbb{R}^m)} \\ &\le \|F'(x) - F'(y)\|_{L(\mathbb{R}^n,\mathbb{R}^m)}\|F(x) - y^\delta\|_{\ell_2} \\ &\quad + \|F'(y)\|_{L(\mathbb{R}^n,\mathbb{R}^m)}\|F(x) - F(y)\|_{\ell_2}. \end{aligned} \tag{4.13}$$

Since

$$\|F(x) - y^\delta\|_{\ell_2} \le \sup_{x\in B_R}\|F(x)\|_{\ell_2} + \|y^\delta\|_{\ell_2}$$

and $\|F'(y)\|_{L(\mathbb{R}^n,\mathbb{R}^m)} \le \sup_{x\in B_R}\|F'(x)\|_{L(\mathbb{R}^n,\mathbb{R}^m)}$, (4.12) follows from (4.13) with a constant

$$\begin{aligned} L_{\phi'}(R) = L_{F'}(R)&\left(\sup_{x\in B_R}\|F(x)\|_{\ell_2} + \|y^\delta\|_{\ell_2}\right) \\ &+ \left(2L_{F'}(R)\,R + \|F'(0)\|_{L(\mathbb{R}^n,\mathbb{R}^m)}\right)\sup_{x\in B_R}\|F'(x)\|_{L(\mathbb{R}^n,\mathbb{R}^m)}. \end{aligned}$$

This completes the proof. ■

Theorem 4.11. *Choose* $C_1 > 0$, $\bar{x}$, $\beta > 0$, *and*

$$\frac{1}{2}\|F(0) - y^\delta\|_{\ell_2} \le \mathcal{J}_{\alpha,\beta}^\delta(\bar{x}) \le \infty.$$

Define

$$\begin{aligned} M_1 &:= \sup\left\{\|x\|_{\ell_2} \mid \mathcal{J}_{\alpha,\beta}^\delta(x) \le \mathcal{J}_{\alpha,\beta}^\delta(\bar{x})\right\}, \\ M_2 &:= C_1^{-1}\|\phi'(0)\|_{L(\mathbb{R}^n,\mathbb{R}^m)} + \left(1 + C_1^{-1}L_{\phi'(M_1)}\right)M_1 + C_1^{-1}\beta. \end{aligned}$$

If

$$\lambda > \max\{L_{\phi'}(M_2), C_1\},$$

Algorithm 2 with a constant step size $s^k = 1$ produces a sequence such that

$$\mathcal{J}^\delta_{\alpha,\beta}(x^{k+1}) < \mathcal{J}^\delta_{\alpha,\beta}(x^k).$$

as long as x^k does not satisfy the first order optimality condition (4.5) for a nonzero minimizer x^* of $\mathcal{J}^\delta_{\alpha,\beta}$.

Proof. If $x^k = 0$, by the proof of Lemma 4.2, we have $\mathcal{J}^\delta_{\alpha,\beta}(x^{k+1}) \le \mathcal{J}^\delta_{\alpha,\beta}(x^k)$. Meanwhile, $\|x^k\|_{\ell_2} \le M_1$ and

$$\mathcal{J}^\delta_{\alpha,\beta}(x^k) = \frac{1}{2}\|F(0) - y^\delta\|_{\ell_2} \le \mathcal{J}^\delta_{\alpha,\beta}(\bar{x}).$$

If $x^k \neq 0$, we assume that $\|x^k\|_{\ell_2} \le M_1$ and $\mathcal{J}^\delta_{\alpha,\beta}(x^k) \le \mathcal{J}^\delta_{\alpha,\beta}(\bar{x})$. Next we show it still holds for x^{k+1}. If $s^k = 1$,

$$x^{k+1} = \mathbb{S}_{\alpha/\lambda}\left(\left(\frac{\beta}{\lambda\|x^k\|_{\ell_2}} + 1\right)x^k - \frac{1}{\lambda}F'(x^k)^*(F(x^k) - y^\delta)\right).$$

By the contraction property of $\mathbb{S}_{\alpha/\lambda}$ [13],

$$\begin{aligned}\|x^{k+1}\|_{\ell_2} &\le \lambda^{-1}\|\phi'(x^k)\|_{L(\mathbb{R}^n,\mathbb{R}^m)} + \|x^k\|_{\ell_2} + \lambda^{-1}\beta\\ &\le \lambda^{-1}\|\phi'(0)\|_{L(\mathbb{R}^n,\mathbb{R}^m)} + \lambda^{-1}L_{\phi'}(M_1)\|x^k\|_{\ell_2} + \|x^k\|_{\ell_2} + \lambda^{-1}\beta\\ &\le C_1^{-1}\|\phi'(0)\|_{L(\mathbb{R}^n,\mathbb{R}^m)} + (1 + C_1^{-1}L_{\phi'}(M_1))M_1 + C_1^{-1}\beta = M_2.\end{aligned}$$

Define

$$G(x) := H(x) - \beta\|x\|_{\ell_2},$$

where

$$H(x) := \frac{1}{2}\|F(x) - y^\delta\|^2_{\ell_2} - \frac{\lambda}{2}\|x\|^2_{\ell_2}.$$

By the Taylor expansion, we have

$$\begin{aligned}G(x^{k+1}) - G(x^k) &= (H(x^{k+1}) - H(x^k)) - (\beta\|x^{k+1}\|_{\ell_2} - \beta\|x^k\|_{\ell_2})\\ &= \int_0^1 \langle H'(x^k + t(x^{k+1} - x^k)), x^{k+1} - x^k\rangle dt - (\beta\|x^{k+1}\|_{\ell_2} - \beta\|x^k\|_{\ell_2})\\ &= \int_0^1 \langle H'(x^k + t(x^{k+1} - x^k)) - H'(x^k), x^{k+1} - x^k\rangle dt + \langle G'(x^k), x^{k+1} - x^k\rangle\\ &\quad + \left\langle \frac{\beta x^k}{\|x^k\|_{\ell_2}}, x^{k+1} - x^k\right\rangle - (\beta\|x^{k+1}\|_{\ell_2} - \beta\|x^k\|_{\ell_2}).\end{aligned} \tag{4.14}$$

Since x^k does not fulfill the first order optimality condition (4.5), we have

$$\langle G'(x^k), x^{k+1} - x^k\rangle < \Phi(x^k) - \Phi(x^{k+1}). \tag{4.15}$$

A combination of (4.14) and (4.15) implies that

$$\begin{aligned}&(G(x^{k+1}) + \Phi(x^{k+1})) - (G(x^k) + \Phi(x^k))\\ &< \int_0^1 \langle H'(x^k + t(x^{k+1} - x^k)) - H'(x^k), x^{k+1} - x^k\rangle dt\\ &\quad + \left\langle \frac{\beta x^k}{\|x^k\|_{\ell_2}}, x^{k+1} - x^k\right\rangle - (\beta\|x^{k+1}\|_{\ell_2} - \beta\|x^k\|_{\ell_2})\\ &= \int_0^1 \langle \phi'(x^k + t(x^{k+1} - x^k)) - \phi'(x^k), x^{k+1} - x^k\rangle dt - \frac{\lambda}{2}\|x^{k+1} - x^k\|^2_{\ell_2}\\ &\quad + \left\langle \frac{\beta x^k}{\|x^k\|_{\ell_2}}, x^{k+1} - x^k\right\rangle - (\beta\|x^{k+1}\|_{\ell_2} - \beta\|x^k\|_{\ell_2}).\end{aligned} \tag{4.16}$$

By Lemma 4.10 and Cauchy–Schwarz inequality, it follows from (4.16) that

$$\begin{aligned}&(G(x^{k+1}) + \Phi(x^{k+1})) - (G(x^k) + \Phi(x^k))\\ &< \int_0^1 L_{\phi'}(M_2)t\|x^{k+1} - x^k\|^2_{\ell_2}dt - \frac{\lambda}{2}\|x^{k+1} - x^k\|^2_{\ell_2} + \left\langle \frac{\beta x^k}{\|x^k\|_{\ell_2}}, x^{k+1}\right\rangle - \beta\|x^{k+1}\|_{\ell_2}\\ &\le \frac{1}{2}\left(L_{\phi'}(M_2) - \lambda\right)\|x^{k+1} - x^k\|^2_{\ell_2} < 0.\end{aligned}$$

This implies that

$$\mathcal{J}^{\delta}_{\alpha,\beta}(x^{k+1}) < \mathcal{J}^{\delta}_{\alpha,\beta}(x^k) \leq \mathcal{J}^{\delta}_{\alpha,\beta}(\bar{x}).$$

By the definition of M_1, we have $\|x^{k+1}\|_{\ell_2} \leq M_1$. ■

4.3. Determining a solution z^k

In Algorithm 2, when $x^k \neq 0$, a crucial issue is how to determine the direction z^k. The Fréchet derivative of $G(x)$ is given by

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

In Algorithm 2, z^k is given as a solution of the minimization problem

$$\min_z \langle F'(x^k)^*(F(x^k) - y^\delta) - \lambda x^k - \frac{\beta x^k}{\|x^k\|_{\ell_2}}, z\rangle + \frac{\lambda}{2}\|z\|^2_{\ell_2} + \alpha\|z\|_{\ell_1}. \tag{4.17}$$

The minimizer of (4.17) can be computed by the iterative soft thresholding operator, cf. [10,13] for the details of iterative soft thresholding operator.

Lemma 4.12. *If $x^k \neq 0$, then z^k is the minimizer of problem (4.17) if and only if*

$$z^k = \mathbb{S}_{\alpha/\lambda}\left(\left(\frac{\beta}{\lambda\|x^k\|_{\ell_2}} + 1\right)x^k - \frac{1}{\lambda}F'(x^k)^*(F(x^k) - y^\delta)\right). \tag{4.18}$$

Proof. The problem (4.17) is equivalent to the problem

$$\min_z \frac{1}{2}\left\| z - \left(\left(\frac{\beta}{\lambda\|x^k\|_{\ell_2}} + 1\right)x^k - \frac{1}{\lambda}F'(x^k)^*(F(x^k) - y^\delta)\right)\right\|^2_{\ell_2} + \frac{\alpha}{\lambda}\|z\|_{\ell_1}. \tag{4.19}$$

The proximal mapping P of $\alpha\lambda^{-1}\|z\|_{\ell_1}$ is defined by

$$P_{\alpha\lambda^{-1}\|\cdot\|_{\ell_1}}(x) := \arg\min_z \frac{1}{2}\|z - x\|^2_{\ell_2} + \frac{\alpha}{\lambda}\|z\|_{\ell_1}.$$

For penalty functionals of the form $\alpha\lambda^{-1}\|z\|_{\ell_1}$, we have the well known equivalence, see, e.g., [15,40],

$$P_{\alpha\lambda^{-1}\|\cdot\|_{\ell_1}}(x) = (I + \frac{\alpha}{\lambda}\partial\|\cdot\|_{\ell_1})^{-1}(x) = \mathbb{S}_{\alpha/\lambda}(x).$$

Then we can obtain (4.18). ■

The componentwise form of $\omega \in \partial\|\cdot\|_{\ell_1}(x)$ is: $\omega_i = \mathrm{sgn}(x_i)$ if $x_i \neq 0$, $\omega_i \in [-1, 1]$ if $x_i = 0$, for any component subscript i.

5. Numerical experiments

Though GCGM can be applied to solve $\alpha\ell_1 - \beta\ell_2$ sparsity regularization, it is challenging to determine the optimal step size. In analogy to that in [31], in this section, we implement Algorithm 2 with constant step size $s = 1$ for a nonlinear compressive sensing (CS) problem [41–44]. Here we are interested in the sparse recovery for a CS problem where the observed signal is measured with some nonlinear system. The research of nonlinear CS is not only important in theoretical analysis but also in many applications, where the observation system is often nonlinear. For example, in diffraction imaging, charge coupled device (CCD) records the amplitude of the Fourier transform of the original signal. So one only obtains the nonlinear measurements of the original signal. Fortunately, in [42], it is shown that if the system satisfies some nonlinear conditions then recovery should still be possible.

Under the nonlinear CS frame, the measurement system is nonlinear. Assume, therefore, that the observation model is

$$y = F(x) + \delta, \tag{5.1}$$

where $\delta \in \mathbb{R}^m$ is a noise level, $x \in \mathbb{R}^n$ and $F : \mathbb{R}^n \to \mathbb{R}^m$ is a nonlinear operator. It can be shown that if the linearization of F at an exact solution $x^\dagger$ satisfies the restricted isometry property (RIP), then the convergence property of the iterative hard thresholding algorithm (IHTA) is guaranteed [42]. Next we illustrate the efficiency of the proposed algorithm by a nonlinear CS example of the form

$$y = F(x) := \hat{a}(A\hat{b}(x)) \tag{5.2}$$

which was introduced in [42], where A is a CS matrix, $\hat{a}$ and $\hat{b}$ are nonlinear operators, respectively. Here, $\hat{a}$ encodes nonlinearity after mixing by A as well as nonlinear "crosstalk" between mixed elements, and $\hat{b}$ encodes the same system properties for the inputs before mixing. For simplicity, we write $\hat{a}(x) = x + a(x)$ and $\hat{b}(x) = x + b(x)$, where again a and b are nonlinear maps. In particular, let $a(x) = x^c$ and $b(x) = x^d$, where $c, d \in \mathbb{N}_+$ and x^c and x^d should be understood in a componentwise sense.

Next, we consider the nonlinearity of the operator $F(x)$. In [42], it is shown that the Jacobian matrix of $\hat{a}(A\hat{b}(x))$ is of the form

$$F'(x) = (I + a')[A(I + b')].$$

Table 1
SNR of iterative solution x^* with different λ.

λ	Iteration number	Computation time	SNR
10	149	5.2690s	35.0692
50	806	29.9565	35.0240
100	1628	58.5911	35.0146
200	3272	115.9108	35.0089
500	8203	438.6556	35.0002
1000	16 423	636.3524	35.0011
2000	32 862	1661.3564	35.0004

We assume that $\|x\|_{\ell_2}$ is bounded, then a' is bounded on bounded sets. Then, there exists a constant $c > 0$ such that $\|F'(x_1) - F'(x_2)\|_{L(\mathbb{R}^n,\mathbb{R}^m)} \le c\|x_1 - x_2\|_{\ell_2}$, i.e. $F'(x)$ is Lipschitz continuous, cf. [42, Lemmas 3 & 4].

We present several numerical tests which demonstrate the efficiency of the proposed method. To make Algorithm 2 clear to the reader, we study the influence of the parameters λ, η, s and the nonlinear maps a and b on the iterative result x^*. Note that if $\eta = 0$ i.e. $\beta = 0$, (1.2) reduces to the convex ℓ_1 sparsity regularization. Then (4.18) reduces to the form

$$z^k = \mathbb{S}_{\alpha/\lambda}\left(x^k - \frac{1}{\lambda}F'(x^k)^*(F(x^k) - y^\delta)\right).$$

For the numerical simulation, we use a setting that A is a Gaussian random measurement matrix. The nonlinear CS problem is of the form $\hat{a}(A_{m\times n}\hat{b}(x_n)) = y_m$, where $A_{m\times n}$ is a Gaussian random measurement matrix. The exact solution $x^\dagger$ is s-sparse. The exact data $y^\dagger$ is obtained by $y^\dagger = \hat{a}(A\hat{b}(x^\dagger))$. White Gaussian noise is added to the exact data $y^\dagger$ and δ is the noise level, measured in dB. The iterative solution is denoted by x^*. The performance of the iterative solution x^* is evaluated by signal-to-noise ratio (SNR) which is defined by

$$\text{SNR} := -10\log_{10}\frac{\|x^* - x^\dagger\|_{\ell_2}^2}{\|x^\dagger\|_{\ell_2}^2}.$$

We utilize the discrepancy principle to choose the regularization parameter α. Starting with an initial guess of the regularization parameter, if $\|F(x^*) - y\|_{\ell_2} > \delta$, then we keep halving the value of the regularization parameter until $\|F(x^*) - y\|_{\ell_2} > \delta$.

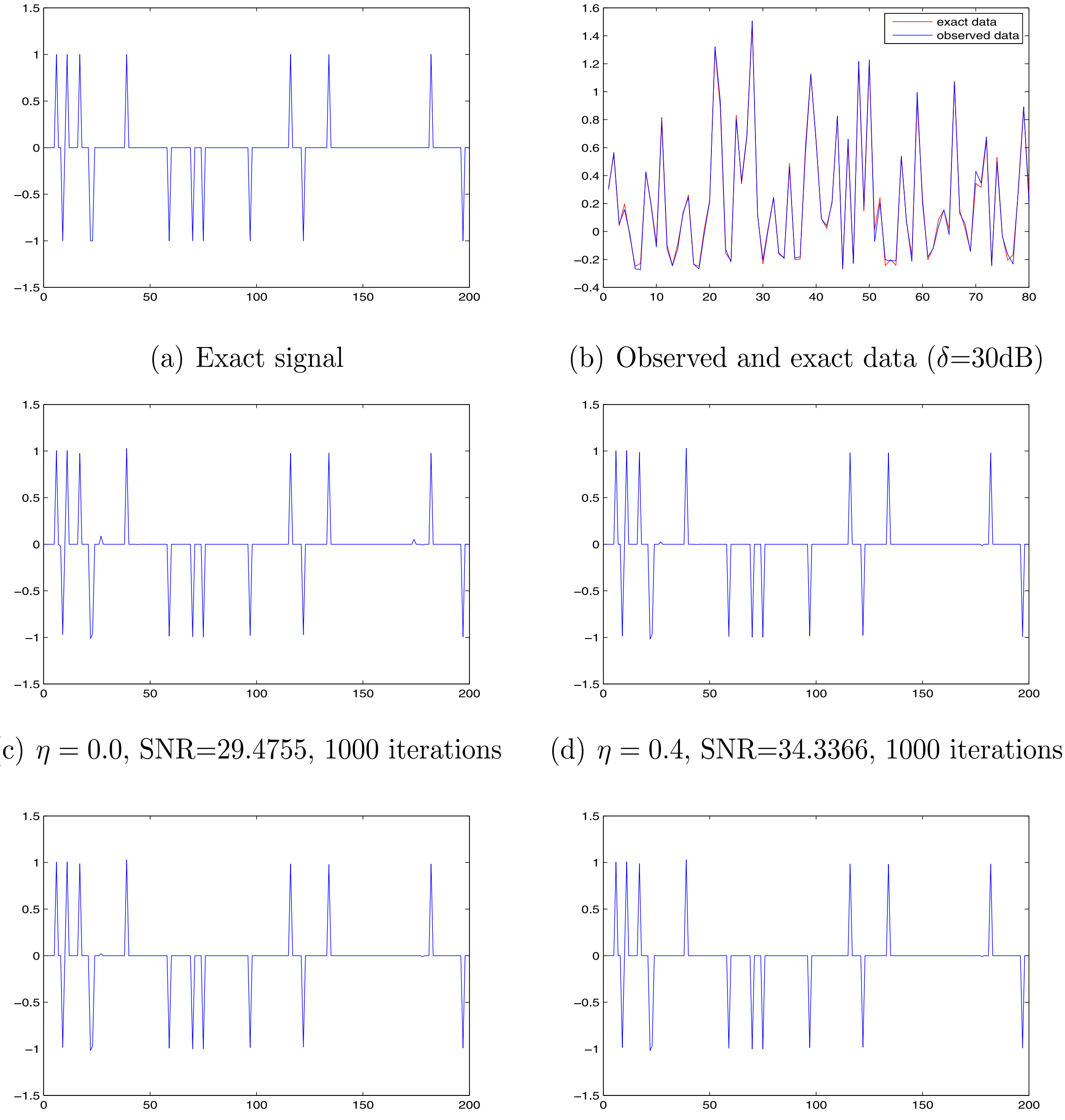
Let $n = 200$, $m = 0.4n$, $p = 0.2m$, where p is the number of the impulses in the true solution. For the sparsity regularization of linear ill-posed problems, the value of $\|A_{m\times n}\|_2$ needs to be less than 1 [13]. This requirement is also needed for the nonlinear CS problem (5.1). The value of $\|A_{m\times n}\|_2$ is around 20, and we re-scale the matrix $A_{m\times n}$ by $A_{m\times n} \to 0.05A_{m\times n}$. The initial value x^0 in Algorithm 2 is generated by calling $x^0 = 1\text{e-}6 * \text{ones}(n,1)$ in MATLAB. Actually, for sparse recovery, one natural choice for the initial value x^0 is $\mathbf{0}$ vector, i.e. $\text{zeros}(n,1)$. If $x^0 = \text{zeros}(n,1)$, we compute x^1 by the classical ℓ_1 sparsity regularization and the number of iterations is 10. All numerical experiments were tested in MATLAB R2010 on an i7-6500U 2.50 GHz workstation with 8 Gb RAM.

In the first test, we discuss the convergence and convergence rate of the proposed algorithm. We let $c = 2$, $d = 3$ and $s = 1$. The noise level δ is 30dB. We choose different parameters η to test its influence on the iterative solution x^*. Theoretically, for Algorithm 2 with a constant step size $s = 1$, we require that the condition in Theorem 4.11 holds, i.e. λ is large enough. Next, we test whether λ satisfies this condition. In Theorem 4.11, we let $c_1 = 1$, $\bar{x} = x^\dagger$, $\alpha = 0.125$ and $\beta = 0.05$. Then, $\mathcal{J}^\delta_{\alpha,\beta}(\bar{x}) = 0.0316$, M_1 is around 16, $\|\phi'(0)\|_{\ell_2}$ is around 3.5706 and $L_{\phi'}(M_1)$ is around 4.5794. Hence, M_2 is around 9.2. The value of $L_{\phi'}(M_2)$ is around 18.3176. So we let $\lambda = 20$. Fig. 1 shows the graphs of the iterative solution x^* when the regularization parameter $\alpha = 0.125$. It is obvious that the results of iteration get better with η increasing, which shows that the non-convex regularization with $\eta > 0$ has better performance than the classical ℓ_1 regularization. Fig. 2 displays graphs of the iterative solution x^* with respect to iteration number k when $\eta = 1.0$. It shows a good convergence pattern. Fig. 3 shows the convergence rate of iterative solution x^* with respect to iteration number k. We use the relative error $e = \|x^* - x^\dagger\|_{\ell_2}/\|x^\dagger\|_{\ell_2}$ to evaluate the accuracy of x^*.

In the second test, we examine the effect of the parameter λ. Let $n = 200$, $m = 0.6n$, $p = 0.2m$. By (4.1), it is obvious that the iterative results do not change with respect to λ. However, numerical results show that larger λ leads to more iterations number and iterations time. Actually, from the formulation of Algorithm 2, we see that a larger value of λ admits a smaller value of the threshold α/λ. Theoretically, if λ is sufficiently large, the convergence speed could be arbitrarily slow and it is computationally expensive. In Table 1, we set $\eta = 1.0$ and let α, δ, c and d same as that in test 1 and give the iterative results of Algorithm 2 with different λ. If the norm value of the iterative solution does not change for three subsequent iterations, then we stop the iteration. It is shown that one needs more iterations and higher computation time with the value of λ increasing.

In the third test, we study the stability of Algorithm 2. To test the influence of δ, we choose several different noise levels which are added to the exact data $y^\dagger$. Table 2 displays the iterative results. Obviously, the SNR of Inversion solution x^* increase with the noise level decreasing. It is shown that we can obtain satisfactory result when the noise level $\delta \ge 20$dB. Meanwhile, Table 2 shows that Algorithm 2 has good stability corresponding to the noise level when the parameter η is fixed. This implies that the constant step-size variant of Algorithm 2 is not sensitive with respect to η.

In the last test, we discuss the influence of the nonlinearity of F, i.e. the parameter c and d on the iterative solution x^*. The nonlinearity of the CS problem (5.2) depends on the parameters c and d. In particular, the degree of nonlinearity of F increases with the parameter c and d increasing. In Table 3, we set $\eta = 1.0$ and let α, δ and s same as that in Test 1. It is obvious that the iterative results are stable with respect to the parameter c. The SNR of the iterative solution x^* are similar with different parameter



(a) Exact signal

(b) Observed and exact data (δ=30dB)

(c) $\eta = 0.0$, SNR=29.4755, 1000 iterations

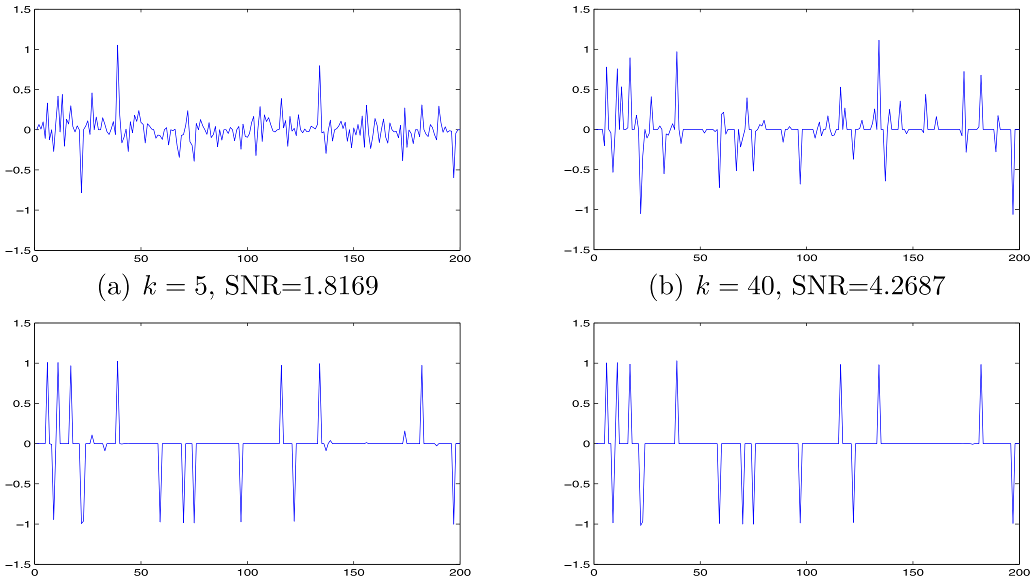
(d) $\eta = 0.4$, SNR=34.3366, 1000 iterations

(e) $\eta = 0.8$, SNR=35.0716, 1000 iterations

(f) $\eta = 1.0$, SNR=35.3914, 1000 iterations

Fig. 1. (a) Exact signal. (b) Observed and exact data. (c)–(f) The iterative solution x^* with different η at a fixed regularization parameter $\alpha = 0.125$.

c. However, the iterative results are sensitive with respect to the parameter d. When $d \geq 7$, we cannot get satisfactory results. In particular, Algorithm 2 is invalid when the parameter d is an even number. Fig. 4 shows the iterative solution x^* with respect to the iterations at the fixed parameter $c = 2$ and $d = 4$. Actually, Algorithm 2 can only identify the positive impulses and it fails to recovery the negative impulses when d is an even number. Theoretically, due to the non-convexity of $\mathcal{J}^{\delta}_{\alpha,\beta}$ in (1.2), the minimizer of $\mathcal{J}^{\delta}_{\alpha,\beta}$ may be non-unique. Nevertheless, in numerical experiments, convergence is still observed and the limit depends on the choice of the initial vector x^0. When d is an even number, we cannot obtain the desired iterative results when x^0 =1e-6$*$ones(n,1). Nevertheless, we still obtain satisfactory iterative results when we let $x^0 = 0.9 * x^{\dagger}$.



(a) $k = 5$, SNR=1.8169 (b) $k = 40$, SNR=4.2687

(c) $k = 80$, SNR=23.8996 (d) $k = 100$, SNR=35.3914

Fig. 2. The iterative solution x^* with different iteration number k at a fixed parameter $\eta = 1.0$.

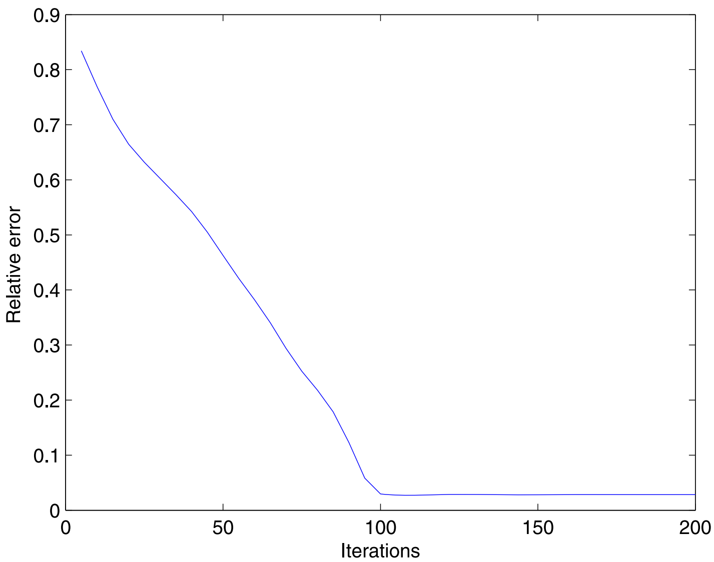


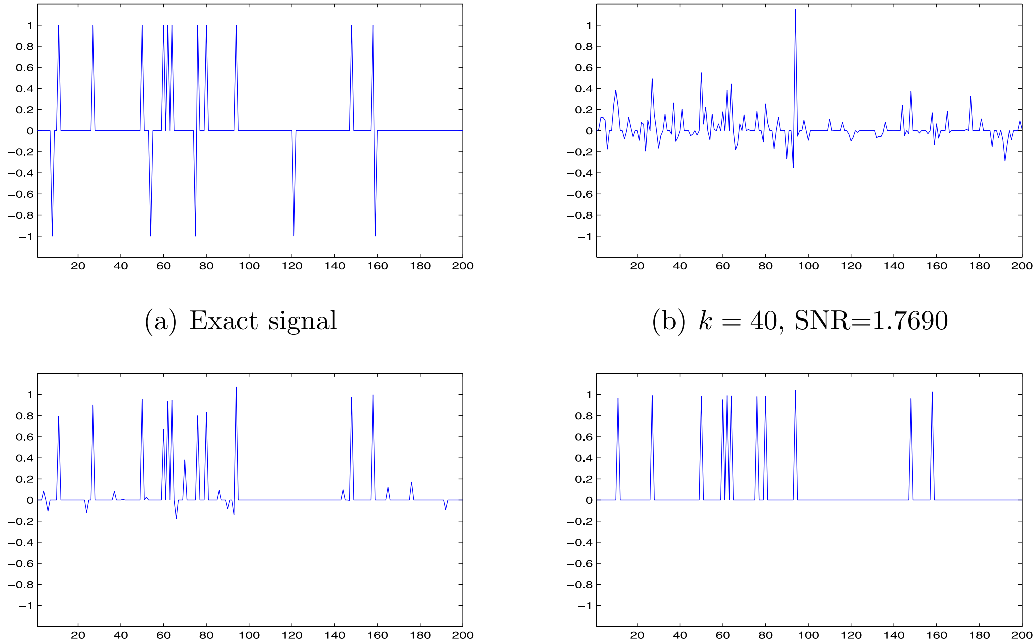
Fig. 3. Convergence rate of the iterative solution x^* with respect to iteration number k at a fixed parameter $\eta = 1.0$.

Table 2
SNR of iterative solution x^* with several noise levels, iterations number 1000.

	$\eta = 0$	$\eta = 0.2$	$\eta = 0.4$	$\eta = 0.6$	$\eta = 0.8$	$\eta = 1.0$
Noise free, $\alpha = 0.015$	44.5479	45.0044	45.4848	45.9917	46.5286	47.0994
$\delta =$ 50 dB, $\alpha = 0.031$	43.3370	45.7636	46.2097	46.6770	47.1678	47.6844
$\delta =$ 40 dB, $\alpha = 0.062$	36.2226	37.6898	38.1772	38.6865	39.2201	39.7804
$\delta =$ 30 dB, $\alpha = 0.125$	29.4775	32.1682	34.3366	34.9682	35.0716	35.3914
$\delta =$ 20 dB, $\alpha = 0.125$	22.4699	24.1333	25.5513	25.7172	25.7959	25.9081
$\delta =$ 10 dB, $\alpha = 0.250$	−1.5015	−1.5146	−1.5307	−1.5438	−1.5546	−1.5641

Table 3
SNR of iterative solution x^* with different parameters c and d, iteration number 1000.

	$d=1$	$d=2$	$d=3$	$d=4$	$d=5$	$d=6$	$d=7$	$d=8$	$d=9$
$c=1$	46.2683	4.1589	41.8642	3.2158	41.2564	2.5784	13.5876	NaN	NaN
$c=2$	40.6830	4.2591	42.0519	3.0102	43.8885	1.2486	11.9139	NaN	NaN
$c=3$	39.0531	3.0102	40.5454	1.2494	33.1849	2.0412	14.1032	NaN	NaN
$c=4$	39.4280	2.0409	39.5473	2.0311	33.4812	3.0022	18.8408	6.0172	NaN
$c=5$	39.5284	3.0097	40.5428	2.0412	31.6905	4.2200	17.0556	2.0336	NaN
$c=6$	39.8423	3.0095	39.9291	2.0411	35.7897	1.2494	16.8327	2.0403	NaN
$c=15$	39.8423	1.2493	42.2085	3.0101	38.4862	2.0412	19.9787	3.0045	NaN
$c=20$	39.8423	1.2492	38.8362	2.0412	39.2417	3.0103	18.6429	2.0412	NaN
$c=50$	40.5934	2.1863	39.7846	2.1957	39.7341	2.9472	19.1584	2.6893	NaN



(a) Exact signal (b) $k=40$, SNR=1.7690

(c) $k=80$, SNR=4.5921 (d) $k=100$, SNR=5.0444

Fig. 4. (a) Exact signal. (b)–(d) The iterative solution x^* with different iteration number k at a fixed constant $c=2$ and $d=4$.

6. Conclusion

We analyzed the $\alpha\ell_1-\beta\ell_2$ $(\alpha\geq\beta\geq 0)$ sparsity regularization for nonlinear ill-posed problems. For the well-posedness of the regularization, compared to the case $\alpha>\beta\geq 0$, we only obtained the weak convergence for the case $\alpha=\beta>0$. If the nonlinear operator F' is Lipschitz continuous, we proved that the regularized solution is sparse. Two different convergence rates $O(\delta^{\frac{1}{2}})$ and $O(\delta)$ were obtained under two widely adopted nonlinear conditions. A soft thresholding algorithm ST-$(\alpha\ell_1-\beta\ell_2)$ can be extended to solving the non-convex $\alpha\ell_1-\beta\ell_2$ $(\alpha\geq\beta\geq 0)$ sparsity regularization for nonlinear ill-posed problems. Numerical experiments show that the proposed method is convergent and stable. However, for some particular nonlinear CS problems, we can only identify the positive impulses.

Data availability

No data was used for the research described in the article.

References

[1] L. Ding, W. Han, $\alpha\ell_1-\beta\ell_2$ Regularization for sparse recovery, Inverse Problems 35 (2019) 125009.
[2] P. Li, W. Chen, H. Ge, M.K. Ng, ℓ_1- $\alpha\ell_2$ Minimization methods for signal and image reconstruction with impulsive noise removal, Inverse Problems 36 (2020) 055009.

[3] Y. Lou, M. Yan, Fast L1-l2 minimization via a proximal operator, J. Sci. Comput. 74 (2018) 767–785.
[4] L. Yan, Y. Shin, D. Xiu, Sparse approximation using ℓ_1-ℓ_2 minimization and its application to stochastic collocation, SIAM J. Sci. Comput. 39 (2017) A229–254.
[5] P. Yin, Y. Lou, Q. He, J. Xin, Minimization of ℓ_{1-2} for compressed sensing, SIAM J. Sci. Comput. 37 (2015) A536–A563.
[6] D.A. Lorenz, E. Resmerita, Flexible sparse regularization, Inverse Problems 33 (2017) 014002.
[7] Fornasier M (Ed.), Theoretical Foundations and Numerical Methods for Sparse Recovery, De Gruyter, 2010.
[8] O. Scherzer, M. Grasmair, H. Grossauer, M. Haltmeier, F. Lenzen, Variational Methods in Imaging, Applied Mathematical Sciences, Springer, New York, 2009.
[9] M. Benning, M. Burger, Modern regularization methods for inverse problems, Acta Numer. (2018) 1–111.
[10] I. Daubechies, M. Defrise, C. De Mol, Sparsity-enforcing regularisation and ISTA revisited, Inverse Problems 32 (2016) 104001.
[11] B. Jin, P. Maass, Sparsity regularization for parameter identification problems, Inverse Problems 28 (2012) 123001.
[12] B. Jin, P. Maass, O. Scherzer, Sparsity regularization in inverse problems, Inverse Problems 33 (2017) 060301.
[13] I. Daubechies, M. Defrise, C. De Mol, An iterative thresholding algorithm for linear inverse problems with a sparsity constraint, Comm. Pure Appl. Math. 57 (2004) 1413–1457.
[14] Q. Jin, Landweber-kaczmarz method in Banach spaces with inexact inner solvers, Inverse Problems 32 (2016) 104005.
[15] D.A. Lorenz, P. Maass, P.Q. Muoi, Gradient descent for Tikhonov functionals with sparsity constraints: theory and numerical comparison of step size rules, Electron. Trans. Numer. Anal. 39 (2012) 437–463.
[16] R. Ramlau, G. Teschke, Tikhonov replacement functionals for iteratively solving nonlinear operator equations, Inverse Problems 21 (2005) 1571–1592.
[17] R. Ramlau, G. Teschke, A Tikhonov-based projection iteration for nonlinear ill-posed problems with sparsity constraints, Numer. Math. 104 (2006) 177–203.
[18] G. Teschke, C. Borries, Accelerated projected steepest descent method for nonlinear inverse problems with sparsity constraints, Inverse Problems 26 (2010) 025007.
[19] M. Zhong, W. Wang, A global minimization algorithm for Tikhonov functionals with p-convex ($p \geq 2$) penalty terms in Banach spaces, Inverse Problems 32 (2016) 104008.
[20] M. Grasmair, Well-posedness and convergence rates for sparse regularization with sublinear ℓ^q penalty term, Inverse Probl. Imaging 3 (2009) 383–387.
[21] W. Wang, S. Lu, B. Hofmann, 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.
[22] W. Wang, S. Lu, H. Mao, J. Cheng, Multi-parameter Tikhonov regularization with the ℓ_0 sparsity constrain, Inverse Problems 29 (2013) 065018.
[23] C.A. Zarzer, On Tikhonov regularization with non-convex sparsity constraints, Inverse Problems 25 (2009) 025006.
[24] Y. Wang, W. Yin, J. Zeng, Global convergence of ADMM in nonconvex nonsmooth optimization, J. Sci. Comput. 78 (2019) 29–63.
[25] Z. Xu, H. Zhang, Y. Wang, X. Chang, Y. Liang, $L_{1/2}$ Regularization, Sci. China Inf. Sci. 53 (2010) 1159–1169.
[26] K. Ito, K. Kunisch, A variational approach to sparsity optimization based on Lagrange multiplier theory, Inverse Problems 30 (2014) 015001.
[27] T. Blumensath, M.E. Davies, Iterative hard thresholding for compressed sensing, Appl. Comput. Harmon. Anal. 27 (2009) 265–274.
[28] R. Ramlau, C.A. Zarzer, On the minimization of a Tikhonov functional with a non-convex sparsity constraint, Electron. Trans. Numer. Anal. 39 (2012) 476–507.
[29] K. Bredies, D. Lorenz, S. Reiterer, Minimization of non-smooth, non-convex functionals by iterative thresholding, J. Optim. Theory Appl. 165 (2015) 78–112.
[30] A. Beck, M. Teboulle, A fast iterative shrinkage-thresholding algorithm for linear inverse problems, SIAM J. Imaging Sci. 2 (2009) 183–202.
[31] T. Bonesky, K. Bredies, D.A. Lorenz, P. Maass, A generalized conditional gradient method for nonlinear operator equations with sparsity constraints, Inverse Problems 23 (2007) 2041–2058.
[32] K. Bredies, D.A. Lorenz, P. Maass, A generalized conditional gradient method and its connection to an iterative shrinkage method, Comput. Optim. Appl. 42 (2009) 173–193.
[33] J. Appell, P.P. Zabrejko, Some elementary examples in nonlinear operator theory, Math. Comput. Modelling 32 (2000) 1367–1376.
[34] K. Bredies, D.A. Lorenz, Regularization with non-convex separable constraints, Inverse Problems 25 (2009) 085011.
[35] M. Grasmair, M. Haltmeier, O. Scherzer, Sparse regularization with ℓ^q penalty term, Inverse Problems 24 (2008) 055020.
[36] B. Hofmann, B. Kaltenbacher, C. Pöschl, O. Scherzer, A convergence rates result for Tikhonov regularization in Banach spaces with non-smooth operators, Inverse Problems 23 (2007) 987–1010.
[37] T. Scherzer, B. Kaltenbacher, B. Hofmann, K. Kazimierski, Regularization Methods in Banach Spaces, De Gruyter, 2012.
[38] H.W. Engl, M. Hanke, A. Neubauer, Regularization of Inverse Problems, Mathematics and Its Applications, vol. 375, Kluwer, Dordrecht, 1996.
[39] B. Kaltenbacher, A. Neubauer, O. Scherzer, Iterative Regularization Methods for Nonlinear Ill-Posed Problems, De Gruyter, 2012.
[40] P.L. Combettes, V.R. Wajs, Signal recovery by proximal forward–backward splitting, Multiscale Model. Simul. 4 (2005) 1168–1200.
[41] A. Beck, Y.C. Eldar, Sparsity constrained nonlinear optimization: optimality conditions and algorithms, SIAM J. Optim. 23 (2013) 1480–1509.
[42] T. Blumensath, Compressed sensing with nonlinear observations and related nonlinear optimization problems, IEEE Trans. Inform. Theory 59 (2013) 3466–3474.
[43] S.P. Chepuri, G. Leus, Sparsity-promoting sensor selection for non-linear measurement models, IEEE Trans. Signal Process. 63 (2015) 684–698.
[44] S. Yang, M. Wang, P. Li, L. Jin, B. Wu, L. Jiao, Compressive hyperspectral imaging via sparse tensor and nonlinear compressed sensing, IEEE Trans. Geosci. Remote Sens. 53 (2015) 5943–5957.