ORIGINAL PAPER

Proximal algorithm for minimization problems in l_0-regularization for nonlinear inverse problems

Pham Quy Muoi[1] (ORCID) · Duong Xuan Hiep[1]

Abstract

In this paper, we study a proximal method for the minimization problem arising from l_0-regularization for nonlinear inverse problems. First of all, we prove the existence of solutions and give an optimality condition for solutions to the minimization problem. Then, we propose and prove the convergence of the proximal method for this minimization problem, which is controlled by step size conditions. Finally, we illustrate performance of the proximal method by applying it to a numerical example.

Keywords Proximal method · l_0-regularization · Nonlinear inverse problems · Nonsmooth optimization

1 Introduction

In this paper, H is denoted to be a real separable Hilbert space, $\langle \cdot, \cdot \rangle$ and $\| \cdot \|$ are respectively the scalar product and the norm in H. We assume that $\{\varphi_i\}_{i \in \Lambda}$ ($\Lambda \subset \mathbb{N}$) is an orthonormal basis of the Hilbert space H. For each $u \in H$, we denote $u = (u_i)_{i \in \Lambda}$, where $u_i = \langle u, \varphi_i \rangle$ is the ith component of u with respect to the basis $\{\varphi_i\}_{i \in \Lambda}$. Note that all results from this paper are true not only for an orthonormal basis but also for a frame in the real separable Hilbert space H. However, for simplicity, we assume that $\{\varphi_i\}_{i \in \Lambda}$ ($\Lambda \subset \mathbb{N}$) is an orthonormal basis of the Hilbert space H. We study the proximal method for the minimization problem

$$\min_{u \in H} \Theta(u) := F(u) + \alpha \Phi(u), \tag{1}$$

✉ Pham Quy Muoi
  phamquymuoi@gmail.com

  Duong Xuan Hiep
  dxhiep1994@gmail.com

[1] Department of Mathematics, The University of Danang - University of Science and Education, 459 Ton Duc Thang, Danang, Vietnam

where $F : H \to \mathbb{R}$ is a Fréchet differentiable functional and $\Phi : H \to \bar{\mathbb{R}} := \mathbb{R} \cup \{+\infty\}$ is defined by

$$\Phi(u) := \|u\|_0 = \sum_{i \in \Lambda} |u_i|_0, \tag{2}$$

with

$$|x|_0 = \begin{cases} 1, & \text{if } x \neq 0 \\ 0, & \text{if } x = 0. \end{cases}$$

Problem (1) appears in l_0-regularization of inverse problems, i.e., we want to find a solution u of the operator equation

$$K(u) = y, \tag{3}$$

where $K : H \to H$ is a Fréchet differentiable operator (linear or nonlinear), y is the exact data, which is unknown and we only obtain a noisy data y^δ of y with

$$\|y - y^\delta\|_H \leq \delta. \tag{4}$$

Further, we assume that problem (3) is ill-posed, which the solutions to the problem do not depend continuously on the data [5]. Therefore, the real problem is finding an approximation of the solution given the operator K and a noisy data y^δ. Applying l_0-regularization for this problem the regularized approximate solutions to (3)–(4) are seen to be minimizers of the minimization problem

$$\min_{u \in H} \Theta(u) := F(u; y^\delta) + \alpha\Phi(u), \tag{5}$$

where $F(\cdot; y^\delta) : H \to \mathbb{R}$ measuring the difference between $K(u)$ and y^δ is a Fréchet differentiable functional. Note that problem (5) is an example of problem (1).

The advantage of l_0-regularization is preserving the edges between homogeneous regions and nonhomogeneous. Thus, l_0-regularization is used in many applications such as signal and image analysis, wavelet frame, compressive sensing, dictionary building, machine learning and actuator location problems. For linear inverse problems, Blumensath and Davies in [2, 3] have proposed an iterative thresholding scheme and proven the local convergence of this scheme. Needell and Tropp have presented the compressive sampling matching pursuit (CoSaMP). Foucart [7] has also improved CoSaMP by introducing hard thresholding pursuit based on a combination of iterative hard thresholding and compressive sampling matching pursuit algorithm. The other method is suggested by Lu et al. in [9]. Nikolova and E. Soubies have studied in detail about global solutions, see [13, 14]. Recently, Wachsmuth has studied the hard-iterative thresholding to optimal control problems with $L^0(\Omega)$ control cost in space $L^2(\Omega)$ in [15]. But for nonlinear inverse problems, there are few algorithms proposed to solve problem (1) in a real separable Hilbert space.

In this paper, we will propose a proximal method for problem (5) in a real separable Hilbert space based on the ideas of the proximal method in [8, 11]. Note that since the functional Φ is nonconvex, problem (1) is different from the minimization problem arising from l_1-regularization in [8, 11]. Therefore, the results and the methods of proof for the convergence of the proximal method in our paper are different from those in [8, 11]. Furthermore, the *iterative thresholding* [2] or the *iterative hard*

thresholding [3] are special cases of our proposed method since they are only applied to linear inverse problem, i.e., to problem (5) with $F(u; y^\delta) = \|Ku - y^\delta\|^2$ and K is a linear operator. Thus, the results in this paper are extensions of those in [2, 3] to nonlinear inverse problems.

The main results in our paper are

- The existence and the optimality condition for solutions to problem (1).
- The convergence of the proximal method for problem (1) and the practical choice for step sizes.

The paper is organized as follows: In Section 2 we give explicit formula for the unique minimizer of the quadratic approximation of the cost functional. In Section 3, we study the existence and the optimality condition for solutions to problem (1). Section 4 is devoted to prove the convergence of the proximal method. A method of choosing step sizes is given in Section 5. Finally, we illustrate the performance of the proximal method by applying it to an identification problem in elliptic equations in Section 6.

2 First-order condition for global minimizers for quadratic approximation of the cost functional

The proximal methods developed in [2–4, 11] are based on quadratic approximate functionals and the proximal operators. In order to present the proximal method as well as the first-order optimality condition of problem (1), we need to introduce the proximal operator for l_0-regularization and the quadratic approximate functionals.

The *proximal operator* (also known as the *hard-thresholding operator*) is defined by

$$\mathbb{P}_\lambda(u) = (P_\lambda(u_1), P_\lambda(u_2), \ldots, P_\lambda(u_n), \ldots), u \in H, \qquad (6)$$

where

$$P_\lambda(x) = \begin{cases} 0 & \text{if } |x| < \sqrt{\lambda} \\ \{0, x\} & \text{if } |x| = \sqrt{\lambda} \\ x & \text{if } |x| > \sqrt{\lambda}, \end{cases} \quad (x \in \mathbb{R}). \qquad (7)$$

Note that $\mathbb{P} : H \rightrightarrows H$ is a multi-valued mapping.

For each fixed $s > 0$, the so called *quadratic approximate functional* of $\Theta(v)$ at a given point u is defined as follows:

$$\Theta_s(v, u) := F(u) + \langle \nabla F(u), v - u \rangle + \tfrac{s}{2}\|v - u\|^2 + \alpha\Phi(v). \qquad (8)$$

Using the proximal operator, we can show that the minimization problem, $\min_{v \in H} \Theta_s(v, u)$, has at least one solution. This result is given in the following theorem.

Theorem 2.1 *Let $F \in C^1(H)$. Then, for any $u \in H$ fixed, the problem*

$$\min_{v \in H} \Theta_s(v, u)$$

has at least one solution and u^+ is a solution if and only if

$$u^+ \in \mathbb{P}_{\frac{2\alpha}{s}} \left(u - \frac{1}{s} \nabla F(u) \right).$$

Proof By simple calculation, which ignores constant term in u, the minimization problem, $\min_{v \in H} \Theta_s(v, u)$, is equivalent to

$$\min_{v \in H} \left\{ \frac{1}{2} \left\| v - \left(u - \frac{1}{s} \nabla F(u) \right) \right\|^2 + \frac{\alpha}{s} \Phi(v) \right\}.$$

This problem can be rewritten as

$$\min_{v \in H} \left\{ \sum_{i \in \Lambda} \frac{1}{2} \left(v_i - \left(u_i - \frac{1}{s} (\nabla F(u))_i \right) \right)^2 + \frac{\alpha}{s} |v_i|_0 \right\}. \qquad (9)$$

Therefore, the minimum of the problem, $\min_{v \in H} \Theta_s(v, u)$, can be calculated by minimizing with respect to v_i individually. For each i we distinguish two cases, $v_i = 0$ and $v_i \neq 0$. In the first case, by the definition of $|.|_0$, we have

$$\frac{1}{2} \left(v_i - \left(u_i - \frac{1}{s} (\nabla F(u))_i \right) \right)^2 + \frac{\alpha}{s} |v_i|_0 = \frac{1}{2} \left(u_i - \frac{1}{s} (\nabla F(u))_i \right)^2.$$

In the second case, we have

$$\frac{1}{2} \left(v_i - \left(u_i - \frac{1}{s} (\nabla F(u))_i \right) \right)^2 + \frac{\alpha}{s} |v_i|_0 \geq \frac{\alpha}{s},$$

where the equality sign is obtained when $v_i = u_i - \frac{1}{s} (\nabla F(u))_i$.

Comparing the values for both cases, the minimum of component i of problem (9) is

$$\min \left(\frac{1}{2} \left(u_i - \frac{1}{s} (\nabla F(u))_i \right)^2, \frac{\alpha}{s} \right)$$

that is equal to

$$\begin{cases} \frac{1}{2} \left(u_i - \frac{1}{s} (\nabla F(u))_i \right)^2 & , \text{ if } |u_i - \frac{1}{s} (\nabla F(u))_i| < \sqrt{\frac{2\alpha}{s}}, \\ \frac{1}{2} \left(u_i - \frac{1}{s} (\nabla F(u))_i \right)^2 = \frac{\alpha}{s} & , \text{ if } |u_i - \frac{1}{s} (\nabla F(u))_i| = \sqrt{\frac{2\alpha}{s}}, \\ \frac{\alpha}{s} & , \text{ if } |u_i - \frac{1}{s} (\nabla F(u))_i| > \sqrt{\frac{2\alpha}{s}}. \end{cases}$$

Note that the first value is obtained when $v_i = 0$, the second one is obtained when $v_i = 0$ or $v_i = u_i - \frac{1}{s} (\nabla F(u))_i$, and the last value is obtained when $v_i = u_i - \frac{1}{s} (\nabla F(u))_i \neq 0$, i.e., the minimum of component i is attained at

$$v_i \in P_{\frac{2\alpha}{s}} \left(u_i - \frac{1}{s} (\nabla F(u))_i \right).$$

Therefore, the problem, $\min_{v \in H} \Theta_s(v, u)$, has a solution

$$u^+ = (v_1, v_2, \ldots, v_n, \ldots),$$

or

$$u^+ \in \mathbb{P}_{\frac{2\alpha}{s}} \left(u - \frac{1}{s} \nabla F(u) \right).$$

$\square$

3 Existence of optimal solutions and optimality conditions

Before presenting the results of the existence of optimal solutions and optimality conditions to problem (1), we recall some basic notions which are not often used. The other basic notions such as convexity, Lipschitz continuity, Lipschitz Fréchet differentiability are well-known in calculus or convex analysis, so we do not recall them here.

Definition 3.1 Let a functional $G : H \to \mathbb{R}$.

1) G is called bounded from below if there exists a constant C such that $G(u) \geq C$ for all $u \in H$.
2) G is called coercive if $G(u) \to \infty$ as $\|u\| \to \infty$.
3) G is called lower semi-continuous if for every $u \in H$ and every sequence $\{u^k\}, u^k \to u$ as $k \to \infty$, we have

$$G(u) \leq \liminf_{k \to \infty} G(u^k).$$

In our paper, the functionals, F and Φ, in problem (1) have some properties of G. In fact, F is assumed to be bounded from below, coercive and lower semi-continuous (see Assumption 1) and Φ is lower semi-continuous (see Lemma 3.1). In order to prove the existence of solutions and give the optimality conditions to problem (1) as well as the convergence of the proximal method, we need the following assumption.

Assumption 1 *Assume that*

a)   *F is coercive, bounded from below and lower semi-continuous.*
b)   *F is Lipschitz Fréchet differentiable with the Lipschitz constant L, i.e.,*

$$\|\nabla F(u) - \nabla F(v)\| \leq L\|u - v\|, \forall u, v \in H.$$

Note that In Assumption 1, Condition a) is used to make sure the existence of a solution to problem (1), and Condition b) is used to obtain the optimality condition of solutions to problem (1) as well as the convergence of the proximal method for this problem. In fact, by Condition b) we have for $s \geqslant L$

$$|F(v) - F(u) - \langle \nabla F(u), v - u \rangle| \leqslant \frac{s}{2}\|v - u\|^2.$$

This is the key inequality used to prove the main results of the paper.

Lemma 3.1 *The functional Φ defined by (2) is non-negative, nonconvex, and lower semi-continuous.*

Proof The first two properties are clear. The last property is followed by the lower semi-continuity of $|x|_0$. $\square$

Note that functional Φ in l_0-regularization (given by (2)) is nonconvex and not coercive. Thus, problem (1) is nonconvex as well. To make sure the existence of a solution to problem (1) we need some properties of the functional F. Such sufficient conditions for F are given in the following lemma.

Lemma 3.2 *Assume that F satisfies Condition a) of Assumption 1. Then, problem* (1) *has at least a solution (global minimizer).*

Proof The proof is standard and it is left to the readers. $\square$

Theorem 3.1 *Assume that F satisfies Condition b) of Assumption 1. Then, u^* is a solution (global minimizer) of problem* (1) *if and only if*

$$u^* \in \mathbb{P}_{\frac{2\alpha}{s^*}}\left(u^* - \frac{1}{s^*}\nabla F(u^*)\right), \tag{10}$$

for any fixed $s^* \geq L$.

Proof Since u^* is a global minimizer of Θ, we have $\Theta(u^*) \leq \Theta(u^* + h)$ for any $h \in H$. For $s^* \geq L$, we have

$$\Theta_{s*}(u^* + h, u^*) - \Theta_{s*}(u^*, u^*) \geq \Theta_{s*}(u^* + h, u^*) - \Theta(u^* + h)$$
$$= F(u^*) - F(u^* + h) + \langle \nabla F(u^*), h \rangle + \frac{s^*}{2}\|h\|^2 \geq 0.$$

Therefore, $\Theta_{s*}(u^* + h, u^*) \geq \Theta_{s*}(u^*, u^*)$ for any $h \in H$. It means that u^* is the global minimizer of $\Theta_{s*}(\cdot, u^*)$. Then, the proof deduces from Theorem 2.1. $\square$

Definition 3.2 Any point u^* satisfying (10) with $s^* \geq L$ is called a *stationary point* of Θ.

Remark 3.1 The condition $s^* \geq L$ is necessary to obtain the last inequality in the proof of Theorem 3.1. For $s^* < L$ we could not prove such a result. This is different from l_1-regularization, which we can prove a similar result for any $s^* > 0$, see, e.g., in [1, 10]. Furthermore, Condition (10) is only true for global minimizers of problem (1). We could not prove such a result for local minimizers. This is also the other difference when it is compared with l_1-regularization, see, e.g., [10].

4 Proximal method

The main idea of the proximal method is based on the quadratic approximation method. In fact, at the kth iteration, u^{k+1} is determined by solving the quadratic

approximation of Θ at u^k, i.e.,

$$u^{k+1} \in \mathrm{argmin}_{v \in H} \Theta_{s^k}(v, u^k)$$

and then we need to propose some conditions for $s^k > 0$ such that $\{u^k\}$ (with a subsequence) converges to a stationary point of the original problem. Here, the functional $\Theta_s(v, u)$ is defined by (8).

By Theorem 2.1, the sequence $\{u^{k+1} \in \mathrm{argmin}_{v \in H} \Theta_{s^k}(v, u^k)\}$ is later given by

$$u^{k+1} \in \mathbb{P}_{\frac{2\alpha}{s^k}}\left(u^k - \frac{1}{s^k}\nabla F(u^k)\right). \tag{11}$$

Now we prove the convergence of this sequence. We separate the result into several following lemmas and theorems.

Lemma 4.1 *Let Assumption 1 hold and $\{u^k\}$ is a sequence defined by (11), where the sequence $\{s^k\}$ satisfies $s^k \geq L + \epsilon$ with $\epsilon > 0$ for all $k \in \mathbb{N}$. Then, the sequence $\{\Theta(u^k)\}$ is monotonically decreasing, the sequence $\{u^k\}$ is bounded and $\lim_{k \to \infty} \|u^{k+1} - u^k\| = 0$.*

Proof Since F is Lipschitz differentiable with Lipschitz constant L. Therefore, for $s \geq L$ we have

$$|F(v) - F(u) - \langle \nabla F(u), v - u \rangle| \leq \frac{s}{2}\|v - u\|^2.$$

Hence, with $s \geq L$ we obtain

$$\Theta(v) = F(v) + \Phi(v) \leq F(u) + \langle \nabla F(u), v - u \rangle + \frac{s}{2}\|v - u\|^2 + \Phi(v) = \Theta_s(v, u).$$

Consequently, we have (since $s^k > L$)

$$\Theta(u^{k+1}) \leq \Theta_{s^k}(u^{k+1}, u^k).$$

In addition, since u^{k+1} is the solution to $\Theta_{s^k}(\cdot, u^k)$ (see (11)), we have

$$\Theta(u^{k+1}) \leq \Theta_{s^k}(u^{k+1}, u^k) \leq \Theta_{s^k}(u^k, u^k) = \Theta(u^k).$$

Thus, the sequence $\{\Theta(u^k)\}$ is monotonically decreasing.

The boundedness of $\{u^k\}$ is a consequence of the decreasing of $\{\Theta(u^k)\}$, the boundedness from below and the coercivity of Θ.

For each k, since L is the Lipschitz constant of ∇F and $s^k > L$, we have

$$\begin{aligned}
\Theta(u^{k+1}) &= F(u^{k+1}) + \alpha \Phi(u^{k+1}) \\
&\leq F(u^k) + \langle \nabla F(u^k), u^{k+1} - u^k \rangle + \frac{L}{2}\|u^{k+1} - u^k\|^2 + \alpha \Phi(u^{k+1}) \\
&\leq F(u^k) + \langle \nabla F(u^k), u^{k+1} - u^k \rangle + \frac{s^k}{2}\|u^{k+1} - u^k\|^2 + \alpha \Phi(u^{k+1}) \\
&= \min_{u \in H} \Theta_{s^k}(u, u^k) \\
&\leq F(u^k) + \alpha \Phi(u^k) = \Theta(u^k).
\end{aligned}$$

This implies that

$$\Theta(u^k) - \Theta(u^{k+1}) \geq \frac{s^k - L}{2} \|u^{k+1} - u^k\|^2 \geq \epsilon \|u^{k+1} - u^k\|^2$$

or

$$\frac{1}{\epsilon} \left(\Theta(u^k) - \Theta(u^{k+1}) \right) \geq \|u^{k+1} - u^k\|^2.$$

Summing the above inequalities over $k = 0, 1, ..., m$ yields

$$\frac{1}{\epsilon} \left(\Theta(u^0) - \Theta(u^{m+1}) \right) \geq \sum_{k=0}^{m} \|u^{k+1} - u^k\|^2, \qquad \forall m.$$

One infers that the series $\sum_{k=0}^{\infty} \|u^{k+1} - u^k\|^2$ converges. Therefore, $\lim_{k \to \infty} \|u^{k+1} - u^k\| = 0$. □

Theorem 4.1 (**Convergence**) *Let F satisfy Assumption 1 and $\{u^k\}$ is a sequence defined by (11), where $s^k \in [L + \epsilon, \overline{s}]$, $\forall k$, $\epsilon > 0$ and $\overline{s} \geq L + \epsilon$. Then, there exists a subsequence $\{u^{k_j}\}$ of $\{u^k\}$ weakly converging to some $u^* \in H$ as $j \to \infty$ and u^* is a stationary point of Θ.*

Proof By Lemma 4.1, the sequence $\{\Theta(u^k)\}$ is decreasing and the sequence $\{u^k\}$ is bounded. Since ∇F is Lipschitz continuous, the sequence $\{\nabla F(u^k)\}_k$ is bounded as well. Therefore, there exists a subsequence $\{u^{k_j}\}$ of $\{u^k\}$ such that $\{u^{k_j}\}$ converges weakly to some $u^* \in H$, $\{\nabla F(u^{k_j})\}$ converges weakly to $\nabla F(u^*) \in H$, and $\{s^{k_j}\}$ converges to $s^* \in [L + \epsilon, \overline{s}]$ as $j \to \infty$. By Lemma 4.1 $\{u^{k_j+1}\}$ also converges weakly to u^*. We define

$$\Gamma_0 = \left\{ i \, \big| \, |u_i^* - \frac{1}{s^*}(\nabla F(u^*))_i| < \sqrt{\frac{2\alpha}{s^*}} \right\},$$

$$\Gamma_1 = \left\{ i \, \big| \, |u_i^* - \frac{1}{s^*}(\nabla F(u^*))_i| > \sqrt{\frac{2\alpha}{s^*}} \right\}, \Gamma = \Lambda \backslash (\Gamma_0 \cup \Gamma_1)$$

and

$$\Gamma_0^j = \left\{ i \, \big| \, |u_i^{k_j} - \frac{1}{s^{k_j}}(\nabla F(u^{k_j}))_i| < \sqrt{\frac{2\alpha}{s^{k_j}}} \right\},$$

$$\Gamma_1^j = \left\{ i \, \big| \, |u_i^{k_j} - \frac{1}{s^{k_j}}(\nabla F(u^{k_j}))_i| > \sqrt{\frac{2\alpha}{s^{k_j}}} \right\}, \Gamma^j = \Lambda \backslash (\Gamma_0^j \cup \Gamma_1^j).$$

We consider three cases:

Case 1 (for each $i \in \Gamma_0$): Since $\{u^{k_j} - \frac{1}{s^{k_j}} \nabla F(u^{k_j})\}$ converges to $u^* - \frac{1}{s^*} \nabla F(u^*)$ and $\{s^{k_j}\}$ converges to s^*, there exists $j^* \in \mathbb{N}$ such that for all $j \geq j^*$ we have $i \in \Gamma_0^j$. Therefore, if $j \geq j^*$, then

$$u_i^{k_j+1} = P_{\frac{2\alpha}{s^{k_j}}} \left(u_i^{k_j} - \frac{1}{s^{k_j}}(\nabla F(u^{k_j}))_i \right) = 0.$$

Let $j \to \infty$, we have
$$u_i^* = 0.$$

This implies that
$$0 = u_i^* = P_{\frac{2\alpha}{s^*}}\left(u_i^* - \frac{1}{s^*}(\nabla F(u^*))_i\right).$$

Case 2 (for each $i \in \Gamma_1$): Since $\{u^{k_j} - \frac{1}{s^{k_j}}\nabla F(u^{k_j})\}$ converges to $u^* - \frac{1}{s^*}\nabla F(u^*)$ and $\{s^{k_j}\}$ converges to s^*, there exists $j^* \in \mathbb{N}$ such that for all $j \geq j^*$ we have $i \in \Gamma_1^j$. Therefore, if $j \geq j^*$, then
$$u_i^{k_j+1} = P_{\frac{2\alpha}{s^{k_j}}}\left(u_i^{k_j} - \frac{1}{s^{k_j}}(\nabla F(u^{k_j}))_i\right) = u_i^{k_j} - \frac{1}{s^{k_j}}(\nabla F(u^{k_j}))_i.$$

Let $j \to \infty$, we have
$$u_i^* = u_i^* - \frac{1}{s^*}(\nabla F(u^*))_i.$$

This implies that
$$(\nabla F(u^*))_i = 0, |u_i^*| > \sqrt{\frac{2\alpha}{s^*}} \text{ and } u_i^* = P_{\frac{2\alpha}{s^*}}\left(u_i^* - \frac{1}{s^*}(\nabla F(u^*))_i\right).$$

Case 3 (for each $i \in \Gamma$): We have
$$u_i^{k_j+1} \in P_{\frac{2\alpha}{s^{k_j}}}\left(u_i^{k_j} - \frac{1}{s^{k_j}}(\nabla F(u^{k_j}))_i\right) \subset \left\{0, u_i^{k_j} - \frac{1}{s^{k_j}}(\nabla F(u^{k_j}))_i\right\}.$$

Let $j \to \infty$, we have
$$u_i^* \in \left\{0, u_i^* - \frac{1}{s^*}(\nabla F(u^*))_i\right\}.$$

Thus
$$u_i^* \in P_{\frac{2\alpha}{s^*}}\left(u_i^* - \frac{1}{s^*}(\nabla F(u^*))_i\right).$$

Combining three cases, we have
$$u^* \in \mathbb{P}_{\frac{2\alpha}{s^*}}\left(u^* - \frac{1}{s^*}\nabla F(u^*)\right).$$

$\square$

5 Step size choice and realization of the proximal method

From Lemma 4.1 the condition for the convergence of proximal method is that parameters s^k must satisfy $s^k \geq L + \epsilon$ for some fixed $\epsilon > 0$. In practice, there are two following cases:

(i)   In the first case, the Lipschitz constant L is known.
(ii) In the second one, the Lipschitz constant L is unknown or very hard to find.

To make sure the convergence of the proximal method: in the first case we can choice $s^k = L + \epsilon$ with $\epsilon > 0$ for all k. However, the bigger the Lipschitz constant is,

the smaller of step size is. As a result, the algorithm converges slower. Thus, we do not want to choose s^k too large for all k. To overcome this situation and for the second case, we propose a practical way to select step size as follow: we can choose s^0 that is not too large and choose the parameter s^k that satisfies the condition $\Theta(u^{k+1}) \leq \Theta_{s^k}(u^{k+1}, u^k)$ by the back-tracking rule. By this method, the sequence $\{\Theta(u^k)\}$ is monotonically decreasing (see the proof of Lemma 4.1). With this choice rule, there are two cases: In the first case we have $s^k \leq L$ for all k and thus the conditions in Lemma 4.1 are not satisfied, i.e., the proximal method may not converge. In the second case there exists a k^* such that $s^{k^*} > L, \forall k \geq k^*$, i.e., the proximal method converges. In practice, by the backtracking rule the first case rarely happen, especially when the number of iterations is large since the sequence s^k is increasing. Thus, the above method of choosing step sizes is practical.

Since the proximal operator is a multi-valued operator, there are many sequences $\{u^k\}$ generated by the proximal method. In practice we only need one of them. Thus, we use one of two modified proximal operators as follows:

$$\mathbb{P}^l_\lambda(u) = \left(P^l_\lambda(u_1), P^l_\lambda(u_2), \ldots, P^l_\lambda(u_n), \ldots \right), u \in H, l \in \{1, 2\} \qquad (12)$$

where

$$P^1_\lambda(x) = \begin{cases} 0 & \text{if } |x| \leq \sqrt{\lambda} \\ x & \text{if } |x| > \sqrt{\lambda} \end{cases}, \quad P^2_\lambda(x) = \begin{cases} 0 & \text{if } |x| < \sqrt{\lambda} \\ x & \text{if } |x| \geq \sqrt{\lambda} \end{cases} \quad (x \in \mathbb{R}). \qquad (13)$$

Combining the choice of step sizes and the modified proximal operator, the iteration (8), is presented in Algorithm 1. We will illustrate Algorithm 1 with different modified proximal operators in Section 6.

Algorithm 1 Proximal algorithm.

Input: Initial guess $u^0 : \Phi(u^0) < \infty, q \in (1, \infty)$ and $s^0 \in [\underline{s}, \overline{s}]$ $(0 < \underline{s} \leq L \leq \overline{s})$
1: **for** $k = 0, 1, 2, \ldots$ **do**
2:     **repeat**
3:         $u^{k+1} \leftarrow \mathbb{P}^l_{\frac{2\alpha}{s^k}}(u^k - \frac{1}{s^k}\nabla F(u^k))$
4:         **if** $\Theta(u^{k+1}) > \Theta_{s^k}(u^{k+1}, u^k)$ **then**
5:             $s^k \leftarrow \min\{s^k.q, \overline{s}\}$
6:         **end if**
7:     **until** $\Theta(u^{k+1}) \leqslant \Theta_{s^k}(u^{k+1}, u^k)$ or $s^k \notin [\underline{s}, \overline{s}]$
8:     $s^{k+1} \leftarrow s^k$
9: **end for**
Output: $\{u^k\}$

---

6 Application to an inverse source problem

For a numerical example we consider the inverse problem of estimating the source $y \in L^2(\Omega)$ in the elliptic boundary problem

$$- \operatorname{div}(a \nabla u) = y \text{ in } \Omega \subset \mathbb{R}^2, \ u = u_0 \text{ on } \partial\Omega \qquad (14)$$

from a noisy data u^δ of the exact solution.

We assume that Ω is a Lipschitz domain and $a \in \mathcal{A}$, where

$$\mathcal{A} = \{a \in L^\infty(\Omega) : 0 < \underline{a} < a < \overline{a}\}$$

Then, for each $y \in L^2(\Omega)$, $a \in \mathcal{A}$ and $u_0 \in H^{1/2}(\partial\Omega)$ problem (14) has a unique weak solution $u \in H^1$, see, e.g., [6].

Let y^* be the parameter that needs to recover and u^* be the solution to (14) with fixed parameters u_0, a and y^*. Let $K : L^2(\Omega) \to H^1(\Omega), y \mapsto u$, the solution operator of (14) with fixed a and u_0. The inverse problem is equivalent to solving the equation

$$K(y) = u^*, \qquad (15)$$

where only a noisy approximation u^δ of u^* is available with

$$\|u^* - u^\delta\|_{L^2(\Omega)} \le \delta.$$

Using least square approach together with l_0-regularization, the approximation to y^* is the solution to the following problem

$$\min_{y \in L^2(\Omega)} \Theta(y) := F(y) + \alpha \sum_{i=1}^{\infty} |y_i|_0, \qquad (16)$$

where

$$F(y) = \frac{1}{2}\|K(y) - u^\delta\|_{L^2(\Omega)}^2, \qquad (17)$$

$\alpha > 0$ is a regularization parameter and $y_i = \langle y, \varphi_i \rangle$ with $\{\varphi_i\}_{i=1}^{\infty}$ is an orthonormal basis of $L^2(\Omega)$.

The wellposedness of problem (16), the convergence and convergence rates of its solutions to solutions of operator (15) are still open (by our best knowledge), but it can be obtained by standard approaches in theory of regularization for inverse problems. For examples, Wang et al. have proved some such results relating to l_0-regularization for general operator equations in [16, 17]. Here, as an example we try to solve problem (16) by our proximal method. The numerical results will show the wellposedness of problem (16) and the convergence of its solutions to operator (15). Note that the proximal method can be applied for problem (16) since F satisfies Assumption 1, where the Lipschitz differentiability of F is given in the following lemma.

Lemma 6.1 *The functional $F : L^2(\Omega) \to \mathbb{R}$ is Fréchet differentiable at any point $y \in L^2(\Omega)$, Fréchet derivative F' is Lipschitz continuous and $F'(y) = p$, where p is*

the weak solution to the following problem

$$-div(a\nabla p) = u - u^\delta \text{ in } \Omega, \ p = 0 \text{ on } \partial\Omega, \tag{18}$$

where $u = K(y)$.

Proof We first recall the following well-known results:

(a) $H_0^1(\Omega)$ is embedding compactly into $L^2(\Omega)$. This implies that there exists a constant $c > 0$ such that

$$\|v\|_{L^2(\Omega)} \le c\|v\|_{H_0^1(\Omega)}, \ \forall v \in H_0^1(\Omega). \tag{19}$$

(b) In $H_0^1(\Omega)$ the norm $\|\|v\|\| := \sqrt{\int_\Omega a\nabla v.\nabla v dx}$ is equivalent to $\|v\|_{H_0^1(\Omega)}$, i.e., there exist constants $c_1, c_2 > 0$ such that

$$c_1\|v\|_{H_0^1(\Omega)} \le \|\|v\|\| \le c_2\|v\|_{H_0^1(\Omega)}, \ \forall v \in H_0^1(\Omega). \tag{20}$$

Using above results, we now prove the Fréchet differentiability of F at $y \in L^2(\Omega)$. Indeed, for $\delta y \in L^2(\Omega)$ we have

$$
\begin{aligned}
2(F(y + \delta y) - F(y)) &= \langle K(y + \delta y) - K(y), K(y + \delta y) + K(y) - 2u^\delta \rangle_{L^2(\Omega)} \\
&= 2\langle u^1 - u, u - u^\delta \rangle_{L^2(\Omega)} + \|u^1 - u\|_{L^2(\Omega)}^2 \\
&\overset{(19)}{\le} 2\langle u^1 - u, u - u^\delta \rangle_{L^2(\Omega)} + c\|u^1 - u\|_{H_0^1(\Omega)}^2 \\
&\overset{(20)}{\le} 2\langle u^1 - u, u - u^\delta \rangle_{L^2(\Omega)} + \tfrac{c}{c_1}\|\|u^1 - u\|\|^2,
\end{aligned} \tag{21}
$$

where $u = K(y), u^1 = K(y + \delta y)$. Furthermore, by the formula of weak solutions we have

$$\int_\Omega a\nabla u.\nabla v dx = \int_\Omega y v dx, \ \forall v \in H_0^1(\Omega), \tag{22}$$

$$\int_\Omega a\nabla u^1.\nabla v dx = \int_\Omega (y + \delta y) v dx, \ \forall v \in H_0^1(\Omega). \tag{23}$$

Subtracting both sides of equations (22) and (23) we have

$$\int_\Omega a\nabla(u^1 - u).\nabla v dx = \int_\Omega \delta y v dx, \ \forall v \in H_0^1(\Omega). \tag{24}$$

Inserting $v = u^1 - u$ and $v = p$ into (24) we obtain

$$\|\|u^1 - u\|\|^2 = \int_\Omega \delta y(u^1 - u)dx, \tag{25}$$

$$\int_\Omega a\nabla(u^1 - u).\nabla p dx = \int_\Omega \delta y p dx. \tag{26}$$

By the formula of weak solution to (18) we have

$$\int_\Omega a\nabla p.\nabla v dx = \int_\Omega (u - u^\delta)v dx, \ \forall v \in H_0^1(\Omega). \tag{27}$$

Inserting $v = u^1 - u \in H_0^1(\Omega)$ into (27) we have

$$\int_\Omega a\nabla p.\nabla(u^1 - u)dx = \int_\Omega (u - u^\delta)(u^1 - u)dx. \tag{28}$$

From (21), (25), (26) and (28) we have

$$
\begin{aligned}
2(F(y + \delta y) - F(y)) &\leq 2\int_\Omega a\nabla p.\nabla(u^1 - u)dx + \frac{c}{c_1}\int_\Omega \delta y(u^1 - u)dx \\
&= 2\int_\Omega \delta y p dx + \frac{c}{c_1}\int_\Omega \delta y(u^1 - u)dx \\
&\leq 2\int_\Omega \delta y p dx + \frac{c}{c_1}\|\delta y\|_{L^2(\Omega)}\|u^1 - u\|_{L^2(\Omega)} \\
&= 2\int_\Omega \delta y p dx + 0(\|\delta y\|_{L^2(\Omega)}) \tag{29}
\end{aligned}
$$

since $\|u^1 - u\|_{L^2(\Omega)} \to 0$ as $\delta y \to 0$ (due to the continuity of the operator K). The inequality (29) implies that F is Fréchet differentiable and $F'(y) = p$, where p is the weak solution to (18).

To finish the proof, we need to prove the Lipschitz continuity of F'. Indeed, let $p_1 = F'(y_1)$ and $p_2 = F'(y_2)$. Then, p_1, $p_2 \in H_0^1(\Omega)$ and they respectively satisfy

$$\int_\Omega a\nabla p_1.\nabla v dx = \int_\Omega (u_1 - u^\delta)v dx, \ \forall v \in H_0^1(\Omega) \tag{30}$$

and

$$\int_\Omega a\nabla p_2.\nabla v dx = \int_\Omega (u_2 - u^\delta)v dx, \ \forall v \in H_0^1(\Omega), \tag{31}$$

where $u_1 = K(y_1)$ and $u_2 = K(y_2)$. Subtracting both sides of (30) and (31) we have

$$\int_\Omega a\nabla(p_1 - p_2).\nabla v dx = \int_\Omega (u_1 - u_2)v dx, \ \forall v \in H_0^1(\Omega). \tag{32}$$

Inserting $v = p_1 - p_2$ into (32) we obtain

$$\int_\Omega a\nabla(p_1 - p_2).\nabla(p_1 - p_2)dx = \int_\Omega (u_1 - u_2)(p_1 - p_2)dx. \tag{33}$$

Similarly, subtracting both sides of the formula of the weak solutions u_1 and u_2 we have

$$\int_\Omega a\nabla(u_1 - u_2).\nabla v dx = \int_\Omega (y_1 - y_2)v dx, \ \forall v \in H_0^1(\Omega). \tag{34}$$

Inserting $v = u_1 - u_2$ into (32) we obtain

$$\int_\Omega a\nabla(u_1 - u_2).\nabla(u_1 - u_2)dx = \int_\Omega (y_1 - y_2)(u_1 - u_2)dx. \tag{35}$$

From (19), (20) and (33) we have

$$
\|p_1 - p_2\|_{L^2(\Omega)}^2 \overset{(19)}{\leq} c^2 \|p_1 - p_2\|_{H^2(\Omega)}^2 \overset{(20)}{\leq} \frac{c^2}{c_1^2} \|\!|p_1 - p_2|\!\|^2
$$

$$
\overset{(33)}{=} \frac{c^2}{c_1^2} \int_\Omega (u_1 - u_2)(p_1 - p_2)dx
$$

$$
\leq \frac{c^2}{c_1^2} \|u_1 - u_2\|_{L^2(\Omega)} \|p_1 - p_2\|_{L^2(\Omega)}
$$

$$
\Rightarrow \|p_1 - p_2\|_{L^2(\Omega)} \leq \frac{c^2}{c_1^2} \|u_1 - u_2\|_{L^2(\Omega)}. \tag{36}
$$

Similarly, from (19), (20) and (35) we have

$$
\|u_1 - u_2\|_{L^2(\Omega)} \leq \frac{c^2}{c_1^2} \|y_1 - y_2\|_{L^2(\Omega)}. \tag{37}
$$

Finally, from (36) and (36) we deduce

$$
\|p_1 - p_2\|_{L^2(\Omega)} \leq \frac{c^4}{c_1^4} \|y_1 - y_2\|_{L^2(\Omega)}. \tag{38}
$$

This implies that F' is Lipschitz continuous with Lipschitz constant $\frac{c^4}{c_1^4}$.   □

7 Numerical solution

In this section we illustrate the performance of the proximal method for the inverse source problem in previous section. We assume that $\Omega = [-1, 1]^2$ and

$$
a(x_1, x_2) = 1 + x_1^2 + x_2^2, \ u_0 = 1, \ y^*(x_1, x_2) = e^{-10(x_1^2 + x_2^2)}.
$$

We discretize problems (14) and (16) by the finite element method (FEM). The domain Ω is divided into triangles and define the following finite element spaces:

$$
V^N = \{u \in H^1(\Omega) : u|_K \in \mathcal{P}_1(K), \forall K \in \mathcal{T}\},
$$
$$
V_0^N = \{u \in H_0^1(\Omega) : u|_K \in \mathcal{P}_1(K), \forall K \in \mathcal{T}\},
$$

where $\mathcal{P}_1(K)$ denotes the set of linear polynomials in K, N and $\mathcal{T}$ are respectively the number of nodes and the set of triangles of the mesh. Let $\{\varphi_k\}_{k=1,\ldots,N}$ is the basis consisting of piecewise linear finite elements in discretized space V^N. Then, the discretization of problem (14) is: find $u \in V^N$ such that

$$
\int_\Omega a\nabla u.\nabla v dx = \int_\Omega yv dx, \ \forall v \in V_0^N \text{ and } u = Pu_0 \text{ on } \partial\Omega, \tag{39}
$$

where P is the interpolation in FEM.

Similarly, the discretization of problem (16) is

$$
\min_{y \in V^N} \Theta(y) = F(y) + \alpha\Phi(y), \tag{40}
$$

where

$$
F(y) = \frac{1}{2} \|K(y) - u^\delta\|_{L^2(\Omega)}, \ \Phi(y) = \sum_{i=1}^N |y_i|_0 \tag{41}
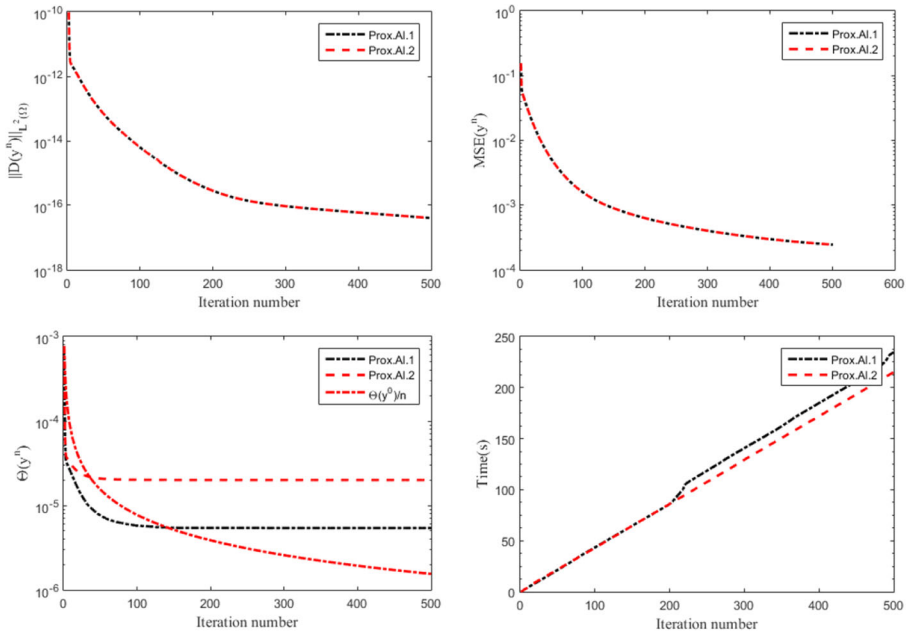$$

Fig. 1 The value of $\|D(y^n)\|_{L^2(\Omega)}$, $MSE(y^n)$, $\Theta(y^n)$ and computational time in *Prox.Al.1* and *Prox.Al.2*

with $y_i = \langle y, \varphi_i \rangle_{L^2(\Omega)}$.

To obtain u^* and u^δ, we solve (39) on a mesh with $N = 887$ nodes and 1680 triangles. The solution to (39) and all the parameters a, u_0 and y are represented by piecewise linear finite elements. Problem (40) is solved by the proximal method, Algorithm 5.1.

We measure the convergence of the computed minimizers to the true parameter y^* by the mean square error sequence

$$MSE(y^k) = \int_\Omega (y^k - y^*)^2 dx,$$

and

$$\|D(y^k)\| = \|y^k - \mathbb{P}^l_{\frac{2\alpha}{s^*}}\left(y^k - \frac{1}{s^*}\nabla F(y^k)\right)\|_{L^2(\Omega)}, \ l = 1, 2$$

with $s^* = 100$.

7.1 Numerical experiments with $\delta = 0$

We first discuss results without noise, i.e., $u^\delta = u^*$. Let *Prox.Al.1* and *Prox.Al.2* be Algorithm 5.1 with $l = 1$ and $l = 2$ respectively. In these algorithms we set $q = 0.5$, $[\underline{s}, \overline{s}] := [5 \cdot 10^{-4}, 5 \cdot 10^4]$ and $\alpha = 10^{-6}$.

Convergence and convergence rate In Fig. 1 we see that the sequences $\{\|D(y^k)\|_{L^2(\Omega)}\}$ generated by *Prox.Al.1* and *Prox.Al.2* are decreasing to zero, i.e.,

the sequences $\{y^k\}$ converge to stationary points of $\Theta(\cdot)$. The behavior of $MSE(y^k)$ indicate that the sequences $\{y^k\}$ in two algorithms converge to the unknown parameter y^*. We also see that the objective functionals $\{\Theta(y^k)\}$ decrease monotonically in two algorithms. These observations are suitable with the theory results in Lemma 4.1. The computational times in two algorithms are similar. Figure 3 shows that after the first iteration, the step sizes $\frac{1}{s^n}$ do not change. Thus, we believe that they are smaller than $\frac{1}{L}$, the Lipschitz constant of F'.

Note that we do not have any result about convergence rate of the proximal algorithm in this paper. Since problem (9) is nonconvex and nondifferentiable, such a result is very hard to prove. However, in [12] for the convex problem of the same form of (9), the convergence rate is proven to be $O(\frac{1}{n})$, where n is the number of iterations of the proximal method. This implies that the convergence rate of the proximal algorithm for problem (9) do not better than $O(\frac{1}{n})$. Figure 1 shows this statement is true since the sequences $\Theta(y^n)$ in two algorithms are decreasing slower than $\Theta(y^0)/n$.

Quality of recovered solutions Figure 2 illustrates y^* and y^n with $n = 500$ in two algorithms. The parameter y^* has been reconstructed very well. The mean square
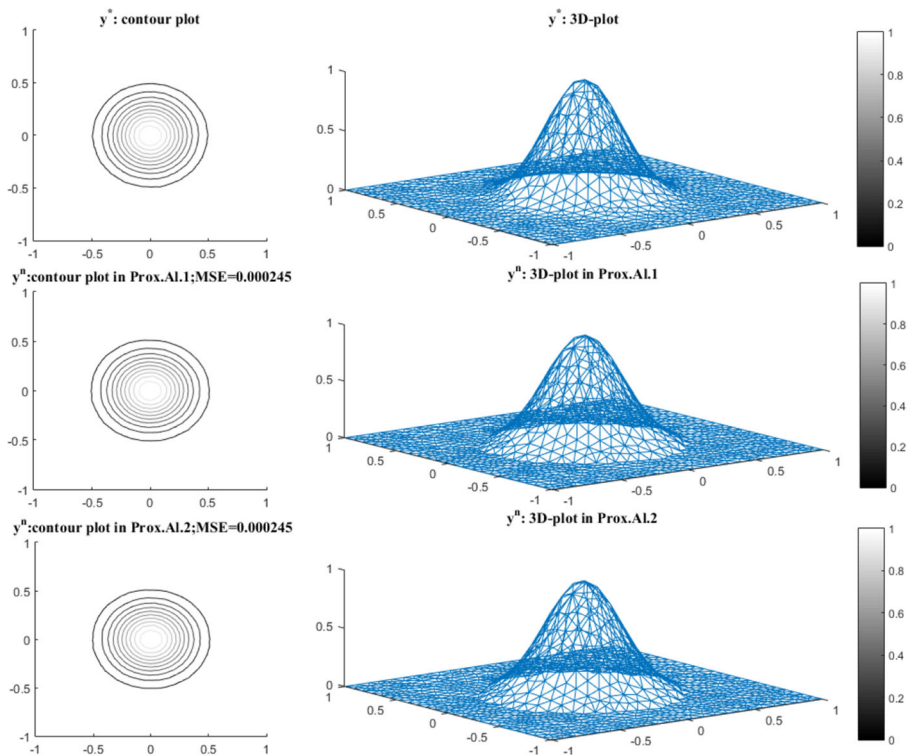

Fig. 2 3D plots and contour plots: exact y^* (top); y^n with $n = 500$ in *Prox.Al.1* (middle) and in *Prox.Al.2* (bottom)
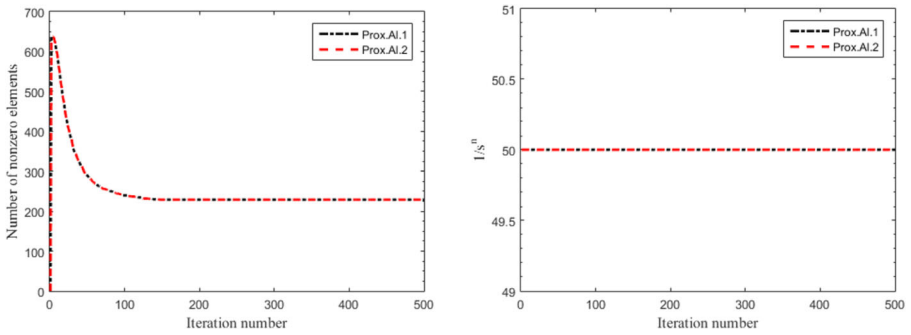
Fig. 3 Number of nonzero elements of y^n (left) and step size s^n (right) in *Prox.Al.1* and *Prox.Al.2*

errors in two algorithms are the same. Note that y^* is always greater than zero (there are many components close to zero), but two reconstructions, y^n with $n = 500$ in two algorithms have many zero components, which are shown in Fig. 3.

7.2 Numerical experiments with noisy data

We now consider noisy data. To obtain $u^\delta \in L^2(\Omega)$, we compute $u^\delta = u^* + \epsilon \frac{R}{\|R\|_{L^2(\Omega)}}$, where R is computed with the MATLAB routine $randn(size(y))$ with setting $randn('state', 0)$. Here, we choose $\epsilon = 10^{-2}$. Figure 4 illustrates the graph of $u^\delta - u^*$. It shows that the data is noisy.

With noisy data we also use the same setting in *Prox.Al.1* and *Prox.Al.2* , i.e., all parameters are the same as those in the case of no noise. From Fig. 5 we have the similar observations on convergence and convergence rate as well as computational time. However, for $MSE(y^n)$ they are decreasing in the first number of iterations, but they are increasing after that. This is always happened in regularization of inverse problems. With noisy data we need to study a rule of stopping the algorithms. Such a rule is out of the aim of this paper. Here, we choose y^k such that $MSE(y^k)$ is the smallest in the sequences of $MSE(y^n)$ generated by two algorithms. These approximate solutions are illustrated in Fig. 6.
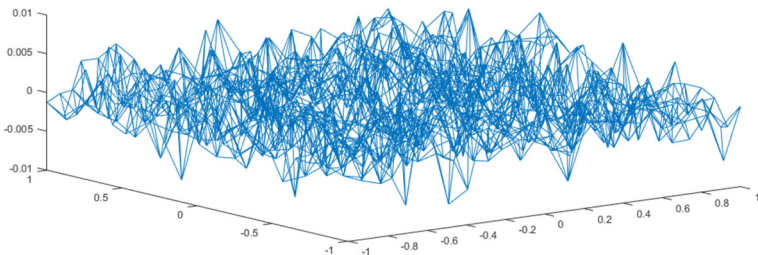

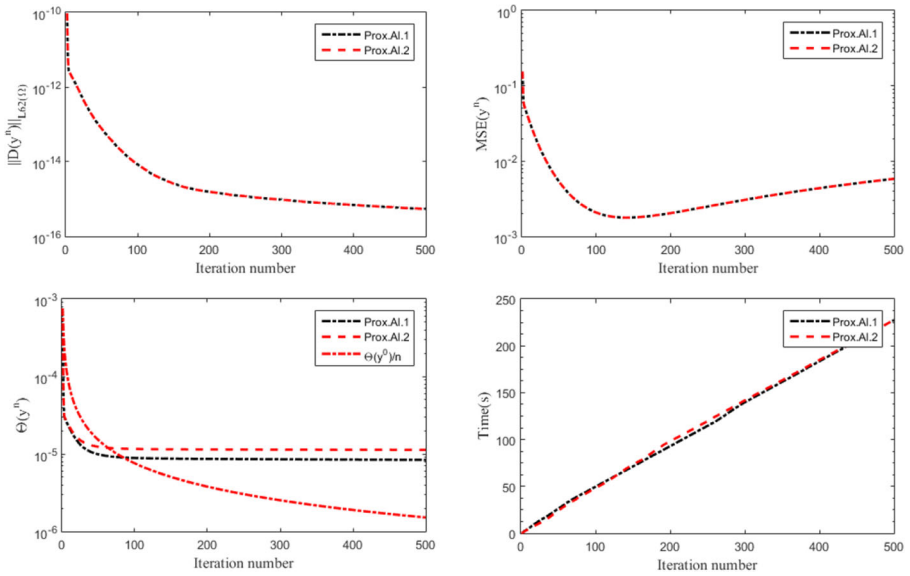**Fig. 4** 3D graph of $u^\delta - u^*$

Fig. 5 The value of $\|D(y^n)\|_{L^2(\Omega)}$, $MSE(y^n)$, $\Theta(y^n)$ and computational time in *Prox.Al.1* and *Prox.Al.2*
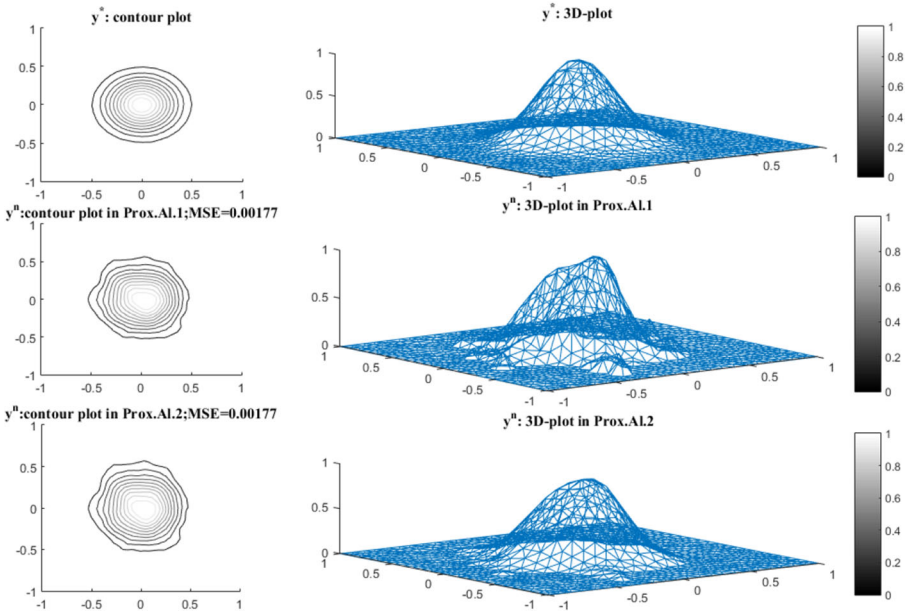

Fig. 6 3D plots and contour plots: exact y^* (top); y^n with $n = 139$ in *Prox.Al.1* (middle) and in *Prox.Al.2* (bottom)

7.3 Numerical experiments with a nonsmooth source functional

We close this section by considering an inverse source problem with a nonsmooth source functional . Here, we use the same setting as two previous subsections, but replace the source functional by another nonsmooth functional, i.e., we assume that $\Omega = [-1, 1]^2$ and

$$a(x_1, x_2) = 1 + x_1^2 + x_2^2, \ u_0 = 1,$$

$$y^*(x_1, x_2) = \begin{cases} 2, & \text{if } (x_1, x_2) \in [-0.6, -0.2] \times [-0.6, -0.2], \\ 2, & \text{if } (x_1, x_2) \in [0.2, 0.6] \times [0.2, 0.6], \\ 0, & \text{otherwise.} \end{cases} .$$

We generate noisy data as previous subsection with $\epsilon = 10^{-2}$. For *Prox.Al.1* and *Prox.Al.2* , we set $q = 0.5$, $[\underline{s}, \overline{s}] := [5 \cdot 10^{-4}, 5 \cdot 10^4]$ and $\alpha = 10^{-6}$. In Fig. 7, the sequences $\{\|D(y^n)\|\}$ show that the sequences $\{y^n\}$ converge to stationary points of Θ, the sequences $\{y^n\}$ also approach to the exact parameter y^* but the sequences $MSE(y^n)$ have higher values than those in the smooth case (see Fig. 5). Figure 8 shows the sparsity of recovered solutions y^n and the behaviors of the stepsizes $\frac{1}{s_n}$, which are suitable with theoretical analysis. It is harder to recover a nonsmooth parameter than a smooth parameter. The qualities of reconstructions are worse in this situation (see Fig. 9).
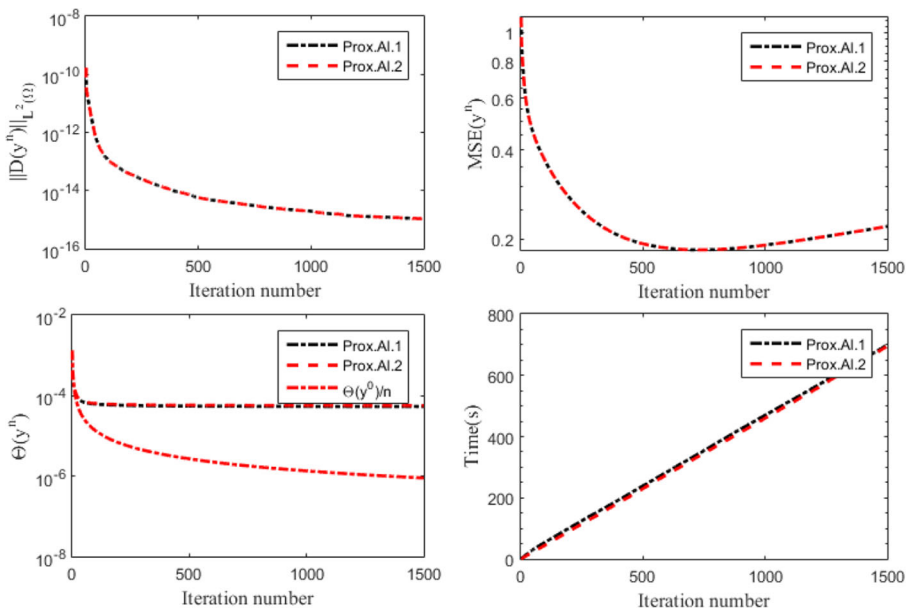

Fig. 7 The value of $\|D(y^n)\|_{L^2(\Omega)}$, $MSE(y^n)$, $\Theta(y^n)$ and computational time in *Prox.Al.1* and *Prox.Al.2*
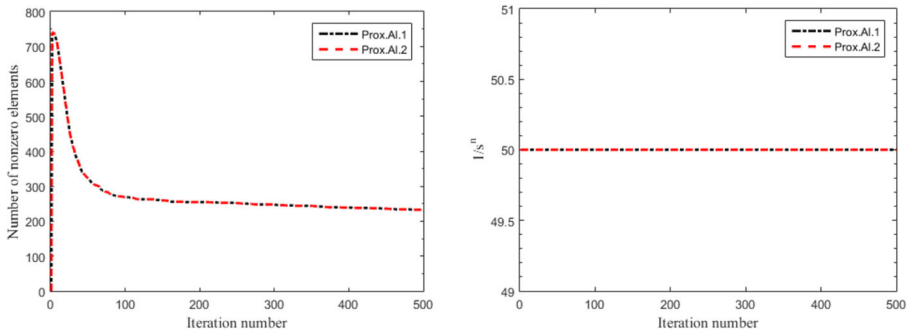
Fig. 8 Number of nonzero elements of y^n (left) and step size s^n (right) in *Prox.Al.1* and *Prox.Al.2*

From three situations, we see that *Prox.Al.1* and *Prox.Al.2* converge to stationary points of Θ and the sequences of the objective functional are decreasing monotonically. *Prox.Al.1* is faster than *Prox.Al.2*, but they have the same order of convergence rates. Furthermore, the computational time for each iteration seems to be constant in both algorithms. Finally, at the same noise level the reconstructions for smooth unknown parameters are better than those for nonsmooth unknown parameters.
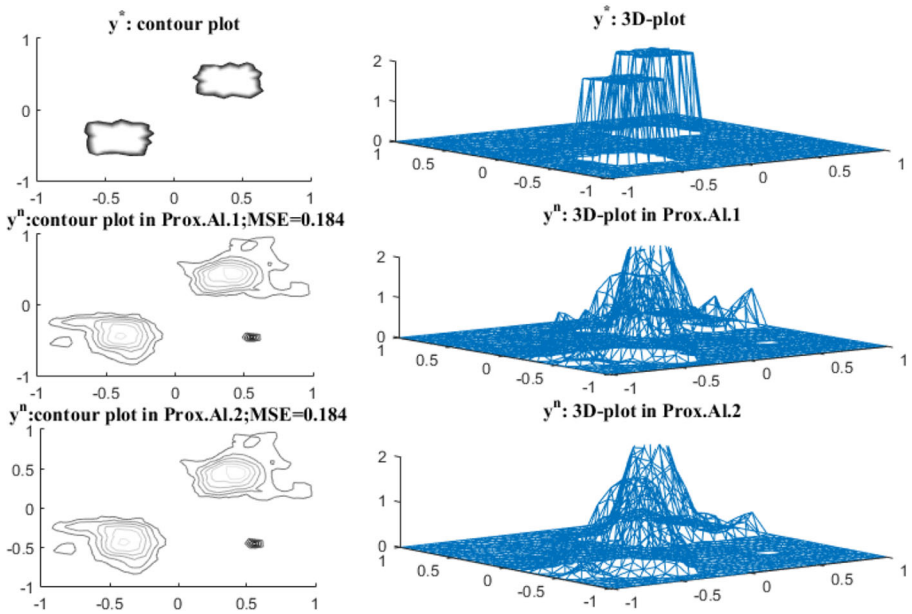

Fig. 9 3D plots and contour plots: exact y^* (top); y^n with $n = 139$ in *Prox.Al.1* (middle) and in *Prox.Al.2* (bottom)

8 Conclusion and future work

We have considered the minimization problem

$$\min_{u \in H} F(u) + \alpha \Phi(u),$$

where $F : H \to \mathbb{R}$ is Fréchet differentiable, F' is Lipschitz continuous and $\Phi(u) = \|u\|_0$.

Under Assumption 1 and the condition on step sizes s^k, Theorem 4.1 has shown that the iteration

$$u^{k+1} \in \mathbb{P}_{\frac{2\alpha}{s^k}}\left(u^k - \frac{1}{s^k}\nabla F(u^k)\right). \tag{42}$$

converges (under a subsequence) to a stationary point u^*, i.e., u^* satisfies

$$u^* \in \mathbb{P}_{\frac{2\alpha}{s^*}}\left(u^* - \frac{1}{s^*}\nabla F(u^*)\right),$$

for some $s^* \geq L$.

Note that the proximal method falls in class of gradient type methods. Thus, it converges quite slow and has linear convergence rate. Therefore, we will study some better methods in the next step, e.g., some algorithms fall into class of Newton method. To make this point clearer, we note that instead of solving the above minimization problem, we can concentrate on solving its first-order optimal condition equation:

$$0 \in D(u) := u - \mathbb{P}_{\frac{2\alpha}{s}}\left(u - \frac{1}{s}\nabla F(u)\right),$$

for some fixed $s \geq L$. Therefore, we can extend Newton method or quasi-Newton method to solve the last equation. It is valuable to show that the last equation is not continuous. As a result, it is most difficult part of extending Newton and quasi-Newton method to deal with this situation. We will propose our ideas to overcome this issue in future.

There are some other issues which are needed to study such as the rules of choosing regularization parameter with respect to noise level, the rules of stopping the algorithms in the case of noisy data and the algorithms for the constrained minimization problem

$$\min_{u \in D} F(u) + \alpha \Phi(u),$$

where $F : D \subset H \to \mathbb{R}$ is Fréchet differentiable, F' is Lipschitz continuous and $\Phi(u) = \|u\|_0$.

Declarations

Conflict of interest The authors declare no competing interests.

References

1. Beck, A., Teboulle, M.: A fast iterative shrinkage-thresholding algorithm for linear inverse problems. SIAM J. Imag. Sci. **2**(1), 183–202 (2009)
2. Blumensath, T., Davies, M.E.: Iterative thresholding for sparse approximations. J. Fourier Anal. Appl. **14**(5–6), 629–654 (2008)
3. Blumensath, T., Yaghoobi, M., Davies, M.E.: l_0 regularisation. In: Iterative hard thresholding IEEE International Conference on Acoustics, Speech and Signal Processing-ICASSP'07, vol. 3, pp. III–877. IEEE (2007)
4. Daubechies, I., Defrise, M., De Mol, C.: An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. Commun. Pure Appl. Math. **57**(11), 1413–1457 (2004)
5. Engl, H.W., Hanke, M., Neubauer, A.: Regularization of Inverse Problems. Kluwer Academic publishers (1996)
6. Evans, L.C.: Partial differential equations. Graduate Studies in Mathematics, 19(2) (1998)
7. Foucart, S.: Hard thresholding pursuit: An algorithm for compressive sensing. SIAM J. Numer. Anal. **49**(6), 2543–2563 (2011)
8. Lorenz, D.A., Maass, P., Muoi, P.Q.: Gradient descent for Tikhonov functionals with sparsity constraints: Theory and numerical comparison of step size rules. Electron. Trans. Numer. Anal. **39**, 437–463 (2012)
9. Lu, Z., Zhang, Y.: Sparse approximation via penalty decomposition methods. SIAM J. Optim. **23**(4), 2448–2478 (2013)
10. Quy Muoi, P., Nho Hào, D., Maass, P., Pidcock, M.: Semismooth Newton and quasi-newton methods in weighted l^1-regularization. J. Inverse Ill-Posed Problems **21**(5), 665–693 (2013)
11. Quy Muoi, P., Nho Hào, D., Maass, P., Pidcock, M.: Descent gradient methods for nonsmooth minimization problems in ill-posed problems. J. Comput. Appl. Math. **298**, 105–122 (2016)
12. Nesterov, Y.: Gradient methods for minimizing composite functions. Math. Program. **140**(1), 125–161 (2013)
13. Nikolova, M.: Description of the minimizers of least squares regularized with ℓ_0-norm. Uniqueness of the global minimizer. SIAM J. Imaging Sci. **6**(2), 904–937 (2013)
14. Soubies, E., Blanc-Féraud, L., Aubert, G.: A continuous exact ℓ_0 penalty (cel0) for least squares regularized problem. SIAM J. Imaging Sci. **8**(3), 1607–1639 (2015)
15. Wachsmuth, D.: Iterative hard-thresholding applied to optimal control problems with $l^0(\omega)$ control cost. SIAM J. Control. Optim. **57**(2), 854–879 (2019)
16. Wang, W., Lu, S., Hofmann, B., Cheng, J.: Tikhonov regularization with l^0-term complementing a convex penalty: l^1-convergence under sparsity constraints. J. Inverse Ill-posed Problems **27**(4), 575–590 (2019)
17. Wang, W., Lu, S., Mao, H., Cheng, J.: Multi-parameter Tikhonov regularization with the l^0 sparsity constraint. Inverse Problems **29**(6), 065018 (2013)

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.