

Levenberg–Marquardt method in Banach spaces with general convex regularization terms

Qinian Jin¹ · Hongqi Yang²

Received: 28 December 2014 / Revised: 9 July 2015 / Published online: 8 September 2015
© Springer-Verlag Berlin Heidelberg 2015

Abstract We propose a Levenberg–Marquardt method with general uniformly convex regularization terms to solve nonlinear inverse problems in Banach spaces, which is an extension of the scheme proposed by Hanke in (Inverse Probl 13:79–95, 1997) in Hilbert space setting. The method is so designed that it can be used to detect the features of the sought solutions such as sparsity or piecewise constancy. It can also be used to deal with the situation that the data is contaminated by noise containing outliers. By using tools from convex analysis in Banach spaces, we establish the convergence of the method. Numerical simulations are reported to test the performance of the method.

Mathematics Subject Classification 65J15 · 65J20 · 47H17

1 Introduction

Inverse problems can arise from many applications in natural sciences. Most of inverse problems are usually ill-posed in the sense that their solutions do not depend continuously on the data. Thus, the stable reconstruction of solutions of inverse problems requires regularization techniques [2, 8, 25].

For solving nonlinear inverse problems in Hilbert spaces, Hanke introduced in [11] his regularizing Levenberg–Marquardt scheme which is a stable iterative procedure

✉ Qinian Jin
Qinian.Jin@anu.edu.au
Hongqi Yang
mcsyhq@mail.sysu.edu.cn

¹ Mathematical Sciences Institute, Australian National University, Canberra, ACT 2601, Australia

² Guangdong Provincial Key Lab of Computational Science, School of Data and Computer Science, Sun Yat-sen University, Guangzhou 510275, China

with the Lagrangian multipliers in the Levenberg–Marquardt method being updated by an adaptive strategy. To describe this method more precisely, consider nonlinear inverse problems that can be formulated as the form

$$F(x) = y, \quad (1.1)$$

where $F : \mathcal{X} \rightarrow \mathcal{Y}$ is a nonlinear Fréchet differentiable operator between two Hilbert spaces \mathcal{X} and \mathcal{Y} whose Fréchet derivative at x is denoted by $F'(x)$. Assuming that y^δ is the only available noisy data, the regularizing Levenberg–Marquardt scheme in [11] constructs the next iterate x_{n+1} from a current iterate x_n by first regularizing the linearized equation

$$F'(x_n)(x - x_n) = y^\delta - F(x_n)$$

at x_n via the Tikhonov regularization

$$x_n(\alpha, y^\delta) := \arg \min_{x \in \mathcal{X}} \left\{ \frac{1}{2} \|y^\delta - F(x_n) - F'(x_n)(x - x_n)\|^2 + \frac{\alpha}{2} \|x - x_n\|^2 \right\} \quad (1.2)$$

using a quadratic penalty term and then define $x_{n+1} := x_n(\alpha_n, y^\delta)$ with $\alpha_n > 0$ being a parameter satisfying

$$\|y^\delta - F(x_n) - F'(x_n)(x_n(\alpha_n, y^\delta) - x_n)\| = \mu \|y^\delta - F(x_n)\|, \quad (1.3)$$

where $0 < \mu < 1$ is a preassigned number. It has been shown in [11] that this defines a regularization method as long as the iteration is terminated by the discrepancy principle. The regularizing Levenberg–Marquardt scheme has had far reaching impact on the development of iterative regularization methods for solving nonlinear inverse problems and has stimulated considerable subsequent work, see [13, 17, 18, 27, 30] and the references therein.

In order to deal with the situations where the sought solutions are sparse or piecewise constant and where the data are contaminated by general noise, one has to use the sparse promoting functionals or total variational like functionals as regularization terms and use general fidelity terms to fit data. This leads to the consideration on regularization methods in Banach spaces with general regularization terms which has emerged as a highly active research field in recent years. The monograph [31] collects some of such research works including the variational regularization of Tikhonov type and some iterative regularization methods in Banach spaces. One may further refer to [14, 18, 20–24, 26, 28] for more recent works.

The purpose of the present paper is to extend the regularizing Levenberg–Marquardt scheme of Hanke to solve nonlinear inverse problems in Banach spaces using general convex regularization terms. Thus, we will consider nonlinear inverse problems modelled by (1.1) with F being a nonlinear operator between two Banach spaces \mathcal{X} and \mathcal{Y} whose dual spaces are denoted as \mathcal{X}^* and \mathcal{Y}^* respectively. Let $\Theta : \mathcal{X} \rightarrow (-\infty, \infty]$ be a proper, lower semi-continuous, uniformly convex function which is chosen accord-

ing to the a prior information on the feature of the sought solution. Using a current iterate $(x_n, \xi_n) \in \mathcal{X} \times \mathcal{X}^*$ with $\xi_n \in \partial\Theta(x_n)$, the subdifferential of Θ at x_n , we define

$$x_n(\alpha, y^\delta) := \arg \min_{x \in \mathcal{X}} \left\{ \frac{1}{r} \|y^\delta - F(x_n) - F'(x_n)(x - x_n)\|^r + \alpha D_{\xi_n} \Theta(x, x_n) \right\}, \tag{1.4}$$

where $1 < r < \infty$ is a given number and $D_{\xi_n} \Theta(x, x_n)$ denotes the Bregman distance induced by Θ at x_n in the direction ξ_n . We then define the next iterate x_{n+1} by $x_{n+1} = x_n(\alpha_n, y^\delta)$, where $\alpha_n > 0$ is chosen such that

$$\mu_0 \|y^\delta - F(x_n)\| \leq \|y^\delta - F(x_n) - F'(x_n)(x_n(\alpha_n, y^\delta) - x_n)\| \leq \mu_1 \|y^\delta - F(x_n)\|$$

with preassigned numbers $0 < \mu_0 \leq \mu_1 < 1$. This choice of α_n is a relaxation of (1.3) and has more flexibility when the root of the Eq. (1.3) is difficult to determine exactly. We need to update the subgradient ξ_n to $\xi_{n+1} \in \partial\Theta(x_{n+1})$ which, according to the minimization property of x_{n+1} , can be taken as

$$\xi_{n+1} = \xi_n + \frac{1}{\alpha_n} F'(x_n)^* J_r^{\mathcal{Y}}(y^\delta - F(x_n) - F'(x_n)(x_{n+1} - x_n)),$$

where $F'(x_n)^* : \mathcal{Y}^* \rightarrow \mathcal{X}^*$ denotes the adjoint of $F'(x_n)$ and $J_r^{\mathcal{Y}} : \mathcal{Y} \rightarrow \mathcal{Y}^*$ denotes the duality mapping on \mathcal{Y} with gauge function $t \rightarrow t^{r-1}$. We repeat the above procedure until a discrepancy principle is satisfied. The precise description of the above extension will be given in Algorithm 1. The main result of this paper shows that the above extension of the Levenberg–Marquardt scheme is well-defined and exhibits the regularization property when used to solve ill-posed problems.

The introduction of a general convex function Θ into the algorithm presents many challenging issues on its convergence analysis. Unlike (1.2) whose minimizer can be determined explicitly, the minimizer of (1.4) does not have a closed form. The possible non-smoothness of Θ and the non-Hilbertian structure of \mathcal{X} and \mathcal{Y} prevent us from using the classical techniques. Instead we have to utilize tools from convex optimization and non-smooth analysis, including the subdifferential calculus and the Bregman distance. The convergence analysis becomes even more subtle when considering the regularization property. The main obstacle comes from the stability issue; an iterative sequence constructed using noisy data can split into many possible noise-free iterative sequences as the noise level tends zero, due to the non-unique determination of α_n . We will conquer this difficulty by borrowing an idea from [10], which is based on the diagonal sequence argument, to show that all these noise-free sequences actually converge uniformly in certain sense. On the other hand, unlike the variational regularization of Tikhonov [31] and the non-stationary iterated Tikhonov regularization [22] whose numerical implementation requires to solve several non-convex minimization problems, our method involves only convex minimization problems and therefore has the advantage of being implemented efficiently by convex optimization techniques.

Under a prior choice of $\{\alpha_n\}$, a Levenberg–Marquardt method was considered in [1] to solve (1.1) with $\mathcal{X} = \mathbf{L}^2(\Omega)$ and \mathcal{Y} a Hilbert space using the convex penalty function $\Theta(x) = a\|x\|_{\mathbf{L}^2}^2 + \int_{\Omega} |Dx|$, where $\int_{\Omega} |Dx|$ denotes the total variation of x .

The analysis in [1], however, is somewhat preliminary since it provides only weak convergence along subsequences. In contrast, our method chooses $\{\alpha_n\}$ adaptively, and the whole iterative sequence converges to a solution of (1.1) strongly.

In recent years the iteratively regularized Gauss–Newton method has been extended to solve nonlinear inverse problems in Banach spaces [21,24,26] in which it defines $\{x_n\}$ by

$$x_{n+1} = \arg \min_{x \in \mathcal{X}} \left\{ \frac{1}{r} \|y^\delta - F(x_n) - F'(x_n)(x - x_n)\|^r + \alpha_n D_{\xi_0} \Theta(x, x_0) \right\}$$

It looks similar to (1.4) but in fact they are essentially different. The iteratively regularized Gauss–Newton method always defines x_n in a neighborhood of the initial guess x_0 , while the Levenberg–Marquardt method defines x_{n+1} in a region around x_n for each $n \geq 0$. From the optimization point of view, Levenberg–Marquardt method is more favorable in nature.

The rest of this paper is organized as follows. In Sect. 2 we collect some basic results from convex analysis in Banach spaces and prove a continuous dependence result of minimizers of uniformly convex functionals on various parameters. We then present in Sect. 3 the Levenberg–Marquardt method in Banach spaces with general convex regularization terms and show that the method is well-defined. The detailed convergence analysis is given in Sect. 4. Finally in Sect. 5 we report various numerical results to indicate the performance of the method.

2 Preliminaries

Let \mathcal{X} and \mathcal{Y} be two Banach spaces whose norms are denoted by $\|\cdot\|$. We use \mathcal{X}^* and \mathcal{Y}^* to denote their dual spaces respectively. Given $x \in \mathcal{X}$ and $x^* \in \mathcal{X}^*$ we write $\langle x^*, x \rangle = x^*(x)$ for the duality pair. We use “ \rightarrow ” and “ \rightharpoonup ” to denote the strong convergence and the weak convergence respectively. By $\mathcal{L}(\mathcal{X}, \mathcal{Y})$ we denote for the space of all continuous linear operators from \mathcal{X} to \mathcal{Y} . For any $A \in \mathcal{L}(\mathcal{X}, \mathcal{Y})$, we use $A^* : \mathcal{Y}^* \rightarrow \mathcal{X}^*$ to denote its adjoint, i.e. $\langle A^*y^*, x \rangle = \langle y^*, Ax \rangle$ for any $x \in \mathcal{X}$ and $y^* \in \mathcal{Y}^*$. We use $\mathcal{N}(A) = \{x \in \mathcal{X} : Ax = 0\}$ to denote the null space of A and define

$$\mathcal{N}(A)^\perp := \{\xi \in \mathcal{X}^* : \langle \xi, x \rangle = 0 \text{ for all } x \in \mathcal{N}(A)\}.$$

When \mathcal{X} is reflexive, there holds $\mathcal{N}(A)^\perp = \overline{\mathcal{R}(A^*)}$, where $\mathcal{R}(A^*)$ denotes the range space of A^* and $\overline{\mathcal{R}(A^*)}$ denotes the closure of $\mathcal{R}(A^*)$ in \mathcal{X}^* .

For a convex function $\Theta : \mathcal{X} \rightarrow (-\infty, \infty]$, we use $\mathcal{D}(\Theta) := \{x \in \mathcal{X} : \Theta(x) < +\infty\}$ to denote its effective domain. We call Θ proper if $\mathcal{D}(\Theta) \neq \emptyset$. Given $x \in \mathcal{X}$ we define

$$\partial\Theta(x) := \{\xi \in \mathcal{X}^* : \Theta(\bar{x}) - \Theta(x) - \langle \xi, \bar{x} - x \rangle \geq 0 \text{ for all } \bar{x} \in \mathcal{X}\}.$$

Any element $\xi \in \partial\Theta(x)$ is called a subgradient of Θ at x . The multi-valued mapping $\partial\Theta : \mathcal{X} \rightarrow 2^{\mathcal{X}^*}$ is called the subdifferential of Θ . It could happen that $\partial\Theta(x) = \emptyset$ for some $x \in \mathcal{D}(\Theta)$. Let

$$\mathcal{D}(\partial\Theta) := \{x \in \mathcal{D}(\Theta) : \partial\Theta(x) \neq \emptyset\}.$$

For $x \in \mathcal{D}(\partial\Theta)$ and $\xi \in \partial\Theta(x)$ we define

$$D_\xi\Theta(\bar{x}, x) := \Theta(\bar{x}) - \Theta(x) - \langle \xi, \bar{x} - x \rangle, \quad \forall \bar{x} \in \mathcal{X}$$

which is called the Bregman distance induced by Θ at x in the direction ξ . Clearly $D_\xi\Theta(\bar{x}, x) \geq 0$. By straightforward calculation one can see that

$$D_{\xi_2}\Theta(x, x_2) - D_{\xi_1}\Theta(x, x_1) = -D_{\xi_1}\Theta(x_2, x_1) + \langle \xi_2 - \xi_1, x_2 - x \rangle \tag{2.1}$$

for all $x_1, x_2 \in \mathcal{D}(\partial\Theta)$, $\xi_1 \in \partial\Theta(x_1)$, $\xi_2 \in \partial\Theta(x_2)$ and $x \in \mathcal{X}$.

Bregman distance can be used to obtain information under the Banach space norm when Θ has stronger convexity. A proper convex function $\Theta : \mathcal{X} \rightarrow (-\infty, \infty]$ is called uniformly convex if there is a strictly increasing function $\varphi : [0, \infty) \rightarrow [0, \infty)$ with $\varphi(0) = 0$ such that

$$\Theta(\lambda\bar{x} + (1 - \lambda)x) + \lambda(1 - \lambda)\varphi(\|\bar{x} - x\|) \leq \lambda\Theta(\bar{x}) + (1 - \lambda)\Theta(x) \tag{2.2}$$

for all $\bar{x}, x \in \mathcal{X}$ and $\lambda \in [0, 1]$. It is easily seen that if Θ is uniformly convex in the sense of (2.2) then

$$D_\xi\Theta(\bar{x}, x) \geq \varphi(\|\bar{x} - x\|) \tag{2.3}$$

for all $\bar{x} \in \mathcal{X}$, $x \in \mathcal{D}(\partial\Theta)$ and $\xi \in \partial\Theta(x)$. Moreover, it follows from [32, Proposition 3.5.8] that any proper, weakly lower semi-continuous, uniform convex function $\Theta : \mathcal{X} \rightarrow (-\infty, \infty]$ is coercive in the sense that

$$\liminf_{\|x\| \rightarrow \infty} \frac{\Theta(x)}{\|x\|^2} > 0. \tag{2.4}$$

On a Banach space \mathcal{X} , we consider for $1 < r < \infty$ the convex function $x \rightarrow \|x\|^r/r$. Its subdifferential at x is given by

$$J_r^\mathcal{X}(x) := \{\xi \in \mathcal{X}^* : \|\xi\| = \|x\|^{r-1} \text{ and } \langle \xi, x \rangle = \|x\|^r\}$$

which gives the duality mapping $J_r^\mathcal{X} : \mathcal{X} \rightarrow 2^{\mathcal{X}^*}$ with gauge function $t \rightarrow t^{r-1}$. We call \mathcal{X} uniformly convex if its modulus of convexity

$$\delta_\mathcal{X}(t) := \inf\{2 - \|\bar{x} + x\| : \|\bar{x}\| = \|x\| = 1, \quad \|\bar{x} - x\| \geq t\}$$

satisfies $\delta_\mathcal{X}(t) > 0$ for all $0 < t \leq 2$. We call \mathcal{X} uniformly smooth if its modulus of smoothness

$$\rho_\mathcal{X}(s) := \sup\{\|\bar{x} + x\| + \|\bar{x} - x\| - 2 : \|\bar{x}\| = 1, \quad \|x\| \leq s\}$$

satisfies $\lim_{s \searrow 0} \frac{\rho_\mathcal{X}(s)}{s} = 0$. It is well known [6] that any uniformly convex or uniformly smooth Banach space is reflexive. On a uniformly smooth Banach space \mathcal{X} , every

duality mapping $J_r^{\mathcal{X}}$ with $1 < r < \infty$ is single valued and uniformly continuous on bounded sets. Furthermore, on a uniformly convex Banach space, any sequence $\{x_n\}$ satisfying $x_n \rightharpoonup x$ and $\|x_n\| \rightarrow \|x\|$ must satisfy $x_n \rightarrow x$ as $n \rightarrow \infty$. This property can be generalized for uniformly convex functions which we state in the following result.

Lemma 2.1 [22] *Let $\Theta : \mathcal{X} \rightarrow (-\infty, \infty]$ be a proper, weakly lower semi-continuous, and uniformly convex function. Then Θ admits the Kadec property, i.e. for any sequence $\{x_n\} \subset \mathcal{X}$ satisfying $x_n \rightharpoonup x \in \mathcal{X}$ and $\Theta(x_n) \rightarrow \Theta(x) < \infty$ there holds $x_n \rightarrow x$ as $n \rightarrow \infty$.*

We conclude this section by providing a continuous dependence result of minimizers for uniformly convex cost functionals on various parameters. This result is crucial in the forthcoming sections.

Lemma 2.2 *Let \mathcal{X} and \mathcal{Y} be Banach spaces with \mathcal{X} being reflexive. Let $\{A(x) : \mathcal{X} \rightarrow \mathcal{Y}\}_{x \in \mathcal{D}} \subset \mathcal{L}(\mathcal{X}, \mathcal{Y})$ be such that $x \rightarrow A(x)$ is continuous on $\mathcal{D} \subset \mathcal{X}$. Let $\Theta : \mathcal{X} \rightarrow (-\infty, \infty]$ be a proper, weakly lower semi-continuous, uniformly convex function. Assume that the sequences $\{\alpha^{(\ell)}\} \subset (0, \infty)$, $\{b^{(\ell)}\} \subset \mathcal{Y}$, $\{x^{(\ell)}\} \subset \mathcal{D}$ and $\{\xi^{(\ell)}\} \subset \mathcal{X}^*$ with $\xi^{(\ell)} \in \partial\Theta(x^{(\ell)})$ satisfy*

$$\alpha^{(\ell)} \rightarrow \bar{\alpha}, \quad b^{(\ell)} \rightarrow \bar{b}, \quad x^{(\ell)} \rightarrow \bar{x}, \quad \xi^{(\ell)} \rightarrow \bar{\xi} \quad \text{as } \ell \rightarrow \infty \tag{2.5}$$

for some $\bar{\alpha} > 0$, $\bar{b} \in \mathcal{Y}$, $\bar{x} \in \mathcal{D}$ and $\bar{\xi} \in \mathcal{X}^*$ with $\bar{\xi} \in \partial\Theta(\bar{x})$. For $1 \leq r < \infty$ let

$$z^{(\ell)} := \arg \min_{z \in \mathcal{X}} \left\{ \frac{1}{r} \|b^{(\ell)} - A(x^{(\ell)})z\|^r + \alpha^{(\ell)} D_{\xi^{(\ell)}} \Theta(z, x^{(\ell)}) \right\}.$$

Then $z^{(\ell)} \rightarrow \bar{z}$ and $\Theta(z^{(\ell)}) \rightarrow \Theta(\bar{z})$ as $\ell \rightarrow \infty$, where

$$\bar{z} := \arg \min_{z \in \mathcal{X}} \left\{ \frac{1}{r} \|\bar{b} - A(\bar{x})z\|^r + \bar{\alpha} D_{\bar{\xi}} \Theta(z, \bar{x}) \right\}.$$

Proof It is easy to see that $z^{(\ell)}$ and \bar{z} are uniquely defined since the corresponding cost functionals are weakly lower semi-continuous and uniformly convex and hence coercive, see (2.4). Because $\xi^{(\ell)} \in \partial\Theta(x^{(\ell)})$ and $\bar{\xi} \in \partial\Theta(\bar{x})$, we may use the condition (2.5) to obtain

$$\begin{aligned} \liminf_{\ell \rightarrow \infty} \Theta(x^{(\ell)}) &\geq \liminf_{\ell \rightarrow \infty} (\Theta(\bar{x}) + \langle \bar{\xi}, x^{(\ell)} - \bar{x} \rangle) = \Theta(\bar{x}), \\ \limsup_{\ell \rightarrow \infty} \Theta(x^{(\ell)}) &\leq \limsup_{\ell \rightarrow \infty} (\Theta(\bar{x}) - \langle \bar{\xi}^{(\ell)}, \bar{x} - x^{(\ell)} \rangle) = \Theta(\bar{x}). \end{aligned}$$

Therefore

$$\lim_{\ell \rightarrow \infty} \Theta(x^{(\ell)}) = \Theta(\bar{x}). \tag{2.6}$$

To show $z^{(\ell)} \rightarrow \bar{z}$ we will adapt the argument in [9, 16]. By the definition of $z^{(\ell)}$ we have

$$\frac{1}{r} \|b^{(\ell)} - A(x^{(\ell)})z^{(\ell)}\|^r + \alpha^{(\ell)} D_{\xi^{(\ell)}} \Theta(z^{(\ell)}, x^{(\ell)}) \leq \frac{1}{r} \|b^{(\ell)} - A(x^{(\ell)})x^{(\ell)}\|^r.$$

By the given conditions, the right hand side is bounded and thus $\{D_{\xi^{(\ell)}} \Theta(z^{(\ell)}, x^{(\ell)})\}$ is bounded. Consequently, $\{z^{(\ell)}\}$ is bounded in \mathcal{X} by the uniform convexity of Θ . Since \mathcal{X} is reflexive, $\{z^{(\ell)}\}$ has a subsequence, denoted by the same notation, such that $z^{(\ell)} \rightharpoonup z_*$ as $\ell \rightarrow \infty$ for some $z_* \in \mathcal{X}$. Since $x \rightarrow A(x)$ is continuous and $A(\bar{x}) \in \mathcal{L}(\mathcal{X}, \mathcal{Y})$, we have $A(x^{(\ell)})z^{(\ell)} \rightharpoonup A(\bar{x})z_*$ as $\ell \rightarrow \infty$. By using (2.6) and the weak lower semi-continuity of Θ and the Banach space norm, it follows that

$$\|\bar{b} - A(\bar{x})z_*\| \leq \liminf_{\ell \rightarrow \infty} \|b^{(\ell)} - A(x^{(\ell)})z^{(\ell)}\|, \tag{2.7}$$

$$D_{\bar{\xi}} \Theta(z_*, \bar{x}) \leq \liminf_{\ell \rightarrow \infty} D_{\xi^{(\ell)}} \Theta(z^{(\ell)}, x^{(\ell)}). \tag{2.8}$$

Consequently

$$\begin{aligned} & \frac{1}{r} \|\bar{b} - A(\bar{x})z_*\|^r + \bar{\alpha} D_{\bar{\xi}} \Theta(z_*, \bar{x}) \\ & \leq \frac{1}{r} \liminf_{\ell \rightarrow \infty} \|b^{(\ell)} - A(x^{(\ell)})z^{(\ell)}\|^r + \liminf_{\ell \rightarrow \infty} \alpha^{(\ell)} D_{\xi^{(\ell)}} \Theta(z^{(\ell)}, x^{(\ell)}) \\ & \leq \liminf_{\ell \rightarrow \infty} \left(\frac{1}{r} \|b^{(\ell)} - A(x^{(\ell)})z^{(\ell)}\|^r + \alpha^{(\ell)} D_{\xi^{(\ell)}} \Theta(z^{(\ell)}, x^{(\ell)}) \right) \\ & \leq \limsup_{\ell \rightarrow \infty} \left(\frac{1}{r} \|b^{(\ell)} - A(x^{(\ell)})z^{(\ell)}\|^r + \alpha^{(\ell)} D_{\xi^{(\ell)}} \Theta(z^{(\ell)}, x^{(\ell)}) \right) \\ & \leq \limsup_{\ell \rightarrow \infty} \left(\frac{1}{r} \|b^{(\ell)} - A(x^{(\ell)})\bar{z}\|^r + \alpha^{(\ell)} D_{\xi^{(\ell)}} \Theta(\bar{z}, x^{(\ell)}) \right) \\ & = \frac{1}{r} \|\bar{b} - A(\bar{x})\bar{z}\|^r + \bar{\alpha} D_{\bar{\xi}} \Theta(\bar{z}, \bar{x}), \end{aligned}$$

where, for the last inequality we used the minimality of $z^{(\ell)}$ and for the last equality we used (2.5) and (2.6). By using the minimality and uniqueness of \bar{z} , we can conclude $z_* = \bar{z}$ and

$$\begin{aligned} & \lim_{\ell \rightarrow \infty} \left(\frac{1}{r} \|b^{(\ell)} - A(x^{(\ell)})z^{(\ell)}\|^r + \alpha^{(\ell)} D_{\xi^{(\ell)}} \Theta(z^{(\ell)}, x^{(\ell)}) \right) \\ & = \frac{1}{r} \|\bar{b} - A(\bar{x})\bar{z}\|^r + \bar{\alpha} D_{\bar{\xi}} \Theta(\bar{z}, \bar{x}). \end{aligned} \tag{2.9}$$

Next we show that

$$\lim_{\ell \rightarrow \infty} D_{\xi^{(\ell)}} \Theta(z^{(\ell)}, x^{(\ell)}) = D_{\bar{\xi}} \Theta(\bar{z}, \bar{x}). \tag{2.10}$$

According to (2.8) and $z_* = \bar{z}$, it suffices to show that

$$\gamma_0 := \limsup_{\ell \rightarrow \infty} D_{\xi^{(\ell)}} \Theta(z^{(\ell)}, x^{(\ell)}) \leq D_{\bar{\xi}} \Theta(\bar{z}, \bar{x}) =: \gamma_1.$$

Suppose this is not true, i.e. $\gamma_0 > \gamma_1$. By taking a subsequence if necessary, we may assume that

$$\gamma_0 = \lim_{\ell \rightarrow \infty} D_{\xi^{(\ell)}} \Theta(z^{(\ell)}, x^{(\ell)}).$$

Then from (2.9) it follows that

$$\lim_{\ell \rightarrow \infty} \frac{1}{r} \|b^{(\ell)} - A(x^{(\ell)})z^{(\ell)}\|^r = \frac{1}{r} \|\bar{b} - A(\bar{x})\bar{z}\|^r + \bar{\alpha}(\gamma_1 - \gamma_0) < \frac{1}{r} \|\bar{b} - A(\bar{x})\bar{z}\|^r$$

which is a contradiction to (2.7). We thus obtain (2.10).

In view of (2.10), (2.5) and $z^{(\ell)} \rightarrow \bar{z}$ as $\ell \rightarrow \infty$ we have

$$\lim_{\ell \rightarrow \infty} (\Theta(z^{(\ell)}) - \Theta(x^{(\ell)})) = \Theta(\bar{z}) - \Theta(\bar{x}).$$

This together with (2.6) implies that $\lim_{\ell \rightarrow \infty} \Theta(z^{(\ell)}) = \Theta(\bar{z})$. Since $z^{(\ell)} \rightarrow \bar{z}$, we may use Lemma 2.1 to conclude that $z^{(\ell)} \rightarrow \bar{z}$ as $\ell \rightarrow \infty$. □

3 The method

We consider the Eq. (1.1) arising from nonlinear inverse problems, where $F : \mathcal{D}(F) \subset \mathcal{X} \rightarrow \mathcal{Y}$ is a nonlinear operator between two Banach spaces \mathcal{X} and \mathcal{Y} with domain $\mathcal{D}(F)$. We will assume that (1.1) has a solution. In general (1.1) may have many solutions. In order to find the desired one, some selection criterion should be enforced. According to a prior information on the sought solution, we choose a proper, weakly lower semi-continuous, uniformly convex function $\Theta : \mathcal{X} \rightarrow (-\infty, \infty]$. By taking $x_0 \in \mathcal{D}(\Theta) \cap \mathcal{D}(F)$ and $\xi_0 \in \partial\Theta(x_0)$ as the initial guess, we define x^\dagger to be the solution of (1.1) with the property

$$D_{\xi_0} \Theta(x^\dagger, x_0) = \min_{x \in \mathcal{D}(\Theta) \cap \mathcal{D}(F)} \{D_{\xi_0} \Theta(x, x_0) : F(x) = y\}. \tag{3.1}$$

We are interested in developing algorithms to find the solution x^\dagger of (1.1). We will work under the following standard conditions.

Assumption 1

- (a) \mathcal{X} is a reflexive Banach space and \mathcal{Y} is a uniformly smooth Banach space;
- (b) F is weakly closed on $\mathcal{D}(F)$, i.e. for any sequence $\{x_n\} \subset \mathcal{D}(F)$ satisfying $x_n \rightharpoonup x \in \mathcal{X}$ and $F(x_n) \rightarrow v \in \mathcal{Y}$ there hold $x \in \mathcal{D}(F)$ and $F(x) = v$;
- (c) There is $\rho > 0$ such that $B_{2\rho}(x_0) \subset \mathcal{D}(F)$ and (1.1) has a solution in $B_\rho(x_0) \cap \mathcal{D}(\Theta)$, where $B_\rho(x_0) := \{x \in \mathcal{X} : \|x - x_0\| \leq \rho\}$;

- (d) There exists $\{L(x) : \mathcal{X} \rightarrow \mathcal{Y}\}_{x \in B_{2\rho}(x_0)} \subset \mathcal{L}(\mathcal{X}, \mathcal{Y})$ such that $x \rightarrow L(x)$ is continuous on $B_{2\rho}(x_0)$ and there is $0 \leq \eta < 1$ such that

$$\|F(\bar{x}) - F(x) - L(x)(\bar{x} - x)\| \leq \eta \|F(\bar{x}) - F(x)\|$$

for all $\bar{x}, x \in B_{2\rho}(x_0)$.

In Assumption 1, the uniform smoothness of \mathcal{Y} in (a) is used to guarantee that the duality mapping $J_r^{\mathcal{Y}}$ is single-valued and continuous for each $1 < r < \infty$. We do not require F to be Fréchet differentiable; in case F is Fréchet differentiable, we may take $L(x) = F'(x)$, where $F'(x)$ denotes the Fréchet derivative of F at x . The condition in (d) is the so called tangential cone condition which has been widely used in the analysis of regularization methods for nonlinear inverse problems and has been verified for several important applications [11–13, 20, 22–24, 26, 30]. How to replace the tangential cone condition by a weaker condition is a challenging issue. Under certain smoothness conditions on the solution, a class of Newton methods in Hilbert spaces has been shown in [19] to be order optimal under merely the Lipschitz condition on the Fréchet derivative of F when the methods are terminated by a discrepancy principle. How to extend such result to the Banach space setting remains an open problem. In view of (d) in Assumption 1, it is easily seen that

$$\|F(\bar{x}) - F(x)\| \leq \frac{1}{1 - \eta} \|L(x)(\bar{x} - x)\|, \quad \bar{x}, x \in B_{2\rho}(x_0)$$

which shows that $x \rightarrow F(x)$ is continuous on $B_{2\rho}(x_0)$. When \mathcal{X} is a reflexive Banach space, by using the uniform convexity of Θ and weak closedness of F , it is standard to show that x^\dagger exists. Moreover, [20, Lemma 3.2] gives the following local uniqueness result.

Lemma 3.1 *Let Assumption 1 hold. If $x^\dagger \in B_\rho(x_0) \cap \mathcal{D}(\Theta)$, then x^\dagger is the unique solution of (1.1) in $B_{2\rho}(x_0) \cap \mathcal{D}(\Theta)$ satisfying (3.1).*

In practical applications, instead of y we only have noisy data y^δ satisfying

$$\|y^\delta - y\| \leq \delta \tag{3.2}$$

with a small noise level $\delta > 0$. We will use y^δ to construct an approximate solution to (1.1). To formulate our Levenberg–Marquardt algorithm, assuming x_n^δ has been constructed we consider the linearized equation

$$L(x_n^\delta)(x - x_n^\delta) = y^\delta - F(x_n^\delta)$$

and apply to it the Tikhonov regularization whose regularizing term is the Bregman distance induced by Θ at x_n^δ . The next iterate x_{n+1}^δ is then constructed by chosen the regularization parameter adaptively. This leads to the following Levenberg–Marquardt algorithm.

Algorithm 1 (Levenberg–Marquardt method with noisy data)

1. Take $x_0 \in \mathcal{X}$ and $\xi_0 \in \mathcal{X}^*$ such that $\xi_0 \in \partial\Theta(x_0)$. Pick $0 < \mu_0 \leq \mu_1 < 1$ and $\tau > 1$.
2. Let $x_0^\delta := x_0$ and $\xi_0^\delta := \xi_0$. Assume that x_n^δ and ξ_n^δ are well-defined, we then define x_{n+1}^δ and ξ_{n+1}^δ as follows:
 - (a) For each $\alpha > 0$ we define $x_n(\alpha, y^\delta)$ and $\xi_n(\alpha, y^\delta)$ as

$$x_n(\alpha, y^\delta) = \arg \min_{x \in \mathcal{X}} \left\{ \frac{1}{r} \|y^\delta - F(x_n^\delta) - L(x_n^\delta)(x - x_n^\delta)\|^r + \alpha D_{\xi_n^\delta} \Theta(x, x_n^\delta) \right\},$$

$$\xi_n(\alpha, y^\delta) = \xi_n^\delta + \frac{1}{\alpha} L(x_n^\delta)^* J_r^{\mathcal{Y}} (y^\delta - F(x_n^\delta) - L(x_n^\delta)(x_n(\alpha, y^\delta) - x_n^\delta));$$

- (b) Take $\alpha_n(y^\delta) > 0$ to be a number α such that

$$\mu_0 \|y^\delta - F(x_n^\delta)\| \leq \|y^\delta - F(x_n^\delta) - L(x_n^\delta)(x_n(\alpha, y^\delta) - x_n^\delta)\| \leq \mu_1 \|y^\delta - F(x_n^\delta)\|;$$

- (c) Define $x_{n+1}^\delta := x_n(\alpha_n(y^\delta), y^\delta)$ and $\xi_{n+1}^\delta := \xi_n(\alpha_n(y^\delta), y^\delta)$.

3. Let n_δ be the first integer such that $\|y^\delta - F(x_{n_\delta}^\delta)\| \leq \tau \delta$ and use $x_{n_\delta}^\delta$ as an approximate solution.

When \mathcal{X} and \mathcal{Y} are Hilbert spaces, $r = 2$ and $\Theta(x) = \|x\|^2/2$, Algorithm 1 reduces to the regularizing Levenberg–Marquardt scheme in [11] and each minimizer $x_n(\alpha, y^\delta)$ can be written explicitly. In the setting of Algorithm 1 in Banach spaces with general convex regularization terms, $x_n(\alpha, y^\delta)$ does not have an explicit formula. This increases the difficulty in convergence analysis. By making use of tools from convex analysis, in this section we will show that Algorithm 1 is well-defined, and in Sect. 4 we will show that $x_{n_\delta}^\delta$ indeed converges to a solution of (1.1) as $\delta \rightarrow 0$.

In Algorithm 1, we need to pick $\xi_0 \in \mathcal{X}^*$ and $x_0 \in \mathcal{X}$ such that $\xi_0 \in \partial\Theta(x_0)$. This can be achieved as follows: pick $\xi_0 \in \mathcal{X}^*$ and define

$$x_0 = \arg \min_{x \in \mathcal{X}} \{ \Theta(x) - \langle \xi_0, x \rangle \},$$

then $\xi_0 \in \partial\Theta(x_0)$ holds automatically; in applications, we usually have $\Theta \geq 0$ and $\Theta(0) = 0$, then we can simply take $x_0 = 0$ and $\xi_0 = 0$.

From the definition of $x_n(\alpha, y^\delta)$ in Algorithm 1 we can see that

$$0 \in -L(x_n^\delta)^* J_r^{\mathcal{Y}} (y^\delta - F(x_n^\delta) - L(x_n^\delta)(x_n(\alpha, y^\delta) - x_n^\delta)) + \alpha (\partial\Theta(x_n(\alpha, y^\delta)) - \xi_n^\delta).$$

The definition of $\xi_n(\alpha, y^\delta)$ is exactly motivated by this fact so that

$$\xi_n(\alpha, y^\delta) \in \partial\Theta(x_n(\alpha, y^\delta)) \quad \text{for all } \alpha > 0.$$

Moreover, by the minimality of $x_n(\alpha, y^\delta)$, we always have

$$\|y^\delta - F(x_n^\delta) - L(x_n^\delta)(x_n(\alpha, y^\delta) - x_n^\delta)\| \leq \|y^\delta - F(x_n^\delta)\|, \quad \forall \alpha > 0. \tag{3.3}$$

In order to prove that Algorithm 1 is well-defined, we need to show that the number $\alpha_n(y^\delta)$ used to define x_{n+1}^δ from x_n^δ exists, each x_n^δ is in $\mathcal{D}(F)$, and the iteration terminates after $n_\delta < \infty$ steps. We achieve these via a series of results.

Lemma 3.2 *Let \mathcal{X} be reflexive and let $\Theta : \mathcal{X} \rightarrow (-\infty, \infty]$ be proper, weakly lower semi-continuous and uniformly convex. Then, for each $\alpha > 0$, $x_n(\alpha, y^\delta)$ is uniquely determined. Moreover, the mapping $\alpha \rightarrow x_n(\alpha, y^\delta)$ is continuous over $(0, \infty)$, and the function*

$$\alpha \rightarrow \|y^\delta - F(x_n^\delta) - L(x_n^\delta)(x_n(\alpha, y^\delta) - x_n^\delta)\| \tag{3.4}$$

is continuous and monotonically increasing over $(0, \infty)$.

Proof All assertions, except the monotonicity of (3.4), follow from Lemma 2.2. The monotonicity of (3.4) can be proved by a standard argument, see [3, Lemma 9.2.1] or [7, Lemma 6.1] for instance. \square

Lemma 3.3 *Let $\Theta : \mathcal{X} \rightarrow (-\infty, \infty]$ be a proper, weakly lower semi-continuous function that is uniformly convex in the sense of (2.2). Let Assumption 1 hold with $0 \leq \eta < 1$. Let $\eta < \mu_0 < 1$ and $\tau > (1 + \eta)/(\mu_0 - \eta)$. Assume that x_n^δ and ξ_n^δ are well-defined for some $0 \leq n < n_\delta$ with*

$$D_{\xi_n^\delta} \Theta(x^\dagger, x_n^\delta) \leq D_{\xi_0} \Theta(x^\dagger, x_0) \leq \varphi(\rho). \tag{3.5}$$

Then for any $\alpha > 0$ such that

$$\|y^\delta - F(x_n^\delta) - L(x_n^\delta)(x_n(\alpha, y^\delta) - x_n^\delta)\| \geq \mu_0 \|y^\delta - F(x_n^\delta)\| \tag{3.6}$$

there hold $x_n(\alpha, y^\delta) \in B_{2\rho}(x_0)$ and

$$D_{\xi_n(\alpha, y^\delta)} \Theta(\hat{x}, x_n(\alpha, y^\delta)) - D_{\xi_n^\delta} \Theta(\hat{x}, x_n^\delta) \leq -\frac{c_0 \mu_0^r}{\alpha} \|y^\delta - F(x_n^\delta)\|^r \tag{3.7}$$

for any solution \hat{x} of (1.1) in $B_{2\rho}(x_0) \cap \mathcal{D}(\Theta)$, where $c_0 := 1 - (1 + \eta + \tau\eta)/(\tau\mu_0)$.

Proof For simplicity of exposition, we write

$$x_n(\alpha) := x_n(\alpha, y^\delta), \quad \xi_n(\alpha) := \xi_n(\alpha, y^\delta) \quad \text{and} \quad L_n := L(x_n^\delta).$$

By using the identity (2.1) and the nonnegativity of the Bregman distance, we obtain

$$D_{\xi_n(\alpha)} \Theta(\hat{x}, x_n(\alpha)) - D_{\xi_n^\delta} \Theta(\hat{x}, x_n^\delta) \leq \langle \xi_n(\alpha) - \xi_n^\delta, x_n(\alpha) - \hat{x} \rangle.$$

By the definition of $\xi_n(\alpha)$ we then have

$$\begin{aligned} & D_{\xi_n(\alpha)} \Theta(\hat{x}, x_n(\alpha)) - D_{\xi_n^\delta} \Theta(\hat{x}, x_n^\delta) \\ & \leq \frac{1}{\alpha} \langle J_r^{\mathcal{Y}}(y^\delta - F(x_n^\delta) - L_n(x_n(\alpha) - x_n^\delta)), L_n(x_n(\alpha) - \hat{x}) \rangle. \end{aligned}$$

We can write

$$L_n(x_n(\alpha) - \hat{x}) = [y^\delta - F(x_n^\delta) - L_n(\hat{x} - x_n^\delta)] - [y^\delta - F(x_n^\delta) - L_n(x_n(\alpha) - x_n^\delta)].$$

Then, by virtue of the property of the duality mapping $J_r^{\mathcal{Y}}$, we obtain

$$\begin{aligned} D_{\xi_n(\alpha)}\Theta(\hat{x}, x_n(\alpha)) - D_{\xi_n^\delta}\Theta(\hat{x}, x_n^\delta) &\leq -\frac{1}{\alpha}\|y^\delta - F(x_n^\delta) - L_n(x_n(\alpha) - x_n^\delta)\|^r \\ &+ \frac{1}{\alpha}\|y^\delta - F(x_n^\delta) - L_n(x_n(\alpha) - x_n^\delta)\|^{r-1}\|y^\delta - F(x_n^\delta) - L_n(\hat{x} - x_n^\delta)\|. \end{aligned}$$

In view of (3.5) and (2.3), we have $\|x_n^\delta - x^\dagger\| \leq \rho$ and $\|x^\dagger - x_0\| \leq \rho$ which implies that $x_n^\delta \in B_{2\rho}(x_0)$. Thus we may use (3.2) and Assumption 1(d) to derive that

$$\|y^\delta - F(x_n^\delta) - L_n(\hat{x} - x_n^\delta)\| \leq (1 + \eta)\delta + \eta\|y^\delta - F(x_n^\delta)\|.$$

Since $n < n_\delta$ we have $\|F(x_n^\delta) - y^\delta\| \geq \tau\delta$. Thus

$$\|y^\delta - F(x_n^\delta) - L_n(\hat{x} - x_n^\delta)\| \leq \frac{1 + \eta + \tau\eta}{\tau}\|y^\delta - F(x_n^\delta)\|. \tag{3.8}$$

Therefore

$$\begin{aligned} D_{\xi_n(\alpha)}\Theta(\hat{x}, x_n(\alpha)) - D_{\xi_n^\delta}\Theta(\hat{x}, x_n^\delta) &\leq -\frac{1}{\alpha}\|y^\delta - F(x_n^\delta) - L_n(x_n(\alpha) - x_n^\delta)\|^r \\ &+ \frac{1 + \eta + \tau\eta}{\tau\alpha}\|y^\delta - F(x_n^\delta) - L_n(x_n(\alpha) - x_n^\delta)\|^{r-1}\|y^\delta - F(x_n^\delta)\|. \end{aligned}$$

In view of the inequality (3.6), we thus obtain

$$D_{\xi_n(\alpha)}\Theta(\hat{x}, x_n(\alpha)) - D_{\xi_n^\delta}\Theta(\hat{x}, x_n^\delta) \leq -\frac{c_0}{\alpha}\|y^\delta - F(x_n^\delta) - L_n(x_n(\alpha) - x_n^\delta)\|^r,$$

where $c_0 := 1 - (1 + \eta + \tau\eta)/(\tau\mu_0)$. According to the conditions on μ_0 and τ , we have $c_0 > 0$. Thus, in view of (3.6) again, we obtain (3.7).

Finally, by using (3.7) with $\hat{x} = x^\dagger$ and (3.5) we have

$$D_{\xi_n(\alpha)}\Theta(x^\dagger, x_n(\alpha)) \leq D_{\xi_n^\delta}\Theta(x^\dagger, x_n^\delta) \leq D_{\xi_0}\Theta(x^\dagger, x_0) \leq \varphi(\rho).$$

This together with (2.3) and $\|x_0 - x^\dagger\| \leq \rho$ implies that $x_n(\alpha) \in B_{2\rho}(x_0)$. □

Proposition 3.4 *Let $\Theta : \mathcal{X} \rightarrow (-\infty, \infty]$ be proper, lower semi-continuous and uniformly convex in the sense of (2.2). Let Assumption 1 hold with $0 \leq \eta < 1/3$ and let $\eta < \mu_0 \leq \mu_1 < 1 - 2\eta$ and $\tau > (1 + \eta)/(\mu_0 - \eta)$. Assume that*

$$D_{\xi_0}\Theta(x^\dagger, x_0) \leq \varphi(\rho). \tag{3.9}$$

Then x_n^δ are well-defined for all $0 \leq n \leq n_\delta$ and Algorithm 1 terminates after $n_\delta < \infty$ iterations with $n_\delta = O(1 + |\log \delta|)$. Moreover, for any solution \hat{x} of (1.1) in $B_{2\rho}(x_0) \cap \mathcal{D}(\Theta)$ there hold

$$D_{\xi_{n+1}^\delta} \Theta(\hat{x}, x_{n+1}^\delta) \leq D_{\xi_n^\delta} \Theta(\hat{x}, x_n^\delta) \tag{3.10}$$

and

$$\frac{1}{\alpha_n(y^\delta)} \|y^\delta - F(x_n^\delta)\|^r \leq C_0(D_{\xi_n^\delta} \Theta(\hat{x}, x_n^\delta) - D_{\xi_{n+1}^\delta} \Theta(\hat{x}, x_{n+1}^\delta)) \tag{3.11}$$

for all $0 \leq n < n_\delta$, where $C_0 = 1/(c_0\mu_0^r)$.

Proof We first show that if

$$D_{\xi_n^\delta} \Theta(x^\dagger, x_n^\delta) \leq D_{\xi_0} \Theta(x^\dagger, x_0) \leq \varphi(\rho) \tag{3.12}$$

for some $0 \leq n < n_\delta$, then there exists $\alpha_n(y^\delta) > 0$ such that

$$\mu_0 \|y^\delta - F(x_n^\delta)\| \leq f(\alpha_n(y^\delta)) \leq \mu_1 \|y^\delta - F(x_n^\delta)\|, \tag{3.13}$$

where

$$f(\alpha) = \|y^\delta - F(x_n^\delta) - L(x_n^\delta)(x_n(\alpha, y^\delta) - x_n^\delta)\|$$

which is continuous and monotonically increasing, see Lemma 3.2. To see this, we may use the minimality of $x_n(\alpha, y^\delta)$ to obtain

$$\alpha D_{\xi_n^\delta} \Theta(x_n(\alpha, y^\delta), x_n^\delta) \leq \frac{1}{r} \|y^\delta - F(x_n^\delta)\|^r, \quad \forall \alpha > 0.$$

This implies that $\lim_{\alpha \rightarrow \infty} D_{\xi_n^\delta} \Theta(x_n(\alpha, y^\delta), x_n^\delta) = 0$ and hence $\lim_{\alpha \rightarrow \infty} \|x_n(\alpha, y^\delta) - x_n^\delta\| = 0$ by the uniform convexity of Θ . Consequently

$$\lim_{\alpha \rightarrow \infty} f(\alpha) = \|y^\delta - F(x_n^\delta)\| > \mu_1 \|y^\delta - F(x_n^\delta)\|.$$

To show the existence of a finite $\alpha_n(y^\delta)$ satisfying (3.13), it suffices to show that

$$\lim_{\alpha \rightarrow 0} f(\alpha) < \mu_0 \|y^\delta - F(x_n^\delta)\|.$$

Suppose this is not true, then

$$f(\alpha) \geq \mu_0 \|y^\delta - F(x_n^\delta)\|, \quad \forall \alpha > 0.$$

Thus, we may use Lemma 3.3 to obtain

$$\frac{c_0\mu_0^r}{\alpha} \|y^\delta - F(x_n^\delta)\|^r \leq D_{\xi_n^\delta} \Theta(x^\dagger, x_n^\delta), \quad \forall \alpha > 0.$$

Taking $\alpha \rightarrow 0$ gives $y^\delta = F(x_n^\delta)$ which is absurd since $\|y^\delta - F(x_n^\delta)\| > \tau\delta$ for $n < n_\delta$.

We next show (3.12) by induction on n . It is trivial for $n = 0$. Assume that it is true for $n = m$ with $m < n_\delta$. Since $x_{m+1}^\delta = x_m(\alpha_m(y^\delta), y^\delta)$ and $\xi_{m+1}^\delta = \xi_m(\alpha_m(y^\delta), y^\delta)$, we may use Lemma 3.3 to conclude that

$$D_{\xi_{m+1}^\delta} \Theta(x^\dagger, x_{m+1}^\delta) \leq D_{\xi_m^\delta} \Theta(x^\dagger, x_m^\delta)$$

which together with the induction hypothesis shows (3.12) for $n = m + 1$.

Since (3.12) holds true for all $0 \leq n < n_\delta$, we may use Lemma 3.3 to obtain (3.10) and (3.11) immediately.

Finally, the finiteness of n_δ can be proved by a standard argument from [11]. For completeness we include the argument here. By Assumption 1(d) and the definition of x_{n+1}^δ we have for all $0 \leq n < n_\delta$ that

$$\begin{aligned} \|y^\delta - F(x_{n+1}^\delta)\| &\leq \|y^\delta - F(x_n^\delta) - L(x_n^\delta)(x_{n+1}^\delta - x_n^\delta)\| \\ &\quad + \|F(x_{n+1}^\delta) - F(x_n^\delta) - L(x_n^\delta)(x_{n+1}^\delta - x_n^\delta)\| \\ &\leq \mu_1 \|y^\delta - F(x_n^\delta)\| + \eta \|F(x_{n+1}^\delta) - F(x_n^\delta)\| \\ &\leq (\mu_1 + \eta) \|y^\delta - F(x_n^\delta)\| + \eta \|y^\delta - F(x_{n+1}^\delta)\|. \end{aligned}$$

This implies that

$$\|y^\delta - F(x_{n+1}^\delta)\| \leq q \|y^\delta - F(x_n^\delta)\| \quad \text{with } q = \frac{\mu_1 + \eta}{1 - \eta} < 1 \tag{3.14}$$

and hence

$$\|y^\delta - F(x_n^\delta)\| \leq q^n \|y^\delta - F(x_0)\|, \quad 0 \leq n < n_\delta. \tag{3.15}$$

If $n_\delta = \infty$, then we must have $\|y^\delta - F(x_n^\delta)\| > \tau\delta$ for all n . But the inequality (3.15) implies $\|y^\delta - F(x_n^\delta)\| \rightarrow 0$ as $n \rightarrow \infty$. Therefore $n_\delta < \infty$. Now we take $n = n_\delta - 1$ in (3.15) and obtain $q^{n_\delta-1} \|y^\delta - F(x_0)\| > \tau\delta$. This implies $n_\delta = O(1 + |\log \delta|)$. \square

4 Convergence analysis

In this section we will show that Algorithm 1 is a regularization method for solving (1.1), that is, we will show that if $(x_n^\delta, \xi_n^\delta) \in \mathcal{X} \times \mathcal{X}^*$, $0 \leq n \leq n_\delta$, are defined by Algorithm 1 using noisy data y^δ , then $x_{n_\delta}^\delta$ converges to x^\dagger as $y^\delta \rightarrow y$. To this end, it is necessary to investigate for each fixed n the behavior of x_n^δ as $y^\delta \rightarrow y$. This leads us to consider the counterpart of Algorithm 1 with exact data which is formulated as Algorithm 2 below. Due to the non-unique determination of α_n , this algorithm actually defines many distinct iterative sequences. We will show that every sequence defined by Algorithm 2 is convergent. This convergence result however is not enough for showing the regularization property of Algorithm 1. Indeed, for each fixed $n \geq 1$, when taking $y^\delta \rightarrow y$, the sequence $\{\alpha_{n-1}(y^\delta)\}$ used to define x_n^δ from x_{n-1}^δ may split into many

convergent subsequences and so does $\{x_n^\delta, \xi_n^\delta\}$ with limits defined by Algorithm 2. This forces us to establish a uniform convergence result for all the possible sequences defined by Algorithm 2. The regularization property of Algorithm 1 is then established by using a stability result which connects Algorithms 1 and 2.

4.1 The method with exact data

We start with the formulation of the counterpart of Algorithm 1 when the exact data is used.

Algorithm 2 (Levenberg–Marquardt method with exact data)

1. Let $0 < \mu_0 \leq \mu_1 < 1$ and $(x_0, \xi_0) \in \mathcal{X} \times \mathcal{X}^*$ be the same as in Algorithm 1.
2. Assume that x_n and ξ_n are defined. If $F(x_n) = y$, we define $x_{n+1} = x_n$ and $\xi_{n+1} = \xi_n$; otherwise, we define x_{n+1} and ξ_{n+1} as follows:
 - (a) For each $\alpha > 0$ we define $x_n(\alpha, y)$ and $\xi_n(\alpha, y)$ as

$$x_n(\alpha, y) = \arg \min_{x \in \mathcal{X}} \left\{ \frac{1}{r} \|y - F(x_n) - L(x_n)(x - x_n)\|^r + \alpha D_{\xi_n} \Theta(x, x_n) \right\},$$

$$\xi_n(\alpha, y) = \xi_n + \frac{1}{\alpha} L(x_n)^* J_r^{\mathcal{Y}} (y - F(x_n) - L(x_n)(x_n(\alpha, y) - x_n));$$

- (b) Take $\alpha_n > 0$ to be a number such that

$$\mu_0 \|y - F(x_n)\| \leq \|y - F(x_n) - L(x_n)(x_n(\alpha_n, y) - x_n)\| \leq \mu_1 \|y - F(x_n)\|;$$

- (c) Define $x_{n+1} := x_n(\alpha_n, y)$ and $\xi_{n+1} := \xi_n(\alpha_n, y)$.

In the formulation of Algorithm 2, we take $\alpha_n > 0$ to be any number satisfying (b) when defining x_{n+1}, ξ_{n+1} from x_n, ξ_n . There might have many possible choices of α_n ; different choice of $\{\alpha_n\}$ may lead to different iterative sequence. We will use $\Gamma_{\mu_0, \mu_1}(x_0, \xi_0)$ to denote the set of all possible sequence $\{(x_n, \xi_n)\}$ in $\mathcal{X} \times \mathcal{X}^*$ constructed from (x_0, ξ_0) by Algorithm 2 with $\alpha_n > 0$ chosen to be any number satisfying (b). By using the same argument in the proof of Proposition 3.4, we can obtain the following result which shows that each sequence in $\Gamma_{\mu_0, \mu_1}(x_0, \xi_0)$ is well-defined and admits certain monotonicity property.

Lemma 4.1 *Let Assumption 1 hold with $0 \leq \eta < 1$ and let $\eta < \mu_0 \leq \mu_1 < 1$. Then any sequence $\{(x_n, \xi_n)\} \in \Gamma_{\mu_0, \mu_1}(x_0, \xi_0)$ is well-defined and for any solution \hat{x} of (1.1) in $B_{2\rho}(x_0) \cap \mathcal{D}(\Theta)$ there hold*

$$D_{\xi_{n+1}} \Theta(\hat{x}, x_{n+1}) \leq D_{\xi_n} \Theta(\hat{x}, x_n), \tag{4.1}$$

$$\frac{1}{\alpha_n} \|y - F(x_n)\|^r \leq \frac{1}{\mu_0 - \eta} (D_{\xi_n} \Theta(\hat{x}, x_n) - D_{\xi_{n+1}} \Theta(\hat{x}, x_{n+1})) \tag{4.2}$$

for all $n \geq 0$.

In order to derive the convergence of every sequence $\{(x_n, \xi_n)\}$ in $\Gamma_{\mu_0, \mu_1}(x_0, \xi_0)$, we will use the following result which gives a general convergence criterion.

Proposition 4.2 *Consider the Eq. (1.1) for which Assumption 1 holds. Let $\Theta : \mathcal{X} \rightarrow (-\infty, \infty]$ be a proper, lower semi-continuous and uniformly convex function. Let $\{x_n\} \subset B_{2\rho}(x_0) \cap \mathcal{D}(\Theta)$ and $\{\xi_n\} \subset \mathcal{X}^*$ be such that*

- (i) $\xi_n \in \partial\Theta(x_n)$ for all n ;
- (ii) for any solution \hat{x} of (1.1) in $B_{2\rho}(x_0) \cap \mathcal{D}(\Theta)$ the sequence $\{D_{\xi_n}\Theta(\hat{x}, x_n)\}$ is monotonically decreasing;
- (iii) $\lim_{n \rightarrow \infty} \|F(x_n) - y\| = 0$.
- (iv) there is a constant C such that for all $k > n$ and any solution \hat{x} of (1.1) in $B_{2\rho}(x_0) \cap \mathcal{D}(\Theta)$ there holds

$$|\langle \xi_k - \xi_n, x_k - \hat{x} \rangle| \leq C(D_{\xi_n}\Theta(\hat{x}, x_n) - D_{\xi_k}\Theta(\hat{x}, x_k)). \tag{4.3}$$

Then there exists a solution x_* of (1.1) in $B_{2\rho}(x_0) \cap \mathcal{D}(\Theta)$ such that

$$\lim_{n \rightarrow \infty} \|x_n - x_*\| = 0, \quad \lim_{n \rightarrow \infty} \Theta(x_n) = \Theta(x_*) \quad \text{and} \quad \lim_{n \rightarrow \infty} D_{\xi_n}\Theta(x_*, x_n) = 0.$$

If, in addition, $x^\dagger \in B_\rho(x_0) \cap \mathcal{D}(\Theta)$ and $\xi_{n+1} - \xi_n \in \overline{\mathcal{R}(L(x^\dagger)^*)}$ for all n , then $x_* = x^\dagger$.

Proof This result follows from [20, Proposition 3.6] and its proof. □

Now we can prove the main convergence result on Algorithm 2.

Theorem 4.3 *Let Assumption 1 hold with $0 \leq \eta < 1/3$ and let $\eta < \mu_0 \leq \mu_1 < 1 - 2\eta$. Let $\Theta : \mathcal{X} \rightarrow (-\infty, \infty]$ be proper, lower semi-continuous and uniformly convex in the sense of (2.2), and let (3.9) be satisfied. Then for any $\{(x_n, \xi_n)\} \in \Gamma_{\mu_0, \mu_1}(x_0, \xi_0)$ there exists a solution x_* of (1.1) in $B_{2\rho}(x_0) \cap \mathcal{D}(\Theta)$ such that*

$$\lim_{n \rightarrow \infty} \|x_n - x_*\| = 0, \quad \lim_{n \rightarrow \infty} \Theta(x_n) = \Theta(x_*) \quad \text{and} \quad \lim_{n \rightarrow \infty} D_{\xi_n}\Theta(x_*, x_n) = 0.$$

If in addition $\mathcal{N}(L(x^\dagger)) \subset \mathcal{N}(L(x))$ for all $x \in B_{2\rho}(x_0) \cap \mathcal{D}(F)$, then $x_* = x^\dagger$.

Proof We will use Proposition 4.2. By the definition of ξ_n in Algorithm 2 we always have $\xi_n \in \partial\Theta(x_n)$ for all $n \geq 0$ which shows (i) in Proposition 4.2. Lemma 4.1 shows (ii) in Proposition 4.2. By the similar argument for deriving (3.14) we can show that

$$\|y - F(x_{n+1})\| \leq q\|y - F(x_n)\|, \quad \forall n \geq 0 \tag{4.4}$$

with $q = (\mu_1 + \eta)/(1 - \eta) < 1$. This implies (iii) in Proposition 4.2. In order to show the convergence result, it remains only to show (iv) in Proposition 4.2.

To this end, for $0 \leq l < m < \infty$ we may use the definition of ξ_n and the property of the duality mapping $J_r^{\mathcal{Y}}$ to obtain

$$\begin{aligned} |\langle \xi_m - \xi_l, x_m - x^\dagger \rangle| &= \left| \sum_{n=l}^{m-1} \langle \xi_{n+1} - \xi_n, x_m - x^\dagger \rangle \right| \\ &= \left| \sum_{n=l}^{m-1} \frac{1}{\alpha_n} \langle J_r^{\mathcal{Y}}(y - F(x_n) - L(x_n)(x_{n+1} - x_n)), L(x_n)(x_m - x^\dagger) \rangle \right| \\ &\leq \sum_{n=l}^{m-1} \frac{1}{\alpha_n} \|y - F(x_n) - L(x_n)(x_{n+1} - x_n)\|^{r-1} \|L(x_n)(x_m - x^\dagger)\|. \end{aligned}$$

By the triangle inequality $\|L(x_n)(x_m - x^\dagger)\| \leq \|L(x_n)(x_n - x^\dagger)\| + \|L(x_n)(x_m - x_n)\|$ and Assumption 1(d), we have

$$\begin{aligned} \|L(x_n)(x_m - x^\dagger)\| &\leq (1 + \eta) (\|y - F(x_n)\| + \|F(x_n) - F(x_m)\|) \\ &\leq (1 + \eta) (2\|y - F(x_n)\| + \|y - F(x_m)\|). \end{aligned}$$

Since the inequality (4.4) implies that $\{\|y - F(x_n)\|\}$ monotonically decreasing, we have

$$\|L(x_n)(x_m - x^\dagger)\| \leq 3(1 + \eta)\|y - F(x_n)\|, \quad 0 \leq n \leq m.$$

Therefore

$$\begin{aligned} |\langle \xi_m - \xi_l, x_m - x^\dagger \rangle| &\leq 3(1 + \eta) \sum_{n=l}^{m-1} \frac{1}{\alpha_n} \|y - F(x_n) - L(x_n)(x_{n+1} - x_n)\|^{r-1} \|y - F(x_n)\| \\ &\leq 3(1 + \eta) \sum_{n=l}^{m-1} \frac{1}{\alpha_n} \|y - F(x_n)\|^r. \end{aligned}$$

In view of (4.2) in Lemma 4.1, we obtain with $C_1 := 3(1 + \eta)/(\mu_0 - \eta)$ that

$$|\langle \xi_m - \xi_l, x_m - x^\dagger \rangle| \leq C_1(D_{\xi_l}\Theta(x^\dagger, x_l) - D_{\xi_m}\Theta(x^\dagger, x_m))$$

which shows (iv) in Proposition 4.2.

To show the last part under the condition $\mathcal{N}(L(x^\dagger)) \subset \mathcal{N}(L(x))$ for $x \in B_{2\rho}(x_0) \cap \mathcal{D}(\Theta)$, we observe from the definition of ξ_n that

$$\xi_{n+1} - \xi_n \in \mathcal{R}(L(x_n)^*) \subset \mathcal{N}(L(x_n))^\perp \subset \mathcal{N}(L(x^\dagger))^\perp = \overline{\mathcal{R}(L(x^\dagger)^*)}.$$

Thus, we may use the second part of Proposition 4.2 to conclude the proof. □

4.2 A uniform convergence result

In Theorem 4.3 we have shown the convergence of every sequence in $\Gamma_{\mu_0, \mu_1}(x_0, \xi_0)$. In this subsection we will strengthen this result by showing the following uniform convergence result for all sequences in $\Gamma_{\mu_0, \mu_1}(x_0, \xi_0)$ which will be crucial in establishing the regularization property of Algorithm 1.

Proposition 4.4 *Assume all the conditions in Theorem 4.3 hold. Assume also that*

$$\mathcal{N}(L(x^\dagger)) \subset \mathcal{N}(L(x)), \quad \forall x \in B_{2\rho}(x_0) \cap \mathcal{D}(\Theta). \tag{4.5}$$

Then, for any $\varepsilon > 0$, there is an integer $n(\varepsilon)$ such that for any sequence $\{(\xi_n, x_n)\} \in \Gamma_{\mu_0, \mu_1}(\xi_0, x_0)$ there holds $D_{\xi_n} \Theta(x^\dagger, x_n) < \varepsilon$ for all $n \geq n(\varepsilon)$.

The proof of Proposition 4.4 is based on some preliminary results. It is easily seen that, to complete the proof, we only need to consider the case $F(x_0) \neq y$. The following result shows that in this case we always have $F(x_n) \neq y$ for all $n \geq 0$ for any sequence $\{(x_n, \xi_n)\} \in \Gamma_{\mu_0, \mu_1}(x_0, \xi_0)$. Thus, when defining x_{n+1} from x_n in Algorithm 2 we always use a finite number $\alpha_n > 0$.

Lemma 4.5 *Let all the conditions in Proposition 4.4 hold. For any sequence $\{(x_n, \xi_n)\} \in \Gamma_{\mu_0, \mu_1}(x_0, \xi_0)$, if $F(x_n) = y$ for some $n \geq 1$, then $F(x_0) = y$.*

Proof It suffices to show that if $F(x_k) = y$ for some $1 \leq k \leq n$, then $F(x_{k-1}) = y$. By using Assumption 1(d) and $F(x_k) = y$ we have $L(x^\dagger)(x_k - x^\dagger) = 0$. Thus, in view of (4.5), we have

$$x_k - x^\dagger \in \mathcal{N}(L(x^\dagger)) \subset \mathcal{N}(L(x_{k-1})).$$

Consequently $L(x_{k-1})(x_k - x^\dagger) = 0$. If $F(x_{k-1}) \neq y$, then by the definition of x_k and Assumption 1(d) we have

$$\begin{aligned} \mu_0 \|y - F(x_{k-1})\| &\leq \|y - F(x_{k-1}) - L(x_{k-1})(x_k - x_{k-1})\| \\ &= \|y - F(x_{k-1}) - L(x_{k-1})(x^\dagger - x_{k-1})\| \\ &\leq \eta \|y - F(x_{k-1})\| \end{aligned}$$

which is impossible since $\mu_0 > \eta$. The proof is thus complete. □

The next result shows that, if $F(x_n) \neq y$, then we can give the upper and lower bounds on the number α_n used to define x_{n+1} from x_n .

Lemma 4.6 *Let all the conditions in Theorem 4.3 hold. If $F(x_n) \neq y$, then for any α satisfying*

$$\mu_0 \|y - F(x_n)\| \leq \|y - F(x_n) - L(x_n)(x_n(\alpha, y) - x_n)\| \leq \mu_1 \|y - F(x_n)\| \tag{4.6}$$

there holds $0 < \underline{\alpha}_n \leq \alpha \leq \bar{\alpha}_n < \infty$, where

$$\underline{\alpha}_n := \frac{(\mu_0^r - \eta^r) \|y - F(x_n)\|^r}{r\varphi(\rho)} \quad \text{and} \quad \bar{\alpha}_n := \frac{\|y - F(x_n)\|^r}{r\varphi((1 - \mu_1)\|y - F(x_n)\|/\|L(x_n)\|)}.$$

Proof By the definition of $x_n(\alpha, y)$ and the uniform convexity of Θ , we have

$$\alpha\varphi(\|x_n(\alpha, y) - x_n\|) \leq \alpha D_{\xi_n} \Theta(x_n(\alpha, y), x_n) \leq \frac{1}{r} \|y - F(x_n)\|^r.$$

In view of the second inequality in (4.6) we can obtain

$$\|L(x_n)\| \|x_n(\alpha, y) - x_n\| \geq \|L(x_n)(x_n(\alpha, y) - x_n)\| \geq (1 - \mu_1) \|y - F(x_n)\|.$$

Consequently

$$\alpha\varphi\left(\frac{(1 - \mu_1)\|y - F(x_n)\|}{\|L(x_n)\|}\right) \leq \frac{1}{r} \|y - F(x_n)\|^r \tag{4.7}$$

which implies that $\alpha \leq \bar{\alpha}_n$. On the other hand, by the definition of $x_n(\alpha, y)$ we have

$$\begin{aligned} & \frac{1}{r} \|y - F(x_n) - L(x_n)(x_n(\alpha, y) - x_n)\|^r \\ & \leq \frac{1}{r} \|y - F(x_n) - L(x_n)(x^\dagger - x_n)\|^r + \alpha D_{\xi_n} \Theta(x^\dagger, x_n). \end{aligned}$$

In view of the first inequality in (4.6), Assumption 1(d), and the inequality $D_{\xi_n} \Theta(x^\dagger, x_n) \leq D_{\xi_0} \Theta(x^\dagger, x_0) \leq \varphi(\rho)$ from Lemma 4.1, it follows that

$$\mu_0^r \|y - F(x_n)\|^r \leq \eta^r \|y - F(x_n)\|^r + r\alpha\varphi(\rho).$$

This implies that $\alpha \geq \underline{\alpha}_n$. □

Now we are ready to give the proof of Proposition 4.4. We will use an idea from [10] which is based on the well-known diagonal sequence argument.

Proof of Proposition 4.4 We may assume that $F(x_0) \neq y$. We will use a contradiction argument. Assume that the result is not true. Then there is an $\varepsilon_0 > 0$ such that for any $\ell \geq 1$ there exist $\{(x_n^{(\ell)}, \xi_n^{(\ell)})\} \in \Gamma_{\mu_0, \mu_1}(x_0, \xi_0)$ and $n_\ell > \ell$ such that

$$D_{\xi_n^{(\ell)}} \Theta(x^\dagger, x_{n_\ell}^{(\ell)}) \geq \varepsilon_0. \tag{4.8}$$

We will construct, for each $n = 0, 1, \dots$, a strictly increasing subsequence $\{\ell_{n,k}\}$ of positive integers and $(\bar{x}_n, \bar{\xi}_n) \in \mathcal{X} \times \mathcal{X}^*$ such that

- (i) $\{(\bar{x}_n, \bar{\xi}_n)\} \in \Gamma_{\mu_0, \mu_1}(x_0, \xi_0)$;
- (ii) for each fixed n there hold $x_n^{(\ell_{n,k})} \rightarrow \bar{x}_n$ and $\xi_n^{(\ell_{n,k})} \rightarrow \bar{\xi}_n$ as $k \rightarrow \infty$. Moreover, for all k there hold

$$D_{\xi_n^{(\ell_{n,k})}} \Theta(\bar{x}_n, x_n^{(\ell_{n,k})}) \leq \varepsilon_0/4 \quad \text{and} \quad |\langle \xi_n^{(\ell_{n,k})} - \bar{\xi}_n, \bar{x}_n - x^\dagger \rangle| \leq \varepsilon_0/4.$$

Assume that the above construction is available, we will derive a contradiction. According to (i), we may use Theorem 4.3 to conclude that $D_{\bar{\xi}_n} \Theta(x^\dagger, \bar{x}_n) \rightarrow 0$ as $n \rightarrow \infty$. Thus we can pick a large integer \hat{n} such that

$$D_{\bar{\xi}_{\hat{n}}} \Theta(x^\dagger, \bar{x}_{\hat{n}}) < \varepsilon_0/2.$$

Let $\hat{\ell} := \ell_{\hat{n}, \hat{n}}$ and consider the sequence $\{(x_n^{(\hat{\ell})}, \xi_n^{(\hat{\ell})})\}$. According to (ii), we have

$$\begin{aligned} \varepsilon_0/2 &> (D_{\bar{\xi}_{\hat{n}}} \Theta(x^\dagger, \bar{x}_{\hat{n}}) - D_{\xi_{\hat{n}}^{(\hat{\ell})}} \Theta(x^\dagger, x_{\hat{n}}^{(\hat{\ell})})) + D_{\xi_{\hat{n}}^{(\hat{\ell})}} \Theta(x^\dagger, x_{\hat{n}}^{(\hat{\ell})}) \\ &= -D_{\xi_{\hat{n}}^{(\hat{\ell})}} \Theta(\bar{x}_{\hat{n}}, x_{\hat{n}}^{(\hat{\ell})}) + \langle \bar{\xi}_{\hat{n}} - \xi_{\hat{n}}^{(\hat{\ell})}, \bar{x}_{\hat{n}} - x^\dagger \rangle + D_{\xi_{\hat{n}}^{(\hat{\ell})}} \Theta(x^\dagger, x_{\hat{n}}^{(\hat{\ell})}) \\ &\geq -\varepsilon_0/4 - \varepsilon_0/4 + D_{\xi_{\hat{n}}^{(\hat{\ell})}} \Theta(x^\dagger, x_{\hat{n}}^{(\hat{\ell})}). \end{aligned}$$

Since $\{\ell_{\hat{n}, k}\}$ is strictly increasing, we have $n_{\hat{\ell}} > \hat{\ell} = \ell_{\hat{n}, \hat{n}} \geq \hat{n}$. Therefore, we may use Lemma 4.1 to obtain

$$D_{\xi_{n_{\hat{\ell}}}^{(\hat{\ell})}} \Theta(x^\dagger, x_{n_{\hat{\ell}}}^{(\hat{\ell})}) \leq D_{\xi_{\hat{n}}^{(\hat{\ell})}} \Theta(x^\dagger, x_{\hat{n}}^{(\hat{\ell})}) < \varepsilon_0$$

which is a contradiction to (4.8) with $\ell = \hat{\ell}$.

We turn to the construction of $\{\ell_{n,k}\}$ and $(\bar{x}_n, \bar{\xi}_n)$, for each $n = 0, 1, \dots$, such that (i) and (ii) hold. For $n = 0$, we take $(\bar{x}_0, \bar{\xi}_0) = (x_0, \xi_0)$ and $\ell_{0,k} = k$ for all k . Since $\xi_0^{(k)} = \xi_0$ and $x_0^{(k)} = x_0$, (ii) holds automatically for $n = 0$.

Next, assume that we have constructed $\{\ell_{n,k}\}$ and $(\bar{x}_n, \bar{\xi}_n)$ for all $0 \leq n \leq m$. We will construct $\{\ell_{m+1,k}\}$ and $(\bar{x}_{m+1}, \bar{\xi}_{m+1})$. Since $F(x_0) \neq y$, we have from Lemma 4.5 that $F(\bar{x}_m) \neq y$ and $F(x_m^{(\ell)}) \neq y$ for all ℓ . Let $\alpha_m^{(\ell)} > 0$ be the number used to define $(x_{m+1}^{(\ell)}, \xi_{m+1}^{(\ell)})$ from $(x_m^{(\ell)}, \xi_m^{(\ell)})$. From Lemma 4.6 and the induction hypothesis $x_m^{(\ell_{m,k})} \rightarrow \bar{x}_m$ we can conclude that there are two positive numbers $\underline{\alpha}_m$ and $\bar{\alpha}_m$ independent of k such that $\underline{\alpha}_m \leq \alpha_m^{(\ell_{m,k})} \leq \bar{\alpha}_m$ for all k . Thus $\{\ell_{m,k}\}$ must have a subsequence, denoted as $\{\ell_{m+1,k}\}$, such that $\{\alpha_m^{(\ell_{m+1,k})}\}$ converges to some number $\alpha_m \in (0, \infty)$ as $k \rightarrow \infty$. We define

$$\bar{x}_{m+1} = \arg \min_{x \in \mathcal{X}} \left\{ \frac{1}{r} \|y - F(\bar{x}_m) - L(\bar{x}_m)(x - \bar{x}_m)\|^r + \alpha_m D_{\bar{\xi}_m} \Theta(x, \bar{x}_m) \right\},$$

$$\bar{\xi}_{m+1} = \bar{\xi}_m + \frac{1}{\alpha_m} L(\bar{x}_m)^* J_r^{\mathcal{Y}}(y - F(\bar{x}_m) - L(\bar{x}_m)(\bar{x}_{m+1} - \bar{x}_m)).$$

It is clear that $\bar{x}_{m+1} = \bar{x}_m(\alpha_m, y)$ and $\bar{\xi}_{m+1} = \bar{\xi}_m(\alpha_m, y)$. In view of the induction hypotheses $x_m^{(\ell_{m+1,k})} \rightarrow \bar{x}_m$ and $\xi_m^{(\ell_{m+1,k})} \rightarrow \bar{\xi}_m$, the continuity of $x \rightarrow F(x)$ and $x \rightarrow L(x)$, and $\alpha_m^{(\ell_{m+1,k})} \rightarrow \alpha_m$, we may use Lemma 2.2 and the continuity of the duality mapping $J_r^{\mathcal{Y}}$ to conclude that

$$x_{m+1}^{(\ell_{m+1,k})} \rightarrow \bar{x}_{m+1}, \quad \Theta(x_{m+1}^{(\ell_{m+1,k})}) \rightarrow \Theta(\bar{x}_{m+1}) \quad \text{and} \quad \xi_{m+1}^{(\ell_{m+1,k})} \rightarrow \bar{\xi}_{m+1} \quad (4.9)$$

as $k \rightarrow \infty$. According to the choice of $\alpha_m^{(\ell_{m+1,k})}$ and thus the definition of $x_{m+1}^{(\ell_{m+1,k})}$, we have

$$\begin{aligned} \mu_0 \|y - F(x_m^{(\ell_{m+1,k})})\| &\leq \|y - F(x_m^{(\ell_{m+1,k})}) - L(x_m^{(\ell_{m+1,k})})(x_{m+1}^{(\ell_{m+1,k})} - x_m^{(\ell_{m+1,k})})\| \\ &\leq \mu_1 \|y - F(x_m^{(\ell_{m+1,k})})\|. \end{aligned}$$

Letting $k \rightarrow \infty$ gives

$$\mu_0 \|y - F(\bar{x}_m)\| \leq \|y - F(\bar{x}_m) - L(\bar{x}_m)(\bar{x}_{m+1} - \bar{x}_m)\| \leq \mu_1 \|y - F(\bar{x}_m)\|.$$

Thus $\bar{x}_{m+1} = \bar{x}_m(\alpha_m, y)$ satisfies the desired requirement. We therefore complete the construction of $\{\ell_{m+1,k}\}$ and $(\bar{x}_{m+1}, \bar{\xi}_{m+1})$. We need to show that \bar{x}_{m+1} and $\bar{\xi}_{m+1}$ satisfy the estimates in (ii) for $n = m + 1$. We may use (4.9) to obtain

$$\lim_{k \rightarrow \infty} \langle \xi_{m+1}^{(\ell_{m+1,k})} - \bar{\xi}_{m+1}, \bar{x}_{m+1} - x^{\dagger} \rangle = 0, \quad \lim_{k \rightarrow \infty} D_{\xi_{m+1}^{(\ell_{m+1,k})}} \Theta(\bar{x}_{m+1}, x_{m+1}^{(\ell_{m+1,k})}) = 0$$

Consequently, by taking a subsequence of $\{\ell_{m+1,k}\}$ if necessary, which is still denoted by the same notation, we can guarantee (ii) for $n = m + 1$. □

4.3 Regularization property

In this section we will establish the regularization property of Algorithm 1 which is stated in the following result.

Theorem 4.7 *Let $\Theta : \mathcal{X} \rightarrow (-\infty, \infty]$ be a proper, lower semi-continuous function that is uniformly convex in the sense of (2.2), and let Assumption 1 hold with $0 \leq \eta < 1/3$. Let $\eta < \mu_0 \leq \mu_1 < 1 - 2\eta$ and $\tau > (1 + \eta)/(\mu_0 - \eta)$, and let (3.9) be satisfied. Assume further that*

$$\mathcal{N}(L(x^{\dagger})) \subset \mathcal{N}(L(x)), \quad \forall x \in B_{2\rho}(x_0) \cap \mathcal{D}(\Theta).$$

Then for $x_{n_\delta}^\delta \in \mathcal{X}$ and $\xi_{n_\delta}^\delta \in \mathcal{X}^*$ defined by Algorithm 1 there hold

$$\lim_{\delta \rightarrow 0} \|x_{n_\delta}^\delta - x^\dagger\| = 0, \quad \lim_{\delta \rightarrow 0} \Theta(x_{n_\delta}^\delta) = \Theta(x^\dagger) \quad \text{and} \quad \lim_{\delta \rightarrow 0} D_{\xi_{n_\delta}^\delta} \Theta(x^\dagger, x_{n_\delta}^\delta) = 0.$$

In order to prove Theorem 4.7, we will need to establish certain stability result to connect Algorithms 1 and 2 so that Proposition 4.4 can be used. The following stability result is sufficient for our purpose.

Lemma 4.8 *Let $F(x_0) \neq y$ and let all the conditions in Theorem 4.7 hold. Let $\{y^{\delta_l}\}$ be a sequence of noisy data satisfying $\|y^{\delta_l} - y\| \leq \delta_l$ with $\delta_l \rightarrow 0$ as $l \rightarrow \infty$. Let $x_n^{\delta_l}$ and $\xi_n^{\delta_l}$, $0 \leq n \leq n_{\delta_l}$, be defined by Algorithm 1. Then for any finite $n \leq \liminf_{l \rightarrow \infty} n_{\delta_l}$, by taking a subsequence of $\{y^{\delta_l}\}$ if necessary, there is a sequence $\{(x_m, \xi_m)\} \in \Gamma_{\mu_0, \mu_1}(x_0, \xi_0)$ such that*

$$x_m^{\delta_l} \rightarrow x_m, \quad \xi_m^{\delta_l} \rightarrow \xi_m \quad \text{and} \quad \Theta(x_m^{\delta_l}) \rightarrow \Theta(x_m) \quad \text{as } l \rightarrow \infty$$

for all $0 \leq m \leq n$.

Proof Since $F(x_0) \neq y$, we must have $\liminf_{l \rightarrow \infty} n_{\delta_l} \geq 1$. We will use an induction argument on n .

When $n = 0$, nothing needs to prove since $x_0^{\delta_l} = x_0$ and $\xi_0^{\delta_l} = \xi_0$. Assume next that, for some $0 \leq n < \liminf_{l \rightarrow \infty} n_{\delta_l}$, the result is true for some sequence $\{(x_m, \xi_m)\} \in \Gamma_{\mu_0, \mu_1}(x_0, \xi_0)$ with $0 \leq m \leq n$. In order to show the result is also true for $n + 1$, we will obtain a sequence from $\Gamma_{\mu_0, \mu_1}(x_0, \xi_0)$ by retaining the first $n + 1$ terms in $\{(x_m, \xi_m)\}$ and modifying the remaining terms. It suffices to redefine x_{n+1} and ξ_{n+1} since then we can apply Algorithm 2 to produce the remaining terms.

Since $F(x_0) \neq y$, we have from Lemma 4.5 that $F(x_n) \neq y$. Let $\alpha_n(y^{\delta_l})$ be the number used to define $x_{n+1}^{\delta_l}$ and $\xi_{n+1}^{\delta_l}$. Since the induction hypothesis $x_n^{\delta_l} \rightarrow x_n$ and the fact $y^{\delta_l} \rightarrow y$ imply $\|F(x_n^{\delta_l}) - y^{\delta_l}\| \rightarrow \|F(x_n) - y\| > 0$ as $l \rightarrow \infty$, we may use the similar argument in the proof of Lemma 4.6 to conclude that

$$\underline{\alpha}_n \leq \alpha_n(y^{\delta_l}) \leq \bar{\alpha}_n$$

for two numbers $0 < \underline{\alpha}_n \leq \bar{\alpha}_n < \infty$ independent of l . Therefore, by taking a subsequence of $\{y^{\delta_l}\}$ if necessary, we may assume that

$$\alpha_n(y^{\delta_l}) \rightarrow \alpha_n \quad \text{as } l \rightarrow \infty$$

for some number $\alpha_n \in (0, \infty)$. We define

$$x_{n+1} = \arg \min_{x \in \mathcal{X}} \left\{ \frac{1}{r} \|y - F(x_n) - L(x_n)(x - x_n)\|^r + \alpha_n D_{\xi_n} \Theta(x, x_n) \right\},$$

$$\xi_{n+1} = \xi_n + \frac{1}{\alpha_n} L(x_n)^* J_r^{\mathcal{Y}}(y - F(x_n) - L(x_n)(x_{n+1} - x_n)).$$

In view of the induction hypotheses and the continuity of $x \rightarrow F(x)$ and $x \rightarrow L(x)$, we may use Lemma 2.2 and the continuity of J_r^y to conclude that

$$x_{n+1}^{\delta_l} \rightarrow x_{n+1}, \quad \Theta(x_{n+1}^{\delta_l}) \rightarrow \Theta(x_{n+1}) \quad \text{and} \quad \xi_{n+1}^{\delta_l} \rightarrow \xi_{n+1} \quad \text{as } l \rightarrow \infty.$$

Moreover, since

$$\mu_0 \|y^{\delta_l} - F(x_n^{\delta_l})\| \leq \|y^{\delta_l} - F(x_n^{\delta_l}) - L(x_n^{\delta_l})(x_{n+1}^{\delta_l} - x_n^{\delta_l})\| \leq \mu_1 \|y^{\delta_l} - F(x_n^{\delta_l})\|,$$

by taking $l \rightarrow \infty$ we can conclude that

$$\mu_0 \|y - F(x_n)\| \leq \|y - F(x_n) - L(x_n)(x_{n+1} - x_n)\| \leq \mu_1 \|y - F(x_n)\|.$$

Therefore $x_{n+1} = x_n(\alpha_n, y)$ and $\xi_{n+1} = \xi_n(\alpha_n, y)$ are the desired elements to be defined. The proof is thus complete. □

The next result will be used to prove $\lim_{\delta \rightarrow 0} \Theta(x_{n_\delta}^\delta) = \Theta(x^\dagger)$ in Theorem 4.7.

Lemma 4.9 *Let all the conditions in Theorem 4.7 hold and let $\{(x_n^\delta, \xi_n^\delta)\}_{0 \leq n \leq n_\delta}$ be defined by Algorithm 1. Then for all $0 \leq l \leq n_\delta$ there holds*

$$|\langle \xi_{n_\delta}^\delta - \xi_l^\delta, x^\dagger - x_{n_\delta}^\delta \rangle| \leq C_2 D_{\xi_l^\delta} \Theta(x^\dagger, x_l^\delta),$$

where $C_2 = (1 + \eta)(3\tau + 1)/(\tau c_0 \mu_0^r)$.

Proof By the definition of ξ_n^δ , the property of J_r^y and (3.3), we can obtain

$$\begin{aligned} |\langle \xi_{n_\delta}^\delta - \xi_l^\delta, x^\dagger - x_{n_\delta}^\delta \rangle| &\leq \sum_{n=l}^{n_\delta-1} |\langle \xi_{n+1}^\delta - \xi_n^\delta, x^\dagger - x_{n_\delta}^\delta \rangle| \\ &\leq \sum_{n=l}^{n_\delta-1} \frac{1}{\alpha_n(y^\delta)} \|y^\delta - F(x_n^\delta) - L(x_n^\delta)(x_{n+1}^\delta - x_n^\delta)\|^{r-1} \|L(x_n^\delta)(x^\dagger - x_{n_\delta}^\delta)\| \\ &\leq \sum_{n=l}^{n_\delta-1} \frac{1}{\alpha_n(y^\delta)} \|y^\delta - F(x_n^\delta)\|^{r-1} \|L(x_n^\delta)(x^\dagger - x_{n_\delta}^\delta)\|. \end{aligned}$$

By Assumption 1(d) and (3.14) we can derive that

$$\|L(x_n^\delta)(x^\dagger - x_{n_\delta}^\delta)\| \leq (1 + \eta)(\delta + 3\|y^\delta - F(x_n^\delta)\|), \quad 0 \leq n < n_\delta.$$

Since $\|y^\delta - F(x_n^\delta)\| > \tau\delta$ for $0 \leq n < n_\delta$, we therefore have

$$\|L(x_n^\delta)(x^\dagger - x_{n_\delta}^\delta)\| \leq \frac{(1 + \eta)(3\tau + 1)}{\tau} \|y^\delta - F(x_n^\delta)\|, \quad 0 \leq n < n_\delta.$$

Consequently

$$|\langle \xi_{n_\delta}^\delta - \xi_l^\delta, x^\dagger - x_{n_\delta}^\delta \rangle| \leq \frac{(1 + \eta)(3\tau + 1)}{\tau} \sum_{n=l}^{n_\delta-1} \frac{1}{\alpha_n(y^\delta)} \|y^\delta - F(x_n^\delta)\|^r.$$

This together with (3.11) in Proposition 3.4 implies the desired estimate. □

We are now ready to prove Theorem 4.7, the main result of this paper.

Proof of Theorem 4.7. We may assume that $F(x_0) \neq y$. We first claim that

$$\lim_{\delta \rightarrow 0} n_\delta = \infty. \tag{4.10}$$

Suppose that this is not true. Then there exists $\{y^{\delta_l}\}$ satisfying $\|y^{\delta_l} - y\| \leq \delta_l$ with $\delta_l \rightarrow 0$ such that $n_{\delta_l} \rightarrow \hat{n}$ as $l \rightarrow \infty$ for some finite integer \hat{n} . Thus $n_{\delta_l} = \hat{n}$ for large l . By the definition of n_{δ_l} we have

$$\|F(x_{\hat{n}}^{\delta_l}) - y^{\delta_l}\| \leq \tau \delta_l. \tag{4.11}$$

In view of Lemma 4.8, by taking a subsequence of $\{y^{\delta_l}\}$ if necessary, we can find $\{(x_n, \xi_n)\} \in \Gamma_{\mu_0, \mu_1}(x_0, \xi_0)$ such that $x_{\hat{n}}^{\delta_l} \rightarrow x_{\hat{n}}$ as $l \rightarrow \infty$. Letting $l \rightarrow \infty$ in (4.11) gives $F(x_{\hat{n}}) = y$. Consequently, by Lemma 4.5, we must have $F(x_0) = y$ which is a contradiction.

We next show the convergence result. We first prove that

$$\lim_{\delta \rightarrow 0} D_{\xi_{n_\delta}^\delta} \Theta(x^\dagger, x_{n_\delta}^\delta) = 0 \tag{4.12}$$

by a contradiction argument. Suppose that (4.12) is not true. Then there exist a number $\varepsilon > 0$ and a sequence $\{y^{\delta_l}\}$ satisfying $\|y^{\delta_l} - y\| \leq \delta_l$ with $\delta_l \rightarrow 0$ as $l \rightarrow \infty$ such that

$$D_{\xi_{n_l}^{\delta_l}} \Theta(x^\dagger, x_{n_l}^{\delta_l}) \geq \varepsilon \quad \text{for all } l, \tag{4.13}$$

where $n_l := n_{\delta_l}$. According to Proposition 4.4, there is an integer $n(\varepsilon)$ such that

$$D_{\xi_{n(\varepsilon)}} \Theta(x^\dagger, x_{n(\varepsilon)}) < \varepsilon, \quad \forall \{(x_n, \xi_n)\} \in \Gamma_{\mu_0, \mu_1}(x_0, \xi_0). \tag{4.14}$$

For this $n(\varepsilon)$, by using Lemma 4.8 and by taking a subsequence of $\{y^{\delta_l}\}$ if necessary, we can find $\{(x_n, \xi_n)\} \in \Gamma_{\mu_0, \mu_1}(x_0, \xi_0)$ such that

$$x_n^{\delta_l} \rightarrow x_n, \quad \xi_n^{\delta_l} \rightarrow \xi_n \quad \text{and} \quad \Theta(x_n^{\delta_l}) \rightarrow \Theta(x_n) \quad \text{as } l \rightarrow \infty \tag{4.15}$$

for $0 \leq n \leq n(\varepsilon)$. Since (4.10) implies that $n_l > n(\varepsilon)$ for large l , by using Proposition 3.4 we have

$$D_{\xi_{n_l}^{\delta_l}} \Theta(x^\dagger, x_{n_l}^{\delta_l}) \leq D_{\xi_{n(\varepsilon)}} \Theta(x^\dagger, x_{n(\varepsilon)}^{\delta_l}) = \Theta(x^\dagger) - \Theta(x_{n(\varepsilon)}^{\delta_l}) - \langle \xi_{n(\varepsilon)}^{\delta_l}, x^\dagger - x_{n(\varepsilon)}^{\delta_l} \rangle.$$

In view of (4.15) and (4.14), we therefore obtain

$$\begin{aligned} \limsup_{l \rightarrow \infty} D_{\xi_{n_l}^{\delta_l}} \Theta(x^\dagger, x_{n_l}^{\delta_l}) &\leq \Theta(x^\dagger) - \lim_{l \rightarrow \infty} \Theta(x_{n(\varepsilon)}^{\delta_l}) - \lim_{l \rightarrow \infty} \langle \xi_{n(\varepsilon)}^{\delta_l}, x^\dagger - x_{n(\varepsilon)}^{\delta_l} \rangle \\ &= \Theta(x^\dagger) - \Theta(x_{n(\varepsilon)}) - \langle \xi_{n(\varepsilon)}, x^\dagger - x_{n(\varepsilon)} \rangle \\ &= D_{\xi_{n(\varepsilon)}} \Theta(x^\dagger, x_{n(\varepsilon)}) < \varepsilon. \end{aligned}$$

This is a contradiction to (4.13). We thus obtain (4.12). By virtue of the uniform convexity of Θ , we then have $\lim_{\delta \rightarrow 0} \|x_{n_\delta}^\delta - x^\dagger\| = 0$.

It remains only to show that $\lim_{\delta \rightarrow 0} \Theta(x_{n_\delta}^\delta) = \Theta(x^\dagger)$. In view of (4.12), it suffices to show that

$$\lim_{\delta \rightarrow 0} \langle \xi_{n_\delta}^\delta, x^\dagger - x_{n_\delta}^\delta \rangle = 0. \tag{4.16}$$

We again use a contradiction argument by assuming that there is a number $\varepsilon > 0$ and a sequence $\{y^{\delta_l}\}$ satisfying $\|y^{\delta_l} - y\| \leq \delta_l$ with $\delta_l \rightarrow 0$ as $l \rightarrow \infty$ such that

$$|\langle \xi_{n_l}^{\delta_l}, x^\dagger - x_{n_l}^{\delta_l} \rangle| \geq C_2 \varepsilon \quad \text{for all } l, \tag{4.17}$$

where C_2 is the constant defined in Lemma 4.9. For this ε , we may use Proposition 4.4 and Lemma 4.8 to find an integer $n(\varepsilon)$ such that (4.14) and (4.15) hold. In view of Lemma 4.9, we have

$$\begin{aligned} |\langle \xi_{n_l}^{\delta_l}, x^\dagger - x_{n_l}^{\delta_l} \rangle| &\leq |\langle \xi_{n_l}^{\delta_l} - \xi_{n(\varepsilon)}^{\delta_l}, x^\dagger - x_{n_l}^{\delta_l} \rangle| + |\langle \xi_{n(\varepsilon)}^{\delta_l}, x^\dagger - x_{n_l}^{\delta_l} \rangle| \\ &\leq C_2 D_{\xi_{n(\varepsilon)}^{\delta_l}} \Theta(x^\dagger, x_{n(\varepsilon)}^{\delta_l}) + |\langle \xi_{n(\varepsilon)}^{\delta_l}, x^\dagger - x_{n_l}^{\delta_l} \rangle|. \end{aligned}$$

By taking $l \rightarrow \infty$, using $x_{n_l}^{\delta_l} \rightarrow x^\dagger$, (4.15) and (4.14), we can obtain

$$\limsup_{l \rightarrow \infty} |\langle \xi_{n_l}^{\delta_l}, x^\dagger - x_{n_l}^{\delta_l} \rangle| \leq C_2 \lim_{l \rightarrow \infty} D_{\xi_{n(\varepsilon)}^{\delta_l}} \Theta(x^\dagger, x_{n(\varepsilon)}^{\delta_l}) = C_2 D_{\xi_{n(\varepsilon)}} \Theta(x^\dagger, x_{n(\varepsilon)}) < C_2 \varepsilon$$

which contradicts (4.17). We therefore obtain (4.16) and complete the proof of Theorem 4.7. □

5 Numerical results

We consider the identification of the parameter c in the boundary value problem

$$\begin{cases} -\Delta u + cu = f & \text{in } \Omega, \\ u = g & \text{on } \partial\Omega \end{cases} \tag{5.1}$$

from the measurements of the state variable u in Ω , where $\Omega \subset \mathbb{R}^d$, $d \leq 2$, is a bounded domain with Lipschitz boundary $\partial\Omega$, $f \in H^{-1}(\Omega)$ and $g \in H^{1/2}(\partial\Omega)$. This

is a benchmark example of nonlinear inverse problems. It is well known that (5.1) has a unique solution $u = u(c) \in H^1(\Omega)$ for each c in the domain

$$\mathcal{D} := \{c \in L^2(\Omega) : \|c - \hat{c}\|_{L^2(\Omega)} \leq \gamma_0 \text{ for some } \hat{c} \geq 0, \text{ a.e.}\}$$

with some $\gamma_0 > 0$. By the Sobolev embedding $H^1(\Omega) \hookrightarrow L^r(\Omega)$, it makes sense to define the parameter-to-solution map $F : \mathcal{D} \subset L^2(\Omega) \rightarrow L^r(\Omega)$ with $F(c) = u(c)$ for any $1 < r < \infty$. We consider the problem of identifying $c \in L^2(\Omega)$ from an $L^r(\Omega)$ -measurement of u . This is amount to solving $F(c) = u$. It is known that F is Fréchet differentiable; the Fréchet derivative and its Banach space adjoint are given respectively by

$$\begin{aligned} F'(c)h &= -A(c)^{-1}(hu(c)), & h \in L^2(\Omega), \\ F'(c)^*w &= -u(c)A(c)^{-1}w, & w \in L^{r^*}(\Omega), \end{aligned} \tag{5.2}$$

where r^* is the number conjugate to r , i.e. $1/r^* + 1/r = 1$, and $A(c) : H^2(\Omega) \cap H_0^1(\Omega) \rightarrow L^2(\Omega)$ is defined by $A(c)u = -\Delta u + cu$. Recall that in the space $L^r(\Omega)$ with $1 < r < \infty$ the duality mapping $J_r : L^r(\Omega) \rightarrow L^{r^*}(\Omega)$ is given by

$$J_r(\varphi) := |\varphi|^{r-1} \text{sign}(\varphi), \quad \varphi \in L^r(\Omega).$$

For this parameter identification problem, the tangential cone condition has been verified for $r \geq 2$ in [12,26]; the verification for $1 < r < 2$, however, is not yet available.

In the following we will report some numerical results for this inverse problem to indicate the performance of Algorithm 1 with various choices of the convex function Θ and the Banach spaces \mathcal{X} and \mathcal{Y} . The main computational cost stems from solving the convex minimization problems involved in the algorithm which requires numerical solutions of differential equations related to calculating the Fréchet derivatives and their adjoint. We use BFGS—one of the most popular quasi-Newton methods—for Example 5.1 and a restarted nonlinear CG method for Examples 5.2 and 5.3 below, see [29]. Some fast algorithms have been discovered for solving convex optimization problems in recent years, including the fast proximal gradient method [4] and the primal dual hybrid gradient methods [5,33]. These methods are powerful to deal with problems for which fast solvers such as `fft` are applicable. In case fast algorithms are not applicable as in our computation, they might not have much advantage over other type methods.

Example 5.1 Consider the one-dimensional problem on the interval $\Omega = (0, 1)$ with the source term $f(t) = 100e^{-10(t-0.5)^2}$ and the boundary data $u(0) = 1$ and $u(1) = 2$. We will identify the sought solution

$$c^\dagger(t) = 5t^2(1 - t) + \sin^2(2\pi t)$$

using noisy data that contains a few data points, called outliers, which are highly inconsistent with other data points. The appearance of outliers may arise from procedural measurement errors.

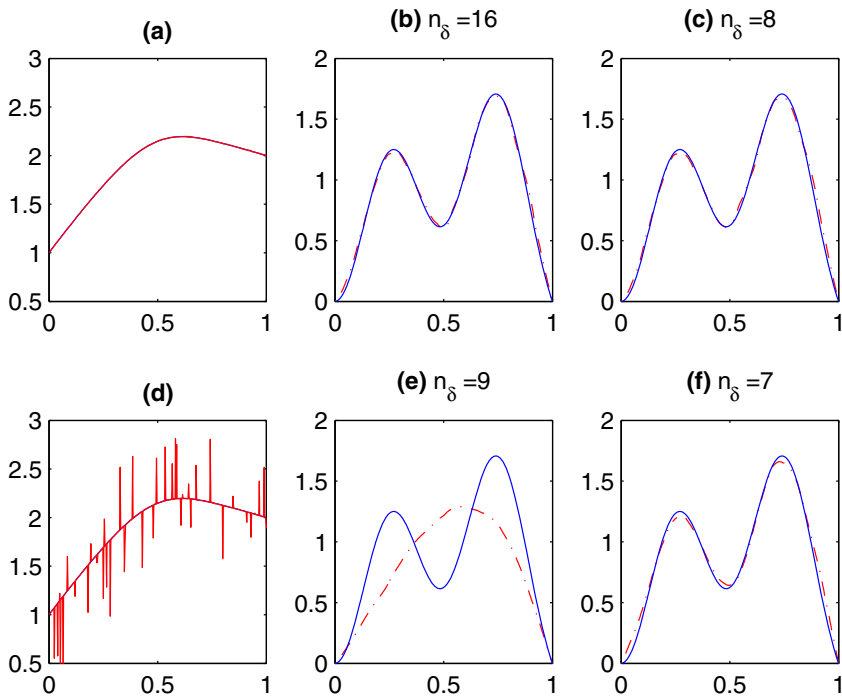


Fig. 1 Numerical results for Example 5.1: **a, d** data with noise; **b, e** reconstruction results by Algorithm 1 with $\mathcal{X} = \mathcal{Y} = L^2[0, 1]$ and $\Theta(c) = \frac{1}{2}\|c\|_2^2$; **c, f** reconstruction result by Algorithm 1 with $\mathcal{X} = L^2[0, 1]$, $\mathcal{Y} = L^{1.1}[0, 1]$ and $\Theta(c) = \frac{1}{2}\|c\|_2^2$

In Fig. 1 we present the numerical results by Algorithm 1 with $\tau = 1.1$, $\mu_0 = 0.90$ and $\mu_1 = 0.96$ using the initial guess $c_0 = 0$ and $\xi_0 = 0$. In order to carry out the computation, the differential equations involved are solved by a finite difference method by dividing $\Omega = (0, 1)$ into 400 subintervals of equal length. The minimization problems involved in the algorithm are solved by performing 50 iterations of the BFGS method. Figure 1a and d show the plots of the noisy data; the one in (a) contains only Gaussian noise, while the one in (d) contains not only Gaussian noise but also 10% outliers. Figure 1b and e present the reconstruction results by the regularizing Levenberg–Marquardt scheme of Hanke, i.e. Algorithm 1 with $\mathcal{X} = \mathcal{Y} = L^2[0, 1]$ and $\Theta(c) = \frac{1}{2}\|c\|_{L^2}^2$; it shows that the method is highly susceptible to the existence of outliers. In Fig. 1c and f we present the reconstruction results by Algorithm 1 with $\mathcal{X} = L^2[0, 1]$, $\mathcal{Y} = L^r[0, 1]$ with $r = 1.1$ and $\Theta(c) = \frac{1}{2}\|c\|_{L^2}^2$. It can be seen that the method is robust enough to prevent being affected by outliers. Using the L^r misfit data terms, with $r \geq 1$ close to 1, to exclude the outliers has been investigated for several other regularization methods, see [14, 15, 18, 23].

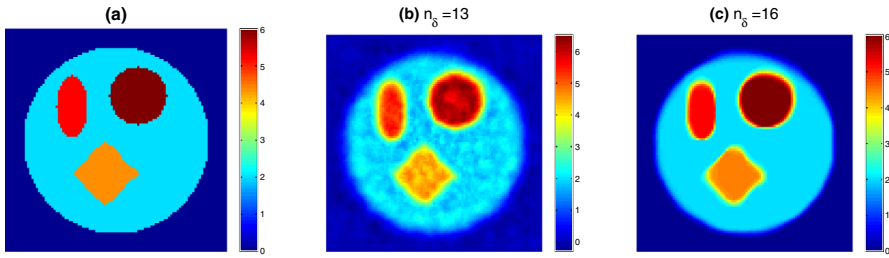


Fig. 2 Reconstruction results for Example 5.2: **a** exact solution; **b** Algorithm 1 with $\Theta(c) = \frac{1}{2} \|c\|_2^2$; **c** Algorithm 1 with $\Theta(c) = \frac{1}{2} \|c\|_2^2 + |c|_{TV,\varepsilon}$ and $\varepsilon = 10^{-4}$

Example 5.2 We next consider the two dimensional problem on $\Omega = [0, 1] \times [0, 1]$ with the source term

$$f(x, y) = 200e^{-10(x-0.5)^2-10(y-0.5)^2}$$

and the boundary data $g \equiv 1$ on $\partial\Omega$. The sought solution is a piecewise constant function as shown in Fig. 2a. We reconstruct the sought solution using Algorithm 1 with $\mathcal{X} = \mathcal{Y} = L^2(\Omega)$ and different choices of Θ . In order to carry out the computation, we divide Ω into 100×100 small squares of equal size. All partial differential equations involved are solved approximately by a finite difference method. When using Algorithm 1, we use $\tau = 1.01, \mu_0 = 0.90, \mu_1 = 0.96$ and take $\xi_0 = c_0 = 0$ as an initial guess. The minimization problem to determine c_n^δ for each $n \geq 1$ is solved by a restart CG method after 1000 iterations.

In Fig. 2 we report the numerical results using measurement data that is corrupted by Gaussian noise with noise level $\delta = 0.001$. Figure 2b presents the reconstruction result using $\Theta(c) = \frac{1}{2} \|c\|_{L^2}^2$. Due to the over-smoothing effect, the reconstruction result turns out to contain unsatisfactory artifacts. Figure 2c reports the reconstruction result using $\Theta(c) = \lambda \|c\|_{L^2}^2 + |c|_{TV,\varepsilon}$ with $\lambda = 1/2$, where $\varepsilon = 10^{-4}$ and $|c|_{TV,\varepsilon} = \int_{\Omega} \sqrt{|\nabla c|^2 + \varepsilon}$ which can be considered as a smoothed approximation of the total variation functional $\int_{\Omega} |\nabla c|$. Clearly the result in (c) significantly improves the one in (b) by efficiently removing the undesired artifacts.

Example 5.3 We use the same setup in Example 5.2 but now the sought solution is sparse. The domain Ω is divided into 120×120 small squares of equal size in order to solve the associated partial differential equations. In Fig. 3 we report the reconstruction results of Algorithm 1 using measurement data contaminated by Gaussian noise with noise level $\delta = 0.001$. We use $\tau = 1.001, \mu_0 = 0.90, \mu_1 = 0.96$ and take $c_0 = \xi_0 = 0$ as an initial guess. The minimization problems involved in the algorithm are solved again by a restart CG after 300 iterations. The true solution is plot in Fig. 3a. Figure 3b presents the numerical result of Algorithm 1 using $\Theta(c) = \frac{1}{2} \|c\|_{L^2}^2$. Figure 3c reports the numerical result of Algorithm 1 using $\Theta(c) = \lambda \|c\|_{L^2}^2 + \int_{\Omega} \sqrt{|c|^2 + \varepsilon}$ with $\lambda = 0.01$ and $\varepsilon = 10^{-4}$ which can be regarded as a smoothed approximation of the L^1 norm. A comparison on the results in (b) and (c) clearly shows that the sparsity

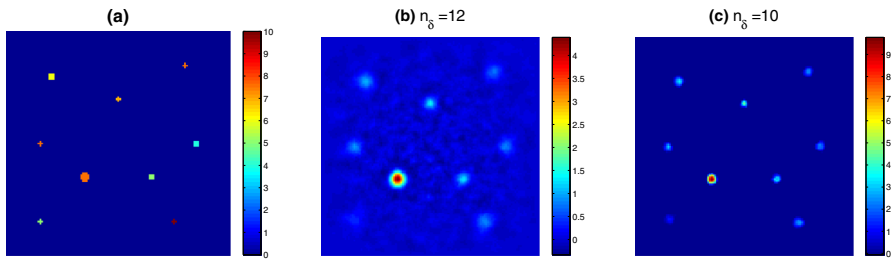


Fig. 3 Reconstruction results for Example 5.3: **a** exact solution; **b** Algorithm 1 with $\Theta(c) = \frac{1}{2} \|c\|_{L^2}^2$; **c** Algorithm 1 with $\Theta(c) = \lambda \|c\|_{L^2}^2 + \int_{\Omega} \sqrt{|c|^2 + \varepsilon}$, where $\lambda = 0.01$ and $\varepsilon = 10^{-4}$

of the sought solution is significantly reconstructed in (c). Therefore, a proper use of a convex function close to the L^1 norm can improve the reconstruction of sparse functions dramatically.

Acknowledgments Q Jin is partly supported by the ARC discovery project grant DP150102345 and H Yang is partly supported by the Natural Science Foundation of China under grant 11071264 and 11571386.

References

1. Bachmayr, M., Burger, M.: Iterative total variation schemes for nonlinear inverse problems. *Inverse Probl.* **25**(10), 105004 (2009)
2. Bakushinsky, A.B., Kokurin, M.Y.: *Iterative Methods for Approximate Solutions of Inverse Problems. Mathematics and Its Applications (New York)*, vol. 577. Springer, Dordrecht (2004)
3. Bazaraa, M.S., Sherali, H.D., Shetty, C.M.: *Nonlinear Programming, Theory and Algorithms*, 3rd edn. Wiley-Interscience, New York (2006)
4. Beck, A., Teboulle, M.: A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sci.* **2**, 183–202 (2009)
5. Chambolle, A., Pock, T.: A first-order primal-dual algorithm for convex problems with applications to imaging. *J. Math. Imaging Vis.* **40**, 120–145 (2011)
6. Cioranescu, I.: *Geometry of Banach Spaces, Duality Mappings and Nonlinear Problems*. Kluwer, Dordrecht (1990)
7. Colonius, F., Kunisch, K.: Stability for parameter estimation in two point boundary value problems. *J. Reine Angew. Math.* **370**, 1–29 (1986)
8. Engl, H.W., Hanke, M., Neunauer, A.: *Regularization of Inverse Problems*. Kluwer, Dordrecht (1996)
9. Engl, H.W., Kunisch, K., Neubauer, A.: Convergence rates of Tikhonov regularization of nonlinear ill-posed problems. *Inverse Probl.* **5**, 523–540 (1989)
10. Hanke, M.: Regularizing properties of a truncated Newton-CG algorithm for nonlinear inverse problems. *Numer. Funct. Anal. Optim.* **18**(9–10), 971–993 (1997)
11. Hanke, M.: A regularizing Levenberg–Marquardt scheme with applications to inverse groundwater filtration problems. *Inverse Probl.* **13**, 79–95 (1997)
12. Hanke, M., Neubauer, A., Scherzer, O.: A convergence analysis of the Landweber iteration for nonlinear ill-posed problems. *Numer. Math.* **72**, 21–37 (1995)
13. Hettrich, F., Rundell, W.: A second degree method for nonlinear inverse problems. *SIAM J. Numer. Anal.* **37**(2), 587–620 (2000)
14. Hohage, T., Werner, F.: Iteratively regularized Newton-type methods for general data misfit functionals and applications to Poisson data. *Numer. Math.* **123**(4), 745–779 (2013)
15. Hohage, T., Werner, F.: Convergence rates for inverse problems with impulsive noise. *SIAM J. Numer. Anal.* **52**(3), 1203–1221 (2014)

16. Jin, Q.: Applications of the modified discrepancy principle to Tikhonov regularization of nonlinear ill-posed problems. *SIAM J. Numer. Anal.* **36**(2), 475–490 (1999)
17. Jin, Q.: On the order optimality of the regularization via inexact Newton iterations. *Numer. Math.* **121**, 237–260 (2012)
18. Jin, Q.: Inexact Newton–Landweber iteration for solving nonlinear inverse problems in Banach spaces. *Inverse Probl.* **28**, 065002 (15pp) (2012)
19. Jin, Q., Tautenhahn, U.: On the discrepancy principle for some Newton type methods for solving nonlinear inverse problems. *Numer. Math.* **111**, 509–558 (2009)
20. Jin, Q., Wang, W.: Landweber iteration of Kaczmarz type with general non-smooth convex penalty functionals. *Inverse Probl.* **29**, 085011 (22pp) (2013)
21. Jin, Q., Zhong, M.: On the iteratively regularized Gauss–Newton method in Banach spaces with applications to parameter identification problems. *Numer. Math.* **124**, 647–683 (2013)
22. Jin, Q., Zhong, M.: Nonstationary iterated Tikhonov regularization in Banach spaces with uniformly convex penalty terms. *Numer. Math.* **127**, 485–513 (2014)
23. Kaltenbacher, B.: A convergence rates result for an iteratively regularized Gauss–Newton–Halley method in Banach space. *Inverse Probl.* **31**, 015007 (20pp) (2015)
24. Kaltenbacher, B., Hofmann, B.: Convergence rates for the iteratively regularized Gauss–Newton method in Banach spaces. *Inverse Probl.* **26**, 035007 (21pp) (2010)
25. Kaltenbacher, B., Neubauer, A., Scherzer, O.: *Iterative Regularization Methods for Nonlinear Ill-Posed Problems*. de Gruyter, Berlin (2008)
26. Kaltenbacher, B., Schöpfer, F., Schuster, T.: Iterative methods for nonlinear ill-posed problems in Banach spaces: convergence and applications to parameter identification problems. *Inverse Probl.* **25**, 065003 (19pp) (2009)
27. Lechleiter, A., Rieder, A.: Towards a general convergence theory for inexact Newton regularizations. *Numer. Math.* **114**, 521–548 (2010)
28. Margotti, F., Rieder, A., Leitão, A.: A Kaczmarz version of the REGINN-Landweber iteration for ill-posed problems in Banach spaces. *SIAM J. Numer. Anal.* **52**, 1439–1465 (2014)
29. Nocedal, J., Wright, S.J.: *Numerical Optimization*, 2nd edn. Springer, New York (2006)
30. Rieder, A.: On the regularization of nonlinear ill-posed problems via inexact Newton iterations. *Inverse Probl.* **15**, 309–327 (1999)
31. Schuster, T., Kaltenbacher, B., Hofmann, B., Kazimierski, K.S.: *Regularization Methods in Banach Spaces*. Walter de Gruyter, Berlin (2012)
32. Zălinescu, C.: *Convex Analysis in General Vector Spaces*. World Scientific Publishing Co., Inc., River Edge (2002)
33. Zhu, M., Chan, T.F.: An efficient primaldual hybrid gradient algorithm for total variation image restoration. CAM Report 08-34, UCLA (2008)