

Nonstationary iterated Tikhonov regularization in Banach spaces with uniformly convex penalty terms

Qinian Jin · Min Zhong

Abstract We consider the nonstationary iterated Tikhonov regularization in Banach spaces which defines the iterates via minimization problems with uniformly convex penalty term. The penalty term is allowed to be non-smooth to include L^1 and total variation (TV) like penalty functionals, which are significant in reconstructing special features of solutions such as sparsity and discontinuities in practical applications. We present the detailed convergence analysis and obtain the regularization property when the method is terminated by the discrepancy principle. In particular we establish the strong convergence and the convergence in Bregman distance which sharply contrast with the known results that only provide weak convergence for a subsequence of the iterative solutions. Some numerical experiments on linear integral equations of first kind and parameter identification in differential equations are reported.

Mathematics Subject Classification (2000) 65J15 · 65J20 · 47H17

1 Introduction

We are interested in solving inverse problems which can be formulated as the operator equation

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

where $F : D(F) \subset \mathcal{X} \mapsto \mathcal{Y}$ is an operator between two Banach spaces \mathcal{X} and \mathcal{Y} with domain $D(F) \subset \mathcal{X}$; the norms in \mathcal{X} and \mathcal{Y} are denoted by the same notation $\|\cdot\|$ that should be clear from the context. A characteristic property of inverse problems is their ill-posedness in the sense that their solutions do not depend continuously on the data. Due to errors in the measurements, one never has the exact data in practical applications; instead only noisy data are available. If one uses the algorithms developed for well-posed problems directly, it usually fails to produce any

Qinian Jin

Mathematical Sciences Institute, Australian National University, Canberra, ACT 0200, Australia
E-mail: Qinian.Jin@anu.edu.au

Min Zhong

School of Mathematical Sciences, Fudan University, Shanghai 200433, China
E-mail: 09110180007@fudan.edu.cn

useful information since noise could be amplified by an arbitrarily large factor. Let y^δ be the only available noisy data to y satisfying

$$\|y^\delta - y\| \leq \delta \quad (1.2)$$

with a given small noise level $\delta > 0$. How to use y^δ to produce a stable approximate solution to (1.1) is a central topic, and regularization methods should be taken into account.

When both \mathcal{X} and \mathcal{Y} are Hilbert spaces, a lot of regularization methods have been proposed to solve inverse problems in the Hilbert space framework ([4, 15]). In case $F : \mathcal{X} \rightarrow \mathcal{Y}$ is a bounded linear operator, nonstationary iterated Tikhonov regularization is an attractive iterative method in which a sequence $\{x_n^\delta\}$ of regularized solutions is defined successively by

$$x_n^\delta := \arg \min_{x \in \mathcal{X}} \left\{ \frac{1}{2} \|Fx - y^\delta\|^2 + \frac{\alpha_n}{2} \|x - x_{n-1}^\delta\|^2 \right\},$$

where $x_0^\delta := x_0 \in \mathcal{X}$ is an initial guess and $\{\alpha_n\}$ is a preassigned sequence of positive numbers. Since $\{x_n^\delta\}$ can be written explicitly as

$$x_n^\delta = x_{n-1}^\delta - (\alpha_n I + F^* F)^{-1} F^* (F x_{n-1}^\delta - y^\delta),$$

where $F^* : \mathcal{Y} \rightarrow \mathcal{X}$ denotes the adjoint of $F : \mathcal{X} \rightarrow \mathcal{Y}$, the complete analysis of the regularization property has been established (see [8] and references therein) when $\{\alpha_n\}$ satisfies suitable property and the discrepancy principle is used to terminate the iteration. This method has been extended in [12, 13] to solve nonlinear inverse problems in Hilbert spaces.

Regularization methods in Hilbert spaces can produce good results when the sought solution is smooth. However, because such methods have a tendency to over-smooth solutions, they may not produce good results in applications where the sought solution has special features such as sparsity or discontinuities. In order to capture the special features, the methods in Hilbert spaces should be modified by incorporating the information of suitable adapted penalty functionals, for which the theories in Hilbert space setting are no longer applicable.

The nonstationary iterated Tikhonov regularization has been extended in [14] for solving linear inverse problems in Banach spaces setting by defining x_n^δ as the minimizer of the convex minimization problem

$$\min_{x \in \mathcal{X}} \left\{ \frac{1}{r} \|Fx - y^\delta\|^r + \alpha_n \Delta_p(x, x_{n-1}^\delta) \right\}$$

for $n \geq 1$ successively, where $1 \leq r < \infty$, $1 < p < \infty$ and $\Delta_p(\cdot, \cdot)$ denotes the Bregman distance on \mathcal{X} induced by the convex function $x \rightarrow \|x\|^p/p$. When \mathcal{X} is uniformly smooth and uniformly convex, and when the method is terminated by the discrepancy principle, the regularization property has been established if $\{\alpha_n\}$ satisfies $\sum_{n=1}^{\infty} \alpha_n^{-1} = \infty$. The numerical simulations in [14] indicate that the method is efficient in sparsity reconstruction when choosing $\mathcal{X} = L^p$ with $p > 1$ close to 1 on one hand, and provides robust estimator in the presence of outliers in the noisy data when choosing $\mathcal{Y} = L^1$ on the other hand. However, since \mathcal{X} is required to be uniformly smooth and uniformly convex and since $\Delta_p(\cdot, \cdot)$ is induced by the power of the norm in \mathcal{X} , the result in [14] does not apply to regularization methods with L^1 and total variation like penalty terms that are important for reconstructing sparsity and discontinuities of sought solutions.

The total variational regularization was introduced in [18], its importance was recognized immediately and many successive works were conducted in the last two decades. In [16] an iterative regularization method based on Bregman distance and total variation was introduced to enhance the multi-scale nature of reconstruction. The method solves (1.1) with $F : \mathcal{X} \rightarrow \mathcal{Y}$ linear and \mathcal{Y} a Hilbert space by defining $\{x_n^\delta\}$ in the primal space \mathcal{X} and $\{\xi_n^\delta\}$ in the dual space \mathcal{X}^* via

$$\begin{aligned} x_n^\delta &:= \arg \min_{x \in \mathcal{X}} \left\{ \|Fx - y^\delta\|^2 + \alpha_n D_{\xi_{n-1}^\delta} \Theta(x, x_{n-1}^\delta) \right\}, \\ \xi_n^\delta &:= \xi_{n-1}^\delta - \frac{1}{\alpha_n} F^*(Fx_n^\delta - y^\delta), \end{aligned} \quad (1.3)$$

where $\Theta : \mathcal{X} \rightarrow (-\infty, \infty]$ is a proper convex function, $x_0^\delta \in \mathcal{X}$ is an initial guess, $\xi_0^\delta \in \mathcal{X}^*$ is in the sub-gradient of Θ at x_0^δ , and $D_\xi \Theta(\cdot, \cdot)$ denotes the Bregman distance induced by Θ . This method was extended in [2] to solve nonlinear inverse problems. Extensive numerical simulations were reported in [16, 2] and convergence analysis was given, with special attention to the case that $\mathcal{X} = L^2(\Omega)$ and $\Theta(x) = a\|x\|_{L^2}^2 + \int_\Omega |Dx|$, where $\int_\Omega |Dx|$ denotes the total variation, when the iteration is terminated by a discrepancy principle and $\{\alpha_n\}$ satisfies the condition $\underline{\alpha} \leq \alpha_n \leq \bar{\alpha}$ for two positive constants $\bar{\alpha} \geq \underline{\alpha} > 0$. The analysis in [16, 2], however, is somewhat preliminary since it provides only the boundedness of $\{\Theta(x_{n_\delta}^\delta)\}$ which guarantees only weak convergence for a subsequence of $\{x_{n_\delta}^\delta\}$, where n_δ denotes the stopping index determined by the discrepancy principle. It is natural to ask if the whole sequence converges strongly and in Bregman distance.

We point out that the method (1.3) is equivalent to the augmented Lagrangian method introduced originally in [10, 17] and developed further in various directions, see [11] and reference therein. One may refer to [6] for some results on convergence and convergence rates of the augmented Lagrangian method applied to linear inverse problems in Hilbert spaces with general convex penalty term. When \mathcal{X} and \mathcal{Y} are Hilbert spaces and $\Theta(x) = \|x\|^2$, (1.3) is exactly the nonstationary iterated Tikhonov regularization. In this paper we formulate an extension of the nonstationary iterated Tikhonov regularization in the spirit of (1.3) to solve (1.1) with both \mathcal{X} and \mathcal{Y} being Banach spaces and present the detailed convergence analysis when the method is terminated by the discrepancy principle. In the method we allow $\{\alpha_n\}$ to vary in various ways so that geometric decreasing sequence can be included; this makes it possible to terminate the method in fewer iterations. Moreover, we allow the penalty term Θ to be general uniformly convex functions on \mathcal{X} so that the method can be used for sparsity reconstruction and discontinuity detection. Most importantly, we obtain

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

and give a characterization of the limit x^\dagger , which significantly improve the known convergence results.

This paper is organized as follows. In section 2 we give some preliminary results on Banach spaces and convex analysis. In section 3, we then formulate the method in Banach spaces with uniformly convex penalty term for solving linear and nonlinear inverse problems, and present the main convergence results. In section 4 we first prove a convergence result for the method when the data is given exactly; we then show that, if the data contains noise, the method is well-defined and admits some stability property; by combining these results we finally obtain the proof of the main convergence theorems. Finally, in section 5 we present some numerical simulations

on linear integral equations of first kind and parameter identification problems in partial differential equations to test the performance of the method.

2 Preliminaries

Let \mathcal{X} be a Banach space with norm $\|\cdot\|$. We use \mathcal{X}^* to denote its dual space. Given $x \in \mathcal{X}$ and $\xi \in \mathcal{X}^*$ we write $\langle \xi, x \rangle = \xi(x)$ for the duality pairing. We use “ \rightarrow ” and “ \rightharpoonup ” to denote the strong convergence and weak convergence respectively. If \mathcal{Y} is another Banach space and $A : \mathcal{X} \rightarrow \mathcal{Y}$ is a bounded linear operator, we use $A^* : \mathcal{Y}^* \rightarrow \mathcal{X}^*$ to denote its adjoint, i.e. $\langle A^*\zeta, x \rangle = \langle \zeta, Ax \rangle$ for any $x \in \mathcal{X}$ and $\zeta \in \mathcal{Y}^*$. 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

$$\mathcal{N}(A)^\perp = \overline{\mathcal{R}(A^*)}, \quad (2.1)$$

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 $D(\Theta) := \{x \in \mathcal{X} : \Theta(x) < +\infty\}$ to denote its effective domain. We call Θ proper if $D(\Theta) \neq \emptyset$. 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 $x \in D(\Theta)$. Let

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

For $x \in D(\partial\Theta)$ 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

$$D_\xi\Theta(x_2, x) - D_\xi\Theta(x_1, x) = D_{\xi_1}\Theta(x_2, x_1) + \langle \xi_1 - \xi, x_2 - x_1 \rangle \quad (2.2)$$

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

A proper convex function $\Theta : \mathcal{X} \rightarrow (-\infty, \infty]$ is called uniformly convex if there is a continuous function $h : [0, \infty) \rightarrow [0, \infty)$, with the property that $h(t) = 0$ implies $t = 0$, such that

$$\Theta(\lambda\bar{x} + (1-\lambda)x) + \lambda(1-\lambda)h(\|\bar{x} - x\|) \leq \lambda\Theta(\bar{x}) + (1-\lambda)\Theta(x) \quad (2.3)$$

for all $\bar{x}, x \in \mathcal{X}$ and $\lambda \in (0, 1)$. If h in (2.3) can be taken as $h(t) = ct^p$ for some $c > 0$ and $p \geq 2$, then Θ is called p -uniformly convex. It can be shown ([20, Theorem 3.5.10]) that Θ is uniformly convex if and only if there is a strictly increasing continuous function $\varphi : [0, \infty) \rightarrow [0, \infty)$ with $\varphi(0) = 0$ such that

$$D_\xi\Theta(\bar{x}, x) \geq \varphi(\|\bar{x} - x\|) \quad (2.4)$$

for all $\bar{x} \in \mathcal{X}$, $x \in D(\partial\Theta)$ and $\xi \in \partial\Theta(x)$.

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

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

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

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

satisfies $\delta_{\mathcal{X}}(t) > 0$ for all $0 < t \leq 2$. If there are $c > 0$ and $r > 1$ such that $\delta_{\mathcal{X}}(t) \geq ct^r$ for all $0 < t \leq 2$, then \mathcal{X} is called r -uniformly convex. We call \mathcal{X} uniformly smooth if its modulus of smoothness

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

satisfies $\lim_{s \searrow 0} \frac{\rho_{\mathcal{X}}(s)}{s} = 0$. One can refer to [1,3] for many examples of Banach spaces, including the sequence spaces l^r , the Lebesgue spaces L^r , the Sobolev spaces $W^{k,r}$ and the Besov spaces $B^{s,r}$ with $1 < r < \infty$, that are both uniformly convex and uniformly smooth.

It is well known that any uniformly convex or uniformly smooth Banach space is reflexive. On a uniformly smooth Banach space \mathcal{X} , every duality mapping J_r with $1 < r < \infty$ is single valued and uniformly continuous on bounded sets; for each $1 < r < \infty$ we use

$$\Delta_r(\bar{x}, x) = \frac{1}{r}\|\bar{x}\|^r - \frac{1}{r}\|x\|^r - \langle J_r(x), \bar{x} - x \rangle, \quad \forall \bar{x}, x \in \mathcal{X}$$

to denote the Bregman distance induced by the convex function $\Theta(x) = \|x\|^r/r$.

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 easily generalized for uniformly convex functions which we state in the following result.

Lemma 2.1 *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 $n \rightarrow \infty$.*

Proof Assume the result is not true. Then, by taking a subsequence if necessary, there is an $\epsilon > 0$ such that $\|x_n - x\| \geq \epsilon$ for all n . In view of the uniform convexity of Θ , there is a $\gamma > 0$ such that $\Theta((x_n + x)/2) \leq (\Theta(x_n) + \Theta(x))/2 - \gamma$. Using $\Theta(x_n) \rightarrow \Theta(x)$ we then obtain

$$\limsup_{n \rightarrow \infty} \Theta\left(\frac{x_n + x}{2}\right) \leq \Theta(x) - \gamma.$$

On the other hand, observing that $(x_n + x)/2 \rightharpoonup x$, we have from the weakly lower semi-continuity of Θ that

$$\Theta(x) \leq \liminf_{n \rightarrow \infty} \Theta\left(\frac{x_n + x}{2}\right).$$

Therefore $\Theta(x) \leq \Theta(x) - \gamma$, which is a contradiction. \square

In many practical applications, proper, weakly lower semi-continuous, uniformly convex functions can be easily constructed. For instance, consider $\mathcal{X} = L^p(\Omega)$, where $2 \leq p < \infty$ and Ω is a bounded domain in \mathbb{R}^d . It is known that the functional $\Theta_0(x) := \int_{\Omega} |x(\omega)|^p d\omega$ is uniformly convex on $L^p(\Omega)$ (it is in fact p -uniformly convex). Consequently we obtain on $L^p(\Omega)$ the uniformly convex functions

$$\Theta(x) := \mu \int_{\Omega} |x(\omega)|^p d\omega + a \int_{\Omega} |x(\omega)| d\omega + b \int_{\Omega} |Dx|, \quad (2.5)$$

where $\mu > 0$, $a, b \geq 0$, and $\int_{\Omega} |Dx|$ denotes the total variation of x over Ω that is defined by ([7])

$$\int_{\Omega} |Dx| := \sup \left\{ \int_{\Omega} x \operatorname{div} f d\omega : f \in C_0^1(\Omega; \mathbb{R}^N) \text{ and } \|f\|_{L^\infty(\Omega)} \leq 1 \right\}.$$

For $a = 1$ and $b = 0$ the corresponding function is useful for sparsity reconstruction ([19]); while for $a = 0$ and $b = 1$ the corresponding function is useful for detecting the discontinuities, in particular, when the solutions are piecewise-constant ([18]).

3 The method and main results

We now return to (1.1), where $F : \mathcal{X} \rightarrow \mathcal{Y}$ is an operator between two Banach spaces \mathcal{X} and \mathcal{Y} . We will always assume that \mathcal{X} is reflexive, \mathcal{Y} is uniformly smooth, and (1.1) has a solution. In general, the equation (1.1) may have many solutions. In order to find the desired one, some selection criteria should be enforced. Choosing a proper convex function Θ , we pick $x_0 \in D(\partial\Theta)$ and $\xi_0 \in \partial\Theta(x_0)$ as the initial guess, which may incorporate some available information on the sought solution. We define x^\dagger to be the solution of (1.1) with the property

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

We will work under the following conditions on the convex function Θ and the operator F .

Assumption 3.1 Θ is a proper, weakly lower semi-continuous and uniformly convex function such that (2.4) holds, i.e. there is a strictly increasing continuous function $\varphi : [0, \infty) \rightarrow [0, \infty)$ with $\varphi(0) = 0$ such that

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

for $\bar{x} \in \mathcal{X}$, $x \in D(\partial\Theta)$ and $\xi \in \partial\Theta(x)$.

Assumption 3.2 (a) $D(F)$ is convex, and F is weakly closed, i.e. for any sequence $\{x_n\} \subset D(F)$ satisfying $x_n \rightharpoonup x \in \mathcal{X}$ and $F(x_n) \rightharpoonup v \in \mathcal{Y}$ there hold $x \in D(F)$ and $F(x) = v$;
 (b) There is $\rho > 0$ such that (1.1) has a solution in $B_\rho(x_0) \cap D(F) \cap D(\Theta)$, where $B_\rho(x_0) := \{x \in \mathcal{X} : \|x - x_0\| \leq \rho\}$;
 (c) F is Fréchet differentiable on $D(F)$, and $x \rightarrow F'(x)$ is continuous on $D(F)$, where $F'(x)$ denotes the Fréchet derivative of F at x ;
 (d) There exists $0 \leq \eta < 1$ such that

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

for all $\bar{x}, x \in B_{3\rho}(x_0) \cap D(F)$.

When \mathcal{X} is a reflexive Banach space, by using the weakly closedness of F and the weakly lower semi-continuity and uniformly convexity of Θ it is standard to show that x^\dagger exists. The following result shows that x^\dagger is in fact uniquely defined.

Lemma 3.1 *Let \mathcal{X} be reflexive, Θ satisfy Assumption 3.1, and F satisfy Assumption 3.2. If x^\dagger is a solution of $F(x) = y$ satisfying (3.1) with*

$$D_{\xi_0}\Theta(x^\dagger, x_0) \leq \varphi(\rho), \quad (3.2)$$

then x^\dagger is uniquely defined.

Proof Assume that (1.1) has two distinct solutions \hat{x} and x^\dagger satisfying (3.1). Then it follows from (3.2) that

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

By using Assumption 3.1 on Θ we obtain $\|\hat{x} - x_0\| \leq \rho$ and $\|x^\dagger - x_0\| \leq \rho$. Since $F(\hat{x}) = F(x^\dagger)$, we can use Assumption 3.2 (d) to derive that $F'(x^\dagger)(\hat{x} - x^\dagger) = 0$. Let $x_\lambda = \lambda\hat{x} + (1 - \lambda)x^\dagger$ for $0 < \lambda < 1$. Then $x_\lambda \in B_\rho(x_0) \cap D(\Theta) \cap D(F)$ and $F'(x^\dagger)(x_\lambda - x^\dagger) = 0$. Thus we can use Assumption 3.2 (d) to conclude that

$$\|F(x_\lambda) - F(x^\dagger)\| \leq \eta \|F(x_\lambda) - F(x^\dagger)\|.$$

Since $0 \leq \eta < 1$, this implies that $F(x_\lambda) = F(x^\dagger) = y$. Consequently, by the minimal property of x^\dagger we have

$$D_{\xi_0}\Theta(x_\lambda, x_0) \geq D_{\xi_0}\Theta(x^\dagger, x_0). \quad (3.3)$$

On the other hand, it follows from the strictly convexity of Θ that

$$D_{\xi_0}\Theta(x_\lambda, x_0) < \lambda D_{\xi_0}\Theta(\hat{x}, x_0) + (1 - \lambda)D_{\xi_0}\Theta(x^\dagger, x_0) = D_{\xi_0}\Theta(x^\dagger, x_0)$$

for $0 < \lambda < 1$ which is a contradiction to (3.3). \square

We are now ready to formulate the nonstationary iterated Tikhonov regularization with penalty term induced by the uniformly convex function Θ . For the initial guess $x_0^\delta := x_0 \in D(\partial\Theta) \cap D(F)$ and $\xi_0^\delta := \xi_0 \in \partial\Theta(x_0)$, we take a sequence of positive numbers $\{\alpha_n\}$ and define the iterative sequences $\{x_n^\delta\}$ and $\{\xi_n^\delta\}$ successively by

$$\begin{aligned} x_n^\delta &\in \arg \min_{x \in D(F)} \left\{ \frac{1}{r} \|F(x) - y^\delta\|^r + \alpha_n D_{\xi_{n-1}^\delta} \Theta(x, x_{n-1}^\delta) \right\}, \\ \xi_n^\delta &= \xi_{n-1}^\delta - \frac{1}{\alpha_n} F'(x_n^\delta)^* J_r(F(x_n^\delta) - y^\delta) \end{aligned} \quad (3.4)$$

for $n \geq 1$, where $1 < r < \infty$ and $J_r : \mathcal{Y} \rightarrow \mathcal{Y}^*$ denotes the duality mapping of \mathcal{Y} with gauge function $t \rightarrow t^{r-1}$ which is single-valued and continuous because \mathcal{Y} is assumed to be uniformly smooth. At each step, the existence of x_n^δ is guaranteed by the reflexivity of \mathcal{X} and \mathcal{Y} , the weakly lower semi-continuity and uniformly convexity of Θ , and the weakly closedness of F . However, x_n^δ might not be unique when F is nonlinear; we will take x_n^δ to be any one of the minimizers. In view of the minimality of x_n^δ , we have $\xi_n^\delta \in \partial\Theta(x_n^\delta)$. From the definition of x_n^δ , it is straightforward to see that

$$\|F(x_n^\delta) - y^\delta\| \leq \|F(x_{n-1}^\delta) - y^\delta\|, \quad n = 1, 2, \dots \quad (3.5)$$

We will terminate the iteration by the discrepancy principle

$$\|F(x_{n_\delta}^\delta) - y^\delta\| \leq \tau\delta < \|F(x_n^\delta) - y^\delta\|, \quad 0 \leq n < n_\delta \quad (3.6)$$

with a given constant $\tau > 1$. The output $x_{n_\delta}^\delta$ will be used to approximate a solution of (1.1).

In order to understand the convergence property of $x_{n_\delta}^\delta$, it is necessary to consider the noise-free iterative sequences $\{x_n\}$ and $\{\xi_n\}$, where each x_n and ξ_n with $n \geq 1$ are defined by (3.4) with y^δ replaced by y , i.e.,

$$\begin{aligned} x_n &\in \arg \min_{x \in D(F)} \left\{ \frac{1}{r} \|F(x) - y\|^r + \alpha_n D_{\xi_{n-1}} \Theta(x, x_{n-1}) \right\}, \\ \xi_n &= \xi_{n-1} - \frac{1}{\alpha_n} F'(x_n)^* J_r(F(x_n) - y) \in \partial \Theta(x_n). \end{aligned} \quad (3.7)$$

In section 4.1 we will give a detailed convergence analysis on $\{x_n\}$; in particular, we will show that $\{x_n\}$ strongly converges to a solution of (1.1). In order to connect such result with the convergence property of $x_{n_\delta}^\delta$, we will make the following assumption.

Assumption 3.3 x_n is uniquely defined for each n .

We will give some sufficient condition for the validity of Assumption 3.3. This assumption enables us to establish some stability results connecting x_n^δ and x_n so that we can finally obtain the convergence property of $x_{n_\delta}^\delta$ in the following result.

Theorem 3.1 Let \mathcal{X} be reflexive and \mathcal{Y} be uniformly smooth, let Θ satisfy Assumption 3.1, and let F satisfy Assumptions 3.2 and 3.3. Assume that $1 < r < \infty$, $\tau > (1 + \eta)/(1 - \eta)$ and that $\{\alpha_n\}$ is a sequence of positive numbers satisfying $\sum_{n=1}^{\infty} \alpha_n^{-1} = \infty$ and $\alpha_n \leq c_0 \alpha_{n+1}$ for all n with some constant $c_0 > 0$. Assume further that

$$D_{\xi_0} \Theta(x^\dagger, x_0) \leq \frac{\tau^r - 1}{\tau^r - 1 + c_0} \varphi(\rho). \quad (3.8)$$

Then, the discrepancy principle (3.6) terminates the method (3.4) after $n_\delta < \infty$ steps. Moreover, there is a solution $x_* \in D(\Theta)$ of (1.1) such that

$$x_{n_\delta}^\delta \rightarrow x_*, \quad \Theta(x_{n_\delta}^\delta) \rightarrow \Theta(x_*) \quad \text{and} \quad D_{\xi_{n_\delta}^\delta} \Theta(x_*, x_{n_\delta}^\delta) \rightarrow 0 \quad (3.9)$$

as $\delta \rightarrow 0$. If, in addition, $\mathcal{N}(F'(x^\dagger)) \subset \mathcal{N}(F'(x))$ for all $x \in B_{3\rho}(x_0) \cap D(F)$, then $x_* = x^\dagger$.

In this result, the closeness condition (3.8) is used to guarantee that x_n^δ is in $B_{3\rho}(x_0)$ for $0 \leq n \leq n_\delta$ so that Assumption 3.2 (d) can be applied. This issue does not appear when $F : \mathcal{X} \rightarrow \mathcal{Y}$ is a bounded linear operator. Furthermore, Assumption 3.3 holds automatically for linear problems when Θ is strictly convex. Consequently, we have the following convergence result for linear inverse problems.

Theorem 3.2 Let $F : \mathcal{X} \rightarrow \mathcal{Y}$ be a bounded linear operator with \mathcal{X} being reflexive and \mathcal{Y} being uniformly smooth, let Θ be proper, weakly lower semi-continuous, and uniformly convex, let $1 < r < \infty$, and let $\{\alpha_n\}$ be such that $\sum_{n=1}^{\infty} \alpha_n^{-1} = \infty$ and $\alpha_n \leq c_0 \alpha_{n+1}$ for all n with $c_0 > 0$. Then, the discrepancy principle (3.6) with $\tau > 1$ terminates the method after $n_\delta < \infty$ steps. Moreover, there hold

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

as $\delta \rightarrow 0$.

In the next section, we will give the detailed proof of Theorem 3.1. It should be pointed out that the convergence $x_{n_\delta}^\delta \rightarrow x_*$ does not imply $\Theta(x_{n_\delta}^\delta) \rightarrow \Theta(x_*)$ directly since Θ is not necessarily continuous. The proof of $\Theta(x_{n_\delta}^\delta) \rightarrow \Theta(x_*)$ relies on additional observation.

When applying our convergence result to the situation that $\mathcal{X} = L^2(\Omega)$ and $\Theta(x) = \mu \int_\Omega |x(\omega)|^2 d\omega + \int_\Omega |Dx|$ with $\mu > 0$, we can obtain

$$\|x_{n_\delta}^\delta - x^\dagger\|_{L^2(\Omega)} \rightarrow 0 \quad \text{and} \quad \int_\Omega |Dx_{n_\delta}^\delta| \rightarrow \int_\Omega |Dx| \quad \text{as } \delta \rightarrow 0.$$

This significantly improves the result in [2] in which only the boundedness of $\Theta(x_{n_\delta}^\delta)$ was derived and hence only weak convergence for a subsequence of $\{x_{n_\delta}^\delta\}$ can be guaranteed.

We conclude this section with some sufficient condition to guarantee the validity of Assumption 3.3.

Assumption 3.4 *There exist $C_0 \geq 0$ and $1/r \leq \kappa < 1$ such that*

$$\|F(\bar{x}) - F(x) - F'(x)(\bar{x} - x)\| \leq C_0 [D_\xi \Theta(\bar{x}, x)]^{1-\kappa} [\Delta_r(F(\bar{x}) - y, F(x) - y)]^\kappa$$

for all $\bar{x}, x \in B_{3\rho}(x_0) \cap D(\Theta) \cap D(F)$ with $x \in D(\partial\Theta)$ and $\xi \in \partial\Theta(x)$, where $\Delta_r(\cdot, \cdot)$ denotes the Bregman distance on \mathcal{Y} induced by the convex function $\|y\|^r/r$.

When \mathcal{Y} is a r -uniformly convex Banach space, Θ is a p -uniformly convex function on \mathcal{X} with $p \geq 2$, and $1/p + 1/r \leq 1$, Assumption 3.4 holds with $\kappa = 1 - 1/p$ if there is a constant $C_1 \geq 0$ such that

$$\|F(\bar{x}) - F(x) - F'(x)(\bar{x} - x)\| \leq C_1 \|\bar{x} - x\| \|F(\bar{x}) - F(x)\| \quad (3.10)$$

for $\bar{x}, x \in B_{3\rho}(x_0) \cap D(F)$, which is a slightly strengthened version of Assumption 3.2 (d).

Lemma 3.2 *Let \mathcal{X} be reflexive and \mathcal{Y} be uniformly smooth, let $1 < r < \infty$, let Θ satisfy Assumption 3.1, let F satisfy Assumptions 3.2 and 3.4, and let $\{\alpha_n\}$ satisfy $\sum_{n=1}^\infty \alpha_n^{-1} = \infty$. Assume that*

$$D_{\xi_0} \Theta(x^\dagger, x_0) \leq \varphi(\rho) \quad \text{and} \quad \bar{C}_0 [D_{\xi_0} \Theta(x^\dagger, x_0)]^{1-\frac{1}{r}} < 1 \quad (3.11)$$

with $\bar{C}_0 := C_0 \kappa^\kappa (1 - \kappa)^{1-\kappa} (1 - \eta)^{\frac{1-r}{r}} \alpha_1^{\kappa-\frac{1}{r}}$. Then Assumption 3.3 holds, i.e. x_n is uniquely defined for each n .

We will prove Lemma 3.2 at the end of Section 4.1 by using some useful estimates that will be derived during the proof of the convergence of $\{x_n\}$.

4 Convergence analysis

We prove Theorem 3.1 in this section. We first obtain a convergence result for the noise-free iterative sequences $\{x_n\}$ and $\{\xi_n\}$. We then consider the sequences $\{x_n^\delta\}$ and $\{\xi_n^\delta\}$ corresponding to the noisy data case, and show that the discrepancy principle indeed terminates the iteration in finite steps. We further establish a stability result which in particular implies that $x_n^\delta \rightarrow x_n$ as $\delta \rightarrow 0$ for each fixed n . Combining all these results we finally obtain the proof of Theorem 3.1.

4.1 Convergence result for noise-free case

We first consider the noise-free iterative sequences $\{x_n\}$ and $\{\xi_n\}$ defined by (3.7) and obtain a convergence result that is crucial for proving Theorem 3.1. Our proof is inspired by [9, 14].

Theorem 4.1 *Let \mathcal{X} be reflexive and \mathcal{Y} be uniformly smooth, let $1 < r < \infty$, let Θ satisfy Assumption 3.1, let F satisfy Assumption 3.2, and let $\{\alpha_n\}$ satisfy $\sum_{n=1}^{\infty} \alpha_n^{-1} = \infty$. Assume that*

$$D_{\xi_0} \Theta(x^\dagger, x_0) \leq \varphi(\rho). \quad (4.1)$$

Then there exists a solution x_ of (1.1) in $B_{3\rho}(x_0) \cap D(\Theta)$ 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 $\mathcal{N}(F'(x^\dagger)) \subset \mathcal{N}(F'(x))$ for all $x \in B_{3\rho}(x_0) \cap D(F)$, then $x_ = x^\dagger$.*

Proof We first show by induction that for any solution \hat{x} of (1.1) in $B_{3\rho}(x_0) \cap D(\Theta)$ there holds

$$D_{\xi_n} \Theta(\hat{x}, x_n) \leq D_{\xi_0} \Theta(\hat{x}, x_0), \quad n = 0, 1, \dots \quad (4.2)$$

This is trivial for $n = 0$. Assume that it is true for $n = m - 1$ for some $m \geq 1$, we will show that it is also true for $n = m$. From (2.2) we have

$$D_{\xi_m} \Theta(\hat{x}, x_m) - D_{\xi_{m-1}} \Theta(\hat{x}, x_{m-1}) = -D_{\xi_{m-1}} \Theta(x_m, x_{m-1}) + \langle \xi_{m-1} - \xi_m, \hat{x} - x_m \rangle.$$

By dropping the first term on the right which is non-positive and using the definition of ξ_m we can obtain

$$D_{\xi_m} \Theta(\hat{x}, x_m) - D_{\xi_{m-1}} \Theta(\hat{x}, x_{m-1}) \leq \frac{1}{\alpha_m} \langle J_r(F(x_m) - y), F'(x_m)(\hat{x} - x_m) \rangle.$$

In view of the properties of the duality mapping J_r it follows that

$$\begin{aligned} & D_{\xi_m} \Theta(\hat{x}, x_m) - D_{\xi_{m-1}} \Theta(\hat{x}, x_{m-1}) \\ & \leq -\frac{1}{\alpha_m} \|F(x_m) - y\|^r + \frac{1}{\alpha_m} \|F(x_m) - y\|^{r-1} \|F(x_m) - y + F'(x_m)(\hat{x} - x_m)\|. \end{aligned} \quad (4.3)$$

In order to proceed further, we need to show that $x_m \in B_{3\rho}(x_0)$ so that Assumption 3.2 (d) on F can be employed. Using the minimizing property of x_m , the induction hypothesis, and (4.1) we obtain

$$D_{\xi_{m-1}} \Theta(x_m, x_{m-1}) \leq D_{\xi_{m-1}} \Theta(x^\dagger, x_{m-1}) \leq D_{\xi_0} \Theta(x^\dagger, x_0) \leq \varphi(\rho).$$

With the help of Assumption 3.1 on Θ , we have

$$\|x_m - x_{m-1}\| \leq \rho, \quad \|x^\dagger - x_{m-1}\| \leq \rho \quad \text{and} \quad \|x^\dagger - x_0\| \leq \rho.$$

Therefore $x_m \in B_{3\rho}(x_0)$. Thus we may use Assumption 3.2 (d) to obtain from (4.3) that

$$D_{\xi_m} \Theta(\hat{x}, x_m) - D_{\xi_{m-1}} \Theta(\hat{x}, x_{m-1}) \leq -\frac{1-\eta}{\alpha_m} \|F(x_m) - y\|^r. \quad (4.4)$$

This and the induction hypothesis imply (4.2) with $n = m$.

As an immediate consequence of (4.2), we know that (4.4) is true for all m . Consequently

$$D_{\xi_n} \Theta(\hat{x}, x_n) \leq D_{\xi_{n-1}} \Theta(\hat{x}, x_{n-1}), \quad n = 1, 2, \dots \quad (4.5)$$

and

$$\frac{1-\eta}{\alpha_n} \|F(x_n) - y\|^r \leq D_{\xi_{n-1}} \Theta(\hat{x}, x_{n-1}) - D_{\xi_n} \Theta(\hat{x}, x_n). \quad (4.6)$$

By using the monotonicity of $\|F(x_n) - y\|$ with respect to n , we obtain

$$\|F(x_n) - y\|^r \sum_{j=1}^n \frac{1}{\alpha_j} \leq \sum_{j=1}^n \frac{1}{\alpha_j} \|F(x_j) - y\|^r \leq \frac{1}{1-\eta} D_{\xi_0} \Theta(\hat{x}, x_0).$$

Since $\sum_{j=1}^n \alpha_j^{-1} \rightarrow \infty$ as $n \rightarrow \infty$, we have $\|F(x_n) - y\| \rightarrow 0$ as $n \rightarrow \infty$.

Next we show that $\{x_n\}$ converges to a solution of (1.1). To this end, we show that $\{x_n\}$ is a Cauchy sequence in \mathcal{X} . For $0 \leq l < m < \infty$ we have from (2.2) that

$$D_{\xi_l} \Theta(x_m, x_l) = D_{\xi_l} \Theta(\hat{x}, x_l) - D_{\xi_m} \Theta(\hat{x}, x_m) + \langle \xi_m - \xi_l, x_m - \hat{x} \rangle.$$

By the definition of ξ_n we have

$$\begin{aligned} |\langle \xi_m - \xi_l, x_m - \hat{x} \rangle| &= \left| \sum_{n=l+1}^m \langle \xi_n - \xi_{n-1}, x_m - \hat{x} \rangle \right| \\ &= \left| \sum_{n=l+1}^m \frac{1}{\alpha_n} \langle J_r(F(x_n) - y), F'(x_n)(x_m - \hat{x}) \rangle \right| \\ &\leq \sum_{n=l+1}^m \frac{1}{\alpha_n} \|F(x_n) - y\|^{r-1} \|F'(x_n)(x_m - \hat{x})\|. \end{aligned} \quad (4.7)$$

By using Assumption 3.2 (d) on F and the monotonicity of $\|F(x_n) - y\|$ we can obtain

$$\begin{aligned} \|F'(x_n)(x_m - \hat{x})\| &\leq \|F'(x_n)(x_n - \hat{x})\| + \|F'(x_n)(x_m - x_n)\| \\ &\leq (1 + \eta) (\|F(x_n) - y\| + \|F(x_m) - F(x_n)\|) \\ &\leq 3(1 + \eta) \|F(x_n) - y\|. \end{aligned} \quad (4.8)$$

Therefore, by using (4.6), we have with $c_0 := 3(1 + \eta)/(1 - \eta)$ that

$$\begin{aligned} |\langle \xi_m - \xi_l, x_m - \hat{x} \rangle| &\leq 3(1 + \eta) \sum_{n=l+1}^m \frac{1}{\alpha_n} \|F(x_n) - y\|^r \\ &\leq c_0 (D_{\xi_l} \Theta(\hat{x}, x_l) - D_{\xi_m} \Theta(\hat{x}, x_m)). \end{aligned} \quad (4.9)$$

Consequently

$$D_{\xi_l} \Theta(x_m, x_l) \leq (1 + c_0) (D_{\xi_l} \Theta(\hat{x}, x_l) - D_{\xi_m} \Theta(\hat{x}, x_m)).$$

Since $\{D_{\xi_n} \Theta(\hat{x}, x_n)\}$ is monotonically decreasing, we obtain $D_{\xi_l} \Theta(x_m, x_l) \rightarrow 0$ as $l, m \rightarrow \infty$. In view of the uniform convexity of Θ , we can conclude that $\{x_n\}$ is a Cauchy sequence in \mathcal{X} . Thus $x_n \rightarrow x_*$ for some $x_* \in \mathcal{X}$ as $n \rightarrow \infty$. Since $\|F(x_n) - y\| \rightarrow 0$ as $n \rightarrow \infty$, we may use the weakly closedness of F to conclude that $x_* \in D(F)$ and $F(x_*) = y$. We remark that $x_* \in B_{3\rho}(x_0)$ because $x_n \in B_{3\rho}(x_0)$.

Next we show that

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

From the convexity of Θ and $\xi_n \in \partial\Theta(x_n)$ it follows that

$$\Theta(x_n) \leq \Theta(\hat{x}) + \langle \xi_n, x_n - \hat{x} \rangle. \quad (4.10)$$

In view of (4.9) we have

$$\Theta(x_n) \leq \Theta(\hat{x}) + \langle \xi_0, x_n - \hat{x} \rangle + c_0 D_{\xi_0} \Theta(\hat{x}, x_0).$$

Since $x_n \rightarrow x_*$ as $n \rightarrow \infty$, by using the weakly lower semi-continuity of Θ we obtain

$$\Theta(x_*) \leq \liminf_{n \rightarrow \infty} \Theta(x_n) \leq \Theta(\hat{x}) + \langle \xi_0, x_* - \hat{x} \rangle + c_0 D_{\xi_0} \Theta(\hat{x}, x_0) < \infty. \quad (4.11)$$

This implies that $x_* \in D(\Theta)$. We next use (4.9) to derive for $l < n$ that

$$|\langle \xi_n, x_n - x_* \rangle| \leq c_0 (D_{\xi_l} \Theta(x_*, x_l) - D_{\xi_n} \Theta(x_*, x_n)) + |\langle \xi_l, x_n - x_* \rangle|.$$

By taking $n \rightarrow \infty$ and using $x_n \rightarrow x_*$ we can derive that

$$\limsup_{n \rightarrow \infty} |\langle \xi_n, x_n - x_* \rangle| \leq c_0 (D_{\xi_l} \Theta(x_*, x_l) - \varepsilon_0),$$

where $\varepsilon_0 := \lim_{n \rightarrow \infty} D_{\xi_n} \Theta(x_*, x_n)$ whose existence is guaranteed by the monotonicity of $\{D_{\xi_n} \Theta(x_*, x_n)\}$. Since the above inequality holds for all l , by taking $l \rightarrow \infty$ we obtain

$$\limsup_{n \rightarrow \infty} |\langle \xi_n, x_n - x_* \rangle| \leq c_0 (\varepsilon_0 - \varepsilon_0) = 0. \quad (4.12)$$

Using (4.10) with \hat{x} replaced by x_* we thus obtain $\limsup_{n \rightarrow \infty} \Theta(x_n) \leq \Theta(x_*)$. Combining this with (4.11) we therefore obtain $\lim_{n \rightarrow \infty} \Theta(x_n) = \Theta(x_*)$. This together with (4.12) then implies that $\lim_{n \rightarrow \infty} D_{\xi_n} \Theta(x_*, x_n) = 0$.

Finally we prove $x_* = x^\dagger$ under the additional condition $\mathcal{N}(F'(x^\dagger)) \subset \mathcal{N}(F'(x))$ for $x \in B_{3\rho}(x_0) \cap D(F)$. We use (4.10) with \hat{x} replaced by x^\dagger to obtain

$$D_{\xi_0} \Theta(x_n, x_0) \leq D_{\xi_0} \Theta(x^\dagger, x_0) + \langle \xi_n - \xi_0, x_n - x^\dagger \rangle. \quad (4.13)$$

By using (4.9), for any $\varepsilon > 0$ we can find l_0 such that

$$|\langle \xi_n - \xi_{l_0}, x_n - x^\dagger \rangle| < \frac{\varepsilon}{2}, \quad n \geq l_0.$$

We next consider $\langle \xi_{l_0} - \xi_0, x_n - x^\dagger \rangle$. According to the definition of ξ_n we have $\xi_j - \xi_{j-1} \in \mathcal{R}(F'(x_j)^*)$. Since \mathcal{X} is reflexive and $\mathcal{N}(F'(x^\dagger)) \subset \mathcal{N}(F'(x_j))$, we have from (2.1) that $\overline{\mathcal{R}(F'(x_j)^*)} \subset \overline{\mathcal{R}(F'(x^\dagger)^*)}$. Thus we can find $v_j \in \mathcal{Y}^*$ and $\beta_j \in \mathcal{X}^*$ such that

$$\xi_j - \xi_{j-1} = F'(x^\dagger)^* v_j + \beta_j \quad \text{and} \quad \|\beta_j\| \leq \frac{\varepsilon}{3l_0 M}, \quad 1 \leq j \leq l_0,$$

where $M > 0$ is a constant such that $\|x_n - x^\dagger\| \leq M$ for all n . Consequently

$$\begin{aligned}
|\langle \xi_{l_0} - \xi_0, x_n - x^\dagger \rangle| &= \left| \sum_{j=1}^{l_0} \langle \xi_j - \xi_{j-1}, x_n - x^\dagger \rangle \right| \\
&= \left| \sum_{j=1}^{l_0} [\langle v_j, F'(x^\dagger)(x_n - x^\dagger) \rangle + \langle \beta_j, x_n - x^\dagger \rangle] \right| \\
&\leq \sum_{j=1}^{l_0} (\|v_j\| \|F'(x^\dagger)(x_n - x^\dagger)\| + \|\beta_j\| \|x_n - x^\dagger\|) \\
&\leq (1 + \eta) \sum_{j=1}^{l_0} \|v_j\| \|F(x_n) - y\| + \frac{\varepsilon}{3}.
\end{aligned}$$

Since $\|F(x_n) - y\| \rightarrow 0$ as $n \rightarrow \infty$, we can find $n_0 \geq l_0$ such that

$$|\langle \xi_{l_0} - \xi_0, x_n - x^\dagger \rangle| < \frac{\varepsilon}{2}, \quad \forall n \geq n_0.$$

Therefore $|\langle \xi_n - \xi_0, x_n - x^\dagger \rangle| < \varepsilon$ for all $n \geq n_0$. Since $\varepsilon > 0$ is arbitrary, we obtain $\lim_{n \rightarrow \infty} \langle \xi_n - \xi_0, x_n - x^\dagger \rangle = 0$. By taking $n \rightarrow \infty$ in (4.13) and using $\Theta(x_n) \rightarrow \Theta(x_*)$ we obtain

$$D_{\xi_0} \Theta(x_*, x_0) \leq D_{\xi_0} \Theta(x^\dagger, x_0).$$

According to the definition of x^\dagger we must have $D_{\xi_0} \Theta(x_*, x_0) = D_{\xi_0} \Theta(x^\dagger, x_0)$. A direct application of Lemma 3.1 gives $x_* = x^\dagger$. \square

As a byproduct, now we can use some estimates established in the proof of Theorem 4.1 to prove Lemma 3.2.

Proof of Lemma 3.2. We assume that the minimization problem in (3.7) has two minimizers x_n and \hat{x}_n . Then it follows that

$$\begin{aligned}
0 &= \frac{1}{r} \|F(\hat{x}_n) - y\|^r + \alpha_n D_{\xi_{n-1}} \Theta(\hat{x}_n, x_{n-1}) - \frac{1}{r} \|F(x_n) - y\|^r \\
&\quad - \alpha_n D_{\xi_{n-1}} \Theta(x_n, x_{n-1}) \\
&= \Delta_r (F(\hat{x}_n) - y, F(x_n) - y) + \langle J_r(F(x_n) - y), F(\hat{x}_n) - F(x_n) \rangle \\
&\quad + \alpha_n (\Theta(\hat{x}_n) - \Theta(x_n) - \langle \xi_{n-1}, \hat{x}_n - x_n \rangle).
\end{aligned}$$

With the help of the definition of ξ_n we can write

$$\begin{aligned}
&\Theta(\hat{x}_n) - \Theta(x_n) - \langle \xi_{n-1}, \hat{x}_n - x_n \rangle \\
&= \Theta(\hat{x}_n) - \Theta(x_n) - \langle \xi_n, \hat{x}_n - x_n \rangle + \langle \xi_n - \xi_{n-1}, \hat{x}_n - x_n \rangle \\
&= D_{\xi_n} \Theta(\hat{x}_n, x_n) - \frac{1}{\alpha_n} \langle J_r(F(x_n) - y), F'(\xi_n)(\hat{x}_n - x_n) \rangle.
\end{aligned}$$

Therefore

$$\begin{aligned}
0 &= \Delta_r (F(\hat{x}_n) - y, F(x_n) - y) + \alpha_n D_{\xi_n} \Theta(\hat{x}_n, x_n) \\
&\quad + \langle J_r(F(x_n) - y), F(\hat{x}_n) - F(x_n) - F'(\xi_n)(\hat{x}_n - x_n) \rangle.
\end{aligned}$$

Since $x_n, \hat{x}_n \in B_{3\rho}(x_0)$ as shown in the proof of Theorem 4.1, we may use Assumption 3.4 and the Young's inequality to obtain

$$\begin{aligned} 0 &\geq \Delta_r(F(\hat{x}_n) - y, F(x_n) - y) + \alpha_n D_{\xi_n} \Theta(\hat{x}_n, x_n) \\ &\quad - C_0 \|F(x_n) - y\|^{r-1} [D_{\xi_n} \Theta(\hat{x}_n, x_n)]^{1-\kappa} [\Delta_r(F(\hat{x}_n) - y, F(x_n) - y)]^\kappa \\ &\geq \alpha_n D_{\xi_n} \Theta(\hat{x}_n, x_n) - (1 - \kappa) \kappa^{\frac{\kappa}{1-\kappa}} C_0^{\frac{1}{1-\kappa}} \|F(x_n) - y\|^{\frac{r-1}{1-\kappa}} D_{\xi_n} \Theta(\hat{x}_n, x_n). \end{aligned}$$

Recall that in the proof of Theorem 4.1 we have established

$$\|F(x_n) - y\|^r \leq \frac{1}{1-\eta} s_n^{-1} D_{\xi_0} \Theta(x^\dagger, x_0) \quad \text{with } s_n := \sum_{j=1}^n \alpha_j^{-1}.$$

Since $s_n^{-1} \leq \min\{\alpha_1, \alpha_n\}$ and $\kappa \geq 1/r$, we therefore obtain

$$0 \geq \left(1 - \bar{C}_0^{\frac{1}{1-\kappa}} D_{\xi_0} \Theta(x^\dagger, x_0)^{\frac{r-1}{r(1-\kappa)}}\right) \alpha_n D_{\xi_n} \Theta(\hat{x}_n, x_n)$$

with $\bar{C}_0 := C_0 \kappa^\kappa (1 - \kappa)^{1-\kappa} (1 - \eta)^{\frac{1-r}{r}} \alpha_1^{\kappa - \frac{1}{r}}$. Thus we may use the second condition in (3.11) to conclude that $D_{\xi_n} \Theta(\hat{x}_n, x_n) = 0$ and hence $\hat{x}_n = x_n$. \square

4.2 Justification of the method

In this subsection we show that the method is well-defined, in particular we prove that, when the data contains noise, the discrepancy principle (3.6) terminates the iteration in finite steps, i.e. $n_\delta < \infty$.

Lemma 4.1 *Let \mathcal{X} be reflexive and \mathcal{Y} be uniformly smooth, let Θ satisfy Assumption 3.1, and let F satisfy Assumption 3.2. Let $1 < r < \infty$ and $\tau > (1 + \eta)/(1 - \eta)$, and let $\{\alpha_n\}$ be such that $\sum_{n=1}^\infty \alpha_n^{-1} = \infty$. Assume that (4.1) holds. Then the discrepancy principle (3.6) terminates the iteration after $n_\delta < \infty$ steps. If $n_\delta \geq 2$, then for $1 \leq n < n_\delta$ there hold*

$$D_{\xi_n^\delta} \Theta(\hat{x}, x_n^\delta) \leq D_{\xi_{n-1}^\delta} \Theta(\hat{x}, x_{n-1}^\delta), \quad (4.14)$$

$$\frac{1}{\alpha_n} \|F(x_n^\delta) - y^\delta\|^r \leq C_1 \left(D_{\xi_{n-1}^\delta} \Theta(\hat{x}, x_{n-1}^\delta) - D_{\xi_n^\delta} \Theta(\hat{x}, x_n^\delta) \right). \quad (4.15)$$

If, in addition, $\alpha_n \leq c_0 \alpha_{n+1}$ for all n with some constant $c_0 > 0$ and

$$D_{\xi_0} \Theta(x^\dagger, x_0) \leq \frac{\tau^r - 1}{\tau^r - 1 + c_0} \varphi(\rho), \quad (4.16)$$

then there holds

$$D_{\xi_{n_\delta}^\delta} \Theta(\hat{x}, x_{n_\delta}^\delta) \leq D_{\xi_{n_\delta-1}^\delta} \Theta(\hat{x}, x_{n_\delta-1}^\delta) + (1 + \eta) \tau^{r-1} \frac{\delta^r}{\alpha_{n_\delta}}, \quad (4.17)$$

where \hat{x} denotes any solution of (1.1) in $B_{3\rho}(x_0) \cap D(\Theta)$ and $C_1 := \tau/[(1 - \eta)\tau - 1 - \eta]$.

Proof To prove the first part, we first show by induction that

$$x_n^\delta \in B_{2\rho}(x_0) \quad \text{and} \quad D_{\xi_n^\delta} \Theta(x^\dagger, x_n^\delta) \leq D_{\xi_0} \Theta(x^\dagger, x_0), \quad 0 \leq n < n_\delta. \quad (4.18)$$

This is trivial for $n = 0$. Next we assume that (4.18) is true for $n = m - 1$ for some $m < n_\delta$ and show that (4.18) is also true for $n = m$. By the minimizing property of x_m^δ and the induction hypothesis we have

$$\begin{aligned} \frac{1}{r} \|F(x_m^\delta) - y^\delta\|^r + \alpha_m D_{\xi_{m-1}^\delta} \Theta(x_m^\delta, x_{m-1}^\delta) &\leq \frac{1}{r} \delta^r + \alpha_m D_{\xi_{m-1}^\delta} \Theta(x^\dagger, x_{m-1}^\delta) \\ &\leq \frac{1}{r} \delta^r + \alpha_m D_{\xi_0} \Theta(x^\dagger, x_0). \end{aligned} \quad (4.19)$$

Since $\|F(x_m^\delta) - y^\delta\| > \tau\delta$, we can obtain

$$\frac{\tau^r}{r} \delta^r + \alpha_m D_{\xi_{m-1}^\delta} \Theta(x_m^\delta, x_{m-1}^\delta) \leq \frac{1}{r} \delta^r + \alpha_m D_{\xi_0} \Theta(x^\dagger, x_0).$$

Because $\tau > 1$, this implies that

$$\alpha_m \geq \frac{(\tau^r - 1)\delta^r}{r D_{\xi_0} \Theta(x^\dagger, x_0)} \quad \text{and} \quad D_{\xi_{m-1}^\delta} \Theta(x_m^\delta, x_{m-1}^\delta) \leq D_{\xi_0} \Theta(x^\dagger, x_0). \quad (4.20)$$

By Assumption 3.1 and the condition (4.1), we can derive that $\|x_m^\delta - x_{m-1}^\delta\| \leq \rho$. In view of the induction hypothesis we also have $\|x_{m-1}^\delta - x_0\| \leq 2\rho$. Thus $x_m^\delta \in B_{3\rho}(x_0)$.

We are now able to use Assumption 3.2 (d) and the similar argument for deriving (4.3) to obtain that

$$\begin{aligned} D_{\xi_m^\delta} \Theta(\hat{x}, x_m^\delta) - D_{\xi_{m-1}^\delta} \Theta(\hat{x}, x_{m-1}^\delta) &\leq \langle \xi_m^\delta - \xi_{m-1}^\delta, x_m^\delta - \hat{x} \rangle = -\frac{1}{\alpha_m} \langle J_r(F(x_m^\delta) - y^\delta), F'(x_m^\delta)(x_m^\delta - \hat{x}) \rangle \\ &\leq -\frac{1}{\alpha_m} \|F(x_m^\delta) - y^\delta\|^r + \frac{1}{\alpha_m} \|F(x_m^\delta) - y^\delta\|^{r-1} (\delta + \eta \|F(x_m^\delta) - y^\delta\|) \\ &\leq -\frac{1-\eta}{\alpha_m} \|F(x_m^\delta) - y^\delta\|^r + \frac{1+\eta}{\alpha_m} \|F(x_m^\delta) - y^\delta\|^{r-1} \delta. \end{aligned} \quad (4.21)$$

Using again $\|F(x_m^\delta) - y^\delta\| > \tau\delta$, we can conclude that

$$D_{\xi_m^\delta} \Theta(\hat{x}, x_m^\delta) - D_{\xi_{m-1}^\delta} \Theta(\hat{x}, x_{m-1}^\delta) \leq -\frac{1}{\alpha_m} \left(1 - \eta - \frac{1+\eta}{\tau}\right) \|F(x_m^\delta) - y^\delta\|^r. \quad (4.22)$$

Since $\tau > (1 + \eta)/(1 - \eta)$, we obtain

$$D_{\xi_m^\delta} \Theta(\hat{x}, x_m^\delta) \leq D_{\xi_{m-1}^\delta} \Theta(\hat{x}, x_{m-1}^\delta).$$

In view of this inequality with $\hat{x} = x^\dagger$ and the induction hypothesis, we obtain the second result in (4.18) with $n = m$. By using again Assumption 3.1 and (4.1) we have $\|x_m^\delta - x^\dagger\| \leq \rho$ and $\|x^\dagger - x_0\| \leq \rho$ which imply that $x_m^\delta \in B_{2\rho}(x_0)$. We therefore complete the proof of (4.18). As a direct consequence, we can see that (4.22) holds for all $1 \leq m < n_\delta$ which implies (4.14) and (4.15).

In view of (4.15) and the monotonicity (3.5) of $\|F(x_n^\delta) - y^\delta\|$ with respect to n , it follows that

$$\|F(x_n^\delta) - y^\delta\|^r \sum_{j=1}^n \frac{1}{\alpha_j} \leq \sum_{j=1}^n \frac{1}{\alpha_j} \|F(x_j^\delta) - y^\delta\|^r \leq \frac{\tau}{(1-\eta)\tau - 1 - \eta} D_{\xi_0} \Theta(\hat{x}, x_0).$$

Since $\|F(x_n^\delta) - y^\delta\| > \tau\delta$ for $1 \leq n < n_\delta$ and $\sum_{j=1}^n \alpha_j^{-1} \rightarrow \infty$ as $n \rightarrow \infty$, we can conclude that n_δ is a finite integer.

Finally we prove the second part, i.e. the inequality (4.17). Since (4.19) is true for $m = n_\delta$, we have

$$D_{\xi_{n_\delta-1}^\delta} \Theta(x_{n_\delta}^\delta, x_{n_\delta-1}^\delta) \leq \frac{\delta^r}{r\alpha_{n_\delta}} + D_{\xi_0} \Theta(x^\dagger, x_0).$$

Recall from (4.20) that $\alpha_{n_\delta-1} \geq (\tau^r - 1)\delta^r / (rD_{\xi_0} \Theta(x^\dagger, x_0))$. Since $\alpha_{n_\delta-1} \leq c_0\alpha_{n_\delta}$, we can derive that

$$D_{\xi_{n_\delta-1}^\delta} \Theta(x_{n_\delta}^\delta, x_{n_\delta-1}^\delta) \leq \frac{\tau^r - 1 + c_0}{\tau^r - 1} D_{\xi_0} \Theta(x^\dagger, x_0).$$

It then follows from Assumption 3.1 and (4.16) that $\|x_{n_\delta}^\delta - x_{n_\delta-1}^\delta\| \leq \rho$. Since $x_{n_\delta-1}^\delta \in B_{2\rho}(x_0)$ we obtain $x_{n_\delta}^\delta \in B_{3\rho}(x_0)$. Thus we can employ Assumption 3.2 (d) to conclude that (4.21) is also true for $m = n_\delta$. By setting $m = n_\delta$ in (4.21) and using $\|F(x_{n_\delta}^\delta) - y^\delta\| \leq \tau\delta$, we can obtain (4.17). \square

As a byproduct of the proof of Lemma 4.1, we have the following result which will be used to show $\lim_{\delta \rightarrow 0} \Theta(x_{n_\delta}^\delta) = \Theta(x_*)$ in the proof of Theorem 3.1.

Lemma 4.2 *Let all the conditions in Lemma 4.1 hold, and let \hat{x} be any solution of (1.1) in $B_{3\rho}(x_0) \cap D(\Theta)$. Then for all $0 \leq l < n_\delta$ there holds*

$$|\langle \xi_{n_\delta}^\delta - \xi_l^\delta, \hat{x} - x_{n_\delta}^\delta \rangle| \leq C_2 \frac{\delta^r}{\alpha_{n_\delta}} + C_3 D_{\xi_l^\delta} \Theta(\hat{x}, x_l^\delta), \quad (4.23)$$

where $C_2 := 3(1 + \eta)\tau^{r-1}(1 + \tau)$ and $C_3 := 3(1 + \eta)(1 + \tau)/[(1 - \eta)\tau - 1 - \eta]$.

Proof By the definition of ξ_n^δ and the property of the duality mapping J_r , we can obtain, using the similar argument for deriving (4.7), that

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

With the help of Assumption 3.2 (d) and the monotonicity (3.5) of $\|F(x_n^\delta) - y^\delta\|$ with respect to n , similar to the derivation of (4.8) we have for $n \leq n_\delta$ that

$$\|F'(x_n^\delta)(\hat{x} - x_{n_\delta}^\delta)\| \leq 3(1 + \eta) (\|F(x_n^\delta) - y^\delta\| + \delta).$$

Therefore

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

Since $\|F(x_{n_\delta}^\delta) - y^\delta\| \leq \tau\delta$ and $\|F(x_n^\delta) - y^\delta\| > \tau\delta$ for $0 \leq n < n_\delta$, we thus obtain

$$\begin{aligned} & |\langle \xi_{n_\delta}^\delta - \xi_l^\delta, \hat{x} - x_{n_\delta}^\delta \rangle| \\ & \leq 3(1 + \eta)\tau^{r-1}(1 + \tau) \frac{\delta^r}{\alpha_{n_\delta}} + \frac{3(1 + \eta)(1 + \tau)}{\tau} \sum_{n=l+1}^{n_\delta-1} \frac{1}{\alpha_n} \|F(x_n^\delta) - y^\delta\|^r. \end{aligned} \quad (4.24)$$

In view of (4.15) in Lemma 4.1, we can see that

$$\sum_{n=l+1}^{n_\delta-1} \frac{1}{\alpha_n} \|F(x_n^\delta) - y^\delta\|^r \leq \frac{\tau}{(1 - \eta)\tau - 1 - \eta} D_{\xi_l^\delta} \Theta(\hat{x}, x_l^\delta).$$

Combining this inequality with (4.24) gives the desired estimate. \square

4.3 Stability

We will prove some stability results on the method which connect $\{x_n^\delta\}$ with $\{x_n\}$. These results enable us to use Theorem 4.1 to complete the proof of Theorem 3.1.

Lemma 4.3 *Let \mathcal{X} be reflexive and \mathcal{Y} be uniformly smooth, let Θ satisfy Assumption 3.1, and let F satisfy Assumptions 3.2 and 3.3. Then for each fixed n there hold*

$$x_n^\delta \rightarrow x_n, \quad \Theta(x_n^\delta) \rightarrow \Theta(x_n) \quad \text{and} \quad \xi_n^\delta \rightarrow \xi_n \quad (4.25)$$

as $y^\delta \rightarrow y$.

Proof We show this result by induction. It is trivial when $n = 0$ since $x_0^\delta = x_0$ and $\xi_0^\delta = \xi_0$. In the following we assume that the result is proved for $n = m - 1$ and show that the result holds also for $n = m$.

We will adapt the argument from [5]. Let $\{y^{\delta_i}\}$ be a sequence of data satisfying $\|y^{\delta_i} - y\| \leq \delta_i$ with $\delta_i \rightarrow 0$. By the minimizing property of $x_m^{\delta_i}$ we have

$$\frac{1}{r} \|F(x_m^{\delta_i}) - y^{\delta_i}\|^r + \alpha_m D_{\xi_{m-1}^{\delta_i}} \Theta(x_m^{\delta_i}, x_{m-1}^{\delta_i}) \leq \frac{1}{r} \|F(x_{m-1}^{\delta_i}) - y^{\delta_i}\|^r.$$

By the induction hypothesis, we can see that the right hand side of the above inequality is uniformly bounded with respect to i . Therefore both $\{\|F(x_m^{\delta_i}) - y^{\delta_i}\|\}$ and $\{D_{\xi_{m-1}^{\delta_i}} \Theta(x_m^{\delta_i}, x_{m-1}^{\delta_i})\}$ are uniformly bounded with respect to i . Consequently $\{F(x_m^{\delta_i})\}$ is bounded in \mathcal{Y} and $\{x_m^{\delta_i}\}$ is bounded in \mathcal{X} ; here we used the uniformly convexity of Θ . Since both \mathcal{X} and \mathcal{Y} are reflexive, by taking a subsequence if necessary, we may assume that $x_m^{\delta_i} \rightharpoonup \bar{x}_m \in \mathcal{X}$ and $F(x_m^{\delta_i}) \rightharpoonup \bar{y}_m \in \mathcal{Y}$ as $i \rightarrow \infty$. Since F is weakly closed, we have $\bar{x}_m \in D(F)$ and $F(\bar{x}_m) = \bar{y}_m$. In view of the weakly lower semi-continuity of Banach space norm we have

$$\|F(\bar{x}_m) - y\| \leq \liminf_{i \rightarrow \infty} \|F(x_m^{\delta_i}) - y^{\delta_i}\|. \quad (4.26)$$

Moreover, by using $x_m^{\delta_i} \rightharpoonup \bar{x}_m$, the weakly lower semi-continuity of Θ , and the induction hypothesis, we have

$$\begin{aligned} \liminf_{i \rightarrow \infty} D_{\xi_{m-1}^{\delta_i}} \Theta(x_m^{\delta_i}, x_{m-1}^{\delta_i}) &= \liminf_{i \rightarrow \infty} \Theta(x_m^{\delta_i}) - \Theta(x_{m-1}) - \langle \xi_{m-1}, \bar{x}_m - x_{m-1} \rangle \\ &\geq \Theta(\bar{x}_m) - \Theta(x_{m-1}) - \langle \xi_{m-1}, \bar{x}_m - x_{m-1} \rangle \\ &= D_{\xi_{m-1}} \Theta(\bar{x}_m, x_{m-1}). \end{aligned} \quad (4.27)$$

The inequalities (4.26) and (4.27) together with the minimizing property of $x_m^{\delta_i}$ and the induction hypothesis imply

$$\begin{aligned} &\frac{1}{r} \|F(\bar{x}_m) - y\|^r + \alpha_m D_{\xi_{m-1}} \Theta(\bar{x}_m, x_{m-1}) \\ &\leq \liminf_{i \rightarrow \infty} \left\{ \frac{1}{r} \|F(x_m^{\delta_i}) - y^{\delta_i}\|^r + \alpha_m D_{\xi_{m-1}^{\delta_i}} \Theta(x_m^{\delta_i}, x_{m-1}^{\delta_i}) \right\} \\ &\leq \limsup_{i \rightarrow \infty} \left\{ \frac{1}{r} \|F(x_m^{\delta_i}) - y^{\delta_i}\|^r + \alpha_m D_{\xi_{m-1}^{\delta_i}} \Theta(x_m^{\delta_i}, x_{m-1}^{\delta_i}) \right\} \\ &\leq \limsup_{i \rightarrow \infty} \left\{ \frac{1}{r} \|F(x_m) - y^{\delta_i}\|^r + \alpha_m D_{\xi_{m-1}^{\delta_i}} \Theta(x_m, x_{m-1}^{\delta_i}) \right\} \\ &= \frac{1}{r} \|F(x_m) - y\|^r + \alpha_m D_{\xi_{m-1}} \Theta(x_m, x_{m-1}). \end{aligned}$$

According to the definition of x_m and Assumption 3.3, we must have $\bar{x}_m = x_m$. Therefore $x_m^{\delta_i} \rightharpoonup x_m$, $F(x_m^{\delta_i}) \rightharpoonup F(x_m)$, and

$$\begin{aligned} \lim_{i \rightarrow \infty} \left\{ \frac{1}{r} \|F(x_m^{\delta_i}) - y^{\delta_i}\|^r + \alpha_m D_{\xi_{m-1}^{\delta_i}} \Theta(x_m^{\delta_i}, x_{m-1}^{\delta_i}) \right\} \\ = \frac{1}{r} \|F(x_m) - y\|^r + \alpha_m D_{\xi_{m-1}} \Theta(x_m, x_{m-1}). \end{aligned} \quad (4.28)$$

Next we will show that

$$\lim_{i \rightarrow \infty} D_{\xi_{m-1}^{\delta_i}} \Theta(x_m^{\delta_i}, x_{m-1}^{\delta_i}) = D_{\xi_{m-1}} \Theta(x_m, x_{m-1}). \quad (4.29)$$

Let

$$a := \limsup_{i \rightarrow \infty} D_{\xi_{m-1}^{\delta_i}} \Theta(x_m^{\delta_i}, x_{m-1}^{\delta_i}) \quad \text{and} \quad b := D_{\xi_{m-1}} \Theta(x_m, x_{m-1}).$$

In view of (4.27), it suffices to show $a \leq b$. Assume to the contrary that $a > b$. By taking a subsequence if necessary, we may assume that

$$a = \lim_{i \rightarrow \infty} D_{\xi_{m-1}^{\delta_i}} \Theta(x_m^{\delta_i}, x_{m-1}^{\delta_i}).$$

It then follows from (4.28) that

$$\frac{1}{r} \lim_{i \rightarrow \infty} \|F(x_m^{\delta_i}) - y^{\delta_i}\|^r = \frac{1}{r} \|F(x_m) - y\|^r + \alpha_m (b - a) < \frac{1}{r} \|F(x_m) - y\|^r$$

which is a contradiction to (4.26). We therefore obtain (4.29).

By using the induction hypothesis and $x_m^{\delta_i} \rightharpoonup x_m$, we obtain from (4.29) that

$$\lim_{i \rightarrow \infty} \Theta(x_m^{\delta_i}) = \Theta(x_m).$$

Since $x_m^{\delta_i} \rightharpoonup x_m$ and since Θ has the Kadec property, see Lemma 2.1, we obtain that $x_m^{\delta_i} \rightarrow x_m$ as $i \rightarrow \infty$. Finally, from the definition of $\xi_m^{\delta_i}$, the induction hypothesis, and the continuity of the map $x \rightarrow F'(x)$, and the continuity of the duality mapping J_r , it follows that $\xi_m^{\delta_i} \rightarrow \xi_m$ as $i \rightarrow \infty$.

The above argument shows that for any sequence $\{y^{\delta_i}\}$ converging to y , the sequence $\{x_m^{\delta_i}\}$ always has a subsequence, still denoted as $x_m^{\delta_i}$, such that $x_m^{\delta_i} \rightarrow x_m$, $\Theta(x_m^{\delta_i}) \rightarrow \Theta(x_m)$ and $\xi_m^{\delta_i} \rightarrow \xi_m$ as $i \rightarrow \infty$. Therefore, we obtain (4.25) with $n = m$ as $y^{\delta_i} \rightarrow y$. The proof is complete. \square

4.4 Proof of Theorem 3.1

Since other parts have been proved in Lemma 4.1, it remains only to show the convergence result (3.9), where x_* is the limit of $\{x_n\}$ which exists by Theorem 4.1.

Assume first that $\{y^{\delta_i}\}$ is a sequence satisfying $\|y^{\delta_i} - y\| \leq \delta_i$ with $\delta_i \rightarrow 0$ such that $n_{\delta_i} \rightarrow n_0$ as $i \rightarrow \infty$ for some integer n_0 . We may assume $n_{\delta_i} = n_0$ for all i . From the definition of $n_{\delta_i} = n_0$, we have

$$\|F(x_{n_0}^{\delta_i}) - y^{\delta_i}\| \leq \tau \delta_i.$$

Since Lemma 4.3 implies $x_{n_0}^{\delta_i} \rightarrow x_{n_0}$, by letting $i \rightarrow \infty$ we have $F(x_{n_0}) = y$. This together with the definition of x_n implies that $x_n = x_{n_0}$ for all $n \geq n_0$. Since

Theorem 4.1 implies $x_n \rightarrow x_*$ as $n \rightarrow \infty$, we must have $x_{n_0} = x_*$. Consequently, we have from Lemma 4.3 that $x_{n_{\delta_i}}^{\delta_i} \rightarrow x_*$, $\Theta(x_{n_{\delta_i}}^{\delta_i}) = \Theta(x_{n_0}^{\delta_i}) \rightarrow \Theta(x_{n_0}) = \Theta(x_*)$ and

$$D_{\xi_{n_{\delta_i}}^{\delta_i}} \Theta(x_*, x_{n_{\delta_i}}^{\delta_i}) = D_{\xi_{n_0}^{\delta_i}} \Theta(x_{n_0}, x_{n_0}^{\delta_i}) \rightarrow 0$$

as $i \rightarrow \infty$.

Assume next that $\{y^{\delta_i}\}$ is a sequence satisfying $\|y^{\delta_i} - y\| \leq \delta_i$ with $\delta_i \rightarrow 0$ such that $n_i := n_{\delta_i} \rightarrow \infty$ as $i \rightarrow \infty$. We first show that

$$D_{\xi_{n_i-2}^{\delta_i}} \Theta(x_*, x_{n_i-2}^{\delta_i}) \rightarrow 0 \quad \text{as } i \rightarrow \infty. \quad (4.30)$$

Let $\epsilon > 0$ be an arbitrary number. Since Theorem 4.1 implies $D_{\xi_n} \Theta(x_*, x_n) \rightarrow 0$ as $n \rightarrow \infty$, there exists an integer $n(\epsilon)$ such that $D_{\xi_{n(\epsilon)}} \Theta(x_*, x_{n(\epsilon)}) < \epsilon/2$. On the other hand, since Lemma 4.3 implies $x_{n(\epsilon)}^{\delta_i} \rightarrow x_{n(\epsilon)}$, $\Theta(x_{n(\epsilon)}^{\delta_i}) \rightarrow \Theta(x_{n(\epsilon)})$ and $\xi_{n(\epsilon)}^{\delta_i} \rightarrow \xi_{n(\epsilon)}$ as $i \rightarrow \infty$, we can pick an integer $i(\epsilon)$ large enough such that for all $i \geq i(\epsilon)$ there hold $n_i - 2 \geq n(\epsilon)$ and

$$\left| D_{\xi_{n(\epsilon)}^{\delta_i}} \Theta(x_*, x_{n(\epsilon)}^{\delta_i}) - D_{\xi_{n(\epsilon)}} \Theta(x_*, x_{n(\epsilon)}) \right| < \frac{\epsilon}{2}.$$

Therefore, it follows from Lemma 4.1 that

$$D_{\xi_{n_i-2}^{\delta_i}} \Theta(x_*, x_{n_i-2}^{\delta_i}) \leq D_{\xi_{n(\epsilon)}^{\delta_i}} \Theta(x_*, x_{n(\epsilon)}^{\delta_i}) \leq D_{\xi_{n(\epsilon)}} \Theta(x_*, x_{n(\epsilon)}) + \frac{\epsilon}{2} < \epsilon$$

for all $i \geq i(\epsilon)$. Since $\epsilon > 0$ is arbitrary, we thus obtain (4.30). With the help of (4.14), we then obtain

$$D_{\xi_{n_i-1}^{\delta_i}} \Theta(x_*, x_{n_i-1}^{\delta_i}) \rightarrow 0 \quad \text{as } i \rightarrow \infty. \quad (4.31)$$

In view of (4.15) we have

$$\frac{1}{\alpha_{n_i-1}} \|F(x_{n_i-1}^{\delta_i}) - y^{\delta_i}\|^r \leq \frac{\tau}{(1-\eta)\tau - 1 - \eta} D_{\xi_{n_i-2}^{\delta_i}} \Theta(x_*, x_{n_i-2}^{\delta_i}).$$

Since $\|F(x_{n_i-1}^{\delta_i}) - y^{\delta_i}\| > \tau\delta_i$, we can conclude from (4.30) that $\delta_i^r/\alpha_{n_i-1} \rightarrow 0$. Since $\alpha_{n_i-1} \leq c_0\alpha_{n_i}$, we must have $\delta_i^r/\alpha_{n_i} \rightarrow 0$ as $i \rightarrow \infty$. In view of (4.17) and (4.31), we can obtain

$$D_{\xi_{n_i}^{\delta_i}} \Theta(x_*, x_{n_i}^{\delta_i}) \rightarrow 0 \quad \text{as } i \rightarrow \infty, \quad (4.32)$$

which together with the uniformly convexity of Θ implies that $x_{n_i}^{\delta_i} \rightarrow x_*$ as $i \rightarrow \infty$.

Finally we show that $\Theta(x_{n_i}^{\delta_i}) \rightarrow \Theta(x_*)$ as $i \rightarrow \infty$. In view of (4.32), it suffices to show that

$$\langle \xi_{n_i}^{\delta_i}, x_* - x_{n_i}^{\delta_i} \rangle \rightarrow 0 \quad \text{as } i \rightarrow \infty. \quad (4.33)$$

Recall that $\Theta(x_n) \rightarrow \Theta(x_*)$ and $\langle \xi_n, x_* - x_n \rangle \rightarrow 0$ as $n \rightarrow \infty$ which have been established in Theorem 4.1 and its proof. Thus, for any $\epsilon > 0$, we can pick an integer l_0 such that

$$|\Theta(x_{l_0}) - \Theta(x_*)| < \epsilon \quad \text{and} \quad |\langle \xi_{l_0}, x_* - x_{l_0} \rangle| < \epsilon. \quad (4.34)$$

Then, using (4.23) in Lemma 4.2, we can derive

$$\begin{aligned} |\langle \xi_{n_i}^{\delta_i}, x_* - x_{n_i}^{\delta_i} \rangle| &\leq \left| \langle \xi_{l_0}^{\delta_i}, x_* - x_{n_i}^{\delta_i} \rangle \right| + \left| \langle \xi_{n_i}^{\delta_i} - \xi_{l_0}^{\delta_i}, x_* - x_{n_i}^{\delta_i} \rangle \right| \\ &\leq \left| \langle \xi_{l_0}^{\delta_i}, x_* - x_{n_i}^{\delta_i} \rangle \right| + C_2 \frac{\delta_i^r}{\alpha_{n_i}} + C_3 D_{\xi_{l_0}^{\delta_i}} \Theta(x_*, x_{l_0}^{\delta_i}). \end{aligned}$$

By using the definition of Bregman distance and (4.34) we have

$$\begin{aligned} D_{\xi_{l_0}^{\delta_i}} \Theta(x_*, x_{l_0}^{\delta_i}) &= [\Theta(x_*) - \Theta(x_{l_0})] + [\Theta(x_{l_0}) - \Theta(x_{l_0}^{\delta_i})] - \langle \xi_{l_0}, x_* - x_{l_0} \rangle \\ &\quad - \langle \xi_{l_0}, x_{l_0} - x_{l_0}^{\delta_i} \rangle - \langle \xi_{l_0}^{\delta_i} - \xi_{l_0}, x_* - x_{l_0}^{\delta_i} \rangle \\ &\leq 2\epsilon + \left| \Theta(x_{l_0}) - \Theta(x_{l_0}^{\delta_i}) \right| + \left| \langle \xi_{l_0}, x_{l_0} - x_{l_0}^{\delta_i} \rangle \right| + \left| \langle \xi_{l_0}^{\delta_i} - \xi_{l_0}, x_* - x_{l_0}^{\delta_i} \rangle \right|. \end{aligned}$$

Therefore

$$\begin{aligned} |\langle \xi_{n_i}^{\delta_i}, x_* - x_{n_i}^{\delta_i} \rangle| &\leq 2C_3\epsilon + C_2 \frac{\delta_i^r}{\alpha_{n_i}} + \left| \langle \xi_{l_0}^{\delta_i}, x_* - x_{n_i}^{\delta_i} \rangle \right| + C_3 \left| \Theta(x_{l_0}) - \Theta(x_{l_0}^{\delta_i}) \right| \\ &\quad + C_3 \left| \langle \xi_{l_0}, x_{l_0} - x_{l_0}^{\delta_i} \rangle \right| + C_3 \left| \langle \xi_{l_0}^{\delta_i} - \xi_{l_0}, x_* - x_{l_0}^{\delta_i} \rangle \right|. \end{aligned}$$

In view of Lemma 4.3 and the facts that $\delta_i^r/\alpha_{n_i} \rightarrow 0$ and $x_{n_i}^{\delta_i} \rightarrow x_*$ as $i \rightarrow \infty$ which we have established in the above, we can conclude that there is an integer $i_0(\epsilon)$ such that for all $i > i_0(\epsilon)$ there hold $n_i > l_0$ and $|\langle \xi_{n_i}^{\delta_i}, x_* - x_{n_i}^{\delta_i} \rangle| \leq 3C_3\epsilon$. Since $\epsilon > 0$ is arbitrary, we thus obtain (4.33).

4.5 A variant of the discrepancy principle

When n_δ denotes the integer determined by the discrepancy principle (3.6), from Lemma 4.1 we can see that the Bregman distance $D_{\xi_n^\delta} \Theta(x^\dagger, x_n^\delta)$ is decreasing up to $n = n_\delta - 1$. This monotonicity, however, may not hold at $n = n_\delta$. Therefore, it seems reasonable to consider the following variant of the discrepancy principle.

Rule 4.1 *Let $\tau > 1$ be a given number. If $\|F(x_0) - y^\delta\| \leq \tau\delta$, we define $n_\delta := 0$; otherwise we define*

$$n_\delta := \max \{n : \|F(x_n^\delta) - y^\delta\| \geq \tau\delta\},$$

i.e., n_δ is the integer such that

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

We point out that the argument for proving Theorem 3.1 can be used to prove the convergence property of $x_{n_\delta}^\delta$ for n_δ determined by Rule 4.1, we can even drop the condition $\alpha_n \leq c_0\alpha_{n+1}$ on $\{\alpha_n\}$ in Theorem 3.1. In fact we have the following result.

Theorem 4.2 *Let \mathcal{X} be reflexive and \mathcal{Y} be uniformly smooth, Θ satisfy Assumption 3.1, and F satisfy Assumptions 3.2 and 3.3. Let $1 < r < \infty$ and $\tau > (1+\eta)/(1-\eta)$, and let $\{\alpha_n\}$ be such that $\sum_{n=1}^\infty \alpha_n^{-1} = \infty$. Assume further that*

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

Then, the integer n_δ defined by Rule 4.1 is finite. Moreover, there is a solution $x_* \in D(\Theta)$ of (1.1) such that

$$x_{n_\delta}^\delta \rightarrow x_*, \quad \Theta(x_{n_\delta}^\delta) \rightarrow \Theta(x_*) \quad \text{and} \quad D_{\xi_{n_\delta}^\delta} \Theta(x_*, x_{n_\delta}^\delta) \rightarrow 0 \quad (4.35)$$

as $\delta \rightarrow 0$. If, in addition, $\mathcal{N}(F'(x^\dagger)) \subset \mathcal{N}(F'(x))$ for all $x \in B_{3\rho}(x_0) \cap D(F)$, then $x_* = x^\dagger$.

Proof The proof of Lemma 4.1 can be used without change to show that $n_\delta < \infty$ and that (4.14) and (4.15) hold for $1 \leq n \leq n_\delta$. Consequently, (4.23) in Lemma 4.2 becomes

$$|\langle \xi_{n_\delta}^\delta - \xi_l^\delta, x_* - x_{n_\delta}^\delta \rangle| \leq \frac{3(1+\eta)(1+\tau)}{(1-\eta)\tau - 1 - \eta} D_{\xi_l^\delta} \Theta(x_*, x_l^\delta), \quad 0 \leq l < n_\delta. \quad (4.36)$$

In order to prove the convergence result (4.35), as in the proof of Theorem 3.1 we consider two cases.

Assume first that $\{y^{\delta_i}\}$ is a sequence satisfying $\|y^{\delta_i} - y\| \leq \delta_i$ with $\delta_i \rightarrow 0$ such that $n_{\delta_i} \rightarrow n_0$ as $i \rightarrow \infty$ for some integer n_0 . We may assume $n_{\delta_i} = n_0$ for all i . By Rule 4.1 we always have $\|F(x_{n_0+1}^{\delta_i}) - y^{\delta_i}\| \leq \tau\delta_i$. By letting $i \rightarrow \infty$, we obtain $F(x_{n_0+1}) = y$. This together with the definition of x_n implies that $x_n = x_{n_0+1}$ for all $n \geq n_0 + 1$. It then follows from Theorem 4.1 that $x_* = x_{n_0+1}$. We claim that $x_{n_0+1} = x_{n_0}$. To see this, by using the definition of ξ_{n_0+1} , we have

$$\xi_{n_0+1} = \xi_{n_0} - \frac{1}{\alpha_{n_0+1}} F'(x_{n_0+1})^* J_r(F(x_{n_0+1}) - y) = \xi_{n_0}.$$

Therefore

$$\begin{aligned} D_{\xi_{n_0}} \Theta(x_{n_0+1}, x_{n_0}) &\leq D_{\xi_{n_0}} \Theta(x_{n_0+1}, x_{n_0}) + D_{\xi_{n_0+1}} \Theta(x_{n_0}, x_{n_0+1}) \\ &= \langle \xi_{n_0+1} - \xi_{n_0}, x_{n_0+1} - x_{n_0} \rangle = 0. \end{aligned}$$

This and the strictly convexity of Θ imply that $x_{n_0+1} = x_{n_0}$. Consequently $x_{n_0} = x_*$. A simple application of Lemma 4.3 then gives the desired conclusion.

Assume next that $\{y^{\delta_i}\}$ is a sequence satisfying $\|y^{\delta_i} - y\| \leq \delta_i$ with $\delta_i \rightarrow 0$ such that $n_{\delta_i} \rightarrow \infty$ as $i \rightarrow \infty$. We can follow the argument for deriving (4.30) to show that $D_{\xi_{n_i}^{\delta_i}} \Theta(x_*, x_{n_i}^{\delta_i}) \rightarrow 0$ which in turn implies that $x_{n_i}^{\delta_i} \rightarrow x_*$ by the uniformly convexity of Θ . Then we can use (4.36) and follow the same procedure in the proof of Theorem 3.1 to obtain $\Theta(x_{n_i}^{\delta_i}) \rightarrow \Theta(x_*)$ as $i \rightarrow \infty$. \square

5 Numerical examples

In this section we present some numerical simulations to test the performance of our method by considering a linear integral equation of the first kind and a nonlinear problem arising from the parameter identification in partial differential equations.

Example 5.1 We consider the linear integral equation of the form

$$Ax(s) := \int_0^1 K(s, t)x(t)dt = y(s) \quad \text{on } [0, 1], \quad (5.1)$$

where

$$K(s, t) = \begin{cases} 40s(1-t), & s \leq t, \\ 40t(1-s), & s \geq t. \end{cases}$$

It is clear that $A : \mathcal{X} := L^2[0, 1] \rightarrow \mathcal{Y} := L^2[0, 1]$ is a compact operator. Our goal is to find the solution of (5.1) by using some noisy data y^δ instead of y . We assume that the exact solution is

$$x^\dagger(t) = \begin{cases} 0.5, & t \in [0.292, 0.300], \\ 1, & t \in [0.500, 0.508], \\ 0.7, & t \in [0.700, 0.708], \\ 0, & \text{elsewhere} \end{cases}$$

Let $y = Ax^\dagger$ which is the exact data. For a given noise level $\delta > 0$, we add random Gaussian noise to y to obtain y^δ satisfying $\|y - y^\delta\|_{L^2[0,1]} = \delta$ which is used to reconstruct x^\dagger when the iteration is terminated by the discrepancy principle (3.6).

In our numerical simulations, we take $x_0 = 0$ and $\xi_0 = 0$, we divide $[0, 1]$ into $N = 400$ subintervals of equal length, approximate any integrals by the trapezoidal rule, and solve the involved minimization problems by the modified Fletcher-Reeves CG method in [21]. In Figure 5.1 we present the reconstruction results by taking $\delta = 0.5 \times 10^{-3}$ and $\alpha_n = 2^{-n}$ with $\tau = 1.02$ in the discrepancy principle (3.6). Figure 5.1(a) reports the result via the method with $\Theta(x) = \|x\|_{L^2}^2$. It is clear that the reconstructed solution is rather oscillatory and fails to capture the sparsity of the exact solution x^\dagger . Figure 5.1(b) gives the result of the method with $\Theta(x) = \mu\|x\|_{L^2}^2 + \|x\|_{L^1}$ and $\mu = 0.01$. During the computation, $\|x\|_{L^1}$ is replaced by a smooth one $\int_0^1 \sqrt{|x|^2 + \epsilon}$ with $\epsilon = 10^{-6}$. The sparsity reconstruction is significantly improved.

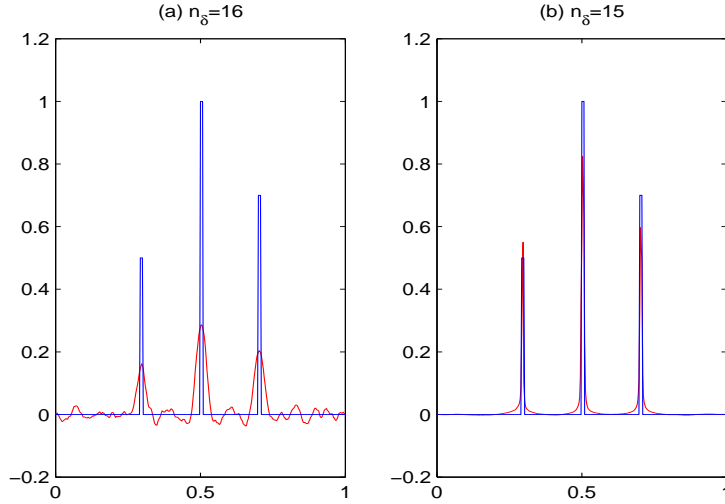


Fig. 5.1 Reconstruction results for Example 5.1: (a) $\Theta(x) = \|x\|_{L^2}^2$; (b) $\Theta(x) = \mu\|x\|_{L^2}^2 + \|x\|_{L^1}$ with $\mu = 0.01$.

Example 5.2 We next consider the identification of the parameter c in the boundary value problem

$$\begin{cases} -\Delta u + cu = f & \text{in } \Omega, \\ u = g & \text{on } \partial\Omega \end{cases} \quad (5.2)$$

from an $L^2(\Omega)$ -measurement of the state u , where $\Omega \subset \mathbb{R}^d$, $d \leq 3$, is a bounded domain with Lipschitz boundary, $f \in L^2(\Omega)$ and $g \in H^{3/2}(\partial\Omega)$. We assume that the sought solution c^\dagger is in $L^2(\Omega)$. This problem reduces to solving an equation of the form (1.1) if we define the nonlinear operator $F : L^2(\Omega) \rightarrow L^2(\Omega)$ by $F(c) := u(c)$, where $u(c) \in H^2(\Omega) \subset L^2(\Omega)$ denotes the unique solution of (5.2). This operator F is well defined on

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

for some positive constant $\gamma_0 > 0$. It is well known that F is Fréchet differentiable; the Fréchet derivative of F and its adjoint are given by

$$F'(c)h = -A(c)^{-1}(hF(c)) \quad \text{and} \quad F'(c)^*w = -u(c)A(c)^{-1}w$$

for $h, w \in L^2(\Omega)$, where $A(c) : H^2 \cap H_0^1 \rightarrow L^2$ is defined by $A(c)u = -\Delta u + cu$ which is an isomorphism uniformly in a ball $B_\rho(c_0) \cap D(F)$ for any $c_0 \in D(F)$ with small $\rho > 0$. It has been shown (see [4]) that for any $\bar{c}, c \in B_\rho(c_0)$ there holds

$$\|F(\bar{c}) - F(c) - F'(c)(\bar{c} - c)\|_{L^2(\Omega)} \leq C\|\bar{c} - c\|_{L^2(\Omega)}\|F(\bar{c}) - F(c)\|_{L^2(\Omega)}.$$

Therefore, Assumption 3.2 and the condition (3.10) hold if $\rho > 0$ is small enough.

In our numerical simulation, we consider the two dimensional problem with $\Omega = [0, 1] \times [0, 1]$ and

$$c^\dagger(x, y) = \begin{cases} 1, & \text{if } (x - 0.3)^2 + (y - 0.7)^2 \leq 0.2^2; \\ 0.5, & \text{if } (x, y) \in [0.6, 0.8] \times [0.2, 0.5]; \\ 0, & \text{elsewhere.} \end{cases}$$

We assume $u(c^\dagger) = x + y$ and add noise to produce the noisy data u^δ satisfying $\|u^\delta - u(c^\dagger)\|_{L^2(\Omega)} = \delta$. We take $\delta = 0.1 \times 10^{-3}$ and $\alpha_n = 2^{-n}$. The partial differential equations involved are solved approximately by a finite difference method by dividing Ω into 40×40 small squares of equal size. and the involved minimization problems are solved by the modified nonlinear CG method in [21]. we take the initial guess $c_0 = 0$ and $\xi_0 = 0$, and terminate the iteration by the discrepancy principle (3.6) with $\tau = 1.05$.

Figure 5.2(a) plots the exact solution $c^\dagger(x, y)$. Figure 5.2(b) shows the result for the method with $\Theta(c) = \|c\|_{L^2}^2$. Figure 5.2 (c) and (d) report the reconstruction results for the method with $\Theta(c) = \mu\|c\|_{L^2}^2 + \int_\Omega |Dc|$ for $\mu = 0.01$ and $\mu = 1.0$ respectively; the term $\int_\Omega |Dc|$ is replaced by a smooth one $\int_\Omega \sqrt{|Dc|^2 + \epsilon}$ with $\epsilon = 10^{-6}$ during computation. The reconstruction results in (c) and (d) significantly improve the one in (b) by efficiently removing the notorious oscillatory effect and indicate that the method is robust with respect to μ . We remark that, due to the smaller value of μ , the reconstruction result in (d) is slightly better than the one in (c) as can be seen from the plots; the computational time for (d), however, is longer.

Acknowledgements Q Jin is partly supported by the grant DE120101707 of Australian Research Council, and M Zhong is partly supported by the National Natural Science Foundation of China (No.11101093).

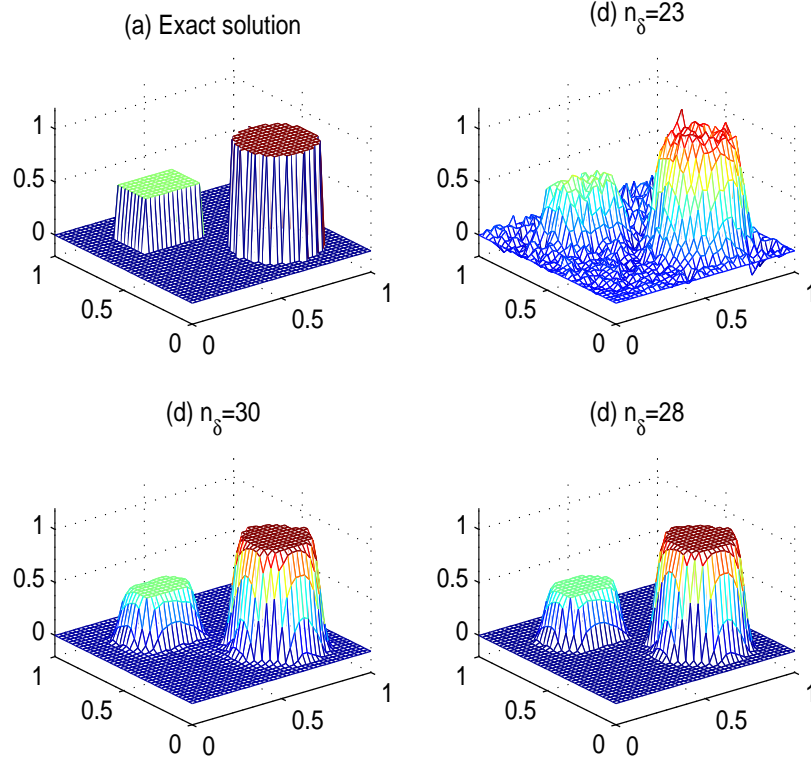


Fig. 5.2 Reconstruction results for Example 5.2: (a) Exact solution; (b) $\Theta(c) = \|c\|_{L^2}^2$; (c) and (d) $\Theta(c) = \mu \|c\|_{L^2}^2 + \int_{\Omega} |Dc|$ with $\mu = 0.01$ and $\mu = 1$ respectively.

References

1. R. A. Adams and J. Fournier, *Sobolev Spaces*, Second edition, Pure and Applied Mathematics (Amsterdam), Vol. 140, Elsevier/Academic Press, Amsterdam, 2003.
2. M. Bachmayr and M. Burger, *Iterative total variation schemes for nonlinear inverse problems*, Inverse Problems, 25 (2009), no. 10, 105004, 26 pp.
3. I. Cioranescu, *Geometry of Banach Spaces, Duality Mappings and Nonlinear Problems*, Dordrecht: Kluwer, 1990.
4. H. W. Engl, M. Hanke and A. Neubauer, *Regularization of Inverse Problems*, Mathematics and its Applications, 375. Kluwer Academic Publishers Group, Dordrecht, 1996.
5. H. W. Engl, K. Kunisch and A. Neubauer, *Convergence rates of Tikhonov regularization of nonlinear ill-posed problems*, Inverse Problems, 5 (1989), 523–540.
6. K. Frick and M. Grasmair, *Regularization of linear ill-posed problems by the augmented Lagrangian method and variational inequalities*, Inverse Problems, 28(2012), 104005.
7. E. Giusti, *Minimal Surfaces and Functions of Bounded Variation*, Basel: Birkhäuser, 1984.
8. M. Hanke and C. W. Groetsch, *Nonstationary iterated Tikhonov regularization*, J. Optim. Theory Appl. 97(1998), 37–53.
9. M. Hanke, A. Neubauer and O. Scherzer, *A convergence analysis of the Landweber iteration for nonlinear ill-posed problems*, Numer. Math., 72 (1995), 21–37.
10. M. R. Hestenes, *Multiplier and gradient methods*, J. Optim. Theory Appl., 4 (1969), 303–320.
11. K. Ito and K. Kunisch, *Lagrange Multiplier Approach to Variational Problems and Applications*, SIAM, Philadelphia, 2008.
12. Q. Jin, *On a regularized Levenberg-Marquardt method for solving nonlinear inverse problems*, Numer. Math., 115 (2010), 229–259.

13. Q. Jin, *A general convergence analysis of some Newton-type methods for nonlinear inverse problems*, SIAM J. Numer. Anal., 49 (2011), 549–573.
14. Q. Jin and L. Stals, *Nonstationary iterated Tikhonov regularization for ill-posed problems in Banach spaces*, Inverse Problems, 28(2012), 104011(15pp).
15. B. Kaltenbacher, A. Neubauer and O. Scherzer, *Iterative Regularization Methods for Nonlinear Ill-Posed Problems*, de Gruyter, Berlin 2008.
16. S. Osher, M. Burger, D. Goldfarb, J. Xu and W. Yin, *An iterative regularization method for total variation-based image restoration*, Multiscale Model. Simul. 4 (2005), no. 2, 460–489.
17. M. J. D. Powell, *A method for nonlinear constraints in minimization problems*, in Optimization ed. by R. Fletcher, Academic Press, New York, NY, 1969, 283–298.
18. L. Rudin, S. Osher, and C. Fatemi, *Nonlinear total variation based noise removal algorithm*, Phys. D, 60 (1992), pp. 259–268.
19. R. Tibshirani, *Regression shrinkage and selection via the lasso*, J. Roy. Statist. Soc. Ser. B, 58 (1996), no. 1, 267–288.
20. C. Zălinescu, *Convex Analysis in General Vector Spaces*, World Scientific Publishing Co., Inc., River Edge, New Jersey, 2002.
21. L. Zhang, W. Zhou and D. Li, *Global convergence of a modified Fletcher-Reeves conjugate gradient method with Armijo-type line search*, Numer. Math., 104 (2006), 561–572.