Numerische Mathematik (2019) 143:713–747
https://doi.org/10.1007/s00211-019-01068-0

Numerische Mathematik

Regularization of inverse problems by two-point gradient methods in Banach spaces

Min Zhong[1] · Wei Wang[2] · Qinian Jin[3]

Received: 27 December 2018 / Revised: 6 May 2019 / Published online: 29 July 2019

Abstract
In this paper, we propose and analyze a two-point gradient method for solving inverse problems in Banach spaces which is based on the Landweber iteration and an extrapolation strategy. The method allows to use non-smooth penalty terms, including the L^1-like and the total variation-like penalty functionals, which are significant in reconstructing special features of solutions such as sparsity and piecewise constancy in practical applications. The design of the method involves the choices of the step sizes and the combination parameters which are carefully discussed. Numerical simulations are presented to illustrate the effectiveness of the proposed method.

Mathematics Subject Classification 65J15 · 65J20 · 47H17

1 Introduction

In this paper we are interested in solving inverse problems of the form

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

where $F : \mathscr{D}(F) \subset \mathcal{X} \to \mathcal{Y}$ is an operator between two Banach spaces $\mathcal{X}$ and $\mathcal{Y}$. Throughout this paper we will assume (1.1) has a solution, which is not necessarily

✉ Wei Wang
weiwang_math@126.com

Min Zhong
min.zhong@seu.edu.cn

Qinian Jin
Qinian.Jin@anu.edu.au

1 School of Mathematics, Southeast University, Nanjing 210096, Jiangsu, People's Republic of China

2 College of Mathematics, Physics and Information Engineering, Jiaxing University, Jiaxing 314001, Zhejiang, People's Republic of China

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

unique. Such inverse problems are ill-posed in the sense of unstable dependence of solutions on small perturbations of the data. Instead of exact data y, in practice we are given only noisy data y^δ satisfying

$$\|y - y^\delta\| \leqslant \delta. \tag{1.2}$$

Consequently, it is necessary to apply regularization methods to solve (1.1) approximately [6].

Landweber iteration is one of the most prominent regularization methods for solving inverse problems formulated in Hilbert spaces. A complete analysis on this method for linear problems as well as nonlinear problems can be found in [6,10]. This method has received tremendous attention due to its simple implementation and robustness with respect to noise.

The classical Landweber iteration in Hilbert spaces, however, has the tendency to over-smooth solutions, which makes it difficult to capture special features of the sought solution such as sparsity and discontinuity. To overcome this drawback, various reformulations of Landweber iteration either in Banach spaces or in a manner of incorporating general non-smooth convex penalty functionals have been proposed, see [4,16,18,23,27,29]. Assuming the Fréchet differentiability of the forward operator F, by applying a gradient method for solving

$$\min \frac{1}{s}\|F(x) - y^\delta\|^s, \tag{1.3}$$

the authors in [23,27] proposed the Landweber iteration of the form

$$\begin{aligned} \xi_{n+1}^\delta &= \xi_n^\delta - \mu_n^\delta F'(x_n^\delta)^* J_s^{\mathcal{Y}}(F(x_n^\delta) - y^\delta), \\ x_{n+1}^\delta &= J_q^{\mathcal{X}^*}(\xi_{n+1}^\delta), \end{aligned}$$

for solving linear as well as nonlinear inverse problems in Banach spaces, assuming suitable smoothness and convexity on $\mathcal{X}$ and $\mathcal{Y}$, where $F'(x)$ and $F'(x)^*$ denote the Fréchet derivative of F at x and its adjoint, μ_n^δ denotes the step size, and $J_s^{\mathcal{Y}} : \mathcal{Y} \to \mathcal{Y}^*$ and $J_q^{\mathcal{X}^*} : \mathcal{X}^* \to \mathcal{X}$ with $1 < s, q < \infty$ denote the duality mappings with gauge functions $t \to t^{s-1}$ and $t \to t^{q-1}$ respectively. This formulation of Landweber iteration, however, exclude the use of the L^1 and the total variation like penalty functionals. A Landweber-type iteration with general uniform convex penalty functionals was introduced in [4] for solving linear inverse problems and was extended in [18] for solving nonlinear inverse problems. Let $\Theta : \mathcal{X} \to (-\infty, \infty]$ be a proper lower semi-continuous uniformly convex functional, the method in [4,18] can be formulated as

$$\begin{aligned} \xi_{n+1}^\delta &= \xi_n^\delta - \mu_n^\delta F'(x_n^\delta)^* J_s^{\mathcal{Y}}(F(x_n^\delta) - y^\delta), \\ x_{n+1}^\delta &= \arg\min_{x \in \mathcal{X}} \left\{\Theta(x) - \langle \xi_{n+1}^\delta, x\rangle\right\}. \end{aligned} \tag{1.4}$$

The advantage of this method is the freedom on the choice of Θ so that it can be utilized in detecting different features of the sought solution.

It is well known that Landweber iteration is a slowly convergent method. As alternatives to Landweber iteration, one may consider the second order iterative methods, such as the Levenberg–Maquardt method [9,19], the iteratively regularized Gauss-Newton method [20,22], or the nonstationary iterated Tikhonov regularization [21]. The advantage of these methods is that they require less number of iterations to satisfy the respective stopping rule than the Landweber iteration, however they always require to spend more computational time in dealing with each iteration step. Therefore, it becomes more desirable to accelerate Landweber iteration by preserving its simple implementation feature.

For linear inverse problems in Hilbert spaces, a family of accelerated Landweber iterations were proposed in [8] using the orthogonal polynomials and the spectral theory of self-adjoint operators. The acceleration strategy using orthogonal polynomials is no longer available for Landweber iteration in Banach spaces with general convex penalty functionals. In [13] an acceleration of Landweber iteration in Banach spaces was considered based on choosing optimal step size in each iteration step. In [12,28] the sequential subspace optimization strategy was employed to accelerate the Landweber iteration.

The Nesterov's strategy was proposed in [26] to accelerate gradient method. It has played an important role on the development of fast first order methods for solving well-posed convex optimization problems [1,2]. Recently, an accelerated version of Landweber iteration based on Nesterov's strategy was proposed in [17] which includes the following method

$$
\begin{aligned}
z_n^\delta &= x_n^\delta + \frac{n}{n+\alpha}(x_n^\delta - x_{n-1}^\delta),\\
x_{n+1}^\delta &= z_n^\delta + \mu_n^\delta F'(z_n^\delta)^*(y^\delta - F(z_n^\delta))
\end{aligned}
\tag{1.5}
$$

with $\alpha \geqslant 3$ as a special case for solving ill-posed inverse problems (1.1) in Hilbert spaces, where $x_{-1}^\delta = x_0^\delta = x_0$ is an initial guess. Although no convergence analysis for (1.5) could be given, the numerical results presented in [17,32] clearly demonstrate its usefulness and acceleration effect. By replacing $n/(n+\alpha)$ in (1.5) by a general connection parameters λ_n^δ, a so called two-point gradient method was proposed in [14] and a convergence result was proved under a suitable choice of $\{\lambda_n^\delta\}$. Furthermore, based on the assumption of local convexity of the residual functional around the sought solution, a weak convergence result on (1.5) was proved in [15] recently.

In this paper, by incorporating an extrapolation step into the iteration scheme (1.4), we propose a two-point gradient method for solving inverse problems in Banach spaces with a general uniformly convex penalty term Θ, which takes the form

$$
\begin{aligned}
\zeta_n^\delta &= \xi_n^\delta + \lambda_n^\delta(\xi_n^\delta - \xi_{n-1}^\delta),\\
z_n^\delta &= \arg\min_{z\in\mathcal{X}} \left\{\Theta(z) - \langle \zeta_n^\delta, z\rangle\right\},\\
\xi_{n+1}^\delta &= \zeta_n^\delta - \mu_n^\delta F'(z_n^\delta)^* J_s^{\mathcal{Y}}(F(z_n^\delta) - y^\delta)
\end{aligned}
$$

with suitable step sizes μ_n^δ and combination parameters λ_n^δ; after terminated by a discrepancy principle, we then use

$$x_n^\delta := \arg\min_{x\in\mathcal{X}} \left\{\Theta(x) - \langle \xi_n^\delta, x\rangle\right\}$$

as an approximate solution. We note that when $\lambda_n^\delta = 0$, our method becomes the Landweber iteration of the form (1.4) and when $\lambda_n^\delta = n/(n+\alpha)$ it becomes a refined version of the Nesterov acceleration of Landweber iteration proposed in [17]. We note also that, when both $\mathcal{X}$ and $\mathcal{Y}$ are Hilbert spaces and $\Theta(x) = \|x\|^2/2$, our method becomes the two-point gradient methods introduced in [14] for solving inverse problems in Hilbert spaces. Unlike the method in [14], our method not only works for inverse problems in Banach spaces, but also allows the use of general convex penalty functions including the L^1 and the total variation like functions. Due to the possible non-smoothness of Θ and the non-Hilbertian structures of $\mathcal{X}$ and $\mathcal{Y}$, we need to use tools from convex analysis and subdifferential calculus to carry out the convergence analysis. Under certain conditions on the combination parameters $\{\lambda_n^\delta\}$, we obtain a convergence result on our method. In order to find nontrivial λ_n^δ, we adapt the discrete backtracking search (DBTS) algorithm in [14] to our situation and provide a complete convergence analysis of the corresponding method. Our analysis in fact improves the convergence result in [14] by removing the technical conditions on $\{\lambda_n^\delta\}$ chosen by the DBTS algorithm.

The paper is organized as follows, In Sect. 2, we give some preliminaries from convex analysis. In Sect. 3, we formulate our two-point gradient method with a general uniformly convex penalty term and present the detailed convergence analysis. We also discuss the choices of the combination parameters by a discrete backtracking search algorithm. Finally in Sect. 4, numerical simulations are given to test the performance of the method.

2 Preliminaries

In this section, we introduce some necessary concepts and properties related to Banach space and convex analysis, we refer to [31] for more details.

Let $\mathcal{X}$ be a Banach space whose norm is denoted by $\|\cdot\|$, 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. For a bounded linear operator $A : \mathcal{X} \to \mathcal{Y}$ between two Banach spaces $\mathcal{X}$ and $\mathcal{Y}$, we use $\mathcal{N}(A)$ and $A^* : \mathcal{Y}^* \to \mathcal{X}^*$ to denote its null space and its adjoint respectively. We also use

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

to denote the annihilator of $\mathcal{N}(A)$.

Given a convex function $\Theta : \mathcal{X} \to (-\infty, \infty]$, we use $\partial\Theta(x)$ to denote the subdifferential of Θ at $x \in \mathcal{X}$, i.e.

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

Let $\mathscr{D}(\Theta) := \{x \in \mathcal{X} : \Theta(x) < \infty\}$ be its effective domain and let

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

The Bregman distance induced by Θ at x in the direction $\xi \in \partial\Theta(x)$ is defined by

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

which is always nonnegative and satisfies the identity

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

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

A proper convex function $\Theta : \mathcal{X} \to (-\infty, \infty]$ is called uniformly convex if there exists a strictly increasing function $h : [0, \infty) \to [0, \infty)$ with $h(0) = 0$ such that

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

for all $\bar{x}, x \in \mathcal{X}$ and $\lambda \in [0, 1]$. If $h(t) = c_0 t^p$ for some $c_0 > 0$ and $p > 1$ in (2.2), then Θ is called p-convex. It can be shown that Θ is p-convex if and only if

$$D_\xi\Theta(\bar{x}, x) \geqslant c_0\|x - \bar{x}\|^p, \quad \forall \bar{x} \in \mathcal{X},\ x \in \mathscr{D}(\partial\Theta),\ \xi \in \partial\Theta(x). \tag{2.3}$$

For a proper lower semi-continuous convex function $\Theta : \mathcal{X} \to (-\infty, \infty]$, its Legendre–Fenchel conjugate is defined by

$$\Theta^*(\xi) := \sup_{x \in \mathcal{X}} \{\langle \xi, x \rangle - \Theta(x)\}, \quad \xi \in \mathcal{X}^*$$

which is also proper, lower semi-continuous, and convex. If $\mathcal{X}$ is reflexive, then

$$\xi \in \partial\Theta(x) \Longleftrightarrow x \in \partial\Theta^*(\xi) \Longleftrightarrow \Theta(x) + \Theta^*(\xi) = \langle \xi, x \rangle. \tag{2.4}$$

Moreover, if Θ is p-convex with $p > 1$ then it follows from [31, Corollary 3.5.11] that $\mathscr{D}(\Theta^*) = \mathcal{X}^*$, Θ^* is Fréchet differentiable and its gradient $\nabla\Theta^* : \mathcal{X}^* \to \mathcal{X}$ satisfies

$$\|\nabla\Theta^*(\xi_1) - \nabla\Theta^*(\xi_2)\| \leqslant \left(\frac{\|\xi_1 - \xi_2\|}{2c_0}\right)^{\frac{1}{p-1}}, \quad \forall \xi_1, \xi_2 \in \mathcal{X}^*. \tag{2.5}$$

Consequently, it follows from (2.4) that

$$x = \nabla\Theta^*(\xi) \Longleftrightarrow \xi \in \partial\Theta(x) \Longleftrightarrow x = \arg\min_{z \in \mathcal{X}} \{\Theta(z) - \langle \xi, z \rangle\}. \tag{2.6}$$

Lemma 2.1 *If Θ is p-convex with $p > 1$, then for any pairs (x, ξ) and $(\bar{x}, \bar{\xi})$ with $x, \bar{x} \in \mathscr{D}(\partial\Theta)$, $\xi \in \partial\Theta(x)$, $\bar{\xi} \in \partial\Theta(\bar{x})$, we have*

$$D_\xi \Theta(\bar{x}, x) \leqslant \frac{1}{p^*(2c_0)^{p^*-1}} \|\xi - \bar{\xi}\|^{p^*}, \tag{2.7}$$

where $p^* := p/(p-1)$ is the number conjugate to p.

Proof Applying (2.4), $\bar{x} = \nabla\Theta^*(\bar{\xi})$ and (2.5), it follows that

$$\begin{aligned} D_\xi \Theta(\bar{x}, x) &= \Theta^*(\xi) - \Theta^*(\bar{\xi}) - \langle \xi - \bar{\xi}, \nabla\Theta^*(\bar{\xi})\rangle \\ &= \int_0^1 \langle \xi - \bar{\xi}, \nabla\Theta^*(\bar{\xi} + t(\xi - \bar{\xi})) - \nabla\Theta^*(\bar{\xi})\rangle dt \\ &\leqslant \|\xi - \bar{\xi}\| \int_0^1 \|\nabla\Theta^*(\bar{\xi} + t(\xi - \bar{\xi})) - \nabla\Theta^*(\bar{\xi})\| dt \\ &\leqslant \frac{1}{p^*(2c_0)^{p^*-1}} \|\xi - \bar{\xi}\|^{p^*} \end{aligned}$$

which shows the result. □

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

$$J_s^{\mathcal{X}}(x) := \left\{\xi \in \mathcal{X}^* : \|\xi\| = \|x\|^{s-1} \text{ and } \langle \xi, x\rangle = \|x\|^s\right\} \tag{2.8}$$

which gives the duality mapping $J_s^{\mathcal{X}} : \mathcal{X} \to 2^{\mathcal{X}^*}$ of $\mathcal{X}$ with gauge function $t \to t^{s-1}$. If $\mathcal{X}$ is uniformly smooth in the sense that its modulus of smoothness

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

satisfies $\lim_{t \searrow 0} \frac{\rho_{\mathcal{X}}(t)}{t} = 0$, then the duality mapping $J_s^{\mathcal{X}}$, for each $1 < s < \infty$, is single valued and uniformly continuous on bounded sets. There are many examples of uniformly smooth Banach spaces, e.g., sequence space ℓ^s, Lebesgue space L^s, Sobolev space $W^{k,s}$ and Besov space $B^{q,s}$ with $1 < s < \infty$.

3 The two-point gradient method

We consider

$$F(x) = y, \tag{3.1}$$

where $F : \mathscr{D}(F) \subset \mathcal{X} \to \mathcal{Y}$ is an operator between two Banach spaces $\mathcal{X}$ and $\mathcal{Y}$. Throughout this paper, we will always assume that $\mathcal{X}$ is reflexive, $\mathcal{Y}$ is uniformly smooth, and (3.1) has a solutions. In order to capture the special feature of the sought solution, we will use a general convex function $\Theta : \mathcal{X} \to (-\infty, \infty]$ as a penalty term. We will need a few assumptions concerning Θ and F.

Assumption 1 $\Theta : \mathcal{X} \to (-\infty, \infty]$ is proper, weak lower semi-continuous, and p-convex with $p > 1$ such that the condition (2.3) is satisfied for some $c_0 > 0$.

Assumption 2 (a) There exists $\rho > 0$, $x_0 \in \mathcal{X}$ and $\xi_0 \in \partial\Theta(x_0)$ such that $B_{3\rho}(x_0) \subset \mathscr{D}(F)$ and (3.1) has a solution $x^* \in \mathscr{D}(\Theta)$ with

$$D_{\xi_0}\Theta(x^*, x_0) \leqslant c_0\rho^p,$$

where $B_\rho(x_0)$ denotes the closed ball around x_0 with radius ρ.

(b) The operator F is weakly closed on $\mathscr{D}(F)$.

(c) There exists a family of bounded linear operators $\{L(x) : \mathcal{X} \to \mathcal{Y}\}_{x \in B_{3\rho}(x_0)\cap\mathscr{D}(\Theta)}$ such that $x \to L(x)$ is continuous on $B_{3\rho}(x_0) \cap \mathscr{D}(\Theta)$ and there is $0 \leqslant \eta < 1$ such that

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

for all $x, \bar{x} \in B_{3\rho}(x_0) \cap \mathscr{D}(\Theta)$. Moreover, there is a constant $C_0 > 0$ such that

$$\|L(x)\|_{\mathcal{X}\to\mathcal{Y}} \leqslant C_0, \quad \forall x \in B_{3\rho}(x_0).$$

All the conditions in Assumption 3.2 are standard. The condition (c) is called the tangential condition and is widely used in the analysis of iterative regularization methods for nonlinear ill-posed inverse problems [10]. The weak closedness of F in condition (b) means that for any sequence $\{x_n\}$ in $\mathscr{D}(F)$ satisfying $x_n \rightharpoonup x$ and $F(x_n) \to v$, then $x \in \mathscr{D}(F)$ and $F(x) = v$, where we use "$\rightharpoonup$" to denote the weak convergence.

Remark 3.1 The condition $B_{3\rho}(x_0) \subset \mathscr{D}(F)$ in Assumption 2(a) can be replaced by $\mathscr{D}(\Theta) \subset \mathscr{D}(F)$ which is automatically satisfied by replacing Θ by $\Theta + \iota_{\mathscr{D}(F)}$ in case $\mathscr{D}(F)$ is closed and convex, where $\iota_{\mathscr{D}(F)}$ denotes the indicator function of $\mathscr{D}(F)$, i.e. $\iota_{\mathscr{D}(F)}(x) = 0$ for $x \in \mathscr{D}(F)$ and $\iota_{\mathscr{D}(F)}(x) = +\infty$ otherwise. The corresponding convergence analysis can be performed by the same argument in the paper without any change.

Using the p-convex function Θ specified in Assumption 1, we may pick among solutions of (3.1) the one with the desired feature. We define $x^\dagger$ to be a solution of (3.1) with the property

$$D_{\xi_0}\Theta(x^\dagger, x_0) = \min_{x\in\mathscr{D}(\Theta)\cap\mathscr{D}(F)} \left\{D_{\xi_0}\Theta(x, x_0) : F(x) = y\right\}. \tag{3.2}$$

When $\mathcal{X}$ is reflexive, by using the weak closedness of F, the weak lower semi-continuity of Θ and the p-convexity of Θ, it is standard to show that $x^\dagger$ exists. According to Assumption 2(a), we always have

$$D_{\xi_0}\Theta(x^\dagger, x_0) \leqslant c_0\rho^p$$

which together with Assumption 1 implies that $\|x^\dagger - x_0\| \leqslant \rho$. The following lemma shows that $x^\dagger$ is uniquely defined.

Lemma 3.1 *Under Assumptions 1 and 2, the solution $x^\dagger$ of (3.1) satisfying (3.2) is uniquely defined.*

Proof This is essentially proved in [18, Lemma 3.2]. □

In order to construct an approximate solution to (3.1), we will formulate a two-point gradient method with penalty term induced by the p-convex function Θ. Let $\tau > 1$ be a given number. By picking $x_{-1}^\delta = x_0^\delta := x_0 \in \mathcal{X}$ and $\xi_{-1}^\delta = \xi_0^\delta := \xi_0 \in \partial\Theta(x_0)$ as initial guess, for $n \geqslant 0$ we define

$$\begin{aligned} \zeta_n^\delta &= \xi_n^\delta + \lambda_n^\delta(\xi_n^\delta - \xi_{n-1}^\delta), \\ z_n^\delta &= \nabla\Theta^*(\zeta_n^\delta), \\ \xi_{n+1}^\delta &= \zeta_n^\delta - \mu_n^\delta L(z_n^\delta)^* J_s^{\mathcal{Y}}(r_n^\delta), \\ x_{n+1}^\delta &= \nabla\Theta^*(\xi_{n+1}^\delta), \end{aligned} \tag{3.3}$$

where $r_n^\delta = F(z_n^\delta) - y^\delta$, $\lambda_n^\delta \geqslant 0$ is the combination parameter, μ_n^δ is the step sizes defined by

$$\mu_n^\delta = \begin{cases} \min\left\{ \dfrac{\bar{\mu}_0 \|r_n^\delta\|^{p(s-1)}}{\|L(z_n^\delta)^* J_s^{\mathcal{Y}}(r_n^\delta)\|^p}, \bar{\mu}_1 \right\} \|r_n^\delta\|^{p-s} & \text{if } \|r_n^\delta\| > \tau\delta \\ 0 & \text{if } \|r_n^\delta\| \leqslant \tau\delta \end{cases} \tag{3.4}$$

for some positive constants $\bar{\mu}_0$ and $\bar{\mu}_1$, and the mapping $J_s^{\mathcal{Y}} : \mathcal{Y} \to \mathcal{Y}^*$ with $1 < s < \infty$ denotes the duality mapping of $\mathcal{Y}$ with gauge function $t \to t^{s-1}$, which is single-valued and continuous because $\mathcal{Y}$ is assumed to be uniformly smooth. We remark that when $\lambda_n^\delta = 0$ for all n, the method (3.3) reduces to the Landweber iteration considered in [18] where a detailed convergence analysis has been carried out. When $\lambda_n^\delta = n/(n+\alpha)$ with $\alpha \geqslant 3$ for all n, the method (3.3) becomes a refined version of the Nesterov acceleration of Landweber iteration proposed in [17]; although there is no convergence theory available, numerical simulations in [17] demonstrate its usefulness and acceleration effect. In this paper we will consider (3.3) with λ_n^δ satisfying suitable conditions to be specified later. Note that our method (3.3) requires the use of the previous two iterations at every iteration step; on the other hand, our method allows the use of a general p-convex penalty function Θ, which could be non-smooth, to reconstruct solutions with special features such as sparsity and discontinuities.

3.1 Convergence

In order to use our two point gradient method (3.3) to produce a useful approximate solution to (3.1), the iteration must be terminated properly. We will use the discrepancy principle with respect to z_n^δ, i.e., for a given $\tau > 1$, we will terminate the iteration after n_δ steps, where $n_\delta := n(\delta, y^\delta)$ is the integer such that

$$\|F(z_{n_\delta}^\delta) - y^\delta\| \leqslant \tau\delta < \|F(z_n^\delta) - y^\delta\|, \quad 0 \leqslant n < n_\delta, \tag{3.5}$$

and use $x_{n_\delta}^\delta$ as an approximate solution. To carry out the convergence analysis of $x_{n_\delta}^\delta$ as $\delta \to 0$, we are going to show that, for any solution $\hat{x}$ of (3.1) in $B_{2\rho}(x_0) \cap \mathscr{D}(\Theta)$, the Bregman distance $D_{\xi_n^\delta}\Theta(\hat{x}, x_n^\delta)$, $0 \leqslant n \leqslant n_\delta$, is monotonically decreasing with respect to n. To this end, we introduce

$$\Delta_n := D_{\xi_n^\delta}\Theta(\hat{x}, x_n^\delta) - D_{\xi_{n-1}^\delta}\Theta(\hat{x}, x_{n-1}^\delta). \tag{3.6}$$

We will show that, under suitable choice of $\{\lambda_n^\delta\}$, there holds $\Delta_n \leqslant 0$ for $0 \leqslant n \leqslant n_\delta$.

Lemma 3.2 *Let $\mathcal{X}$ be reflexive, let $\mathcal{Y}$ be uniformly smooth, and let Assumptions 1 and 2 hold. Then, for any solution $\hat{x}$ of (3.1) in $B_{2\rho}(x_0) \cap \mathscr{D}(\Theta)$, there holds*

$$\begin{aligned} &D_{\zeta_n^\delta}\Theta(\hat{x}, z_n^\delta) - D_{\xi_n^\delta}\Theta(\hat{x}, x_n^\delta) \\ &\qquad \leqslant \lambda_n^\delta \Delta_n + \frac{1}{p^*(2c_0)^{p^*-1}}\left(\lambda_n^\delta + \left(\lambda_n^\delta\right)^{p^*}\right)\|\xi_n^\delta - \xi_{n-1}^\delta\|^{p^*}. \end{aligned} \tag{3.7}$$

If $z_n^\delta \in B_{3\rho}(x_0)$ then

$$\begin{aligned} D_{\xi_{n+1}^\delta}\Theta(\hat{x}, x_{n+1}^\delta) - D_{\zeta_n^\delta}\Theta(\hat{x}, z_n^\delta) \leqslant &-\left(1 - \eta - \frac{1}{p^*}\left(\frac{\bar{\mu}_0}{2c_0}\right)^{p^*-1}\right)\mu_n^\delta\|F(z_n^\delta) - y^\delta\|^s \\ &+ (1+\eta)\mu_n^\delta\|F(z_n^\delta) - y^\delta\|^{s-1}\delta. \end{aligned} \tag{3.8}$$

Proof By using the identity (2.1), Lemma 2.1 and the definition of ζ_n^δ, we can obtain

$$\begin{aligned} D_{\zeta_n^\delta}\Theta(\hat{x}, z_n^\delta) - D_{\xi_n^\delta}\Theta(\hat{x}, x_n^\delta) &= D_{\zeta_n^\delta}\Theta(x_n^\delta, z_n^\delta) + \langle \zeta_n^\delta - \xi_n^\delta, x_n^\delta - \hat{x}\rangle \\ &\leqslant \frac{1}{p^*(2c_0)^{p^*-1}}\|\zeta_n^\delta - \xi_n^\delta\|^{p^*} + \langle \zeta_n^\delta - \xi_n^\delta, x_n^\delta - \hat{x}\rangle \\ &= \frac{1}{p^*(2c_0)^{p^*-1}}(\lambda_n^\delta)^{p^*}\|\xi_{n-1}^\delta - \xi_n^\delta\|^{p^*} + \langle \zeta_n^\delta - \xi_n^\delta, x_n^\delta - \hat{x}\rangle. \end{aligned}$$

By using again the definition of ζ_n^δ, (2.1) and Lemma 2.1, and referring to the definition of Δ_n and $\lambda_n^\delta \geqslant 0$, we have

$$\begin{aligned} \langle \zeta_n^\delta - \xi_n^\delta, x_n^\delta - \hat{x}\rangle &= \lambda_n^\delta\langle \xi_n^\delta - \xi_{n-1}^\delta, x_n^\delta - \hat{x}\rangle \\ &= \lambda_n^\delta\left(D_{\xi_n^\delta}\Theta(\hat{x}, x_n^\delta) - D_{\xi_{n-1}^\delta}\Theta(\hat{x}, x_{n-1}^\delta) + D_{\xi_{n-1}^\delta}\Theta(x_n^\delta, x_{n-1}^\delta)\right) \\ &\leqslant \lambda_n^\delta\Delta_n + \frac{1}{p^*(2c_0)^{p^*-1}}\lambda_n^\delta\|\xi_n^\delta - \xi_{n-1}^\delta\|^{p^*}. \end{aligned}$$

The combination of the above two estimates yields (3.7).

To derive (3.8), we first use the identity (2.1) and Lemma 2.1 to obtain

$$\begin{aligned}
&D_{\xi_{n+1}^\delta}\Theta(\hat{x}, x_{n+1}^\delta) - D_{\zeta_n^\delta}\Theta(\hat{x}, z_n^\delta) \\
&\quad = D_{\xi_{n+1}^\delta}\Theta(z_n^\delta, x_{n+1}^\delta) + \langle \xi_{n+1}^\delta - \zeta_n^\delta, z_n^\delta - \hat{x}\rangle \\
&\quad \leqslant \frac{1}{p^*(2c_0)^{p^*-1}}\|\xi_{n+1}^\delta - \zeta_n^\delta\|^{p^*} + \langle \xi_{n+1}^\delta - \zeta_n^\delta, z_n^\delta - \hat{x}\rangle. \qquad (3.9)
\end{aligned}$$

Recall the definition of ξ_{n+1}^δ in (3.3), we have

$$\|\xi_{n+1}^\delta - \zeta_n^\delta\|^{p^*} = (\mu_n^\delta)^{p^*}\|L(z_n^\delta)^* J_s^{\mathcal{Y}}(F(z_n^\delta) - y^\delta)\|^{p^*}.$$

According to the definition (3.4) of μ_n^δ, one can see that

$$\mu_n^\delta \leqslant \frac{\bar{\mu}_0\|F(z_n^\delta) - y^\delta\|^{s(p-1)}}{\|L(z_n^\delta)^* J_s^{\mathcal{Y}}(F(z_n^\delta) - y^\delta)\|^p},$$

which implies that

$$\begin{aligned}
(\mu_n^\delta)^{p^*-1}\|L(z_n^\delta)^* J_s^{\mathcal{Y}}(F(z_n^\delta) - y^\delta)\|^{p^*} &\leqslant \bar{\mu}_0^{p^*-1}\|F(z_n^\delta) - y^\delta\|^{s(p-1)(p^*-1)} \\
&= \bar{\mu}_0^{p^*-1}\|F(z_n^\delta) - y^\delta\|^s.
\end{aligned}$$

Therefore, the first term on the right hand side of (3.9) can be estimated as

$$\frac{1}{p^*(2c_0)^{p^*-1}}\|\xi_{n+1}^\delta - \zeta_n^\delta\|^{p^*} \leqslant \frac{1}{p^*}\left(\frac{\bar{\mu}_0}{2c_0}\right)^{p^*-1}\mu_n^\delta\|F(z_n^\delta) - y^\delta\|^s. \qquad (3.10)$$

For the second term on the right hand side of (3.9), we may use the definition of ξ_{n+1}^δ, the property of $J_s^{\mathcal{Y}}$, and the definition of μ_n^δ to derive that

$$\begin{aligned}
\langle \xi_{n+1}^\delta - \zeta_n^\delta, z_n^\delta - \hat{x}\rangle &= -\mu_n^\delta\langle J_s^{\mathcal{Y}}(F(z_n^\delta) - y^\delta), L(z_n^\delta)(z_n^\delta - \hat{x})\rangle \\
&= -\mu_n^\delta\langle J_s^{\mathcal{Y}}(F(z_n^\delta) - y^\delta), y^\delta - F(z_n^\delta) - L(z_n^\delta)(\hat{x} - z_n^\delta)\rangle \\
&\quad - \mu_n^\delta\|F(z_n^\delta) - y^\delta\|^s \\
&\leqslant \mu_n^\delta\|F(z_n^\delta) - y^\delta\|^{s-1}\left(\delta + \|y - F(z_n^\delta) - L(z_n^\delta)(\hat{x} - z_n^\delta)\|\right) \\
&\quad - \mu_n^\delta\|F(z_n^\delta) - y^\delta\|^s.
\end{aligned}$$

Recall that $z_n^\delta \in B_{3\rho}(x_0)$, we may use Assumption 2(c) to further obtain

$$\begin{aligned}
&\langle \xi_{n+1}^\delta - \zeta_n^\delta, z_n^\delta - \hat{x}\rangle \\
&\quad \leqslant \mu_n^\delta\|F(z_n^\delta) - y^\delta\|^{s-1}\left(\delta + \eta\|F(z_n^\delta) - y\|\right) - \mu_n^\delta\|F(z_n^\delta) - y^\delta\|^s \\
&\quad \leqslant (1+\eta)\mu_n^\delta\|F(z_n^\delta) - y^\delta\|^{s-1}\delta - (1-\eta)\mu_n^\delta\|F(z_n^\delta) - y^\delta\|^s. \qquad (3.11)
\end{aligned}$$

The combination of above two estimates (3.10) and (3.11) with (3.9) yields (3.8). □

Lemma 3.3 *Let $\mathcal{X}$ be reflexive, let $\mathcal{Y}$ be uniformly smooth, and let Assumptions 1 and 2 hold. Assume that $\tau > 1$ and $\bar{\mu}_0 > 0$ are chosen such that*

$$c_1 := 1 - \eta - \frac{1+\eta}{\tau} - \frac{1}{p^*}\left(\frac{\bar{\mu}_0}{2c_0}\right)^{p^*-1} > 0. \tag{3.12}$$

If $z_n^\delta \in B_{3\rho}(x_0)$, then, for any solution $\hat{x}$ of (3.1) in $B_{2\rho}(x_0) \cap \mathscr{D}(\Theta)$, there holds

$$\begin{aligned}\Delta_{n+1} \leqslant \lambda_n^\delta \Delta_n + \frac{1}{p^*(2c_0)^{p^*-1}}\left(\lambda_n^\delta + (\lambda_n^\delta)^{p^*}\right)\|\xi_n^\delta - \xi_{n-1}^\delta\|^{p^*} \\ - c_1\mu_n^\delta\|F(z_n^\delta) - y^\delta\|^s,\end{aligned} \tag{3.13}$$

where Δ_n is defined by (3.6).

Proof By using the definition of μ_n^δ it is easily seen that $\mu_n^\delta\delta \leqslant \mu_n^\delta\|F(z_n^\delta) - y^\delta\|/\tau$. It then follows from (3.8) that

$$D_{\xi_{n+1}^\delta}\Theta(\hat{x}, x_{n+1}^\delta) - D_{\zeta_n^\delta}\Theta(\hat{x}, z_n^\delta) \leqslant -c_1\mu_n^\delta\|F(z_n^\delta) - y^\delta\|^s.$$

Combining this estimate with (3.7) yields (3.13). □

We will use Lemmas 3.2 and 3.3 to show that $z_n^\delta \in B_{3\rho}(x_0)$ and $\Delta_n \leqslant 0$ for all $n \geqslant 0$ and that the integer n_δ determined by (3.5) is finite. To this end, we need to place conditions on $\{\lambda_n^\delta\}$. We assume that $\{\lambda_n^\delta\}$ is chosen such that

$$\frac{1}{p^*(2c_0)^{p^*-1}}\left((\lambda_n^\delta)^{p^*} + \lambda_n^\delta\right)\|\xi_n^\delta - \xi_{n-1}^\delta\|^{p^*} \leqslant c_0\rho^p \tag{3.14}$$

and

$$\frac{1}{p^*(2c_0)^{p^*-1}}\left((\lambda_n^\delta)^{p^*} + \lambda_n^\delta\right)\|\xi_n^\delta - \xi_{n-1}^\delta\|^{p^*} - \frac{c_1}{\nu}\mu_n^\delta\|F(z_n^\delta) - y^\delta\|^s \leqslant 0 \tag{3.15}$$

for all $n \geqslant 0$, where $\nu > 1$ is a constant independent of δ and n. We will discuss how to choose $\{\lambda_n^\delta\}$ to satisfy (3.14) and (3.15) shortly.

Proposition 3.4 *Let $\mathcal{X}$ be reflexive, let $\mathcal{Y}$ be uniformly smooth, and let Assumptions 1 and 2 hold. Let $\tau > 1$ and $\bar{\mu}_0 > 0$ be chosen such that (3.12) holds. If $\{\lambda_n^\delta\}$ is chosen such that (3.14) and (3.15) hold, then*

$$z_n^\delta \in B_{3\rho}(x_0) \quad \textit{and} \quad x_n^\delta \in B_{2\rho}(x_0) \quad \textit{for } n \geqslant 0. \tag{3.16}$$

Moreover, for any solution $\hat{x}$ of (3.1) in $B_{2\rho}(x_0) \cap \mathscr{D}(\Theta)$ there hold

$$D_{\xi_n^\delta}\Theta(\hat{x}, x_n^\delta) \leqslant D_{\xi_{n-1}^\delta}\Theta(\hat{x}, x_{n-1}^\delta) \tag{3.17}$$

and

$$\sum_{m=0}^{n} \mu_m^\delta \|F(z_m^\delta) - y^\delta\|^s \leqslant \frac{\nu}{(\nu - 1)\, c_1} D_{\xi_0} \Theta(\hat{x}, x_0) \tag{3.18}$$

for all $n \geqslant 0$. Let n_δ be chosen by the discrepancy principle (3.5), then n_δ must be a finite integer.

Proof We will show (3.16) and (3.17) by induction. Since $x_{-1}^\delta = x_0^\delta = x_0$, $\xi_{-1}^\delta = \xi_0^\delta = \xi_0$, and $z_0^\delta = \nabla\Theta^*(\zeta_0^\delta) = \nabla\Theta^*(\xi_0) = x_0$, they are trivial for $n = 0$. Now we assume that (3.16) and (3.17) hold for all $0 \leqslant n \leqslant m$ for some integer $m \geqslant 0$, we will show that they are also true for $n = m+1$. By the induction hypotheses $z_m^\delta \in B_{3\rho}(x_0)$, we may use Lemma 3.3 and (3.15) to derive that

$$\Delta_{m+1} \leqslant \lambda_m^\delta \Delta_m - \left(1 - \frac{1}{\nu}\right) c_1 \mu_m^\delta \|F(z_m^\delta) - y^\delta\|^s.$$

Since $\lambda_m^\delta \geqslant 0$ and $\nu > 1$, this together with the induction hypothesis $\Delta_m \leqslant 0$ implies that

$$\Delta_{m+1} \leqslant -\left(1 - \frac{1}{\nu}\right) c_1 \mu_m^\delta \|F(z_m^\delta) - y^\delta\|^s \leqslant 0 \tag{3.19}$$

which shows (3.17) for $n = m + 1$. Consequently, by taking $\hat{x} = x^\dagger$ and using Assumption 2(a), we have

$$D_{\xi_{m+1}^\delta} \Theta(x^\dagger, x_{m+1}^\delta) \leqslant D_{\xi_m^\delta} \Theta(x^\dagger, x_m^\delta) \leqslant \cdots \leqslant D_{\xi_0} \Theta(x^\dagger, x_0) \leqslant c_0 \rho^p.$$

By virtue of Assumption 1, we then have $c_0 \|x_{m+1}^\delta - x^\dagger\|^p \leqslant c_0 \rho^p$ which together with $x^\dagger \in B_\rho(x_0)$ implies that $x_{m+1}^\delta \in B_{2\rho}(x_0)$. Now we may use (3.7) in Lemma 3.2, (3.14) and $\Delta_{m+1} \leqslant 0$ to derive that

$$\begin{aligned} D_{\zeta_{m+1}^\delta} \Theta(x^\dagger, z_{m+1}^\delta) &\leqslant D_{\xi_{m+1}^\delta} \Theta(x^\dagger, x_{m+1}^\delta) + \lambda_{m+1}^\delta \Delta_{m+1} + c_0 \rho^p \\ &\leqslant D_{\xi_{m+1}^\delta} \Theta(x^\dagger, x_{m+1}^\delta) + c_0 \rho^p \\ &\leqslant D_{\xi_0} \Theta(x^\dagger, x_0) + c_0 \rho^p \\ &\leqslant 2 c_0 \rho^p. \end{aligned}$$

This together with Assumption 1 yields $\|x^\dagger - z_{m+1}^\delta\| \leqslant 2^{1/p} \rho \leqslant 2\rho$, and consequently $z_{m+1}^\delta \in B_{3\rho}(x_0)$. We therefore complete the proof of (3.16) and (3.17).

Since (3.16) and (3.17) are valid, the inequality (3.19) holds for all $m \geqslant 0$. Thus

$$\left(1 - \frac{1}{\nu}\right) c_1 \mu_m^\delta \|F(z_m^\delta) - y^\delta\|^s \leqslant D_{\xi_m^\delta} \Theta(\hat{x}, x_m^\delta) - D_{\xi_{m+1}^\delta} \Theta(\hat{x}, x_{m+1}^\delta) \tag{3.20}$$

for $m \geqslant 0$. Hence, for any integer $n \geqslant 0$ we have

$$\left(1-\frac{1}{\nu}\right) c_1 \sum_{m=0}^{n} \mu_m^\delta \|y^\delta - F(z_m^\delta)\|^s \leqslant D_{\xi_0}\Theta(\hat{x}, x_0) - D_{\xi_{n+1}^\delta}\Theta(\hat{x}, x_{n+1}^\delta) \leqslant D_{\xi_0}\Theta(\hat{x}, x_0) \tag{3.21}$$

which shows (3.18).

If n_δ is not finite, then $\|F(z_m^\delta) - y^\delta\| > \tau\delta$ for all integers m and consequently, by using $\|L(x)\| \leqslant C_0$ from Assumption 2(c) and the property of $J_s^{\mathcal{Y}}$, we have

$$\begin{aligned}\mu_m^\delta &= \min\left\{\frac{\bar{\mu}_0\|F(z_m^\delta) - y^\delta\|^{p(s-1)}}{\|L(z_m^\delta)^* J_s^{\mathcal{Y}}(F(z_m^\delta) - y^\delta)\|^p}, \bar{\mu}_1\right\}\|F(z_m^\delta) - y^\delta\|^{p-s} \\ &\geqslant \min\left\{\frac{\bar{\mu}_0}{C_0^p}, \bar{\mu}_1\right\}\|F(z_m^\delta) - y^\delta\|^{p-s}. \end{aligned} \tag{3.22}$$

Therefore, it follows from (3.18) that

$$\begin{aligned}\frac{\nu}{(\nu-1)c_1} D_{\xi_0}\Theta(\hat{x}, x_0) &\geqslant \min\left\{\frac{\bar{\mu}_0}{C_0^p}, \bar{\mu}_1\right\}\sum_{m=0}^{n}\|F(z_m^\delta) - y^\delta\|^p \\ &\geqslant \min\left\{\frac{\bar{\mu}_0}{C_0^p}, \bar{\mu}_1\right\}(n+1)\tau^p\delta^p\end{aligned}$$

for all $n \geqslant 0$. By taking $n \to \infty$ we derive a contradiction. Thus n_δ must be finite. □

Remark 3.2 In the proof of Proposition 3.4, the condition (3.15) plays a crucial role. Note that, by the definition of our method (3.3), z_n^δ depends on λ_n^δ. Therefore, it is not immediately clear how to choose λ_n^δ to make (3.15) satisfied. One may ask if there exists λ_n^δ such that (3.15) holds. Obviously $\lambda_n^\delta = 0$ satisfies the inequality, which correspond to the Landweber iteration. In order to achieve acceleration, it is necessary to find nontrivial λ_n^δ. Note that when $\|F(z_n^\delta) - y^\delta\| \leqslant \tau\delta$ occurs, (3.15) forces $\lambda_n^\delta = 0$ because $\mu_n^\delta = 0$. Therefore we only need to consider the case $\|F(z_n^\delta) - y^\delta\| > \tau\delta$. By using (3.22) we can derive a sufficient condition

$$\frac{1}{p^*(2c_0)^{p^*-1}}\left(\lambda_n^\delta + (\lambda_n^\delta)^{p^*}\right)\|\xi_n^\delta - \xi_{n-1}^\delta\|^{p^*} \leqslant M\tau^p\delta^p,$$

where

$$M := \frac{c_1}{\nu}\min\left\{\frac{\bar{\mu}_0}{C_0^p}, \bar{\mu}_1\right\}.$$

Considering the particular case when $p = 2$, this thus leads to the choice

$$\lambda_n^\delta := \min \left\{ -\frac{1}{2} + \sqrt{\frac{1}{4} + \frac{4c_0 M \tau^2 \delta^2}{\|\xi_n^\delta - \xi_{n-1}^\delta\|^2}}, \frac{n}{n+\alpha} \right\}, \tag{3.23}$$

where $\alpha \geqslant 3$ is a given number. Note that in the above formula for λ_n^δ, inside the "min" the second argument is taken to be $n/(n+\alpha)$ which is the combination parameter used in Nesterov's acceleration strategy; in case the first argument is large, this formula may lead to $\lambda_n^\delta = n/(n+\alpha)$ and consequently the acceleration effect of Nesterov can be utilized. For general $p > 1$, by placing the requirement $0 \leqslant \lambda_n^\delta \leqslant n/(n+\alpha) \leqslant 1$, one may choose λ_n^δ to satisfy

$$\frac{2\lambda_n^\delta}{p^*(2c_0)^{p^*-1}} \|\xi_n^\delta - \xi_{n-1}^\delta\|^{p^*} \leqslant M\tau^p\delta^p$$

which leads to the choice

$$\lambda_n^\delta = \min \left\{ \frac{\gamma_0 \delta^p}{\|\xi_n^\delta - \xi_{n-1}^\delta\|^{p^*}}, \frac{n}{n+\alpha} \right\}, \quad \gamma_0 := \frac{1}{2}(2c_0)^{p^*-1} p^* M \tau^p. \tag{3.24}$$

We remark that the choices of λ_n^δ given in (3.23) and (3.24) may decrease to 0 as $\delta \to 0$, consequently the acceleration effect could also decrease for $\delta \to 0$. Since for small values of δ the acceleration is needed most, other strategies should be explored. We will give a further consideration on the choice of λ_n^δ in the next subsection.

In order to establish the regularization property of the method (3.3), we need to consider its noise-free counterpart. By dropping the superscript δ in all the quantities involved in (3.3), it leads to the following formulation of the two-point gradient method for the noise-free case:

$$\begin{aligned} \zeta_n &= \xi_n + \lambda_n(\xi_n - \xi_{n-1}), \\ z_n &= \nabla\Theta^*(\zeta_n), \\ \xi_{n+1} &= \zeta_n - \mu_n L(z_n)^* J_s^{\mathcal{Y}}(r_n), \\ x_{n+1} &= \nabla\Theta^*(\xi_{n+1}) \end{aligned} \tag{3.25}$$

with $\xi_{-1} = \xi_0$, where $r_n := F(z_n) - y$, $\lambda_n \geqslant 0$ is the combination parameter, and μ_n is the step size given by

$$\mu_n = \begin{cases} \min\left\{ \dfrac{\bar{\mu}_0 \|r_n\|^{p(s-1)}}{\|L(z_n)^* J_s^{\mathcal{Y}}(r_n)\|^p}, \bar{\mu}_1 \right\} \|r_n\|^{p-s} & \text{if } F(z_n) \neq y, \\ 0 & \text{if } F(z_n) = y. \end{cases} \tag{3.26}$$

We will first establish a convergence result for (3.25). The following result plays a crucial role in the argument.

Proposition 3.5 *Consider the Eq.* (3.1) *for which Assumption* 2 *holds. Let* $\Theta : \mathcal{X} \to (-\infty, \infty]$ *be a proper, lower semi-continuous and uniformly convex function. Let* $\{x_n\} \subset B_{2\rho}(x_0) \cap \mathscr{D}(\Theta)$ *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* (3.1) *in* $B_{2\rho}(x_0) \cap \mathscr{D}(\Theta)$ *the sequence* $\{D_{\xi_n}\Theta(\hat{x}, x_n)\}$ *is monotonically decreasing;*
(iii) $\lim_{n\to\infty} \|F(x_n) - y\| = 0.$
(iv) *there is a subsequence* $\{n_k\}$ *with* $n_k \to \infty$ *such that for any solution* $\hat{x}$ *of* (3.1) *in* $B_{2\rho}(x_0) \cap \mathscr{D}(\Theta)$ *there holds*

$$\lim_{l\to\infty} \sup_{k\geqslant l} |\langle \xi_{n_k} - \xi_{n_l}, x_{n_k} - \hat{x}\rangle| = 0. \tag{3.27}$$

Then there exists a solution x_* *of* (3.1) *in* $B_{2\rho}(x_0) \cap \mathscr{D}(\Theta)$ *such that*

$$\lim_{n\to\infty} D_{\xi_n}\Theta(x_*, x_n) = 0.$$

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

Proof This result essentially follows from [18, Proposition 3.6] and its proof. □

Theorem 3.6 *Let* $\mathcal{X}$ *be reflexive, let* $\mathcal{Y}$ *be uniformly smooth, and let Assumptions* 1 *and* 2 *hold. Assume that* $\bar{\mu}_0 > 0$ *is chosen such that*

$$1 - \eta - \frac{1}{p^*}\left(\frac{\bar{\mu}_0}{2c_0}\right)^{p^*-1} > 0$$

and the combination parameters $\{\lambda_n\}$ *are chosen to satisfy the counterparts of* (3.14) *and* (3.15) *with* $\delta = 0$ *and*

$$\sum_{n=0}^{\infty} \lambda_n \|\xi_n - \xi_{n-1}\| < \infty. \tag{3.28}$$

Then, there exists a solution x_* *of* (3.1) *in* $B_{2\rho}(x_0) \cap \mathscr{D}(\Theta)$ *such that*

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

If, in addition, $\mathcal{N}(L(x^\dagger)) \subset \mathcal{N}(L(x))$ *for all* $x \in B_{3\rho}(x_0)$*, then* $x_* = x^\dagger$*.*

Proof We will use Proposition 3.5 to prove the result. By the definition $x_n = \nabla\Theta^*(\xi_n)$ we have $\xi_n \in \partial\Theta(x_n)$ which shows (i) in Proposition 3.5. By using the same argument for proving Proposition 3.4 we can show that $z_n \in B_{3\rho}(x_0)$ and $x_n \in B_{2\rho}(x_0)$ for all n with

$$D_{\xi_{n+1}}\Theta(\hat{x}, x_{n+1}) \leqslant D_{\xi_n}\Theta(\hat{x}, x_n) \tag{3.29}$$

for any solution $\hat{x}$ of (3.1) in $B_{2\rho}(x_0) \cap \mathscr{D}(\Theta)$ and

$$\sum_{n=0}^{\infty} \mu_n \|F(z_n) - y\|^s < \infty. \tag{3.30}$$

From (3.29) it follows that (ii) in Proposition 3.5 holds. Moreover, by using the definition of μ_n and the similar derivation for (3.22) we have

$$\min\left\{\frac{\bar{\mu}_0}{C_0^p}, \bar{\mu}_1\right\} \|F(z_n) - y\|^p \leqslant \mu_n \|F(z_n) - y\|^s \leqslant \bar{\mu}_1 \|F(z_n) - y\|^p.$$

Thus it follows from (3.30) that

$$\sum_{n=0}^{\infty} \|F(z_n) - y\|^p < \infty.$$

Consequently

$$\lim_{n\to\infty} \|F(z_n) - y\| = 0. \tag{3.31}$$

By using Assumption 2(c), (3.25), (2.5) and (3.15) with $\delta = 0$, we have

$$\begin{aligned}
\|F(x_n) - F(z_n)\| &\leqslant \frac{1}{1-\eta} \|L(z_n)(x_n - z_n)\| \leqslant \frac{C_0}{1-\eta} \|x_n - z_n\| \\
&= \frac{C_0}{1-\eta} \|\nabla\Theta^*(\xi_n) - \nabla\Theta^*(\zeta_n)\| \\
&\leqslant \frac{C_0}{(1-\eta)(2c_0)^{p^*-1}} \|\xi_n - \zeta_n\|^{p^*-1} \\
&= \frac{C_0}{(1-\eta)(2c_0)^{p^*-1}} \lambda_n^{p^*-1} \|\xi_n - \xi_{n-1}\|^{p^*-1} \\
&\leqslant \frac{C_0}{1-\eta} \left(\frac{c_1 p^*}{2c_0 \nu}\right)^{1/p} (\mu_n \|F(z_n) - y\|^s)^{1/p} \\
&\leqslant \frac{C_0}{1-\eta} \left(\frac{c_1 p^* \bar{\mu}_1}{2c_0 \nu}\right)^{1/p} \|F(z_n) - y\|.
\end{aligned} \tag{3.32}$$

The combination of (3.31) and (3.32) implies that $\|F(x_n) - y\| \to 0$ as $n \to \infty$ which shows (iii) in Proposition 3.5.

In order to establish the convergence result, it remains only to show (iv) in Proposition 3.5. To this end, we consider $\|F(z_n) - y\|$. It is known that $\|F(z_n) - y\| \to 0$ as $n \to \infty$. If $\|F(z_n) - y\| = 0$ for some n, then (3.15) with $\delta = 0$ forces $\lambda_n(\xi_n - \xi_{n-1}) = 0$. Thus $\zeta_n = \xi_n$ by (3.25). On the other hand, we also have $\mu_n = 0$ and hence $\xi_{n+1} = \zeta_n$. Consequently $\xi_{n+1} = \zeta_n = \xi_n$ and

$$\zeta_{n+1} = \xi_{n+1} + \lambda_{n+1}(\xi_{n+1} - \xi_n) = \xi_{n+1} = \zeta_n.$$

Thus

$$z_{n+1} = \nabla\Theta^*(\zeta_{n+1}) = \nabla\Theta^*(\zeta_n) = z_n$$

which implies that $F(z_{n+1}) = F(z_n) = y$. By repeating the argument one can see that $F(z_m) = y$ for all $m \geqslant n$. Based on the above facts, we therefore can choose a strictly increasing sequence $\{n_k\}$ of integers by letting $n_0 = 0$ and for each $k \geqslant 1$, letting n_k be the first integer satisfying

$$n_k \geqslant n_{k-1} + 1 \quad \text{and} \quad \|F(z_{n_k}) - y\| \leqslant \|F(z_{n_{k-1}}) - y\|.$$

For such chosen strictly increasing sequence $\{n_k\}$, it is easily seen that

$$\|F(z_{n_k}) - y\| \leqslant \|F(z_n) - y\|, \quad 0 \leqslant n < n_k. \tag{3.33}$$

For any integers $0 \leqslant l < k < \infty$, we consider

$$\langle \xi_{n_k} - \xi_{n_l}, x_{n_k} - \hat{x}\rangle = \sum_{n=n_l}^{n_k-1} \langle \xi_{n+1} - \xi_n, x_{n_k} - \hat{x}\rangle.$$

By using the definition of ξ_{n+1} we have

$$\xi_{n+1} - \xi_n = \lambda_n(\xi_n - \xi_{n-1}) - \mu_n L(z_n)^* J_s^{\mathcal{Y}}(F(z_n) - y).$$

Therefore, by using the property of $J_s^{\mathcal{Y}}$, we have

$$\begin{aligned}
\left|\langle \xi_{n_k} - \xi_{n_l}, x_{n_k} - \hat{x}\rangle\right| &\leqslant \sum_{n=n_l}^{n_k-1} \lambda_n |\langle \xi_n - \xi_{n-1}, x_{n_k} - \hat{x}\rangle| \\
&\quad + \sum_{n=n_l}^{n_k-1} \mu_n |\langle J_s^{\mathcal{Y}}(F(z_n) - y), L(z_n)(x_{n_k} - \hat{x})\rangle| \\
&\leqslant \sum_{n=n_l}^{n_k-1} \lambda_n \|\xi_n - \xi_{n-1}\| \|x_{n_k} - \hat{x}\| \\
&\quad + \sum_{n=n_l}^{n_k-1} \mu_n \|F(z_n) - y\|^{s-1} \|L(z_n)(x_{n_k} - \hat{x})\|.
\end{aligned} \tag{3.34}$$

By using Assumption 2(c) and (3.33), we obtain for $n < n_k$ that

$$\begin{aligned}
\|L(z_n)(x_{n_k} - \hat{x})\| &\leqslant \|L(z_n)(x_{n_k} - z_n)\| + \|L(z_n)(z_n - \hat{x})\| \\
&\leqslant (1+\eta)\left(\|F(x_{n_k}) - F(z_n)\| + \|F(z_n) - y\|\right)
\end{aligned}$$

$$\begin{aligned}
&\leqslant 2(1+\eta)\|F(z_n)-y\| + (1+\eta)\|F(x_{n_k})-y\| \\
&\leqslant 2(1+\eta)\|F(z_n)-y\| \\
&\quad + (1+\eta)\left(\|F(x_{n_k})-F(z_{n_k})\| + \|F(z_{n_k})-y\|\right) \\
&\leqslant 3(1+\eta)\|F(z_n)-y\| + (1+\eta)\|F(x_{n_k})-F(z_{n_k})\|.
\end{aligned}$$

By using (3.32) and (3.33), we have for $n < n_k$ that

$$\begin{aligned}
\|F(x_{n_k})-F(z_{n_k})\| &\leqslant \frac{C_0}{1-\eta}\left(\frac{c_1 p^* \bar{\mu}_1}{2c_0\nu}\right)^{1/p}\|F(z_{n_k})-y\| \\
&\leqslant \frac{C_0}{1-\eta}\left(\frac{c_1 p^* \bar{\mu}_1}{2c_0\nu}\right)^{1/p}\|F(z_n)-y\|.
\end{aligned}$$

Therefore

$$\|L(z_n)(x_{n_k}-\hat{x})\| \leqslant C_1\|F(z_n)-y\| \tag{3.35}$$

for $n < n_k$, where $C_1 := 3(1+\eta) + \frac{(1+\eta)C_0}{1-\eta}\left(\frac{c_1 p^* \bar{\mu}_1}{2c_0\nu}\right)^{1/p}$. Combining (3.35) with (3.34) and using $x_{n_k} \in B_{2\rho}(x_0)$ we obtain

$$\begin{aligned}
\left|\langle \xi_{n_k}-\xi_{n_l}, x_{n_k}-\hat{x}\rangle\right| &\leqslant \sum_{n=n_l}^{n_k-1} \lambda_n\|\xi_n-\xi_{n-1}\|\|x_{n_k}-\hat{x}\| + C_1\sum_{n=n_l}^{n_k-1}\mu_n\|F(z_n)-y\|^s \\
&\leqslant 4\rho\sum_{n=n_l}^{n_k-1}\lambda_n\|\xi_n-\xi_{n-1}\| + C_1\sum_{n=n_l}^{n_k-1}\mu_n\|F(z_n)-y\|^s.
\end{aligned}$$

By making use of (3.20) with $\delta = 0$, we obtain, with $C_2 := \nu C_1/((\nu-1)c_1)$, that

$$\begin{aligned}
&\left|\langle \xi_{n_k}-\xi_{n_l}, x_{n_k}-\hat{x}\rangle\right| \\
&\quad\leqslant 4\rho\sum_{n=n_l}^{n_k-1}\lambda_n\|\xi_n-\xi_{n-1}\| + C_2\sum_{n=n_l}^{n_k-1}\left(D_{\xi_n}\Theta(\hat{x},x_n) - D_{\xi_{n+1}}\Theta(\hat{x},x_{n+1})\right) \\
&\quad= 4\rho\sum_{n=n_l}^{n_k-1}\lambda_n\|\xi_n-\xi_{n-1}\| + C_2\left(D_{\xi_{n_l}}\Theta(\hat{x},x_{n_l}) - D_{\xi_{n_k}}\Theta(\hat{x},x_{n_k})\right).
\end{aligned} \tag{3.36}$$

Let $\gamma := \lim_{n\to\infty} D_{\xi_n}\Theta(\hat{x},x_n)$ whose existence is guaranteed by the monotonicity of $\{D_{\xi_n}\Theta(\hat{x},x_n)\}$. Then, for each fixed l, we have

$$\sup_{k\geqslant l}\left|\langle \xi_{n_k}-\xi_{n_l}, x_{n_k}-\hat{x}\rangle\right| \leqslant 4\rho\sum_{n=n_l}^{\infty}\lambda_n\|\xi_n-\xi_{n-1}\| + C_2\left(D_{\xi_{n_l}}\Theta(\hat{x},x_{n_l}) - \gamma\right).$$

Thus it follows from (3.28) that

$$\lim_{l\to\infty}\sup_{k\geqslant l}\left|\langle \xi_{n_k}-\xi_{n_l}, x_{n_k}-\hat{x}\rangle\right| \leqslant C_2\left(\lim_{l\to\infty} D_{\xi_{n_l}}\Theta(\hat{x}, x_{n_l})-\gamma\right)=0$$

which verifies (iv) in Proposition 3.5.

To show $x_* = x^\dagger$ under the additional condition $\mathcal{N}(L(x^\dagger)) \subset \mathcal{N}(L(x))$ for all $x \in B_{3\rho}(x_0)$, we observe from (3.25) and $\xi_0 - \xi_{-1} = 0$ that

$$\begin{aligned}\xi_{n+1}-\xi_n &= -\mu_n L(z_n)^*(F(z_n)-y)+\lambda_n(\xi_n-\xi_{n-1})\\ &= -\sum_{k=0}^{n}\left(\prod_{i=k+1}^{n}\lambda_i\right)\mu_k L(z_k)^*(F(z_k)-y).\end{aligned}$$

Since $\mathcal{X}$ is reflexive and $\mathcal{N}(L(x^\dagger)) \subset \mathcal{N}(L(x))$, we have $\overline{\mathcal{R}(L(x)^*)} \subset \overline{\mathcal{R}(L(x^\dagger)^*)}$ for all $x \in B_{3\rho}(x_0)$. Recall that $z_k \in B_{3\rho}(x_0)$. It thus follows from the above formula that $\xi_{n+1}-\xi_n \in \overline{\mathcal{R}(L(x^\dagger)^*)}$. Therefore we may use the second part of Proposition 3.5 to conclude the proof. □

Next, we are going to show that, using the discrepancy principle (3.5) as a stopping rule, our method (3.3) becomes a convergent regularization method, if we additionally assume that λ_n^δ depends continuously on δ in the sense that $\lambda_n^\delta \to \lambda_n$ as $\delta \to 0$ for all n. We need the following stability result.

Lemma 3.7 *Let $\mathcal{X}$ be reflexive, let $\mathcal{Y}$ be uniformly smooth, and let Assumptions 1 and 2 hold. Assume that $\tau > 1$ and $\bar{\mu}_0 > 0$ are chosen to satisfy (3.12). Assume also that the combination parameters $\{\lambda_n^\delta\}$ are chosen to depend continuously on δ as $\delta \to 0$ and satisfy (3.14), (3.15) and (3.28). Then for all $n \geqslant 0$ there hold*

$$\zeta_n^\delta \to \zeta_n, \quad z_n^\delta \to z_n, \quad \xi_n^\delta \to \xi_n \quad \textit{and} \quad x_n^\delta \to x_n \quad \textit{as } \delta \to 0.$$

Proof The result is trivial for $n = 0$. We next assume that the result is true for all $0 \leqslant n \leqslant m$ and show that the result is also true for $n = m+1$. We consider two cases.

Case 1: $F(z_m) = y$. In this case we have $\mu_m = 0$ and $\|F(z_m^\delta)-y^\delta\| \to 0$ as $\delta \to 0$ by the continuity of F and the induction hypothesis $z_m^\delta \to z_m$. Thus

$$\xi_{m+1}^\delta - \xi_{m+1} = \zeta_m^\delta - \zeta_m - \mu_m^\delta L(z_m^\delta)^* J_s^{\mathcal{Y}}(F(z_m^\delta)-y^\delta),$$

which together with the definition of μ_m^δ and the induction hypothesis $\zeta_m^\delta \to \zeta_m$ implies that

$$\|\xi_{m+1}^\delta - \xi_{m+1}\| \leqslant \|\zeta_m^\delta - \zeta_m\| + C_0\bar{\mu}_1\|F(z_m^\delta)-y^\delta\|^{p-1} \to 0 \quad \text{as } \delta \to 0.$$

Consequently, by using the continuity of $\nabla\Theta^*$ we have $x_{m+1}^\delta = \nabla\Theta^*(\xi_{m+1}^\delta) \to \nabla\Theta^*(\xi_{m+1}) = x_{m+1}$ as $\delta \to 0$. Recall that

$$\zeta_{m+1}^\delta = \xi_{m+1}^\delta + \lambda_{m+1}^\delta(\xi_{m+1}^\delta - \xi_m^\delta), \qquad z_{m+1}^\delta = \nabla\Theta^*(\zeta_{m+1}^\delta).$$

We may use the condition $\lambda_{m+1}^\delta \to \lambda_{m+1}$ to conclude that $\zeta_{m+1}^\delta \to \zeta_{m+1}$ and $z_{m+1}^\delta \to z_{m+1}$ as $\delta \to 0$.

Case 2: $F(z_m) \neq y$. In this case we have $\|F(z_m^\delta) - y^\delta\| \geqslant \tau\delta$ for small $\delta > 0$. Therefore

$$\mu_m^\delta = \min\left\{\frac{\bar{\mu}_0\|F(z_m^\delta) - y^\delta\|^{p(s-1)}}{\|L(z_m^\delta)^* J_s^{\mathcal{Y}}(F(z_m^\delta) - y^\delta)\|^p}, \bar{\mu}_1\right\}\|F(z_m^\delta) - y^\delta\|^{p-s},$$

$$\mu_m = \min\left\{\frac{\bar{\mu}_0\|F(z_m) - y\|^{p(s-1)}}{\|L(z_m)^* J_s^{\mathcal{Y}}(F(z_m) - y)\|^p}, \bar{\mu}_1\right\}\|F(z_m) - y\|^{p-s}.$$

If $L(z_m)^* J_s^{\mathcal{Y}}(F(z_m) - y) \neq 0$, then, by the induction hypothesis on z_m^δ, it is easily seen that $\mu_m^\delta \to \mu_m$ as $\delta \to 0$. If $L(z_m)^* J_s^{\mathcal{Y}}(F(z_m) - y) = 0$, then $\mu_m = \bar{\mu}_1\|F(z_m) - y\|^{p-s}$ and $\mu_m^\delta = \bar{\mu}_1\|F(z_m^\delta) - y^\delta\|^{p-s}$ for small $\delta > 0$. This again implies that $\mu_m^\delta \to \mu_m$ as $\delta \to 0$. Consequently, by utilizing the continuity of F, L, $J_s^{\mathcal{Y}}$ and $\nabla\Theta^*$ and the induction hypotheses, we can conclude that $\xi_{m+1}^\delta \to \xi_{m+1}$, $x_{m+1}^\delta \to x_{m+1}$, $\zeta_{m+1}^\delta \to \zeta_{m+1}$ and $z_{m+1}^\delta \to z_{m+1}$ as $\delta \to 0$. □

We are now in a position to give the main convergence result on our method (3.3).

Theorem 3.8 *Let $\mathcal{X}$ be reflexive, let $\mathcal{Y}$ be uniformly smooth, and let Assumptions 1 and 2 hold. Assume that $\tau > 1$ and $\bar{\mu}_0 > 0$ are chosen to satisfy (3.12). Assume also that the combination parameters $\{\lambda_n^\delta\}$ are chosen to depend continuously on δ as $\delta \to 0$ and satisfy (3.14), (3.15) and (3.28). Let n_δ be chosen according to the discrepancy principle (3.5). Then there exists a solution x_* of (3.1) in $B_{2\rho}(x_0) \cap \mathscr{D}(\Theta)$ such that*

$$\lim_{\delta\to 0}\|x_{n_\delta}^\delta - x_*\| = 0 \quad \textit{and} \quad \lim_{\delta\to 0} D_{\xi_{n_\delta}^\delta}\Theta(x_*, x_{n_\delta}^\delta) = 0.$$

If, in addition, $\mathcal{N}(L(x^\dagger)) \subset \mathcal{N}(L(x))$ for all $x \in B_{3\rho}(x_0)$, then $x_* = x^\dagger$.

Proof This result can be proved by the same argument in the proof of Theorem 3.10. So we omit the details. □

3.2 DBTS: the choice of λ_n^δ

In this section we will discuss the choice of the combination parameter λ_n^δ which leads to a convergent regularization method.

In Remark 3.2 we have briefly discussed how to choose the combination parameter leading to the formulae (3.23) and (3.24). However, these choices of λ_n^δ decrease to 0 as $\delta \to 0$, and consequently the acceleration effect will decrease as $\delta \to 0$ as well. Therefore, it is necessary to find out other strategy for generating λ_n^δ such that (3.14) and (3.15) hold. We will adapt the discrete backtracking search (DBTS) algorithm introduced in [14] to our situation. To this end, we take a function $q : \mathbb{N}\cup\{0\} \to (0, \infty)$ that is non-increasing and

$$\sum_{i=0}^{\infty} q(i) < \infty. \tag{3.37}$$

The DBTS algorithm for choosing the combination parameter λ_n^δ in our method (3.3) is formulated in Algorithm 1 below. Comparing with the one in [14], there are two modifications: The first modification is the definition of β_n in which we place $\beta_n(i) \leqslant n/(n+\alpha)$ instead of $\beta_n(i) \leqslant 1$; this modification gives the possibility to speed up convergence by making use of the Nesterov's acceleration strategy. The second modification is in the "**Else**" part, where instead of setting $\lambda_n^\delta = 0$ we calculate λ_n^δ by (3.24); this modification can provide additional acceleration to speed up convergence.

Algorithm 1 Discrete backtracking search (DBTS) algorithm for λ_n^δ, $n \geqslant 1$.

Given $\xi_n^\delta, \xi_{n-1}^\delta, \tau, \delta, c_1, \nu, q : \mathbb{N} \to \mathbb{N}, i_{n-1}^\delta \in \mathbb{N}, j_{\max} \in \mathbb{N}$
Set $\gamma_1 = c_1 p^*(2c_0)^{p^*-1}/\nu$
Calculate $\|\xi_n^\delta - \xi_{n-1}^\delta\|$ and define, with $\alpha \geqslant 3$,

$$\beta_n(i) = \min\left\{ \frac{q(i)}{\|\xi_n^\delta - \xi_{n-1}^\delta\|}, \frac{p^*(2c_0)^{p^*}\rho^p}{4\|\xi_n^\delta - \xi_{n-1}^\delta\|^{p^*}}, \frac{n}{n+\alpha} \right\}.$$

For $j = 1, \ldots, j_{\max}$
  Set $\lambda_n^\delta = \beta_n(i_{n-1}^\delta + j)$;
  Calculate $\zeta_n^\delta = \xi_n^\delta + \lambda_n^\delta(\xi_n^\delta - \xi_{n-1}^\delta)$ and $z_n^\delta = \nabla\Theta^*(\zeta_n^\delta)$;
  Calculate μ_n^δ by (3.4);
  If $\|y^\delta - F(z_n^\delta)\| \leqslant \tau\delta$
    $\lambda_n^\delta = 0$;
    $i_n^\delta = i_{n-1}^\delta + j$;
    break;
  Else if $(\lambda_n^\delta + \left(\lambda_n^\delta\right)^{p^*})\|\xi_n^\delta - \xi_{n-1}^\delta\|^{p^*} \leqslant \gamma_1\mu_n^\delta\|F(z_n^\delta) - y^\delta\|^s$
    $i_n^\delta = i_{n-1}^\delta + j$;
    break;
  Else
    calculate λ_n^δ by (3.24);
    $i_n^\delta = i_{n-1}^\delta + j_{\max}$;
  End If
End For
Output: $\lambda_n^\delta, i_n^\delta$

We need to show that the combination parameter λ_n^δ chosen by Algorithm 1 satisfies (3.14) and (3.15). From Algorithm 1 it is easily seen that $0 \leqslant \lambda_n^\delta \leqslant \beta(i_n^\delta)$. Therefore (3.14) holds automatically. When $\|F(z_n^\delta) - y^\delta\| \leqslant \tau\delta$, Algorithm 1 gives $\lambda_n^\delta = 0$ which ensures (3.15) hold. When $\|F(z_n^\delta) - y^\delta\| > \tau\delta$, Algorithm 1 either finds λ_n^δ of the form $\beta_n(i_n^\delta)$ to satisfy (3.15) or gives λ_n^δ by (3.24) which again satisfies (3.15). Thus Algorithm 1 always produces a λ_n^δ satisfying (3.15).

We can not use Theorem 3.8 to conclude the regularization property of the two-point gradient method (3.3) when the combination parameter is determined by Algorithm 1 because the produced parameter λ_n^δ is not necessarily continuously dependent on δ as $\delta \to 0$. In fact, λ_n^δ may have many different cluster points as $\delta \to 0$. Using these different cluster points as the combination parameter in (3.25) may lead to many different iterative sequences for noise-free case. We need to consider all these possible iterative sequences altogether. We will use $\Gamma_{\bar{\mu}_0,\bar{\mu}_1,\nu,q}(\xi_0, x_0)$ to denote the set consisting of all the iterative sequences $\{(\xi_n, x_n, \zeta_n, z_n)\} \subset \mathcal{X}^* \times \mathcal{X} \times \mathcal{X}^* \times \mathcal{X}$ defined by (3.25), where the combination parameters $\{\lambda_n\}$ are chosen to satisfy

$$\frac{1}{p^*(2c_0)^{p^*-1}}\left(\lambda_n^{p^*} + \lambda_n\right)\|\xi_n - \xi_{n-1}\|^{p^*} \leqslant \frac{c_1}{\nu}\mu_n\|F(z_n) - y\|^s \tag{3.38}$$

and

$$0 \leqslant \lambda_n \leqslant \min\left\{\frac{q(i_n)}{\|\xi_n - \xi_{n-1}\|}, \frac{p^*(2c_0)^{p^*}\rho^p}{4\|\xi_n - \xi_{n-1}\|^{p^*}}, \frac{n}{n+\alpha}\right\} \tag{3.39}$$

with a sequence $\{i_n\}$ of integers satisfying $i_0 = 0$ and $1 \leqslant i_n - i_{n-1} \leqslant j_{\max}$ for all n.

Given a sequence $\{(\xi_n, x_n, \zeta_n, z_n)\} \in \Gamma_{\bar{\mu}_0,\bar{\mu}_1,\nu,q}(\xi_0, x_0)$, we can check that the corresponding combination parameters $\{\lambda_n\}$ satisfy (3.14), (3.15) and (3.28) with $\delta = 0$. Indeed, (3.38) is exactly (3.15). Since $0 \leqslant \lambda_n \leqslant \frac{p^*(2c_0)^{p^*}\rho^p}{4\|\xi_n - \xi_{n-1}\|^{p^*}}$ and $\lambda_n \leqslant \frac{n}{n+\alpha} < 1$, we have

$$\frac{1}{p^*(2c_0)^{p^*-1}}\left(\lambda_n^{p^*} + \lambda_n\right)\|\xi_n - \xi_{n-1}\|^{p^*} \leqslant \frac{2}{p^*(2c_0)^{p^*-1}}\lambda_n\|\xi_n - \xi_{n-1}\|^{p^*} \leqslant c_0\rho^p$$

which shows (3.14). Moreover, we note that $\lambda_n \leqslant q(i)/\|\xi_n - \xi_{n-1}\|$. By the definition of i_n one can see that $i_n \geqslant i_{n-1} + 1$ and thus $i_n \geqslant n$. Therefore, by using the monotonicity of q, we have

$$\sum_{n=0}^{\infty}\lambda_n\|\xi_n - \xi_{n-1}\| \leqslant \sum_{n=0}^{\infty}\beta_n(i_n)\|\xi_n - \xi_{n-1}\| \leqslant \sum_{n=0}^{\infty}q(i_n) \leqslant \sum_{n=0}^{\infty}q(n) < \infty.$$

Hence (3.28) is satisfied. Thus we may use Theorem 3.6 to conclude the convergence of $\{x_n\}$.

We have the following stability result on the two point gradient method (3.3) with the combination parameters chosen by Algorithm 1.

Lemma 3.9 *Let $\mathcal{X}$ be reflexive, let $\mathcal{Y}$ be uniformly smooth, and let Assumptions 1 and 2 hold. Let $\tau > 1$ and $\bar{\mu}_0 > 0$ be chosen to satisfy (3.12). Let $\{y^{\delta_l}\}$ be a sequence of noisy data satisfying $\|y^{\delta_l} - y\| \leqslant \delta_l$ with $\delta_l \to 0$ as $l \to \infty$. Assume that the combination parameters $\{\lambda_n^{\delta_l}\}$ are chosen by Algorithm 1 with $i_0^{\delta_l} = 0$. Then, by taking a subsequence of $\{y^{\delta_l}\}$ if necessary, there is a sequence $\{(\xi_n, x_n, \zeta_n, z_n)\} \in \Gamma_{\bar{\mu}_0,\bar{\mu}_1,\nu,q}(\xi_0, x_0)$ such that*

$$\xi_n^{\delta_l} \to \xi_n, \quad x_n^{\delta_l} \to x_n, \quad \zeta_n^{\delta_l} \to \zeta_n \quad \text{and} \quad z_n^{\delta_l} \to z_n \quad \text{as } l \to \infty$$

for all $n \geqslant 0$.

Proof Note that $\xi_0^{\delta_l} = \xi_0$, $x_0^{\delta_l} = x_0$, $\zeta_0^{\delta_l} = \zeta_0$, $z_0^{\delta_l} = z_0$ and $i_0^{\delta_l} = i_0$. Therefore, by the diagonal sequence argument, it suffices to show that, for each integer $n \geqslant 1$, if ξ_{n-1}, x_{n-1}, ζ_{n-1}, z_{n-1} and i_{n-1} are constructed such that

$$\xi_{n-1}^{\delta_l} \to \xi_{n-1}, \quad x_{n-1}^{\delta_l} \to x_{n-1}, \quad \zeta_{n-1}^{\delta_l} \to \zeta_{n-1}, \quad z_{n-1}^{\delta_l} \to z_{n-1} \quad \text{and} \quad i_{n-1}^{\delta_l} = i_{n-1} \tag{3.40}$$

as $l \to \infty$, then, by taking a subsequence of $\{y^{\delta_l}\}$ if necessary, we can construct ξ_n, x_n, ζ_n, z_n and i_n with the desired properties.

To this end, we set

$$\xi_n := \zeta_{n-1} - \mu_{n-1} L(z_{n-1})^* J_r(F(z_{n-1}) - y) \quad \text{and} \quad x_n := \nabla\Theta^*(\xi_n),$$

where μ_{n-1} is defined by (3.26) with n replaced by $n-1$. By the similar argument in the proof of Lemma 3.7, we can show that

$$\xi_n^{\delta_l} \to \xi_n \quad \text{and} \quad x_n^{\delta_l} \to x_n \quad \text{as } l \to \infty. \tag{3.41}$$

Note that the combination parameter $\lambda_n^{\delta_l}$ determined by Algorithm 1 satisfies

$$\frac{1}{p^*(2c_0)^{p^*-1}} \left(\left(\lambda_n^{\delta_l}\right)^{p^*} + \lambda_n^{\delta_l} \right) \|\xi_n^{\delta_l} - \xi_{n-1}^{\delta_l}\|^{p^*} \leqslant \frac{c_1}{\nu} \mu_n^{\delta_l} \|F(z_n^{\delta_l}) - y^{\delta_l}\|^s \tag{3.42}$$

and

$$\begin{aligned} &0 \leqslant \lambda_n^{\delta_l} \leqslant \min\left\{ \frac{q(i_n^{\delta_l})}{\|\xi_n^{\delta_l} - \xi_{n-1}^{\delta_l}\|}, \frac{p^*(2c_0)^{p^*}\rho^p}{4\|\xi_n^{\delta_l} - \xi_{n-1}^{\delta_l}\|^{p^*}}, \frac{n}{n+\alpha} \right\} \quad \text{or} \\ &\lambda_n^{\delta_l} = \min\left\{ \frac{(2c_0)^{p^*-1} p^* M \tau^p \delta_l^p}{2\|\xi_n^{\delta_l} - \xi_{n-1}^{\delta_l}\|^{p^*}}, \frac{n}{n+\alpha} \right\} \end{aligned} \tag{3.43}$$

with $1 \leqslant i_n^{\delta_l} - i_{n-1}^{\delta_l} \leqslant j_{\max}$. Since $0 \leqslant \lambda_n^{\delta_l} \leqslant n/(n+\alpha)$ and $n \leqslant i_n^{\delta_l} \leqslant n j_{\max}$, by taking a subsequence of $\{y^{\delta_l}\}$ again if necessary, we have

$$\lim_{l\to\infty} \lambda_n^{\delta_l} = \lambda_n \quad \text{and} \quad i_n^{\delta_l} = i_n \quad \text{for all } l \tag{3.44}$$

for some number λ_n and some integer i_n. We set

$$\zeta_n := \xi_n + \lambda_n(\xi_n - \xi_{n-1}) \quad \text{and} \quad z_n := \nabla\Theta^*(\zeta_n).$$

By using (3.40), (3.41), (3.44) and the continuity of $\nabla\Theta^*$, we can obtain $\zeta_n^{\delta_l} \to \zeta_n$ and $z_n^{\delta_l} \to z_n$ as $l \to \infty$. Now we define μ_n by (3.26). By using the similar argument in the proof of Lemma 3.7 again we have

$$\mu_n^{\delta_l}\|F(z_n^{\delta_l})-y^{\delta_l}\|^s\to\mu_n\|F(z_n)-y\|^s \quad \text{as } l\to\infty.$$

Therefore, by taking $l\to\infty$ in (3.42) and (3.43), we can see that λ_n satisfies (3.38) and (3.39). Furthermore, by using $1\leqslant i_n^{\delta_l}-i_{n-1}^{\delta_l}\leqslant j_{\max}$, $i_{n-1}^{\delta_l}=i_{n-1}$ and $i_n^{\delta_l}=i_n$, we immediately have $1\leqslant i_n-i_{n-1}\leqslant j_{\max}$. We thus complete the proof. □

We are now ready to show the regularization property of the method (3.3) when the combination parameter λ_n^δ is chosen by Algorithm 1.

Theorem 3.10 *Let $\mathcal{X}$ be reflexive, let $\mathcal{Y}$ be uniformly smooth, and let Assumptions 1 and 2 hold. Let $\tau>1$ and $\bar{\mu}_0>0$ be chosen to satisfy (3.12). Let $\{y^\delta\}$ be a family of noisy data satisfying $\|y^\delta-y\|\leqslant\delta\to 0$. Assume that the combination parameters $\{\lambda_n^\delta\}$ are chosen by Algorithm 1 with $i_0^\delta=0$. Let n_δ be the integer determined by the discrepancy principle (3.5).*

(a) *For any subsequence $\{y^{\delta_l}\}$ of $\{y^\delta\}$ with $\delta_l\to 0$ as $l\to\infty$, by taking a subsequence of $\{y^{\delta_l}\}$ if necessary, there hold*

$$\lim_{l\to\infty}\|x_{n_{\delta_l}}^{\delta_l}-x_*\|=0 \quad \textit{and} \quad \lim_{l\to\infty}D_{\xi_{n_{\delta_l}}^{\delta_l}}\Theta(x_*,x_{n_{\delta_l}}^{\delta_l})=0$$

for some solution x_* of (3.1) in $B_{2\rho}(x_0)\cap\mathscr{D}(\Theta)$.

(b) *If, in addition, $\mathcal{N}(L(x^\dagger))\subset\mathcal{N}(L(x))$ for all $x\in B_{3\rho}(x_0)\cap\mathscr{D}(\Theta)$, then*

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

Proof Let $\{y^{\delta_l}\}$ be a sequence of noisy data satisfying $\|y^{\delta_l}-y\|\leqslant\delta_l\to 0$ as $l\to\infty$. Let $N:=\liminf_{l\to\infty}n_{\delta_l}$. By taking a subsequence of $\{y^{\delta_l}\}$ if necessary, we may assume $N=\lim_{l\to\infty}n_{\delta_l}$, and according to Lemma 3.9, we can find a sequence $\{(\xi_n,x_n,\zeta_n,z_n)\}\in\Gamma_{\bar{\mu}_0,\bar{\mu}_1,\nu,q}(\xi_0,x_0)$ such that

$$\xi_n^{\delta_l}\to\xi_n \quad \text{and} \quad x_n^{\delta_l}\to x_n \quad \text{as } l\to\infty \tag{3.45}$$

for all $n\geqslant 0$. Due to the properties of the sequences in $\Gamma_{\bar{\mu}_0,\bar{\mu}_1,\nu,q}(\xi_0,x_0)$, we can apply Theorem 3.6 to conclude that $D_{\xi_n}\Theta(x_*,x_n)\to 0$ as $n\to\infty$ for some solution x_* of (3.1) in $B_{2\rho}(x_0)\cap\mathscr{D}(\Theta)$, and if, in addition, $\mathcal{N}(L(x^\dagger))\subset\mathcal{N}(L(x))$ for all $x\in B_{3\rho}(x_0)\cap\mathscr{D}(\Theta)$, then $x_*=x^\dagger$. We will show that

$$\lim_{l\to\infty}D_{\xi_{n_{\delta_l}}^{\delta_l}}\Theta(x_*,x_{n_{\delta_l}}^{\delta_l})=0. \tag{3.46}$$

Case 1: $N<\infty$. We have $n_{\delta_l}=N$ for large l. According to the definition of n_{δ_l}, there holds

$$\|F(z_N^{\delta_l})-y^{\delta_l}\|\leqslant\tau\delta_l.$$

By using the similar argument for deriving (3.32) we have

$$\|F(x_N^{\delta_l}) - F(z_N^{\delta_l})\| \leqslant C\|F(z_N^{\delta_l}) - y^{\delta_l}\|$$

for some universal constant C. Thus

$$\|F(x_N^{\delta_l}) - y^{\delta_l}\| \leqslant (1+C)\|F(z_N^{\delta_l}) - y^{\delta_l}\| \leqslant (1+C)\tau\delta_l.$$

Taking $l \to \infty$ and using the continuity of F gives $F(x_N) = y$. Thus x_N is a solution of (3.1) in $B_{2\rho}(x_0) \cap \mathscr{D}(\Theta)$. By the monotonicity of $\{D_{\xi_n}\Theta(x_N, x_n)\}$ with respect to n, we then obtain

$$D_{\xi_n}\Theta(x_N, x_n) \leqslant D_{\xi_N}\Theta(x_N, x_N) = 0, \quad \forall n \geqslant N.$$

Therefore $x_n = x_N$ for all $n \geqslant N$. Since $x_n \to x_*$ as $n \to \infty$, we must have $x_N = x_*$ and thus $x_{n_{\delta_l}}^{\delta_l} = x_N^{\delta_l} \to x_N = x_*$ as $l \to \infty$. This together with the lower semi-continuity of Θ shows that

$$\begin{aligned} 0 &\leqslant \liminf_{l\to\infty} D_{\xi_{n_{\delta_l}}^{\delta_l}}\Theta(x_*, x_{n_{\delta_l}}^{\delta_l}) \leqslant \limsup_{l\to\infty} D_{\xi_{n_{\delta_l}}^{\delta_l}}\Theta(x_*, x_{n_{\delta_l}}^{\delta_l}) \\ &\leqslant \Theta(x_*) - \liminf_{l\to\infty}\Theta(x_{n_{\delta_l}}^{\delta_l}) - \lim_{l\to\infty}\langle \xi_{n_{\delta_l}}^{\delta_l}, x_* - x_{n_{\delta_l}}^{\delta_l}\rangle \\ &\leqslant \Theta(x_*) - \Theta(x_*) = 0 \end{aligned}$$

which shows (3.46)

Case 2: $N = \infty$. Let n be any fixed integer, then $n_{\delta_l} > n$ for large l. It then follows from Lemma 3.4 that

$$D_{\xi_{n_{\delta_l}}^{\delta_l}}\Theta(x_*, x_{n_{\delta_l}}^{\delta_l}) \leqslant D_{\xi_n^{\delta_l}}\Theta(x_*, x_n^{\delta_l}) = \Theta(x_*) - \Theta(x_n^{\delta_l}) - \langle \xi_n^{\delta_l}, x_* - x_n^{\delta_l}\rangle.$$

By using (3.45) and the lower semi-continuity of Θ we obtain

$$\begin{aligned} 0 &\leqslant \liminf_{l\to\infty} D_{\xi_{n_{\delta_l}}^{\delta_l}}\Theta(x_*, x_{n_{\delta_l}}^{\delta_l}) \leqslant \limsup_{l\to\infty} D_{\xi_{n_{\delta_l}}^{\delta_l}}\Theta(x_*, x_{n_{\delta_l}}^{\delta_l}) \\ &\leqslant \Theta(x_*) - \liminf_{l\to\infty}\Theta(x_n^{\delta_l}) - \lim_{l\to\infty}\langle \xi_n^{\delta_l}, x_* - x_n^{\delta_l}\rangle \\ &\leqslant \Theta(x_*) - \Theta(x_n) - \langle \xi_n, x_* - x_n\rangle \\ &= D_{\xi_n}\Theta(x_*, x_n). \end{aligned}$$

Since n can be arbitrary and $\lim_{n\to\infty} D_{\xi_n}\Theta(x_*, x_n) = 0$, by taking $n \to \infty$ in the above equation we therefore obtain (3.46) again.

If, in addition, $\mathcal{N}(L(x^\dagger)) \subset \mathcal{N}(L(x))$ for all $x \in B_{3\rho}(x_0) \cap \mathscr{D}(\Theta)$, we have $x_* = x^\dagger$. Thus, the above argument shows that any subsequence $\{y^{\delta_l}\}$ of $\{y^\delta\}$ has a subsequence, denoted by the same notation, such that $D_{\xi_{n_{\delta_l}}^{\delta_l}}\mathcal{R}(x^\dagger, x_{n_{\delta_l}}^{\delta_l}) \to 0$ as $l \to \infty$. Therefore $D_{\xi_{n_\delta}^\delta}\mathcal{R}(x^\dagger, x_{n_\delta}^\delta) \to 0$ as $\delta \to 0$. □

Remark 3.3 A two-point gradient method in Hilbert spaces was considered in [14] in which the combination parameter is chosen by a discrete backtracking search (DBTS) algorithm. The regularization property was proved under the condition that the noise-free counterpart of the method never terminates at a solution of (3.1) in finite many steps if the combination parameter is chosen by the DBTS algorithm. This technical condition seems difficult to be verified because the exact data y is unavailable. In Theorem 3.10 we removed this condition by using a stability result established in Lemma 3.9.

4 Numerical simulations

In this section we will present numerical simulations on our TPG-DBTS method, i.e. the two point gradient method (3.3) with the combination parameter λ_n^δ chosen by the DBTS algorithm (Algorithm 1). In order to illustrate the performance of TPG-DBTS algorithm, we will compare the computational results with the ones obtained by the Landweber iteration (1.4) and the Nesterov acceleration of Landweber iteration, i.e. the method (3.3) with $\lambda_n^\delta = n/(n+\alpha)$ for some $\alpha \geqslant 3$. In order to be fair, the step sizes μ_n^δ involved in all these methods are computed by (3.4) and all the iterations are terminated by the discrepancy principle with $\tau = 1.05$.

A key ingredient for the numerical implementation is the determination of $x = \nabla\Theta^*(\xi)$ for any given $\xi \in \mathcal{X}^*$ which is equivalent to solving the minimization problem

$$x = \arg\min_{z\in\mathcal{X}} \{\Theta(z) - \langle \xi, z\rangle\}. \tag{4.1}$$

For some choices of Θ, this minimization problem can be easily solved numerically. For instance, when $\mathcal{X} = L^2(\Omega)$ and the sought solution is piecewise constant, we may choose

$$\Theta(x) = \frac{1}{2\beta}\|x\|_2^2 + |x|_{TV} \tag{4.2}$$

with a constant $\beta > 0$, where $|x|_{TV}$ denotes the total variation of x. Then the minimization problem (4.1) becomes the total variation denoising problem

$$x = \arg\min_{z\in L^2(\Omega)} \left\{\frac{1}{2\beta}\|z - \beta\xi\|_2^2 + |z|_{TV}\right\} \tag{4.3}$$

which is nonsmooth and convex. Note that for this Θ, Assumption 1 holds with $p = 2$ and $c_0 = \frac{1}{2\beta}$. Many efficient algorithms have been developed for solving (4.3), including the fast iterative shrinkage-thresholding algorithm [2,3], the alternating direction method of multipliers [5], and the primal dual hybrid gradient (PDHG) method [33].

In the following numerical simulations we will only consider the situation that the sought solution is piecewise constant. We will use the PDHG method to solve (4.3) iteratively. Our simulations are performed via MATLAB R2012a on a Lenovo laptop with Intel(R) Core(TM) i5 CPU 2.30GHz and 6GB memory.

4.1 Computed tomography

Computed tomography (CT) consists in determining the density of cross sections of a human body by measuring the attenuation of X-rays as they propagate through the biological tissues. Mathematically, it requires to determine a function supported on a bounded domain from its line integrals [25]. In order to apply our method to solve the CT problems, we need a discrete model. In our numerical experiment, we assume that the image is supported on a rectangular domain in $\mathbb{R}^2$ which is divided into $I \times J$ pixels so that the discrete image has size $I \times J$ and can be represented by a vector $x \in \mathbb{R}^N$ with $N = I \times J$. We further assume that there are n_θ projection directions and in each direction there are p X-rays emitted. We want to reconstruct the image by using the measurement data of attenuation along the rays which can be represented by a vector $b \in \mathbb{R}^M$ with $M = n_\theta \times p$. According to a standard discretization of the Radon transform [11], we arrive at a linear algebraic system

$$Fx = b,$$

where F is a sparse matrix of size $M \times N$ whose form depends on the scanner geometry.

In the numerical simulations we consider only test problems that model the standard 2D parallel-beam tomography. The true image is taken to be the modified Shepp-Logan phantom of size 256×256 generated by MATLAB. This phantom is widely used in evaluating tomographic reconstruction algorithms. We use the full angle model with 45 projection angles evenly distributed between 1 and 180 degrees, with 367 lines per projection. The function `paralleltomo` in MATLAB package AIR TOOLS [11] is used to generate the sparse matrix F, which has the size $M = 16{,}515$ and $N = 66{,}536$. Let $x^\dagger$ denote the vector formed by stacking all the columns of the true image and let $b = Fx^\dagger$ be the true data. We add Gaussian noise on b to generate a noisy data b^δ with relative noise level $\delta_{rel} = \|b^\delta - b\|_2/\|b\|_2$ so that the noise level is $\delta = \delta_{rel}\|b\|_2$. We will use b^δ to reconstruct $x^\dagger$. In order to capture the feature of the sought image, we take Θ to be the form (4.2) with $\beta = 1$. In our numerical simulations we will use $\xi_0 = 0$ as an initial guess. According to (3.12) we need $\bar{\mu}_0 < 2(1 - 1/\tau)/\beta$. Therefore we take the parameters $\bar{\mu}_0$ and $\bar{\mu}_1$ in the definition of μ_n^δ to be $\bar{\mu}_0 = 1.8(1 - 1/\tau)/\beta$ and $\bar{\mu}_1 = 20{,}000$. For implementing TPG-DBTS method with λ_n^δ chosen by Algorithm 1, we take $j_{\max} = 1$, $\alpha = 5$, $\gamma_0 = 10$ in (3.24), $\gamma_1 = 1$; we also choose the function $q : \mathbb{N} \to \mathbb{N}$ by $q(m) = m^{-1.1}$. For implementing the Nesterov acceleration of Landweber iteration, we take $\lambda_n^\delta = n/(n+\alpha)$ with $\alpha = 5$. During the computation, the total variation denoising problem (4.3) involved in each iteration step is solved approximately by the PDHG method after 100 iterations.

The computational results by TPG-DBTS, Landweber, and Nesterov acceleration of Landweber are reported in Table 1, including the number of iterations n_δ, the CPU running time and the relative errors $\|x_{n_\delta}^\delta - x^\dagger\|_2/\|x^\dagger\|_2$, using noisy data with various relative noise level $\delta_{rel} > 0$. Table 1 shows that both TPG-DBTS and Nesterov acceleration, terminated by the discrepancy principle, lead to a considerable decrease in the number of iterations and the amount of computational time, which demonstrates that these two methods have the striking acceleration effect. Moreover, both TPG-DBTS and Nesterov acceleration produce more accurate results than Landwe-

Table 1 Numerical results for computed tomography ($\beta = 1$, $\tau = 1.05$)

δ_{rel}	Method	n_δ	CPU time (s)	$\|x_n^\delta - x^\dagger\|_2/\|x^\dagger\|_2$
0.05	Landweber	111	29.94	0.19216
	Nesterov	34	7.45	0.17573
	TPG-DBTS	34	11.99	0.17573
0.01	Landweber	489	106.97	0.06410
	Nesterov	79	16.25	0.05923
	TPG-DBTS	79	29.51	0.05923
0.005	Landweber	879	197.65	0.03578
	Nesterov	109	22.25	0.03165
	TPG-DBTS	109	43.02	0.03165
0.001	Landweber	3299	775.10	0.00694
	Nesterov	247	51.05	0.00521
	TPG-DBTS	247	105.29	0.00521
0.0005	Landweber	5703	1328.31	0.00326
	Nesterov	351	82.93	0.00199
	TPG-DBTS	351	153.79	0.00199

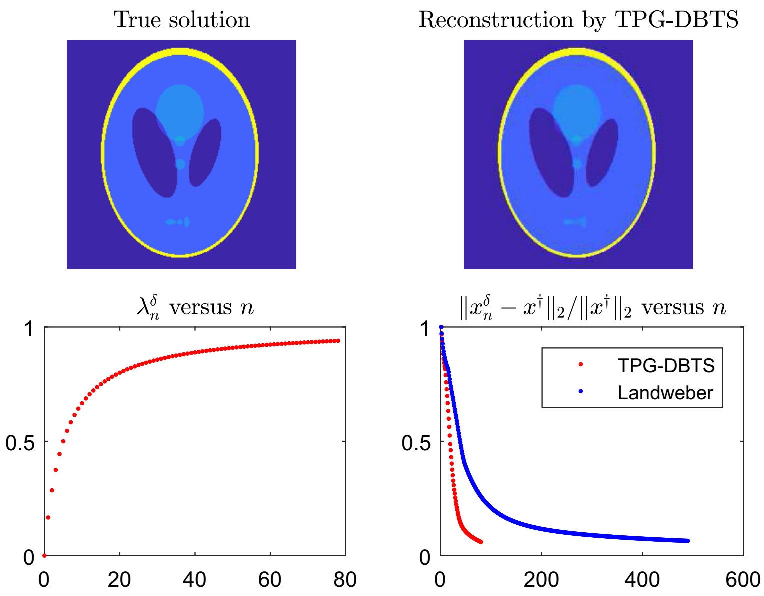

Fig. 1 The computed tomography using noisy data with relative noise level $\delta_{rel} = 0.01$

ber iteration. With the above setup, our computation shows that TPG-DBTS produces the combination parameter λ_n^δ which is exactly same as the combination parameter $n/(n+\alpha)$ in Nesterov acceleration in each iteration step. Therefore, TPG-DBTS and Nesterov acceleration require the same number of iterations and produce the same reconstruction result. Because TPG-DBTS spends more time on determining λ_n^δ, the Nesterov acceleration requires less amount of computational time than TPG-DBTS. However, unlike TPG-DBTS, there exists no convergence result concerning Nesterov acceleration for ill-posed inverse problems.

In order to visualize the reconstruction accuracy of the TPG-DBTS method, we plot in Fig. 1 the true image, the reconstruction result by TPG-DBTS using noisy data with relative noise level $\delta_{rel} = 0.01$, the curve of λ_n^δ versus n, and the relative error $\|x_n^\delta - x^\dagger\|_2/\|x^\dagger\|_2$ versus n for TPG-DBTS and Landweber iteration.

4.2 Elliptic parameter identification

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

$$\begin{cases} -\triangle u + cu = f & \text{in } \Omega, \\ u = g & \text{on } \partial\Omega \end{cases} \tag{4.4}$$

from an $L^2(\Omega)$-measurement of the state u, where $\Omega \subset \mathbb{R}^d$ with $d \leqslant 3$ is a bounded domain with Lipschitz boundary $\partial\Omega$, $f \in H^{-1}(\Omega)$ and $g \in H^{1/2}(\Omega)$. We assume that the sought parameter $c^\dagger$ is in $L^2(\Omega)$. This problem reduces to solving $F(c) = u$ if we define the nonlinear operator $F : L^2(\Omega) \to L^2(\Omega)$ by

$$F(c) := u(c), \tag{4.5}$$

where $u(c) \in H^1(\Omega) \subset L^2(\Omega)$ is the unique solution of (4.4). This operator F is well defined on

$$\mathscr{D} := \left\{ c \in L^2(\Omega) : \|c - \hat{c}\|_{L^2(\Omega)} \leqslant \varepsilon_0 \text{ for some } \hat{c} \geqslant 0, \text{ a.e.} \right\}$$

for some positive constant $\varepsilon_0 > 0$. It is well-known [6] that the operator F is weakly closed and Fréchet differentiable with

$$F'(c)h = v \quad \text{and} \quad F'(c)^*\sigma = -u(c)w$$

for $c \in \mathscr{D}$ and $h, \sigma \in L^2(\Omega)$, where $v, w \in H^1(\Omega)$ are the unique solutions of the problems

$$\begin{cases} -\triangle v + cv = -hu(c) & \text{in } \Omega, \\ v = 0 \quad \text{on } \partial\Omega \end{cases} \quad \text{and} \quad \begin{cases} -\triangle w + cw = \sigma & \text{in } \Omega, \\ w = 0 \quad \text{on } \partial\Omega \end{cases}$$

respectively. Moreover, F satisfies Assumption 2(c).

In our numerical simulation, we consider the two-dimensional problem with $\Omega = [0, 1] \times [0, 1]$ and the sought parameter is assumed to be

$$c^\dagger(x, y) = \begin{cases} 1, & \text{if } (x - 0.65)^2 + (y - 0.36)^2 \leqslant 0.18^2, \\ 0.5, & \text{if } (x - 0.35)^2 + 4(y - 0.75)^2 \leqslant 0.2^2, \\ 0, & \text{elsewhere.} \end{cases}$$

Table 2 Numerical results for 2-dimensional elliptic parameter estimation ($\beta = 10$, $\tau = 1.05$)

δ	Method	n_δ	CPU time (s)	$\|c_n^\delta - c^\dagger\|_{L^2}$
0.005	Landweber	27	8.27	0.27821
	Nesterov	16	5.06	0.22922
	TPG-DBTS	16	4.90	0.22922
0.001	Landweber	135	38.65	0.13500
	Nesterov	57	16.53	0.11433
	TPG-DBTS	57	17.34	0.11433
0.0005	Landweber	201	59.66	0.11200
	Nesterov	92	26.86	0.09416
	TPG-DBTS	91	33.95	0.104
0.0001	Landweber	799	250.03	0.08113
	Nesterov	190	56.60	0.07774
	TPG-DBTS	266	96.54	0.077217
0.00005	Landweber	1689	528.06	0.07189
	Nesterov	358	107.46	0.06546
	TPG-DBTS	484	171.31	0.070146

Assuming $u(c^\dagger) = x + y$, we add random Gaussian noise to produce noisy data u^δ satisfying $\|u^\delta - u(c^\dagger)\|_{L^2(\Omega)} \leqslant \delta$ with various noise level $\delta > 0$. We will use u^δ to reconstruct $c^\dagger$. In order to capture the feature of the sought parameter, we take Θ to be the form (4.2) with $\beta = 10$. We will use the initial guess $\xi_0 = 0$ to carry out the iterations. The parameters $\bar{\mu}_0$ and $\bar{\mu}_1$ in the definition of μ_n^δ are taken to be $\bar{\mu}_0 = 1.8(1 - 1/\tau)/\beta$ and $\bar{\mu}_1 = 20000$. For implementing TPG-DBTS method with λ_n^δ chosen by Algorithm 1, we take $j_{\max} = 1$, $\alpha = 5$, $\gamma_0 = 10$ in (3.24), $\gamma_1 = 1$; we also choose the function $q : \mathbb{N} \to \mathbb{N}$ by $q(m) = m^{-1.1}$. For implementing the Nesterov acceleration of Landweber iteration, we take $\lambda_n^\delta = n/(n + \alpha)$ with $\alpha = 5$. In order to carry out the computation, we divide Ω into 128×128 small squares of equal size and solve all partial differential equations involved approximately by a multigrid method [7] via finite difference discretization. The total variation denoising problem (4.3) involved in each iteration step is solved by the PDHG method after 200 iterations.

In Table 2 we report the computational results by TPG-DBTS, Landweber, and Nesterov acceleration of Landweber, including the number of iterations n_δ, the CPU running time and the absolute errors $\|c_{n_\delta}^\delta - c^\dagger\|_{L^2(\Omega)}$, for various noise level $\delta > 0$. Table 2 shows that both TPG-DBTS and Nesterov acceleration, terminated by the discrepancy principle, reduce the number of iterations and the amount of computational time significantly, and produce more accurate results than Landweber iteration. This demonstrates that these two methods have a remarkable acceleration effect. In order to visualize the reconstruction accuracy of the TPG-DBTS method, we plot in Fig. 2 the true solution, the reconstruction result by TPG-DBTS with noise level $\delta = 0.0001$, the curve of λ_n^δ versus n, and the error $\|c_n^\delta - c^\dagger\|_{L^2(\Omega)}$ versus n for TPG-DBTS and Landweber iteration.

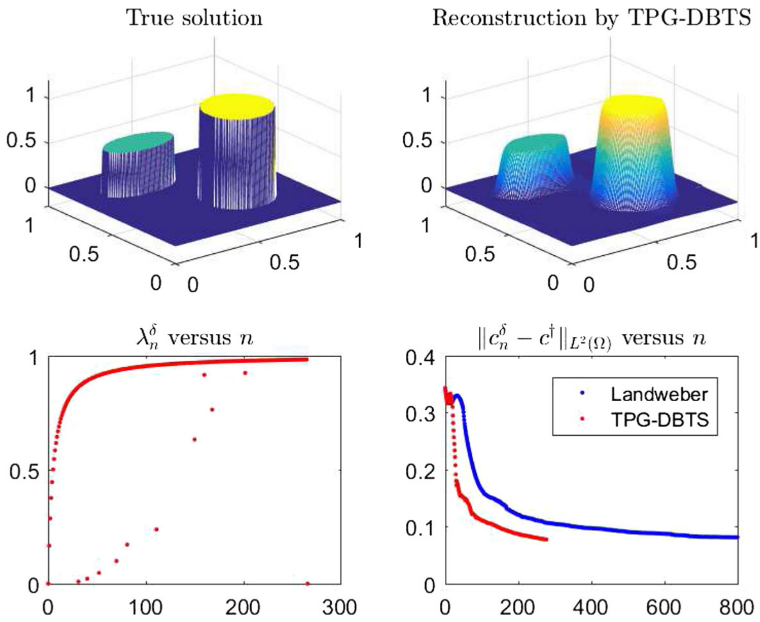

Fig. 2 The 2-dimensional elliptic parameter identification using noisy data with noise level $\delta = 0.0001$

4.3 Robin coefficient reconstruction

We consider the heat conduction process in a homogeneous solid rod located on the interval $[0, \pi]$. If the endpoints of the rod contacts with liquid media, then the convective heat transfer occurs. The temperature field of $u(x, t)$ during a time interval $[0, T]$ with a fixed time of interest $T > 0$ can be modeled by

$$\begin{cases} u_t - a^2 u_{xx} = 0, & x \in (0, \pi),\ t \in (0, T); \\ u_x(0, t) = f(t), \quad u_x(\pi, t) + \sigma(t) u(\pi, t) = \varphi(t), & t \in [0, T]; \\ u(x, 0) = u_0(x), & x \in [0, \pi]. \end{cases} \tag{4.6}$$

The function $\sigma(t) \geqslant 0$ represents the corrosion damage, which is interpreted as a Robin coefficient of energy exchange. We will assume f, φ and u_0 are all continuous. Notice that if $\sigma(t)$ is given, the problem (4.6) is a well-posed direct problem. The inverse problem of identifying the Robin coefficient $\sigma(t)$ requires additional data to be specified. We consider the reconstruction of $\sigma(t)$ from the temperature information measured at the boundary

$$u(0, t) = g(t), \quad t \in [0, T].$$

Define

$$\mathscr{D} := \{\sigma \in L^2[0, T], 0 < \sigma_- \leqslant \sigma \leqslant \sigma_+,\ a.e. \text{ in } [0, T]\}, \tag{4.7}$$

and define the nonlinear operator $F : \sigma \in \mathscr{D} \to u[\sigma](0,t) \in L^2[0,T]$, where $u[\sigma]$ denotes the unique solution of (4.6). Then the above Robin coefficient inversion problem reduces to solving the equation $F(\sigma) = g$. We refer to [30] for the well-posedness of F and the uniqueness of the inverse problem in the L^2 sense. By the standard theory of parabolic equation, one can show that F is Fréchet differentiable in the sense that

$$\|F(\sigma+h) - F(\sigma) - F'(\sigma)h\|_{L^2(0,T)} = o(\|h\|_{L^2(0,T)})$$

for all $\sigma, \sigma + h \in \mathscr{D}$, where $[F'(\sigma)h](t) = w(0,t)$ and w is the unique solution of

$$\begin{cases} w_t - a^2 w_{xx} = 0, & x \in (0,\pi),\ t \in (0,T); \\ w_x(0,t) = 0,\quad w_x(\pi,t) + \sigma(t)w(\pi,t) = -h(t)u[\sigma](\pi,t), & t \in [0,T]; \\ w(x,0) = 0, & x \in [0,\pi]. \end{cases}$$

In addition, the adjoint of the Fréchet derivative is given by

$$[F'(\sigma)^*\zeta](t) = u[\sigma](\pi,t)v(\pi,\tau),$$

where $v(x,t)$ solves the adjoint system

$$\begin{cases} -v_t - a^2 v_{xx} = 0, & x \in (0,\pi),\ t \in (0,T); \\ v_x(0,t) = \zeta(t),\quad v_x(\pi,t) + \sigma(t)v(\pi,t) = 0, & t \in [0,T]; \\ v(x,T) = 0, & x \in [0,\pi]. \end{cases}$$

In our numerical simulations, we take $a = 5$, $T = 1$, and assume the sought Robin coefficient is

$$\sigma^\dagger(t) = \begin{cases} 1.5, & 0 \leqslant t \leqslant 0.1563, \\ 2, & 0.1563 < t \leqslant 0.3125, \\ 1.2, & 0.3125 < t \leqslant 0.5469, \\ 2.5, & 0.5469 < t \leqslant 0.6250, \\ 1.8, & 0.6250 < t \leqslant 0.7813, \\ 1, & 0.7813 < t \leqslant 1. \end{cases}$$

We also assume that the exact solution of the forward problem (4.6) with $\sigma = \sigma^\dagger$ is

$$u(x,t) = e^{-a^2 t}\sin x + x^2 + 2a^2 t \tag{4.8}$$

through which we can obtain the expression of $(f(t), u_0(x), \varphi(t))$ and the inversion input $g(t) := u(0,t)$. We add random Gaussian noise on g to produce noisy data g^δ satisfying $\|g^\delta - g\|_{L^2(0,T)} \leqslant \delta$ with various noise level $\delta > 0$. We will use g^δ to reconstruct $\sigma^\dagger$. In order to capture the feature of the sought Robin coefficient, we take Θ to be the form (4.2) with $\beta = 1$. We will use the initial guess $\xi_0 = 0$ to carry out the computation. The parameters $\bar{\mu}_0$ and $\bar{\mu}_1$ in the definition of μ_n^δ are taken to be $\bar{\mu}_0 = 1.8(1 - 1/\tau)/\beta$ and $\bar{\mu}_1 = 20000$. For implementing TPG-DBTS method with

Table 3 Numerical results for Robin coefficient reconstruction ($\beta = 1$, $\tau = 1.05$)

δ	Method	n_δ	CPU time (s)	$\|\sigma_n^\delta - \sigma^\dagger\|_{L^2}$
0.1	Landweber	242	1.18	0.120974
	Nesterov	76	0.56	0.119851
	TPG-DBTS	76	1.03	0.119851
0.01	Landweber	1431	5.33	0.042751
	Nesterov	240	1.14	0.038788
	TPG-DBTS	223	2.30	0.040195
0.001	Landweber	12,033	42.14	0.007127
	Nesterov	706	2.76	0.002754
	TPG-DBTS	763	7.20	0.0023295
0.0001	Landweber	31,021	109.85	0.000702
	Nesterov	959	3.65	0.000236
	TPG-DBTS	1866	15.67	1.7341e−05

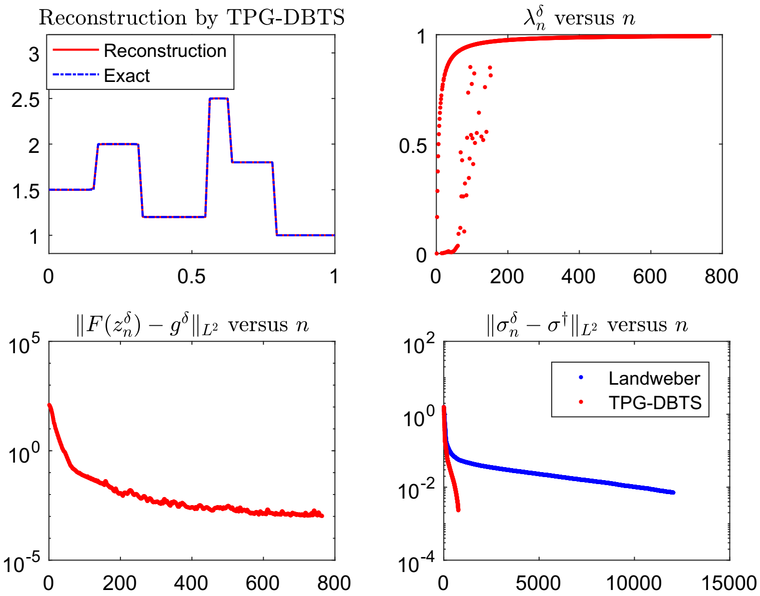

Fig. 3 Robin coefficient reconstruction using noisy data with noise level $\delta = 0.001$

λ_n^δ chosen by Algorithm 1, we take $j_{\max} = 2$, $\alpha = 5$, $\gamma_0 = 10$ in (3.24), $\gamma_1 = 1$; we also choose the function $q : \mathbb{N} \to \mathbb{N}$ by $q(m) = m^{-1.1}$. For implementing the Nesterov acceleration of Landweber iteration, we take $\lambda_n^\delta = n/(n + \alpha)$ with $\alpha = 5$. During the computation, the initial-boundary value problems for parabolic equation are transformed into integral equations by the potential theory [24] and then solved by a boundary element method by dividing $[0, T]$ into $N = 64$ subintervals of equal length. The total variation denoising problem (4.3) involved in each iteration step is solved approximately by the PDHG method after 200 iterations.

In Table 3 we report the computational results by TPG-DBTS, Landweber, and Nesterov acceleration of Landweber, using noisy data for various noise level $\delta > 0$, which clearly demonstrates the acceleration effect of TPG-DBTS and Nesterov acceleration and shows that these two methods have superior performance over Landweber iteration. In Fig. 3 we also plot the computational results by TPG-DBTS using noisy data with noise level $\delta = 0.001$. We note that the combination parameter λ_n^δ produced by TPG-DBTS may be different from $n/(n+\alpha)$ for some n, but eventually λ_n^δ becomes the same as the combination parameter $n/(n+\alpha)$ in Nesterov acceleration.

Acknowledgements The work of M. Zhong is supported by the National Natural Science Foundation of China (Nos. 11871149, 11671082) and and supported by Zhishan Youth Scholar Program of SEU. The work of W. Wang is partially supported by the National Natural Science Foundation of China (No. 11871180) and Natural Science Foundation of Zhejiang Province (No. LY19A010009). The work of Q Jin is partially supported by the Future Fellowship of the Australian Research Council (FT170100231).

References

1. Attouch, H., Peypouquet, J.: The rate of convergence of Nesterov accelerated forward–backward method is actually faster than $O(1/k^2)$. SIAM J. Optim. **26**, 1824–1834 (2016)
2. Beck, A., Teboulle, M.: A fast iterative shrinkage–thresholding algorithm for linear inverse problems. SIAM J. Imaging Sci. **2**, 183–202 (2009)
3. Beck, A., Teboulle, M.: Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems. IEEE Trans. Image Process. **18**, 2419–2434 (2009)
4. Bot, R., Hein, T.: Iterative regularization with a geeral penalty term|theory and applications to L^1 and TV regularization. Inverse Probl. **28**, 104010 (2012)
5. Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J.: Distributed optimization and statistical learning via the alternating direction method of multipliers. Found. Trends® Mach. Learn. **3**, 1–122 (2011)
6. Engl, H.W., Hanke, M., Neubauer, A.: Regularization of Inverse Problems. Kluwer, Dordrecht (1996)
7. Hackbusch, W.: Iterative Solution of Large Sparse Systems of Equations, Second Edition, Applied Mathematical Sciences, vol. 95. Springer, Berlin (2016)
8. Hanke, M.: Accelerated Landweber iterations for the solution of ill-posed equations. Numer. Math. **60**, 341–373 (1991)
9. Hanke, M.: A regularizing Levenberg–Marquardt scheme with applications to inverse groundwater filtration problems. Inverse Probl. **13**, 79–95 (1997)
10. Hanke, M., Neubauer, A., Scherzer, O.: A convergence analysis of the Landweber iteration for nonlinear ill-posed problems. Numer. Math. **72**, 21–37 (1995)
11. Hansen, P.C., Saxild-Hansen, M.: AIR tools—a MATLAB package of algebraic iterative reconstruction methods. J. Comput. Appl. Math. **236**, 2167–2178 (2012)
12. Hegland, M., Jin, Q., Wang, W.: Accelerated Landweber iteration with convex penalty for linear inverse problems in Banach spaces. Appl. Anal. **94**, 524–547 (2015)
13. Hein, T., Kazimierski, K.S.: Accelerated Landweber iteration in Banach spaces. Inverse Probl. **26**, 1037–1050 (2010)
14. Hubmer, S., Ramlau, R.: Convergence analysis of a two-point gradient method for nonlinear ill-posed problems. Inverse Probl. **33**, 095004 (2017)
15. Hubmer, S., Ramlau, R.: Nesterov's accelerated gradient method for nonlinear ill-posed problems with a locally convex residual functional. Inverse Probl. **34**, 095003 (2018)
16. Jin, Q.: Inexact Newton–Landweber iteration for solving nonlinear inverse problems in Banach spaces. Inverse Probl. **28**, 065002 (2012)
17. Jin, Q.: Landweber–Kaczmarz method in Banach spaces with inexact inner solvers. Inverse Probl. **32**, 104005 (2016)
18. Jin, Q., Wang, W.: Landweber iteration of Kaczmarz type with general non-smooth convex penalty functionals. Inverse Probl. **29**, 085011 (2013)

19. Jin, Q., Yang, H.: Levenberg–Marquardt method in Banach spaces with general convex regularization terms. Numer. Math. **133**, 655–684 (2016)
20. 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)
21. Jin, Q., Zhong, M.: Nonstationary iterated Tikhonov regularization in Banach spaces with uniformly convex penalty terms. Numer. Math. **127**, 485–513 (2014)
22. Kaltenbacher, B.: A convergence rates result for an iteratively regularized Gauss–Newton Halley method in Banach space. Inverse Probl. **31**, 015007 (2015)
23. 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 (2009)
24. Kress, R.: Linear Integral Equations. Springer, Berlin (1989)
25. Natterer, F.: The Mathematics of Computerized Tomography. SIAM, Philadelphia (2001)
26. Nesterov, Y.: A method of solving a convex programming problem with convergence rate $O(1/k^2)$. Sov. Math. Dokl. **27**, 372–376 (1983)
27. Schöpfer, F., Louis, A.K., Schuster, T.: Nonlinear iterative methods for linear ill-posed problems in Banach spaces. Inverse Probl. **22**, 311–329 (2006)
28. Schöpfer, F., Schuster, T.: Fast regularizing sequential subspace optimization in Banach spaces. Inverse Probl. **25**, 015013 (2009)
29. Wang, J., Wang, W., Han, B.: An iteration regularizaion method with general convex penalty for nonlinear inverse problems in Banach spaces. J. Comput. Appl. Math. **361**, 472–486 (2019)
30. Wang, Y.C., Liu, J.J.: Identification of non-smooth boundary heat dissipation by partial boundary data. Appl. Math. Lett. **69**, 42–48 (2017)
31. Zălinscu, C.: Convex Analysis in General Vector Spaces. World Scientific Publishing Co., Inc., River Edge (2002)
32. Zhong, M., Wang, W.: A regularizing multilevel approach for nonlinear inverse problems. Appl. Numer. Math. **32**, 297–331 (2019)
33. Zhu, M., Chan, T.F.: An efficient primaldual hybrid gradient algorithm for total variation image restoration. CAM Report 08-34, UCLA (2008)