



Hanke–Raus rule for Landweber iteration in Banach spaces

Rommel R. Real¹

Received: 15 December 2022 / Revised: 16 October 2023 / Accepted: 26 October 2023 /
Published online: 18 December 2023

© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2023

Abstract

We consider the Landweber iteration for solving linear as well as nonlinear inverse problems in Banach spaces. Based on the discrepancy principle, we propose a heuristic parameter choice rule for choosing the regularization parameter which does not require the information on the noise level, so it is purely data-driven. According to a famous veto, convergence in the worst-case scenario cannot be expected in general. However, by imposing certain conditions on the noisy data, we establish a new convergence result which, in addition, requires neither the Gâteaux differentiability of the forward operator nor the reflexivity of the image space. Therefore, we also expand the applied range of the Landweber iteration to cover non-smooth ill-posed inverse problems and to handle the situation that the data is contaminated by various types of noise. Numerical simulations are also reported.

Mathematics Subject Classification 65J20 Numerical solutions of ill-posed problems in abstract spaces; regularization

1 Introduction

Consider ill-posed problems of the form

$$F(x) = y^\dagger \quad (1.1)$$

where $F : \mathcal{D}(F) \subset \mathcal{X} \rightarrow \mathcal{Y}$, is a linear or a nonlinear operator, with domain $\mathcal{D}(F)$. We assume (1.1) has a solution. To find a solution of (1.1) with specific properties, we may use a proper, lower semi-continuous, uniformly convex function $\mathcal{R} : \mathcal{X} \rightarrow (-\infty, \infty]$.

✉ Rommel R. Real
rrreal@up.edu.ph

¹ Department of Mathematics, Physics, and Computer Science, University of the Philippines Mindanao, Davao City 8022, Philippines

Let $\partial\mathcal{R}$ denote the subdifferential of \mathcal{R} , i.e.

$$\partial\mathcal{R}(x) := \left\{ \xi \in \mathcal{X}^* : \mathcal{R}(\bar{x}) \geq \mathcal{R}(x) + \langle \xi, \bar{x} - x \rangle_{\mathcal{X}^*, \mathcal{X}} \text{ for all } \bar{x} \in \mathcal{X} \right\}$$

for all x in \mathcal{X} , with \mathcal{X}^* denoting the dual space of a Banach space \mathcal{X} and $\langle \cdot, \cdot \rangle_{\mathcal{X}^*, \mathcal{X}}$ the duality pairing between \mathcal{X}^* and \mathcal{X} . Also we denote by $\|\cdot\|_{\mathcal{X}}$ the norm on \mathcal{X} . We use

$$D_{\xi}\mathcal{R}(\bar{x}, x) := \mathcal{R}(\bar{x}) - \mathcal{R}(x) - \langle \xi, \bar{x} - x \rangle_{\mathcal{X}^*, \mathcal{X}}, \quad \bar{x} \in \mathcal{X}$$

to denote the Bregman distance induced by \mathcal{R} at x in the direction ξ . By picking $x_0 \in \mathcal{D}(\partial\mathcal{R})$ and $\xi_0 \in \partial\mathcal{R}(x_0)$ as initial guesses, we define x^\dagger to be a solution of (1.1) with the property

$$D_{\xi_0}\mathcal{R}(x^\dagger, x_0) := \min_{x \in \mathcal{D}(F)} \left\{ D_{\xi_0}\mathcal{R}(x, x_0) : F(x) = y^\dagger \right\}. \tag{1.2}$$

The exact data y^\dagger is either inaccessible or not precisely available. Instead, a noisy measurement $y^\delta \in \mathcal{Y}$ is available that satisfies

$$\|y^\delta - y^\dagger\|_{\mathcal{Y}} \leq \delta, \tag{1.3}$$

where δ is the noise level. Then the Landweber iteration in Banach spaces has the form

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

where $\xi_0^\delta = \xi_0$, $x_0^\delta = x_0$, the step size is given by

$$\mu_n^\delta := \tilde{\mu}_n^\delta \|F(x_n^\delta) - y^\delta\|_{\mathcal{Y}}^{p-r}, \text{ where } \tilde{\mu}_n^\delta := \min \left\{ \frac{\mu_0 \|F(x_n^\delta) - y^\delta\|_{\mathcal{Y}}^{p(r-1)}}{\|L(x_n^\delta)^* j_r^{\mathcal{Y}}(F(x_n^\delta) - y^\delta)\|_{\mathcal{X}^*}^p}, \mu_1 \right\}, \tag{1.5}$$

for some positive constants μ_0 and μ_1 , p is a positive constant depending on the convexity of \mathcal{R} , $\{L(x) : x \in \mathcal{D}(F)\}$ is a family of bounded linear operators from \mathcal{X} to \mathcal{Y} satisfying certain properties, and $j_r^{\mathcal{Y}} : \mathcal{Y} \rightarrow \mathcal{Y}^*$, with $1 \leq r < \infty$ is a selection of the (possibly) multi-valued mapping $J_r^{\mathcal{Y}} : \mathcal{Y} \rightarrow 2^{\mathcal{Y}^*}$, defined as the subdifferential of the convex function $y \mapsto \|y\|_{\mathcal{Y}}^r / r$.

The iteration must be terminated appropriately to produce a useful approximate solution. The authors in [19] used the discrepancy principle i.e., the iteration (1.4) is

terminated after $n_\delta := n(\delta, y^\delta)$ steps with

$$\|y^\delta - F(x_{n_\delta}^\delta)\|_{\mathcal{Y}} \leq \tau\delta < \|y^\delta - F(x_n^\delta)\|_{\mathcal{Y}}, \quad 0 \leq n < n_\delta, \quad (1.6)$$

where τ is an appropriately chosen constant, and the regularization property of x_n^δ as $\delta \rightarrow 0$ has been extensively studied [3, 15, 19, 20, 26]. All previous convergence analysis of Landweber iteration (under discrepancy principle) requires standard assumptions on the forward operator F , as well as the following key conditions:

- (a) the Banach space \mathcal{Y} is uniformly smooth and $1 < r < \infty$;
- (b) the mapping $x \rightarrow L(x)$ from \mathcal{X} to $\mathcal{L}(\mathcal{X}, \mathcal{Y})$ is continuous,

where $\mathcal{L}(\mathcal{X}, \mathcal{Y})$ denotes the Banach space of all bounded linear operators from the Banach space \mathcal{X} to the Banach space \mathcal{Y} . The paper [24] relaxes conditions (a) and (b) to address situations where the noisy data is contaminated by non-Gaussian noise such as the impulsive noise or the uniform distributed noise.

The discrepancy principle (1.6) requires knowledge of the noise level, which is not always available or reliable in real world applications. Overestimation or underestimation of noise level may lead to significant loss of reconstruction accuracy when using the discrepancy principle. Therefore, it is necessary to develop heuristic rules for Landweber iteration that do not use any knowledge of the noise level. Based on the discrepancy principle, we will propose a heuristic rule for Landweber iteration in the spirit of the work of Hanke and Raus [10]. Our heuristic rule determines an integer $n_* := n_*(y^\delta)$ by minimizing

$$\Theta(n, y^\delta) := (n + a) \|F(x_n^\delta) - y^\delta\|_{\mathcal{Y}}^p$$

over the iteration, where $a \geq 1$ is a fixed number. The Bakushinskii's veto [2] states that heuristic rules can not lead to convergence in the sense of worst case scenario for any regularization methods. Despite this veto, by imposing certain conditions on the noisy data, convergence results under heuristic rules are obtained for various regularization methods [8, 12, 16, 18, 22, 30]. The Landweber iteration is computationally cheap to implement, thus it is worth analyzing under heuristic rules.

In this paper we will present a convergence analysis of the Landweber iteration in Banach spaces under the Hanke–Raus heuristic rule. To do this, we first present the assumptions lifted from previous convergence analysis of the iteration, including a compactness condition. Then we discuss the Hanke–Raus rule for the iteration and the special assumption for our analysis, the noise condition. Next, we present the main result on the convergence analysis of the Landweber iteration under the Hanke–Raus rule. Finally, numerical examples are presented to illustrate the theoretical results.

2 Landweber iteration in Banach spaces

In analyzing the Landweber iteration in Banach spaces defined by (1.4)–(1.5), we will assume the following standard assumptions concerning the regularization functional F and the forward operator F , as also used in [19, 24].

Assumption 2.1 $\mathcal{R} : \mathcal{X} \rightarrow (-\infty, \infty]$ is proper, lower semi-continuous and p -convex with $p > 1$ in the sense that there is a constant $c_0 > 0$ such that

$$\mathcal{R}(t\bar{x} + (1 - t)x) + c_0t(1 - t)\|\bar{x} - x\|_{\mathcal{X}}^p \leq t\mathcal{R}(\bar{x}) + (1 - t)\mathcal{R}(x)$$

for all $\bar{x}, x \in \mathcal{X}$ and $0 \leq t \leq 1$.

Assumption 2.2 (a) F is weakly closed over $\mathcal{D}(F)$.

(b) There exist $\rho > 0, x_0 \in \mathcal{X}$ and $\xi_0 \in \partial\mathcal{R}(x_0)$ such that $B_{3\rho}(x_0) \subset \mathcal{D}(F)$ and (1.1) has a solution x^* satisfying $D_{\xi_0}\mathcal{R}(x^*, x_0) \leq c_0\rho^p$.

(c) There is a family of bounded linear operators $\{L(x) : \mathcal{X} \rightarrow \mathcal{Y}\}_{x \in B_{3\rho}(x_0)}$ such that $\|L(x)\| \leq B$ for all $x \in B_{3\rho}(x_0)$ for some constant B , and there is a constant $0 \leq \eta < 1$ such that

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

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

It can be easily checked that the p -convexity of \mathcal{R} in Assumption 2.1 implies that

$$D_{\xi}R(\bar{x}, x) \geq c_0\|\bar{x} - x\|_{\mathcal{X}}^p$$

for all $\bar{x}, x \in \mathcal{X}$ and $\xi \in \partial\mathcal{R}(x)$. Let \mathcal{R}^* denote the Legendre-Fenchel conjugate of \mathcal{R} , i.e.

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

then \mathcal{R}^* is Fréchet differentiable over \mathcal{X}^* and

$$\|\nabla\mathcal{R}^*(\bar{\xi}) - \nabla\mathcal{R}^*(\xi)\|_{\mathcal{X}^*} \leq \left(\frac{\|\bar{\xi} - \xi\|_{\mathcal{X}^*}}{2c_0} \right)^{p^* - 1},$$

where p^* is the number conjugate to p , i.e., $1/p + 1/p^* = 1$, see [33]. Using the definition of x_n^δ and the subdifferential calculus, we have $x_n^\delta = \nabla\mathcal{R}^*(\xi_n^\delta)$.

Assumption 2.2 (a) means that for any sequence $\{x_n\} \in \mathcal{D}(F)$ satisfying $x_n \rightarrow x \in \mathcal{X}$ and $F(x_n) \rightarrow y \in \mathcal{Y}$ there hold $x \in \mathcal{D}(F)$ and $F(x) = y$ (here we denote weak convergence and strong convergence by " \rightharpoonup " and " \rightarrow " respectively). Assumption 2.2 (c) is a version of the tangential cone condition, widely used in convergence analysis of iterative regularization methods for nonlinear inverse problems [9]. Notice that it is expressed in terms of a bounded linear operator $L(x)$ which does not have to be the Fréchet derivative of F . It is easy to see that Assumption 2.2 (c) implies that F is continuous over $B_{3\rho}(x_0)$.

When \mathcal{X} is a reflexive Banach space, by using the p -convexity and the weakly lower semi-continuity of \mathcal{R} together with the weakly closedness of F , x^\dagger exists and

is uniquely defined [19, Lemma 3.2]. Moreover, Assumption 2.2 (b) implies that

$$D_{\xi_0} \mathcal{R}(x^\dagger, x_0) \leq c_0 \rho^p,$$

which, together with Assumption 2.1, implies that $\|x_0 - x^\dagger\|_{\mathcal{X}} \leq \rho$.

In order to make the method more transparent, it is necessary to give more explanation on the mapping $j_r^{\mathcal{Y}} : \mathcal{Y} \rightarrow \mathcal{Y}^*$ for $1 \leq r < \infty$. We use $J_r^{\mathcal{Y}} : \mathcal{Y} \rightarrow 2^{\mathcal{Y}^*}$ to denote the subdifferential of the convex functional $y \rightarrow \|y\|_{\mathcal{Y}}^r/r$ over \mathcal{Y} . By the Asplund’s theorem [5], $J_r^{\mathcal{Y}}$ with $1 < r < \infty$ is exactly the duality mapping on \mathcal{Y} with the gauge function $t \rightarrow t^{r-1}/r$, i.e.

$$J_r^{\mathcal{Y}}(y) = \{y^* \in \mathcal{Y}^* : \|y^*\|_{\mathcal{Y}^*} = \|y\|_{\mathcal{Y}}^{r-1} \text{ and } \langle y^*, y \rangle_{\mathcal{Y}^*, \mathcal{Y}} = \|y\|_{\mathcal{Y}}^r\}. \tag{2.1}$$

While for $r = 1$, one has

$$J_1(y) = \begin{cases} \{y^* \in \mathcal{Y}^* : \|y^*\|_{\mathcal{Y}^*} = 1 \text{ and } \langle y^*, y \rangle_{\mathcal{Y}^*, \mathcal{Y}} = \|y\|_{\mathcal{Y}}\} & \text{if } y \neq 0, \\ \{y^* \in \mathcal{Y}^* : \|y^*\|_{\mathcal{Y}^*} \leq 1\} & \text{if } y = 0. \end{cases}$$

Note that $J_r^{\mathcal{Y}}$ in general is multi-valued. The mapping $j_r^{\mathcal{Y}}$ in (1.4) denotes a selection of $J_r^{\mathcal{Y}}$ which is defined to be a single-valued mapping from \mathcal{Y} to \mathcal{Y}^* with the property $j_r^{\mathcal{Y}}(y) \in J_r^{\mathcal{Y}}(y)$ for each $y \in \mathcal{Y}$. It is easily seen that

$$\langle j_r^{\mathcal{Y}}(y), y \rangle_{\mathcal{Y}^*, \mathcal{Y}} = \|y\|_{\mathcal{Y}}^r \quad \text{and} \quad \|j_r^{\mathcal{Y}}(y)\|_{\mathcal{Y}^*} \leq \|y\|_{\mathcal{Y}}^{r-1} \tag{2.2}$$

for all $y \in \mathcal{Y}$ and $1 \leq r < \infty$. Here we adopt the convention $0^0 = 1$. For a bounded linear operator $A : \mathcal{X} \rightarrow \mathcal{Y}$, we use $\mathcal{N}(A)$ to denote its null space and we also use

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

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

When F is a linear operator, the Landweber iteration (1.4)–(1.5) was analyzed in [3], in which the domain of \mathcal{R} is required to have a nonempty interior. This condition excludes many important properties of the regularization functional \mathcal{R} , so this was removed in [19], extending the method to cover nonlinear inverse problems.

In order to show that the Landweber iteration (1.4)–(1.5) under the discrepancy principle (1.6) is a regularization method for solving (1.1), besides the standard conditions specified in Assumptions 2.1 and 2.2, the convergence analysis in [19] requires the following additional conditions.

- Assumption 2.3** (a) The Banach space \mathcal{Y} is uniformly smooth and $1 < r < \infty$;
 (b) The mapping $x \rightarrow L(x)$ from \mathcal{X} to $\mathcal{L}(\mathcal{X}, \mathcal{Y})$ is continuous.

Here a Banach space \mathcal{Y} is called *uniformly smooth* in the sense that its modulus of smoothness

$$\rho_{\mathcal{Y}}(t) := \sup\{\|\bar{y} + y\|_{\mathcal{Y}} + \|\bar{y} - y\|_{\mathcal{Y}} - 2 : \|\bar{y}\|_{\mathcal{Y}} = 1 \text{ and } \|y\|_{\mathcal{Y}} \leq t\}$$

satisfies $\lim_{r \searrow 0} \rho_{\mathcal{Y}}(t)/t = 0$. The uniform smoothness of \mathcal{Y} in Assumption 2.3 (a) guarantees that the duality mapping $J_r^{\mathcal{Y}}$, for each $1 < r < \infty$, is single-valued and uniformly continuous on bounded sets [5]. Therefore, Assumption 2.3 is crucial for establishing the stability of Landweber iteration in [19, lemma 3.8].

Under the discrepancy principle, the convergence result of Landweber iteration (1.4)–(1.6) with $1 < r < \infty$ is established in [19, theorem 3.9]. However, this depends heavily on the uniform smoothness of \mathcal{Y} and the continuity of the mapping $\mathcal{X} \ni x \mapsto L(x) \in \mathcal{L}(\mathcal{X}, \mathcal{Y})$. There are situations where the noisy data is contaminated by non-Gaussian noise such as the impulsive noise or the uniform distributed noise; in such situations one may choose \mathcal{Y} to be the L^1 -space or the L^∞ -space to effectively remove the effects of noise [1, 6]. Note that both the L^1 -space and the L^∞ -space are not uniformly smooth. On the other hand, there exist ill-posed inverse problems where the forward operator F is non-smooth, i.e. not necessarily Gâteaux differentiable [7]. For such non-smooth ill-posed problems, one needs to choose $L(x)$ carefully as the replacement of the Gâteaux derivative; for instance, one may choose $L(x)$ to be a Bouligand subderivative of F at x which is defined as a limit of Fréchet derivative of F in differentiable points. The Bouligand subderivative mapping in general is not continuous unless the forward operator is Gâteaux differentiable. The paper [24] revisits the iteration (1.4)–(1.6) to provide a new convergence result that requires neither the reflexivity of \mathcal{Y} nor the continuity of the mapping $x \rightarrow L(x)$. To replace Assumption 2.3, the following compactness condition was used.

Assumption 2.4 There exists a Banach space Z such that

$$\bigcup_{x \in B_{3\rho}(x_0)} \text{Range}(L(x)^*) \subset Z$$

and Z is compactly embedded into \mathcal{X}^* . Moreover, there is a constant \hat{B} such that $\|L(x)^*\|_{\mathcal{L}(\mathcal{Y}^*, Z)} \leq \hat{B}$ for all $x \in B_{3\rho}(x_0)$.

Assumption 2.4 places a condition on the smoothing properties of $L(x)^*$ for $x \in B_{3\rho}(x_0)$ in a uniform sense. This assumption is inspired by the work [7] and is an adaptation of their assumption in Hilbert spaces to Banach space setting. Also, this assumption can be independent of the reflexivity of the Banach space \mathcal{Y} , as illustrated in [24, example 2.2] which uses a nonlinear inverse problem.

In order to show the convergence of $x_{n_\delta}^\delta$ to a solution of (1.1), it is necessary to investigate for each n the behaviour of $(\xi_n^\delta, x_n^\delta)$ as $\delta \rightarrow 0$ which is an important step. Since the mapping $x \rightarrow L(x)$ is not assumed to be continuous, and because the mapping $y \in \mathcal{Y} \rightarrow j_r^{\mathcal{Y}}(y) \in \mathcal{Y}^*$ is no longer continuous, one may not expect the convergence of $(\xi_n^\delta, x_n^\delta)$ to a single point as $\delta \rightarrow 0$. For each n , $(\xi_n^\delta, x_n^\delta)$ may have many limit points as $\delta \rightarrow 0$. By picking any one of the limit points for each n , we may form an iterative sequence $\{(\xi_n, x_n)\}$ in $\mathcal{X}^* \times \mathcal{X}$ corresponding to the noise-free case. It turns out that all these iterative sequences obey certain properties which will be specified in the following. We will use $\Gamma_{\mu_0, \mu_1}(\xi_0, x_0)$ to denote the set of all possible

sequences $\{(\xi_n, x_n)\}$ in $\mathcal{X}^* \times \mathcal{X}$ constructed from (ξ_0, x_0) by

$$\xi_{n+1} = \xi_n - \mu_n \psi_n, \quad \text{and} \quad x_{n+1} = \arg \min_{x \in \mathcal{X}} \{ \mathcal{R}(x) - \langle \xi_{n+1}, x \rangle_{\mathcal{X}^*, \mathcal{X}} \}, \quad (2.3)$$

where

$$\mu_n = \begin{cases} \min \left\{ \frac{\mu_0 \|F(x_n) - y^\dagger\|_{\mathcal{Y}}^{p(r-1)}}{\|\psi_n\|_{\mathcal{X}^*}^p}, \mu_1 \right\} \|F(x_n) - y^\dagger\|_{\mathcal{Y}}^{p-r} & \text{if } F(x_n) \neq y^\dagger \\ 0 & \text{if } F(x_n) = y^\dagger \end{cases} \quad (2.4)$$

and $\psi_n \in \mathcal{X}^*$ satisfies the properties

$$\begin{aligned} \langle \psi_n, \hat{x} - x_n \rangle_{\mathcal{X}^*, \mathcal{X}} &\leq -(1 - \eta) \|F(x_n) - y^\dagger\|_{\mathcal{Y}}^r, \\ |\langle \psi_n, \hat{x} - x \rangle_{\mathcal{X}^*, \mathcal{X}}| &\leq (1 + \eta) \|F(x_n) - y^\dagger\|_{\mathcal{Y}}^{r-1} \left(2 \|F(x_n) - y^\dagger\|_{\mathcal{Y}} + \|F(x) - y^\dagger\|_{\mathcal{Y}} \right), \end{aligned} \quad (2.5)$$

for any integer n , any $x \in B_{3\rho}(x_0)$ and any solution \hat{x} of (1.1) in $B_{3\rho}(x_0)$. As shown in [24], (2.5) hold when $\psi_n = L(x_n)^* J_r^{\mathcal{Y}}(F(x_n) - y)$, ensuring the well-definedness of (2.3) and (2.4). Indeed, for this choice of ψ_n , one may use Assumption 2.1, Assumption 2.2 and the argument in [19] to show that $x_n \in B_{3\rho}(x_0)$ for all $n \geq 0$. Hence $\Gamma_{\mu_0, \mu_1}(\xi_0, x_0)$ is non-empty.

The performance of the discrepancy principle depends heavily on the accurate knowledge of noise level. Such noise level information, however, is not always available or reliable in applications. Incorrect estimation on noise level may lead to significant loss of reconstruction accuracy when using the discrepancy principle. Therefore, it is necessary to develop heuristic rules for Landweber iteration that do not use any knowledge of the noise level in case a reliable noise level information is unavailable.

Based on modifying the discrepancy principle, we propose the following heuristic rule for Landweber iteration in Banach spaces in the spirit of [10]. The Hanke–Raus rule has been studied in variational regularization in Banach spaces [13], and has been generalized for variational regularization in general topological spaces [22].

Rule 2.1 *Let $a \geq 1$ be a fixed number and let*

$$\Theta(n, y^\delta) := (n + a) \|F(x_n^\delta) - y^\delta\|_{\mathcal{Y}}^p.$$

We define $n_ := n_*(y^\delta)$ be an integer such that*

$$n_* \in \operatorname{argmin} \{ \Theta(n, y^\delta) : 0 \leq n \leq n_\infty \},$$

where $n_\infty := n_\infty(y^\delta)$ is the largest integer such that $x_n^\delta \in \mathcal{D}(F)$ for all $0 \leq n \leq n_\infty$.

Rule 2.1 can be easily implemented. During the iteration, the value of $\Theta(n, y^\delta)$ is recorded versus n . After performing a large number of iterations, we stop and choose n_* to be the integer that minimizes $\Theta(n, y^\delta)$. Setting an upper limit in the number of iterations due to the local convergence is an issue with the nonlinear Landweber iteration [11]. This upper limit must be also large enough to prevent badly estimating n_* by the first local minimum, which is hardly the global minimum [11].

Using rule 2.1 requires the choice of a fixed number a , a scheme first introduced in [16] which uses a rule just like rule 2.1 for the augmented Lagrangian method in solving linear inverse problems. To use rule 2.1, we suggest choosing a to be suitably large to generate an accurate solution.

Using a fixed constant similar to a in rule 2.1 prevents the regularization parameter n_* from becoming too small. This was first illustrated in the numerical simulations in [16] for the augmented Lagrangian method. Meanwhile the paper [30] reported a convergence analysis of Hanke–Raus rule for non-stationary iterated Tikhonov regularization using a fixed constant $a \geq 0$. The choice of the lower bound for this fixed constant for the Hanke–Raus rule could depend on the regularization method to guarantee convergence. For the Landweber iteration, $a \geq 1$ in rule 2.1 is crucial, as illustrated later in the proof of Lemma 2.1.

With the integer $n_* := n_*(y^\delta)$ determined by Rule 2.1, we will use $x_{n_*(y^\delta)}^\delta$ as an approximate solution. A natural question is, for a family of noisy data $\{y^\delta\}$ with $y^\delta \rightarrow y$ as $\delta \rightarrow 0$, if it is possible to guarantee the convergence of $x_{n_*(y^\delta)}^\delta$ to a solution of (1.1) as $\delta \rightarrow 0$. The answer in general is no, according to the Bakushinskii's veto which states that heuristic rules can not lead to convergence in the sense of worst case scenario for any regularization method [2]. Therefore, to guarantee a convergence result, one must impose certain conditions on the noisy data. We will use the following *noise condition*.

Assumption 2.5 $\{y^\delta\}$ is a family of noisy data satisfying $0 < \|y^\delta - y^\dagger\|_Y \rightarrow 0$ as $\delta \rightarrow 0$ and there is a constant $\kappa > 0$ such that

$$\|y^\delta - F(x)\|_Y \geq \kappa \|y^\delta - y^\dagger\|_Y$$

for every y^δ and every $x \in S(y^\delta)$, where $S(y^\delta) := \{x_n^\delta : 0 \leq n \leq n_\infty\}$ generated by (1.4).

Under Assumption 2.5 it is immediate to see that Rule 2.1 defines a finite integer n_* . Indeed, if n_∞ is finite, then there is nothing to prove. So we only consider $n_\infty = \infty$. It then follows from Assumption 2.5 that

$$\Theta(n, y^\delta) = (n + a) \|F(x_n^\delta) - y^\delta\|_Y^p \geq (n + a)\kappa^p \|y^\dagger - y^\delta\|_Y^p \rightarrow \infty$$

as $n \rightarrow \infty$. This implies that there must exist a finite integer n_* achieving the minimum of $\Theta(n, y^\delta)$.

As of writing, Assumption 2.5 is a rather abstract condition and difficult to verify. However, this is the best thing that can be done for now in order to come up with

some analysis. The noise condition was introduced in the seminal paper [10] for linear problems in Hilbert spaces. Under the Hanke–Raus rule, such condition was extended in the convergence analyses for variational regularization [12, 13, 24], augmented Lagrangian method [16], and non-stationary iterated Tikhonov regularization [30]. We can only illustrate this condition via numerical examples. Meanwhile, a very comprehensive numerical study of heuristic rules, aside from the Hanke–Raus rule, for nonlinear Landweber iteration can be found in [11].

2.1 Analysis of rule 2.1

In this section, we will show that

$$\Theta(n_*(y^\delta), y^\delta) \rightarrow 0 \quad \text{as } \delta \rightarrow 0 \tag{2.6}$$

To achieve this result, we introduce an auxiliary integer \hat{n}_δ defined by the stopping rule

$$\|F(x_{\hat{n}_\delta}^\delta) - y^\delta\|_Y^p + \frac{(2^P - 1)c_0\rho^P}{2\mu_1(\hat{n}_\delta + a)^2} \leq \tau^P \|y^\delta - y\|_Y^p < \|F(x_n^\delta) - y^\delta\|_Y^p + \frac{(2^P - 1)c_0\rho^P}{2\mu_1(n + a)^2} \tag{2.7}$$

for $0 \leq n < \hat{n}_\delta$, where $\tau > 1$ is a large number (not to be confused with the one in (1.6)). This is a slight modification of the discrepancy principle (1.6) with an extra term $\frac{(2^P - 1)c_0\rho^P}{2\mu_1(n + a)^2}$. This extra term ensures that $\hat{n}_\delta \rightarrow \infty$ as $\delta \rightarrow 0$, an important step in establishing (2.6).

Lemma 2.1 *Let Assumptions 2.1 and 2.2 hold. Let μ_0 be chosen such that*

$$c_1 := 1 - \eta - \frac{1}{p^*} \left(\frac{\mu_0}{2c_0} \right)^{p^* - 1} > 0. \tag{2.8}$$

If $\tau > (1 + \eta)/c_1$, then (2.7) defines a finite integer \hat{n}_δ with $\hat{n}_\delta \rightarrow \infty$ as $\delta \rightarrow 0$. Moreover $x_n^\delta \in B_{3\rho}(x_0)$ for all $0 \leq n \leq \hat{n}_\delta$.

Proof Using induction, we show first that $x_n^\delta \in B_{3\rho}(x_0)$ for $0 \leq n \leq \hat{n}_\delta$. This is trivial for $n = 0$. Now, suppose $x_k^\delta \in B_{3\rho}(x_0)$ for all $0 \leq k \leq n$ for some $n < \hat{n}_\delta$, we will show that $x_{n+1}^\delta \in B_{3\rho}(x_0)$. Noting that we are considering the noisy case, by a similar argument in [19, lemma 3.4], we have for $0 \leq k \leq n$ that

$$\begin{aligned} & D_{\xi_{k+1}^\delta} \mathcal{R}(x^\dagger, x_{k+1}^\delta) - D_{\xi_k^\delta} \mathcal{R}(x^\dagger, x_k^\delta) \\ & \leq \frac{1}{p^*(2c_0)^{p^* - 1}} \|\xi_{k+1}^\delta - \xi_k^\delta\|_{\mathcal{X}^*}^{p^*} - \mu_k^\delta \left\langle j_r^\mathcal{Y}(F(x_k^\delta) - y^\delta), L(x_k^\delta)(x_k^\delta - x^\dagger) \right\rangle_{\mathcal{Y}^*, \mathcal{Y}} \end{aligned} \tag{2.9}$$

Based on (1.4) and (2.1), we have

$$\|\xi_{k+1}^\delta - \xi_k^\delta\|_{\mathcal{X}^*}^{p^*} = (\mu_n^\delta)^{p^*} \left\| L(x_n^\delta)^* j_r^{\mathcal{Y}}(F(x_n^\delta) - y^\delta) \right\|_{\mathcal{X}^*}^{p^*} \leq \mu_0^{p^*-1} \mu_n^\delta \|F(x_n^\delta) - y^\delta\|_{\mathcal{Y}}^r$$

Moreover, by invoking the iterative Eq. (1.4) and Assumption 2.2 (c), we have

$$\begin{aligned} & - \langle j_r^{\mathcal{Y}}(F(x_k^\delta) - y^\delta), L(x_k^\delta)(x_k^\delta - x^\dagger) \rangle_{\mathcal{Y}^*, \mathcal{Y}} \\ &= - \langle j_r^{\mathcal{Y}}(F(x_k^\delta) - y^\delta), F(x_k^\delta) - y^\delta \rangle_{\mathcal{Y}^*, \mathcal{Y}} + \langle j_r^{\mathcal{Y}}(F(x_k^\delta) - y^\delta), y^\dagger - y^\delta \rangle_{\mathcal{Y}^*, \mathcal{Y}} \\ & \quad - \langle j_r^{\mathcal{Y}}(F(x_k^\delta) - y^\delta), F(x^\dagger) - F(x_k^\delta) - L(x_k^\delta)(x^\dagger - x_k^\delta) \rangle_{\mathcal{Y}^*, \mathcal{Y}} \\ &\leq - \|F(x_k^\delta) - y^\delta\|_{\mathcal{Y}}^r + \|F(x_k^\delta) - y^\delta\|_{\mathcal{Y}}^{r-1} \|y^\dagger - y^\delta\|_{\mathcal{Y}} \\ & \quad + \eta \|F(x_k^\delta) - y^\delta\|_{\mathcal{Y}}^{r-1} \|F(x^\dagger) - F(x_k^\delta)\|_{\mathcal{Y}} \\ &\leq - \|F(x_k^\delta) - y^\delta\|_{\mathcal{Y}}^r + \|F(x_k^\delta) - y^\delta\|_{\mathcal{Y}}^{r-1} \|y^\dagger - y^\delta\|_{\mathcal{Y}} \\ & \quad + \eta \|F(x_k^\delta) - y^\delta\|_{\mathcal{Y}}^{r-1} (\|F(x^\dagger) - y^\delta\|_{\mathcal{Y}} + \|y^\delta - F(x_k^\delta)\|_{\mathcal{Y}}) \\ &= -(1 - \eta) \|F(x_k^\delta) - y^\delta\|_{\mathcal{Y}}^r + (1 + \eta) \|F(x_k^\delta) - y^\delta\|_{\mathcal{Y}}^{r-1} \|y^\dagger - y^\delta\|_{\mathcal{Y}} \end{aligned}$$

Therefore, with the definition of μ_n^δ in (1.5), we have

$$\begin{aligned} & D_{\xi_{k+1}^\delta} \mathcal{R}(x^\dagger, x_{k+1}^\delta) - D_{\xi_k^\delta} \mathcal{R}(x^\dagger, x_k^\delta) \\ & \leq -c_1 \tilde{\mu}_k^\delta \|F(x_k^\delta) - y^\delta\|_{\mathcal{Y}}^p + (1 + \eta) \tilde{\mu}_k^\delta \|F(x_k^\delta) - y^\delta\|_{\mathcal{Y}}^{p-1} \|y^\dagger - y^\delta\|_{\mathcal{Y}}. \end{aligned} \tag{2.10}$$

According to the definition of \hat{n}_δ , for $k < \hat{n}_\delta$ we have

$$\|F(x_k^\delta) - y^\delta\|_{\mathcal{Y}}^{p-1} \|y^\dagger - y^\delta\|_{\mathcal{Y}} \leq \frac{1}{\tau} \|F(x_k^\delta) - y^\delta\|_{\mathcal{Y}}^{p-1} \left(\|F(x_k^\delta) - y^\delta\|_{\mathcal{Y}}^p + \frac{(2^p - 1)c_0 \rho^p}{2\mu_1(k+a)^2} \right)^{1/p},$$

and by Young’s inequality, we further arrive at

$$\begin{aligned} & \|F(x_k^\delta) - y^\delta\|_{\mathcal{Y}}^{p-1} \|y^\dagger - y^\delta\|_{\mathcal{Y}} \\ & \leq \frac{1}{\tau} \left(\frac{1}{p} \left(\|F(x_k^\delta) - y^\delta\|_{\mathcal{Y}}^p + \frac{(2^p - 1)c_0 \rho^p}{2\mu_1(k+a)^2} \right) + \frac{1}{p^*} \|F(x_k^\delta) - y^\delta\|_{\mathcal{Y}}^p \right) \\ & \leq \frac{1}{\tau} \left(\frac{1}{p} \left(\|F(x_k^\delta) - y^\delta\|_{\mathcal{Y}}^p + \frac{(2^p - 1)c_0 \rho^p}{2\mu_1(k+a)^2} \right) + \frac{1}{p^*} \left(\|F(x_k^\delta) - y^\delta\|_{\mathcal{Y}}^p + \frac{(2^p - 1)c_0 \rho^p}{2\mu_1(k+a)^2} \right) \right) \\ & = \frac{1}{\tau} \left(\|F(x_k^\delta) - y^\delta\|_{\mathcal{Y}}^p + \frac{(2^p - 1)c_0 \rho^p}{2\mu_1(k+a)^2} \right). \end{aligned}$$

Since $\tilde{\mu}_k^\delta \leq \mu_1$, we can conclude that

$$\begin{aligned} & D_{\xi_{k+1}^\delta} \mathcal{R}(x^\dagger, x_{k+1}^\delta) - D_{\xi_k^\delta} \mathcal{R}(x^\dagger, x_k^\delta) \\ & \leq - \left(c_1 - \frac{1 + \eta}{\tau} \right) \tilde{\mu}_k^\delta \|F(x_k^\delta) - y^\delta\|_{\mathcal{Y}}^p + \frac{(1 + \eta)(2^p - 1)c_0\rho^p}{2\tau(k + a)^2}. \end{aligned}$$

Then by taking the sum from $k = 0$ to $k = n < \hat{n}_\delta$ on both sides, we further obtain

$$\begin{aligned} & D_{\xi_{n+1}^\delta} \mathcal{R}(x^\dagger, x_{n+1}^\delta) + \left(c_1 - \frac{1 + \eta}{\tau} \right) \sum_{k=0}^n \tilde{\mu}_k^\delta \|F(x_k^\delta) - y^\delta\|_{\mathcal{Y}}^p \\ & \leq D_{\xi_0} \mathcal{R}(x^\dagger, x_0) + \frac{(1 + \eta)(2^p - 1)c_0\rho^p}{2\tau} \sum_{k=0}^n \frac{1}{(k + a)^2} \\ & \leq c_0\rho^p + \frac{(2^p - 1)c_0\rho^p}{2} \sum_{k=0}^\infty \frac{1}{(k + a)^2}, \end{aligned}$$

where we used Assumption 2.2 (b) and the condition $\tau > (1 + \eta)/c_1$ which implies that $(1 + \eta)/\tau < c_1 < 1$. Noting that $a \geq 1$, and since $\sum_{k=0}^\infty \frac{1}{(k+1)^2} = \pi^2/6 < 2$, we can obtain

$$D_{\xi_{n+1}^\delta} \mathcal{R}(x^\dagger, x_{n+1}^\delta) + \left(c_1 - \frac{1 + \eta}{\tau} \right) \sum_{k=0}^n \tilde{\mu}_k^\delta \|F(x_k^\delta) - y^\delta\|_{\mathcal{Y}}^p \leq c_0(2\rho)^p. \tag{2.11}$$

By using the p -convexity of \mathcal{R} , it follows that

$$c_0 \|x_{n+1}^\delta - x^\dagger\|_{\mathcal{X}}^p \leq D_{\xi_{n+1}^\delta} \mathcal{R}(x^\dagger, x_{n+1}^\delta) \leq c_0(2\rho)^p$$

which shows that $\|x_{n+1}^\delta - x^\dagger\|_{\mathcal{X}} \leq 2\rho$. Since $\|x_0 - x^\dagger\|_{\mathcal{X}} \leq \rho$, we therefore have $x_{n+1}^\delta \in B_{3\rho}(x_0)$.

Next, we prove that \hat{n}_δ is finite. By contradiction, suppose \hat{n}_δ is infinite. Then the previous argument shows that $x_n^\delta \in B_{3\rho}(x_0)$ for all $n \geq 0$. Consequently it follows from (2.11) that

$$\left(c_1 - \frac{1 + \eta}{\tau} \right) \sum_{k=0}^n \tilde{\mu}_k^\delta \|F(x_k^\delta) - y^\delta\|_{\mathcal{Y}}^p \leq c_0(2\rho)^p.$$

for all $n \geq 0$. From the definition of $\tilde{\mu}_n^\delta$, we have

$$\tilde{\mu}_n^\delta \geq \min \{ \mu_0 B^{-p}, \mu_1 \}.$$

Then, for all $n \geq 0$ we have

$$c_3 \sum_{k=0}^n \|F(x_k^\delta) - y^\delta\|_{\mathcal{Y}}^p \leq c_0(2\rho)^p, \tag{2.12}$$

where $c_3 := \left(c_1 - \frac{1+\eta}{\tau}\right) \min\{\mu_0 B^{-p}, \mu_1\} > 0$. In addition, the definition of \hat{n}_δ tells us that

$$\|F(x_k^\delta) - y^\delta\|_{\mathcal{Y}}^p \geq \tau^p \|y^\delta - y^\dagger\|_{\mathcal{Y}}^p - \frac{(2^p - 1)c_0\rho^p}{2\mu_1(k+a)^2}$$

for all $k \geq 0$. Summing up both sides of the previous inequality, and then using (2.12), we further obtain

$$\begin{aligned} c_3 \tau^p \|y^\delta - y^\dagger\|_{\mathcal{Y}}^p (n+1) &\leq 2^p c_0 \rho^p + \frac{(2^p - 1)c_0 \rho^p}{2\mu_1} \sum_{k=0}^n \frac{1}{(k+a)^2} \\ &\leq 2^p c_0 \rho^p + \frac{\pi^2(2^p - 1)c_0 \rho^p}{12\mu_1} < \infty, \end{aligned}$$

and by taking $n \rightarrow \infty$, we obtain a contradiction. Finally, from the definition of \hat{n}_δ in (2.7),

$$\frac{(2^p - 1)c_0 \rho^p}{2\mu_1(\hat{n}_\delta + a)^2} \leq \tau^p \|y^\delta - y^\dagger\|_{\mathcal{Y}}^p \rightarrow 0 \text{ as } \delta \rightarrow 0.$$

Therefore we must have $\hat{n}_\delta \rightarrow \infty$ as $\delta \rightarrow 0$. □

Lemma 2.2 *Let Assumptions 2.1 and 2.2 hold. Let $\mu_0 > 0$ be chosen such that (2.8) holds, and let $\{y^\delta\}$ be a family of noisy data satisfying Assumption 2.5. Let $n_* := n_*(y^\delta)$ be the integer defined by rule 2.1. Then $\Theta(n_*(y^\delta), y^\delta) \rightarrow 0$ as $\delta \rightarrow 0$. Consequently, $\|F(x_{n_*(y^\delta)}^\delta) - y^\delta\|_{\mathcal{Y}} \rightarrow 0$ and $n_*(y^\delta) \|y^\delta - y^\dagger\|_{\mathcal{Y}}^p \rightarrow 0$ as $\delta \rightarrow 0$.*

Proof Let \hat{n}_δ be the integer defined by (2.7). From the proof of Lemma 2.1 we have

$$\sum_{n=0}^{\hat{n}_\delta-1} \|F(x_n^\delta) - y^\delta\|_{\mathcal{Y}}^p \leq \frac{2^p c_0 \rho^p}{c_3}$$

By the minimality of $\Theta(n_*(y^\delta), y^\delta)$, it follows that

$$\|F(x_n^\delta) - y^\delta\|_{\mathcal{Y}}^p = \frac{\Theta(n, y^\delta)}{n+a} \geq \frac{\Theta(n_*(y^\delta), y^\delta)}{n+a}.$$

Hence,

$$\Theta(n_*(y^\delta), y^\delta) \sum_{n=0}^{\hat{n}_\delta-1} \frac{1}{n+a} \leq \sum_{n=0}^{\hat{n}_\delta-1} \|F(x_n^\delta) - y^\delta\|_{\mathcal{Y}}^p \leq \frac{2^p c_0 \rho^p}{c_3}.$$

Note that

$$\sum_{n=0}^{\hat{n}_\delta-1} \frac{1}{n+a} \geq \sum_{n=0}^{\hat{n}_\delta-1} \int_n^{n+1} \frac{1}{t+a} dt = \int_0^{\hat{n}_\delta} \frac{1}{t+a} dt = \log \frac{\hat{n}_\delta + a}{a}.$$

Thus

$$\Theta(n_*(y^\delta), y^\delta) \leq \frac{2^p c_0 \rho^p}{c_3 \log(\hat{n}_\delta + a)/a}.$$

Since $\hat{n}_\delta \rightarrow \infty$ we must have $\Theta(n_*(y^\delta), y^\delta) \rightarrow 0$ as $\delta \rightarrow 0$. Note that

$$a \|F(x_{n_*(y^\delta)}^\delta) - y^\delta\|_{\mathcal{Y}}^p \leq \Theta(n_*(y^\delta), y^\delta),$$

and Assumption 2.5 implies that

$$(n_*(y^\delta) + a) \kappa^p \|y^\delta - y^\dagger\|^p \leq \Theta(n_*(y^\delta), y^\delta).$$

Hence, by the recent claim, $\|F(x_{n_*(y^\delta)}^\delta) - y^\delta\|_{\mathcal{Y}} \rightarrow 0$ and $n_*(y^\delta) \|y^\delta - y^\dagger\|_{\mathcal{Y}}^p \rightarrow 0$ as $\delta \rightarrow 0$. □

Lemma 2.3 *Let Assumptions 2.1 and 2.2 hold. Let $\mu_0 > 0$ be chosen such that (2.8) holds, and let $\{y^\delta\}$ be a family of noisy data satisfying Assumption 2.5. Let $n_* := n_*(y^\delta)$ be the integer defined by Rule 2.1. Then $x_n^\delta \in B_{3\rho}(x_0)$ for all $0 \leq n \leq n_*$ if δ is sufficiently small.*

Proof We use an induction argument. The result is trivial for $n = 0$. Now we assume that $x_n^\delta \in B_{3\rho}(x_0)$ for some $n < n_*(y^\delta)$. We will use (2.10) to prove that $x_{n+1}^\delta \in B_{3\rho}(x_0)$. By the Young’s inequality we have

$$(1 + \eta) \|F(x_n^\delta) - y^\delta\|_{\mathcal{Y}}^{p-1} \|y^\delta - y^\dagger\|_{\mathcal{Y}} \leq \frac{c_1}{p^*} \|F(x_n^\delta) - y^\delta\|_{\mathcal{Y}}^p + \frac{(1 + \eta)^p}{pc_1^{p-1}} \|y^\delta - y^\dagger\|_{\mathcal{Y}}^p.$$

Combining this with (2.10), we have

$$D_{\xi_{n+1}^\delta} \mathcal{R}(x^\dagger, x_{n+1}^\delta) - D_{\xi_n^\delta} \mathcal{R}(x^\dagger, x_n^\delta) \leq -\frac{c_1}{p} \tilde{\mu}_n^\delta \|F(x_n^\delta) - y^\delta\|_{\mathcal{Y}}^p + c_4 \|y^\delta - y^\dagger\|_{\mathcal{Y}}^p,$$

where $c_4 := (1 + \eta)^p \mu_1 / (pc_1^{p-1})$. Therefore

$$D_{\xi_{n+1}^\delta} \mathcal{R}(x^\dagger, x_{n+1}^\delta) + \frac{c_1}{p} \sum_{k=0}^n \tilde{\mu}_k^\delta \|F(x_k^\delta) - y^\delta\|_Y^p \leq D_{\xi_0} \mathcal{R}(x^\dagger, x_0) + c_4(n + 1)\|y^\delta - y^\dagger\|_Y^p \leq c_0\rho^p + c_4n_*(y^\delta)\|y^\delta - y^\dagger\|_Y^p$$

By Lemma 2.2, we can guarantee that

$$c_4n_*(y^\delta)\|y^\delta - y^\dagger\|_Y^p \leq (2^p - 1)c_0\rho^p$$

for sufficiently small δ . Then

$$D_{\xi_{n+1}^\delta} \mathcal{R}(x^\dagger, x_{n+1}^\delta) \leq c_0\rho^p + (2^p - 1)c_0\rho^p = c_0(2\rho)^p.$$

By the p -convexity of \mathcal{R} we have $c_0\|x_{n+1}^\delta - x^\dagger\|_{\mathcal{X}}^p \leq c_0(2\rho)^p$ which show that $\|x_{n+1}^\delta - x^\dagger\|_{\mathcal{X}} \leq 2\rho$. Since $\|x_0 - x^\dagger\|_{\mathcal{X}} \leq \rho$, we therefore have $x_{n+1}^\delta \in B_{3\rho}(x_0)$. \square

We lift a previous result regarding the noise-free algorithm (2.3)–(2.5), which we will use later.

Lemma 2.4 *Let Assumptions 2.1 and 2.2 hold. Let $\mu_0 > 0$ be chosen such that (2.8) holds. The for any sequence $\{(\xi_n, x_n)\} \in \Gamma_{\mu_0, \mu_1}(\xi_0, x_0)$ there exists a solution x_* of (1.1) such that $D_{\xi_n} \mathcal{R}(x_*, x_n) \rightarrow 0$ as $n \rightarrow \infty$. If, in addition, $\xi_{n+1} - \xi_n \in \mathcal{N}(L(x^\dagger))^\perp$ for all n , then $x_* = x^\dagger$.*

Proof This is [24, Lemma 2.5]. \square

2.2 Main result

Now we are about to prove the main result of this paper. We need two stability results based also on Assumption 2.3 and Assumption 2.4. These stability results will link the method (1.4) to the noise-free algorithm (2.3)–(2.5) so that lemma 2.4 can be applied.

Lemma 2.5 *Let Assumptions 2.1 and 2.2 hold. Let $\mu_0 > 0$ be chosen such that (2.8) holds, and let $\{y^\delta\}$ be a family of noisy data satisfying Assumption 2.5. Let $n_* := n_*(y^\delta)$ be the integer defined by Rule 2.1.*

(a) *If Assumption 2.3 holds and $1 < r < \infty$, then there is a sequence $\{(\xi_n, x_n)\} \in \Gamma_{\mu_0, \mu_1}(\xi_0, x_0)$ such that*

$$x_n^\delta \rightarrow x_n \quad \text{and} \quad \xi_n^\delta \rightarrow \xi_n \quad \text{as } \delta \rightarrow 0$$

for all $0 \leq n \leq \liminf_{\delta \rightarrow 0} n_(y^\delta)$.*

(a) If Assumption 2.4 holds and $1 \leq r < \infty$, then for any subsequence $\{y^{\delta_l}\}$, with $\delta_l \rightarrow 0$ as $l \rightarrow \infty$, of $\{y^\delta\}$, by taking a subsequence if necessary, there is a sequence $\{(\xi_n, x_n)\} \in \Gamma_{\mu_0, \mu_1}(\xi_0, x_0)$ such that

$$x_n^{\delta_l} \rightarrow x_n \quad \text{and} \quad \xi_n^{\delta_l} \rightarrow \xi_n \quad \text{as } l \rightarrow \infty$$

for all $0 \leq n \leq \liminf_{l \rightarrow \infty} n_*(y^{\delta_l})$.

If in addition $\mathcal{N}(L(x^\dagger)) \subset \mathcal{N}(L(x))$ for all $x \in B_{3\rho}(x_0)$, then there also holds $\xi_{n+1} - \xi_n \in \mathcal{N}(L(x^\dagger))^\perp$ for all n .

Proof (a) The proof is similar to that of [19, lemma 3.8]. Note that since Assumption 2.3 holds, the mappings $y \mapsto j_r^{\mathcal{Y}}(y)$ and $x \mapsto L(x)$ are continuous [5].

(b) Since Assumption 2.4 holds, the proof follows from that of [24, lemma 2.3] with $N := \liminf_{l \rightarrow \infty} n_*(y^{\delta_l})$. □

Now we are ready to prove the main result of this paper concerning the method (1.4)–(1.5).

Theorem 2.6 *Let Assumption 2.1 and 2.2 hold, and let μ_0 be chosen such that (2.8) holds. Let $\{y^\delta\}$ be a sequence of noisy data satisfying assumption 2.5 and let $n_*(y^\delta)$ be the integer determined by Rule 2.1.*

(a) *If Assumption 2.3 holds and $1 < r < \infty$, then there is a solution x_* of (1.1) such that*

$$\left\| x_{n_*(y^\delta)}^\delta - x_* \right\|_{\mathcal{X}} \rightarrow 0 \quad \text{and} \quad D_{\xi_{n_*(y^\delta)}^\delta} \mathcal{R}(x_*, x_{n_*(y^\delta)}^\delta) \rightarrow 0$$

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

(b) *If Assumption 2.4 holds and $1 \leq r < \infty$, then for any subsequence $\{y^{\delta_l}\}$, with $\delta_l \rightarrow 0$ as $l \rightarrow \infty$, of $\{y^\delta\}$, by taking a subsequence of $\{y^{\delta_l}\}$ if necessary, there hold*

$$\left\| x_{n_*(y^{\delta_l})}^{\delta_l} - x_* \right\|_{\mathcal{X}} \rightarrow 0 \quad \text{and} \quad D_{\xi_{n_*(y^{\delta_l})}^{\delta_l}} \mathcal{R}(x_*, x_{n_*(y^{\delta_l})}^{\delta_l}) \rightarrow 0$$

as $l \rightarrow \infty$ for some solution x_* of (1.1) in $B_{3\rho}(x_0)$. If in addition $\mathcal{N}(L(x^\dagger)) \subset \mathcal{N}(L(x))$ for all $x \in B_{3\rho}(x_0)$, then

$$\left\| x_{n_*(y^\delta)}^\delta - x^\dagger \right\|_{\mathcal{X}} \rightarrow 0 \quad \text{and} \quad D_{\xi_{n_*(y^\delta)}^\delta} \mathcal{R}(x^\dagger, x_{n_*(y^\delta)}^\delta) \rightarrow 0$$

as $\delta \rightarrow 0$.

Proof We will only prove (b) since (a) can be proved similarly. Let $\{y^{\delta_l}\}$ be a subsequence of $\{y^\delta\}$, and let $N := \liminf_{l \rightarrow \infty} n_*(y^{\delta_l})$. By taking a subsequence of $\{y^{\delta_l}\}$

if necessary, we may assume $N = \lim_{l \rightarrow \infty} n_*(y^{\delta_l})$ and, according to Lemma 2.5, we can find a sequence $\{(\xi_n, x_n)\} \in \Gamma_{\mu_0, \mu_1}(\xi_0, x_0)$ such that

$$\xi_n^{\delta_l} \rightarrow \xi_n \quad \text{and} \quad x_n^{\delta_l} \rightarrow x_n \quad \text{as } l \rightarrow \infty \tag{2.13}$$

for all $0 \leq n \leq N$. Let x_* be the limit of $\{x_n\}$ which is a solution of (1.1) in $B_{3\rho}(x_0)$. We show that

$$\lim_{l \rightarrow \infty} D_{\xi_{n_*(y^{\delta_l})}^{\delta_l}} \mathcal{R}(x_*, x_{n_*(y^{\delta_l})}^{\delta_l}) = 0. \tag{2.14}$$

If $N < \infty$, then (2.14) can be similarly proven using the argument for the corresponding case in the proof of [24, theorem 2.6] since there requires only $\|F(x_{n_*(y^{\delta_l})}^{\delta_l}) - y^{\delta_l}\|_{\mathcal{Y}} \rightarrow 0$ as $l \rightarrow \infty$, which is guaranteed by Lemma 2.2. If $N = \infty$ then for any fixed integer $n \geq 1$ we have $n_*(y^{\delta_l}) > n$ for large l . According to the proof of Lemma 2.3 we have

$$D_{\xi_{k+1}^{\delta_l}} \mathcal{R}(x_*, x_{k+1}^{\delta_l}) - D_{\xi_k^{\delta_l}} \mathcal{R}(x_*, x_k^{\delta_l}) \leq C \|y^{\delta_l} - y^\dagger\|_{\mathcal{Y}}^p$$

for $0 \leq k < n_*(y^{\delta_l})$, where C is a positive constant independent of l . Choose n such that $0 \leq n < n_*(y^{\delta_l})$. By taking the sum both sides from $k = n$ to $k = n_*(y^{\delta_l}) - 1$, the previous inequality implies that

$$\begin{aligned} D_{\xi_{n_*(y^{\delta_l})}^{\delta_l}} \mathcal{R}(x_*, x_{n_*(y^{\delta_l})}^{\delta_l}) &\leq D_{\xi_n^{\delta_l}} \mathcal{R}(x_*, x_n^{\delta_l}) + C(n_*(y^{\delta_l}) - n + 1) \|y^{\delta_l} - y^\dagger\|_{\mathcal{Y}}^p \\ &\leq D_{\xi_n^{\delta_l}} \mathcal{R}(x_*, x_n^{\delta_l}) + Cn_*(y^{\delta_l}) \|y^{\delta_l} - y^\dagger\|_{\mathcal{Y}}^p. \end{aligned}$$

We follow the argument for the corresponding case in the proof of [24, theorem 2.6]. From the proof of Lemma 2.3, the monotonicity of $\{D_{\xi_n^{\delta_l}} \mathcal{R}(x_*, x_n^{\delta_l})\}$ also holds. Using this, (2.13), and the lower semi-continuity of \mathcal{R} , we can obtain

$$\begin{aligned} 0 &\leq \limsup_{l \rightarrow \infty} D_{\xi_{n_*(y^{\delta_l})}^{\delta_l}} \mathcal{R}(x_*, x_{n_*(y^{\delta_l})}^{\delta_l}) \leq \limsup_{l \rightarrow \infty} D_{\xi_n^{\delta_l}} \mathcal{R}(x_*, x_n^{\delta_l}) \\ &\leq \mathcal{R}(x_*) - \liminf_{l \rightarrow \infty} \mathcal{R}(x_n^{\delta_l}) - \lim_{l \rightarrow \infty} \langle \xi_n^{\delta_l}, x_* - x_n^{\delta_l} \rangle \\ &\leq \mathcal{R}(x_*) - \mathcal{R}(x_n) - \langle \xi_n, x_* - x_n \rangle \\ &\leq D_{\xi_n} \mathcal{R}(x_*, x_n). \end{aligned}$$

Since Lemma 2.4 implies that $D_{\xi_n} \mathcal{R}(x_*, x_n) \rightarrow 0$ as $n \rightarrow \infty$, we can conclude (2.14) again. □

For the Eq. (1.1) with a bounded linear operator $F : \mathcal{X} \rightarrow \mathcal{Y}$, Assumption 2.4 can be replaced by the compactness of F . This leads to the following convergence result.

Corollary 2.7 Consider the Eq. (1.1) where $F : \mathcal{X} \rightarrow \mathcal{Y}$ is a compact bounded linear operator. Let \mathcal{R} satisfy Assumption 2.1 and let $\mu_0 > 0$ be chosen such that

$$1 - \frac{1}{p^*} \left(\frac{\mu_0}{2c_0} \right)^{p^*-1} > 0.$$

Then for the Landweber iteration (1.4) with the integer $n_*(y^\delta)$ determined by rule 2.1 there hold

$$\left\| x_{n_*(y^\delta)}^\delta - x^\dagger \right\|_{\mathcal{X}} \rightarrow 0 \quad \text{and} \quad D_{\xi_{n_*(y^\delta)}^\delta} \mathcal{R}(x^\dagger, x_{n_*(y^\delta)}^\delta) \rightarrow 0$$

as $\delta \rightarrow 0$.

3 Numerical examples

Example 3.1 Now we consider an example with a nonsmooth forward operator. To start, consider an open bounded subset $\Omega \subset \mathbb{R}^d$, $d \in \{1, 2\}$, with a Lipschitz boundary $\partial\Omega$, and consider the nonsmooth semilinear equation

$$\begin{cases} -\Delta y + y^+ = u & \text{in } \Omega, \\ y = 0 & \text{on } \partial\Omega, \end{cases} \tag{3.1}$$

with $u \in L^2(\Omega)$ and $y^+(x) := \max(y(x), 0)$ for almost every $x \in \Omega$. Equation (3.1) arises in a number of applications, such as modelling the deflection of a stretched thin membrane partially covered by water (see [21]). It also appears in free boundary problems for a confined plasma; see [21, 23, 28] for some examples. For each $u \in L^2(\Omega)$, a unique solution y_u in $H_0^1(\Omega) \cap C(\Omega)$ exists [29, theorem 4.7], so we consider the inverse problem of estimating the source term u from noisy measurements of y_u .

Define the forward operator $F : L^2 \rightarrow H_0^1(\Omega) \cap C(\Omega)$ where $y_u = F(u)$ for $u \in L^2(\Omega)$. As shown in [4, proposition 2.1, theorem 2.2], F is globally Lipschitz continuous, and has a directional derivative $F'(u; h)$ in the direction $h \in L^2(\Omega)$ given by $v \in H_0^1(\Omega)$ which solves

$$\begin{cases} -\Delta v + \mathbb{1}_{\{y_u=0\}} v^+ + \mathbb{1}_{\{y_u>0\}} v = h & \text{in } \Omega, \\ v = 0 & \text{on } \partial\Omega. \end{cases}$$

The operator F is Gâteaux differentiable in $u \in L^2(\Omega)$ if and only if the Lebesgue measure of the set $\{u : y_u = 0\}$ is zero [7, proposition 3.4]. Hence, in general F not is Gâteaux differentiable. Moreover, computing the directional derivative of F could be difficult, and a more convenient alternative is using the Bouligand subdifferential [4, proposition 3.16]. Given $u \in L^2(\Omega)$, a specific Bouligand subderivative of F is given by the solution operator $G_u : L^2(\Omega) \rightarrow H_0^1(\Omega) \hookrightarrow L^2(\Omega)$ which maps $h \in L^2(\Omega)$

to the unique solution $v \in H_0^1(\Omega)$ of

$$\begin{cases} -\Delta v + \mathbb{1}_{\{y_u > 0\}} v = h & \text{in } \Omega, \\ v = 0 & \text{on } \partial\Omega, \end{cases} \tag{3.2}$$

where $y_u = F(u)$. In fact G_u is uniformly bounded for all $u \in L^2(\Omega)$ [7, lemmas 3.7 and 3.9], and consequently satisfies the generalized tangential cone condition. Hence, Assumption 2.2(c) holds for G_u .

Now we can use Landweber iteration (1.4) to solve the inverse problem of recovering $u \in L^2(\Omega)$ from $y_u = F(u)$ satisfying (3.1). Since F is a mapping between Hilbert spaces, we have $r = 2$ and $J_2 \equiv I$. Since the convex quadratic penalty $\mathcal{R}(u) = \|u\|_{L^2(\Omega)} / 2$ is 2-convex, set $p = 2$. By choosing $L = G_u$, the iterative equation in (1.4) reduces to the Bouligand-Landweber iteration [7] given by

$$u_{n+1}^\delta = u_n^\delta - \mu_n^\delta G_{u_n^\delta}^* (F(u_n^\delta) - y^\delta). \tag{3.3}$$

where u_n^δ solves (3.1) given $y = y_n^\delta$, and μ_n^δ is a given stepsize. Under the discrepancy principle (1.6), the authors in [7] established the convergence of (3.3) using a constant stepsize

$$\mu_n^\delta = \frac{2 - 2\eta}{\bar{L}^2}, \text{ where } \bar{L} = 5 \cdot 10^{-2}, \tag{3.4}$$

where $\eta = 0.1$ is a tangential constant estimate in Assumption 2.2(c).

For the computation of the nonsmooth forward operator F , we discretize the non-smooth semilinear elliptic problem (3.1) using standard continuous piecewise linear finite elements, and then solve the resulting non-smooth nonlinear equation using a semi-smooth Newton method. The same discretization scheme is also done for computing the Bouligand subdifferential G_u in terms of (3.1). For more details, we refer the reader to [7] and the reference therein.

We consider the two-dimensional problem with $\Omega = (0, 1) \times (0, 1) \subset \mathbb{R}^2$ and use a uniform triangular Friedrichs-Keller triangulation with 128^2 vertices. The discretization scheme for computing the operator F and G_u involve standard continuous piecewise linear finite elements (see [7] for more details). We used the Python implementation available in <https://github.com/clason/bouligandlandweber>. The exact solution to be reconstructed is defined as

$$\begin{aligned} u^\dagger(x_1, x_2) &= \max(y^\dagger(x_1, x_2), 0) \\ &+ [4\pi^2 y^\dagger(x_1, x_2) - 2((2x_1 - 1)^2 + 2(x_1 - 1 + \beta)(x_1 - \beta)) \sin(2\pi x_2)] \\ &\times \mathbb{1}_{(\beta, 1-\beta]}(x_1) \end{aligned}$$

where

$$y^\dagger(x_1, x_2) = \left[(x_1 - \beta)^2 (x_1 - 1 + \beta)^2 \sin(2\pi x_2) \right] \mathbb{1}_{(\beta, 1-\beta]}(x_1)$$

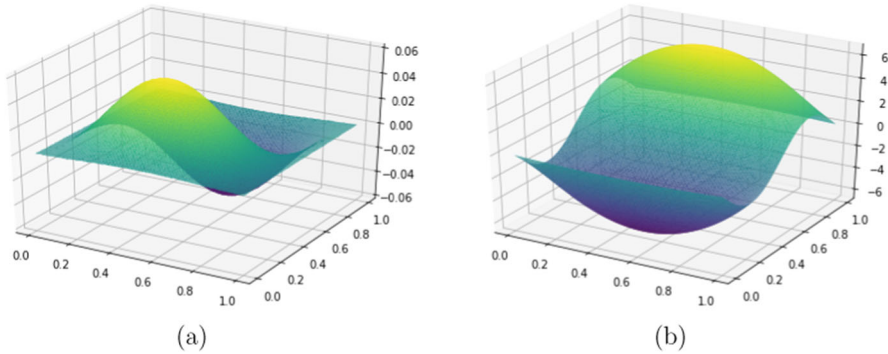


Fig. 1 **a** plots the noisy data y_h^δ with $\delta = 3.875 \times 10^{-3}$; **b** plots the initial guess $u_0 = \bar{u}$

with $\beta = 0.005$. Figure 2(i) shows a plot of u^\dagger . The function $y^\dagger \in H^2(\Omega) \cup H_0^1(\Omega)$ satisfies (3.1) for the right-hand side u^\dagger and that y^\dagger vanishes on a set of measure 2β , so the forward operator F is not Gâteaux differentiable at u^\dagger [7, proposition 3.4]. Hence we can use the Bouligand–Landweber iteration (3.3).

Given the projection of y^\dagger to the finite element space, denoted by y_h^\dagger , random Gaussian noise is added to obtain the noisy data y_h^δ , so that $\delta = \|y_h^\delta - y_h^\dagger\|_{L^2(\Omega)}$. Figure 1b shows a noisy data with $\delta = 1.043 \times 10^{-4}$.

Now we test the Landweber iteration in solving the problem under the discrepancy principle (1.6) and the Hanke–Raus rule (Rule 2.1). We used two initial solutions: the trivial point $u_0 \equiv 0$ and the function

$$\bar{u} := u^\dagger - 10 \sin(\pi x_1) \sin(2\pi x_2),$$

plotted in Fig. 1a.

Now we test the iteration under rule 2.1, with $\mu_0 = 0.6$, $\mu_1 = 5.0 \times 10^5$. For comparison, we also test under discrepancy principle, by choosing the appropriate parameter τ (see [19, theorem 3.9]). We terminated the iteration either when it satisfies (1.6) or when it exceeds 5000 iterations.

Figures 2 and 3 report the numerical results using rule 2.1 and the discrepancy principle. Using rule 2.1 for various choices of the parameter a , Figs. 2d–f and 3d–f shows the reconstruction using $u_0 \equiv 0$ and $u_0 = \bar{u}$, respectively. To illustrate that the noise condition in Assumption 2.5 is satisfied, we plot the residual against the iteration number in Figs. 2g and 3g. Figure 2(i) plots the exact solution u^\dagger . Notice the iteration (3.3) converges faster and generates better reconstruction using the initial point \bar{u} since it satisfies the general source condition with the exact solution u^\dagger [7]. The faster convergence is illustrated by the plots of $\Theta(n, u^\delta)$ versus n in Fig. 2a, b for $u_0 \equiv 0$, and Fig. 3a, b for $u_0 = \bar{u}$. Since our convergence analysis do not assume any source condition, the trivial initial point $u_0 \equiv 0$ can still produce accurate reconstruction under rule 2.1. For comparison, Figs. 2h and 3h plot the reconstructions using discrepancy principle for the two different initial points u_0 .

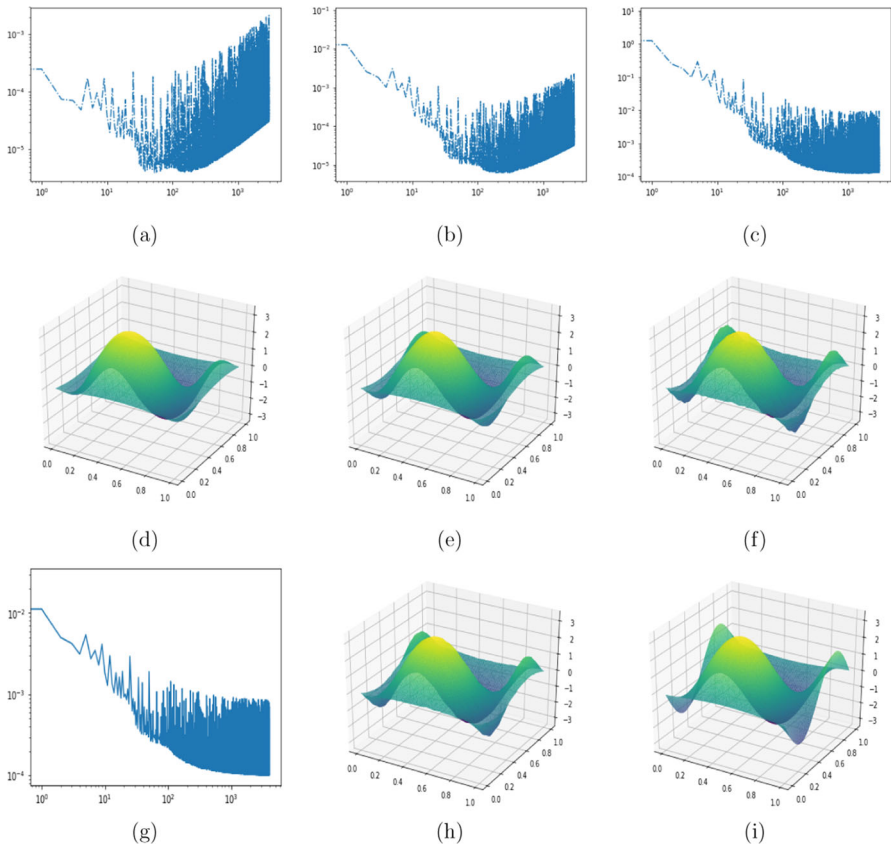


Fig. 2 The reconstruction for Example 3.1 with noise level $\delta = 3.875e-3$ and the trivial initial solution $u_0 \equiv 0$: **a**, **b**, and **c** plot the relationship between $\Theta(n, u^\delta)$ and n for $a = 1, 100$ and 10000 , respectively; **d**, **e**, and **f** plot the reconstructed solutions via rule 2.1 for $a = 1, 100$ and 10000 , respectively; **g** plots the residual $\|F(u_n^\delta) - y^\delta\|_{L^2(\Omega)}$ versus n ; **h** plots the reconstructed solution via discrepancy principle; and **i** plots the exact solution

Tables 1 and 2 gives more details in the numeric results for rule 2.1 for various noise levels using $u_0 \equiv 0$ and $u_0 = \bar{u}$, respectively. The regularization parameter n_* and the relative errors

$$E_n^\delta := \frac{\|u_h^\delta - u^\dagger\|_{L^2(\Omega)}}{\|u^\dagger\|_{L^2(\Omega)}}$$

are shown. For consistency we also show the values $\delta_{rel} = \|y_h^\delta - y_h^\dagger\|_{L^2(\Omega)} / \|y_h^\dagger\|_{L^2(\Omega)}$. For comparison, Table 3 report the stopping indices n_δ via discrepancy principle for both $u_0 \equiv 0$ and $u_0 = \bar{u}$ as well as the relative errors E_{n_δ} . As expected using the variable step (1.5) reduces the number of iterations to achieve convergence than using the constant step (3.4). The tabular results further verifies the previously described effect

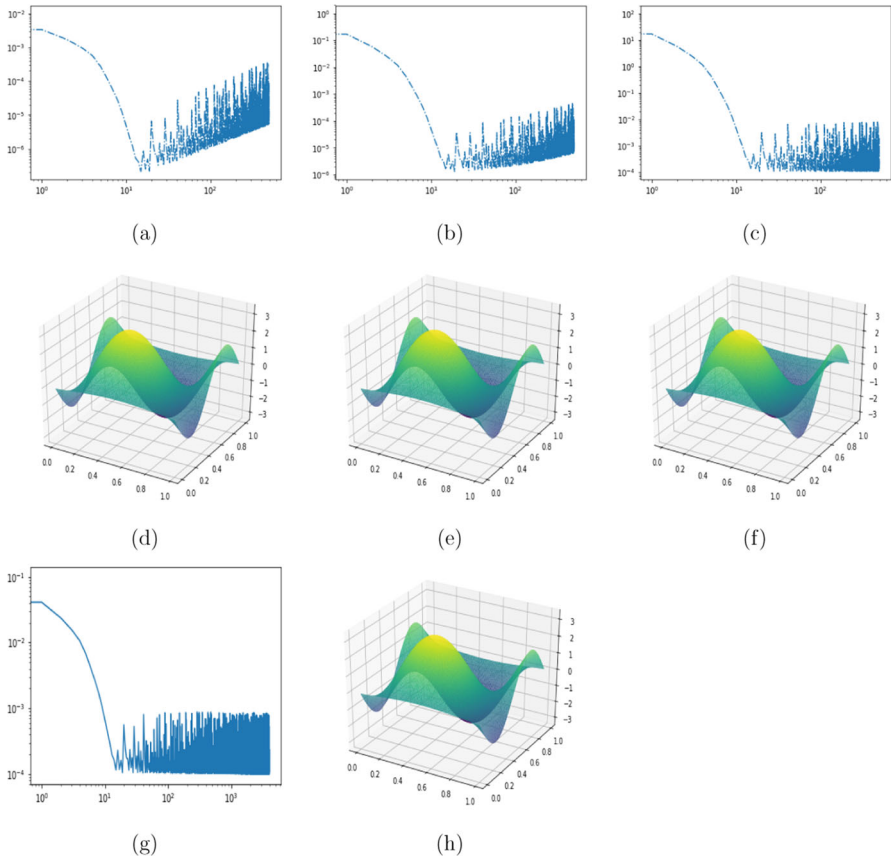


Fig. 3 The reconstruction for Example 3.1 with noise level $\delta = 3.875e-3$ and $u_0 = \bar{u}$: **a**, **b**, and **c** plot the relationship between $\Theta(n, u^\delta)$ and n for $a = 1, 100$ and 10000 , respectively; **(d)**, **(e)**, and **(f)** plot the reconstructed solutions via Hanke–Raus rule for $a = 1, 100$ and 10000 , respectively; **g** plots the residual $\|F(u_n^\delta) - y^\delta\|_{L^2(\Omega)}$ versus n ; and **h** plots the solution via discrepancy principle

using $u_0 = \bar{u}$. Just More importantly, the figures and tables show that even without information on the noise level, rule 2.1 can still produce accurate reconstructions.

Example 3.2 The problem of identifying the source or coefficient term/s in partial differential equations arises in a number of applications. Here we consider a known benchmark problem for regularizing nonlinear inverse problems.

Suppose we want to solve for the space-dependent source function c in the following elliptic boundary value problem

$$\begin{cases} -\Delta u + cu = f & \text{in } \Omega, \\ u = g & \text{on } \partial\Omega, \end{cases} \tag{3.5}$$

Table 1 Results for the Bouligand-Landweber iteration with variable stepsize for Example 3.1 under Rule 2.1 with $u_0 \equiv 0$

δ_{rel}	Rule 2.1 for various a									
	1e0		1e1		1e2		1e3		1e4	
	n_*	$E_{n_*}^\delta$	n_*	$E_{n_*}^\delta$	n_*	$E_{n_*}^\delta$	n_*	$E_{n_*}^\delta$	n_*	$E_{n_*}^\delta$
3.875e-1	1	5.129e-1	6	3.466e-1	6	3.466e-1	29	2.452e-1	59	2.840e-1
3.875e-2	7	3.142e-1	30	2.357e-1	54	2.108e-1	104	1.763e-1	255	1.571e-1
3.875e-3	53	1.855e-1	153	1.553e-1	192	1.479e-1	339	1.315e-1	1270	1.050e-1

Table 2 Results for the Bouligand-Landweber iteration with variable stepsize for Example 3.1 under Rule 2.1 with $u_0 = \bar{u}$

δ_{rel}	Rule 2.1 for various a									
	1e0		1e1		1e2		1e3		1e4	
	n_*	$E_{n_*}^\delta$	n_*	$E_{n_*}^\delta$	n_*	$E_{n_*}^\delta$	n_*	$E_{n_*}^\delta$	n_*	$E_{n_*}^\delta$
3.875e-1	5	1.627e-1	5	1.627e-2	14	3.401e-2	14	3.401e-2	64	2.203e-1
3.875e-2	11	2.932e-3	11	2.932e-3	11	2.932e-3	11	2.932e-3	90	3.087e-2
3.875e-3	15	1.590e-4	19	9.823e-4	19	9.823e-4	19	9.823e-4	117	3.580e-3

given a measurement u in Ω , where $\Omega \subset \mathbb{R}^m$ with $m \in \mathbb{N}$ is a smooth bounded domain. Given spaces \mathcal{X} and \mathcal{Y} , which are to be specified later, the forward operator F

$$F : \mathcal{D}(F) \subseteq \mathcal{X} \rightarrow \mathcal{Y}, \tag{3.6}$$

its derivative F' , and the adjoint of F' , can be formally written as

$$F(c) = A(c)^{-1} f, \quad F'(c)h = -A(c)^{-1}(h \cdot F(c)), \quad F'(c)^*w = -u(c)A(c)^{-1}w$$

for $h, w \in L^2(\Omega)$, where

$$A(c) : H^2(\Omega) \cap H_0^1(\Omega) \rightarrow L^2(\Omega) \\ u \mapsto -\Delta u + cu.$$

To preserve ellipticity, a straightforward choice of the domain $\mathcal{D}(F)$ is

$$\mathcal{D}(F) = \{c \in \mathcal{X} \mid c \geq 0 \text{ a.e. } \|c\|_X \leq \rho\} \tag{3.7}$$

for some sufficiently small $\rho > 0$. For situations requiring a nonempty interior of $\mathcal{D}(F)$ in \mathcal{X} , the choice

$$\mathcal{D}(F) = \{c \in \mathcal{X} \mid \exists \hat{c} \in L^\infty(\Omega), \hat{c} \geq 0 \text{ a.e. } \|c - \hat{c}\|_X \leq \rho\} \tag{3.8}$$

for some sufficiently small $\rho > 0$ has been devised [9].

Table 3 The reconstruction for Example 3.1 using the Bouligand-Landweber iteration (3.3) using the discrepancy principle

δ_{rel}	Using $u_0 \equiv 0$				Using $u_0 = \bar{u}$			
	DP w/ constant step (3.4)		DP w/ variable step (1.5)		DP w/ constant step (3.4)		DP w/ variable step (1.5)	
	n_δ	$E_{n_\delta}^\delta$	n_δ	$E_{n_\delta}^\delta$	n_δ	$E_{n_\delta}^\delta$	n_δ	$E_{n_\delta}^\delta$
3.875e-1	7	4.007e-1	6	3.466e-1	9	1.551e-1	9	3.854e-2
3.875e-2	118	2.200e-1	68	1.982e-1	16	1.401e-2	11	2.932e-3
3.875e-3	3288	1.227e-1	809	1.135e-1	23	1.286e-3	19	9.823e-4

So far, the problem has been studied in the context of regularization in Hilbert spaces, by setting the preimage and image spaces \mathcal{X} and \mathcal{Y} to be $L^2(\Omega)$ [9, 17, 18]. By treating the same \mathcal{X} and \mathcal{Y} as Banach spaces, the problem has also been studied in a more general setting by incorporating a non-smooth convex penalty functional to recover a non-smooth solution [13, 19, 30, 31]. However, as previously argued, $\mathcal{Y} = L^\infty(\Omega)$ or $\mathcal{Y} = L^1(\Omega)$ are more suitable choices for \mathcal{Y} , especially for a practically relevant situation of impulsive noise. Hence, we treat this example in a more general setting by using

$$\mathcal{X} = L^p(\Omega), \quad \mathcal{Y} = L^r(\Omega)$$

for $p, r \in [1, \infty]$. The result in [27, proposition 1.2] guarantees that in this more general setting the forward operator F and its derivative F' are still well-defined.

In addition, item (c) of Assumption 2.2 holds for small $\rho > 0$ [9]. Hence, we can consider $\mathcal{X} = L^2(\Omega)$ and $\mathcal{Y} = L^1(\Omega)$ for numerical simulation. We use finite differences to discretize the problem. We consider the two-dimensional problem with $\Omega = [0, 1] \times [0, 1]$ divided into $N \times N$ small squares of equal size. This results to a grid with $N + 1$ grid points in both x - and y -directions with step size $h = (N + 1)^{-1}$. We also define the sought parameter c^\dagger as

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

Since c^\dagger is piecewise constant, we use the Landweber iteration in (1.4)–(1.5) with the 2-convex TV-like functional

$$\mathcal{R}(x) := \frac{1}{2\beta} \|x\|_{L^2(\Omega)}^2 + |x|_{TV} \tag{3.9}$$

where $\beta > 0$ and

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

denotes the total variation of f over Ω . For our chosen image and preimage spaces \mathcal{X} and \mathcal{Y} , implementing the iteration (1.4) requires solving a minimization problem of the form

$$x = \arg \min_{z \in L^2(\Omega)} \{ \mathcal{R}(z) - \langle \xi, z \rangle_{L^2(\Omega)} \}$$

for any $\xi \in L^2(\Omega)$. For our choice of \mathcal{R} given by (3.9), this minimization problem is equivalent to solving

$$x = \arg \min_{z \in L^2(\Omega)} \left\{ \frac{1}{2\beta} \|z - \beta\xi\|_{L^2(\Omega)}^2 + |x|_{TV} \right\}, \tag{3.10}$$

which is the total variation denoising problem, also known as the ROF model [25]. Since there is no explicit formula for the minimisation step in (1.4), numerical solvers are used; we will use the primal-dual hybrid gradient (PDHG) algorithm [32]. The penalty functional (3.9) is discretized first before applying PDHG, as illustrated in [14]. After setting the exact data to be $u(c^\dagger) = x + y$, we add random uniform noise to produce noisy data u^δ with noise level $\delta := \|u^\delta - u(c^\dagger)\|_{L^2}$.

By relaxing the uniform smoothness on \mathcal{Y} , we use $r = 1.0$ in the duality mapping $j_r^\mathcal{Y}$, in contrast to the implementation in [19]. For $r \geq 1$ the duality mapping $j_r^\mathcal{Y}(y)$ for each $y \in \mathcal{Y} = L^r [0, 1]$ has the pointwise expression

$$\left[j_r^\mathcal{Y}(y) \right] (t) := |y(t)|^{r-1} \text{sign}(x(t)), \quad t \in [0, 1].$$

Next we apply the Landweber iteration in (1.4)–(1.5) under the discrepancy principle (1.6) and the Hanke–Raus rule (Rule 2.1). We used the parameters

$$\beta = 12.0, \quad r = 1.0, \quad p = 2.0, \quad \eta = 0.01, \quad \tau = 1.001 \cdot \frac{1 + \eta}{1 - \eta},$$

and $\xi_0 = x_0 = 0$ as initial guess. For the step size (1.5), we choose

$$\mu_0 = 3.9 \left(1 - \eta - \frac{1 + \eta}{\tau} \right) \text{ and } \mu_1 = 10000$$

for both parameter choice rules, to ensure that $\mu_0 > 0$ and (2.8) holds. We picked a large μ_1 to ensure fast convergence. Since the Fréchet derivative of F satisfies Assumption 2.2(c) for small $\rho > 0$ (see [9]), take $L = F'$.

Figure 4 shows the numerical results. The solution obtained using discrepancy principle is plotted in Fig. 4b. Moreover, the plot in Fig. 4c illustrates that the noise condition in assumption 2.5 is satisfied. The solution obtained in Fig. 4e shows that Rule 2.1 can provide accurate reconstruction in the absence of information on the noise level.

On the other hand, we also apply the Landweber iteration using a data filled with impulse noise. Figure 5 shows the reconstruction results. The noisy data in Fig. 5a

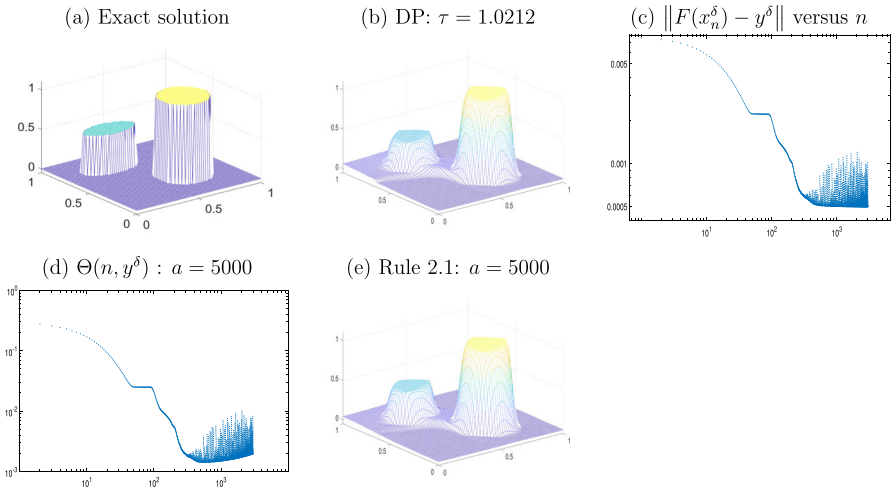


Fig. 4 The reconstruction for Example 3.2 with a Gaussian noisy data with $\delta = 0.0005$: **a** plots the exact solution c^\dagger ; **b** plots the reconstructed solution by discrepancy principle; **c** plots the relationship between the residual $\|F(x_n^\delta) - y^\delta\|_{L^2}$ and n ; **d** plots the relationship between $\Theta(n, y^\delta)$ and n for $a = 5000$ respectively; and **e** plots the reconstructed solutions via Hanke–Raus rule for $a = 5000$

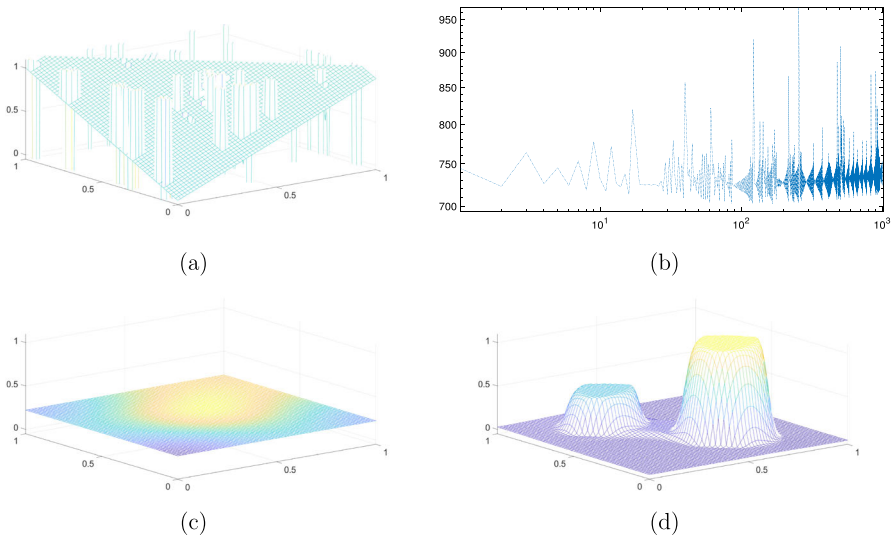


Fig. 5 The reconstruction for example 3.2 using noisy data with impulse noise: **a** plots the noisy data y^δ ; **b** plots the relationship between $\Theta(n, y^\delta)$ and n for $a = 50000$; **c** and **d** plots the reconstructions via discrepancy principle and rule 2.1, respectively

contains some outliers due to the impulse noise, making the noise level harder to estimate. We used a large number of iterations to ensure we get n_* as shown by the plot in Fig. 5b.

The inaccurate reconstruction in Fig. 5c shows that the discrepancy principle becomes ineffective with a poorly estimated noise level. However, the reconstruction in Fig. 5d shows that rule 2.1 can overcome this scenario, since it does not depend on the noise level.

Example 3.3 Now we consider an image deblurring problem, where an unknown image $x^\dagger \in \mathbb{R}^{M \times N}$ is to be reconstructed from an observed image $\tilde{y} = Fx^\dagger + v$ downgraded by a linear convolution blurring operator F and a salt-and-pepper-noise v .

Since the image to be reconstructed has some periodic boundary conditions, we use the total variation (TV) deblurring. The function $\mathcal{R}(x)$ is chosen as

$$\mathcal{R}(x) = |x|_{TV} + \frac{\beta}{2} \|x\|_F^2$$

with $\beta = 0.001$, which is a small perturbation of the TV function. Here $|x|_{TV}$ and $\|x\|_F$ denote the total variation of x and the Frobenius norm of x respectively. To remove the salt-and-pepper noise efficiently, we use the data fidelity term as

$$\frac{1}{r} \|Fx - \tilde{y}\|_r^r$$

with $r = 1.0$.

Next we apply the Landweber iteration via (1.4)–(1.5) with $r = 1.0$, $p = 2$, $\mu_0 = 0.008$, $\mu_1 = 100$. We choose $\xi_0 = x_0 = 0$ as initial guesses. In addition, the minimization problem in (3.10) is solved by a primal-dual hybrid gradient method.

Figure 6 shows the reconstructions using the Boats (576×720) motion blur (`fspecial('motion', 35, 50)` in MATLAB). The noisy observed image was obtained by adding salt-and-pepper noise with noise density 0.4. To measure the quality of the reconstructed image, we show the computed PSNR (peak signal-to-noise ratio) defined by

$$\text{PSNR} = 10 \cdot \log_{10} \frac{255^2}{\text{MSE}} \text{ (dB)}$$

where the MSE stands for the mean-squared error per pixel. We present the reconstruction of Landweber iteration via discrepancy principle and via rule 2.1. Figure 6g illustrates that assumption 2.5 is satisfied. Moreover, Fig. 6h shows an eventual rising of $\Theta(n, y^\delta)$, illustrating the existence of n_* via rule 2.1. Both parameter choice rules give satisfactory rules, and notice that rule 2.1 gives better reconstruction for a sufficiently large fixed constant a . For this type of noisy data, the Hanke–Raus rule can indeed provide accurate reconstruction.

4 Conclusion

A general convergence analysis for Landweber iteration for inverse problems under the Hanke–Raus rule was established. By using the so-called noise condition, and the

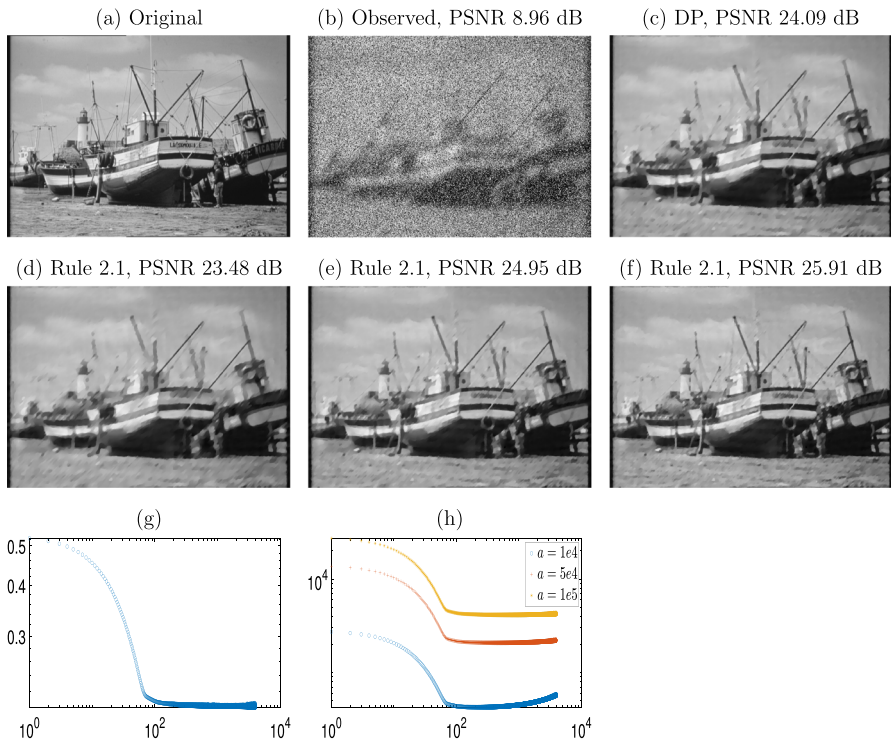


Fig. 6 **a** The original image; **b** observed image with 40% salt-and-pepper noise; **c** reconstruction result via discrepancy principle with $\tau = 1.01$; **d–f** are the reconstruction results via rule 2.1 with $a = 10000$, 50000, and 100000, respectively; **g** plots the residual $\|Fx_n^\delta - \tilde{y}\|_r$ versus n ; **h** plots $\Theta(n, \tilde{y})$ versus n for $a = 10000$, 50000, and 100000

compactness condition, we obtain the convergence result. More importantly, since the Hanke–Raus rule do not rely on any information about the noise level, this makes the Landweber iteration purely data driven, while expanding its applied range towards inverse problems with a nonsmooth forward operator and problems whose data is corrupted by various types of noise.

Acknowledgements The author also would like to thank Dr. Qinian Jin for the guidance in this research.

References

1. Aster, R.C., Borchers, B., Thurber, C.H.: Parameter Estimation and Inverse Problems, vol. 90. Elsevier Academic Press, Amsterdam (2005)
2. Bakushinskii, A.B.: Remarks on choosing a regularization parameter using the quasioptimality and ratio criterion. USSR Comput. Math. Math. Phys. **24**(4), 181–182 (1984)
3. Boj, R.I., Hein, T.: Iterative regularization with a general penalty term—theory and applications to L1 and TV regularization. Inverse Probl. **28** (2012)
4. Christof, C., Meyer, C., Walther, S., Clason, C.: Optimal control of a non-smooth semilinear elliptic equation. Math. Control Relat. Fields **8**(1), 247–276 (2018)

5. Cioranescu, I.: *Geometry of Banach Spaces, Duality Mappings, and Nonlinear Problems*. Kluwer, Dordrecht (1990)
6. Clason, C.: L^∞ fitting for inverse problems with uniform noise. *Inverse Probl.* **28**(10), 104007 (2012)
7. Clason, C., Nhu, V.H.: Bouligand–Landweber iteration for a non-smooth ill-posed problem. *Numer. Math.* **142**(4), 789–832 (2019)
8. Fu, Z., Jin, Q., Zhang, Z., Han, B., Chen, Y.: Analysis of a heuristic rule for the IRGNM in Banach spaces with convex regularization terms. *Inverse Probl.* **36**(7), 75002 (2020)
9. Hanke, M., Neubauer, A., Scherzer, O.: A convergence analysis of the Landweber iteration for nonlinear ill-posed problems. *Numer. Math.* **72**(1), 21–37 (1995)
10. Hanke, M., Raus, T.: A general heuristic for choosing the regularization parameter in ill-posed problems. *SIAM J. Sci. Comput.* **17**(4), 956–972 (1996)
11. Hubmer, S., Sherina, E., Kindermann, S., Raik, K.: A numerical comparison of some heuristic stopping rules for nonlinear Landweber iteration. *Electron. Trans. Numer. Anal.* **57**, 216–241 (2022). https://doi.org/10.1553/etna_vol57s216
12. Jin, B., Lorenz, D.A.: Heuristic parameter-choice rules for convex variational regularization based on error estimates. *SIAM J. Numer. Anal.* **48**(3), 1208–1229 (2010)
13. Jin, Q.: Hanke–Raus heuristic rule for variational regularization in Banach spaces. *Inverse Probl.* **32**(8) (2016)
14. Jin, Q.: Inexact Newton–Landweber iteration in Banach spaces with nonsmooth convex penalty terms. *SIAM J. Numer. Anal.* **53**(5), 2389–2413 (2015)
15. Jin, Q.: Landweber–Kaczmarz method in Banach spaces with inexact inner solvers. *Inverse Probl.* **32**(10), 104005 (2016)
16. Jin, Q.: On a heuristic stopping rule for the regularization of inverse problems by the augmented Lagrangian method. *Numer. Math.* **136**(4), 973–992 (2016)
17. Jin, Q.: On a regularized Levenberg–Marquardt method for solving nonlinear inverse problems. *Numer. Math.* **115**(2), 229–259 (2010)
18. Jin, Q., Wang, W.: Analysis of the iteratively regularized Gauss–Newton method under a heuristic rule. *Inverse Probl.* **34**(3) (2018)
19. Jin, Q., Wang, W.: Landweber iteration of Kaczmarz type with general non-smooth convex penalty functionals. *Inverse Probl.* **29**(8), 1–22 (2013)
20. Kaltenbacher, B., Schöpfer, F., Schuster, T.: Iterative methods for nonlinear inverse ill-posed problems in Banach spaces: convergence and applications to parameter identification problems. *Inverse Probl.* **25**(6) (2009)
21. Kikuchi, F., Nakazato, K., Ushijima, T.: Finite element approximation of a nonlinear eigenvalue problem related to MHD equilibria. *Jpn. J. Appl. Math.* **1**(2), 369–403 (1984)
22. Liu, H., Real, R., Lu, X., Jia, X., Jin, Q.: Heuristic discrepancy principle for variational regularization of inverse problems. *Inverse Probl.* **36**(7) (2020)
23. Rappaz, J.: Approximation of a nondifferentiable nonlinear problem related to MHD equilibria. *Numer. Math.* **45**(1), 117–133 (1984)
24. Real, R., Jin, Q.: A revisit on Landweber iteration. *Inverse Probl.* **36**(7), 75011 (2020)
25. Rudin, L.I., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. *Phys. D Nonlinear Phenom.* **60**(1–4), 259–268 (1992)
26. Schöpfer, F., Louis, A.K., Schuster, T.: Nonlinear iterative methods for linear ill-posed problems in Banach spaces. *Inverse Probl.* **22**(1) (2006)
27. Schuster, T., Kaltenbacher, B., Hofmann, B., Kazimierski, K.: *Regularization Methods in Banach Spaces*, vol. 10. Walter de Gruyter GmbH Co.KG, Berlin (2012)
28. Temam, R.: A non-linear eigenvalue problem: the shape at equilibrium of a confined plasma. *Arch. Ration. Mech. Anal.* **60**(1), 51–73 (1975/76)
29. Tröltzsch, F.: *Optimal Control of Partial Differential Equations*, vol. 112. Graduate Studies in Mathematics. Theory, methods and applications, Translated from the 2005 German original by Jürgen Sprekels. American Mathematical Society, Providence, RI, pp. xvi+399 (2010)
30. Zhang, Z., Jin, Q.: Heuristic rule for non-stationary iterated Tikhonov regularization in Banach spaces. *Inverse Probl.* **34**(11), 115002 (2018)
31. Zhong, M., Wang, W., Jin, Q.: Regularization of inverse problems by two-point gradient methods in Banach spaces. *Numer. Math.* **143**(3), 713–747 (2019)
32. Zhu, M., Chan, T.: An efficient primal-dual hybrid gradient algorithm for total variation image restoration. In: *UCLA CAM Report* (May 2008)

33. Zălinescu, C.: *Convex Analysis in General Vector Spaces*. World Scientific, River Edge, NJ (2002)

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.