An asymptotical regularization with convex constraints for inverse problems

Min Zhong\(^1\), Wei Wang\(^2\),∗ and Shanshan Tong\(^3\)

\(^1\) School of Mathematics, Southeast University, Nanjing, Jiangsu 210096, People’s Republic of China
\(^2\) College of Data Science, Jiaxing University, Zhejiang 314001, People’s Republic of China
\(^3\) School of Mathematics and Statistics, Shaanxi Normal University, Xi’an, Shaanxi Province, 710119, People’s Republic of China

E-mail: min.zhong@seu.edu.cn, weiwang@zjxu.edu.cn and shanshantong@snnu.edu.cn

Received 30 October 2021, revised 2 February 2022
Accepted for publication 16 February 2022
Published 8 March 2022

Abstract
We investigate the method of asymptotical regularization for the stable approximate solution of nonlinear ill-posed problems \(F(x) = y\) in Hilbert spaces. The method consists of two components, an outer Newton iteration and an inner scheme providing increments by solving a local coupling linearized evolution equations. In addition, a non-smooth uniformly convex functional has been embedded in the evolution equations which is allowed to be non-smooth, including \(L^1\)-like and total variation-like penalty terms. We establish convergence properties of the method, derive stability estimates, and perform the convergence rate under the Hölder continuity of the inverse mapping. Furthermore, based on Runge–Kutta (RK) discretization, different kinds of iteration schemes can be developed for numerical realization. In our numerical experiments, four types iterative scheme, including Landweber type, one-stage explicit, implicit Euler and two-stage RK are presented to illustrate the performance of the proposed method.

Keywords: nonlinear inverse problems, asymptotical regularization, non-smooth constraints, convergence rate

(Some figures may appear in colour only in the online journal)

∗Author to whom any correspondence should be addressed.
1. Introduction

We are interested in solving the nonlinear operator equation

\[ F(x) = y, \]  

where \( F : \mathcal{D}(F) \subset \mathcal{X} \to \mathcal{Y} \) is a nonlinear operator acting between two infinite dimensional Hilbert spaces \( \mathcal{X} \) and \( \mathcal{Y} \). Instead of exact data \( y \), in practice we are given only noisy data \( y^\delta \) obeying the deterministic noise model

\[ \| y - y^\delta \| \leq \delta, \]  

with known noise level \( \delta > 0 \). A characteristic property of such equations is their ill-posedness in the sense that the solutions are not stable with respect to the perturbation of data.

The numerical treatment of nonlinear ill-posed problem requires the application of regularization methods. Loosely speaking, two groups of regularization methods exist: variational regularization methods and iterative regularization methods. Tikhonov regularization is certainly the most significant variational regularization method (cf, e.g., [36]) in which a regularized approximation \( x^\delta \) is defined as a solution of the minimization problem

\[ \min_{x \in \mathcal{D}(F)} J_\alpha(x), \quad J_\alpha(x) = \| F(x) - y^\delta \|^2 + \alpha \| x - x_0 \|^2, \]

where \( \alpha > 0 \) is the regularization parameter and \( x_0 \) is suitable initial guess to the sought solution. The Landweber iteration is one of the most prominent iterative regularization approaches [17], in this method a regularized approximation \( x^\delta k \) is obtained from the iteration process

\[ x^{\delta k+1} = x^\delta k - F'(x^\delta k)(F(x^\delta k) - y^\delta), \quad x^\delta 0 = x_0, \]

where the iteration number plays a role of regularization parameter.

The continuous analogue of Landweber iteration can be considered as a first-order evolution equation, which is called Showalter’s method or asymptotical regularization [37], in which a regularized approximation \( x^\delta (T) \) is obtained by solving the initial value problem

\[ \frac{dx^\delta (t)}{dt} = F'(x^\delta (t))(y^\delta - F(x^\delta (t))), \quad 0 < t \leq T, \quad x^\delta (0) = x_0. \]  

The time \( T \) plays the role of the regularization parameter. This method has particular advantages for large-scale problems and for implicit ill-posed parameter identification problems. Moreover, by using Runge–Kutta (RK) integrators, all of the properties of asymptotical regularization (1.3) carry over to its numerical realization [5, 32]. Recently, there has been a great interests in analysing such dynamical flows. In [42], variational source conditions have been established for general spectral regularization methods which yield convergence rates, the asymptotical regularization for linear case has been revisited to illustrate the performance.

In [26], the authors connected the statistical asymptotical regularization method to the white noise. In [41], the authors established a second-order asymptotical regularization method and showed the corresponding convergence rate under source conditions. In [2], the authors applied the general theory of convergence rates to the three examples, including Showalter’s method, heavy ball method and the vanishing viscosity method.

To analyze the convergence rate, the theories related to general spectral analysis, source condition and the conditional stability assumption are the efficient tools. In this paper, we discussed the convergence rate of proposed asymptotical regularization under the conditional
stability assumption, which was first proposed in [6] for Hilbert space, then extended to Banach spaces in [7]. In [9], the authors analyzed a nonlinear Landweber iteration and proved its local convergence rate under the Hölder continuity of the inverse mapping, such stability assumption was developed in [22, 23], in which a finite-dimensional Ivanov’s quasisolution method was investigated. In [11], the authors considered the stable solution of nonlinear inverse problems satisfying a conditional stability estimate by Tikhonov regularization in Hilbert scales. The latest related works were [29, 30], based on the conditional stability, the convergence rate of Landweber-typed iteration, iteratively regularized Landweber iteration and iteratively regularized Gauss–Newton method in Banach spaces were discussed, respectively.

From the perspective of computational efficiency, the Landweber iteration is quite slow. Therefore, various discrete acceleration strategies are investigated. Hanke proposed the truncated Newton-CG algorithm in [15] and the regularizing Levenberg–Marquardt scheme in [16]. Rieder generalized the idea and proposed a family of inexact Newton regularization methods in [33], and gave some convergence analyses in [28, 34]. Jin studied the continuous analogue of Gauss–Newton methods in [18] and derive the rate of convergence. These methods can produce satisfactory reconstruction when the sought solutions are smooth, but always over-smooth the solution so that fail to capture the special features, such as sparsity and discontinuity. Therefore, it is necessary to reformulate these methods either in Banach space setting or in a manner that incorporate non-smooth penalty functionals, such as the $L^1$ or total variation (TV) penalties. In recent years, extensive researches have been carried out [13, 14, 20, 21, 31, 35, 43, 44]. Utilizing a proper, lower semi-continuous, uniformly convex functional $\Theta : X \to (-\infty, +\infty)$ as penalty, these methods can be applied in detecting different features of the sought solution.

In order to understand better the intrinsic properties of iterative regularization methods with uniformly convex functional $\Theta$, we consider in this paper the continuous version in the form of a coupling evolution equation. For a given time $T^\delta_n > 0$ and a constant $\tau > 1$, we assume the previous $n$-step pair $(\xi^\delta_n(T^\delta_n), x^\delta_n(T^\delta_n))$ are already obtained, satisfying $\|F(x^\delta(T^\delta_n)) - y^\delta\| \geq \tau\delta$.

Considering the linearization

$$s^\delta_n(t) = y^\delta - F(x^\delta(T^\delta_n)) - F'(x^\delta(T^\delta_n))(x(t) - x^\delta(T^\delta_n)),$$

we define a $(n + 1)$-step asymptotical regularization by solving the coupling evolution equation

$$\frac{d\xi^\delta_n(t)}{dt} = F'(x^\delta(T^\delta_n))^* s^\delta_n(t), \quad t > T^\delta_n,$$

$$x^\delta(t) = \nabla \Theta^* (\xi^\delta(t)), \quad (1.4)$$

which can be considered as a continuous analogue of inexact Newton–Landweber in [19]. We will establish the convergence and stable properties of such asymptotical regularization method, determine a proper choice of the terminating time and discuss the convergence rate under conditional stability assumption. These analyses provide theoretical guarantees for various iteration schemes. Furthermore, based on RK discretization, different kinds of iterative algorithms are developed and compared through the numerical realizations.

The paper is organized as follows. In section 2, we give some preliminaries of convex analysis. In section 3, the noisy and noise-free asymptotical algorithm are proposed, we analyze the convergence and regularization properties of the such method. In section 4, we derive the convergence rate under conditional stability assumption. In the final section, based on RK discretization, four different iteration algorithms are derived for illustration and comparison.
2. Tools from convex analysis

In this section, we revisit some concepts and properties related to convex analysis, one may refer to [39] for further details.

We restrict our attention to Hilbert spaces $\mathcal{X}$ and $\mathcal{Y}$ with norms $\| \cdot \|$ and inner-products $\langle \cdot, \cdot \rangle$. For a bounded linear operator $A : \mathcal{X} \to \mathcal{Y}$, we use $\mathcal{N}(A)$ and $A^* : \mathcal{Y} \to \mathcal{X}$ to denote its null space and 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 denote $\mathcal{D}(\Theta) := \{ x \in \mathcal{X} : \Theta(x) < \infty \}$ be its effective domain. The function $\Theta$ is called proper if $\mathcal{D}(\Theta) \neq \emptyset$. For $x \in \mathcal{X}$, we use $\partial \Theta(x)$ to denote the subdifferential of $\Theta$ at $x$, i.e.,

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

We also denote the effective domain $\mathcal{D}(\partial \Theta) := \{ x \in \mathcal{D}(\Theta) : \partial \Theta(x) \neq \emptyset \}$.

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

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

be the Bregman distance induced by $\Theta$ at $x$ in the direction $\xi \in \partial \Theta(x)$, it is always nonnegative and satisfies the identity

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

for all $x \in \mathcal{D}(\Theta), x_1, x_2 \in \mathcal{D}(\partial \Theta)$, 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 weakly lower semi-continuous at $x \in \mathcal{X}$ if for arbitrary sequence $x_k \in \mathcal{X}$, the weakly convergence of $x_k \rightharpoonup x$ yields

$$\Theta(x) \leq \liminf_{k \to \infty} \Theta(x_k).$$

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

$$\Theta(\lambda \bar{x} + (1 - \lambda)x) + \lambda(1 - \lambda)h(\| x - \bar{x} \|) \leq \lambda \Theta(\bar{x}) + (1 - \lambda)\Theta(x), \quad (2.2)$$

for all $\bar{x}, x \in \mathcal{X}$ and $\lambda \in [0, 1]$. Especially when $h(t) = c_0 t^2$ for some $c_0 > 0$ in (2.2), $\Theta$ is called strongly convex. It is straightforward to show that $\Theta$ is strongly convex if and only if

$$D_{\xi} \Theta(\bar{x}, x) \geq c_0 \| \bar{x} - x \|^2. \quad (2.3)$$

In addition, if $\Theta$ is strongly convex, then for arbitrary $x, \bar{x} \in \mathcal{D}(\partial \Theta), \xi \in \partial \Theta(x), \tilde{\xi} \in \partial \Theta(\bar{x})$, it can be proved in [44] that

$$D_{\xi} \Theta(\bar{x}, x) \leq \frac{1}{4c_0} \| \xi - \tilde{\xi} \|^2. \quad (2.4)$$

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

$$\Theta^*(\xi) := \sup_{x \in \mathcal{X}} \{ \langle \xi, x \rangle - \Theta(x) \}, \quad \xi \in \mathcal{X}. $$
It is well known that $\Theta^*$ is also proper, lower semi-continuous, and convex. In addition,
\[
\xi \in \partial \Theta(x) \iff x \in \partial \Theta^*(\xi) \iff \Theta(x) + \Theta^*(\xi) = \langle \xi, x \rangle. \tag{2.5}
\]
Moreover, if $\Theta$ is strongly convex, then it follows from [39] (corollary 3.5.11) that $D(\Theta^*) = \mathcal{X}$, $\Theta^*$ is Fréchet differentiable and its gradient $\nabla \Theta^*: \mathcal{X} \to \mathcal{X}$ satisfies
\[
\| \nabla \Theta^*(\xi_1) - \nabla \Theta^*(\xi_2) \| \leq \frac{\| \xi_1 - \xi_2 \|}{2c_0}, \quad \forall \xi_1, \xi_2 \in \mathcal{X}. \tag{2.6}
\]
Consequently, it follows from (2.5) that
\[
x \in \partial \Theta^*(\xi) \iff \xi \in \partial \Theta(x) \iff x = \arg \min_{z \in \mathcal{X}} \{ \Theta(z) - \langle \xi, z \rangle \}. \tag{2.7}
\]
In many practical applications, a proper, weakly lower semi-continuous and strongly convex functional can be easily constructed. For instance, let $\mathcal{X} = L^2(\Omega)$ with $\Omega$ be a bounded domain in $\mathbb{R}^d$. We can create
\[
\Theta(x) = \nu \int_{\Omega} |x(\omega)|^2 d\omega + a \int_{\Omega} |x(\omega)| d\omega + b \int_{\Omega} |Dx|, \tag{2.8}
\]
where $\nu > 0$, $a, b \geq 0$ and $\int_{\Omega} |Dx|$ denotes the TV of $x$ over $\Omega$. With different settings for $a = 1$, $b = 0$ or $a = 0$, $b = 1$, the functional includes $L^2-L^1$ and $L^2$-TV penalties which is useful for detecting the discontinuities and for sparsity reconstruction.

3. Convergence of the asymptotical method

We now return to equation (1.1), where $F: D(F) \subset \mathcal{X} \to \mathcal{Y}$ is a nonlinear operator between Hilbert spaces $\mathcal{X}$ and $\mathcal{Y}$. 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 the functional $\Theta$ and operator $F$.

**Assumption 3.1.** The functional $\Theta: \mathcal{X} \to (-\infty, \infty]$ is proper, weakly lower semi-continuous, and strongly convex such that the condition (2.2) is satisfied for $h(t) = c_0 t^2$ with some $c_0 > 0$.

**Assumption 3.2.** The nonlinear operator $F: D(F) \subset \mathcal{X} \to \mathcal{Y}$ satisfies
(a) There exists $\rho > 0$, $x_0 \in \mathcal{X}$ and $\xi_0 \in \partial \Theta(x_0)$ such that $B_{2\rho}(x_0) := \{ x \in \mathcal{X} : \| x - x_0 \| \leq 2\rho \} \subset D(F)$, and (1.1) has a solution $x^* \in D(\Theta)$ satisfying
\[
D_{\Theta}(x^*, x_0) \leq c_0 \rho^2.
\]
(b) The operator $F$ is weakly closed on $D(F)$.
(c) The operator $F$ is Fréchet differentiable, whose Fréchet derivative is denoted by $F'(x)$. The map $x \to F'(x)$ is continuous on $B_{2\rho}(x_0) \cap D(F)$ and there exists constant $0 \leq \eta < 1$ such that
\[
\| F(x) - F(\bar{x}) - F'(\bar{x})(x - \bar{x}) \| \leq \eta \| F(x) - F(\bar{x}) \|, \quad \forall x, \bar{x} \in B_{2\rho}(x_0). \tag{3.1}
\]
Moreover, there is a constant $L > 0$ such that

$$||F'(x)||_{X\rightarrow Y} \leq L, \quad \forall x \in B_{2\rho}(x_0).$$

All the conditions in assumption 3.2 are standard. The weakly closedness of $F$ in condition (b) means that, if $\{x_n\} \subset D(F)$ weakly converges to some $u \in X$ and $\{F(x_n)\}$ weakly converges to some $v \in Y$, then $u \in D(F)$ and $F(u) = v$. The (3.1) in condition (c) is called the tangential condition and is widely used in the analysis of iterative regularization methods for nonlinear ill-posed inverse problems [24].

In general, the equation (1.1) may have many solutions. In order to find the desired one, we enforce some selection criteria. According to a priori information on the sought solution, using a functional $\Theta$ specified in assumption 3.1, we may pick among solution of (1.1) the one with desired feature. We define the $x^1$ be the solution of (1.1) satisfying

$$D_{\delta_0} \Theta(x^1, x_0) = \min_{x \in D(\Theta) \cap D(F)} \{D_{\delta_0} \Theta(x, x_0), F(x) = y\}. \quad (3.2)$$

Utilizing weakly lower semi-continuity and strongly convexity of $\Theta$ together with the weakly closedness of $F$, it is standard to show the existence of $x^1$. Moreover, combining with condition (a) in assumption 3.2, it is obvious that

$$D_{\delta_0} \Theta(x^1, x_0) \leq c_0 \rho^2,$$

thus $x^1 \in B_{\rho}(x_0)$. It has been shown in [21] (lemma 3.2) that $x^1$ is unique.

3.1. Asymptotical method in noisy case

To formulate the asymptotical method in noisy case, let $y^\delta$ be a available noisy data satisfying (1.2), and pick the initial time $T_0^\delta = T_0$, define the initial guess $\xi^0(T_0^\delta) := \xi_0 \in X$ and $x^\delta(T_0^\delta) := \arg \min_{x \in X} \{\Theta(x) - \langle \xi_0, x \rangle\}$. At a current iterate $x^\delta(T_n^\delta)$ with $n \geq 0$, we have the local linearized equation

$$F'(x^\delta(T_n^\delta))x(t) = y^\delta - F(x^\delta(T_n^\delta)) + F'(x^\delta(T_n^\delta))x^\delta(T_n^\delta).$$

We may consider the asymptotical method to this equation to produce the next time $T_{n+1}^\delta$ and corresponding approximate $x^\delta(T_{n+1}^\delta)$ so that $||y^\delta - F(x^\delta(T_{n+1}^\delta)) - F'(x^\delta(T_n^\delta))(x^\delta(T_{n+1}^\delta) - x^\delta(T_n^\delta))||$ is smaller than a suitable multiple of residual $||y^\delta - F(x^\delta(T_n^\delta))||$. The idea will be formulated in algorithm 1.

The algorithm 1 consists of two components: an outer iteration and an inner scheme providing increments by solving a coupling linearized evolution equation. Note that the subgradient $\xi^\delta(t)$ is determined by $F$ completely without using the functional $\Theta$, and the approximated function $x^\delta(t)$ is defined as the minimizer of a convex functional over $X$, which may independent of $F$, this splitting can make implementation for solving the evolution equation efficiently. By using (2.7) and the differentiability of $\Theta^*$, one can see that

$$x^\delta(t) = \nabla \Theta^*(\xi^\delta(t)). \quad (3.7)$$

We will use this fact in the forthcoming convergence analysis.

In order to show that algorithm 1 is well-defined, we first claim that the coupling linearized evolution equation (3.3) has a unique solution.

Lemma 3.3. Assume $x^\delta(T_n^\delta) \in B_{2\rho}(x_0)$, and define the operator $G : X \rightarrow X$ by

$$G(\xi) = F'(x^\delta(T_n^\delta))^{-1}(y^\delta - F(x^\delta(T_n^\delta)) + F'(x^\delta(T_n^\delta))((\nabla \Theta^*(\xi) - x^\delta(T_n^\delta))),$$

$$G(\xi) = F'(x^\delta(T_n^\delta))^{-1}(y^\delta - F(x^\delta(T_n^\delta)) + F'(x^\delta(T_n^\delta))((\nabla \Theta^*(\xi) - x^\delta(T_n^\delta))),$$
then, the coupling linearized evolution equation (3.3) has a unique solution.

Proof. For \((\xi_1, x_1), (\xi_2, x_2) \in \mathcal{X} \times \mathcal{X}\) with \(x_i = \nabla \Theta^*(\xi_i)\) for \(i = 1, 2\), we have

\[G(\xi_1) - G(\xi_2) = F'(x_1(T_n^\delta))F'(x_2(T_n^\delta))(\nabla \Theta^*(\xi_1) - \nabla \Theta^*(\xi_2)).\]

The application of assumption 3.1 (c) and (2.6) yields

\[\|G(\xi_1) - G(\xi_2)\| \leq \frac{L^2}{2c_0}\|\xi_1 - \xi_2\|.
\]

Therefore the operator \(G(\xi)\) is Lipschitz continuous in \(\mathcal{X}\). The existence and uniqueness are consequently obtained with the application of Cauchy–Lipschitz–Pichard theorem in [8].

Now we provide the following basic results.

Lemma 3.4. Let \(\mathcal{X}\) and \(\mathcal{Y}\) be both Hilbert spaces. The function \(\Theta\) and the operator \(F\) satisfy assumptions 3.1 and 3.2 respectively. Let \((\xi(t), \xi^s(t))\) be the solution of (3.3) for \(T \geq T_n^\delta\). Then \(\xi(t)\) and residual \(\|F(\xi(t)) - y\|\) are continuous with respect to \(T\).

Proof. Referring to (3.7) and (2.6),

\[\lim_{\Delta T \to 0} \|\xi(T + \Delta T) - \xi(T)\| = \lim_{\Delta T \to 0} \|\nabla \Theta^*(\xi(T + \Delta T)) - \nabla \Theta^*(\xi(T))\| \leq \lim_{\Delta T \to 0} \frac{\|\xi(T + \Delta T) - \xi(T)\|}{2c_0} \leq \lim_{\Delta T \to 0} \|\frac{\xi(T + \Delta T) - \xi(T)}{\Delta T}\| \cdot \frac{\Delta T}{2c_0},\]
Since
\[
\lim_{\Delta T \to 0} \left\| \frac{\xi(T + \Delta T) - \xi(T)}{\Delta T} \right\| = \left\| \frac{d\xi(t)}{dt} \right\|_{T = T},
\]
(3.8)
evaluates and \( \lim_{\Delta T \to 0} \frac{\Delta T}{|T_0^\delta|} = 0 \). This implies \( x^\delta(T) \) is continuous. With the continuity of \( F \) the residual is continuous as well.

**Proposition 3.5.** Let \( \mathcal{X} \) and \( \mathcal{Y} \) be both Hilbert spaces. The function \( \Theta \) and the operator \( F \) satisfy assumptions 3.1 and 3.2 respectively. Let \((\xi^\delta(T), x^\delta(T))\) be the solution of (3.3) for \( T \geq T_n^\delta \). For any solution \( \hat{x} \) of (1.1) in \( B_{2\rho}(x_0) \bigcap D(\Theta) \), define the function
\[
\phi(T) := D_{\xi^\delta(T)} \Theta(\hat{x}, x^\delta(T)),
\]
assume \( x^\delta(T_n^\delta) \in B_{2\rho}(x_0) \), then \( \phi(T) \) is differentiable for \( T \geq T_n^\delta \) and
\[
\phi'(T) \leq \left( -\|s_n^\delta(T)\| + \delta + \eta \|\gamma - F(x^\delta(T_n^\delta))\| \right) \|s_n^\delta(T)\|.
\]
(3.9)

**Proof.** Referring to (2.1), for \( T \geq T_n^\delta \), we have
\[
\lim_{\Delta T \to 0} \frac{\phi(T + \Delta T) - \phi(T)}{\Delta T} = \lim_{\Delta T \to 0} \frac{D_{\xi^\delta(T + \Delta T)} \Theta(\hat{x}, x^\delta(T + \Delta T)) - D_{\xi^\delta(T)} \Theta(\hat{x}, x^\delta(T))}{\Delta T} = \lim_{\Delta T \to 0} \frac{D_{\xi^\delta(T + \Delta T)} \Theta(x^\delta(T), x^\delta(T + \Delta T)) + (\xi^\delta(T + \Delta T) - \xi^\delta(T), x^\delta(T) - \hat{x})}{\Delta T}.
\]
The application of (2.4) provides
\[
\lim_{\Delta T \to 0} \frac{D_{\xi^\delta(T + \Delta T)} \Theta(x^\delta(T), x^\delta(T + \Delta T))}{\Delta T} = \lim_{\Delta T \to 0} \frac{|\Delta T| \left\| \frac{\xi^\delta(T + \Delta T) - \xi^\delta(T)}{\Delta T} \right\|^2}{4c_\gamma}.
\]
(3.10)

Referring to (3.8) again, the above estimate converges to 0. Then, since
\[
\lim_{\Delta T \to 0} \left\langle \frac{\xi^\delta(T + \Delta T) - \xi^\delta(T)}{\Delta T}, x^\delta(T) - \hat{x} \right\rangle = \left\langle \frac{d\xi(t)}{dt} \right\|_{T = T}, x^\delta(T) - \hat{x} \right\rangle = \left( s_n^\delta(T), F'(x^\delta(T_n^\delta))(x^\delta(T) - \hat{x}) \right),
\]
and the definition of \( s_n^\delta(T) \) gives
\[
F'(x^\delta(T_n^\delta))(x^\delta(T) - \hat{x}) = -s_n^\delta(T) + (s^\delta - F(x^\delta(T_n^\delta)) - F'(x^\delta(T_n^\delta))(\hat{x} - x^\delta(T_n^\delta))).
\]
Since \( x^i(T^i_n) \in B_{2\rho}(x_0) \), utilizing the assumption 3.2 (c),

\[
\begin{align*}
\left\langle \frac{d x^i(t)}{dt}, x^i(T) - \tilde{x} \right\rangle \\
= \langle \dot{x}^i(T), F'(x^i(T^i_n))(x^i(T) - \tilde{x}) \rangle \\
= -\| \dot{x}^i(T) \|^2 + \langle \dot{x}^i(T), y^\delta - F(x^i(T^i_n)) - F'(x^i(T^i_n))(\tilde{x} - x(T^i_n)) \rangle \\
\leq -\| \dot{x}^i(T) \|^2 + \| \dot{x}^i(T) \| \| F(x^i(T^i_n)) - y^\delta \| \| F'(x^i(T^i_n))(\tilde{x} - x(T^i_n)) \| \\
\leq -\| \dot{x}^i(T) \|^2 + \| \dot{x}^i(T) \| \| F(x^i(T^i_n)) - y^\delta \| \| F'(x^i(T^i_n))(\tilde{x} - x(T^i_n)) \|. 
\end{align*}
\] (3.11)

The combination of the estimates (3.10) and (3.11) provides the result. \( \square \)

**Lemma 3.6.** Let \( \mathcal{X} \) and \( \mathcal{Y} \) be both Hilbert spaces. The function \( \Theta \) and the operator \( F \) satisfy assumptions 3.1 and 3.2 respectively. Let \( (\xi^i(T), x^i(T)) \) be the solution of (3.3) for \( T \geq T^i_n \).

Assume the nonlinearity constant \( \eta \) in (3.1) satisfies \( \eta < \gamma < 1 \) and the discrepancy constant \( \tau \) in (3.6) satisfies \( \tau > \frac{1+\eta}{\gamma} \rho \). Then, we have the following properties

- \( \phi(T) \) is monotonically decreasing in \( T \in [T^i_n, T^i_{n+1}] \) with \( 0 < n < n_\delta \).
- \( x^i(T^i_n) \in B_{2\rho}(x_0) \) with \( 0 \leq n \leq n_\delta \).
- \( T^i_n \) is finite with \( 0 \leq n \leq n_\delta \).

**Proof.** We will prove the conclusion by induction. For \( T \in [T^i_n, T^i_{n+1}] \), since \( n < n_\delta \), we have

\[
\| \dot{x}^i(T) \| \geq \| F(x^i(T^i_n)) - y^\delta \| \quad \text{and} \quad \| F(x^i(T^i_n)) - y^\delta \| > \gamma \delta.
\]

Assume \( x^i(T^i_n) \in B_{2\rho}(x_0) \) is valid for all \( 0 \leq n < n_\delta \), recalling to (3.9), it follows that

\[
\phi(T) \leq \left(-\| \dot{x}^i(T) \| + \delta + \eta \delta + \eta \| F(x^i(T^i_n)) - y^\delta \| \right) \| \dot{x}^i(T) \|
\leq \left(-\| \dot{x}^i(T) \| + \frac{1+\eta}{\gamma} \| F(x^i(T^i_n)) - y^\delta \| \right) \| \dot{x}^i(T) \|
\leq \left(1 - \frac{1+\eta}{\gamma \tau} - \frac{\eta}{\gamma} \right) \| \dot{x}^i(T) \|^2.
\] (3.12)

Since \( \gamma \in (\eta, 1) \) and \( \tau > \frac{1+\eta}{\gamma \eta} \), it is consequently that \( \phi(T) < 0 \). Equivalently,

\[
D_{\xi^i(T)}(\hat{\Theta}(\hat{x}, x^i(T))) < D_{\xi^i(T^i_n)}(\hat{\Theta}(\hat{x}, x^i(T^i_n))), \quad T \in (T^i_n, T^i_{n+1}].
\]

In particular, when \( T = T^i_{n+1} \), we have

\[
D_{\xi^i(T^i_{n+1})}(\hat{\Theta}(\hat{x}, x^i(T^i_{n+1}))) < D_{\xi^i(T^i_n)}(\hat{\Theta}(\hat{x}, x^i(T^i_n))).
\]

Then, by taking \( \hat{x} = x^i \) and using the induction with respect to \( n \), we have

\[
D_{\xi^i(T^i_{n+1})}(\hat{\Theta}(x^i, x^i(T^i_{n+1}))) < \cdots < D_{\xi^i}(\hat{\Theta}(x^i, x_0)) \leq c_0 \rho^2.
\]

This together with \( x^i \in B_{\rho}(x_0) \) yields \( x^i(T^i_{n+1}) \in B_{2\rho}(x_0) \).
In order to show $T_{n+1}^3$ is finite. Assume $T_n^3$ is finite, we integrate the (3.12) from $T_n^3$ to $T$, yielding

$$\phi(T) - \phi(T_n^3) < - \left(1 - \frac{1 + \eta}{\gamma} \right) \frac{\eta}{\gamma} \int_{T_n^3}^{T} \|s_n^3(t)\|^2 dt,$$

(3.13)

which implies

$$\left(1 - \frac{1 + \eta}{\gamma} - \frac{\eta}{\gamma}\right) \int_{T_n^3}^{T} \|s_n^3(t)\|^2 dt < \phi(T_n^3).$$

Consequently,

$$\left(1 - \frac{1 + \eta}{\gamma} - \frac{\eta}{\gamma}\right) (T - T_n^3) \gamma^2 \tau^2 \delta^2 < \phi(T_n^3) = D_{\xi,T_n^3}(\Theta(x^1, x^3(T_n^3))) \leq c_0 \rho^2.$$

The above estimate is valid for all $T \leq T_{n+1}^3$, this shows that $T_{n+1}^3$ is finite. □

In the following, we are willing to show the discrepancy $\|F(x^3(T_n^3)) - y^3\|$ is monotonically decreasing with respect to $n$.

**Lemma 3.7.** Under all the conditions of lemma 3.6, if $0 < \eta < 1/3$ and $\eta < \gamma < 1 - 2\eta$, then the residual $\|F(x^3(T_n^3)) - y^3\|$ is monotonically decreasing with respect to $n \leq n_s$ and $n_s$ is finite.

**Proof.** The conclusion is directly derived by the following estimate:

$$\|F(x^3(T_{n+1}^3)) - y^3\| \leq \|F(x^3(T_{n+1}^3)) - F(x^3(T_n^3)) - F(x^3(T_n^3))(x^3(T_{n+1}^3) - x^3(T_n^3))\| + \|s_n^3(T_{n+1}^3)\|$$

$$\leq \eta\|F(x^3(T_{n+1}^3)) - F(x^3(T_n^3))\| + \gamma\|F(x^3(T_n^3)) - y^3\|$$

$$\leq \eta\|F(x^3(T_{n+1}^3)) - y^3\| + (\eta + \gamma)\|F(x^3(T_n^3)) - y^3\|.$$

It is consequently that

$$\|F(x^3(T_{n+1}^3)) - y^3\| \leq \frac{\eta + \gamma}{1 - \eta}\|F(x^3(T_n^3)) - y^3\|.$$

Since the constant $\frac{\eta + \gamma}{1 - \eta} < 1$, thus the residual $\|F(x^3(T_n^3)) - y^3\|$ is monotonically decreasing, consequently stopping index $n_s$ is finite thus the algorithm is well defined. □

In the following we will give the convergence analysis of algorithm 1. It is necessary to provide the convergence analysis in noise-free case and then study the stability issue.

### 3.2. Asymptotical method in noise-free case

In this subsection, associated with the exact data, we reformulate the algorithm as follows.

By using the same arguments in the proof of lemma 3.6, one can show the following similar statements for noise-free case. It is worth to note that, compared with algorithm 1, we do not demand $T_{n+1}^3$ be the first time satisfying equality (3.16).
**Lemma 3.8.** Let $X$ and $Y$ be both Hilbert spaces. The function $\Theta$ and the operator $F$ satisfy assumptions 3.1 and 3.2 respectively. Let $(\xi(T), x(T))$ be the solution of (3.14) for $T \geq T_n$. Assume $\eta < \gamma < 1$, we have the following properties

- $\varphi(T) := D_{\xi(T)} \Theta(\hat{x}, x(T))$ is monotonically decreasing in $T \in [T_n, T_{n+1}]$, where $\hat{x}$ is arbitrary solution of (1.1) in $B_{\varphi}(x_0) \cap D(\Theta)$.
- $x(T_n) \in B_{\varphi}(x_0)$ for $n \in \mathbb{N}$.
- For $T \geq T_n$, we have
  \[ \varphi(T) - \varphi(T_n) \leq -\frac{\gamma - \eta}{\gamma} \int_{T_n}^{T} \|s_n(t)\|^2 dt. \]  
  (3.17)

- $T_{n+1}$ is finite.

**Proof.** Assume $x(T_n) \in B_{\varphi}(x_0)$, the similar technique in proposition 3.5 can be utilized, which yields
  \[ \varphi'(T) \leq -\|s_n(T)\|^2 + \eta\|s_n(T)\|\|y - F(x_n(T))\|. \]

Since for $T \in [T_n, T_{n+1}]$, we have
  \[ \|s_n(t)\| \geq \gamma\|y - F(x_n(T))\|, \]
then,
  \[ \varphi'(T) \leq -\left(\frac{\gamma - \eta}{\gamma}\right)\|s_n(T)\|^2 < 0. \]  
  (3.18)

This implies $\varphi(T)$ is decreasing in $[T_n, T_{n+1}]$ and consequently $x(T_{n+1}) \in B_{\varphi}(x_0)$. Integrating (3.18) from $T_n$ to $T$, we have (3.17).

Finally, in order to prove $T_{n+1}$ is finite, integral the (3.18) from $T_n$ to $T$, yielding
  \[ \gamma(\gamma - \eta)(T - T_n)\|y - F(x_n(T))\|^2 \leq \varphi(T_n) \leq \cdots \leq c_0 \rho^2, \quad \forall T \in [T_n, T_{n+1}]. \]

Notice that $F(x_n(T)) \neq y$, this shows $T_{n+1}$ is finite.

**Lemma 3.9.** Under all the condition of lemma 3.8, if $0 < \eta < 1/3$ and $\eta < \gamma < 1 - 2\eta$, then the residual $\|F(x(T_n)) - y\|$ is monotonically non-increasing with respect to $n \in \mathbb{N}$.

In order to derive the convergence of $\{x(T_n)\}$ to a solution of (1.1) in Bregman distance, we need the following useful proposition, which was first appeared in [21] for Banach spaces.

**Proposition 3.10.** Let $X$ and $Y$ be both Hilbert spaces. The function $\Theta$ and the operator $F$ satisfy assumptions 3.1 and 3.2 respectively. Then, if $\{x(T_n)\} \subset B_{\varphi}(x_0)$ and $\{\xi(T_n)\} \subset X$ such that

(a) $\xi(T_n) \in \partial \Theta(x(T_n))$ for $n \in \mathbb{N}$.
(b) For arbitrary solution of (1.1) $\hat{x} \in B_{\varphi}(x_0) \cap D(\Theta)$, the Bregman distance $\{D_{\xi(T_n)}\Theta(\hat{x}, x(T_n))\}$ is monotonically decreasing.
(c) $\lim_{n \to \infty} \|F(x(T_n)) - y\| = 0.$
Algorithm 2. Asymptotical method in noise-free case.

Input: Exact data $y$, suitably chosen numbers $0 < \gamma < 1$ and $\tau > 1$.

Initial guess: $\xi(T_0) \vcentcolon= \xi_0 \in \mathcal{X}$ and $x(T_0) \vcentcolon= \arg\min_{x \in \mathcal{X}} \{\Theta(x) - \langle \xi_0, x \rangle\}$.

Set $n = 0$.

Repeat:

1. If $F(x(T_n)) = y$, set $\xi(T_{n+1}) = \xi(T_n)$ and $x(T_{n+1}) = x(T_n)$.
2. If $F(x(T_n)) \neq y$, solving the coupling linearized ODE

\[
\begin{align*}
\frac{d\xi(t)}{dt} &= F'(x^\dagger(T_n))\xi(t), \quad t \geq T_n; \\
\xi(t) &= \arg\min_{x \in \mathcal{X}} \{\Theta(x) - \langle \xi(t), x \rangle\},
\end{align*}
\]

(3.14)

in which $s_n(t) = y - F(x(T_n)) - F'(x(T_n))(x(t) - x(T_n))$.

Determine the $T_{n+1} > T_n$ satisfying

\[
\|s_n(t)\| \geq \gamma \|F(x(T_n)) - y\|, \quad T_n \leq t < T_{n+1},
\]

and define $\xi(T_{n+1})$ and $x(T_{n+1})$.

Set $n = n + 1$.

\[\] 

(d) There exists a subsequence $\{n_j\}$ with $n_j \to \infty$ such that for arbitrary solution of (1.1) \( \hat{x} \in B_{2\rho}(x_0) \cap \mathcal{D}(\Theta) \), there holds

\[
\limsup_{j \to \infty} \|\xi(T_{n_j}) - \xi(T_{n_j}), x(T_{n_j}) - \hat{x}\| = 0. \tag{3.19}
\]

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

\[
\lim_{n \to \infty} \|x_* - x(T_n)\| = 0, \quad \lim_{n \to \infty} D_{\xi(T_n)}(x_*, x(T_n)) = 0.
\]

If in addition, $x^\dagger \in B_{\rho}(x_0) \cap \mathcal{D}(\Theta)$, and $\xi(T_{n+1}) - \xi(T_n) \in \text{Ran}(F'(x^\dagger)^*)$ for all $n \in \mathbb{N}$, then $x_* = x^\dagger$.

Theorem 3.11. Let $\mathcal{X}$ and $\mathcal{Y}$ be both Hilbert spaces. The function $\Theta$ and the operator $F$ satisfy assumptions 3.1 and 3.2 respectively. Assume nonlinearity constant $0 < \eta < 1/3$ and suitable choosing $\eta < \gamma < 1 - 2\eta$. Let \{$(\xi(T_n), x(T_n))$\}$_{n \in \mathbb{N}}$ be the sequence generated by algorithm 2, then there exists a solution $x_* \in B_{2\rho}(x_0) \cap \mathcal{D}(\Theta)$ such that

\[
\lim_{n \to \infty} \|x(T_n) - x_*\| = 0, \quad \lim_{n \to \infty} D_{\xi(T_n)}(x_*, x(T_n)) = 0. \tag{3.20}
\]

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

Proof. We will use proposition 3.10 in the following proof. By definition we always have $x(T_n) = \nabla^\times \Theta^*(\xi(T_n))$ which implies $\xi(T_n) \in \partial \Theta(x(T_n))$ for all $n \in \mathbb{N}$. The lemma 3.8 provides the monotonicity of Bregman distance, and referring to the lemma 3.9 we can also show the monotonicity of Bregman distance and residual. Moreover, the residual converges to zero due to the compressibility factor $\frac{\eta}{1-\eta} < 1$.
According to proposition 3.10, it remains only to show (3.19). To this end, recalling the monotonicity of the residual, we can consider the sequence itself, that

$$\|F(x(T_n)) - y\| \leq \frac{\eta + \gamma}{1 - \eta} |F(x(T_n)) - y|, \quad 0 \leq n \leq \ell.$$ 

Now, for $0 \leq j < \ell < \infty$, we consider

$$\langle \xi(T_j) - \xi(T_j), x(T_j) - \hat{x} \rangle = \sum_{n=j}^{\ell-1} \langle \xi(T_{n+1}) - \xi(T_n), x(T_n) - \hat{x} \rangle$$

$$= \sum_{n=j}^{\ell-1} \int_{T_n}^{T_{n+1}} F'(x(T_n))^\ast s_n(t)dt, x(T_n) - \hat{x} \rangle$$

$$= \sum_{n=j}^{\ell-1} \int_{T_n}^{T_{n+1}} \langle s_n(t), F'(x(T_n))(x(T_n) - \hat{x}) \rangle dt$$

$$\leq \sum_{n=j}^{\ell-1} \int_{T_n}^{T_{n+1}} \| s_n(t) \| \| F'(x(T_n))(x(T_n) - \hat{x}) \| dt.$$

The application of assumption 3.2 (c) yields

$$\| F'(x(T_n))(x(T_n) - \hat{x}) \| \leq \| F'(x(T_n))(x(T_n) - x(T_n)) \| + \| F'(x(T_n))(x(T_n) - \hat{x}) \|$$

$$\leq (1 + \eta) \left( \| F(x(T_n)) - F(x(T_n)) \| + \| F(x(T_n)) - y \| \right)$$

$$\leq (1 + \eta) \left( \| F(x(T_n)) - y \| + 2 \| F(x(T_n)) - y \| \right)$$

$$\leq 3 \| F(x(T_n)) - y \|,$$

then, referring to (3.17),

$$\| \langle \xi(T_n) - \xi(T_j), x(T_n) - \hat{x} \rangle \| \leq \frac{3(1 + \eta)}{\gamma} \sum_{n=j}^{\ell-1} \int_{T_n}^{T_{n+1}} \| s_n(t) \|^2 dt$$

$$\leq \frac{3(1 + \eta)}{\gamma} \left( D_{\xi(T_j)} \Theta(x, x(T_j)) - D_{\xi(T_j)} \Theta(\hat{x}, x(T_j)) \right).$$

In view of the monotonicity of Bregman distance, the above estimate implies (3.19).

To show that $x_n = x^\dagger$ under the condition $\mathcal{N}(F'(x^\dagger)) \subset \mathcal{N}(F'(x))$ for all $x \in B_{\rho}(x_0)$, we observe from the definition of $\xi(T_n)$ that

$$\xi(T_{n+1}) - \xi(T_n) = \int_{T_n}^{T_{n+1}} F'(x(T_n))^\ast s_n(t)dt.$$

Due to the continuity of integrand, it follows that

$$\xi(T_{n+1}) - \xi(T_n) \in \text{Ran}(F'(x(T_n))^\ast) \subset \mathcal{N}(F'(x(T_n))^\ast)^\perp \subset \mathcal{N}(F'(x^\dagger))^\perp$$

$$= \text{Ran}(F'(x^\dagger))^\ast.$$

We may use proposition 3.10 to obtain the conclusion. □
3.3. Stability results

In this subsection, we will show the result

\[ \lim_{\delta \to 0} D \xi(T_n^\delta) \Theta(x, x(T_n^\delta)) = 0 \]

for the sequence \( \{ \xi(T_n^\delta), x(T_n^\delta) \} \) defined by algorithm 1. To this end, we need to establish some stability results so that theorem 3.11 can be applied. We will present two such results: the first one concerns the stability of the inner iteration and the second one concerns the stability of the whole algorithm.

**Lemma 3.12.** Let all the conditions in lemma 3.6 hold. Let \( \{ y^{\delta} \} \) be a sequence of noisy data satisfying \( ||y^{\delta} - y|| \leq \delta \ell \) with \( \delta \ell \to 0 \) as \( \ell \to \infty \). For any integer \( n \leq \liminf_{\ell \to \infty} n_{\delta} \), let \( \{ \xi(T_n^{\delta}), x^{\delta}(T_n^{\delta}) \} \) be the sequence generated by algorithm 1. Assume that there exists \( T_n^{\delta} \) such that

\[ T_n^{\delta} \to T_n, \quad \xi(T_n^{\delta}) \to \xi(T_n), \quad x^{\delta}(T_n^{\delta}) \to x(T_n), \]

as \( \ell \to \infty \), then for any \( T > T_n^{\delta} \), there hold

\[ \xi^{\delta}(T) \to \xi(T) \quad \text{and} \quad x^{\delta}(T) \to x(T) \quad \text{as} \quad \ell \to \infty. \]

**Proof.** Since \( T_n^{\delta} \to T_n \) as \( \ell \to \infty \), there exists sufficiently large \( \ell \) such that \( T_n^{\delta} < T \). For fixed \( n \in \mathbb{N} \), denoting \( A_n^{\delta} = F'(x^{\delta}(T_n^{\delta})) \) and \( A_n = F'(x(T_n)) \) for simplicity, we have

\[
\xi^{\delta}(T) = \xi(T_n) + \int_{T_n}^{T} (A_n^{\delta})^* \left( y^{\delta} - F(x^{\delta}(T_n^{\delta})) - A_n^{\delta} (x^{\delta}(t) - x^{\delta}(T_n^{\delta})) \right) \, dt,
\]

\[
\xi(T) = \xi(T_n) + \int_{T_n}^{T} A_n^* (y - F(x(T_n))) - A_n (x(t) - x(T_n))) \, dt.
\]

Then,

\[
\xi^{\delta}(T) - \xi(T) = \xi(T_n) + (T - T_n)(A_n^{\delta})^* \left( y^{\delta} - F(x^{\delta}(T_n^{\delta})) + A_n^{\delta} x^{\delta}(T_n^{\delta}) \right) - (T - T_n) A_n^* (y - F(x(T_n))) + A_n x(T_n))
\]

\[
+ \int_{T_n}^{T} A_n^* A_n x(t) \, dt - \int_{T_n}^{T} (A_n^{\delta})^* A_n^{\delta} x^{\delta}(t) \, dt
\]

\[= \Sigma_1 + \Sigma_2 + \Sigma_3. \]

Since \( T_n^{\delta} \to T_n \), \( (\xi^{\delta}(T_n), x^{\delta}(T_n)) \to (\xi(T_n), x(T_n)) \) as \( \ell \to \infty \), it follows that,

\[ ||\Sigma_1|| = ||\xi^{\delta}(T_n^{\delta}) - \xi(T_n)|| \to 0. \]

Applying the continuity of \( F \) and \( F' \), it is consequently \( A_n^{\delta} \to A_n \) and \( (A_n^{\delta})^* \to A_n^* \) as \( \ell \to \infty \), thus

\[ ||\Sigma_2|| = ||(T - T_n)(A_n^{\delta})^* \left( y^{\delta} - F(x^{\delta}(T_n^{\delta})) + A_n^{\delta} x^{\delta}(T_n^{\delta}) \right) - (T - T_n) A_n^* (y - F(x(T_n))) + A_n x(T_n)) || \to 0. \]
For the estimate to $\Sigma_3$, without lose of generality, we assume $T_n^{\beta_0} < T_n$, we have

$$
|\Sigma_3| = \left| \int_{T_n}^{T} A_n^* A_n x(t) dt - \int_{T_n}^{T} (A_n^h)^* A_n^h x^h(t) dt \right|
$$

$$
= \left| \int_{T_n}^{T} A_n^* A_n x(t) dt - \int_{T_n}^{T} (A_n^h)^* A_n^h x^h(t) dt - \int_{T_n}^{T} (A_n^h)^* A_n^h x^h(t) dt \right|
$$

$$
\leq \int_{T_n}^{T} \| A_n^* A_n x(t) - (A_n^h)^* A_n^h x^h(t) \| dt + \int_{T_n}^{T} \| (A_n^h)^* A_n^h x^h(t) \| dt
$$

$$
\leq \int_{T_n}^{T} \| (A_n^h)^* A_n^h \| \| x^h(t) - x(t) \| dt + \int_{T_n}^{T} \| (A_n^h)^* A_n^h - A_n^* A_n \| \| x(t) \| dt
$$

$$
+ \int_{T_n}^{T} \| (A_n^h)^* A_n^h \| \| x^h(t) \| dt
$$

$$
\leq L^2 \int_{T_n}^{T} \| \nabla \Theta^* (x^h(t)) - \nabla \Theta^* (x(t)) \| dt
$$

$$
+ B(T - T_n) \| (A_n^h)^* A_n^h - A_n^* A_n \| + L^2 B(T - T_n)
$$

$$
\leq \frac{L^2}{2c_0} \int_{T_n}^{T} \| \xi^h(t) - \xi(t) \| dt + B(T - T_n) \| (A_n^h)^* A_n^h - A_n^* A_n \| + L^2 B(T - T_n),
$$

in which the constant in the last two inequalities are based on the $\| F'(\cdot) \| \leq L$ and $\| x^h(t) \| \leq B := 2\rho + \| x_0 \|$. 

Pay attention that the last two terms in the estimate of $\Sigma_3$ converge to 0 as $\ell \to 0$, thus for arbitrary $\varepsilon > 0$, there exists $\ell_0$ such that for all $\ell \geq \ell_0$,

$$
\| \xi^h(T) - \xi(T) \| \leq \varepsilon + K \int_{T_n}^{T} \| \xi^h(t) - \xi(t) \| dt.
$$

(3.21)

where the constant $K := \frac{L^2}{2c_0}$ for simplicity.

Taking $V(T) = \int_{T_n}^{T} \| \xi^h(t) - \xi(t) \| dt$, then $V'(T) = \| \xi^h(T) - \xi(T) \|$, the (3.21) provides

$$
\frac{V'(T)}{\varepsilon + KV(T)} \leq 1.
$$

Equivalently, it holds that

$$
\int_{T_n}^{T} \frac{KV'(t)}{\varepsilon + KV(t)} dt \leq K(T - T_n^\beta).
$$

This implies

$$
\ln[\varepsilon + KV(t)] \bigg|_{T_n^\beta}^{T} \leq K(T - T_n^\beta),
$$
therefore, by using $V(T_n^h) = 0$,

$$\left\| \xi^h(T) - \xi(T) \right\| \leq \varepsilon + KV(T) \leq \varepsilon e^{T-T_m^h} \to 0, \quad \ell \to \infty,$$

we have $\xi^h(T) \to \xi(T)$ and $x^h(T) \to x(T)$. \hfill \Box

**Theorem 3.13.** Let $\mathcal{X}$ and $\mathcal{Y}$ be both Hilbert spaces. The function $\Theta$ and the operator $F$ satisfy assumptions 3.1 and 3.2 respectively. Let $\{y^h\}$ be a sequence of noisy data satisfying $\|y^h - y\| \leq \delta_l$ with $\delta_l \to 0$ as $\ell \to \infty$. Let $\{(\xi^h(T_m^n), x^h(T_m^n))\}$ with $0 \leq n \leq n_h$ be the sequence defined by algorithm 1. Assume the nonlinearity constant $\eta$ in (3.1) satisfies $\eta < \gamma < 1$ and the discrepancy constant $\tau$ in (3.6) satisfies $\tau > \frac{1+\eta}{\gamma}$. Then, for any integer $n \leq \liminf_{l \to \infty} n_h$, by taking a subsequence of $\{y^h\}$ if necessary, there is a sequence $\{(\xi(T_m), x(T_m))\}$ generated from initial pairs $(\xi_0, x_0)$ by algorithm 2, such that

$$\lim_{\ell \to \infty} D_{\xi^l(T_m^n)} \Theta(x(T_m), x^h(T_m^n)) = 0,$$

for all $0 \leq m \leq n$. Let $T_m$ be the first time such that $F(x(T_m)) = y$, then

$$\lim_{\ell \to \infty} \xi^l(T_m^n) = \xi(T_m), \quad 0 \leq m \leq \min_{\{n, n_r\}}.$$

**Proof.** If $\liminf_{l \to \infty} n_h = 0$, this is trivial since $\xi^h_0 = \xi_0$ and $x_0^h = x_0$. Therefore, we assume $\liminf_{l \to \infty} n_h \geq 1$ and complete the proof by induction.

When $n = 0$, the result is trivial again. Assume for some $0 \leq n < \liminf_{l \to \infty} n_h$, the result is true for some $\{(\xi(T_m), x(T_m))\}$ with $0 \leq m \leq n$. We are willing to show, the result is also valid for $n + 1$. We will obtain a sequence by retaining the first $m = 0, 1, \ldots, n$ terms for $\{(\xi(T_m), x(T_m))\}$ and redefine the $\xi(T_{n+1})$ and $x(T_{n+1})$ since then we can apply the algorithm for exact data to produce the remaining terms. The two cases will be considered:

**Case 1:** $F(x(T_n)) = y$, by algorithm 2 it is obvious that $\xi(T_{n+1}) = \xi(T_n)$ and $x(T_{n+1}) = x(T_n) \in B_2(\Theta)$ then $D(\Theta)$, the result is also valid. Then, we have

$$\lim_{\ell \to \infty} D_{\xi^l(T_{n+1}^a)} \Theta(x(T_{n+1}), x^h(T_{n+1}^a)) = \lim_{\ell \to \infty} D_{\xi^l(T_{n+1}^a)} \Theta(x(T_n), x^h(T_{n+1}^a)) \leq \lim_{\ell \to \infty} D_{\xi^l(T_n)} \Theta(x(T_n), x^h(T_n)) = 0.$$

**Case 2:** $F(x(T_n)) \neq y$, we denote $T_{n+1}^a$ such that

$$\|s_{n}(T_{n+1}^a)\| < \gamma \|y - F(x(T_n))\|,$$

which means $T_{n+1}^a > T_{n+1} > T_n$, therefore by the conclusion of lemma 3.12 that $x^h(T_{n+1}^a) \to x(T_{n+1}^a)$. Moreover, by the continuity of $F$ and $F'$, it is consequently $s_{n}^h(T_{n+1}^a) \to s(T_{n+1}^a)$, thus, for sufficiently large $\ell$ there must hold

$$\|x^h(T_{n+1}^a)\| < \gamma \|y^h - F(x^h(T_{n+1}^a))\|.$$

Consequently, $T_{n+1}^h < T_{n+1}^a$ for large $\ell$. Therefore, by taking a subsequence if necessary, we may denote $\lim_{n \to \infty} T_{n+1}^h = T_{n+1}$. By the definition of $T_{n+1}^h$, it is straight to know

$$\|s_{n}(T_{n+1})\| = \gamma \|y - F(x(T_n))\|,$$
we claim that
\[ \| s_n(T) \| \geq \gamma \| y - F(x(T_n)) \|, \quad T_n \leq T < T_{n+1}. \]
If not, there exists \( T_\epsilon \in [T_n, T_{n+1}) \) such that
\[ \| s_n(T_\epsilon) \| < \gamma \| y - F(x(T_n)) \|. \]
It is trivial that \( T_\epsilon \neq T_n \), thus \( T_\epsilon > T_n \) and the application of lemma 3.12 yields \( s^{\epsilon}(T_\epsilon) \rightarrow x(T_n) \), then since \( y^\epsilon \rightarrow y \) and \( F(x^\delta(T_n^\ell)) \rightarrow F(x(T_n)) \), there exists sufficiently large \( \ell \) such that
\[ \| x^\delta(T_\epsilon) \| < \gamma \| y^\epsilon - F(x^\delta(T_n^\ell)) \|. \]
This yields \( T_\epsilon > T_{n+1}^\delta \) and thus \( T_\epsilon \geq T_{n+1} \) which is a contradiction.

Using this \( T_{n+1} \) we can define \( \xi(T_{n+1}) = \xi(T_{n+1}) \) and \( x(T_{n+1}) = x(T_{n+1}) \), by the lemma 3.12 and the continuity of \( x^\delta(T) \) and \( \xi^\delta(T) \) with respect to \( T \), it is obvious that \( \{ (\xi^\delta(T_{n+1}^\ell)), x^\delta(T_{n+1}^\ell) \} \rightarrow \{ (\xi(T_{n+1})), x(T_{n+1}) \} \) as \( \ell \rightarrow \infty \). Finally,
\[ \limsup_{\ell \rightarrow \infty} D_{\xi(T_{n+1}^\ell)}(x(T_{n+1}), x^\delta(T_{n+1}^\ell)) = \Theta(x(T_{n+1})) \liminf_{\ell \rightarrow \infty} \Theta(x^\delta(T_{n+1}^\ell)) \]
\[ = \Theta(x(T_{n+1})) - \liminf_{\ell \rightarrow \infty} \Theta(x^\delta(T_{n+1}^\ell)) = \Theta(x(T_{n+1})) - \Theta(x(T_{n+1})) = 0. \]
The proof is complete. \( \square \)

**Remark 3.14.** In theorem 3.13, when \( F(x(T_n)) = y \) for some \( n \leq \liminf_{\ell \rightarrow \infty} n_{\delta_\ell} \), we can not guarantee that \( \xi^\delta(T_{n+1}^\ell) = \xi(T_{n+1}) \). However, we can guarantee the convergence in Bregman distance, which is enough for our purpose.

### 4. Convergence rate of the asymptotical method

In this section, we discuss the local convergence and convergence rate of algorithm 1 with noisy data under following Hörder type conditional assumption:

**Assumption 4.1.** Denote \( \nu \in (0, 2] \), assume there exists a constant \( R_F > 0 \) such that
\[ D_\nu(\Theta(x, \bar{x}), \nu) \leq R_F \| F(x) - F(\bar{x}) \|_\nu, \quad (4.1) \]
is valid for all \( \bar{x}, x \in \mathcal{U}_M, \bar{x} \in D(\partial \Theta) \) and \( \bar{x} \in \partial \Theta(\bar{x}) \), in which the set \( \mathcal{U}_M \) is defined as
\[ \mathcal{U}_M := \{ x \in D(F) \cap D(\Theta) : \Theta(x) \leq M \}. \]

**Theorem 4.2.** Let \( \mathcal{X} \) and \( \mathcal{Y} \) be Hilbert spaces, let the functional \( \Theta \) and operator \( F \) satisfy assumptions 3.1 and 3.2 respectively. Assume the nonlinearity constant \( 0 < \eta < 1/3 \) and suitable choosing \( \eta \leq \gamma < 1 - 2\eta \). Let the sequence \( \{ (\xi^\delta(T_n^\ell), x^\delta(T_n^\ell)) \} \) be defined via algorithm 1, and the iterates will be stopped by the discrepancy principle (3.6) with \( \tau > \frac{1}{\gamma - \eta} \). Then for
all $0 \leq n \leq n_\delta$, the iterates $x^\delta(T_n^\delta)$ can be uniformly bounded by a constant $M$. Assume the Hölder type conditional assumption (4.1) is valid, then,

$$\|x^\delta_n - x^\delta\| \leq \sqrt{R_f(\tau + 1)} \delta^{\tilde{\delta}}.$$  

**Proof.** Let $\hat{x}$ be arbitrary solution to (1.1) satisfying $D_{\delta \Theta}(\hat{x}, x_0) \leq c_0 \rho^2$, referring to definition of Bregman distance, for $n \leq n_\delta$,

$$\begin{align*}
\Theta(x^\delta(T_n^\delta)) &= \Theta(\hat{x}) + \langle \xi^\delta(T_n^\delta), x^\delta(T_n^\delta) - \hat{x} \rangle - D_{\xi^\delta(T_n^\delta)}(\Theta(\hat{x}), x^\delta(T_n^\delta)) \\
&\leq \Theta(\hat{x}) + \langle \xi^\delta(T_n^\delta), x^\delta(T_n^\delta) - \hat{x} \rangle \\
&= \Theta(\hat{x}) + \langle \xi^\delta(T_n^\delta) - \xi_0, x^\delta(T_n^\delta) - \hat{x} \rangle + \langle \xi_0, x^\delta(T_n^\delta) - \hat{x} \rangle. 
\end{align*}$$

(4.2)

Note that

$$\begin{align*}
\langle \xi^\delta(T_n^\delta) - \xi_0, x^\delta(T_n^\delta) - \hat{x} \rangle &= \sum_{m=0}^{n-1} \langle \xi^\delta(T_m^\delta) - \xi_0, x^\delta(T_m^\delta) - \hat{x} \rangle \\
&= \sum_{m=0}^{n-1} \int_{T_m^\delta}^{T_{m+1}^\delta} \langle \frac{d\xi^\delta(t)}{dt}, x^\delta(T_m^\delta) - \hat{x} \rangle \, dt \\
&= \sum_{m=0}^{n-1} \int_{T_m^\delta}^{T_{m+1}^\delta} \langle \xi^\delta(x^\delta(T_m^\delta)), x^\delta(T_m^\delta) - \hat{x} \rangle dt.
\end{align*}$$

Recalling that $\tau \delta \leq \|F(x^\delta(T_m^\delta)) - y^\delta\|$ for $0 \leq m < n \leq n_\delta$ and utilizing the monotonicity of the residual, it is consequently that,

$$\begin{align*}
\|F'(x^\delta(T_m^\delta))(x^\delta(T_m^\delta) - \hat{x})\| &\leq \|F'(x^\delta(T_m^\delta))(x^\delta(T_m^\delta) - x^\delta(T_n^\delta))\| \\
&\quad + \|F'(x^\delta(T_n^\delta))(\hat{x} - x^\delta(T_n^\delta))\| \\
&\leq (1 + \eta) \|F'(x^\delta(T_m^\delta) - y^\delta)\| + (1 + \eta)\delta \|F(x^\delta(T_n^\delta) - y^\delta)\| \\
&\leq (1 + \eta)(2 + 1/\gamma)\|F(x^\delta(T_m^\delta)) - y^\delta\| \\
&\quad + (1 + \eta)\|F(x^\delta(T_n^\delta)) - y^\delta\| \\
&\leq (1 + \eta)(3 + 1/\gamma)\|s^\delta(t)\|, \quad t \in [T_m^\delta, T_{m+1}^\delta].
\end{align*}$$

Therefore, combining with estimate (3.13), we have,

$$\begin{align*}
\|\langle \xi^\delta(T_n^\delta) - \xi_0, x^\delta(T_n^\delta) - \hat{x} \rangle\| &\leq \frac{(1 + \eta)(3 + 1)}{\gamma \tau - 1 - \eta - \eta \tau} D_{\delta \Theta}(\hat{x}, x_0) \\
&\leq \frac{(1 + \eta)(3 + 1)}{\gamma \tau - 1 - \eta - \eta \tau} c_0 \rho^2.
\end{align*}$$
Combining with estimate in (4.2), it is follows that,
\[
\Theta(x^i(T_{n_0})) \leq \Theta(\hat{x}) + \frac{(1 + \eta)(3\tau + 1)}{\gamma\tau - 1 - \eta - \eta\tau}c_0\rho^2 + ||\xi_0||||x^i(T_{n_0}) - \hat{x}||
\leq \Theta(\hat{x}) + \frac{(1 + \eta)(3\tau + 1)}{\gamma\tau - 1 - \eta - \eta\tau}c_0\rho^2 + 3\rho||\xi_0||.
\]
(4.3)
This shows that the sequence \( \{\Theta(x^i(T_{n_0}))\}_{n=0}^\infty \) can be uniformly bounded by a constant
\[
M := \Theta(\hat{x}) + \frac{(1 + \eta)(3\tau + 1)}{\gamma\tau - 1 - \eta - \eta\tau}c_0\rho^2 + 3\rho||\xi_0||.
\]
Then, note that \( \hat{x} \in U_M \), the assumption 4.1 can be applied, yielding,
\[
\|x^i - x^i(T_{n_0})\|^2 \leq \frac{1}{c_0}D_R\Theta(x^i(T_{n_0})), x^i\)
\[
\leq \frac{R_e}{c_0}\|F(x^i(T_{n_0})) - F(x^i)\|''
\leq \frac{R_e}{c_0}(\tau + 1)^\delta\delta'.
\]
The proof is complete. \( \square \)

5. Numerical example

5.1. Discretization of asymptotical regularization

In this section we recall the RK type iterative methods which are regarded as the discretization of asymptotical method.

RK methods are a class of one-step methods for the numerical solutions of the initial value problems for ordinary differential equations,
\[
\frac{dw(t)}{dt} = \Psi(t, w(t)), \quad t > 0, \quad w(t) = w_0,
\]
with given \( \Psi : [0, \infty] \times X \rightarrow X \) and \( w_0 \in X \).

A RK method of \( s \)-stage provides approximations \( w_n \) to \( w(t_n) \) with \( t_n = \sum_{k=1}^n \triangle t_k \) such that
\[
w_{n+1} = w_n + \triangle t_n \sum_{i=1}^s b_i \Psi(t + c_i \triangle t_n, k_i), \quad \text{(5.1)}
\]
\[
k_i = w_n + \triangle t_n \sum_{j=1}^s a_{ij} \Psi(t + c_j \triangle t_n, k_j). \quad \text{(5.2)}
\]

The given coefficients \( A = (a_{ij}) \in \mathbb{R}^{s \times s}, \quad b = (b_1, b_2, \ldots, b_s)^T \) and \( c = (c_1, c_2, \ldots, c_s)^T \) determine the particular method. The method is called \textit{explicit} if \( A \) is strictly lower triangular matrix. Otherwise, the method is called \textit{implicit}, since linear or nonlinear algebraic equations have to be solved to calculate \( k_i, i = 1, \ldots, s \). Usually, the specific instance of RK methods can
be presented by Butcher tableau, see table 1 or by triple \((A, b, c)\) in [8]. We only consider consistent \(s\)-stage RK method, i.e., \(\sum_{i=1}^{s} b_i = 1\). We refer [5, 25] for other type RK regularization methods.

Applying RK method (5.1) and (5.2) to autonomous ODE (3.3), i.e.,

\[
Ψ(ξ^n(t)) = F(x^s(T)) (y^s - F(x^s(T)) - F(x^s(T)) (x^s(t) - x^s(T))),
\]

we obtain

\[
ξ_{n,k+1}^δ = ξ_{n,k}^δ + ∆t^δn\sum_{j=1}^{s} b_j F(x^s(T_n^j)) (y^s - F(x^s(T_n^j)) - F(x^s(T_n^j)) (x^s(T_n^j) - x^s(T_n^j))),
\]

\[
k_{n,i}^δ = ξ_{n,k}^δ + ∆t^δn\sum_{j=1}^{s} a_{i,j} F(x^s(T_n^j)) (y^s - F(x^s(T_n^j)) - F(x^s(T_n^j)) (x^s(T_n^j) - x^s(T_n^j))).
\]

(5.3)

Firstly, utilizing the explicit Euler method \((s = 1, A = 0, b = 1)\), we will have the iterative methods, which is consistent with inexact Newton–Landweber iteration [19]. The method is described in details in algorithm 3.

We also consider the explicit two-stage RK method for comparison, for \(s = 2\), there is a family of such methods, parameterized by \(α\) and given in the table 2, in this family, \(α = 1\) gives the midpoint method, and \(α = 1\) is Heun’s method. In our numerical example, we set \(α = 4/5\).

Furthermore, we consider implicit Euler method \((s = 1, A = 1, b = 1)\) to discrete (3.3), yielding

\[
ξ_{n,k+1}^δ = ξ_{n,k}^δ + ∆t^δn F(x^s(T_n^j))^\ast (y^s - F(x^s(T_n^j)))
\]

\[
- F(x^s(T_n^j)) (x_{n,k+1}^δ - x^s(T_n^j)),
\]

\[
x_{n,k+1}^δ = ∇Θ^\ast (ξ_{n,k+1}^δ).
\]

(5.4)

Referring to (2.6), we approximate

\[
x_{n,k+1}^δ - x^s(T_n^j) = ∇Θ^\ast (ξ_{n,k+1}^δ) - ∇Θ^\ast (ξ_{n,k+1}^δ) ≈ \frac{1}{2ξ_0}(ξ_{n,k+1}^δ - ξ_{n,k+1}^δ).
\]

Let \(Π_n = (I + \frac{ξ_{n,k}^δ F(x^s(T_n^j))^\ast F(x^s(T_n^j)))}{ξ_0})\), then we have

\[
ξ_{n,k+1}^δ = ξ_{n,k}^δ + Π_n^{-1} (ξ_{n,k}^δ - ξ_{n,k}^δ + ∆t^δn F(x^s(T_n^j))^\ast (y^s - F(x^s(T_n^j)))).
\]
For the simplicity, we take $k = 0$ then the method is reduced as

$$
\xi^i(T^i_{n+1}) = \xi^i(T^i_n) + \Delta t^i_n \Pi_n^{-1} \left( F^i(x^i(T^i_n))^r (y^i - F^i(x^i(T^i_n))) \right),
$$

which seems like the Levenberg–Marquardt method [5, 13]. The method is described in details in Algorithm 4.

Algorithm 4. Implicit Euler method with convex penalty terms.

Input: Noisy data $y^i$, and $\tau > 1$.
Initial guess: $\xi^i(T^i_0) = \xi_0 \in \mathcal{X}$ and set $x_0 = \arg\min_{x \in \mathcal{X}} \{ \Theta(x) - \langle \xi_0, x \rangle \}$.
Set $n = 0$, $\gamma = 1$.
While $\| F^i(x^i(T^i_n)) - y^i \| > \tau \delta$
    Set $k = 0$, define $\xi^i_n = \xi^i(T^i_n)$.
    While $\| s^i_{nk} \| > \gamma \| y^i - F^i(x^i_n) \|
        \begin{align*}
        x^i_{nk+1} &= y^i - F^i(x^i_n) + \Delta t^i_n \Pi_n^{-1} \left( F^i(x^i_n)^r (y^i - F^i(x^i_n)) \right), \\
        k &= k + 1.
        \end{align*}
    \end{align*}
End
End
Record the terminated index $n_\tau$.
Output: The approximated solution $x^i(T^i_n)$.

Table 2. Particular two-stage RK formula.

| $t$ | $\alpha$ | $\frac{1 - \frac{1}{2\alpha}}{2\alpha}$ |
|-----|----------|----------------------------------|
| 0   |          |                                  |

Algorithm 3. Explicit Euler method (inexact Newton–Landweber iteration) with convex penalty terms.

Input: Noisy data $y^i$, suitably chosen numbers $0 < \gamma < 1$, $\mu_1$, $\mu_2 > 0$ and $\tau > 1$.
Initial guess: $\xi^i(T^i_0) = \xi_0 \in \mathcal{X}$ and set $x_0 = \arg\min_{x \in \mathcal{X}} \{ \Theta(x) - \langle \xi_0, x \rangle \}$.
Set $n = 0$, $\gamma = 1$.
While $\| F^i(x^i(T^i_n)) - y^i \| > \tau \delta$
    Set $k = 0$, define $\xi^i_n = \xi^i(T^i_n)$.
    While $\| s^i_{nk} \| > \gamma \| y^i - F^i(x^i_n) \|$
        \begin{align*}
        x^i_{nk+1} &= y^i - F^i(x^i_n) + \Delta t^i_n \Pi_n^{-1} \left( F^i(x^i_n)^r (y^i - F^i(x^i_n)) \right), \\
        k &= k + 1.
        \end{align*}
End
End
Record the terminated index $n_\tau$.
Output: The approximated solution $x^i(T^i_n)$.
Finally, we consider the autonomous ODE (1.3), inserting a uniformly convex functional $\Theta$ that is Landweber-type scheme
\[
\Psi(\xi^\delta(t)) = F'(x^\delta(t)) (y^\delta - F(x^\delta(t))), \quad x^\delta(t) = \nabla \Theta^*(\xi^\delta(t)). \quad (5.6)
\]
The one-stage explicit RK discretization yields Landweber-type iteration in [21], i.e.,
\[
\xi_{n+1}^\delta = \xi_n^\delta - \Delta t^\delta \delta_n F(x_n^\delta) (F(x_n^\delta) - y^\delta),
\]
\[
x_{n+1}^\delta = \nabla \Theta^*(\xi_{n+1}^\delta).
\]
We will consider these four iteration methods and compare them.

A key ingredient for the numerical implementation is the determination of $x = \nabla \Theta^*(\xi)$ for any given $\xi \in X^*$ which is equivalent to solving the minimization problem
\[
x = \arg \min_{z \in X} \{ \Theta(z) - \langle \xi, z \rangle \}. \quad (5.7)
\]
For some choices of $\Theta$, this minimization problem can be easily solved numerically. For instance, when $X = L^2(\Omega)$ and the sought solution is piecewise constant, we may choose
\[
\Theta(x) = \frac{1}{2} \beta \| x \|_2^2 + |x|_{TV}, \quad (5.8)
\]
with a constant $\beta > 0$, where $|x|_{TV}$ denotes the TV of $x$. Then the minimization problem (5.7) becomes the TV denoising problem
\[
x = \arg \min_{z \in L^2(\Omega)} \left\{ \frac{1}{2\beta} \| z - \beta \xi \|_2^2 + |z|_{TV} \right\}, \quad (5.9)
\]
which is nonsmooth and convex. The fast iterative shrinkage-thresholding algorithm [3], the alternating direction method of multipliers (ADMM) [4], and the primal dual hybrid gradient method [40] can be applied to solve it.

5.2. Parameter identification problem

As a first example, we consider a parameter identification problem arising from reaction diffusion process. If the endpoints $x = 0, x = \pi$ of a homogeneous solid rod contacts with liquid media, then the convective heat transfer occurs. The temperature field $u(x, t)$ of the heat conduction process during time $[0, T]$ can be modeled by
\[
\begin{cases}
\frac{\partial u}{\partial t} - \alpha^2 \frac{\partial^2 u}{\partial x^2} = 0, & x \in (0, \pi), \ t \in (0, T); \\
u_x(0, t) = f(t), & u_x(\pi, t) + \sigma(t) u(\pi, t) = \varphi(t), \quad t \in [0, T]; \\
u(x, 0) = u_0(x), & x \in [0, \pi].
\end{cases} \quad (5.10)
\]
The function $\sigma(t) > 0$ represents the corrosion damage, which is interpreted as a Robin coefficient of energy exchange. We consider the reconstruction of $\sigma(t)$ from
\[
u(0, t) = g^\delta(t), \quad t \in [0, T],
\]
in which $g^\delta$ are noisy observations, satisfying $\| g^\delta - g \|_2 \leq \delta$. 

Define
\[ D := \{ \sigma \in L^2([0, T]), 0 < \sigma_- \leq \sigma \leq \sigma_+, \text{ a.e. in } [0, T] \}, \] (5.11)
and define the nonlinear operator \( F: \sigma \in D \rightarrow u[\sigma](0, t) \in L^2([0, T]) \), the above Robin coefficient inversion problem reduces to solving the equation \( F(\sigma) = g \). The well-posedness of \( F \) and the uniqueness of the inverse problem have been discussed in [38]. The Fréchet derivative \( F'(\sigma)h(t) = w(0, t) \) is given by
\[
\begin{cases}
w_t - \alpha^2 w_{xx} = 0, & x \in (0, \pi), t \in (0, T); \\
w_t(0, t) = 0, & w(\pi, t) + \sigma(t)w(\pi, t) = -h(t)u[\sigma](\pi, t), & t \in [0, T]; \\
w(x, 0) = 0, & x \in [0, \pi].
\end{cases}
\]
In addition, the adjoint of the Fréchet derivative is given by \( F'(\sigma)^* \zeta(t) = u[\sigma](\pi, t)v(\pi, t) \), where \( v(x, t) \) solves the adjoint system
\[
\begin{cases}
-v_t - \alpha^2 v_{xx} = 0, & x \in (0, \pi), t \in (0, T); \\
v_t(0, t) = \zeta(t), & v(\pi, t) + \sigma(t)v(\pi, t) = 0, & t \in [0, T]; \\
v(x, T) = 0, & x \in [0, \pi].
\end{cases}
\]
In our numerical simulations, we take \( a = 5 \), \( T = 1 \), and assume the sought Robin coefficient is
\[
\sigma^\dagger(t) = \begin{cases}
1.5, & 0 \leq t \leq 0.1563, \\
2, & 0.1563 < t \leq 0.3125, \\
1.2, & 0.3125 < t \leq 0.5469, \\
2.5, & 0.5469 < t \leq 0.6250, \\
1.8, & 0.6250 < t \leq 0.7813, \\
1, & 0.7813 < t \leq 1.
\end{cases}
\]
We also assume that the exact solution of the forward problem with \( \sigma = \sigma^\dagger \) is
\[
u(x, t) = e^{-\alpha^2 t} \sin x + x^2 + 2\alpha^2 t.
\] (5.12)
Through such explicit expression we can obtain the expression of \( (f(t), u_0(x), \varphi(t)) \) and the inversion input \( g(t) = u(0, t) \).

In our numerical simulations, we use a discretized version of the operator \( F \) with an equidistant grid of \( N = 65 \) points with coordinates \( \tau \in \mathbb{R}^{65} \). We add random Gaussian noise on \( g \) to produce noisy data \( \tilde{g} \) satisfying \( ||g - \tilde{g}||_2 \leq \delta \) with different noise level. In order to capture the feature of the sought Robin coefficient, we take \( \Theta \) as in (5.8) with the constant \( \beta = 10 \), which implies that \( c_0 = 1/(2\beta) = 0.05 \) in (2.3).

We test Landweber-type (5.6), explicit Euler method (inexact NL iteration, algorithm 3), the two-stage RK method and implicit Euler method (algorithm 4). In order for comparison, we use the same initial guesses \( x_0 = \xi_0 = 0 \) and same stopping parameter \( \tau \) in discrepancy principle (3.6). Two different \( \tau = 2.01 \) and \( \tau = 1.05 \) are both tested. The \( \gamma \) in explicit Euler method and
boundary element method. For the step size, we adaptively choose them via two-stage RK are both chosen as $\gamma = 0.95$. The forward PDE are solved approximately by a boundary element method. For the step size, we adaptively choose them via

$$\Delta t^n = \min \left\{ \frac{\mu_0 \| F(x^n_k) - y^k \|^2}{\| F'(x^n)(F(x^n_k) - y^k) \|^2}, \mu_1 \right\},$$

in Landweber-type method, and

$$\Delta t_{n,k} = \min \left\{ \frac{\mu_0 \| s_{n,k} \|^2}{\| F'(x^n(T^n_k))'(s_{n,k}) \|^2}, \mu_1 \right\},$$

in explicit Euler method and two-stage RK method, where the constants $\mu_0 = (1 - 1/\tau)/\beta$, $\mu_1 = 0.1$, thus $\mu_0 \approx 0.025$ for $\tau = 2.01$ and $\mu_0 \approx 0.0024$ for $\tau = 1.05$. Referring to the A-stable properties of implicit Euler, the discrete step of the constructed iterative algorithm can be chosen arbitrarily large theoretically. However, a larger discrete step can lead to an overshoot solution as [5] observed. We take constant step size $\Delta t^n = 0.04$ in implicit Euler method. One can also consider the choice strategy as in [13]. The computational results and details are contained in the following table 3 and figure 1. Compared with Landweber-type method, the other three methods discretized by our proposed asymptotical regularization (3.3) show the superiorities in both reconstruction results and convergence rates. The two-stage RK method usually spends more time than explicit Euler methods since two-stage RK method needs solve one more minimization problem (5.9) in every inner loop. For the explicit RK methods (explicit Euler method and two-stage RK) the step sizes are limited by some constant and the larger step size can reduce the numbers of iteration steps, in which the residual errors are not monotonically decreasing. The implicit Euler method requires solving a linear system, which is implemented by conjugate gradient method $\text{pcg}$ in Matlab, although the amount of calculation in each step is large, the computational efficiency is still dominant in comparison.

### Table 3. Quantitative analysis of reconstruction results under different noisy data for Robin coefficient inversion.

| $\delta$ (%) | Method        | $\tau = 2.01(\mu_0 = 0.025)$ | $\tau = 1.05(\mu_0 = 0.0024)$ |
|--------------|---------------|-------------------------------|-------------------------------|
|              | $n_s$ | L2-error | Time (s) | $n_s$ | L2-error | Time (s) |
| 1            | Landweber-type | 922 | $4.4038 \times 10^{-2}$ | 3.25 | 3185 | $1.9650 \times 10^{-2}$ | 10.90 |
|              | Explicit Euler | 145 | $4.0898 \times 10^{-2}$ | 3.89 | 207 | $1.6752 \times 10^{-2}$ | 11.75 |
|              | Two-stage RK   | 152 | $4.1670 \times 10^{-2}$ | 5.58 | 234 | $1.5821 \times 10^{-2}$ | 21.02 |
|              | Implicit Euler | 63  | $3.3917 \times 10^{-2}$ | 2.95 | 83  | $1.5938 \times 10^{-2}$ | 3.83 |
| 0.5          | Landweber-type | 1141 | $2.4517 \times 10^{-2}$ | 4.01 | 3523 | $1.1621 \times 10^{-2}$ | 12.18 |
|              | Explicit Euler | 177  | $2.2667 \times 10^{-2}$ | 4.69 | 226 | $1.0197 \times 10^{-2}$ | 13.49 |
|              | Two-stage RK   | 142  | $2.3559 \times 10^{-2}$ | 6.63 | 254 | $1.0021 \times 10^{-2}$ | 22.84 |
|              | Implicit Euler | 73  | $1.9533 \times 10^{-2}$ | 3.34 | 95  | $9.3836 \times 10^{-3}$ | 4.31 |
| 0.1          | Landweber-type | 1462 | $7.3491 \times 10^{-3}$ | 5.16 | 5123 | $2.1480 \times 10^{-3}$ | 17.76 |
|              | Explicit Euler | 285  | $7.1699 \times 10^{-3}$ | 5.98 | 281 | $1.3786 \times 10^{-3}$ | 76.93 |
|              | Two-stage RK   | 335  | $6.9815 \times 10^{-3}$ | 8.71 | 311 | $1.4705 \times 10^{-3}$ | 150.41 |
|              | Implicit Euler | 101  | $5.7318 \times 10^{-3}$ | 4.81 | 2354 | $7.7653 \times 10^{-4}$ | 98.18 |
| 0.05         | Landweber-type | 1789 | $3.8083 \times 10^{-3}$ | 6.90 | 44583 | $5.9489 \times 10^{-4}$ | 154.36 |
|              | Explicit Euler | 334  | $3.6690 \times 10^{-3}$ | 7.25 | 310 | $5.0374 \times 10^{-4}$ | 245.84 |
|              | Two-stage RK   | 392  | $3.5334 \times 10^{-3}$ | 10.60 | 345 | $4.7490 \times 10^{-4}$ | 319.76 |
|              | Implicit Euler | 127  | $2.5268 \times 10^{-3}$ | 5.78 | 2551 | $4.1704 \times 10^{-4}$ | 102.42 |
5.3. Diffuse optical tomography

In the second example, we consider diffuse optical tomography (DOT) problem which is a noninvasive imaging technique and typically used in breast cancer detection or monitoring oxygenation state of blood [1]. The DOT inverse problem is to identify the optical coefficients of the tissues by using the knowledge of impulse near-infrared sources, the model of photon
propagation and external detection measurements. The transport behavior of light in the tissues with high scattering property is usually described by diffusion approximation equation [1]

\[
-\nabla \cdot (\kappa(x)\nabla \Phi(x)) + \mu_a(x)\Phi(x) = q(x), \quad x \in \Omega
\]

\[
\Phi(\zeta) + 2\nu\kappa(\zeta)\frac{\partial \Phi}{\partial \Omega}(\zeta) = 0, \quad \zeta \in \partial \Omega,
\]

(5.13)

where \( \Omega \subset \mathbb{R}^n \), \( (n = 2, 3) \) be a bounded domain with \( C^{1,1} \) boundary \( \partial \Omega \), \( \Phi \) is the photon density, \( \kappa(x) = 1/3(\mu_a(x) + \mu_s'(x)) \) is the diffusion coefficient at position \( x \), \( \mu_a(x) \) and \( \mu_s'(x) \) are absorption and reduced scattering coefficients respectively. The incoming flux is denoted by \( q(x) \), \( n \) is the outer unit normal vector at \( \partial \Omega \) and \( \nu \) is a factor resulting from the refractive index mismatch at the boundary. The measurable quantity in optical tomography is given by \( y(\zeta) = -\kappa(\zeta)\frac{\partial \Phi}{\partial \Omega}(\zeta) \).

In the following experiment, we consider to recover the absorption coefficient \( \mu_a \) assuming that \( \mu_s' \) is known.

Define the admissible set

\[
\mathcal{D}(F) := \{ \mu_a \in L^2(\Omega) : \mu_1 \leq \mu_a \leq \mu_2, \quad \text{for some} \quad \mu_1, \mu_2 > 0 \}. \quad (5.14)
\]

The forward problem in DOT can be described as a nonlinear operator from \( \mu_a \) to measurements \( y \), i.e.,

\[
F : \mathcal{D}(F) \to L^2(\partial \Omega), \mu_a \to y,
\]

(5.15)

which builds a nonlinear equation

\[
F(\mu_a) = y.
\]

The forward operator \( F \) is well-defined, continuous, weakly closed and Lipschitz continuous in a neighborhood of the true parameter \( \mu_a^t \). The derivative of \( F \) at \( \mu_a \) is uniformly bounded [12].

The numerical solutions of (5.13) are obtained through the shared software Nifast [10]. We consider to use 16 equidistant light sources and 16 detectors around the boundary of circular region with radius 43 mm, including one circular anomaly located at the upper right of the circular, shown in figure 2(a). The optical coefficients of background are set as \( \mu_a = 0.01 \) mm\(^{-1} \) and \( \mu_s' = 1 \) mm\(^{-1} \), respectively. The optical coefficients of the anomaly are set as \( \mu_a = 0.03 \) mm\(^{-1} \) and \( \mu_s' = 1 \) mm\(^{-1} \), respectively. The noisy data are constructed by \( y^0 = y(\mu_a^0 + \delta \cdot \epsilon) \), where \( \mu_a^0 \) is the true absorption coefficient, \( y(\mu_a^0) \) denotes the exact measurement data, \( \delta \) is a relative noise level, and the random variable \( \epsilon \sim N(0, 1) \).

For numerical calculation, the finite element method is used to solve the equation (5.13). The circular region is discretized into 3418 triangular elements for forward calculation and 1230 triangular elements for inverse problem to avoid inverse crime. The distribution of \( \mu_a \) has the property of piecewise constant, then we choose the convex penalty as (5.8) with \( \beta = 30 \). Under the unstructured discretization of the circular region, anisotropic TV regularization is introduced to tackle the non-differentiable TV penalty, which is regarded as the discrete differential operators for TV regularization [27]. The anisotropic TV regularization is defined as

\[
\int_{\Omega} (|\partial_x z| + |\partial_y z|) \, dx \, dy = \| D_x z^h \|_1 + \| D_y z^h \|_1,
\]

(5.16)

where \( \partial_x \) and \( \partial_y \) are continuous partial derivatives along the \( x \) and \( y \) directions, respectively. The vector \( z^h \) is the discretization of \( z \) in finite element mesh. \( D_x \) is a matrix which acts on \( z^h \).
Figure 2. Numerical experiment on DOT under $\delta = 0.2\%$. (a) True $\mu_a$; (b) L2-error curves; (c) recovered $\mu_a$ by Landwerber-type method; (d) recovered $\mu_a$ by explicit Euler method; (e) recovered $\mu_a$ by two-stage RK method; (f) recovered $\mu_a$ by implicit Euler method.
Table 4. Absorption coefficient $\mu_a$ reconstruction in DOT ($\beta = 30$, $\tau = 6$).

| $\delta$ (%) | Method             | $n_\delta$ | $\|c_a^\delta - c^1\|_{L_2}$ | CPU time (s) |
|--------------|-------------------|------------|-------------------------------|--------------|
| 0.2          | Landweber-type    | 149        | $8.9931 \times 10^{-2}$      | 50.74        |
|              | Explicit Euler    | 6          | $8.4102 \times 10^{-2}$      | 49.23        |
|              | Two-stage RK      | 4          | $8.4688 \times 10^{-2}$      | 82.71        |
|              | Implicit Euler    | 3          | $7.9566 \times 10^{-2}$      | 2.54         |
| 0.5          | Landweber-type    | 53         | $1.0873 \times 10^{-1}$      | 18.84        |
|              | Explicit Euler    | 4          | $9.7969 \times 10^{-2}$      | 15.47        |
|              | Two-stage RK      | 2          | $9.7267 \times 10^{-2}$      | 18.37        |
|              | Implicit Euler    | 3          | $7.9857 \times 10^{-2}$      | 2.58         |
| 1            | Landweber-type    | 17         | $1.2327 \times 10^{-1}$      | 6.57         |
|              | Explicit Euler    | 3          | $1.1759 \times 10^{-1}$      | 4.79         |
|              | Two-stage RK      | 2          | $1.0831 \times 10^{-1}$      | 12.22        |
|              | Implicit Euler    | 2          | $8.0769 \times 10^{-2}$      | 1.93         |

leads to the discrete partial derivative of $z^h$ along direction $x$. Similarly, $D_y$ is the derivative matrix along direction $y$. The ADMM method is used for minimization problems.

We present the reconstruction results of Landweber-type method, explicit Euler method, two-stage RK method and implicit Euler method for $\mu_a$ identification. The initial guesses $x_0 = 0$, $\xi = 0$ and discrepancy principle $\tau = 6$. The $\gamma$ in inexact NL and two-stage RK are both chosen as $\gamma = 0.4$. The step size for Landweber-type method, explicit Euler and two-stage RK are all be adaptively chosen, the step size in implicit Euler is chosen as $\triangle \mu_a = 10$. As shown in figure 2, all these methods can accurately locate the circular anomaly, well quantify the value of $\mu_a$ and provide clear recovered background. Figure 2(b) and table 4 indicate that implicit Euler method has the advantages in computational efficiency and convergence rate.

Acknowledgments

The authors would like to thank the editors and the anonymous referees for their constructive comments and suggestions, which have helped us to improve this manuscript. The authors are grateful to Qinian Jin (Australian National University, Australia) for some useful comments. The work of M Zhong is supported by the National Natural Science Foundation of China (No. 11871149). The work of W Wang is supported by the National Natural Science Foundation of China (Nos. 12071184, 11871180) and The Top-level Talent Project of Zhejiang Province. The work of S Tong is supported by the National Natural Science Foundation of China (No. 11901373) and the Fundamental Research Funds for the Central Universities (No. GK202003007).

Data availability statement

All data that support the findings of this study are included within the article (and any supplementary files).
ORCID iDs

Min Zhong  ♦  https://orcid.org/0000-0003-0497-7743
Wei Wang  ♦  https://orcid.org/0000-0002-3316-8438
Shanshan Tong  ♦  https://orcid.org/0000-0002-4980-6844

References

[1] Arridge S R 1999 Optical tomography in medical imaging *Inverse Problems* 15 R41
[2] Boj R, Dong G Z, Elbahl P and Scherzer O 2021 Convergence rates of first-and higher-order dynamics for solving linear ill-posed problems *Found. Comput. Math.* [https://doi.org/10.1007/s10208-021-09536-6](https://doi.org/10.1007/s10208-021-09536-6)
[3] Beck A and Teboulle M 2009 Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems *IEEE Trans. Image Process.* 18 2419–34
[4] Boyd S *et al* 2011 Distributed optimization and statistical learning via the alternating direction method of multipliers *Found. Trends Mach. Learn.* 3 1–122
[5] Bckmann C and Pornsawad P 2008 Iterative Runge–Kutta-type methods for nonlinear ill-posed problems *Inverse Problems* 24 025002
[6] Cheng J and Yamamoto M 2000 One new strategy for *a priori* choice of regularizing parameters in Tikhonov’s regularization *Inverse Problems* 16 L31–8
[7] Cheng J, Hofmann B and Lu S 2014 The index function and Tikhonov regularization for ill-posed problems *J. Comput. Appl. Math.* 265 110–9
[8] Diehl P and Bornemann F 2002 *Scientific Computing with Ordinary Differential Equations* (Berlin: Springer)
[9] de Hoop M V, Qiu L and Scherzer O 2012 Local analysis of inverse problems: Hölder stability and iterative reconstruction *Inverse Problems* 28 045001
[10] Dehghani H, Eames M E, Yalavarthy P K, Davis S C, Srinivasan S, Carpenter C M, Pogue B W and Paulsen K D 2009 *Commun. Numer. Methods Eng.* 25 711–32
[11] Egger H and Hofmann B 2018 Tikhonov regularization in Hilbert scales under conditional stability assumptions *Inverse Problems* 34 115015
[12] Egger H and Schlottbom M 2010 Analysis and regularization of problems in diffuse optical tomography *SIAM J. Math. Anal.* 42 1934–48
[13] Fu Z, Han B and Chen Y 2022 Levenberg–Marquardt method with general convex penalty for nonlinear inverse problems *J. Comput. Appl. Math.* 404 113771
[14] Gu R, Han B and Chen Y 2019 Fast subspace optimization method for nonlinear inverse problems in Banach spaces with uniformly convex penalty terms *Inverse Problems* 35 125011
[15] Hanke M 1997 Regularizing properties of a truncated Newton-CG algorithm for nonlinear inverse problems *Numer. Funct. Anal. Optim.* 18 971–93
[16] Hanke M 1997 A regularizing Levenberg–Marquardt scheme, with applications to inverse groundwater filtration problems *Inverse Problems* 13 79–95
[17] Hanke M, Neubauer A and Scherzer O 1995 A convergence analysis of the Landweber iteration for nonlinear ill-posed problems *Numer. Math.* 72 21–37
[18] Jin Q 2013 Solving nonlinear inverse problems by evolution equations based on Gauss–Newton methods *Appl. Anal.* 92 527–48
[19] Jin Q 2015 Inexact Newton–Landweber iteration in Banach spaces with nonsmooth convex penalty terms *SIAM J. Numer. Anal.* 53 2389–413
[20] Jin Q 2016 Landweber–Kaczmarz method in Banach spaces with inexact inner solvers *Inverse Problems* 32 104005
[21] Jin Q and Wang W 2013 Landweber iteration of Kaczmarz type with general non-smooth convex penalty functionals *Inverse Problems* 29 085011
[22] Kokurin M Y 2016 On stable finite dimensional approximation of conditionally well-posed inverse problems *Inverse Problems* 32 105007
[23] Kokurin M Y 2020 On the clustering of stationary points of Tikhonov’s functional for conditionally well-posed inverse problems *J. Inverse Ill-Posed Problems* 28 713–25
[24] Kaltenbacher B, Neubauer A and Scherzer O 2008 *Iterative Regularization Methods for Nonlinear Ill-Posed Problems* (Berlin: de Gruyter and Co)
[25] Li L, Han B and Wang W 2007 R–K type Landweber method for nonlinear ill-posed problems J. Comput. Appl. Math. 206 341–57

[26] Lu S, Niu P and Werner F 2021 On the asymptotical regularization for linear inverse problems in presence of white noise SIAM/ASA J. Uncertain. Quantif. 9 1–28

[27] Lu W, Duan J, Drive-Miguel D, Herve L and Styles I B 2019 Graph- and finite element-based total variation models for the inverse problem in diffuse optical tomography Biomed. Opt. Express 10 2684–707

[28] Lechleiter A and Rieder A 2010 Towards a general convergence theory for inexact Newton regularizations Numer. Math. 114 521–48

[29] Mittal G and Giri A K 2021 Iteratively regularized Landweber iteration method: convergence analysis via Hölder stability Appl. Math. Comput. 392 125744

[30] Mittal G and Giri A K 2022 Convergence rates for iteratively regularized Gauss–Newton method subject to stability constraints J. Comput. Appl. Math. 400 113744

[31] Maaß P and Strehlow R 2016 An iterative regularization method for nonlinear problems based on Bregman projections Inverse Problems 32 115013

[32] Rieder A 2005 Runge–Kutta integrators yield optimal regularization schemes Inverse Problems 21 453–71

[33] Rieder A 1999 On the regularization of nonlinear ill-posed problems via inexact Newton iterations Inverse Problems 15 309–27

[34] Rieder A 2001 On convergence rates of inexact Newton regularizations Numer. Math. 88 347–65

[35] Real R and Jin Q 2020 A revisit on Landweber iteration Inverse Problems 36 075011

[36] Tikhonov A, Leonov A and Yagola A 1998 Nonlinear ill-Posed Problems vols I and II (London: Chapman and Hall)

[37] Tautenhahn U 1994 On the asymptotical regularization of nonlinear ill-posed problems Inverse Problems 10 1405–18

[38] Wang Y and Liu J 2017 Identification of non-smooth boundary heat dissipation by partial boundary data Appl. Math. Lett. 69 42–8

[39] Zălinescu C 2002 Convex Analysis in General Vector Spaces (Singapore: World Scientific)

[40] Zhu M and Chan T F 2008 An efficient primal-dual hybrid gradient algorithm for total variation image restoration CAM Report 08-34 (UCLA)

[41] Zhang Y and Hofmann B 2020 On the second-order asymptotical regularization of linear ill-posed inverse problems Appl. Anal. 99 1000–25

[42] Zhao Y, Mathé P and Lu S 2020 Convergence analysis of asymptotical regularization and Runge–Kutta integrators for linear inverse problems under variational source conditions SIAM Trans. Appl. Math. 1 693–714

[43] Zhong M and Wang W 2020 The two-point gradient methods for nonlinear inverse problems based on Bregman projections Inverse Problems 36 045012

[44] Zhong M, Wang W and Jin Q 2019 Regularization of inverse problems by two-point gradient methods in Banach spaces Numer. Math. 143 713–47