Abstract. By making use of tools from convex analysis, we formulate an inexact Newton-Landweber iteration method for solving nonlinear inverse problems in Banach spaces. The method consists of two components: an outer Newton iteration and an inner scheme. The inner scheme provides increments by applying Landweber iteration with non-smooth uniformly convex penalty terms to local linearized equations. The outer iteration is then terminated by the discrepancy principle. Detailed convergence analysis is present under standard conditions on the nonlinearity. Finally, numerical simulations are reported to test the performance of the method.

Key words. Nonlinear inverse problems, inexact Newton-Landweber iteration, Bregman distance, uniformly convex functions, convergence

AMS subject classifications. 65J15, 65J20, 47H17

1. Introduction. In this paper we consider the nonlinear operator equations

\[ F(x) = y \]

(1.1)
arising from nonlinear inverse problems, where \( F : \mathcal{D}(F) \subset \mathcal{X} \to \mathcal{Y} \) is a nonlinear operator between two Banach spaces \( \mathcal{X} \) and \( \mathcal{Y} \) with domain \( \mathcal{D}(F) \). A characteristic property of such equations is their ill-posedness in the sense that their solutions are not stable with respect to the perturbation of data. Due to errors in the measurements, one never has exact data in practical applications, instead one only has data contaminated by noise. If an algorithm developed for well-posed problems is used directly to ill-posed inverse problems, it usually fails to produce useful information on the sought solution because noise could be amplified by an arbitrarily large factor. How to construct stable approximate solutions of (1.1) from noisy data is therefore an important topic and regularization methods should be taken into account.

Motivated by the inexact Newton methods in [7] for well-posed nonlinear equations, Hanke proposed the truncated Newton-CG algorithm in [9] and the regularizing Levenberg-Marquardt scheme in [10] for solving nonlinear inverse problems in Hilbert spaces. Rieder generalized the idea and proposed in [21] a family of inexact Newton regularization methods as efficient solvers and gave some convergence analysis in a series of papers [21, 22, 19]. The order optimality of certain inexact Newton regularization methods has been confirmed recently, see [11, 13]. These methods can produce good reconstruction results when the sought solutions have certain smoothness property. However, due to their tendency to over-smooth solutions, they fail to capture the special features of the sought solutions such as sparsity and piecewise constancy. Therefore it is necessary to reformulate these methods either in Banach space setting or in a manner that modern non-smooth penalty functions, such as the \( L^1 \) and the total variation like functions, can be incorporated.

To proceed, it is necessary to briefly review the basic idea behind the inexact Newton methods. Assume that \( F \) is Fréchet differentiable whose Fréchet derivative at \( x \) is denoted by \( F'(x) \). When \( x_n \) is a current iterate, we may consider the linearized
equation

\[ F'(x_n)(x - x_n) = y - F(x_n) \]  

(1.2)

of (1.1) at \( x_n \). The classical Newton method builds the next iterate by solving (1.2) using an exact procedure. Unfortunately, for large size problems, it is usually very costly to solve (1.2) exactly. To avoid such expensive computation, the inexact Newton methods suggest to use certain iterative procedure to solve (1.2) inexactly to produce an element \( x_{n+1} \) such that

\[ \|y - F(x_n) - F'(x_n)(x_{n+1} - x_n)\|_Y \leq \gamma \|y - F(x_n)\|_Y \]  

(1.3)

for some \( \gamma < 1 \). This \( x_{n+1} \) is then used as the next iterate, see [7]. For nonlinear inverse problems, the linearized operator \( F'(x_n) \) may not be invertible in general which makes the exact resolution of (1.2) impossible. One has to regularize the linear equation (1.2) by certain regularization procedure to build such an \( x_{n+1} \), see [9, 21].

When \( X \) is a uniformly smooth and uniformly convex Banach space, by making use of the duality mappings, we proposed in [14] an inexact Newton-Landweber iteration method for solving nonlinear inverse problems in Banach spaces. The method in [14] is based on regularizing the linearized equation (1.2) by the Landweber iteration in Banach spaces from [24]. Assuming that \( x_n \) is a current iterate and using \( F'(x)^* : Y^* \rightarrow X^* \) to denote the adjoint of \( F'(x) \), it is proposed in [14] to define

\[ x_{n,k+1} = J_{p^*}^X (J_p^X(x_{n,k}) + \mu_{n,k} F'(x_{n,k})^* J_{p^*}^Y (y - F(x_{n,k}) - F'(x_{n,k})(x_{n,k} - x_n))) \]

for \( k = 0, 1, \ldots \), with \( 1 < p, r < \infty \) and suitable step-length \( \mu_{n,k} \), where \( x_{n,0} = x_n \), \( p^* \) is the number conjugate to \( p \), i.e. \( 1/p^* + 1/p = 1 \), and \( J_{p^*}^X \) is the duality mapping of \( X \) with gauge function \( t \rightarrow t^{p-1} \). The method in [14] then defines the next iterate as \( x_{n+1} = x_{n,k_n} \) such that \( x_{n+1} \) is the first element in the sequence \( \{x_{n,k}\} \) satisfying (1.3).

There are two issues arising from the method in [14]. First, although the strong convergence for the method with exact data is proved, only weak convergence in subsequences is obtained for the method with noisy data. Second, the formulation of the method in [14] depends crucially on the duality mapping \( J_{p^*}^X \) which in fact is the Fréchet derivative of the convex function \( \Theta(x) = \frac{1}{p} \|x\|_p^p \). When \( X = L^2(\Omega) \), where \( \Omega \) denotes a bounded domain in \( \mathbb{R}^d \), the formulation of the method in [14] excludes the use of the non-smooth convex penalty functions

\[ \Theta(x) = \frac{1}{2\beta} \|x\|_2^2 + \|x\|_{L^1} \quad \text{and} \quad \Theta(x) := \frac{1}{2\beta} \|x\|_2^2 + \|x\|_{TV} \]  

(1.4)

which are important for sparsity reconstruction and discontinuity detection respectively, where \( \beta > 0 \) and \( |x|_{TV} \) denotes the total variation of \( x \) over \( \Omega \) defined by

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

Moreover, in some applications the inverse problems are naturally formulated as the equation (1.1) with \( F \) being an operator on \( L^2(\Omega) \). In order to apply the method to such a problem to find sparse solutions, one has to extend \( F \) as an operator on \( L^p(\Omega) \) with \( p > 1 \) being close to 1. This unnatural reformulation makes it difficult to verify the required properties on the operator \( F \).
In order to formulate a version of inexact Newton-Landweber iteration in more general framework so that the convex functions in (1.4) can be incorporated, instead of using the method in [24] one should use a more general method to regularize the linear equation (1.2). To this end, we recall a recent version of Landweber iteration introduced in [4] for solving linear inverse problems $Au = h$, where $A : \mathcal{X} \to \mathcal{Y}$ is a bounded linear operator between two Banach spaces $\mathcal{X}$ and $\mathcal{Y}$. By taking a proper, lower semi-continuous, uniformly convex function $\Theta : \mathcal{X} \to (-\infty, \infty]$, the method in [4] reads

$$
\varphi_{k+1} = \varphi_k + \mu_k A^* J^Y_\rho (h - Au_k), \\
u_{k+1} = \arg\min_{u \in \mathcal{X}} \{ \Theta(u) - \langle \varphi_{k+1}, u \rangle_{\mathcal{X}^*, \mathcal{X}} \},
$$

with suitable initial guess $(\varphi_0, u_0) \in \mathcal{X}^* \times \mathcal{X}$ and step-lengths $\mu_k$, where $\langle \cdot, \cdot \rangle_{\mathcal{X}^*, \mathcal{X}}$ denotes the duality pairing between $\mathcal{X}^*$ and $\mathcal{X}$. Further convergence analysis on (1.5) can be found in [15]. In this paper, by applying (1.5) to the linearized equation (1.2) at each step we will propose an inexact Newton-Landweber iteration method with non-smooth uniformly convex penalty terms. Roughly speaking, starting from a current iterate $(\xi_n, x_n) \in \mathcal{X}^* \times \mathcal{X}$, our method first defines

$$
\xi_{n+1} = \xi_{n,k} + \mu_{n,k} F'(x_n)^* J^Y_\rho (y - F(x_n) - F'(x_n)(x_{n,k} - x_n)), \\
x_{n,k+1} = \arg\min_{x \in \mathcal{X}} \{ \Theta(x) - \langle \xi_{n,k+1}, x \rangle_{\mathcal{X}^*, \mathcal{X}} \}
$$

for $k = 0, 1, \ldots$, where $\xi_{n,0} = \xi_n$ and $x_{n,0} = x_n$. The next iterate is then defined as $\xi_{n+1} = \xi_{n,k} \text{ and } x_{n+1} = x_{n,k_n}$ with the integer $k_n \geq 1$ determined by the inexact Newton criterion (1.3). The detailed description of the method will be given in Section 3. The non-smoothness of $\Theta$ presents many challenges in convergence analysis. We will use tools from convex analysis to settle the challenging issues. For inexact Newton regularization methods the convergence analysis is rather subtle in particular when the data contains noise. The main obstacle comes from the stability issue; when taking the noise level to zero, the method with noisy data can result in many possible noise-free iterative sequences. We will conquer this difficulty by borrowing an idea from [9] to show that all these noise-free sequences converge uniformly in certain sense.

During the last decade, there have been many research activities on regularization methods for solving inverse problems in Banach spaces. Due to its variational formulation, Tikhonov regularization receives much attention and it solves (1.1) by using the minimizer of the minimization problem

$$
x_{\alpha} = \arg\min_{x \in \partial(F)} \left\{ \| F(x) - y \|_Y^2 + \alpha \Theta(x) \right\},
$$

see [25] and the references therein. The performance of Tikhonov regularization requires a good choice of the regularization parameter $\alpha > 0$ and an efficient solver to determine the minimizers. Since the cost function is possibly non-convex and since the possibly non-smooth penalty term $\Theta$ is mixed with $F$, the numerical implementation in general is rather expensive. As alternatives, iterative regularization methods have also been considered for solving nonlinear inverse problems in Banach spaces. One may refer to [18, 16] for the iteratively regularized Gauss-Newton method and to [17] for the nonstationary iterated Tikhonov regularization method. In these iterative methods, every step involves solving a minimization problem like (1.7) or the convex minimization problem

$$
\min_{x \in \mathcal{X}} \left\{ \| y - F(x_n) - F'(x_n)(x - x_n) \|_Y^2 + \alpha \Theta(x) \right\}
$$
in which Θ and F are mixed which increases the difficulty of numerical implementation. Unlike the above methods, our inexact Newton-Landweber iteration method (1.6) has the splitting character in the sense that ξ_{n,k} is determined by F completely without involving Θ and x_{n,k} is defined by Θ without using F. This feature is favorable in numerical computation.

This paper is organized as follows. In section 2 we give some preliminary results from convex analysis in Banach spaces. In section 3, we propose an inexact Newton-Landweber iteration method for solving (1.1) using general non-smooth uniformly convex penalty terms, and present the detailed convergence analysis including the strong convergence when data contain noise. Finally, in section 4 we present some numerical simulations to test the performance of the method.

2. Preliminaries. Let X be a Banach space whose norm is denoted as ||·||_X. We use X^* to denote its dual space, and for any x ∈ X and ξ ∈ X^* we write ⟨ξ, x⟩_{X^∗,X} = ξ(x) for the duality pairing. If Y is another Banach space and A : X → Y is a bounded linear operator, we use A^∗ : Y^∗ → X^* to denote its adjoint, i.e. ⟨A^∗ξ, x⟩_{X^∗,X} = ⟨ξ, Ax⟩_{Y^∗,Y} for any x ∈ X and ξ ∈ Y^*. Let \mathcal{N}(A) = \{x ∈ X : Ax = 0\} be the null space of A and let

\mathcal{N}(A)^\perp := \{ξ ∈ X^∗ : ⟨ξ, x⟩_{X^∗,X} = 0 \text{ for all } x ∈ \mathcal{N}(A)\}

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

For each 1 < r < ∞, the set-valued mapping J_r^X : X → 2^{X^∗} defined by

J_r^X(x) := \{ξ ∈ X^∗ : ||ξ||_{X^∗} = ||x||_X^r \text{ and } ⟨ξ, x⟩_{X^∗,X} = ||x||_X^r\}, \quad ∀x ∈ X

called the duality mapping of X with gauge function t → t^{r−1}. When X is uniformly smooth in the sense that its modulus of smoothness

ς_X(s) := \sup\{||\bar{x} + x||_X + ||\bar{x} − x||_X − 2 : ||\bar{x}||_X = 1, ||x||_X ≤ s\}

satisfies lim_{s→0} \frac{ς_X(s)}{s} = 0, the duality mapping J_r^X is single valued and uniformly continuous on bounded sets.

Given a convex function Θ : X → (−∞, ∞], we use \mathcal{D}(Θ) := \{x ∈ X : Θ(x) < +∞\} to denote its effective domain. It is called proper if \mathcal{D}(Θ) ≠ ∅. The subdifferential of Θ at x ∈ X is defined as

\partial Θ(x) := \{ξ ∈ X^∗ : Θ(z) − Θ(x) − ⟨ξ, z − x⟩_{X^∗,X} ≥ 0 \text{ for all } z ∈ X\}.

For the multi-valued mapping \partial Θ : X → 2^{X^∗}, we set

\mathcal{D}(\partial Θ) := \{x ∈ \mathcal{D}(Θ) : \partial Θ(x) ≠ ∅\}.

Given x ∈ \mathcal{D}(\partial Θ) and ξ ∈ \partial Θ(x), we define

D_ξΘ(z, x) := Θ(z) − Θ(x) − ⟨ξ, z − x⟩_{X^∗,X}, \quad ∀z ∈ X

which is called the Bregman distance induced by Θ at x in the direction ξ ([6]).

Bregman distance can be used to obtain information under the Banach space norm when Θ has stronger convexity. A proper convex function Θ : X → (−∞, ∞] is called
uniformly convex if there is a continuous increasing function \( h : [0, \infty) \to [0, \infty) \), with the property that \( h(t) = 0 \) implies \( t = 0 \), such that
\[
\Theta(\lambda \bar{x} + (1 - \lambda)x) + c_0 \lambda (1 - \lambda) h(\|\bar{x} - x\|_X) \leq \lambda \Theta(\bar{x}) + (1 - \lambda)\Theta(x)
\]
for all \( \bar{x}, x \in \mathcal{X} \) and \( \lambda \in (0, 1) \). If \( h \) can be taken as \( h(t) = c_0 t^p \) for some \( c_0 > 0 \) and \( p \geq 2 \), then \( \Theta \) is called \( p \)-convex. It is straightforward to show that if \( \Theta \) is \( p \)-convex then
\[
D_\xi \Theta(\bar{x}, x) \geq c_0 \|\bar{x} - x\|_X^p
\tag{2.1}
\]
for all \( \bar{x} \in \mathcal{X} \), \( x \in \partial (\partial \Theta) \) and \( \xi \in \partial \Theta(x) \).

For a proper, lower semi-continuous, convex function \( \Theta : \mathcal{X} \to (-\infty, \infty] \), its Legendre-Fenchel conjugate is defined by
\[
\Theta^*(\xi) := \sup_{x \in \mathcal{X}} \{ \langle \xi, x \rangle_{\mathcal{X}^*, \mathcal{X}} - \Theta(x) \}, \quad \xi \in \mathcal{X}^*.
\]

It is well known that \( \Theta^* \) is also proper, lower semi-continuous, and convex. If, in addition, \( \mathcal{X} \) is reflexive, then ([26, Theorem 2.4.2])
\[
\xi \in \partial \Theta(x) \iff x \in \partial \Theta^*(\xi) \iff \Theta(x) + \Theta^*(\xi) = \langle \xi, x \rangle_{\mathcal{X}^*, \mathcal{X}}
\tag{2.2}
\]
and, by the subdifferential calculus, there also holds
\[
x \in \partial \Theta^*(\xi) \iff x = \arg \min_{x \in \mathcal{X}} \{ \Theta(z) - \langle \xi, z \rangle_{\mathcal{X}^*, \mathcal{X}} \}.
\tag{2.3}
\]

From (2.2) it follows that
\[
D_\xi \Theta(\bar{x}, x) = \Theta(x) + \Theta^*(\xi) - \langle \xi, x \rangle_{\mathcal{X}^*, \mathcal{X}}
\tag{2.4}
\]
for all \( \bar{x} \in \mathcal{X} \), \( x \in \partial (\partial \Theta) \) and \( \xi \in \partial \Theta(x) \).

When \( \Theta \) is proper, lower semi-continuous and \( p \)-convex satisfying (2.1) with \( p \geq 2 \), it follows from [26, Corollary 3.5.11] that \( \partial (\Theta^*) = \mathcal{X}^* \), \( \Theta^* \) is Fréchet differentiable and its gradient \( \nabla \Theta^* : \mathcal{X}^* \to \mathcal{X} \) satisfies
\[
\| \nabla \Theta^*(\xi_1) - \nabla \Theta^*(\xi_2) \|_X \leq \left( \frac{\| \xi_1 - \xi_2 \|_{\mathcal{X}^*}}{2c_0} \right)^{1/p^*}, \quad \forall \xi_1, \xi_2 \in \mathcal{X}^*.
\tag{2.5}
\]
	Moreover
\[
\Theta^*(\xi_2) - \Theta^*(\xi_1) - \langle \xi_2 - \xi_1, \nabla \Theta^*(\xi_1) \rangle_{\mathcal{X}^*, \mathcal{X}} \leq \frac{1}{p^*(2c_0)^{p^*}} \| \xi_2 - \xi_1 \|_{X^*}^{p^*}.
\tag{2.6}
\]
for any \( \xi_1, \xi_2 \in \mathcal{X}^* \), where \( p^* \) is the number conjugate to \( p \), i.e. \( 1/p + 1/p^* = 1 \).

3. Inexact Newton-Landweber iteration. We consider the equation (1.1), where \( F : \partial(F) \subset \mathcal{X} \to \mathcal{Y} \) is a nonlinear operator between two Banach spaces \( \mathcal{X} \) and \( \mathcal{Y} \) with domain \( \partial(F) \). We will assume that (1.1) has a solution. In general, (1.1) may have many solutions. In order to find the desired one, some selection criteria should be enforced. According to a prior information on the sought solution, we choose a proper, lower semi-continuous, \( p \)-convex function \( \Theta : \mathcal{X} \to (-\infty, \infty] \). By picking \((\xi_0, x_0) \in \mathcal{X}^* \times \mathcal{X} \) with \( \xi_0 \in \partial \Theta(x_0) \) as an initial guess, we define \( x^0 \) to be the solution of (1.1) with the property
\[
D_{\xi_0} \Theta(x^+, x_0) := \min_{x \in \partial (\partial \Theta) \cap \partial(F)} \{ D_{\xi_0} \Theta(x, x_0) : F(x) = y \}.
\tag{3.1}
\]


We are interested in the situation that (1.1) is ill-posed in the sense that its solution does not depend continuously on the data. Our goal is to develop some iterative regularization method to find \( x^\dagger \) using noisy data \( y^\delta \) satisfying

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

(3.2)

with a small known noise level \( \delta > 0 \). We will work under the following standard conditions on the operator \( F \), where \( B_\delta(x_0) := \{ x \in X : \| x - x_0 \|_X \leq \delta \} \).

**Assumption 3.1.**

(a) There is \( \rho > 0 \) such that \( B_{2\delta}(x_0) \subset \mathcal{D}(F) \) and (1.1) has a solution in \( B_\delta(x_0) \cap \mathcal{D}(\Theta) \);

(b) The operator \( F \) is weakly closed on \( \mathcal{D}(F) \), i.e. if \( \{ u_n \} \subset \mathcal{D}(F) \) converges weakly to some \( u \in X \) and \( \{ F(u_n) \} \) converges weakly to some \( v \in Y \), then \( u \in \mathcal{D}(F) \) and \( F(u) = v \);

(c) There exists a family of bounded linear operators \( \{ L(x) : X \to Y \} \in B_{2\delta}(x_0) \) such that \( x \to L(x) \) is continuous on \( B_{2\delta}(x_0) \) and there is \( 0 \leq \eta < 1 \) such that

\[
\| F(\bar{x}) - F(x) - L(x)(\bar{x} - x) \|_Y \leq \eta \| F(\bar{x}) - F(x) \|_Y
\]

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

In Assumption 3.1, \( F \) is not required to be Fréchet differentiable; in case \( F \) is Fréchet differentiable, we can take \( L(x) = F'(x) \), where \( F'(x) \) denotes the Fréchet derivative of \( F \) at \( x \). By (c) in Assumption 3.1 it is easy to see that

\[
\| F(\bar{x}) - F(x) \|_Y \leq \frac{1}{1 - \eta} \| L(x)(\bar{x} - x) \|_Y, \quad \bar{x}, x \in B_{2\delta}(x_0)
\]

which implies that \( x \to F(x) \) is continuous on \( B_{2\delta}(x_0) \).

When \( X \) is a reflexive Banach space, by using the \( p \)-convexity and the lower semicontinuity of \( \Theta \) together with the weakly closedness of \( F \) it is standard to show that \( x^\dagger \) exists. Moreover, it has been shown in [15, Lemma 3.2] that if \( x^\dagger \) is a solution of (1.1) satisfying (3.1) with

\[
D_{\delta_0}\Theta(x^\dagger, x_0) \leq \gamma_0 \theta^p,
\]

(3.3)

then \( x^\dagger \) is uniquely defined. We emphasize that the weakly closedness of \( F \) over \( \mathcal{D}(F) \) is only used to guarantee the existence of \( x^\dagger \), it will not be used any more in the following analysis.

### 3.1. The method with noisy data. To formulate the method, let \( 1 < r < \infty \) and let \( J^\delta_r \) denote the duality mapping over \( Y \) with gauge function \( t \to t^{r-1} \). Let \( y^\delta \) be the available noisy data satisfying (3.2). Then, instead of (1.2), at a current iterate \( x_n^\delta \) we have the local linearized equation

\[
L(x_n^\delta)x = y^\delta - F(x_n^\delta) + L(x_n^\delta)x_n^\delta.
\]

We may apply the method (1.5) to this equation to produce the next iterate \( x_{n+1}^\delta \) so that \( \| y^\delta - F(x_{n+1}^\delta) - L(x_n^\delta)(x_{n+1}^\delta - x_n^\delta) \|_Y \) is smaller than a suitable multiple of \( \| y^\delta - F(x_n^\delta) \|_Y \). This leads to the inexact Newton-Landweber iteration which can be formulated as follows:

**Algorithm 3.1.** Let \( 0 < \gamma < 1, \mu_0 > 0, \mu_1 > 0 \) and \( \tau > 1 \) be suitably chosen numbers.
(i) Pick \( \xi_0 \in \mathcal{X}^* \) and set \( x_0 := \arg \min_{x \in \mathcal{X}} \{ \Theta(x) - \langle \xi_0, x \rangle_{\mathcal{X}^*} \} \).

(ii) Let \( \xi_0^\delta := \xi_0 \) and \( x_0^\delta := x_0 \). Assume that \( \xi_n^\delta \) and \( x_n^\delta \) are defined for some \( n \geq 0 \), we set \( \xi_{n,0}^\delta = \xi_n^\delta \) and \( x_{n,0}^\delta = x_n^\delta \) and define

\[
\begin{align*}
\xi_{n,k+1}^\delta &= \xi_{n,k}^\delta + \mu_{n,k}^\delta L(x_n^\delta)^* J_r^Y (s_{n,k}^\delta), \\
x_{n,k+1}^\delta &= \arg \min_{x \in \mathcal{X}} \{ \Theta(x) - \langle \xi_{n,k+1}^\delta, x \rangle_{\mathcal{X}^*} \},
\end{align*}
\]

where

\[
s_{n,k}^\delta = y^\delta - F(x_n^\delta) - L(x_n^\delta)(x_n^\delta - x_{n,k}^\delta),
\]

\[
\mu_{n,k}^\delta = \min \left\{ \frac{\mu_0}{\| L(x_n^\delta)^* J_r^Y (s_{n,k}^\delta) \|_{\mathcal{X}^*}}, \mu_1 \right\} \| s_{n,k}^\delta \|_{\mathcal{X}}^{p-r}.
\]

(iii) Let \( k_n^\delta \geq 1 \) be the first integer such that

\[
\| s_{n,k}^\delta \|_Y < \gamma \| y^\delta - F(x_n^\delta) \|_Y.
\]

We then define \( \xi_{n,k}^\delta := \xi_{n,k,n}^\delta \) and \( x_{n,k}^\delta := x_{n,k}^\delta \).

(iv) Let \( n_{\delta} \) be the integer determined by the discrepancy principle

\[
\| F(x_{n,k}^\delta) - y^\delta \|_Y \leq \tau \delta < \| F(x_{n,k}^\delta) - y^\delta \|_Y, \quad 0 \leq n < n_{\delta}
\]

and use \( x_{n,k}^\delta \) as an approximate solution.

The formulation of Algorithm 3.1 consists of two components: an outer Newton iteration and an inner scheme providing increments by regularizing local linearized equations using Landweber iteration penalized by non-smooth convex functions. Note that, for each \( k = 0, 1, \ldots, \xi_{n,k}^\delta \) is determined by \( F \) completely without involving \( \Theta \), and \( x_{n,k}^\delta \) is defined as the minimizer of a convex function over \( \mathcal{X} \) related to \( \Theta \) that is independent of \( F \). This splitting character can make the implementation of the algorithm efficiently. By using (2.3) and the differentiability of \( \Theta^* \), one can see that

\[
x_{n,k}^\delta = \nabla \Theta^*(\xi_{n,k}^\delta)
\]

We will use this fact in the forthcoming convergence analysis.

We first prove the following basic result which shows that Algorithm 3.1 is well-defined.

**Lemma 3.1.** Let \( \mathcal{X} \) and \( \mathcal{Y} \) be Banach spaces with \( \mathcal{Y} \) being uniformly smooth, let \( \Theta : \mathcal{X} \to (-\infty, \infty] \) be a proper, lower semi-continuous and \( p \)-convex function with \( p \geq 2 \) satisfying (2.1) for some \( c_0 > 0 \), let Assumption 3.1 hold with \( 0 \leq \eta < 1 \), and let (3.3) be satisfied. If \( \eta < \gamma < 1 \), \( \mu_0 > 0 \) and \( \tau > 1 \) are chosen such that

\[
c_1 := 1 - \frac{\eta}{\gamma} - \frac{1 + \eta}{\gamma^2} - \frac{p - 1}{p} \left( \frac{\mu_0}{2c_0} \right)^{\frac{1}{p-1}} > 0,
\]

then for Algorithm 3.1 there hold

(i) for each \( 0 \leq n < n_{\delta} \), \( k_n^\delta \) is finite and \( x_{n,k}^\delta \in B_{2\epsilon}(x_0) \) for all \( 0 \leq k \leq k_n^\delta \);

(ii) the method terminates after \( n_{\delta} < \infty \) iteration steps;

(iii) \( D_{n+1}^\delta \Theta(\hat{x}, x_{n+1}^\delta) \leq D_{n}^\delta \Theta(\hat{x}, x_n^\delta) \) for all \( 0 \leq n < n_{\delta} \), where \( \hat{x} \) denotes any solution of (1.1) in \( B_{2\epsilon}(x_0) \cap \mathcal{D}(\Theta) \).
Proof. We first show that, if \( x_n^\delta \in B_{2\varepsilon}(x_0) \) for some \( 0 \leq n < n_\delta \), then \( k_n^\delta \) must be finite and

\[
D_{\xi_n^+, k + 1}^\delta \Theta(\hat{x}, x_{n,k}^\delta) \leq D_{\xi_n^+, k}^\delta \Theta(\hat{x}, x_{n,k}^\delta), \quad \forall 0 \leq k < k_n^\delta. \tag{3.8}
\]

To see this, from (2.4) and the fact \( x_n^\delta = \nabla \Theta^*(\xi_n^\delta) \), we can write

\[
D_{\xi_n^+, k + 1}^\delta \Theta(\hat{x}, x_{n,k+1}^\delta) - D_{\xi_n^+, k}^\delta \Theta(\hat{x}, x_{n,k}^\delta) = \Theta^*(\xi_{n,k+1}^\delta) - \Theta^*(\xi_{n,k}^\delta)
\]

\[
- \langle \xi_{n,k+1}^\delta - \xi_{n,k}^\delta, \nabla \Theta^*(\xi_{n,k}^\delta) \rangle_{X^*,X}
+ \langle \xi_{n,k+1}^\delta - \xi_{n,k}^\delta, x_{n,k}^\delta - \hat{x} \rangle_{X^*,X}.
\]

Since \( \Theta \) is \( p \)-convex, we may use (2.6) and the definition of \( \xi_{n,k}^\delta \) to obtain

\[
D_{\xi_n^+, k + 1}^\delta \Theta(\hat{x}, x_{n,k+1}^\delta) - D_{\xi_n^+, k}^\delta \Theta(\hat{x}, x_{n,k}^\delta) \leq \frac{1}{p^r(2\varepsilon_0)^p} (\mu_{n,k}^\delta)_{\frac{r}{p}} \| L(x_n^\delta)^* J^\delta_Y(s_n^\delta) \|_{X^*}^p
\]

\[
+ \mu_{n,k}^\delta \| J_Y^\delta (s_n^\delta, L(x_n^\delta)(x_n^\delta - \hat{x}))_{Y^*,Y} \|
\]

By writing

\[
L(x_n^\delta)(x_{n,k}^\delta - \hat{x}) = -s_{n,k}^\delta + [y_n^\delta - F(x_n^\delta) - L(x_n^\delta)(\hat{x} - x_n^\delta)],
\]
we may use (3.2), Assumption 3.1 (c), the property of \( J_Y^\delta \) and the choice of \( \mu_{n,k}^\delta \) to obtain

\[
D_{\xi_n^+, k + 1}^\delta \Theta(\hat{x}, x_{n,k+1}^\delta) - D_{\xi_n^+, k}^\delta \Theta(\hat{x}, x_{n,k}^\delta) \leq \frac{1}{p^r(2\varepsilon_0)^p} \left( \frac{\mu_0}{2\varepsilon_0} \right)^{\frac{r}{p} - 1} \mu_{n,k}^\delta \| s_n^\delta \|_{Y}^p - \mu_{n,k}^\delta \| s_n^\delta \|_{Y}^p
\]

\[
+ \mu_{n,k}^\delta \| s_n^\delta \|_{Y} \| X^* \|
\]

Since \( 0 \leq n < n_\delta \) and \( 0 \leq k < k_n^\delta \), we have

\[
\| F(x_n^\delta) - y_n^\delta \|_Y \geq \tau \delta \quad \text{and} \quad \| s_n^\delta \|_Y \geq \gamma \| y_n^\delta - F(x_n^\delta) \|_Y. \tag{3.9}
\]

Consequently, for \( 0 \leq n < n_\delta \) and \( 0 \leq k < k_n^\delta \) we have

\[
D_{\xi_n^+, k + 1}^\delta \Theta(\hat{x}, x_{n,k+1}^\delta) - D_{\xi_n^+, k}^\delta \Theta(\hat{x}, x_{n,k}^\delta) \leq -c_1 \mu_{n,k}^\delta \| s_n^\delta \|_Y.
\]

where \( c_1 > 0 \) is the constant defined by (3.7). This implies (3.8) and

\[
c_1 \sum_{k=0}^l \mu_{n,k}^\delta \| s_n^\delta \|_{Y} \leq D_{\xi_n^\delta} \Theta(\hat{x}, x_{n}^\delta) - D_{\xi_n^\delta, l+1} \Theta(\hat{x}, x_{n,l+1}^\delta) \tag{3.10}
\]

for all \( l < k_n^\delta \). Observing that

\[
\mu_{n,k}^\delta \geq c_2 \| s_n^\delta \|_Y^{p-r} \quad \text{with} \quad c_2 := \min\{ \mu_0 B^{-p}, \mu_1 \},
\]

where \( B > 0 \) is a constant such that \( \| L(x) \|_{X^*,X} \leq B \) for all \( x \in \mathcal{B}_{2\varepsilon}(x_0) \), we obtain from (3.9) that

\[
\mu_{n,k}^\delta \| s_n^\delta \|_{Y} \geq c_2 \| s_n^\delta \|_Y^{p} \geq c_2 \gamma^p \| y_n^\delta - F(x_n^\delta) \|_Y^p \geq c_2 \gamma^p \tau \delta^p. \tag{3.11}
\]
Therefore, it follows from (3.10) that
\[ c_1 c_2 \gamma^p \tau^p \delta^p (l + 1) \leq D_{\xi_n^k} \Theta(\hat{x}, x_n^\delta) < \infty, \quad \forall 0 \leq l < k_n^\delta. \]
This shows that $k_n^\delta$ is finite and hence $x_{n+1}^\delta$ is well-defined. By taking $l = k_n^\delta - 1$ in (3.10) we obtain
\[ c_1 \sum_{k=0}^{k_n^\delta - 1} \mu_{n,k} \|s_{n,k}^\delta\|_Y \leq D_{\xi_n^k} \Theta(\hat{x}, x_n^\delta) - D_{\xi_n^k} \Theta(\hat{x}, x_{n+1}^\delta). \quad (3.12) \]

Next, by using (3.3) and (2.1) we have $\|x_0 - x^\dagger\|_X \leq \varrho$, i.e. $x^\dagger$ is a solution of (1.1) in $B_\varrho(x_0) \cap D(\Theta)$. We may use (3.8) with $\hat{x} = x^\dagger$ inductively to conclude that
\[ D_{\xi_n^k} \Theta(x^\dagger, x_n^\delta) \leq D_{\xi_n^0} \Theta(x^\dagger, x_0) \leq c_0 \varrho^p \]
which together with (2.1) gives $\|x_{n,k}^\delta - x^\dagger\|_X \leq \varrho$ and hence $\|x_{n,k}^\delta - x_0\|_X \leq 2\varrho$ for all $0 \leq n < n_\delta$ and $0 \leq k \leq k_n^\delta$. This in particular implies $x_n^\delta \in B_{2\varrho}(x_0)$ for $0 \leq n < n_\delta$. Consequently, we can conclude from (3.8) that
\[ D_{\xi_n^k} \Theta(\hat{x}, x_{n+1}^\delta) \leq D_{\xi_n^0} \Theta(\hat{x}, x_n^\delta), \quad 0 \leq n < n_\delta. \]

Finally, in order to conclude $n_\delta < \infty$, we may sum the equation (3.12) over $n$ from $n = 0$ to $n = m$ for any $m < n_\delta$ and use (3.11) and $k_n^\delta \geq 1$ to obtain
\[ c_1 c_2 \gamma^p \tau^p \delta^p (m + 1) \leq D_{\xi_n^0} \Theta(\hat{x}, x_0) < \infty. \]
This implies that $n_\delta$ must be finite. \( \Box \)

**Remark 3.1.** The proof of Lemma 3.1 does not require $k_n^\delta$ to satisfy (3.6); it only requires $k_n^\delta \geq 1$ to be an integer such that
\[ \|s_{n,k}^\delta\|_Y \geq \gamma \|y^\delta - F(x_n^\delta)\|_Y, \quad 0 \leq k < k_n^\delta. \quad (3.13) \]
This crucial observation will be used later.

**Remark 3.2.** In Algorithm 3.1, since $k_n^\delta$ is the first integer satisfying (3.6) in each iteration step, if $0 < \eta < 1/3$ and $\eta < \gamma < 1 - 2\eta$ then one can derive that
\[ \|F(x_{n+1}^\delta) - y^\delta\|_Y \leq \sigma \|F(x_n^\delta) - y^\delta\|_Y, \quad 0 \leq n < n_\delta \]
with $\sigma = (\gamma + \eta)/(1 - \eta) < 1$. This together with the definition of $n_\delta$ implies that $n_\delta = O(1 + |\log \delta|)$, see [10, 14] for details. It is this property that makes Algorithm 3.1 an efficient method for solving nonlinear inverse problems.

In the following we will give the convergence analysis of Algorithm 3.1. It is necessary to consider first the counterpart for exact data and then study the stability issue. The counterpart of Algorithm 3.1 for exact data can not be simply formulated by removing the superscript $\delta$ in all quantities since (iii) in Algorithm 3.1 could be violated when taking $\delta \to 0$. A correct formulation should cover all possible cases induced by taking the limit $\delta \to 0$; this will be done in the next subsection.
3.2. The method with exact data. In this subsection, associated with the exact data, we introduce a set that contains all possible iterative sequences arising from Algorithm 3.1 when taking $\delta \to 0$. To this end, let $0 < \gamma < 1$, $\mu_0 > 0$, $\mu_1 > 0$ and $(\xi_0, x_0) \in X^* \times X$ be the same as in Algorithm 3.1. We denote by $\Gamma_{\gamma, \mu_0, \mu_1}(\xi_0, x_0)$ the set consisting of all the sequences $\{ (\xi_n, x_n) \}$ in $X^* \times X$ constructed from $(\xi_0, x_0)$ iteratively as follows: when $(\xi_n, x_n)$ is constructed, one can construct $(\xi_{n+1}, x_{n+1})$ according to the following two cases:

(i) If $F(x_n) = y$, we define $\xi_{n+1} := \xi_n$ and $x_{n+1} := x_n$;
(ii) If $F(x_n) \neq y$, we first construct $\{ \xi_{n,k} \} \subset X^*$ and $\{ x_{n,k} \} \subset X$ iteratively by setting $\xi_{n,0} = \xi_n$ and $x_{n,0} = x_n$ and defining

$$
\xi_{n,k+1} = \xi_{n,k} + \mu_{n,k} L(x_n)^* J_p^Y (s_{n,k}),
$$

$$
x_{n,k+1} = \min_{x \in X} \{ \Theta(x) - \langle \xi_{n,k+1}, x \rangle_{X^*} \},
$$

(3.14)

where

$$
s_{n,k} = y - F(x_n) - L(x_n)(x_{n,k} - x_n),
$$

$$
\mu_{n,k} = \min \left\{ \frac{\mu_0 \| s_{n,k} \|^p r - 1}{\| L(x_n)^* J_p^Y (s_{n,k}) \|_{X^*}^*}, \mu_1 \right\} \| s_{n,k} \|^p r.
$$

(3.15)

Let $k_n \geq 1$ be an integer such that

$$
\| s_{n,k} \|_Y \geq \gamma \| y - F(x_n) \|_Y, \quad 0 \leq k < k_n.
$$

(3.16)

We then define $\xi_{n+1} := \xi_{n,k_n}$ and $x_{n+1} := x_{n,k_n}$.

By using the same argument in the proof of Lemma 3.1, one can show that if $\eta < \gamma < 1$ and $\mu_0 > 0$ are chosen such that

$$
c_3 := 1 - \frac{\eta}{\gamma} - \frac{p - 1}{p} \left( \frac{\mu_0}{2c_0} \right)^{\frac{1}{p-1}} > 0,
$$

(3.17)

then any sequence $\{ (\xi_n, x_n) \} \in \Gamma_{\gamma, \mu_0, \mu_1}(\xi_0, x_0)$ is well-defined and, in case $F(x_n) \neq y$, $k_n$ is a finite integer bounded from above by a number depending only on $D_{\xi_n} \Theta(\hat{x}, x_n)$ and $\| F(x_n) - y \|_Y$. In order to derive the convergence of $\{ x_n \}$ to a solution of (1.1) in Bregman distance, we need the following useful result.

**Proposition 3.2.** Consider the equation (1.1) for which Assumption 3.1 holds. Let $\Theta : X \to (-\infty, \infty]$ be a proper, lower semi-continuous and uniformly convex function. Let $\{ x_n \} \subset B_{2\rho}(x_0)$ and $\{ \xi_n \} \subset X^*$ be such that

(i) $\xi_n \in \partial \Theta(x_n)$ for all $n$;
(ii) for any solution $\hat{x}$ of (1.1) in $B_{2\rho}(x_0) \cap \mathcal{D}(\Theta)$ the sequence $\{ D_{\xi_n} \Theta(\hat{x}, x_n) \}$ is monotonically decreasing;
(iii) $\lim_{n \to \infty} \| F(x_n) - y \|_Y = 0$.

(iv) there is a subsequence $\{ n_j \}$ with $n_j \to \infty$ such that for any solution $\hat{x}$ of (1.1) in $B_{2\rho}(x_0) \cap \mathcal{D}(\Theta)$ there holds

$$
\limsup_{j \to \infty} \| (\xi_{n_j} - \xi_{n_j}, x_{n_j} - \hat{x})_{X^*} \|_X = 0.
$$

(3.18)

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

$$
\lim_{n \to \infty} \| x_n - x_* \|_X = 0 \quad \text{and} \quad \lim_{n \to \infty} D_{\xi_n} \Theta(x_*, x_n) = 0.
$$
If, in addition, \( x^\dagger \in B_\varepsilon(x_0) \cap \mathcal{D}(\Theta) \) and \( \xi_{n+1} - \xi_n \in \mathcal{A}(L(x^\dagger)) \) for all \( n \), then \( x_* = x^\dagger \).

Proof. This is a slight modification of [15, Proposition 3.6], but the same argument applies. \( \square \)

We are now ready to give the convergence result for any sequence \( \{(\xi_n, x_n)\} \in \Gamma_{\gamma, \mu_0, \mu_1}(\xi_0, x_0) \).

**Theorem 3.3.** Let \( \mathcal{X} \) be reflexive and let \( \mathcal{Y} \) be uniformly smooth, let \( \Theta : \mathcal{X} \to (-\infty, \infty] \) be proper, lower semi-continuous and \( p \)-convex with \( p \geq 2 \) satisfying (2.4) for some \( c_0 > 0 \), let Assumption 3.1 hold with \( 0 \leq \eta < 1 \), and let (3.3) be satisfied. Let \( \mu_1 > 0 \) be any number and let \( \eta < \gamma < 1 \) and \( \mu_0 > 0 \) be chosen such that (3.17) holds. Then for any sequence \( \{(\xi_n, x_n)\} \in \Gamma_{\gamma, \mu_0, \mu_1}(\xi_0, x_0) \), there is a solution \( x_* \) of (1.1) in \( B_{2\varepsilon}(x_0) \cap \mathcal{D}(\Theta) \) such that

\[
\lim_{n \to \infty} \|x_n - x_*\|_{\mathcal{X}} = 0 \quad \text{and} \quad \lim_{n \to \infty} D_{\xi_n}\Theta(x_n, x_*) = 0.
\]

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

Proof. We will use Proposition 3.2. By definition we always have \( x_n = \nabla \Theta^*(\xi_n) \) which implies \( \xi_n \in \partial \Theta(x_n) \) for all \( n \geq 0 \). By the same argument in the proof of Lemma 3.1 we can show that

\[
D_{\xi_n+k}\Theta(\hat{x}, x_{n+k}) = D_{\xi_n}\Theta(\hat{x}, x_n), \quad k = 0, \ldots, k_0 - 1
\]

and

\[
c_0 \sum_{k=1}^{k_0-1} \mu_{n,k} \|s_{n,k}\|_{\mathcal{Y}}^p \leq D_{\xi_n}\Theta(\hat{x}, x_n) - D_{\xi_{n+1}}\Theta(\hat{x}, x_{n+1})
\]

for all \( n \geq 0 \), where \( \hat{x} \) denotes any solution of (1.1) in \( B_{2\varepsilon}(x_0) \cap \mathcal{D}(\Theta) \). From (19) it follows immediately that \( \{D_{\xi_n}\Theta(\hat{x}, x_n)\} \) is monotonically decreasing. Moreover, if \( F(x_n) \neq y \), then by using (3.15), (3.16) and the same derivation of (3.11), we have

\[
\mu_{n,k} \|s_{n,k}\|_{\mathcal{Y}}^p \geq c_0 \gamma^p \|y - F(x_n)\|_{\mathcal{Y}}^p.
\]

The combination of (20), (21) and the fact \( k_0 \geq 1 \) gives

\[
c_0 c_0 \gamma^p \|y - F(x_n)\|_{\mathcal{Y}}^p \leq D_{\xi_n}\Theta(\hat{x}, x_n) - D_{\xi_{n+1}}\Theta(\hat{x}, x_{n+1})
\]

which holds automatically when \( F(x_n) = y \). This together with the monotonicity of \( \{D_{\xi_n}\Theta(\hat{x}, x_n)\} \) implies that \( \|F(x_n) - y\|_{\mathcal{Y}} \to 0 \) as \( n \to \infty \).

According to Proposition 3.2, it remains only to show that there is an increasing sequence \( \{n_l\} \) satisfying \( \lim_{l \to \infty} n_l = \infty \) such that (3.18) holds. To this end, we recall that \( \lim_{n \to \infty} \|y - F(x_n)\|_{\mathcal{Y}} = 0 \). Moreover, if \( F(x_n) = y \) for some \( n \), then by definition it follows that \( x_m = x_n \) for all \( m \geq n \). Therefore

\[
\|F(x_n) - y\|_{\mathcal{Y}} = 0 \quad \text{for some} \quad n \quad \Rightarrow \quad \|F(x_m) - y\|_{\mathcal{Y}} = 0 \quad \text{for all} \quad m \geq n.
\]

Thus we can introduce a subsequence \( \{n_l\} \) by setting \( n_0 = 0 \) and letting \( n_l \), for each \( l \geq 1 \), be the first integer satisfying

\[
n_l \geq n_{l-1} + 1 \quad \text{and} \quad \|F(x_{n_l}) - y\|_{\mathcal{Y}} \leq \|F(x_{n_{l-1}}) - y\|_{\mathcal{Y}}.
\]
For such chosen strictly increasing sequence \{n_i\} it is easy to see that
\[
\|F(x_{n_i}) - y\|_Y \leq \|F(x_n) - y\|_Y, \quad 0 \leq n \leq n_i. \tag{3.22}
\]

We now consider \(\langle \xi_{n_i} - \xi_{n_{j}}, x_{n_i} - \hat{x} \rangle_{X^*,X}\) for \(0 \leq j < l < \infty\). Since \(\xi_n = \xi_{n,0}\) and \(\xi_{n+1} = \xi_{n,k_n}\), we can write
\[
\langle \xi_{n_i} - \xi_{n_{j}}, x_{n_i} - \hat{x} \rangle_{X^*,X} = \sum_{n=n_j}^{n_i-1} \langle \xi_{n_{j}+1} - \xi_{n_{j}}, x_{n_i} - \hat{x} \rangle_{X^*,X} \tag{3.23}
\]
By the definition of \(\xi_{n,k}\) and the property of \(J^Y_F\) we have
\[
|\langle \xi_{n,k_n} - \xi_{n,0}, x_{n_l} - \hat{x} \rangle_{X^*,X}| = \sum_{k=0}^{k_n-1} \left| \sum_{n=n_{j}}^{n_{j+1}-1} \langle \xi_{n_{j}} - \xi_{n_{j+1}}, x_{n_i} - \hat{x} \rangle_{X^*,X} \right| \leq \sum_{k=0}^{k_n-1} \mu_{n,k} \|J^Y_F(s_{n,k})L(x_{n})(x_{n_i} - \hat{x})\|_Y, \tag{3.24}
\]
By the condition (c) in Assumption 3.1, we have
\[
\|L(x_{n})(x_{n_i} - \hat{x})\|_Y \leq \|L(x_{n})(x_{n_i} - x_{n})\|_Y + \|L(x_{n_i} - \hat{x})\|_Y \leq (1 + \eta) \|F(x_{n_i}) - F(x_{n})\|_Y + \|F(x_{n}) - y\|_Y \leq (1 + \eta) \|F(x_{n_i}) - y\|_Y + 2\|F(x_{n}) - y\|_Y.
\]
Consequently, we can use (3.22) to derive that
\[
\|L(x_{n})(x_{n_i} - \hat{x})\|_Y \leq 3(1 + \eta)\|F(x_{n_i}) - y\|_Y, \quad 0 \leq n \leq n_i.
\]
This together with (3.16) gives for \(0 \leq n \leq n_i\) and \(0 \leq k < k_n\) that
\[
\|L(x_{n})(x_{n_i} - \hat{x})\|_Y \leq \frac{3(1 + \eta)}{\gamma} \|s_{n,k}\|_Y.
\]
Therefore it follows from (3.24) and (3.20) that
\[
|\langle \xi_{n,k_n} - \xi_{n,0}, x_{n_i} - \hat{x} \rangle_{X^*,X}| \leq \frac{3(1 + \eta)}{\gamma} \sum_{k=0}^{k_n-1} \mu_{n,k} \|s_{n,k}\|_Y \leq c_4 \left( D_{\xi_{n}} \Theta(\hat{x}, x_{n}) - D_{\xi_{n+1}} \Theta(\hat{x}, x_{n+1}) \right),
\]
where \(c_4 := 3(1 + \eta)/(\gamma c_3)\). This together with (3.23) gives
\[
|\langle \xi_{n_i} - \xi_{n_{j}}, x_{n_i} - \hat{x} \rangle_{X^*,X}| \leq c_4 \left( D_{\xi_{n_i}} \Theta(\hat{x}, x_{n_i}) - D_{\xi_{n_{j}} \Theta(\hat{x}, x_{n_{j}})}\right)
\]
which, in view of the monotonicity of \(\{D_{\xi_{n}}(\hat{x}, x_{n})\}\), implies (3.18).
To show the last part under the condition \( \mathcal{N}(L(x^1)) \subset \mathcal{N}(L(x)) \) for all \( x \in B_{2\xi}(x_0) \), we observe from the definition of \( \xi_n \) that
\[
\xi_{n+1} - \xi_n \in \mathcal{R}(L(x_n^*)) \subset \mathcal{N}(L(x_n))^{\bot} \subset \mathcal{N}(L(x^1))^{\bot} = \mathcal{R}(L(x^1)).
\]
Thus, we may use the second part of Proposition 3.2 to conclude the proof. \( \Box \)

The following result shows that we in fact have a certain uniform convergence result for all the sequences \( \{ (\xi_n, x_n) \} \in \Gamma_\gamma,\mu,\rho_1(\xi_0, x_0) \) which will be crucial in proving the regularization property of Algorithm 3.1.

**Lemma 3.4.** Assume all the conditions in Theorem 3.3 hold. Assume also that
\[
\mathcal{N}(L(x^1)) \subset \mathcal{N}(L(x)), \quad \forall x \in B_{2\xi}(x_0).
\] (3.25)

Then, for any \( \varepsilon > 0 \), there is an integer \( n(\varepsilon) \) such that for any sequence \( \{ (\xi_n, x_n) \} \in \Gamma_\gamma,\mu,\rho_1(\xi_0, x_0) \) there holds \( D_{\xi_n} \Theta(x^1, x_n) < \varepsilon \) for all \( n \geq n(\varepsilon) \).

**Proof.** We will use a contradiction argument. Assume that the result is not true. Then there is an \( \varepsilon_0 > 0 \) such that for any \( \ell \geq 1 \) there exist \( \{ (\xi^{(\ell)}_n, x^{(\ell)}_n) \} \in \Gamma_\gamma,\mu,\rho_1(\xi_0, x_0) \) and \( n_\ell \geq \ell \) such that
\[
D_{\xi^{(\ell)}_n} \Theta(x^1_{n_\ell}, x^{(\ell)}_{n_\ell}) \geq \varepsilon_0.
\] (3.26)

We will construct, for each \( n = 0, 1, \ldots \), a strictly increasing subsequence \( \{ \ell_{n,j} \} \) of positive integers and \( (\xi_n, x_n) \in \mathcal{X}^* \times \mathcal{X} \) such that

(i) \( \{ (\xi_{\ell_{n,j}}, x_{\ell_{n,j}}) \} \in \Gamma_\gamma,\mu,\rho_1(\xi_0, x_0) \);

(ii) for each fixed \( n \) there hold \( \xi_{\ell_{n,j}} = \xi_n \) and \( x_{\ell_{n,j}} = x_n \) for all \( j \).

Assume that the above construction is available, we will derive a contradiction. According to (i), it follows from Theorem 3.3 that \( D_{\xi_{\ell_{n,j}}} \Theta(x^1, x_{\ell_{n,j}}) \rightarrow 0 \) as \( n \rightarrow \infty \). Thus we can pick a sufficiently large integer \( n_0 \) such that
\[
D_{\xi_{\ell_{n,j}}} \Theta(x^1, x_{n_0}) < \varepsilon_0.
\]

Let \( \ell_0 := \ell_{n_0,n_0} \) and consider the sequence \( \{ (\xi^{(\ell_0)}_n, x^{(\ell_0)}_n) \} \). According to (ii), the fact \( n_{\ell_0} > \ell_0 \geq n_0 \), and the monotonicity of \( \{ D_{\xi^{(\ell_0)}_n} \Theta(x^1, x^{(\ell_0)}_n) \} \) with respect to \( n \), we have
\[
\varepsilon_0 > D_{\xi_{\ell_{n,j}}} \Theta(x^1, x_{n_0}) = D_{\xi_{\ell_{n,j}}} \Theta(x^1, x^{(\ell_0)}_{n_0}) \geq D_{\xi_{\ell_{n,j}}} \Theta(x^1, x^{(\ell_0)}_{n_{\ell_0}})
\]
which is a contradiction to (3.26) with \( \ell = \ell_0 \).

We turn to the construction of \( \{ \ell_{n,j} \} \) and \( (\xi_n, x_n) \), for each \( n = 0, 1, \ldots \), such that (i) and (ii) hold. For \( n = 0 \), we take \( (\xi_0, x_0) = (\xi_0, x_0) \) and \( \ell_{0,j} = j \) for all \( j \).

Next, assume that we have constructed \( \{ \ell_{m,j} \} \) and \( (\xi_n, x_n) \) for all \( 0 \leq n \leq m \).

We will construct \( \{ \ell_{m+1,j} \} \) and \( (\xi_{m+1}, x_{m+1}) \). If \( F(x_m) = y \), then, by the induction hypothesis on (ii), \( F(x^{(\ell_{m,j})}_m) = y \) for all \( j \). By definition we then have
\[
\xi^{(\ell_{m,j})}_{m+1} = \xi^{(\ell_{m,j})}_m = \xi_m \quad \text{and} \quad x^{(\ell_{m,j})}_{m+1} = x^{(\ell_{m,j})}_m = x_m.
\]
Therefore, we can define
\[
\{ \ell_{m+1,j} \} := \{ \ell_{m,j} \}, \quad \xi_{m+1} := \xi_m \quad \text{and} \quad x_{m+1} := x_m.
\]
So we may assume $F(\bar{x}_m) \neq y$. Let $k_m^{(\ell)} \geq 1$ be the integer used to define $(\xi_m^{(\ell)}, \bar{x}_m^{(\ell)})$ from $(\xi_m^{(\ell)}, \bar{x}_m^{(\ell)})$. From the proof of Lemma 3.1 we can see that $k_m^{(\ell)}$ is bounded from above by a constant depending only on $D_{\xi_m} \Theta(x^\dagger, \bar{x}_m)$ and $\|F(\bar{x}_m) - y\|_Y$ but independent of $j$. Thus $\{\ell_{m,j}\}$ must have a subsequence, denoted as $\{\ell_{m+1,j}\}$, such that all $k_m^{(\ell_{m+1,j})}$ take the same integer value, which is denoted as $\bar{k}_m$. Since $\xi_m^{(\ell_{m+1,j})} = \xi_m$ and $x_m^{(\ell_{m+1,j})} = \bar{x}_m$ by the induction hypothesis, we can conclude by definition that $\xi_{m,k}$ and $x_{m+1,j}$ are independent of $\ell$ for all $0 \leq k \leq \bar{k}_m$. Consequently $\xi_{m+1,j}$ and $x_{m+1,j}$ are independent of $j$. Let $\xi_{m+1} = \xi_{m+1,j}$ and $\bar{x}_{m+1} = x_{m+1,j}$.

From the definition of $(\epsilon_{m+1,j}, \delta_{m+1,j}^{(\ell)})$ it is easy to see that $\xi_{m+1} = \xi_{m,k}$ and $\bar{x}_{m+1} = \bar{x}_{m,k}$ with $\xi_{m,k}$ and $\bar{x}_{m,k}$ being defined similarly by (3.14) starting from $(\xi_m, \bar{x}_m)$; moreover, $\bar{k}_m \geq 1$ is an integer such that

$$\|y - F(\bar{x}_m) - L(\bar{x}_m)(\bar{x}_{m,k} - \bar{x}_m)\|_Y \geq \gamma \|y - F(\bar{x}_m)\|_Y, \quad 0 \leq k \leq \bar{k}_m.$$ 

We therefore complete the construction of $\{\ell_{m+1,j}\}$ and $(\xi_{m+1}, \bar{x}_{m+1})$. The proof is thus complete. $\square$

### 3.3. Convergence

In this subsection we show that $\lim_{\delta \to 0} D_{\xi_n} \Theta(x^\dagger, x_{n,k}^\delta) = 0$ for the sequence $\{(\xi_n^\delta, x_{n,k}^\delta)\}$ defined by Algorithm 3.1. To this end, we need to establish some stability results on the algorithm so that Theorem 3.3 and Lemma 3.4 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 3.5.** Let all the conditions in Lemma 3.1 hold. Let $\{y^\delta_l\}$ be a sequence of noisy data satisfying $\|y^\delta_l - y\|_Y \leq \delta_l$ with $\delta_l \to 0$ as $l \to \infty$. For any integer $n \leq \liminf_{l \to \infty} n_{\delta_l}$, let $\xi_n^\delta_l$ and $x_n^\delta_l$ be defined by Algorithm 3.1. If

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

for some $(\xi_n, x_n) \in \mathcal{X}^* \times \mathcal{X}$ and if we define $\xi_{n,k}$ and $x_{n,k}$ according to (3.14), then for each integer $k = 0, 1, \cdots$ there holds

$$\xi_{n,k}^\delta_l \to \xi_{n,k} \quad \text{and} \quad x_{n,k}^\delta_l \to x_{n,k} \quad \text{as} \ l \to \infty.$$

**Proof.** We prove this result by induction on $k$. It is trivial for $k = 0$ by the condition since $\xi_{n,0} = \xi_n$ and $x_{n,0} = x_n$. We next assume that the result is true for some $k \geq 0$ and show that $\xi_{n,k+1}^\delta_l \to \xi_{n,k+1}$ and $x_{n,k+1}^\delta_l \to x_{n,k+1}$ as $l \to \infty$. We consider two cases.

**Case 1:** $s_{n,k} = 0$. In this case we have $\xi_{n,k+1} = \xi_{n,k}$. Therefore

$$\xi_{n,k+1}^\delta_l - \xi_{n,k+1} = \xi_{n,k}^\delta_l - \xi_{n,k} + \mu_{n,k}^\delta_l L(x_{n,k}^\delta_l)^* J^\mathcal{Y}_l(x_{n,k}^\delta_l).$$

Since $\mu_{n,k}^\delta_l \leq \mu_1 \|s_{n,k}^\delta_l\|^{\rho - \tau}$, we may use the property of $J^\mathcal{Y}_l$ to obtain

$$\|\xi_{n,k+1}^\delta_l - \xi_{n,k+1}\|_{\mathcal{X}^*} \leq \|\xi_{n,k}^\delta_l - \xi_{n,k}\|_{\mathcal{X}^*} + \mu_1 B \|s_{n,k}^\delta_l\|^{\rho - 1}.$$ 

By the induction hypothesis and the continuity of $F$ and $L$, we then have $\xi_{n,k+1} \to \xi_{n,k+1}$ as $l \to \infty$. Consequently, by using the continuity of $\nabla \Theta^*$ we have $x_{n,k+1}^\delta_l = \nabla \Theta^*(s_{n,k+1}^\delta_l) \to \nabla \Theta^*(\xi_{n,k+1}) = x_{n,k+1}$ as $l \to \infty$. 

**Case 2:** $s_{n,k} \neq 0$. In this case we have $\xi_{n,k+1} = \xi_{n,k}$. Therefore
Case 2: $s_{n,k} \neq 0$. We claim that $\mu_{n,k}^{\delta_i} \to \mu_{n,k}$ as $l \to \infty$. To see this, let

$$
\hat{\mu}_{n,k} = \min \left\{ \frac{\mu_0 \| s_{n,k} \|^p}{\| L(J_r^*(\Lambda \Lambda_{n,k})) \|^p_r}, \mu_1 \right\}, \quad \hat{\mu}_{n,k}^{\delta_i} = \min \left\{ \frac{\mu_0 \| s_{n,k} \|^p}{\| L(J_r^*(\Lambda \Lambda_{n,k})) \|^p_r}, \mu_1 \right\}.
$$

If $L(J_r^*(\Lambda \Lambda_{n,k})) = 0$, then, by definition, we must have $\hat{\mu}_{n,k} = \mu_1$ and $\hat{\mu}_{n,k}^{\delta_i} = \mu_1$ for sufficiently large $l$. If $L(J_r^*(\Lambda \Lambda_{n,k})) \neq 0$, then, by the induction hypothesis, we can conclude that $\hat{\mu}_{n,k} \to \mu_{n,k}$ as $l \to \infty$. In any case, we always have $\hat{\mu}_{n,k}^{\delta_i} \to \hat{\mu}_{n,k}$ as $l \to \infty$. Since

$$
\mu_{n,k} = \hat{\mu}_{n,k} \quad \text{and} \quad \mu_{n,k}^{\delta_i} = \hat{\mu}_{n,k}^{\delta_i},
$$

we can obtain $\mu_{n,k}^{\delta_i} \to \mu_{n,k}$ as $l \to \infty$. Consequently, it follows from the induction hypotheses and the continuity of $F$ that $\xi_{n,k+1}^{\delta_i} \to \xi_{n,k+1}$ and hence $x_{n,k+1}^{\delta_i} \to x_{n,k+1}$ as $l \to \infty$ using again the continuity of $\nabla \Theta$.

**Lemma 3.6.** Let all the conditions in Lemma 3.1 hold. Let $\{y^{\delta_i}\}$ be a sequence of noisy data satisfying $\| y^{\delta_i} - y \|_Y \leq \delta_i$ with $\delta_i \to 0$ as $l \to \infty$. Let $\xi_{n,k}^{\delta_i}$ and $x_{n,k}^{\delta_i}$, $0 \leq n \leq n_{\delta_i}$, be defined by Algorithm 3.1. Then for any $n \leq \lim_{l \to \infty} n_{\delta_i}$, by taking a subsequence of $\{y^{\delta_i}\}$ if necessary, there is a sequence $\{(\xi_m, x_m)\} \in (0, x_0)$ such that

$$
\lim_{l \to \infty} D_{\xi_{n,k}^{\delta_i}} \Theta(x_m, x_m^{\delta_i}) = 0
$$

for all $0 \leq m \leq n$. Furthermore, let $n_0$ be the first integer such that $F(x_{n_0}) = y$, then

$$
\xi_{n_0}^{\delta_i} \to \xi_{n_0} \quad \text{as} \quad l \to \infty
$$

for all $0 \leq m \leq \min\{n_0, n_{\delta_i}\}$. 

**Proof.** If $\lim_{l \to \infty} n_{\delta_i} = 0$, nothing needs to prove since $\xi_{n_0}^{\delta_i} = \xi_0$ and $x_{n_0}^{\delta_i} = x_0$. Therefore, we may assume $\lim_{l \to \infty} n_{\delta_i} \geq 1$ and complete the proof by induction. When $n = 0$, the result is again trivial. Assume next that, for some $0 \leq n < \lim_{l \to \infty} n_{\delta_i}$, the result is true for some sequence $\{(\xi_m, x_m)\} \in (0, x_0)$.

In order to show the result is also true for $n + 1$, we will obtain a sequence from $\Gamma_{\gamma,\phi_0,\mu_1}(\xi_0, x_0)$ by retaining the first $n + 1$ terms in $\{(\xi_m, x_m)\}$ and modifying the remaining terms. It suffices to redefine $\xi_{n+1}$ and $x_{n+1}$ since then we can apply the algorithm for exact data to produce the remaining terms. We consider two cases:

**Case 1:** $F(x_n) = y$. By definition we have $\xi_{n+1} = \xi_n$ and $x_{n+1} = x_n$. Since $x_n$ is a solution of $F(x) = y$, we may use the monotonicity result in Lemma 3.1 and the induction hypothesis to derive that

$$
D_{\xi_{n+1}} \Theta(x_{n+1}, x_{n+1}^{\delta_i}) = D_{\xi_{n+1}} \Theta(x_n, x_n^{\delta_i}) \leq D_{\xi_n} \Theta(x_n, x_n^{\delta_i}) \to 0
$$

as $l \to \infty$.

**Case 2:** $F(x_n) \neq y$. Let $\bar{k}_n \geq 1$ be the first integer such that

$$
\| s_{n,k} \|_Y < \gamma \| y - F(x_n) \|_Y
$$

which is finite according to the proof of Lemma 3.1. We claim that $k_n^{\delta_i} \leq \bar{k}_n$ for large $l$. To see this, we observe that $x_{n,k}^{\delta_i} \to x_{n,k_n}^{\delta_i}$ as $l \to \infty$ by virtue of Lemma 3.5.
Therefore, by using the induction hypothesis and the continuity of $F$ and $L$, there must hold
\[ \|s_{n,k}^{\delta_l} \|_{\mathcal{Y}} < \gamma \| y^{\delta_l} - F(x_n) \|_{\mathcal{Y}} \]
for sufficiently large $l$. According to the definition of $k_n^{\delta_l}$ we must have $k_n^{\delta_l} \leq k_n$ for large $l$. This shows that $\{k_n^{\delta_l}\}$ takes only finite many integer values. By taking a subsequence if necessary, we may assume that all $k_n^{\delta_l}$ takes the same integer value, which is denoted as $k_n$. By the definition of $k_n^{\delta_l}$ we have
\[ \|s_{n,k} \|_{\mathcal{Y}} \geq \gamma \| y^{\delta_l} - F(x_n) \|_{\mathcal{Y}}, \quad 0 \leq k < k_n. \]
Letting $l \to \infty$ and using Lemma 3.5 it follows
\[ \|s_{n,k} \|_{\mathcal{Y}} \geq \gamma \| y - F(x_n) \|_{\mathcal{Y}}, \quad 0 \leq k < k_n. \]
Using this $k_n$ we can define $\xi_{n+1} := \xi_{n,k_n}$ and $x_{n+1} := x_{n,k_n}$, In view of Lemma 3.5, it is clear that $\xi_{n+1}^{\delta_l} \to \xi_{n+1}$ and $x_{n+1}^{\delta_l} \to x_{n+1}$ as $l \to \infty$. Moreover, by the lower semi-continuity of $\Theta$ we have
\[
\begin{align*}
\limsup_{l \to \infty} D_{\xi_{n+1}}^{\delta_l} \Theta(x_{n+1}, x_{n+1}^{\delta_l}) \\
= \Theta(x_{n+1}) - \liminf_{l \to \infty} \Theta(x_{n+1}^{\delta_l}) = \lim_{l \to \infty} (\xi_{n+1}^{\delta_l}, x_{n+1})_{\mathcal{X}^*, \mathcal{X}} \\
\leq \Theta(x_{n+1}) - \Theta(x_{n+1}) = 0.
\end{align*}
\]
The proof is therefore complete. $\Box$

**Remark 3.3.** In Lemma 3.6, when $F(x_n) = y$ for some $n < \liminf_{l \to \infty} n^{\delta_l}$, we can not guarantee that $\lim_{l \to \infty} \xi_{n+1} = \xi_{n+1}$ since we do not have a good control on $k_n^{\delta_l}$. However, we can guarantee that $\lim_{l \to \infty} D_{\xi_{n+1}}^{\delta_l} \Theta(x_{n+1}, x_{n+1}^{\delta_l}) = 0$ which is enough for our purpose.

We are now ready to give the main convergence result concerning the sequence $\{(\xi_{n}, x_{n})\}$ defined by Algorithm 3.1.

**Theorem 3.7.** Let $\mathcal{X}$ be reflexive and let $\mathcal{Y}$ be uniformly smooth, let $\Theta : \mathcal{X} \to (-\infty, \infty]$ be proper, lower semi-continuous, and $p$-convex with $p \geq 2$ satisfying (2.1) for some $c_0 > 0$, let Assumption 3.1 hold with $0 \leq \eta < 1$, and let (3.3) be satisfied. Let $\mu_1 > 0$ be a given number, and let $\eta < \gamma < 1$, $\mu_0 > 0$ and $\tau > 1$ be chosen such that (3.7) holds. Assume further that (3.25) holds. Then for Algorithm 3.1 there hold
\[
\lim_{\delta \to 0} \|x_{n, \delta}^\delta - x^\dagger\|_{\mathcal{X}} = 0 \quad \text{and} \quad \lim_{\delta \to 0} D_{\xi_{n, \delta}} \Theta(x^\dagger, x_{n, \delta}^\delta) = 0.
\]

**Proof.** Due to the $p$-convexity of $\Theta$, it suffices to show $\lim_{\delta \to 0} D_{\xi_{n, \delta}} \Theta(x^\dagger, x_{n, \delta}^\delta) = 0$. We complete the proof by considering two cases.

Assume first that $\{y^{\delta_l}\}$ is a sequence satisfying $\| y^{\delta_l} - y \|_{\mathcal{Y}} \leq \delta_l$ with $\delta_l \to 0$ such that $n_l := n_{\delta_l} \to n_0$ as $l \to \infty$ for some finite integer $n_0$. We may assume $n_l = n_0$ for all $l$. According to Lemma 3.6, by taking a subsequence of $\{y^{\delta_l}\}$ if necessary, we can find a sequence $\{(\xi_{n}, x_{n})\} \in \Gamma_{\gamma, \mu_0, \mu_1}(\xi_0, x_0)$ such that
\[
D_{\xi_{n, \delta}} \Theta(x_{n, \delta}, x_{n, \delta}^{\delta_l}) \to 0 \quad \text{as} \quad l \to \infty. \tag{3.27}
\]
This together with the p-convexity of $\Theta$ implies $x_{n_l}^\delta \to x_{n_0}$ as $l \to \infty$. From the definition of $n_0 := n_l$ we have $\|F(x_{n_l}^\delta) - y^\delta\| \leq \tau \delta_l$. By taking $l \to \infty$, we can obtain $F(x_{n_0}) = y$. This then implies $x_n = x_{n_0}$ for all $n \geq n_0$. Since $\{(\xi_n, x_n)\} \in \Gamma_{\gamma, \mu_0, \mu_1}(\xi_0, x_0)$, we can use (3.25) and Theorem 3.3 to conclude that $\lim_{n \to \infty} x_n = x^\dagger$. Thus $x_{n_0} = x^\dagger$ which together with (3.27) gives $\lim_{n \to \infty} D_{\xi_{n_l}} \Theta(x^\dagger, x_{n_0}^\delta) = 0$.

Assume next that $\{y^\delta\}$ is a sequences satisfying $\|y^\delta - y\| \leq \delta_l$ with $\delta_l \to 0$ such that $n_l := n_{\delta_l} \to \infty$ as $l \to \infty$. Let $\varepsilon > 0$ be an arbitrary but fixed number. According to Lemma 3.4, there is an integer $n(\varepsilon)$ such that

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

For this $n(\varepsilon)$, by using Lemma 3.6 and by taking a subsequence of $\{y^\delta\}$ if necessary, we can find $\{(\xi_n, x_n)\} \in \Gamma_{\gamma, \mu_0, \mu_1}(\xi_0, x_0)$ such that

$$\lim_{l \to \infty} D_{\xi_{n_l}} \Theta(x_{n_l}^\delta) = 0$$

for $0 \leq n \leq n(\varepsilon)$. Let $n_\ast$ be the first integer such that $F(x_n) = y$ and let $\hat{n} = \min\{n(\varepsilon), n_\ast\}$. Then by the construction of $\{(\xi_n, x_n)\}$ we have $\xi_{n(\varepsilon)} = \xi_{\hat{n}}$ and $x_{n(\varepsilon)} = x_{\hat{n}}$. Moreover, it follows from Lemma 3.6 that

$$\xi_{\hat{n}}^\delta \to \xi_{\hat{n}} \quad \text{and} \quad x_{\hat{n}}^\delta \to x_{\hat{n}} \quad \text{as} \quad l \to \infty.$$

Since $n_l > \hat{n}$ for large $l$, by the monotonicity result in Lemma 3.1 we have

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

By using the lower semi-continuity of $\Theta$ we obtain

$$\limsup_{l \to \infty} D_{\xi_{n_l}} \Theta(x^\dagger, x_{n_l}^\delta) \leq \Theta(x^\dagger) - \liminf_{l \to \infty} \Theta(x_{n_l}^\delta) - \lim_{l \to \infty} \langle \xi_{n_l}^\delta, x^\dagger - x_{n_l}^\delta \rangle x^*, x.$$

Since $\varepsilon > 0$ can be arbitrarily small, we therefore obtain $\lim_{l \to \infty} D_{\xi_{n_l}} \Theta(x^\dagger, x_{n_l}^\delta) = 0$. The proof is complete. 

**Remark 3.4.** In each inner iteration of Algorithm 3.1, we define $k_\ast^\delta$ to be the first integer satisfying (3.6) which is shown to be finite. In case $k_\ast^\delta$ is huge, the computation could get stuck at that step. To avoid this, we may take a preassigned integer $k_{\max} \geq 1$ and redefined $k_\ast^\delta$ to be the first integer $k \leq k_{\max}$ such that (3.6) holds; if such an integer does not exist, we take $k_\ast^\delta := k_{\max}$. Under this modification of Algorithm 3.1, all the results in this paper still hold without any change.

**Remark 3.5.** In the formulation of Algorithm 3.1, every $x_{n,k}^\delta$ is determined by solving a minimization problem related to $\Theta$. Although the exact minimizers can be found for some special choices of $\Theta$, such minimization problems in general can only be solved inexactly by iterative procedures. Thus, in practical computations, certain iterative procedure should be employed to determine $x_{n,k}^\delta$ satisfying

$$\Theta(x_{n,k}^\delta) - \langle \xi_{n,k}^\delta, x_{n,k}^\delta \rangle x^*, x \leq \min \left\{ \Theta(x) - \langle \xi_{n,k}^\delta, x \rangle x^*, x \right\} - \varepsilon_{n,k}.$$
where $\varepsilon_{n,k} > 0$ is a small number. By replacing the determination of $x_{n,k}$ in Algorithm 3.1 with the above inexact procedure, it is possible to show that the resulting algorithm is still a regularization method if $\sum_{n=1}^{\infty} \sum_{k=1}^{k_n} \varepsilon_{n,k} < \infty$. Since the argument involves new tools such as $\varepsilon$-subdifferential calculus and duality theory, we plan to report such result in a future work.

4. Numerical examples. In this section we present some numerical simulations to test the performance of Algorithm 3.1. Our tests were done by using MATLAB R2012a on a Lenovo laptop with Intel(R) Core(TM) i5 CPU 2.30 GHz and 6 GB memory.

Because implementing Algorithm 3.1 requires efficient solvers to find the minimizer of the minimization problem

$$\bar{x} = \arg \min_{x \in \mathcal{X}} \left\{ \Theta(x) - \langle \bar{\xi}, x \rangle_{\mathcal{X}^*, \mathcal{X}} \right\}$$

(4.1)

for any $\bar{\xi} \in \mathcal{X}^*$, it is necessary to give some discussions on the resolution of (4.1).

(i) When $\mathcal{X}$ is a Hilbert space and $\Theta(x) = \frac{1}{2} \|x\|_{\mathcal{X}}^2$, the minimizer of (4.1) for any $\bar{\xi} \in \mathcal{X}^*$ is given by $\bar{x} = \bar{\xi}$. Thus, when $\mathcal{Y}$ is also a Hilbert space and $J^Y_r = J^Y_2$, Algorithm 3.1 reduces to the inexact Newton-Landweber iteration in Hilbert spaces ([21]).

(ii) When $\mathcal{X} = L^2(\Omega)$ and the sought solution is sparse, we may consider

$$\Theta(x) := \frac{1}{2\beta} \|x\|_{L^2}^2 + \|x\|_{L^1}$$

(4.2)

with $\beta > 0$. For this $\Theta$, the minimizer of (4.1) is given explicitly by

$$\bar{x} = \beta \text{sign}(\bar{\xi}) \max \left\{ |\bar{\xi}| - 1, 0 \right\}.$$

(iii) When the sought solution is piecewise constant, we take $\Theta$ to be total variation (TV) like functions. For numerical simulations, we use the discrete TV function

$$|x|_{TV} = \sum_{i=1}^{M} \|D^T_i x\|_2,$$

where $x \in \mathbb{R}^M$, $M = m^2$, are vectors coming from the column-wise stacking of $m \times m$ square images, and each $D^T_i x$ is the discrete gradient of $x$ at the pixel $i$ defined by

$$D^T_i x = \begin{cases} (x[i + 1] - x[i], x[i + m] - x[i])^T & \text{if } i \mod m \neq 0 \text{ and } i \leq M - m, \\
(0, x[i + m] - x[i])^T & \text{if } i \mod m = 0 \text{ and } i \leq M - m, \\
(x[i + 1] - x[i], 0)^T & \text{if } i \mod m \neq 0 \text{ and } i > M - m, \\
(0, 0)^T & \text{if } i \mod m = 0 \text{ and } i > M - m 
\end{cases}$$

with $x[i]$ denoting the $i$-th component of $x$, and $\| \cdot \|_2$ denoting the 2-norm. We take

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

(4.3)

with $\beta > 0$. Then the corresponding minimization problem (4.1) with $\mathcal{X} = \mathbb{R}^M$ becomes

$$\bar{x} = \arg \min_{x \in \mathbb{R}^M} \left\{ \frac{1}{2\beta} \|x - \beta \bar{\xi}\|_2^2 + |x|_{TV} \right\}.$$  

(4.4)
This is a total variation denoising problem \cite{23} for which many efficient solvers have been developed in recent years; for instance, see \cite{2, 3} for the fast iterative shrinkage/thresholding algorithm and \cite{27} for the primal-dual hybrid gradient (PDHG) method. We will use the PDHG method which can be described as follows. First we can rewrite (4.4) as

\[
\bar{x} = \arg \min_{x \in \mathbb{R}^M} \max_{z \in \mathbb{R}^{2M}} \Phi(x, z), \quad \Phi(x, z) = \frac{1}{2\beta} \|x - \beta \bar{\xi}\|_2^2 + x^T D z - \iota_Z(z),
\]

where \( D = (D_1, \cdots, D_M) \in \mathbb{R}^{M \times 2M} \) and \( \iota_Z \) denotes the indicator function of \( Z \) with \( Z \) defined by

\[
Z = \{ z = (z_1, \cdots, z_M)^T \in \mathbb{R}^{2M} : z_l \in \mathbb{R}^2 \text{ and } \|z_l\|_2 \leq 1 \text{ for } l = 1, \cdots, M \}.
\]

Then the PDHG method takes the form

\[
z^{k+1} = \arg \max_{z \in \mathbb{R}^{2M}} \left\{ \Phi(x^k, z) - \frac{\beta}{2\tau_k} \|z - z^k\|_2^2 \right\},
\]

\[
x^{k+1} = \arg \min_{x \in \mathbb{R}^M} \left\{ \Phi(x, z^{k+1}) + \frac{1 - \theta_k}{2\beta \theta_k} \|x - x^k\|_2^2 \right\}
\]

with suitably chosen step sizes \( \tau_k > 0 \) and \( 0 < \theta_k < 1 \). The above maximization and minimization problems can be solved explicitly, giving the following explicit formula of the PDHG method

\[
z^{k+1} = \Pi_Z \left( z^k + \beta^{-1} \tau_k D^T x^k \right),
\]

\[
x^{k+1} = (1 - \theta_k) x^k + \beta \theta_k \left( \bar{\xi} - D z^{k+1} \right),
\]

where \( \Pi_Z \) denotes the projection onto the set \( Z \) given by

\[
(\Pi_Z(z))_l = \frac{z_l}{\max\{\|z_l\|_2, 1\}}, \quad l = 1, \cdots, M.
\]

To achieve the fast convergence, it was suggested in \cite{27} to choose the step sizes as

\[
\tau_k = 0.2 + 0.08k, \quad \theta_k = \left( 0.5 - \frac{5}{15 + k} \right) / \tau_k.
\]

The convergence of the PDHG method, under such a choice of the step sizes, was confirmed in \cite{5}.

\textbf{4.1. Parameter identification by interior measurements.} We first consider the identification of the parameter \( c \) in the boundary value problem

\[
\begin{align*}
-\Delta u + cu &= f \quad \text{in } \Omega, \\
u &= g \quad \text{on } \partial \Omega
\end{align*}
\]

from an \( L^2(\Omega) \)-measurement of the state \( u \), where \( \Omega \subset \mathbb{R}^2 \) is a bounded domain with Lipschitz boundary, \( f \in L^2(\Omega) \) and \( g \in H^{3/2}(\Omega) \). This is a benchmark example of nonlinear inverse problems. We assume that the exact solution \( c^* \) is in \( L^2(\Omega) \). This problem reduces to solving \( F(c) = u \), if we define the nonlinear operator \( F : L^2(\Omega) \rightarrow L^2(\Omega) \) by \( F(c) := u(c) \), where \( u(c) \in H^2(\Omega) \subset L^2(\Omega) \) is the unique solution of (4.5). This operator \( F \) is well defined on

\[
\mathcal{D}(F) := \{ c \in L^2(\Omega) : \|c - \hat{c}\|_{L^2(\Omega)} \leq \gamma_0 \text{ for some } \hat{c} \geq 0, \text{ a.e.} \}
\]
for some positive constant $\gamma_0 > 0$. It is known that $F$ is Fréchet differentiable; the Fréchet derivative of $F$ and its adjoint are given by

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

(4.6)

for $h, w \in L^2(\Omega)$, where $A(c) : H^2(\Omega) \cap H^1_0(\Omega) \to L^2(\Omega)$ is defined by $A(c)u = -\Delta u + cu$ which is an isomorphism uniformly in ball $B_{2\rho}(c^1)$ for small $\rho > 0$. Moreover, Assumption 3.1 holds for small $\rho > 0$ (see [12]).

![Reconstruction results for the sought solution that is piecewise constant: (a) exact solution; (b) Algorithm 3.1 with $\Theta(c) = \frac{1}{2\mu}||c||_2^2$; (c) Algorithm 3.1 with $\Theta(c) = \frac{1}{2\mu}||c||_2^2 + |c|\tau V$.](image)

We will present two numerical results on $\Omega = [0, 1] \times [0, 1]$: one is to reconstruct the solution that is piecewise constant and the other is to recover the solution that is sparse. We will take $g \equiv 1$ on $\partial \Omega$ and

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

which can be considered as approximation of suitable multiple of Dirac delta function. In order to carry out the computation, we divide $\Omega$ into 120 $\times$ 120 small squares of equal size. All partial differential equations involved are solved approximately by a finite difference method. When using Algorithm 3.1, we always take $\xi_1 = c_0 = 0$ as an initial guess.

In Figure 4.1 we report the reconstruction results of Algorithm 3.1 using noisy data with noise level $\delta = 0.001$ when the sought solution is piecewise constant, see Figure 4.1 (a); we use $L(c) = F'(c)$ and take $\tau = 1.01$, $\gamma = 0.98$ and $\mu_1 = 100$. Figure 4.1 (b) presents the numerical result of Algorithm 3.1 with $\Theta(c) = \frac{1}{2\mu}||c||_2^2$ and $\mu_0 = 0.8$. The computation takes 1658 seconds and is terminated after 48 outer iterations which account for 781 iterations in total. Due to the over-smoothing effect, the reconstruction result turns out to contain unsatisfactory artifacts. Figure 4.1 (c) reports the numerical result of Algorithm 3.1 using $\Theta(c) = \frac{1}{2\mu}||c||_2^2 + |c|\tau V$ with $\beta = 10$ and $\mu_0 = 0.8/\beta$. The minimization problems related to $\Theta$ is solved by the PDHG method which is terminated as long as the error between two successive iterates is smaller than $10^{-4}$ or the number of iterations exceeds 250. The results in (c) is obtained after 64 outer iterations which account for 856 iterations in total. The computation takes 2412 seconds. Clearly the result in (c) significantly improves the one in (b) by efficiently removing the undesired artifacts.

In Figure 4.2 we report the reconstruction result of Algorithm 3.1 using noisy data with noise level $\delta = 10^{-4}$ when the sought solution is sparse, see Figure 4.2 (a) for the true solution. We use $L(c) = F'(c)$ and take $\tau = 1.1$, $\gamma = 0.98$, and $\mu_1 = 4 \times 10^5$. Figure 4.1 (b) presents the numerical result of Algorithm 3.1 using $\Theta(c) = \frac{1}{2\mu}||c||_2^2$ and $\mu_0 = 1$. The result is obtained after 85 outer iterations which
account for 1059 iterations in total and the computation takes 2037 seconds. Figure 4.1 (c) reports the numerical result of Algorithm 3.1 using \( \Theta(c) = \frac{1}{2} \|c\|_2^2 \) with \( \beta = 25 \) and \( \mu_0 = 1/\beta \). The result is obtained after 79 outer iterations which account for 953 iterations in total and the computation takes 1946 seconds. A comparison on the result in (b) and (c) clearly shows that the sparsity of the sought solution is significantly reconstructed in (c).

4.2. De-autoconvolution. We next consider the autoconvolution equation

\[
\int_0^t x(t-s)x(s)ds = y(t)
\]  

(4.7)

defined on the interval \([0, 1]\). This problem has a couple of applications in spectroscopy ([1]) and stochastics ([20]). Some properties of the autoconvolution operator \([F(x)](t) := \int_0^t x(t-s)x(s)ds\) have been discussed in [8]. In particular, as an operator from \(L^2[0, 1]\) to \(L^2[0, 1]\), \(F\) is Fréchet differentiable; its Fréchet derivative and the adjoint are given respectively by

\[
[F'(x)v](t) = 2 \int_0^t x(t-s)v(s)ds, \quad v \in L^2[0, 1],
\]

\[
[F'(x)^*w](s) = 2 \int_s^1 w(t)x(t-s)dt, \quad w \in L^2[0, 1].
\]

We assume that (4.7) has a piecewise constant solution and use a noisy data \(y^\delta\) satisfying \(\|y^\delta - y\|_{L^2[0, 1]} = \delta\) with \(\delta = 0.01\) to reconstruct the solution. In numerical computation, we discretize the problem by dividing \([0, 1]\) into \(N = 400\) subintervals of equal length and approximate integrals involved by the trapezoidal rule. When applying Algorithm 3.1, we take \(L(x) = F'(x)\) and

\[
\Theta(x) = \frac{1}{2\beta} \|x\|_2^2 + |x|_{TV},
\]

where \(|x|_{TV} := \sum_{i=1}^N |x[i+1] - x[i]|\) denotes the discrete 1-dimensional total variation of \(x\). In Figure 4.3 we report the reconstruction result for various values of \(\beta\). For those parameters involved in the algorithm, we take \(\tau = 1.02, \gamma = 0.99, \mu_0 = 10/\beta\) and \(\mu_1 = 100\). We also take the constant function \(\xi_0(t) \equiv 1/\beta\) as an initial guess. The minimization problems related to \(\Theta\) are solved by the 1-dimensional analog of the PDHG method which is terminated as long as the number of iterations exceeds 250 or the error between two successive iterates is smaller than \(10^{-4}\). From Figure...
4.3 we can see that, for larger value \( \beta \), Algorithm 3.1 has better capability to capture the feature of the sought solutions and the reconstruction result has better accuracy. This is because the total variation term \( |x|_{TV} \) dominates the penalty functional \( \Theta \) for large \( \beta \). However, we have to pay the price of more computational time because the convexity of the minimization problem involved becomes weaker and hence more iteration steps are required to get an approximate minimizer within certain accuracy. It is interesting to note that Algorithm 3.1 works well for solving (4.7), although condition (c) in Assumption 3.1 can not be verified.

4.3. Parameter identification by boundary measurements. We consider the identification of the parameter \( c \) in the Neumann boundary value problem

\[
- \Delta u + cu = f \quad \text{in } \Omega, \\
\frac{\partial u}{\partial n} = 0 \quad \text{on } \partial \Omega
\]

from boundary measurements \( u|_{\partial \Omega} \) using multiple sources \( f \in H^{-1}(\Omega) \), where \( \Omega \subset \mathbb{R}^d \) is a bounded domain with Lipschitz boundary \( \partial \Omega \). This is a severely ill-posed inverse problems.

According to the theory of elliptic equations, for each \( \nu_0 > 0 \) there is \( \varepsilon_0 > 0 \) such that for any \( c \) in the set

\[
\mathcal{D} := \{ c \in L^2(\Omega) : \|c - \hat{c}\|_{L^2(\Omega)} \leq \varepsilon_0 \text{ for some } \hat{c} \text{ satisfying } \hat{c} \geq \nu_0 \text{ a.e.} \},
\]
the problem (4.8) with source \( f \in H^{-1}(\Omega) \) has a unique solution in \( H^1(\Omega) \); we denote this solution by \( u_f(c) \) to indicate its dependence on \( c \) and \( f \). Consequently, by the trace embedding theorem, we have \( u_f(c)|_{\partial \Omega} \in H^{1/2}(\partial \Omega) \subset L^2(\partial \Omega) \). Thus, we can define a map \( F_f : \mathcal{D} \subset L^2(\Omega) \to L^2(\partial \Omega) \) by \( F_f(c) := u_f(c) \). It is easy to check that \( F_f \) is Fréchet differentiable and for any \( h \in L^2(\Omega) \) we have \( F_f'(c)h = w|_{\partial \Omega} \), where \( w \) is the weak solution of the problem

\[
-\Delta w + cw = -hu_f(c) \quad \text{in} \quad \Omega \quad \text{and} \quad \frac{\partial w}{\partial n} = 0 \quad \text{on} \quad \partial \Omega,
\]
i.e.

\[
\int_{\Omega} (\nabla w \cdot \nabla \varphi + cw\varphi) \, dx = -\int_{\Omega} hu_f(c)\varphi \, dx, \quad \forall \varphi \in H^1(\Omega).
\]

We need to determine \( F_f'(c)^*g \) for any \( g \in L^2(\partial \Omega) \). To this end, let \( v \in H^1(\Omega) \) be the weak solution of the problem

\[
-\Delta v + cv = 0 \quad \text{in} \quad \Omega \quad \text{and} \quad \frac{\partial v}{\partial n} = g \quad \text{on} \quad \partial \Omega. \tag{4.9}
\]

Then for any \( h \in L^2(\Omega) \) we have

\[
\langle F_f'(c)^*g, h \rangle_{L^2(\Omega)} = \langle g, F_f'(c)h \rangle_{L^2(\partial \Omega)} = \int_{\partial \Omega} gw = \int_{\Omega} (\nabla w \cdot \nabla v + cwv) \, dx = -\int_{\Omega} hu_f(c)v \, dx = \langle -vu_f(c), h \rangle_{L^2(\Omega)}.
\]

Therefore, \( F_f'(c)^*g = -vu_f(c) \), where \( v \in H^1(\Omega) \) is determined by (4.9).

In our numerical simulations, we take \( \Omega = [0, 1] \times [0, 1] \) and use \( N = 25 \) source functions

\[
f_k(x, y) = 200e^{-20(x-i/6)^2-20(y-j/6)^2}, \quad i, j = 1, \ldots, 5,
\]
where \( k = 5(i-1) + j \). Let \( c^\dagger \) be the sought solution and let \( F_k := F_{f_k} \) and \( u_k := u_{f_k}(c^\dagger)|_{\partial \Omega} \). Then the identification of \( c^\dagger \) is equivalent to solving the system

\[
F_k(c) = u_k, \quad k = 1, \ldots, N.
\]

We may reformulate the system into a single equation \( F(c) = u \) by introducing

\[
F = (F_1, \ldots, F_N) : \mathcal{D} \subset L^2(\Omega) \to (L^2(\partial \Omega))^N \quad \text{and} \quad u = (u_1, \ldots, u_N) \in (L^2(\partial \Omega))^N.
\]

When implementing Algorithm 3.1 numerically, we divide \( \Omega \) into \( 100 \times 100 \) small squares of equal size and solve all the partial differential equations involved approximately by the finite difference method.

In Figure 4.4 we report the numerical results by Algorithm 3.1 with random noisy data \( u^\delta_1, \ldots, u^\delta_N \) satisfying \( \|u^\delta_k - u_k\|_{L^2(\partial \Omega)} \leq \delta \) for \( k = 1, \ldots, N \) with noise level \( \delta = 0.001 \) to reconstruct a piecewise constant solution \( c^\dagger \) given in (a). We use \( L(c) = F'(c) := (F_1'(c), \ldots, F_N'(c)) \), \( \xi_0 = 1/\beta \) and take \( \tau = 1.1, \gamma = 0.95 \) and \( \mu_1 = 100 \). Figure 4.4 (b) presents the result by Algorithm 3.1 with \( \Theta(c) = \frac{1}{2}\|c\|_2^2 + |c|_{TV} \) with

\[
\frac{1}{2}\|c\|_2^2 + |c|_{TV} \]
β = 10 and µ0 = 1/β. As before, the minimization problems related to this Θ is solved by the PDHG method which is terminated as long as the error between two successive iterates is smaller than 10^{-4} or the number of iterations exceeds 250. The result is rather satisfactory and it successfully detects the desired feature of the sought solution. Due to the severe ill-posedness of the underlying problem, sufficient long time is required to complete the computation; in fact, the computation for (b) takes 4 hours and 23 minutes, and the computation for (c) takes 5 hours and 47 minutes. It is worthy to point out that parallel computing can be used in implementing Algorithm 3.1 for this problem to significantly reduce the computational time.

REFERENCES


