Landweber iteration of Kaczmarz type with general non-smooth convex penalty functionals

Qinian Jin[1] **and Wei Wang**[2,3]

[1]Mathematical Sciences Institute, Australian National University, Canberra, ACT 0200, Australia
[2]College of Mathematics, Physics and Information Engineering, Jiaxing University, Zhejiang 314001, China
[3]School of Mathematical Sciences, Fudan University, Shanghai 200433, China

E-mail: Qinian.Jin@anu.edu.au and weiwangmath@gmail.com

Abstract. The determination of solutions of many inverse problems usually requires a set of measurements which leads to solving systems of ill-posed equations. In this paper we propose the Landweber iteration of Kaczmarz type with general uniformly convex penalty functional. The method is formulated by using tools from convex analysis. The penalty term is allowed to be non-smooth to include the L^1 and total variation (TV) like penalty functionals, which are significant in reconstructing special features of solutions such as sparsity and piecewise constancy in practical applications. Under reasonable conditions, we establish the convergence of the method. Finally we present numerical simulations on tomography problems and parameter identification in partial differential equations to indicate the performance.

1. Introduction

Landweber iteration is one of the most well-known regularization methods for solving inverse problems formulated in Hilbert spaces. A complete account on this method for linear inverse problems can be found in [5] including the convergence analysis and its various accelerated versions. A nonlinear version of Landweber iteration was proposed in [10] for solving nonlinear inverse problems, where an elegant convergence analysis was present. Although Landweber iteration converges slowly, it still receives a lot of attention because it is simple to implement and is robust with respect to noise.

The classical Landweber iteration in Hilbert spaces, however, has the tendency to over-smooth solutions which makes it difficult to capture the special features of the sought solutions such as sparsity and piecewise constancy. It is therefore necessary to reformulate this method either in Banach space setting or in a manner that modern non-smooth penalty functionals, such as the L^1 and total variation like functionals, can be incorporated.

Let $A : \mathcal{X} \to \mathcal{Y}$ be a linear compact operator between two Banach spaces $\mathcal{X}$ and $\mathcal{Y}$ with norms $\|\cdot\|$ whose dual spaces are denoted by $\mathcal{X}^*$ and $\mathcal{Y}^*$ respectively. Some recent advances on Landweber iteration for linear inverse problems

$$Ax = y \tag{1.1}$$

in Banach space setting have been reported using only the noisy data y^δ satisfying

$$\|y^\delta - y\| \le \delta$$

with a small known noise level $\delta > 0$. In particular, when $\mathcal{X}$ is uniformly smooth and uniformly convex, by virtue of the duality mappings, a version of Landweber iteration for solving (1.1) was proposed in [15]. Although the method excludes the use of the L^1 and total variation like penalty functionals, new ideas were introduced in [15] which promote the study of Landweber iteration in modern setup. Recently a version of Landweber iteration was proposed in [2] using non-smooth uniformly convex penalty functionals. Let $\Theta : \mathcal{X} \to (-\infty, \infty]$ be a proper, lower semi-continuous, uniformly convex functional, then the method in [2] reads as

$$\begin{cases} \xi_{n+1} = \xi_n - \mu_n A^* J_r(Ax_n - y^\delta), \\ x_{n+1} = \arg\min_{x\in\mathcal{X}} \left\{\Theta(x) - \langle \xi_{n+1}, x\rangle\right\}, \end{cases} \tag{1.2}$$

where $A^* : \mathcal{Y}^* \to \mathcal{X}^*$ denotes the adjoint of A, J_r with $1 < r < \infty$ is the duality mapping of $\mathcal{Y}$ with gauge function $t \to t^{r-1}$, $\{\mu_n\}$ are suitable chosen step-lengths, and $\langle \cdot, \cdot\rangle$ denotes the duality pairing between $\mathcal{X}^*$ and $\mathcal{X}$. The method (1.2) reduces to the one in [15] when taking $\Theta(x) := \|x\|^p/p$ with $1 < p < \infty$. However, (1.2) has more freedom on Θ so that it can be used to detect special features of solutions.

The convergence analysis of (1.2) is given in [2] when it is terminated by the discrepancy principle

$$\|Ax_{n_\delta} - y^\delta\| \le \tau\delta < \|Ax_n - y^\delta\|, \qquad 0 \le n < n_\delta \tag{1.3}$$

with $\tau > 1$. The argument in [2], however, requires that $\mathrm{int}(D(\Theta))$, the interior of $D(\Theta)$, must be non-empty and that (1.1) must have a solution in $\mathrm{int}(D(\Theta))$. These conditions are indeed quite restrictive; for instance, the domain of the total variation like functional

$$\Theta(x) := \frac{1}{2\beta}\int_\Omega |x(\omega)|^2 d\omega + \int_\Omega |Dx|$$

with $\beta > 0$ over the space $\mathcal{X} := L^2(\Omega)$ on a bounded domain $\Omega \subset \mathbb{R}^d$ does not have any interior point in $L^2(\Omega)$, where

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

denotes the total variation of x over Ω ([6]). Therefore, the theoretical result in [2] can not be applied to this important penalty functional.

It is natural to ask if the convergence of (1.2) can be proved without assuming $\mathrm{int}(D(\Theta)) \ne \emptyset$. An affirmative answer would theoretically justify the applicability of (1.2) to a wider class of penalty functionals Θ including the total variation like functionals. The control of $\{\xi_n\}$ presents one of the major challenges. The analysis in [2] is based on proving the boundedness of $\{\xi_n\}$ in $\mathcal{X}^*$ which consequently enforces to assume that $\mathrm{int}(D(\Theta)) \ne \emptyset$. We observe that the boundedness of $\{\xi_n\}$ is not essential in the convergence analysis, the most essential ingredient is to control $\langle \xi_n, x_n - \hat{x}\rangle$ for any solution $\hat{x}$ of (1.1). Due to the lack of monotonicity on the residual $\|Ax_n - y^\delta\|$, it turns out to be difficult to consider $\langle \xi_n, x_n - \hat{x}\rangle$ for all $n \ge 0$. Fortunately, with a careful chosen subsequence $\{n_k\}$ of integers, we can derive what we expect on $\langle \xi_{n_k}, x_{n_k} - \hat{x}\rangle$ which together with some monotonicity results enables us to prove a stronger result, i.e. x_{n_δ} converges to a solution of (1.1) in Bregman distance.

Instead of considering (1.2) for solving (1.1) directly, we consider a more general setup in which (1.2) is extended for solving linear as well as nonlinear inverse problems. Instead of studying a single equation, we consider the system

$$F_i(x) = y_i, \qquad i = 0, \cdots, N-1 \tag{1.4}$$

consisting of N equations, where, for each $i = 0, \cdots, N-1$, $F_i : D(F_i) \subset \mathcal{X} \to \mathcal{Y}_i$ is an operator between two Banach spaces $\mathcal{X}$ and $\mathcal{Y}_i$. Such systems arise naturally in many practical applications including various tomography techniques using multiple exterior measurements. By introducing

$$F := (F_0, \cdots, F_{N-1}) : \mathcal{D} = \bigcap_{i=0}^{N-1} D(F_i) \to \mathcal{Y}_0 \times \cdots \times \mathcal{Y}_{N-1}$$

and

$$y := (y_0, \cdots, y_{N-1}),$$

the system (1.4) could be reformulated as a single equation $F(x) = y$. One might consider extending (1.2) to solve $F(x) = y$ directly. This procedure, however, becomes inefficient if N is large because it destroys the special structure of (1.4) and results in an equation requiring huge memory to save the intermediate computational results. Therefore, it seems advantageous to use the Kaczmarz-type methods, which cyclically consider each equation in (1.4) separately and hence require only reasonable memory consumption.

Some Landweber-Kaczmarz methods were formulated in [11, 7] for solving the system (1.4) when $\mathcal{X}$ and $\mathcal{Y}_i$ are Hilbert spaces, and the numerical results indicate that artefacts can appear in the reconstructed solutions due to oversmoothness. Recently a Landweber-Kaczmarz method was proposed in [13] for solving (1.4) in Banach space setting in the spirit of [15] and hence the possible use of the L^1 and total variation like penalty functionals is excluded. Furthermore, the convergence analysis in [13] unfortunately contains an error (see the first line on page 12 in [13]). In this paper, we propose a Landweber iteration of Kaczmarz type in which (1.2) is adapted to solve each equation in (1.4) and thus general non-smooth uniformly convex penalty functionals Θ are incorporated into the method with the hope of removing artefacts and of capturing special features of solutions. We give the detailed convergence analysis of our method. It is worthy pointing out that our analysis does not require the interior of $D(\Theta)$ be nonempty and therefore the convergence result applies for the total variation like penalty functionals.

This paper is organized as follows. In section 2 we give some preliminary results from convex analysis. In section 3, we first formulate the Landweber iteration of Kaczmarz type with general uniformly convex penalty term for solving the system (1.4), and then present the detail convergence analysis. In section 4 we give the proof of an important proposition which plays an important role in section 3. Finally, in section 5 we present some numerical simulations on tomography problems in imaging and parameter identification in partial differential equations to test the performance of the method.

2. Preliminaries

Let $\mathcal{X}$ be a Banach space with norm $\|\cdot\|$. We use $\mathcal{X}^*$ to denote its dual space, and for any $x \in \mathcal{X}$ and $\xi \in \mathcal{X}^*$ we write $\langle \xi, x \rangle = \xi(x)$ for the duality pairing. If

$\mathcal{Y}$ is another Banach space and $A : \mathcal{X} \to \mathcal{Y}$ is a bounded linear operator, we use $A^* : \mathcal{Y}^* \to \mathcal{X}^*$ to denote its adjoint, i.e. $\langle A^*\zeta, x\rangle = \langle \zeta, Ax\rangle$ for any $x \in \mathcal{X}$ and $\zeta \in \mathcal{Y}^*$. Let $\mathcal{N}(A) = \{x \in \mathcal{X} : Ax = 0\}$ be the null space of A and let

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

be the annihilator of $\mathcal{N}(A)$. When $\mathcal{X}$ 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 $\mathcal{X}^*$.

Given a convex function $\Theta : \mathcal{X} \to (-\infty, \infty]$, we use

$$D(\Theta) := \{x \in \mathcal{X} : \Theta(x) < +\infty\}$$

to denote its effective domain. It is called proper if $D(\Theta) \neq \emptyset$. The subgradient of Θ at $x \in \mathcal{X}$ is defined as

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

The multi-valued mapping $\partial\Theta : \mathcal{X} \to 2^{\mathcal{X}^*}$ is called the subdifferential of Θ. We set

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

For $x \in D(\partial\Theta)$ and $\xi \in \partial\Theta(x)$ we define ([3])

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

which is called the Bregman distance induced by Θ at x in the direction ξ. Clearly $D_\xi\Theta(z, x) \geq 0$ and

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

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

Bregman distance can be used to obtain information under the Banach space norm when Θ has stronger convexity. A proper convex function $\Theta : \mathcal{X} \to (-\infty, \infty]$ 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

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

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 Θ is called p-convex. It can be shown that if Θ is uniformly convex then

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

for all $\bar{x} \in \mathcal{X}$, $x \in D(\partial\Theta)$ and $\xi \in \partial\Theta(x)$. In particular, if Θ is p-convex with $h(t) = c_0 t^p$, then

$$D_\xi\Theta(\bar{x}, x) \geq c_0\|\bar{x} - x\|^p \tag{2.2}$$

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

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

It is well known that Θ^* is also proper, lower semi-continuous, and convex. If, in addition, $\mathcal{X}$ is reflexive, then

$$\xi \in \partial\Theta(x) \Longleftrightarrow x \in \partial\Theta^*(\xi) \Longleftrightarrow \Theta(x) + \Theta^*(\xi) = \langle \xi, x\rangle. \tag{2.3}$$

When Θ is p-convex satisfying (2.2) with $p \geq 2$, it follows from [19, Corollary 3.5.11] that $D(\Theta^*) = \mathcal{X}^*$, Θ^* is Fréchet differentiable and its gradient $\nabla\Theta^* : \mathcal{X}^* \to \mathcal{X}$ satisfies

$$\|\nabla\Theta^*(\xi_1) - \nabla\Theta^*(\xi_2)\| \leq \left(\frac{\|\xi_1 - \xi_2\|}{2c_0}\right)^{\frac{1}{p-1}}, \quad \forall \xi_1, \xi_2 \in \mathcal{X}^*. \tag{2.4}$$

Moreover

$$\Theta^*(\xi_2) - \Theta^*(\xi_1) - \langle \xi_2 - \xi_1, \nabla\Theta^*(\xi_1)\rangle \leq \frac{1}{p^*(2c_0)^{p^*-1}}\|\xi_2 - \xi_1\|^{p^*} \tag{2.5}$$

for any $\xi_1, \xi_2 \in \mathcal{X}^*$, where p^* is the number conjugate to p, i.e. $1/p + 1/p^* = 1$. By the subdifferential calculus, there also holds

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

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

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

which gives the duality mapping $J_r^{\mathcal{X}} : \mathcal{X} \to 2^{\mathcal{X}^*}$ of $\mathcal{X}$ with gauge function $t \to t^{r-1}$. The duality mapping $J_r^{\mathcal{X}}$, for each $1 < r < \infty$, is single valued and uniformly continuous on bounded sets if $\mathcal{X}$ is uniformly smooth in the sense that its modulus of smoothness

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

satisfies $\lim_{s\searrow 0} \frac{\rho_{\mathcal{X}}(s)}{s} = 0$.

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

$$\Theta_0(x) := \int_\Omega |x(\omega)|^p d\omega$$

is p-convex on $L^p(\Omega)$. We can construct the new p-convex functions

$$\Theta(x) := \mu \int_\Omega |x(\omega)|^p d\omega + a \int_\Omega |x(\omega)| d\omega + b \int_\Omega |Dx|, \tag{2.7}$$

where $\mu > 0$, $a, b \geq 0$, and $\int_\Omega |Dx|$ denotes the total variation of x over Ω. For $a = 1$ and $b = 0$ the corresponding function is useful for sparsity reconstruction ([17]); while for $a = 0$ and $b = 1$ the corresponding function is useful for detecting the discontinuities, in particular, when the solutions are piecewise-constant ([14]).

3. Landweber iteration of Kaczmarz type

We consider the system (1.4), i.e.

$$F_i(x) = y_i, \qquad i = 0, \cdots, N-1 \tag{3.1}$$

consisting of N equations, where, for each $i = 0, \cdots, N-1$, $F_i : D(F_i) \subset \mathcal{X} \to \mathcal{Y}_i$ is an operator between two Banach spaces $\mathcal{X}$ and $\mathcal{Y}_i$. We will assume that

$$\mathcal{D} := \bigcap_{i=0}^{N-1} D(F_i) \neq \emptyset$$

and each F_i is Fréchet differentiable with the Fréchet derivative denoted by $F_i'(x)$ for $x \in \mathcal{D}$. We will also assume that (3.1) has a solution. In general, (3.1) may have many solutions. In order to find the desired one, some selection criteria should be enforced. We choose a proper, lower semi-continuous, p-convex function $\Theta : \mathcal{X} \to (-\infty, \infty]$. By picking $x_0 \in D(\partial\Theta)$ and $\xi_0 \in \partial\Theta(x_0)$ as the initial guess, which may incorporate some available information on the sought solution, we define $x^\dagger$ to be the solution of (3.1) with the property

$$D_{\xi_0}\Theta(x^\dagger, x_0) := \min_{x \in D(\Theta)\cap\mathcal{D}} \{D_{\xi_0}\Theta(x, x_0) : F_i(x) = y_i, i = 0, \cdots, N-1\}. \tag{3.2}$$

We will work under the following conditions on the operators F_i where $B_\rho(x_0) := \{x \in \mathcal{X} : \|x - x_0\| \le \rho\}$.

Assumption 3.1 *(a) There is $\rho > 0$ such that $B_{2\rho}(x_0) \subset \mathcal{D}$ and (3.1) has a solution in $B_\rho(x_0) \cap D(\Theta)$;*

(b) Each operator F_i is weakly closed on $\mathcal{D}$ and is Fréchet differentiable on $B_{2\rho}(x_0)$, and $x \to F_i'(x)$ is continuous on $B_{2\rho}(x_0)$.

(c) Each F_i is properly scaled so that $\|F_i'(x)\| \le 1$ for $x \in B_{2\rho}(x_0)$.

(d) There exists $0 \le \eta < 1$ such that

$$\|F_i(x) - F_i(\bar{x}) - F_i'(\bar{x})(x - \bar{x})\| \le \eta\|F_i(x) - F_i(\bar{x})\|$$

for all $x, \bar{x} \in B_{2\rho}(x_0)$ and $i = 0, \cdots, N-1$.

All the conditions in Assumption 3.1 are standard. Condition (d) is called the tangential cone condition and is widely used in the analysis of regularization methods for solving nonlinear ill-posed inverse problems ([10]) The weakly closedness of F_i over $\mathcal{D}$ in (b) means that if $\{x_n\} \subset \mathcal{D}$ converges weakly to some $x \in \mathcal{X}$ and $\{F_i(x_n)\}$ converges weakly to some $y_i \in \mathcal{Y}_i$, then $x \in \mathcal{D}$ and $F_i(x) = y_i$.

When $\mathcal{X}$ is a reflexive Banach space, by using the p-convexity and the weakly lower semi-continuity of Θ together with the weakly closedness of F_i for $i = 0, \cdots, N-1$ it is standard to show that $x^\dagger$ exists. The following result shows that $x^\dagger$ is in fact uniquely defined.

Lemma 3.2 *Let $\mathcal{X}$ be reflexive and F_i satisfy Assumption 3.1. If $x^\dagger \in B_\rho(x_0) \cap D(\Theta)$, then $x^\dagger$ is the unique solution of (3.1) in $B_{2\rho}(x_0) \cap D(\Theta)$ satisfying (3.2).*

Proof. Assume that (3.1) has another solution $\hat{x}$ in $B_{2\rho}(x_0) \cap D(\Theta)$ satisfying (3.2) with $\hat{x} \ne x^\dagger$. Since $F_i(\hat{x}) = y_i = F_i(x^\dagger)$ for $i = 0, \cdots, N-1$, we can use Assumption 3.1 (d) to derive that

$$F_i'(x^\dagger)(\hat{x} - x^\dagger) = 0, \qquad i = 0, \cdots, N-1.$$

Let $x_\lambda = \lambda\hat{x} + (1-\lambda)x^\dagger$ for $0 < \lambda < 1$. Then $x_\lambda \in B_{2\rho}(x_0) \cap D(\Theta)$ and

$$F_i'(x^\dagger)(x_\lambda - x^\dagger) = 0, \qquad i = 0, \cdots, N-1.$$

Thus we can use Assumption 3.1 (d) to conclude that

$$\|F_i(x_\lambda) - F_i(x^\dagger)\| \le \eta\|F_i(x_\lambda) - F_i(x^\dagger)\|.$$

Since $0 \le \eta < 1$, this implies that $F_i(x_\lambda) = F_i(x^\dagger) = y_i$ for $i = 0, \cdots, N-1$. Consequently, by the minimal property of $x^\dagger$ we have

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

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

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

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

In practical application, instead of y_i we only have noisy data y_i^δ satisfying

$$\|y_i^\delta - y_i\| \le \delta, \qquad i = 0, \cdots, N-1$$

with a small known noise level $\delta > 0$. We will use y_i^δ, $i = 0, \cdots, N-1$, to construct an approximate solution to (3.1). We assume that each $\mathcal{Y}_i$ is uniformly smooth so that, for each $1 < r < \infty$, the duality mapping $J_r^{\mathcal{Y}_i} : \mathcal{Y}_i \to \mathcal{Y}_i^*$ is single valued and continuous. By introducing a proper, lower-semi continuous, p-convex function $\Theta : (-\infty, \infty] \to \mathcal{X}$ satisfying (2.2) for some constant $c_0 > 0$, we propose the following Landweber iteration of Kaczmarz type:

Algorithm 3.3 *(i) Pick $\xi_0 \in \mathcal{X}^*$ and set $x_0 := \arg\min_{x\in\mathcal{X}} \{\Theta(x) - \langle \xi_0, x\rangle\}$;*

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

$$\begin{aligned} \xi_{n,i+1}^\delta &= \xi_{n,i}^\delta - \mu_{n,i}^\delta F_i'(x_{n,i}^\delta)^* J_r^{\mathcal{Y}_i}(F_i(x_{n,i}^\delta) - y_i^\delta), \\ x_{n,i+1}^\delta &= \arg\min_{x\in\mathcal{X}} \left\{\Theta(x) - \langle \xi_{n,i+1}^\delta, x\rangle\right\} \end{aligned}$$

for $i = 0, \cdots, N-1$, where

$$\mu_{n,i}^\delta = \begin{cases} \mu_0 \|F_i(x_{n,i}^\delta) - y_i^\delta\|^{p-r}, & \text{if } \|F_i(x_{n,i}^\delta) - y_i^\delta\| > \tau\delta, \\ 0, & \text{otherwise} \end{cases}$$

for some $\mu_0 > 0$. We then define $\xi_{n+1}^\delta := \xi_{n,N}^\delta$ and $x_{n+1}^\delta := x_{n,N}^\delta$.

(iii) Let n_δ be the first integer such that

$$\mu_{n_\delta,i}^\delta = 0 \quad \text{for all } i = 0, \cdots, N-1$$

and use $x_{n_\delta}^\delta$ to approximate the solution of (3.1).

In Algorithm 3.3, $x_{n,i+1}^\delta$ is defined as the minimizer of a p-convex functional over $\mathcal{X}$ which is independent of F_i and therefore it could be found by efficient solvers. By using (2.6), one can see that

$$x_{n,i+1}^\delta = \nabla\Theta^*(\xi_{n,i+1}^\delta)$$

which is useful for the forthcoming theoretical analysis.

In this section we will show that Algorithm 3.3 is well-defined by showing that n_δ is finite and establish a convergence result on $x_{n_\delta}^\delta$ as $\delta \to 0$.

Lemma 3.4 *Let Assumption 3.1 hold and let $\Theta : (-\infty, \infty] \to \mathcal{X}$ be a proper, lower-semi continuous, p-convex function with $p \ge 2$ satisfying (2.2) for some $c_0 > 0$. Assume that*

$$D_{\xi_0}\Theta(x^\dagger, x_0) \le c_0 \rho^p. \tag{3.4}$$

Let $\{\xi_n^\delta\}$ and $\{x_n^\delta\}$ be defined by Algorithm 3.3 with $\tau > 1$ and $\mu_0 > 0$ such that

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

Then $n_\delta < \infty$ and $x_{n,i}^\delta \in B_{2\rho}(x_0)$ for all $n \ge 0$ and $i = 0, \cdots, N-1$. Moreover, for any solution $\hat{x}$ of (3.1) in $B_{2\rho}(x_0) \cap D(\Theta)$ and all n there hold

$$D_{\xi_{n+1}^\delta}\Theta(\hat{x}, x_{n+1}^\delta) \le D_{\xi_n^\delta}\Theta(\hat{x}, x_n^\delta), \tag{3.5}$$

$$c_1 \sum_{i=0}^{N-1} \mu_{n,i}^\delta \|F_i(x_{n,i}^\delta) - y_i^\delta\|^r \le D_{\xi_n^\delta}\Theta(\hat{x}, x_n^\delta) - D_{\xi_{n+1}^\delta}\Theta(\hat{x}, x_{n+1}^\delta). \tag{3.6}$$

Proof. In order to obtain (3.5) and (3.6), it suffices to show that $x_{n,i}^\delta \in B_{2\rho}(x_0)$ and

$$D_{\xi_{n,i+1}^\delta}\Theta(\hat{x}, x_{n,i+1}^\delta) - D_{\xi_{n,i}^\delta}\Theta(\hat{x}, x_{n,i}^\delta) \le -c_1\mu_{n,i}^\delta \|F_i(x_{n,i}^\delta) - y_i^\delta\|^r \tag{3.7}$$

for all $n \ge 0$ and $i = 0, \cdots, N-1$. From the definition of Bregman distance and (2.3) it follows that

$$\begin{aligned} &D_{\xi_{n,i+1}^\delta}\Theta(\hat{x}, x_{n,i+1}^\delta) - D_{\xi_{n,i}^\delta}\Theta(\hat{x}, x_{n,i}^\delta) \\ &\qquad = \Theta(x_{n,i}^\delta) - \Theta(x_{n,i+1}^\delta) - \langle \xi_{n,i+1}^\delta, \hat{x} - x_{n,i+1}^\delta\rangle + \langle \xi_{n,i}^\delta, \hat{x} - x_{n,i}^\delta\rangle \\ &\qquad = \Theta^*(\xi_{n,i+1}^\delta) - \Theta^*(\xi_{n,i}^\delta) - \langle \xi_{n,i+1}^\delta - \xi_{n,i}^\delta, \hat{x}\rangle. \end{aligned}$$

Using $x_{n,i}^\delta = \nabla\Theta^*(\xi_{n,i}^\delta)$, we can write

$$\begin{aligned} D_{\xi_{n,i+1}^\delta}\Theta(\hat{x}, x_{n,i+1}^\delta) - D_{\xi_{n,i}^\delta}\Theta(\hat{x}, x_{n,i}^\delta) = &\Theta^*(\xi_{n,i+1}^\delta) - \Theta^*(\xi_{n,i}^\delta) \\ &- \langle \xi_{n,i+1}^\delta - \xi_{n,i}^\delta, \nabla\Theta^*(\xi_{n,i}^\delta)\rangle \\ &+ \langle \xi_{n,i+1}^\delta - \xi_{n,i}^\delta, x_{n,i}^\delta - \hat{x}\rangle. \end{aligned}$$

Since Θ is p-convex, we may use (2.5) to obtain

$$\begin{aligned} D_{\xi_{n,i+1}^\delta}\Theta(\hat{x}, x_{n,i+1}^\delta) - D_{\xi_{n,i}^\delta}\Theta(\hat{x}, x_{n,i}^\delta) \le &\frac{1}{p^*(2c_0)^{p^*-1}}\|\xi_{n,i+1}^\delta - \xi_{n,i}^\delta\|^{p^*} \\ &-\mu_{n,i}^\delta\langle J_r^{\mathcal{Y}_i}(F_i(x_{n,i}^\delta) - y_i^\delta), F_i'(x_{n,i}^\delta)(x_{n,i}^\delta - \hat{x})\rangle, \end{aligned}$$

where $1/p + 1/p^* = 1$. By using the properties of the duality mapping $J_r^{\mathcal{Y}_i}$ and Assumption 3.1 it follows that

$$\begin{aligned} &D_{\xi_{n,i+1}^\delta}\Theta(\hat{x}, x_{n,i+1}^\delta) - D_{\xi_{n,i}^\delta}\Theta(\hat{x}, x_{n,i}^\delta) \\ &\qquad \le \mu_{n,i}^\delta\|F_i(x_{n,i}^\delta) - y_i^\delta\|^{r-1}\|y_i^\delta - F_i(x_{n,i}^\delta) - F_i'(x_{n,i}^\delta)(\hat{x} - x_{n,i}^\delta)\| \\ &\qquad\quad -\mu_{n,i}^\delta\|F_i(x_{n,i}^\delta) - y_i^\delta\|^r + \frac{1}{p^*(2c_0)^{p^*-1}}\|\xi_{n,i+1}^\delta - \xi_{n,i}^\delta\|^{p^*} \\ &\qquad \le (1+\eta)\mu_{n,i}^\delta\|F_i(x_{n,i}^\delta) - y_i^\delta\|^{r-1}\delta - (1-\eta)\mu_{n,i}^\delta\|F_i(x_{n,i}^\delta) - y_i^\delta\|^r \\ &\qquad\quad +\frac{1}{p^*(2c_0)^{p^*-1}}\left(\mu_{n,i}^\delta\right)^{p^*}\|F_i'(x_{n,i}^\delta)^* J_r^{\mathcal{Y}_i}(F_i(x_{n,i}^\delta) - y_i^\delta)\|^{p^*}. \end{aligned} \tag{3.8}$$

According to the definition of $\mu_{n,i}^\delta$, the scaling condition in Assumption 3.1 (c), and the property of $J_r^{\mathcal{Y}_i}$, it is easy to see that

$$\mu_{n,i}^\delta\|F_i(x_{n,i}^\delta) - y_i^\delta\|^{r-1}\delta \le \frac{1}{\tau}\mu_{n,i}^\delta\|F_i(x_{n,i}^\delta) - y_i^\delta\|^r,$$

$$\left(\mu_{n,i}^\delta\right)^{p^*-1}\|F_i'(x_{n,i}^\delta)^* J_r^{\mathcal{Y}_i}(F_i(x_{n,i}^\delta) - y_i^\delta)\|^{p^*} \le \mu_0^{p^*-1}\|F_i(x_{n,i}^\delta) - y_i^\delta\|^r.$$

Combining these two inequalities with (3.8) we can obtain (3.7). To show $x_{n,i+1}^\delta \in B_{2\rho}(x_0)$ we first use (3.7) with $\hat{x} = x^\dagger$ and (3.4) to obtain

$$D_{\xi_{n,i+1}^\delta}\Theta(x^\dagger, x_{n,i+1}^\delta) \le D_{\xi_0}\Theta(x^\dagger, x_0) \le c_0\rho^p.$$

In view of (2.2), we then have $\|x_{n,i+1}^\delta - x^\dagger\| \le \rho$ and $\|x^\dagger - x_0\| \le \rho$. Consequently $x_{n,i+1}^\delta \in B_{2\rho}(x_0)$.

We next show $n_\delta < \infty$. According to the definition of n_δ, for any $n < n_\delta$ there is at least one $i_n \in \{0, \cdots, N-1\}$ such that $\|F_{i_n}(x_{n,i_n}^\delta) - y_{i_n}^\delta\| > \tau\delta$. Consequently

$$\mu_{n,i_n}^\delta = \mu_0 \|F_{i_n}(x_{n,i_n}^\delta) - y_{i_n}^\delta\|^{p-r}$$

and

$$\sum_{i=0}^{N-1} \mu_{n,i}^\delta \|F_i(x_{n,i}^\delta) - y_i^\delta\|^r \ge \mu_0 \|F_{i_n}(x_{n,i_n}^\delta) - y_{i_n}^\delta\|^p > \mu_0 \tau^p \delta^p.$$

By summing (3.6) over n from $n = 0$ to $n = m$ for any $m < n_\delta$ and using the above inequality we obtain $c_1 \mu_0 \tau^p \delta^p (m+1) \le D_{\xi_0}\Theta(\hat{x}, x_0)$. Since this is true for any $m < n_\delta$, we must have $n_\delta < \infty$. □

When using the exact data y_i instead of the noisy data y_i^δ in Algorithm 3.3, we will drop the superscript δ in all the quantities involved, for instance, we will write ξ_n^δ as ξ_n, x_n^δ as x_n, and so on. Observing that

$$\mu_{n,i} \|F_i(x_{n,i}) - y_i\|^r = \mu_0 \|F_i(x_{n,i}) - y_i\|^p.$$

The proof of Lemma 3.4 in fact shows that, under Assumption 3.1, if

$$c_2 := 1 - \eta - \frac{p-1}{p} \left(\frac{\mu_0}{2c_0}\right)^{\frac{1}{p-1}} > 0,$$

then

$$x_{n,i} \in B_{2\rho}(x_0) \qquad \forall n \ge 0 \text{ and } i = 0, \cdots, N-1$$

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

$$D_{\xi_{n+1}}\Theta(\hat{x}, x_{n+1}) \le D_{\xi_n}\Theta(\hat{x}, x_n), \tag{3.9}$$

$$c_2 \mu_0 \sum_{i=0}^{N-1} \|F_i(x_{n,i}) - y_i\|^p \le D_{\xi_n}\Theta(\hat{x}, x_n) - D_{\xi_{n+1}}\Theta(\hat{x}, x_{n+1}). \tag{3.10}$$

These two inequalities imply immediately that

$$\lim_{n\to\infty} \sum_{i=0}^{N-1} \|F_i(x_{n,i}) - y_i\|^p = 0. \tag{3.11}$$

The next result gives an estimate on $\|F_i(x_n) - y_i\|$ and shows that $\|F_i(x_n) - y_i\| \to 0$ as $n \to \infty$ for all $i = 0, \cdots, N-1$.

Lemma 3.5 *Let all the conditions in Lemma 3.4 hold. Then there is a constant C_0 such that for all $n \ge 1$ there hold*

$$\|F_i(x_n) - y_i\| \le C_0 \sum_{j=0}^{i} \|F_j(x_{n,j}) - y_j\|, \quad i = 0, \cdots, N-1.$$

Consequently $\lim_{n\to\infty} \|F_i(x_n) - y_i\| = 0$ *for all* $i = 0, \cdots, N-1$.

Proof. Recall that $x_n = x_{n,0}$, we have

$$\|F_i(x_n) - y_i\| \le \|F_i(x_{n,i}) - y_i\| + \|F_i(x_{n,i}) - F_i(x_{n,0})\|.$$

By using the condition on F we have

$$\begin{aligned}\|F_i(x_n) - y_i\| &\le \|F_i(x_{n,i}) - y_i\| + \sum_{j=0}^{i-1} \|F_i(x_{n,j+1}) - F_i(x_{n,j})\| \\ &\le \|F_i(x_{n,i}) - y_i\| + \frac{1}{1-\eta} \sum_{j=0}^{i-1} \|F_i'(x_{n,j})(x_{n,j+1} - x_{n,j})\| \\ &\le \|F_i(x_{n,i}) - y_i\| + \frac{1}{1-\eta} \sum_{j=0}^{i-1} \|x_{n,j+1} - x_{n,j}\|. \end{aligned} \tag{3.12}$$

Since $x_{n,j} = \nabla\Theta^*(\xi_{n,j})$, we can use the property (2.4) to derive that

$$\begin{aligned}\|x_{n,j+1} - x_{n,j}\| &= \|\nabla\Theta^*(\xi_{n,j+1}) - \nabla\Theta^*(\xi_{n,j})\| \\ &\le \left(\frac{\|\xi_{n,j+1} - \xi_{n,j}\|}{2c_0}\right)^{\frac{1}{p-1}}.\end{aligned}$$

Using the definition of $\xi_{n,j+1}$ and the property of the duality mapping $J_r^{\mathcal{Y}_i}$ we obtain

$$\begin{aligned}\|x_{n,j+1} - x_{n,j}\| &\le (2c_0)^{-\frac{1}{p-1}} \mu_{n,j}^{\frac{1}{p-1}} \|F_j(x_{n,j}) - y_j\|^{\frac{r-1}{p-1}} \\ &= \left(\frac{\mu_0}{2c_0}\right)^{\frac{1}{p-1}} \|F_j(x_{n,j}) - y_j\|.\end{aligned}$$

Combining this with (3.12) gives the desired inequality. □

As the first step toward the proof of convergence on $x_{n_\delta}^\delta$, we need to derive some convergence results on the sequences $\{x_n\}$ and $\{\xi_n\}$. This will be achieved by the following proposition which gives a general convergence criterion on any sequences $\{x_n\} \subset \mathcal{X}$ and $\{\xi_n\} \subset \mathcal{X}^*$ satisfying certain conditions.

Proposition 3.6 *Consider the system (3.1) for which Assumption 3.1 holds. Let $\Theta : \mathcal{X} \to (-\infty, \infty]$ be a proper, lower semi-continuous and uniformly convex function. Let $\{x_n\} \subset B_{2\rho}(x_0)$ and $\{\xi_n\} \subset \mathcal{X}^*$ be such that*

(i) $\xi_n \in \partial\Theta(x_n)$ for all n;

(ii) for any solution $\hat{x}$ of (3.1) in $B_{2\rho}(x_0) \cap D(\Theta)$ the sequence $\{D_{\xi_n}\Theta(\hat{x}, x_n)\}$ is monotonically decreasing;

(iii) $\lim_{n\to\infty} \|F_i(x_n) - y_i\| = 0$ for all $i = 0, \cdots, N-1$.

(iv) there is a subsequence $\{n_k\}$ with $n_k \to \infty$ such that for all $l < k$ and any solution $\hat{x}$ of (3.1) in $B_{2\rho}(x_0) \cap D(\Theta)$ there holds

$$|\langle \xi_{n_k} - \xi_{n_l}, x_{n_k} - \hat{x}\rangle| \le C_1 \left(D_{\xi_{n_l}}\Theta(\hat{x}, x_{n_l}) - D_{\xi_{n_k}}\Theta(\hat{x}, x_{n_k})\right) \tag{3.13}$$

for some constant C_1.

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

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

If, in addition, $x^\dagger \in B_\rho(x_0) \cap D(\Theta)$ and $\xi_{n+1} - \xi_n \in \overline{\mathcal{R}(F_0'(x^\dagger)^*)} + \cdots + \overline{\mathcal{R}(F_{N-1}'(x^\dagger)^*)}$ for all n, then $x_* = x^\dagger$.

Proposition 3.6 may be of independent interest and its proof is deferred to section 4. Now we are ready to give the convergence result on $\{\xi_n\}$ and $\{x_n\}$ defined by the Landweber iteration of Kaczmarz type with exact data.

Lemma 3.7 *Let all the conditions in Lemma 3.4 hold. For the sequences $\{\xi_n\}$ and $\{x_n\}$ defined by Algorithm 3.3 with exact data, there exists a solution $x_* \in B_{2\rho}(x_0) \cap D(\Theta)$ of (3.1) such that*

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

If in addition $\mathcal{N}(F_i'(x^\dagger)) \subset \mathcal{N}(F_i'(x))$ for all $x \in B_{2\rho}(x_0)$ and $i = 0, \cdots, N-1$, then $x_* = x^\dagger$.

Proof. We will use Proposition 3.6 to complete the proof. By the definition of $\{\xi_n\}$ and $\{x_n\}$ we have $\xi_n \in \partial\Theta(x_n)$. The monotonicity of $\{D_{\xi_n}\Theta(\hat{x}, x_n)\}$ is given by (3.9). Lemma 3.5 shows that

$$\lim_{n\to\infty} \|F_i(x_n) - y_i\| = 0 \qquad \text{for all } i = 0, \cdots, N-1.$$

Therefore, in order to derive the convergence result, it suffices to show that there exists a strictly increasing subsequence $\{n_k\}$ such that for any solution $\hat{x}$ of (3.1) and any $l < k$ there holds

$$|\langle \xi_{n_k} - \xi_{n_l}, x_{n_k} - \hat{x}\rangle| \le C\left(D_{\xi_{n_l}}\Theta(\hat{x}, x_{n_l}) - D_{\xi_{n_k}}\Theta(\hat{x}, x_{n_k})\right) \tag{3.14}$$

To this end, let

$$R_n := \sum_{i=0}^{N-1} \|y_i - F_i(x_{n,i})\|^p.$$

It follows from (3.11) that

$$\lim_{n\to\infty} R_n = 0. \tag{3.15}$$

Moreover, if $R_n = 0$ for some n, then $y_i = F_i(x_{n,i})$ for $i = 0, \cdots, N-1$. Consequently it follows from the definition of the method that $x_{m,i} = x_n$ for all $m \ge n$ and $i = 0, \cdots, N-1$. Therefore

$$R_n = 0 \text{ for some } n \Longrightarrow R_m = 0 \text{ for all } m \ge n. \tag{3.16}$$

In view of (3.15) and (3.16), we can introduce a subsequence $\{n_k\}$ by setting $n_0 = 0$ and letting n_k, for each $k \ge 1$, be the first integer satisfying

$$n_k \ge n_{k-1} + 1 \qquad \text{and} \qquad R_{n_k} \le R_{n_{k-1}}.$$

For such chosen strictly increasing sequence $\{n_k\}$ it is easy to see that

$$R_{n_k} \le R_n, \qquad 0 \le n < n_k. \tag{3.17}$$

We now prove (3.14) for the above chosen subsequence $\{n_k\}$. We first use the definition of the method to obtain for $n < n_k$ that

$$\begin{aligned}\langle \xi_{n+1} - \xi_n, x_{n_k} - \hat{x}\rangle &= \sum_{i=0}^{N-1} \langle \xi_{n,i+1} - \xi_{n,i}, x_{n_k} - \hat{x}\rangle \\ &= -\sum_{i=0}^{N-1} \mu_{n,i} \langle J_r^{\mathcal{Y}_i}(F_i(x_{n,i}) - y_i), F_i'(x_{n,i})(x_{n_k} - \hat{x})\rangle.\end{aligned}$$

Using the condition on F_i it is easy to obtain

$$\begin{aligned}\|F_i'(x_{n,i})(x_{n_k}-\hat{x})\| &\le \|F_i'(x_{n,i})(x_{n,i}-\hat{x})\| + \|F_i'(x_{n,i})(x_{n_k}-x_{n,i})\| \\ &\le (1+\eta)\left(\|F_i(x_{n,i})-y_i\| + \|F_i(x_{n_k})-F_i(x_{n,i})\|\right) \\ &\le (1+\eta)\left(2\|F_i(x_{n,i})-y_i\| + \|F_i(x_{n_k})-y_i\|\right).\end{aligned}$$

Therefore, by using the property of the duality mapping $J_r^{\mathcal{Y}_i}$ and the Hölder inequality, we have

$$\begin{aligned}|\langle \xi_{n+1}-\xi_n, x_{n_k}-\hat{x}\rangle| &\le \sum_{i=0}^{N-1} \mu_{n,i}\|F_i(x_{n,i})-y_i\|^{r-1}\|F_i'(x_{n,i})(x_{n_k}-\hat{x})\| \\ &\le (1+\eta)\sum_{i=0}^{N-1} \mu_{n,i}\|F_i(x_{n,i})-y_i\|^{r-1}\|F_i(x_{n_k})-y_i\| + 2(1+\eta)\mu_0 R_n \\ &= (1+\eta)\mu_0 \sum_{i=0}^{N-1} \|F_i(x_{n,i})-y_i\|^{p-1}\|F_i(x_{n_k})-y_i\| + 2(1+\eta)\mu_0 R_n \\ &\le (1+\eta)\mu_0 R_n^{\frac{p-1}{p}} \left(\sum_{i=0}^{N-1}\|F_i(x_{n_k})-y_i\|^p\right)^{\frac{1}{p}} + 2(1+\eta)\mu_0 R_n. \qquad (3.18)\end{aligned}$$

From Lemma 3.5 and the Hölder inequality it follows that

$$\|F_i(x_{n_k})-y_i\| \le C_0\sum_{j=0}^{i}\|F_j(x_{n_k,j})-y_j\| \le C_0(i+1)^{\frac{p-1}{p}}\left(\sum_{j=0}^{i}\|F_j(x_{n_k,j})-y_j\|^p\right)^{\frac{1}{p}}.$$

This implies that

$$\begin{aligned}\sum_{i=0}^{N-1}\|F_i(x_{n_k})-y_i\|^p &\le C_0^p\sum_{i=0}^{N-1}\sum_{j=0}^{i}(i+1)^{p-1}\|F_j(x_{n_k,j})-y_j\|^p \\ &\le C_0^p(N+1)^p\sum_{j=0}^{N-1}\|F_j(x_{n_k,j})-y_j\|^p \\ &= C_0^p(N+1)^p R_{n_k}.\end{aligned}$$

Combining this with (3.18) and using (3.17) we can obtain for $n<n_k$ that

$$\begin{aligned}|\langle \xi_{n+1}-\xi_n, x_{n_k}-\hat{x}\rangle| &\le (1+\eta)\mu_0\left[C_0(N+1)R_n^{\frac{p-1}{p}}R_{n_k}^{\frac{1}{p}} + 2R_n\right] \\ &\le (1+\eta)\mu_0\left[C_0(N+1)+2\right]R_n.\end{aligned}$$

Finally, we can derive that

$$|\langle \xi_{n_k}-\xi_{n_l}, x_{n_k}-\hat{x}\rangle| \le \sum_{n=n_l}^{n_k-1}|\langle \xi_{n+1}-\xi_n, x_{n_k}-\hat{x}\rangle| \le C_2\sum_{n=n_l}^{n_k-1}R_n$$

with $C_2 := (1+\eta)\mu_0\left[C_0(N+1)+2\right]$. In view of (3.10) we therefore obtain (3.14).

In order to show $x_* = x^\dagger$ under the condition $\mathcal{N}(F_i'(x^\dagger)) \subset \mathcal{N}(F_i'(x))$ for all $x \in B_{2\rho}(x_0)$ and $i = 0,\cdots,N-1$, we observe from the definition of ξ_n that

$$\xi_{n+1}-\xi_n = \sum_{i=0}^{N-1}(\xi_{n,i+1}-\xi_{n,i}) \in \mathcal{R}(F_0'(x_{n,0})^*) + \cdots + \mathcal{R}(F_{N-1}'(x_{n,N-1})^*).$$

Since $\mathcal{X}$ is reflexive and $\mathcal{N}(F_i'(x^\dagger)) \subset \mathcal{N}(F_i'(x_{n,i}))$, we have $\overline{\mathcal{R}(F_i'(x_{n,i})^*)} \subset \overline{\mathcal{R}(F_i'(x^\dagger)^*)}$. Therefore

$$\xi_{n+1} - \xi_n \in \overline{\mathcal{R}(F_0'(x^\dagger)^*)} + \cdots + \overline{\mathcal{R}(F_{N-1}'(x^\dagger)^*)} \quad \text{for all } n.$$

Hence, we can use the second part of Proposition 3.6 to conclude that $x_* = x^\dagger$. □

In order to use the above result to prove the convergence of the Landweber iteration of Kaczmarz type described in Algorithm 3.3 with noisy data, we need the following stability result.

Lemma 3.8 *Let $\mathcal{X}$ be reflexive and let $\mathcal{Y}_i$ be uniformly smooth. Let all the conditions in Lemma 3.4 hold. Then for all $n \geq 0$ and $i = 0, \cdots, N-1$ there hold*

$$\xi_{n,i}^\delta \to \xi_{n,i} \quad \text{and} \quad x_{n,i}^\delta \to x_{n,i} \quad \text{as } \delta \to 0.$$

Proof. The result is trivial for $n = 0$ and $i = 0$. We next assume that the result is true for some $n \geq 0$ and some $i \in \{0, \cdots, N-1\}$ and show that $\xi_{n,i+1}^\delta \to \xi_{n,i+1}$ and $x_{n,i+1}^\delta \to x_{n,i+1}$ as $\delta \to 0$. We consider two cases.

Case 1: $F_i(x_{n,i}) = y_i$. In this case we have $\mu_{n,i} = 0$ and $\|F_i(x_{n,i}^\delta) - y_i^\delta\| \to 0$ as $\delta \to 0$ by the continuity of F_i. Thus

$$\xi_{n,i+1}^\delta - \xi_{n,i+1} = \xi_{n,i}^\delta - \xi_{n,i} - \mu_{n,i}^\delta F_i'(x_{n,i}^\delta)^* J_r^{\mathcal{Y}_i}(F_i(x_{n,i}^\delta) - y_i^\delta)$$

which implies that

$$\|\xi_{n,i+1}^\delta - \xi_{n,i+1}\| \leq \|\xi_{n,i}^\delta - \xi_{n,i}\| + \mu_0 \|F_i(x_{n,i}^\delta) - y_i^\delta\|^{p-1}.$$

By the induction hypotheses, we then have $\xi_{n,i+1}^\delta \to \xi_{n,i+1}$ as $\delta \to 0$. Consequently, by using the continuity of $\nabla\Theta^*$ we have $x_{n,i+1}^\delta = \nabla\Theta^*(\xi_{n,i+1}^\delta) \to \nabla\Theta^*(\xi_{n,i+1}) = x_{n,i+1}$ as $\delta \to 0$.

Case 2: $F_i(x_{n,i}) \neq y_i$. In this case we have $\|F_i(x_{n,i}^\delta) - y_i^\delta\| > \tau\delta$ for small $\delta > 0$. Therefore

$$\mu_{n,i}^\delta = \mu_0 \|F_i(x_{n,i}^\delta) - y_i^\delta\|^{p-r} \to \mu_{n,i} = \mu_0 \|F_i(x_{n,i}) - y_i\|^{p-r}$$

as $\delta \to 0$. By Assumption 3.1 (b) and the uniform smoothness of $\mathcal{Y}_i$, we know that F, F' and $J_r^{\mathcal{Y}_i}$ are continuous. It then follows from the induction hypotheses that $\xi_{n,i+1}^\delta \to \xi_{n,i+1}$ and hence $x_{n,i+1}^\delta \to x_{n,i+1}$ as $\delta \to 0$ using again the continuity of $\nabla\Theta^*$. □

We are now in a position to give the main convergence result on the Landweber iteration of Kaczmarz type.

Theorem 3.9 *Let $\mathcal{X}$ be reflexive and let $\mathcal{Y}_i$ be uniformly smooth, let Assumption 3.1 hold with $0 \leq \eta < 1$, and let $\Theta : \mathcal{X} \to (-\infty, \infty]$ be proper, lower semi-continuous, and p-convex function satisfies (2.2). Assume that (3.4) holds. Then for $\{\xi_n^\delta\}$ and $\{x_n^\delta\}$ defined by Algorithm 3.3 with $\tau > 1$ and $\mu_0 > 0$ satisfying*

$$1 - \eta - \frac{1+\eta}{\tau} - \frac{p-1}{p}\left(\frac{\mu_0}{2c_0}\right)^{\frac{1}{p-1}} > 0, \tag{3.19}$$

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

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

If in addition $\mathcal{N}(F_i'(x^\dagger)) \subset \mathcal{N}(F_i'(x))$ for all $x \in B_{2\rho}(x_0) \cap D(F)$ and $i = 0, \cdots, N-1$, then $x_* = x^\dagger$.

Proof. Let x_* be the solution of (3.1) determined in Lemma 3.7. Due to the p-convexity of Θ, it suffices to show that $\lim_{\delta\to 0} D_{\xi^\delta_{n_\delta}}\Theta(x_*, x^\delta_{n_\delta}) = 0$. We complete the proof by considering two cases.

Assume first that $\{y_i^{\delta_k}\}$, $i = 0, \cdots, N-1$, are sequences satisfying $\|y_i^{\delta_k} - y_i\| \le \delta_k$ with $\delta_k \to 0$ such that $n_k := n_{\delta_k} \to \hat{n}$ as $k \to \infty$ for some finite integer $\hat{n}$. We may assume $n_k = \hat{n}$ for all k. From the definition of $\hat{n} := n_k$ we have

$$\|F_i(x^{\delta_k}_{\hat{n}}) - y_i^{\delta_k}\| \le \tau\delta_k, \qquad i = 0, \cdots, N-1. \tag{3.20}$$

By taking $k \to \infty$ and using Lemma 3.8, we can obtain

$$F_i(x_{\hat{n}}) = y_i, \qquad i = 0, \cdots, N-1.$$

Using the definition of $\{\xi_n\}$ and $\{x_n\}$, this implies that $\xi_n = \xi_{\hat{n}}$ and $x_n = x_{\hat{n}}$ for all $n \ge \hat{n}$. Since Lemma 3.7 implies that $x_n \to x_*$ as $n \to \infty$, we must have $x_{\hat{n}} = x_*$. Consequently, by Lemma 3.8, $\xi^{\delta_k}_{n_k} \to \xi_{\hat{n}}$ and $x^{\delta_k}_{n_k} \to x_*$ as $k \to \infty$. This together with the lower semi-continuity of Θ implies that

$$\begin{aligned}
0 &\le \liminf_{k\to\infty} D_{\xi^{\delta_k}_{n_k}}\Theta(x_*, x^{\delta_k}_{n_k}) \le \limsup_{k\to\infty} D_{\xi^{\delta_k}_{n_k}}\Theta(x_*, x^{\delta_k}_{n_k}) \\
&= \Theta(x_*) - \liminf_{k\to\infty}\Theta(x^{\delta_k}_{n_k}) - \lim_{k\to\infty}\langle \xi^{\delta_k}_{n_k}, x_* - x^{\delta_k}_{n_k}\rangle \\
&\le \Theta(x_*) - \Theta(x_*) = 0.
\end{aligned}$$

Therefore $\lim_{k\to\infty} D_{\xi^{\delta_k}_{n_k}}\Theta(x_*, x^{\delta_k}_{n_k}) = 0$.

Assume next that $\{y_i^{\delta_k}\}$, $i = 0, \cdots, N-1$, are sequences satisfying $\|y_i^{\delta_k} - y_i\| \le \delta_k$ with $\delta_k \to 0$ such that $n_k := n_{\delta_k} \to \infty$ as $k \to \infty$. Let n be any fixed integer. then $n_k > n$ for large k. It then follows from (3.5) in Lemma 3.4 that

$$D_{\xi^{\delta_k}_{n_k}}\Theta(x_*, x^{\delta_k}_{n_k}) \le D_{\xi^{\delta_k}_{n}}\Theta(x_*, x^{\delta_k}_{n}) = \Theta(x_*) - \Theta(x^{\delta_k}_n) - \langle \xi^{\delta_k}_n, x_* - x^{\delta_k}_n\rangle.$$

By using Lemma 3.8 and the lower semi-continuity of Θ we obtain

$$\begin{aligned}
0 &\le \liminf_{k\to\infty} D_{\xi^{\delta_k}_{n_k}}\Theta(x_*, x^{\delta_k}_{n_k}) \le \limsup_{k\to\infty} D_{\xi^{\delta_k}_{n_k}}\Theta(x_*, x^{\delta_k}_{n_k}) \\
&\le \Theta(x_*) - \liminf_{k\to\infty}\Theta(x^{\delta_k}_n) - \lim_{k\to\infty}\langle \xi^{\delta_k}_n, x_* - x^{\delta_k}_n\rangle \\
&\le \Theta(x_*) - \Theta(x_n) - \langle \xi_n, x_* - x_n\rangle \\
&= D_{\xi_n}\Theta(x_*, x_n).
\end{aligned}$$

Since n can be arbitrary and since Lemma 3.7 implies that $D_{\xi_n}\Theta(x_*, x_n) \to 0$ as $n \to \infty$, we therefore have $\lim_{k\to\infty} D_{\xi^{\delta_k}_{n_k}}\Theta(x_*, x^{\delta_k}_{n_k}) = 0$. □

Remark 3.10 Without assuming the scaling condition (c) in Assumption 3.1, a good choice of the step length $\mu^\delta_{n,i}$ in Algorithm 3.3 requires the knowledge of $\|F_i'(x^\delta_{n,i})\|$ which is not easy to estimate. In order to avoid this inconvenience, we may consider the alternative choice

$$\mu^\delta_{n,i} = \begin{cases} \tilde{\mu}^\delta_{n,i}\|F_i(x^\delta_{n,i}) - y_i^\delta\|^{p-r} & \text{if } \|F_i(x^\delta_{n,i}) - y_i^\delta\| > \tau\delta, \\ 0, & \text{otherwise,} \end{cases}$$

where

$$\tilde{\mu}^\delta_{n,i} := \min\left\{ \frac{\mu_0\|F_i(x^\delta_{n,i}) - y_i^\delta\|^{p(r-1)}}{\|F_i'(x^\delta_{n,i})^* J_r^{\mathcal{Y}_i}(F_i(x^\delta_{n,i}) - y_i^\delta)\|^p},\ \mu_1 \right\}$$

with two positive constants μ_0 and μ_1. It is easy to see that

$$\min\{\mu_0 B_0^{-p}, \mu_1\} \le \tilde{\mu}_{n,i}^{\delta} \le \mu_1,$$

where $B_0 > 0$ is a constant such that $\|F_i'(x)\| \le B_0$ for all $x \in B_{2\rho}(x_0)$ and $i = 0, \cdots, N-1$. If $\tau > 1$ and $\mu_0 > 0$ are chosen such that (3.19) holds and $\mu_1 > 0$ is chosen to be any number, then, with some obvious modification in the proof of Lemma 3.4, we can obtain $n_\delta < \infty$ and the monotonicity result. The requirement $\tilde{\mu}_{n,i}^{\delta} \le \mu_1$ is to guarantee that the stability result in Lemma 3.8 remains true. Therefore, we can still obtain the same convergence result as in Theorem 3.9. In order to allow large step lengths, we usually take μ_1 to be a large number in practical applications.

Remark 3.11 In Algorithm 3.3 we may replace the step (iii) by the stopping criterion that defines n_δ to be the first integer satisfying

$$\sum_{i=0}^{N-1} \|F_i(x_{n_\delta,i}^{\delta}) - y_i^{\delta}\|^p \le N\tau^p\delta^p.$$

With some minor changes in the above arguments, we can still have the same convergence result as in Theorem 3.9. This new stopping criterion clearly terminates the iteration earlier than (iii) of Algorithm 3.3 and hence provides an opportunity of avoiding computing many additional iterations that do not have essential contribution in the final stage.

4. Proof of Proposition 3.6

Proposition 3.6 plays an important role in the convergence analysis of Algorithm 3.3 in Section 3. It shows that for some sequences $\{x_n\} \subset \mathcal{X}$ and $\{\xi_n\} \subset \mathcal{X}^*$ constructed by a suitable algorithm, the convergence of $\{x_n\}$ can be derived by showing certain monotonicity result together with a result like (3.13) along a suitable chosen subsequence of integers. This result might be useful for analyzing other methods as well. In the following we give the proof.

Proof of Proposition 3.6. We first show the convergence of $\{x_{n_k}\}$. For any $l < k$ we have from (2.1) and (3.13) that

$$\begin{aligned} D_{\xi_{n_l}}\Theta(x_{n_k}, x_{n_l}) &= D_{\xi_{n_l}}\Theta(\hat{x}, x_{n_l}) - D_{\xi_{n_k}}\Theta(\hat{x}, x_{n_k}) + \langle \xi_{n_k} - \xi_{n_l}, x_{n_k} - \hat{x}\rangle \\ &\le (1 + C_1)\left(D_{\xi_{n_l}}\Theta(\hat{x}, x_{n_l}) - D_{\xi_{n_k}}\Theta(\hat{x}, x_{n_k})\right). \end{aligned}$$

By the monotonicity of $\{D_{\xi_n}\Theta(\hat{x}, x_n)\}$ we can conclude that $D_{\xi_{n_l}}\Theta(x_{n_k}, x_{n_l}) \to 0$ as $k, l \to \infty$. In view of the uniformly convexity of Θ, it follows that $\{x_{n_k}\}$ is a Cauchy sequence in $\mathcal{X}$. Thus $x_{n_k} \to x_*$ for some $x_* \in B_{2\rho}(x_0) \subset \mathcal{X}$. Since $\lim_{n\to\infty} \|F_i(x_n) - y_i\| = 0$, we have $F_i(x_*) = y_i$ for all $i = 0, \cdots, N-1$.

In order to show $x_* \in D(\Theta)$, we use $\xi_{n_k} \in \partial\Theta(x_{n_k})$ to obtain

$$\Theta(x_{n_k}) \le \Theta(\hat{x}) + \langle \xi_{n_k}, x_{n_k} - \hat{x}\rangle. \tag{4.1}$$

In view of (3.13) we have

$$\Theta(x_{n_k}) \le \Theta(\hat{x}) + \langle \xi_{n_0}, x_{n_k} - \hat{x}\rangle + C_1 D_{\xi_{n_0}}\Theta(\hat{x}, x_{n_0}).$$

Since $x_{n_k} \to x_*$ as $k \to \infty$, by using the lower semi-continuity of Θ we obtain

$$\Theta(x_*) \le \liminf_{k\to\infty} \Theta(x_{n_k}) \le \Theta(\hat{x}) + \langle \xi_{n_0}, x_* - \hat{x}\rangle + C_1 D_{\xi_{n_0}}\Theta(\hat{x}, x_{n_0}) < \infty. \tag{4.2}$$

This implies that $x_* \in D(\Theta)$.

In order to derive the convergence in Bregman distance, we first use (3.13) to derive for $l < k$ that

$$|\langle \xi_{n_k}, x_{n_k} - x_* \rangle| \le C_1 \left(D_{\xi_{n_l}} \Theta(x_*, x_{n_l}) - D_{\xi_{n_k}} \Theta(x_*, x_{n_k}) \right) + |\langle \xi_{n_l}, x_{n_k} - x_* \rangle|.$$

By taking $k \to \infty$ and using $x_{n_k} \to x_*$ we can derive that

$$\limsup_{k\to\infty} |\langle \xi_{n_k}, x_{n_k} - x_* \rangle| \le C_1 \left(D_{\xi_{n_l}} \Theta(x_*, x_{n_l}) - \varepsilon_0 \right),$$

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

$$\limsup_{k\to\infty} |\langle \xi_{n_k}, x_{n_k} - x_* \rangle| \le C_1 (\varepsilon_0 - \varepsilon_0) = 0. \tag{4.3}$$

Using (4.1) with $\hat{x}$ replaced by x_* we thus obtain $\limsup_{k\to\infty} \Theta(x_{n_k}) \le \Theta(x_*)$. Combining this with (4.2) we therefore obtain

$$\lim_{k\to\infty} \Theta(x_{n_k}) = \Theta(x_*).$$

This together with (4.3) then implies that

$$\lim_{k\to\infty} D_{\xi_{n_k}} \Theta(x_*, x_{n_k}) = 0.$$

Since $\{D_{\xi_n} \Theta(x_*, x_n)\}$ is monotonically decreasing, we can conclude that

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

Finally we show that $x_* = x^\dagger$. We use (4.1) with $\hat{x}$ replaced by $x^\dagger$ to obtain

$$D_{\xi_0} \Theta(x_{n_k}, x_0) \le D_{\xi_0} \Theta(x^\dagger, x_0) + \langle \xi_{n_k} - \xi_0, x_{n_k} - x^\dagger \rangle. \tag{4.4}$$

By using (3.13), for any $\varepsilon > 0$ we can find k_0 such that

$$\left| \langle \xi_{n_k} - \xi_{n_{k_0}}, x_{n_k} - x^\dagger \rangle \right| < \frac{\varepsilon}{2}, \qquad k \ge k_0.$$

We next consider $\langle \xi_{n_{k_0}} - \xi_0, x_{n_k} - x^\dagger \rangle$. Since $\xi_{n+1} - \xi_n \in \overline{\mathcal{R}(F_0'(x^\dagger)^*)} + \cdots + \overline{\mathcal{R}(F_{N-1}'(x^\dagger)^*)}$, we can find $v_{n,i} \in \mathcal{Y}_i^*$ and $\beta_{n,i} \in \mathcal{X}^*$ such that

$$\xi_{n+1} - \xi_n = \sum_{i=0}^{N-1} \left(F_i'(x^\dagger)^* v_{n,i} + \beta_{n,i} \right) \quad \text{and} \quad \|\beta_{n,i}\| \le \frac{\varepsilon}{3NB_1 n_{k_0}}, \quad 0 \le n < n_{k_0},$$

where $B_1 > 0$ is a constant such that $\|x_n - x^\dagger\| \le B_1$ for all n. Consequently

$$\begin{aligned}
\left| \langle \xi_{n_{k_0}} - \xi_0, x_{n_k} - x^\dagger \rangle \right| &= \left| \sum_{n=0}^{n_{k_0}-1} \langle \xi_{n+1} - \xi_n, x_{n_k} - x^\dagger \rangle \right| \\
&= \left| \sum_{n=0}^{n_{k_0}-1} \sum_{i=0}^{N-1} \left[\langle v_{n,i}, F_i'(x^\dagger)(x_{n_k} - x^\dagger) \rangle + \langle \beta_{n,i}, x_{n_k} - x^\dagger \rangle \right] \right| \\
&\le \sum_{n=0}^{n_{k_0}-1} \sum_{i=0}^{N-1} \left(\|v_{n,i}\| \|F_i'(x^\dagger)(x_{n_k} - x^\dagger)\| + \|\beta_{n,i}\| \|x_{n_k} - x^\dagger\| \right) \\
&\le (1+\eta) \sum_{n=0}^{n_{k_0}-1} \sum_{i=0}^{N-1} \|v_{n,i}\| \|F_i(x_{n_k}) - y_i\| + \frac{\varepsilon}{3}.
\end{aligned}$$

Since $\|F_i(x_n) - y_i\| \to 0$ as $n \to \infty$, we can find $k_1 \geq k_0$ such that

$$|\langle \xi_{n_{k_0}} - \xi_0, x_{n_k} - x^\dagger \rangle| < \frac{\varepsilon}{2}, \qquad \forall k \geq k_1.$$

Therefore $|\langle \xi_{n_k} - \xi_0, x_{n_k} - x^\dagger \rangle| < \varepsilon$ for all $k \geq k_1$. Since $\varepsilon > 0$ is arbitrary, we obtain $\lim_{k\to\infty} \langle \xi_{n_k} - \xi_0, x_{n_k} - x^\dagger \rangle = 0$. By taking $k \to \infty$ in (4.4) and using $\Theta(x_{n_k}) \to \Theta(x_*)$ we obtain

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

According to the definition of $x^\dagger$ we must have $D_{\xi_0}\Theta(x_*, x_0) = D_{\xi_0}\Theta(x^\dagger, x_0)$. A direct application of Lemma 3.2 gives $x_* = x^\dagger$. □

Remark 4.1 In the proof of Proposition 3.6, we obtain that $\lim_{k\to\infty} \Theta(x_{n_k}) = \Theta(x_*)$. It is not clear if there holds

$$\lim_{n\to\infty} \Theta(x_n) = \Theta(x_*) \tag{4.5}$$

for the whole sequence $\{x_n\}$. Although the result of Proposition 3.6 implies $x_n \to x_*$, we can not use it to derive (4.5) directly since Θ is not necessarily continuous at x_*.

5. Numerical examples

In this section we will present some numerical simulations on Algorithm 3.3. A key ingredient in this algorithm is the resolution of the minimization problem

$$x = \arg\min_{z\in\mathcal{X}} \{\Theta(z) - \langle \xi, z \rangle\} \tag{5.1}$$

for any given $\xi \in \mathcal{X}^*$. For some choices of Θ, this minimization problem can be easily solved numerically. In particular, when $\mathcal{X} = L^2(\Omega)$ and

$$\Theta(x) := \frac{1}{2\beta} \int_\Omega |x(\omega)|^2 d\omega + \int_\Omega |x(\omega| d\omega \tag{5.2}$$

with $\beta > 0$, the minimizer of (5.1) can be given explicitly by the soft thresholding

$$x(\omega) = \begin{cases} \beta(\xi(\omega) - 1) & \text{if } \xi(\omega) > 1, \\ 0 & \text{if } |\xi(\omega)| \leq 1, \\ \beta(\xi(\omega) + 1) & \text{if } \xi(\omega) < -1. \end{cases}$$

For the total variation like functional

$$\Theta(x) := \frac{1}{2\beta} \int_\Omega |x(\omega)|^2 d\omega + \int_\Omega |Dx| \tag{5.3}$$

with $\beta > 0$ in $\mathcal{X} := L^2(\Omega)$, although there is no explicit formula for the minimizer of (5.1), there are many numerical solvers developed in the literature; in our numerical simulations, we will use the monotone version of the fast iterative shrinkage/thresholding algorithm (MFISTA) introduced in [1].

5.1. Photoacoustic/thermoacoustic tomography

Let $\Omega \subset \mathbb{R}^2$ be a bounded convex domain in $\mathbb{R}^2$ with smooth boundary $\partial\Omega$. We consider the problem of recovering a function $f : \mathbb{R}^2 \to \mathbb{R}$ supported over Ω from its means

$$(\mathcal{M}f)(x,r) := \frac{1}{2\pi}\int_{\mathbb{S}^1} f(x+r\sigma)d\sigma$$

over the unit circle $\mathbb{S}^1$ with centers $x \in \partial\Omega$ and radii $r > 0$. The resolution of this problem is a key ingredient of many modern imaging techniques, including the photoacoustic tomography (PAT) and thermoacoustic tomography (TAT).

PAT/TAT is a hybrid imaging technique based on the photoacoustic effect which refers to the generation of acoustic waves by the absorption of electromagnetic (EM) energy coming from visible light, radio waves, or microwaves. This technique combines the ultrasonic resolution with the high EM contrast to overcome the low contrast of pure ultrasound diagnostics and weak resolution of pure optical imaging methodologies in soft tissues (see [12, 18] and references therein). It provides diagnostics by determining the EM energy absorption function $f(x)$ which strongly depends on the local biological properties of the cells. Upon absorption of a short EM pulse, it results in thermoelastic expansion and thus in acoustic waves $p(x,t)$ that can be measured by ultrasound receivers placed on a surface S around the object. These measured data $p(x,t)$, $x \in S$ are then used to determine $f(x)$. By assuming that the sound velocity is equal to the unit constant, the acoustic wave $p(x,t)$ satisfies the Cauchy problem for the wave equation

$$\begin{cases} \partial_t^2 p - \Delta p = 0 & \text{in} \quad \mathbb{R}^3 \times \{t \geq 0\}, \\ p(x,0) = f(x), & \partial_t p(x,0) = 0, \quad x \in \mathbb{R}^3. \end{cases}$$

The 3D TAT/PAT imaging can be reduced to 2D problems by introducing integrating line detectors [4]. Assuming the parallel detectors are perpendicular to the x_2-axis. Let the direction of detectors be x_3-axis, the detectors measure $\bar{p}(x',t)$, $x' := (x_1, x_2) \in C$, for some detection curve C in the x_1x_2-plane, where $\bar{p}(x',t)$ denotes the projection of $p(x,t)$ in the x_3-direction defined by

$$\bar{p}(x',t) := \int_{-\infty}^{\infty} p(x', x_3, t)dx_3$$

and satisfies the 2D wave equation

$$\begin{cases} \partial_t^2 \bar{p} - \Delta \bar{p} = 0 & \text{in} \quad \mathbb{R}^2 \times \{t \geq 0\}, \\ \bar{p}(x',0) = \bar{f}(x'), & \partial_t \bar{p}(x',0) = 0, \quad x' \in \mathbb{R}^2, \end{cases} \tag{5.4}$$

where $\bar{f}(x') = \int_{-\infty}^{\infty} f(x', x_3)dx_3$. Once $\bar{f}(x')$ can be determined from $\bar{p}(x',t)$, $x' \in C$, we may rotate the detectors around the x_2-axis to obtain $\int_L f(x)d\ell$ for all lines L perpendicular to x_2 axis which enable us to use the inversion of Radon transform to reconstruct $f(x)$.

In order to determine $\bar{f}(x')$ from $\bar{p}(x',t)$, $(x',t) \in C \times [0,\infty)$, we use the representation formula for the solution of (5.4) to obtain ([16])

$$\bar{p}(x',t) = \partial_t \int_0^t \frac{r(\mathcal{M}\bar{f})(x',r)}{\sqrt{t^2 - r^2}}dr.$$

By using the solution formula of Abel integral equation, it follows ([4])

$$(\mathcal{M}\bar{f})(x',r) = \frac{2}{\pi}\int_0^r \frac{\bar{p}(x',t)}{\sqrt{r^2 - t^2}}dt, \quad x' \in C \text{ and } r \geq 0.$$

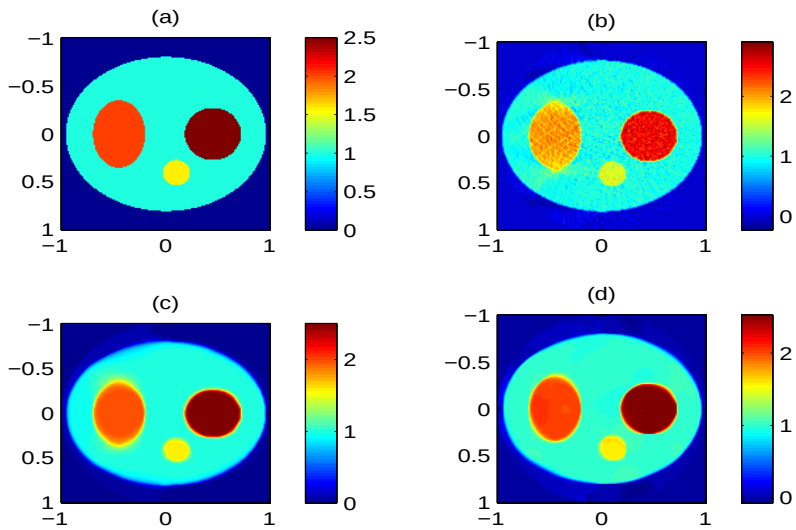

Figure 1. Numerical results with $N = 80$ measurements and $\delta = 0.01$: (a) exact solution; (b) $\Theta(f) = \|f\|^2_{L^2}$; (c) and (d) $\Theta(f) = \frac{1}{2\beta}\|f\|^2_{L^2} + \int_\Omega |Df|$ with $\beta = 1$ and $\beta = 10$ respectively

Therefore we need to determine $\bar{f}(x')$ from its means $(\mathcal{M}\bar{f})(x', r)$, $x' \in C$ and $r \geq 0$.

In numerical simulations, we reconstruct a function f supported on the disk B_R of radius $R = 0.96$ centered at the origin from its means $(\mathcal{M}f)(\mathbf{x}_j, r)$, $r \geq 0$, measured at N points $\mathbf{x}_j = R(\sin(j\pi/N), \cos(j\pi/N))$, $j = 0, \cdots, N-1$ uniformly distributed on the semicircle $S^+ := \{\mathbf{x} \in \partial B_R : x_1 \geq 0\}$. This is equivalent to solving the system

$$\mathcal{M}_i f = g_i, \quad i = 0, \cdots, N-1,$$

where

$$(\mathcal{M}_i f)(r) := (\mathcal{M}f)(\mathbf{x}_i, r) = \frac{1}{2\pi}\int_{\mathbb{S}^1} f(\mathbf{x}_i + r\sigma)d\sigma.$$

It is easy to check [8] that the operators $\mathcal{M}_i$ can be continuously extended to

$$\mathcal{M}_i : L^2(B_R) \to L^2([0, 2R], rdr)$$

with $\|\mathcal{M}_i\| \leq 2\sqrt{\pi}$ and the adjoint $\mathcal{M}_i^* : L^2([0, 2R], rdr) \to L^2(B_R)$ is given by

$$(\mathcal{M}_i^* g)(\mathbf{x}) = 2g(|\mathbf{x}_i - \mathbf{x}|).$$

In Figure 1 we report the numerical results with $N = 80$ measurements. In order to approximate functions, we take 160×160 grid points uniformly distributed on the square $[-1, 1] \times [-1, 1]$. The exact piecewise constant phantom $f^\dagger$ is shown in (a). For the simulations, we add 2% uniformly distributed noise to $\mathcal{M}_i f^\dagger$ to produce the noisy data which are then used to reconstruct $f^\dagger$. Figure 1 (b)–(d) give the reconstruction result by Algorithm 3.3 with initial guess $f_0 = \xi_0 = 0$ and different choices of the uniformly convex functionals Θ, we take $r = 2$, $\tau = 1.2$ and $\mu_0 = (1 - 1/\tau)/(\beta\sqrt{\pi})$ when Θ given by (5.2) and (5.3) are used. Figure 1 (b) presents the reconstruction result by Algorithm 3.3 with $\Theta(f) = \int_{B_R} |f(\mathbf{x})|^2 d\mathbf{x}$. Figure 1 (c) and (d) report the

reconstruction results by Algorithm 3.3 with $\Theta(f)$ defined by (5.3) with $\beta = 1$ and $\beta = 10$ respectively. The results in (c) and (d) significantly improve the one in (b) by efficiently removing the artefacts due to the notorious oscillatory effect and indicate that the results are robust with respect to β.

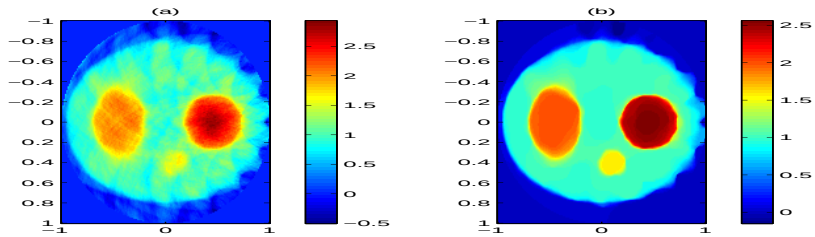

Figure 2. Numerical results with $N = 10$ measurements and $\delta = 0.01$: (a) $\Theta(f) = \|f\|^2_{L^2}$; (b) $\Theta(f) = \frac{1}{2}\|f\|^2_{L^2} + \int_\Omega |Df|$

In Figure 2 we report the computation results by reducing the number of measurements to $N = 10$. The reconstruction ability of the method using $\Theta(f) = \|f\|^2_{L^2}$ becomes worse, the method with $\Theta(f) = \frac{1}{2}\|f\|^2_{L^2} + \int_\Omega |Df|$, however, still has good reconstruction. This reflects the philosophy in compressed sensing: it is possible to reconstruct an image from very few number of measurements by using L^1 penalty term if it is sparse under suitable transformation.

5.2. Parameter identification

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

$$\begin{cases} -\triangle u + cu = f & \text{in } \Omega, \\ u = g & \text{on } \partial\Omega, \end{cases} \tag{5.5}$$

from an $L^2(\Omega)$-measurement of the state u, where $\Omega \subset \mathbb{R}^d$ with $d \le 3$ is a bounded domain with Lipschitz boundary, $f \in L^2(\Omega)$ and $g \in H^{3/2}(\Omega)$. We assume that the exact solution $c^\dagger$ is in $L^2(\Omega)$. This problem reduces to solving $F(c) = u$, i.e. (3.1) with $N = 1$, if we define the nonlinear operator $F : L^2(\Omega) \to L^2(\Omega)$ by

$$F(c) := u(c)\,, \tag{5.6}$$

where $u(c) \in H^2(\Omega) \subset L^2(\Omega)$ is the unique solution of (5.5). This operator F is well defined on

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

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

$$F'(c)h = -A(c)^{-1}(hF(c)) \quad \text{and} \quad F'(c)^* w = -u(c)A(c)^{-1}w \tag{5.7}$$

for $h, w \in L^2(\Omega)$, where $A(c) : H^2 \cap H^1_0 \to L^2$ is defined by $A(c)u = -\triangle u + cu$ which is an isomorphism uniformly in ball $B_\rho(c^\dagger) \cap D(F)$ for small $\rho > 0$. Moreover, Assumption 3.1 holds for small $\rho > 0$ (see [10]).

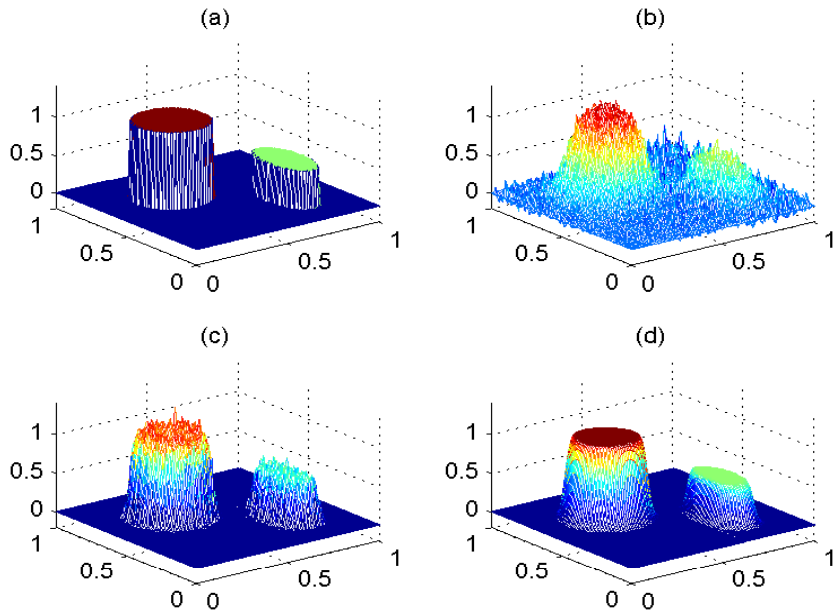

Figure 3. numerical results with $\delta = 0.5 \times 10^{-4}$ and $\tau = 1.1$: (a) exact solution; (b) $\Theta(f) = \|f\|_{L^2}^2$; (c) $\Theta(c) = \frac{1}{2}\|c\|_{L^2}^2 + \|c\|_{L^1}$; (d) $\Theta(f) = \frac{1}{2}\|f\|_{L^2}^2 + \int_\Omega |Df|$

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

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

We assume $u(c^\dagger) = x + y$ and add random noise to produce u^δ satisfying $\|u^\delta - u(c^\dagger)\|_{L^2(\Omega)} = \delta$ with $\delta = 0.5 \times 10^{-4}$. In order to reconstruct $c^\dagger$, we use Algorithm 3.3 with $r = 2$, $N = 1$ and $\tau = 1.1$ and take $c_0 = \xi_0 = 0$ as initial guess; we take the step length $\mu_{n,i}^\delta$ according to Remark 3.10 with

$$\tilde{\mu}_{n,i}^\delta = \min\left\{ \frac{\mu_0 \|F_i(c_{n,i}^\delta) - y_i^\delta\|_{L^2}^2}{\|F_i'(c_{n,i}^\delta)^*(F_i(c_{n,i}^\delta) - y_i^\delta)\|_{L^2}^2}, \mu_1 \right\}$$

with $\mu_0 = (1 - 1/\tau)/\beta$ and $\mu_1 = 4000$ when Θ given by (5.2) and (5.3) are used. The partial differential equations involved are solved approximately by a finite difference method by dividing Ω into 100×100 small squares of equal size.

We report the numerical results in Figure 3. In (a) we plot the exact solution $c^\dagger(x,y)$. In (b) we plot the result of Algorithm 3.3 with $\Theta(c) = \|c\|_{L^2}^2$; although the reconstruction tells something on the sought solution, it does not tell more information such as sparsity, discontinuities and constancy since the result is too oscillatory. In (c) we report the result of Algorithm 3.3 with Θ given by (5.2) with $\beta = 1$. It is clear that the sparsity of the sought solution is significantly reconstructed. The reconstruction

result, however, is still oscillatory on the nonzero parts which is typical for this choice of Θ. In (d) we report the result of Algorithm 3.3 with $\Theta(c)$ given by (5.3) with $\beta = 1$. The reconstruction is rather satisfactory and the notorious oscillatory effect is efficiently removed.

5.3. Schlieren imaging

Consider the problem of reconstructing a function f supported on a bounded domain $D \subset \mathbb{R}^2$ from

$$\mathcal{I}f(s,\sigma) = \left(\int_{\mathbb{R}} f(s\sigma + r\sigma^{\perp})dr\right)^2, \qquad (s,\sigma) \in \mathbb{R} \times \mathbb{S}^1.$$

This problem arises from determining the 3D pressure fields on cross-sections of a water tank generated by an ultrasound transducer from Schlieren data. The data are collected with a Schlieren optical system based on Raman scattering. The Schlieren optical system outputs the intensity of light through the tank which is proportional to the square of the line integral of the pressure along the light path [9].

In our numerical simulations we reconstruct a function f support on $D = [-1,1] \times [-1,1]$ from $N = 100$ recording angles $\sigma_i \in \mathbb{S}^1$, $i = 0, \cdots, N-1$ uniformly distributed on the semicircle. It reduces to solving the system (3.1) by introducing

$$F_i(f)(s) := \mathcal{I}f(s,\sigma_i) := \left(\mathcal{R}_i f(s)\right)^2,$$

where $\mathcal{R}_i f(s) = \int_{\mathbb{R}} f(s\sigma_i + r\sigma_i^{\perp})dr$ is the Radon transform. It is easy to show ([8]) that each $F_i : H_0^1(D) \to L^2(I)$ with $I = [-\sqrt{2}, \sqrt{2}]$ is Fréchet differentiable with

$$F_i'(f)h = \mathcal{R}_i f \cdot \mathcal{R}_i h, \quad \forall h \in H_0^1(D)$$

and the adjoint $F_i'(f)^* : L^2(I) \to H_0^1(D)$ is given by

$$F_i'(f)^* g = (I - \Delta)^{-1}(2\mathcal{R}_i^{\#}(g\mathcal{R}_i f)), \quad \forall g \in L^2(I),$$

where $\mathcal{R}_i^{\#} : L^2(I) \to L^2(D)$ is the adjoint of $\mathcal{R}_i$ given by $(\mathcal{R}_i^{\#} g)(x) = g(\langle x, \sigma_i \rangle)$.

In Figure 4 we report the numerical results with $N = 100$ measurements with the exact solution $f^\dagger$ shown in (a). For the simulations, we use noisy data with noise level $\delta = 0.002$ to reconstruct $f^\dagger$; in order to approximate functions, we take 120×120 grid points uniformly distributed over $[-1,1] \times [-1,1]$. Figure 4 (b)–(d) give the reconstruction results by Algorithm 3.3 implemented with $r = 2$ and $\tau = 1.5$, initial guess $\xi_0 = 0.01$ and different choices of Θ; the step length $\mu_{n,i}^\delta$ is chosen according to Remark 3.10 with $\mu_0 = (1 - 1/\tau)/\beta$ and $\mu_1 = 1000$. Figure 4 (b) clearly contains many artefacts. Figure 4 (c) removes almost all artefacts since it uses the L^1 like penalty functional, the reconstruction on nonzero profile, however, turns out to be unsatisfactory. Because of the use of total variation like penalty functional, Figure 4 (d) efficiently removes the artefacts and reconstructs the profile very well.

Acknowledgement

The work of Q Jin is partially supported by the DECRA grant DE120101707 of Australian Research Council.

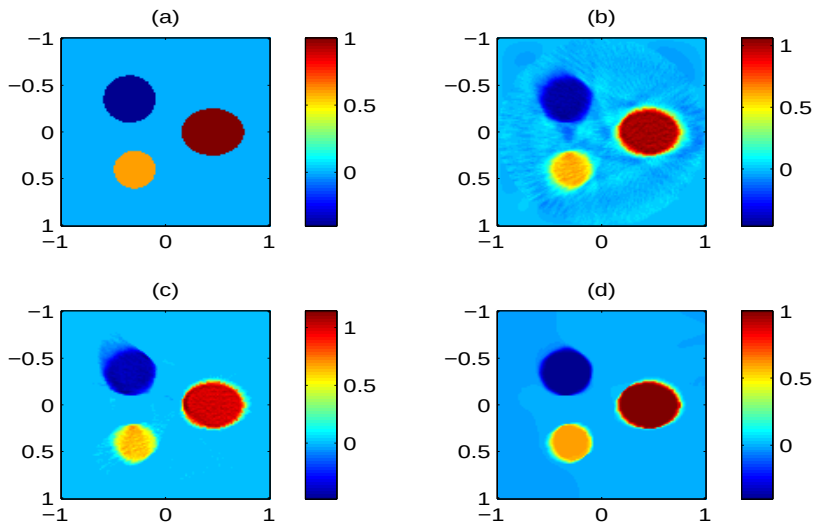

Figure 4. Numerical results with $N = 100$ measurements, $\delta = 0.002$ and $\tau = 1.5$: (a) exact solution; (b) $\Theta(f) = \|f\|^2_{L^2}$; (c) $\Theta(f) = \frac{1}{2}\|f\|^2_{L^2} + \|f\|_{L^1}$; (d) $\Theta(f) = \frac{1}{2}\|f\|^2_{L^2} + \int_\Omega |Df|$

References

[1] A. Beck and M. Teboulle, *Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems*, IEEE Trans. Image Process, 18 (2009), no. 11, 2419–2434.

[2] R. Boţ and T. Hein, *Iterative regularization with a general penalty term – theory and application to L^1 and TV regularization*, Inverse Problems, 28 (2012), 104010(19pp).

[3] L. M. Bregman, *The relaxation method for finding common points of convex sets and its application to the solution of problems in convex programming*, USSR Comput. Math. Math. Phys., 7 (1967), 200–217.

[4] P. Burgholzer, J. Bauer-Marschallinger, H. Grün, M. Haltmeier and G. Paltauf, *Temporal back-projection algorithms for photoacoustic tomography with integrating line detectors*, Inverse Problems, 23 (2007) S65–S80.

[5] H. W. Engl, M. Hanke and A. Neubauer, *Regularization of Inverse Problems*, Kluwer, Dordrecht, 1996.

[6] E. Giusti, *Minimal Surfaces and Functions of Bounded Variation*, Basel: Birkhäuser, 1984.

[7] M. Haltmeier, A. Leitão, O. Scherzer, *Kaczmarz methods for regularizing nonlinear ill-posed equations: I. Convergence analysis*, Inverse Probl. Imaging, 1 (2007), 289–298.

[8] M. Haltmeier, R. Kowar, A. Leitão and O. Scherzer, *Kaczmarz methods for regularizing nonlinear ill-posed equations. II. Applications*, Inverse Probl. Imaging, 1 (2007), no. 3, 507–523.

[9] A. Hanafy and C. I. Zanelli, *Quantitative real-time pulsed Schlieren imaging of ultrasonic waves*, Proc. IEEE Ultrasonics Symposium, 2 (1991), 1223–11227.

[10] M. Hanke, A. Neubauer and O. Scherzer, *A convergence analysis of the Landweber iteration for nonlinear ill-posed problems*, Numer. Math., 72 (1995), 21–37.

[11] R. Kowar and O. Scherzer, *Convergence analysis of a Landweber-Kaczmarz method for solving nonlinear ill-posed problems*, Ill-Posed and Inverse Problems, 253–270, VSP, Zeist, 2002.

[12] P. Kuchment and L. Kunyansky, *Mathematics of thermoacoustic and photoacoustic tomography*, European J. Appl. Math., 19 (2008), 191–224.

[13] A. Leitão and M. M. Alves, *On Landweber-Kaczmarz methods for regularizing systems of ill-posed equations in Banach spaces*, Inverse Problems, 28 (2012), no. 10, 104008, 15 pp.

[14] L. Rudin, S. Osher, and C. Fatemi, *Nonlinear total variation based noise removal algorithm*, Phys. D, 60 (1992), 259–268.

[15] F. Schöpfer, A. K. Louis and T. Schuster, *Nonlinear iterative methods for linear ill-posed problems in Banach spaces*, Inverse Problems, 22 (2006), 311–329.

[16] C. D. Sogge, *Lectures on Nonlinear Wave Equations*, 2nd edition, Monographs in analysis, Internation Press, 2008.

[17] R. Tibshirani, *Regression shrinkage and selection via the lasso*, J. Roy. Statist. Soc. Ser. B, 58 (1996), no. 1, 267–288.

[18] M. Xu and L.V. Wang, *Photoacoustic imaging in biomedicine*, Review of Scientific Instruments, 77 (2006), 041101.

[19] C. Zălinscu, *Convex Analysis in General Vector Spaces*, World Scientific Publishing Co., Inc., River Edge, New Jersey, 2002.