IOP PUBLISHING INVERSE PROBLEMS

Inverse Problems **28** (2012) 104008 (15pp) doi:10.1088/0266-5611/28/10/104008

On Landweber–Kaczmarz methods for regularizing systems of ill-posed equations in Banach spaces

A Leitão and M Marques Alves

Department of Mathematics, Federal University of St. Catarina, PO Box 476, 88040-900 Florianópolis, Brazil

E-mail: acgleitao@gmail.com and maicon@impa.br

Received 15 April 2012, in final form 3 September 2012
Published 1 October 2012
Online at stacks.iop.org/IP/28/104008

Abstract

In this paper, iterative regularization methods of Landweber–Kaczmarz type are considered for solving systems of ill-posed equations modeled (finitely many) by operators acting between Banach spaces. Using assumptions of uniform convexity and smoothness on the parameter space, we are able to prove a monotony result for the proposed method, as well as to establish convergence (for exact data) and stability results (in the noisy data case).

1. Introduction

1.1. Systems of nonlinear ill-posed equations

In this paper, we propose a new method for obtaining regularized approximations of systems of nonlinear ill-posed operator equations in Banach spaces.

The *inverse problem* we are interested in consists in determining an unknown physical quantity $x \in X$ from the set of data $(y_1, \ldots, y_m) \in Y^m$, where X and Y are Banach spaces, with X being uniformly convex and smooth [6], and $m \geqslant 1$.

In practical situations, we do not know the data exactly. Instead, we have only approximate measured data $y_i^\delta \in Y$ satisfying

$$\|y_i^\delta - y_i\| \leqslant \delta_i, \quad i = 1, \ldots, m, \tag{1}$$

with $\delta_i > 0$ (noise level). The finite set of data above is obtained by indirect measurements of the parameter, this process being described by the model

$$F_i(x) = y_i, \quad i = 1, \ldots, m, \tag{2}$$

where $F_i : D_i \subset X \to Y$, and D_i are the corresponding domains of definition.

Standard methods for the solution of system (2) are based on the use of *iterative-type* regularization methods [1, 8, 17, 23, 18] or *Tikhonov-type* regularization methods [8, 20, 25, 22] after rewriting (2) as a single equation $F(x) = y$, where

$$F := (F_1, \ldots, F_m) : \bigcap_{i=1}^{m} D_i =: D \to Y^m \tag{3}$$

and $y := (y_1, \ldots, y_m)$. However, these methods become inefficient if m is large or the evaluations of $F_i(x)$ and $F_i'(x)^*$ are expensive. In such a situation, Kaczmarz-type methods [16, 19, 21] which cyclically consider each equation in (2) separately are much faster [21] and are often the method of choice in practice (see section 1.3).

Example 1.1. A tutorial example of an inverse problem of the form (2) is the identification of the space-dependent coefficient $a(x)$ (bounded away from zero) in the elliptic model

$$-\nabla(a\nabla u) = f, \text{ in } \Omega, \qquad u = 0, \text{ at } \partial\Omega,$$

where $\Omega \subset \mathbb{R}^2$ is an open bounded domain with a regular (smooth) boundary $\partial\Omega$. Available data for the identification problem are $u|_{\Omega_i}$, i.e. the restrictions of the solution u to the given open sets $\Omega_i \subset \Omega$, $i = 1, \ldots, m$.

In the standard Hilbert space setting [8, 2], we have $F_i : H^2(\Omega) = X \supset D_i \ni a \mapsto (\Delta_a^{-1} f)|_{\Omega_i} \in Y_i = L^2(\Omega_i)$, where $\Delta_a : H^2(\Omega) \cap H_0^1(\Omega) \ni u \mapsto -\nabla(a\nabla u) \in L^2(\Omega)$ and $D_i = D := \{a \in X;\, a(x) \geqslant \underline{a} > 0, \text{ a.e. in } \Omega\}$, $i = 1, \ldots, m$.

A possible Banach space setting for this problem is analyzed in [18] (for the case $m = 1$ and $\Omega_1 = \Omega$), where the choice $X = W^{1,q}(\Omega)$, $Y_i = L^r(\Omega_i)$ with $q > 2$ and $r \in (1, \infty)$ is considered. In particular, it follows from [18, corollary 3] that the results of the convergence analysis derived here can be applied to this parameter identification problem (see assumption 3.1).

1.2. Regularization in Banach spaces

Ill-posed operator equations in Banach spaces are a fast-growing area of research. Over the last seven years, several theoretical results have been derived in this field.

- The classical paper on regularization of ill-posed problems in Banach spaces by Resmerita [22].
- Tikhonov's regularization in Banach spaces is also investigated in [4], where two distinct iterative methods for finding the minimizer of norm-based Tikhonov functionals are proposed and analyzed (convergence is proven). Moreover, convergence rates results for Tikhonov's regularization in Banach spaces are considered in [15].
- In [23], a nonlinear extension of the Landweber method to linear operator equations in Banach spaces is investigated using duality mappings. The same authors considered in [24] the solution of convex split feasibility problems in Banach spaces by cyclic projections. See also [14, 13] for the convergence analysis of modified Landweber iterations in Banach spaces.
- In [18], the nonlinear Landweber method and the IRGN method are considered for a single (nonlinear) operator equation in Banach spaces, and convergence results are derived. Moreover, the applicability of the proposed methods to parameter identification problems for elliptic PDEs is investigated.
- The Gauss–Newton method in Banach spaces is considered in [1] for a single operator equation in the special case $X = Y$. A convergence result is obtained and convergence rates (under strong source conditions) are provided.

The starting point of our approach is the Landweber method [23, 18] for solving ill-posed problems in Banach spaces[1]. In the case of a single operator equation, i.e. $m = 1$ in (2), this method is defined by

$$x_n^* = J_p(x_n) - \mu_n F'(x_n)^* J_r\big(F(x_n) - y^\delta\big), \qquad x_{n+1} = J_q\big(x_n^*\big), \tag{4}$$

[1] See also [1, 8, 17] for the analysis of the Landweber method in Hilbert spaces.

where $F'(x)$ is the Fréchet derivative of F at point x, and J_p, J_r and J_q are the duality mappings from X, Y and X^* to their duals, respectively. Moreover, $x_0 \in D$ and $p, q, r \in (1, \infty)$ satisfy $p + q = pq$.

The step size μ_n depends on the constant of the tangential cone condition, the constant of the discrepancy principle, the residual at x_n and a constant describing geometrical properties of the Banach spaces (see [23, section 3]).

The convergence analysis for the linear case $F \in \mathcal{L}(X, Y)$ can be found in [23], while convergence for nonlinear operator equations is derived in [18], where X is assumed to be uniformly smooth and uniformly convex (actually, X is assumed to be p-convex, which is equivalent to the dual being q-smooth, i.e. there exists a constant $C_q > 0$ such that for all x^*, $y^* \in X^*$, $\|x^* - y^*\|^q \leqslant \|x^*\|^q - q\langle J_q(x^*), y^*\rangle + C_q\|y^*\|^q$ follows; see [18, section 2.2]). For a detailed definition of smoothness, uniform smoothness and uniform convexity in Banach spaces, we refer the reader to [6, 23].

1.3. Landweber–Kaczmarz method in Banach spaces

The Landweber–Kaczmarz method in Banach spaces (LKB) consists in incorporating the (cyclic) Kaczmarz strategy to the Landweber method depicted in (4) for solving the system of operator equations in (2).

This strategy is an analogue of the one proposed in [10, 9] regarding the Landweber–Kaczmarz (LK) iteration in Hilbert spaces. See also [7] for the steepest-descent-Kaczmarz (SDK) iteration, [11] for the expectation-maximization-Kaczmarz (EMK) iteration, [3] for the Levenberg–Marquardt–Kaczmarz (LMK) iteration, and [2] for the iterated-Tikhonov–Kaczmarz (ɪTK) iteration.

Motivated by the ideas in the above-mentioned papers (in particular by the approach in [11], where $X = L^1(\Omega)$ and convergence is measured with respect to the Kullback–Leibler distance), we propose in this paper the LBK method, which is sketched as follows:

$$x_n^* = J_p(x_n) - \mu_n F'_{i_n}(x_n)^* J_r\big(F_{i_n}(x_n) - y^\delta_{i_n}\big), \quad x_{n+1} = J_q(x_n^*), \tag{5}$$

for $n = 0, 1, \ldots$. Moreover, $i_n := (n \bmod m) + 1 \in \{1, \ldots, m\}$, and $x_0 \in X\backslash\{0\}$ is an initial guess, possibly incorporating *a priori* knowledge about the exact solution (which may not be unique).

Here, $\mu_n \geqslant 0$ is chosen analogously as in (4) if $\|F_{i_n}(x_n) - y^\delta_{i_n}\| \geqslant \tau\delta_{i_n}$ (see section 3 for the precise definition of μ_n and the discrepancy parameter $\tau > 0$). Otherwise, we set $\mu_n = 0$. Consequently, $x_{n+1} = J_q(x_n^*) = J_q(J_p(x_n)) = x_n$ every time the residual of the iterate x_n w.r.t. the i_nth equation of system (2) drops below the discrepancy level given by $\tau\delta_{i_n}$.

Due to the bang-bang strategy used to define the sequence of parameters (μ_n), the iteration in (5) is alternatively called loping Landweber–Kaczmarz method in Banach spaces.

As usual in Kaczmarz-type algorithms, a group of m subsequent steps (beginning at some integer multiple of m) is called a *cycle*. The iteration should be terminated when, for the first time, all of the residuals $\|F_{i_n}(x_{n+1}) - y^\delta_{i_n}\|$ drop below a specified threshold within a cycle. That is, we stop the iteration at the step

$$\hat{n} := \min\{\ell m + (m-1) : \ell \in \mathbb{N}, \ \ \|F_i(x_{\ell m+i-1}) - y_i^\delta\| \leqslant \tau\delta_i, \ \text{for } 1 \leqslant i \leqslant m\}. \tag{6}$$

In other words, writing $\hat{n} := \hat{\ell} m + (m-1)$, (6) can be interpreted as $\|F_i(x_{\hat{\ell} m+i-1}) - y_i^\delta\| \leqslant \tau\delta_i$, $i = 1, \ldots, m$. In the case of noise-free data ($\delta_i = 0$ in (1)), the stop criteria in (6) may never be reached, i.e. $\hat{n} = \infty$ for $\delta_i = 0$.

Outline of the manuscript

In section 2, we introduce the notation used in this paper and briefly recall some results on the convex analysis and Bregman distances, which are necessary for the analysis presented in the forthcoming sections. In section 3, the LK algorithm for solving systems of nonlinear ill-posed equations in Banach spaces is formulated. Moreover, some preliminary results are derived, namely boundedness and monotony of iteration error and residual. In section 4, the main results of the manuscript are presented. A convergence analysis of the proposed method is given, and stability results are proven. Section 5 is devoted to conclusions and discussion of future work perspectives.

2. Overview of the convex analysis and Bregman distances

2.1. Convex analysis

Let X be a (nontrivial) real Banach space with the topological dual X^*. By $\|\cdot\|$ we denote the norm on X and X^*. The duality product on $X \times X^*$ is a bilinear symmetric mapping, denoted by $\langle \cdot, \cdot \rangle$, and defined as $\langle x, x^* \rangle = x^*(x)$, for all $(x, x^*) \in X \times X^*$.

Let $f : X \to (-\infty, \infty]$ be convex, proper and lower semicontinuous. Recall that f is convex lower semicontinuous when its epigraph $\mathrm{epi}(f) := \{(x, \lambda) \in X \times \mathbb{R} : f(x) \leqslant \lambda\}$ is a closed convex subset of $X \times \mathbb{R}$. Moreover, f is proper when its domain $\mathrm{dom}(f) := \{x \in X : f(x) < \infty\}$ is nonempty. The *subdifferential* of f is the (point-to-set) operator $\partial f : X \to 2^{X^*}$ defined at $x \in X$ by

$$\partial f(x) = \{x^* \in X^* : f(y) \geqslant f(x) + \langle x^*, y - x \rangle \quad \forall y \in X\}. \tag{7}$$

Note that $\partial f(x) = \emptyset$, whenever $x \notin \mathrm{dom}(f)$. The domain of ∂f is the set $\mathrm{dom}(\partial f) = \{x \in X : \partial f(x) \neq \emptyset\}$. Next, we present a very useful characterization of ∂f using the concept of *Fenchel conjugation*. The Fenchel conjugate of f is the lower semicontinuous convex function $f^* : X^* \to (-\infty, \infty]$ defined at $x^* \in X^*$ by

$$f^*(x^*) = \sup_{x \in X} \langle x, x^* \rangle - f(x). \tag{8}$$

It is well known that f^* is also proper whenever f is proper. The *Fenchel–Young* inequality follows directly from (8):

$$f(x) + f^*(x^*) \geqslant \langle x, x^* \rangle \ \ \forall (x, x^*) \in X \times X^*. \tag{9}$$

Proposition 2.1. *Let $f : X \to (-\infty, \infty]$ be proper convex lower semicontinuous and $(x, x^*) \in X \times X^*$. Then, $x^* \in \partial f(x) \iff f(x) + f^*(x^*) = \langle x, x^* \rangle$.*

Proof. The proof is quite straightforward and can be found in [6]. □

An important example considered in this paper is given by $f(x) = p^{-1}\|x\|^p$, where $p \in (1, \infty)$. In this particular case, the following result can be found in [6].

Proposition 2.2. *Let $p \in (1, \infty)$ and $f : X \ni x \mapsto p^{-1}\|x\|^p \in \mathbb{R}$. Then,*

$$f^* : X^* \to \mathbb{R}, \ x^* \mapsto q^{-1}\|x^*\|^q, \ \ \textit{where} \ \ p + q = pq.$$

For $p \in (1, \infty)$, the duality mapping $J_p : X \to 2^{X^*}$ is defined by

$$J_p := \partial p^{-1}\| \cdot \|^p.$$

From the above proposition, we conclude that

$$x^* \in J_p(x) \iff p^{-1}\|x\|^p + q^{-1}\|x^*\|^q = \langle x, x^*\rangle, \qquad p + q = pq.$$

It follows from the above identity that $J_p(0) = \{0\}$. On the other hand, when $x \neq 0$, $J_p(x)$ may not be singleton.

Proposition 2.3. *Let X and the duality mapping J_p be defined as above. The following identities hold:*

$$\begin{aligned} J_p(x) &= \{x^* \in X^* : \|x^*\| = \|x\|^{p-1} \text{ and } \langle x, x^*\rangle = \|x\|\|x^*\|\} \\ &= \{x^* \in X^* : \|x^*\| = \|x\|^{p-1} \text{ and } \langle x, x^*\rangle = \|x\|^p\} \\ &= \{x^* \in X^* : \|x^*\| = \|x\|^{p-1} \text{ and } \langle x, x^*\rangle = \|x^*\|^q\}. \end{aligned}$$

Moreover, $J_p(x) \neq \emptyset$ for all $x \in X$.

Proof. See [6] or [23, section 2]. □

Since $f(x) = p^{-1}\|x\|^p$ is a continuous convex functions, $J_p(x)$ is a singleton at $x \in X$ iff f is Gâteaux differentiable at x [5, corollary 4.2.5]. This motivates us to consider X a *smooth* Banach space, i.e. a Banach space having a Gâteaux differentiable norm $\|\cdot\|_X$ on $X\backslash\{0\}$. As already observed, $J_p(0) = \{0\}$ in any Banach space. In particular, in a smooth Banach space, $f(x) = p^{-1}\|x\|^p$ is Gâteaux differentiable everywhere.

The next theorem describes a coercivity result related to geometrical properties of uniformly smooth Banach spaces. For details on the proof (as well as the precise definition of the constant G_q) we refer the reader to [23, section 2.1] or [26].

Theorem 2.4. *Let X be uniformly convex, $q \in (1, \infty)$ and $\rho_{X^*}(\cdot)$ the smoothness modulus of X^* [6]. There exists a positive constant G_q such that the function*

$$\tilde{\sigma}(x^*, y^*) := qG_q \int_0^1 (\|x^* - ty^*\| \vee \|x^*\|)^q t^{-1} \rho_{X^*}\big(t\|y^*\|/2(\|x^* - ty^*\| \vee \|x^*\|)\big)\, \mathrm{d}t$$

satisfies[2]

$$\|x^*\|^q - q\langle J_q(x^*), y^*\rangle + \tilde{\sigma}_q(x^*, y^*) \geqslant \|x^* - y^*\|^q \qquad \forall\, x^*,\ y^* \in X^*.$$

2.2. Bregman distances

Let $f : X \to (-\infty, \infty]$ be a proper, convex and lower semicontinuous function which is Gâteaux differentiable on $\mathrm{int}(\mathrm{dom}(f))$. Moreover, denote by f' the Gâteaux derivative of f. The *Bregman distance* induced by f is defined as $D_f : \mathrm{dom}(f) \times \mathrm{int}(\mathrm{dom}(f)) \to \mathbb{R}$

$$D_f(y, x) = f(y) - (f(x) + \langle f'(x), y - x\rangle).$$

The following proposition is a useful characterization of Bregman distances using Fenchel conjugate functions.

Proposition 2.5. *Let $f : X \to (-\infty, \infty]$ be a proper lower semicontinuous convex function which is Gâteaux differentiable on* $\mathrm{int}(\mathrm{dom}(f))$. *Then,*

$$D_f(y, x) = f(y) + f^*(f'(x)) - \langle f'(x), y\rangle \quad \forall (y, x) \in \mathrm{dom}(f) \times \mathrm{int}(\mathrm{dom}(f)).$$

[2] We adopt the notation $a \vee b := \max\{a, b\}$, $a \wedge b := \min\{a, b\}$, for $a, b \in \mathbb{R}$.

Based on this proposition, we derive the following two corollaries, which are used in a forthcoming convergence analysis. The proofs of these corollaries follow as a particular case $f(x) = p^{-1}\|x\|^p$ $(p \in (1, \infty))$. We use the notation D_p instead of D_f.

Corollary 2.6. *Let X be a smooth Banach space. Then,* $J_p : X \to X^*$ *is a single-valued mapping for which* $D_p : X \times X \to \mathbb{R}$ *satisfies*

$$D_p(y, x) = p^{-1}\|y\|^p + q^{-1}\|J_p(x)\|^q - \langle y, J_p(x)\rangle = p^{-1}\|y\|^p + q^{-1}\|x\|^p - \langle y, J_p(x)\rangle.$$

Corollary 2.7. *Let X be a smooth Banach space. Then,* $J_p : X \to X^*$ *is a single-valued mapping for which* $D_p : X \times X \to \mathbb{R}$ *satisfies*

$$D_p(y, x) = q^{-1}\,(\|x\|^p - \|y\|^p) + \langle J_p(y) - J_p(x), y\rangle.$$

3. An LK algorithm in Banach spaces

In this section, we introduce an algorithm for solving the system of nonlinear ill-posed equations (2) with data satisfying (1). In the rest of the paper, we assume the Banach space X to be uniformly convex and smooth, e.g., L_p spaces for $p \in (1, \infty)$.[3] These assumptions are crucial for the analysis derived in this section as well as in the forthcoming one.

We denote by

$$\mathcal{B}^1_p(x, r) = \{y \in X : D_p(x, y) \leqslant r\}, \quad \mathcal{B}^2_p(x, r) = \{y \in X : D_p(y, x) \leqslant r\},$$

the balls of radius $r > 0$ with respect to the Bregman distance $D_p(\cdot, \cdot)$.

A solution of (2) is any $\bar{x} \in D$ satisfying simultaneously the operator equation in (2), while a *minimum-norm solution* of (2) in S ($S \subset X$) is any solution $x^\dagger \in S$ satisfying

$$\|x^\dagger\| = \min\{\|x\| : x \in S \text{ is a solution of} (2)\}.$$

Assumption 3.1. Let $p, q, r \in (1, \infty)$ be given with $p + q = pq$. The following assumptions will be required in the forthcoming analysis.

(A0) Each operator F_i is of class C^1 in D. Moreover, the system of operator equations (2) has a solution $\bar{x} \in X$ satisfying $x_0 \in \mathcal{B}^1_p(\bar{x}, \bar{\rho}) \subset D$, for some $\bar{\rho} > 0$. Furthermore, we require $D_p(\bar{x}, x_0) \leqslant p^{-1}\|\bar{x}\|^p$. The element x_0 will be used as an initial guess of the LK algorithm.

(A1) The family $\{F_i\}_{1\leqslant i\leqslant m}$ satisfies the tangential cone condition in $\mathcal{B}^1_p(\bar{x}, \bar{\rho})$, i.e. there exists $\eta \in (0, 1)$, such that

$$\|F_i(y) - F_i(x) - F_i'(x)(y - x)\| \;\leqslant\; \eta\|F_i(y) - F_i(x)\|,$$

for all $x, y \in \mathcal{B}^1_p(\bar{x}, \bar{\rho})$, $i = 1, \dots, m$.

(A2) The family $\{F_i\}_{1\leqslant i\leqslant m}$ satisfies the tangential cone condition in $\mathcal{B}^2_p(x_0, \rho_0) \subset D$ for some $\rho_0 > 0$, i.e. there exists $\eta \in (0, 1)$, such that

$$\|F_i(y) - F_i(x) - F_i'(x)(y - x)\| \;\leqslant\; \eta\|F_i(y) - F_i(x)\|,$$

for all $x, y \in \mathcal{B}^2_p(x_0, \rho_0)$, $i = 1, \dots, m$.

(A3) For every $x \in \mathcal{B}^1_p(\bar{x}, \bar{\rho})$, we have $\|F_i'(x)\| \leqslant 1$, $i = 1, 2, \dots, m$.

In what follows, we formulate our *LK* algorithm for approximating a solution of (2), with data given as in (1).

Algorithm 3.1. Under assumptions (A0) and (A1), choose $c \in (0, 1)$ and $\tau \in (0, \infty)$, such that $\beta := \eta + \tau^{-1}(1 + \eta) < 1$.

[3] Note that L_1 and L_∞ are not uniformly convex [23, example 2.2].

- *Step 0.* Set $n = 0$ and take $x_0 \neq 0$ satisfying (A0).
- *Step 1.* Set $i_n = n(\mathrm{mod}\, m) + 1$ and evaluate the residual $R_n = F_{i_n}(x_n) - y^\delta_{i_n}$;
- *Step 2.* IF ($\|R_n\| \leqslant \tau\delta_{i_n}$) THEN

  $\mu_n := 0;$

  ELSE

  Find $\tau_n \in (0, 1]$ solving the equation

$$\rho_{X^*}(\tau_n)\,\tau_n^{-1} = \left(c(1-\beta)\,\|R_n\|\left[2^q\,G_q(1 \vee \|F'_{i_n}(x_n)\|)\,\|x_n\|\right]^{-1}\right) \wedge \rho_{X^*}(1); \tag{10}$$

  $\mu_n := \tau_n \|x_n\|^{p-1} \,/\, \left[(1 \vee \|F'_{i_n}(x_n)\|)\,\|R_n\|^{r-1}\right];$

  ENDIF

  $x^*_n := J_p(x_n) - \mu_n F'_{i_n}(x_n)^* J_r(F_{i_n}(x_n) - y^\delta_{i_n});$

$$x_{n+1} = J_q(x^*_n); \tag{11}$$

- *Step 3.* IF ($i_n = m$) AND ($x_{n+1} = x_n = \cdots = x_{n-(m-1)}$) THEN STOP;
- *Step 4.* SET $n = n + 1$; GO TO Step 1.

The next remark guarantees that the above algorithm is well defined.

Remark 3.1. It is worth noting that a solution $\tau_n \in (0, 1]$ of equation (10) can always be found. Indeed, since X^* is uniformly smooth, the function $(0, \infty) \ni \tau \mapsto \rho_{X^*}(\tau)/\tau \in (0, 1]$ is continuous and satisfies $\lim_{\tau\to 0} \rho_{X^*}(\tau)/\tau = 0$ (see, e.g., [23, definition 2.1] or [6]). For each $n \in \mathbb{N}$, define

$$\lambda_n := \left(c(1-\beta)\,\|R_n\|\,[2^q\,G_q(1 \vee \|F'_{i_n}(x_n)\|)\,\|x_n\|]^{-1}\right) \wedge \rho_{X^*}(1). \tag{12}$$

It follows from [23, section 2.1] that $\rho_{X^*}(1) \leqslant 1$. Therefore, $\lambda_n \in (0, 1]$, $n \in \mathbb{N}$, and we can find $\sigma_n \in (0, 1]$ satisfying $\rho_{X^*}(\sigma_n)/\sigma_n < \lambda_n \leqslant \rho_{X^*}(1)$. Finally, the mean value theorem guarantees the existence of corresponding $\tau_n \in (0, 1]$, such that $\lambda_n = \rho_{X^*}(\tau_n)/\tau_n$, $n \in \mathbb{N}$.

Algorithm 3.1 should be stopped at the smallest iteration index $\hat{n} \in \mathbb{N}$ of the form $\hat{n} = \hat{\ell}m + (m-1)$, $\hat{\ell} \in \mathbb{N}$, which satisfies

$$\|F_{i_n}(x_n) - y^\delta_{i_n}\| \leqslant \tau\delta_{i_n}, \qquad n = \hat{\ell}m, \ldots, \hat{\ell}m + (m-1) \tag{13}$$

(note that $i_{\hat{n}} = m$). In this case, $x_{\hat{n}} = x_{\hat{n}-1} = \cdots = x_{\hat{n}-(m-1)}$ within the $\hat{\ell}$th cycle. The next result guarantees monotonicity of the iteration error (w.r.t. the Bregman distance D_p) until the discrepancy principle in (13) is reached.

Lemma 3.2 (Monotonicity)**.** *Let assumptions (A0) and (A1) be satisfied and* (x_n) *be a sequence generated by algorithm 3.1. Then,*

$$D_p(\bar{x}, x_{n+1}) \leqslant D_p(\bar{x}, x_n), \qquad n = 0, 1, \ldots, \hat{n},$$

where $\hat{n} = \hat{\ell}m + (m-1)$ *is defined by (13). From the above inequality, it follows that* $x_n \in \mathcal{B}^1_p(\bar{x}, \bar{\rho}) \subset D$, $n = 0, 1, \ldots, \hat{n}$.

Proof. Let $0 \leqslant n \leqslant \hat{n}$ and assume that x_n is a nonzero vector satisfying $x_n \in \mathcal{B}^1_p(\bar{x}, \bar{\rho})$. From assumption (A0), $x_n \in D$ follows.

If $\|R_n\| \leqslant \tau\delta_{i_n}$, then $x_{n+1} = x_n$ and the lemma follows trivially. Otherwise, it follows from corollary 2.6 that

$$D_p(\bar{x}, x_{n+1}) = p^{-1}\|\bar{x}\|^p + q^{-1}\|J_p(x_{n+1})\|^q - \langle \bar{x}, J_p(x_{n+1})\rangle. \tag{14}$$

Since $R_n = F_{i_n}(x_n) - y^\delta_{i_n}$, we conclude from (11) and $J_q = (J_p)^{-1}$ [6] that

$$J_p(x_{n+1}) = J_p(x_n) - \mu_n F'_{i_n}(x_n)^* J_r(R_n).$$

Thus, it follows from theorem 2.4 that

$$\begin{aligned}\|J_p(x_{n+1})\|^q &= \|J_p(x_n) - \mu_n F'_{i_n}(x_n)^* J_r(R_n)\|^q \\ &\leqslant \|J_p(x_n)\|^q - q\mu_n \langle J_q(J_p(x_n)), F'_{i_n}(x_n)^* J_r(R_n)\rangle + \tilde\sigma_q\big(J_p(x_n), \mu_n F'_{i_n}(x_n)^* J_r(R_n)\big) \\ &= \|J_p(x_n)\|^q - q\mu_n \langle x_n, F'_{i_n}(x_n)^* J_r(R_n)\rangle + \tilde\sigma_q(J_p(x_n), \mu_n F'_{i_n}(x_n)^* J_r(R_n)). \qquad (15)\end{aligned}$$

In order to estimate the last term on the right-hand side of (15), note that for all $t \in [0, 1]$ the inequality

$$\begin{aligned}\|J_p(x_n) - t\mu_n F'_{i_n}(x_n)^* J_r(R_n)\| \vee \|J_p(x_n)\| &\leqslant \|x_n\|^{p-1} + \mu_n (1 \vee \|F'_{i_n}(x_n)\|)\|R_n\|^{r-1} \\ &\leqslant (1+\tau_n)\|x_n\|^{p-1} \ \leqslant \ 2\|x_n\|^{p-1}\end{aligned}$$

holds true (to obtain the first inequality we used proposition 2.3).

Moreover, $\|J_p(x_n) - t\mu_n F'_{i_n}(x_n)^* J_r(R_n)\| \vee \|J_p(x_n)\| \geqslant \|J_p(x_n)\| = \|x_n\|^{p-1}$. From the last two inequalities together with the monotonicity of $\rho_{X^*}(t)/t$, it follows that (see theorem 2.4)

$$\tilde\sigma_q(J_p(x_n), \mu_n F'_{i_n}(x_n)^* J_r(R_n)) \ \leqslant \ qG_q \int_0^1 \frac{(2\|x_n\|^{p-1})^q}{t} \rho_{X^*}\left(\frac{t\mu_n (1 \vee \|F'_{i_n}(x_n)\|)\|R_n\|^{r-1}}{\|x_n\|^{p-1}}\right) \mathrm{d}t.$$

Consequently,

$$\begin{aligned}\tilde\sigma_q(J_p(x_n), \mu_n F'_{i_n}(x_n)^* J_r(R_n)) &\leqslant 2^q\, q\, G_q\, \|x_n\|^p \int_0^1 \rho_{X^*}(t\tau_n)/t \,\mathrm{d}t \\ &= 2^q\, q\, G_q\, \|x_n\|^p \int_0^{\tau_n} \rho_{X^*}(t)/t \,\mathrm{d}t \\ &\leqslant 2^q\, q\, G_q\, \rho_{X^*}(\tau_n)/\tau_n \|x_n\|^p \int_0^{\tau_n} \mathrm{d}t \\ &= 2^q\, q\, G_q\, \rho_{X^*}(\tau_n)\|x_n\|^p. \qquad (16)\end{aligned}$$

Now, substituting (16) into (15), we obtain the estimate

$$\|J_p(x_{n+1})\|^q \ \leqslant \ \|J_p(x_n)\|^q - q\mu_n \langle x_n, F'_{i_n}(x_n)^* J_r(R_n)\rangle + q2^q G_q \rho_{X^*}(\tau_n)\|x_n\|^p.$$

From this last inequality, corollary 2.6 and (14), we obtain

$$D_p(\bar x, x_{n+1}) \leqslant D_p(\bar x, x_n) - \mu_n \,\langle x_n - \bar x, F'_{i_n}(x_n)^* J_r(R_n)\rangle + 2^q\, G_q\, \rho_{X^*}(\tau_n)\|x_n\|^p. \qquad (17)$$

Next, we estimate the term $\langle x_n - \bar x, F'_{i_n}(x_n)^* J_r(R_n)\rangle$ in (17). Since $\bar x, x_n \in \mathcal{B}^1_p(\bar x, \bar\rho)$, it follows from (A1) and simple algebraic manipulations (including proposition 2.3) that

$$\begin{aligned}\langle \bar x - x_n, F'_{i_n}(x_n)^* J_r(R_n)\rangle &= \langle y_{i_n} - F_{i_n}(x_n) - F'_{i_n}(x_n)(\bar x - x_n), -J_r(R_n)\rangle - \langle \tilde R_n, J_r(R_n)\rangle \\ &\leqslant \eta \|\tilde R_n\| \|J_r(R_n)\| - \langle R_n, J_r(R_n)\rangle + \langle y_{i_n} - y^\delta_{i_n}, J_r(R_n)\rangle \\ &\leqslant \eta \left(\|R_n\| + \delta_{i_n}\right) \|R_n\|^{r-1} - \|R_n\|^r + \delta_{i_n}\|R_n\|^{r-1} \\ &= \left(\eta(\|R_n\| + \delta_{i_n}\right) + \delta_{i_n})\|R_n\|^{r-1} - \|R_n\|^r \\ &\leqslant [(\eta + \tau^{-1}(1+\eta)]\|R_n\|)\|R_n\|^{r-1} - \|R_n\|^r \\ &= -(1-\beta)\|R_n\|^r,\end{aligned}$$

where $\tilde R_n := F_{i_n}(x_n) - y_{i_n}$ and $\beta > 0$ is defined as in algorithm 3.1. Substituting this last inequality into (17) yields

$$D_p(\bar x, x_{n+1}) \ \leqslant \ D_p(\bar x, x_n) - (1-\beta)\mu_n \|R_n\|^r + 2^q\, G_q\, \rho_{X^*}(\tau_n)\, \|x_n\|^p. \qquad (18)$$

Moreover, from the explicit formula for μ_n and τ_n (see algorithm 3.1), we can estimate the last two terms on the right-hand side of (18) by

$$
\begin{aligned}
-(1-\beta)\mu_n\|R_n\|^r + 2^q G_q\,\rho_{X^*}(\tau_n)\|x_n\|^p &= -(1-\beta)\frac{\tau_n\|x_n\|^{p-1}\|R_n\|}{1\vee\|F'_{i_n}(x_n)\|} + 2^q G_q\,\rho_{X^*}(\tau_n)\|x_n\|^p \\
&= -(1-\beta)\frac{\tau_n\|x_n\|^{p-1}\|R_n\|}{1\vee\|F'_{i_n}(x_n)\|}\left(1 - \frac{2^q G_q(1\vee\|F'_{i_n}(x_n)\|)\|x_n\|}{(1-\beta)\|R_n\|}\,\frac{\rho_{X^*}(\tau_n)}{\tau_n}\right) \\
&\leqslant -(1-\beta)(1-c)\frac{\tau_n\|x_n\|^{p-1}\|R_n\|}{1\vee\|F'_{i_n}(x_n)\|}. \qquad (19)
\end{aligned}
$$

Finally, substituting (19) into (18), we obtain

$$D_p(\bar{x}, x_{n+1}) \leqslant D_p(\bar{x}, x_n) - (1-\beta)(1-c)\tau_n\|x_n\|^{p-1}\|R_n\|\,[1\vee\|F'_{i_n}(x_n)\|]^{-1}, \qquad (20)$$

concluding the proof. □

Remark 3.3. In the proof of lemma 3.2, we used the fact that the elements $x_n \in X$ generated by algorithm 3.1 are nonzero vectors. This can be verified by an inductive argument. Indeed, $x_0 \neq 0$ is chosen in algorithm 3.1. Assume $x_k \neq 0$, $k = 0, \ldots, n$. If $\|R_n\| \leqslant \tau\delta_{i_n}$, then $x_{n+1} = x_n$ is also a nonzero vector. Otherwise, $\|R_n\| > \tau\delta_{i_n} > 0$ and it follows from (20) that $D_p(\bar{x}, x_{n+1}) < D_p(\bar{x}, x_n) \leqslant \cdots \leqslant D_p(\bar{x}, x_0) \leqslant p^{-1}\|\bar{x}\|^p$ (the last inequality follows from the choice of x_0 in (A0)). If x_{n+1} were the null vector, we would have $p^{-1}\|\bar{x}\|^p = D_p(\bar{x}, 0) < D_p(\bar{x}, x_n) \leqslant p^{-1}\|\bar{x}\|^p$ (the identity follows from corollary 2.6), which is clearly a contradiction. Therefore, x_n is a nonzero vector, for $n = 0, 1, \ldots, \hat{n}$.

In the case of exact data ($\delta_i = 0$), we have $x_n \neq 0$, $n \in \mathbb{N}$.

The next lemma guarantees that, in the case of noisy data, algorithm 3.1 is stopped after a finite number of cycles, i.e. $\hat{n} < \infty$, in (13).

Lemma 3.4. *Let assumptions (A0), (A1) and (A3) be satisfied and (x_n) be a sequence generated by algorithm 3.1. Then,*

$$\sum_{n\in\hat{\Sigma}} \tau_n\|x_n\|^{p-1}\|R_n\| \leqslant (1-\beta)^{-1}(1-c)^{-1}D_p(\bar{x}, x_0), \qquad (21)$$

where $\hat{\Sigma} := \{n \in \{0, 1, \ldots, \hat{n}-1\} : \|R_n\| > \tau\delta_{i_n}\}$. Additionally, we have the following.

(i) In the noisy data case, $\min\{\delta_i\}_{1\leqslant i\leqslant m} > 0$, *algorithm 3.1 is stopped after finitely many steps.*

(ii) In the noise-free case, we have $\lim_{n\to\infty}\|R_n\| = 0$.

Proof. Given $n \in \hat{\Sigma}$, it follows from (20) and (A3) that

$$(1-\beta)(1-c)\tau_n\|x_n\|^{p-1}\|R_n\| \leqslant D_p(\bar{x}, x_n) - D_p(\bar{x}, x_{n+1}). \qquad (22)$$

Moreover, if $n \notin \hat{\Sigma}$ and $n < \hat{n}$, we have $0 \leqslant D_p(\bar{x}, x_n) - D_p(\bar{x}, x_{n+1})$. Inequality (21) follows now from a telescopic sum argument using the above inequalities.

Add. (i). Assume by contradiction that algorithm 3.1 is never stopped by the discrepancy principle. Therefore, $\hat{n}$ defined in (13) is not finite. Consequently, $\hat{\Sigma}$ is an infinite set (at least one step is performed in each iteration cycle).

Since $(D_p(\bar{x}, x_n))_{n\in\hat{\Sigma}}$ is bounded, it follows that $(\|x_n\|)_{n\in\hat{\Sigma}}$ is bounded [23, theorem 2.12(b)]. Therefore, the sequence $(\lambda_n)_{n\in\hat{\Sigma}}$ in (12), is bounded away from zero (see (10) and remark 3.1) and from what follows that $(\tau_n)_{n\in\hat{\Sigma}}$ is bounded away from zero as well. From this fact and (21) we obtain

$$\sum_{n\in\hat{\Sigma}} \|x_n\|^{p-1} < \infty.$$

Consequently, $(x_n)_{n\in\hat{\Sigma}}$ converges to zero in X and, arguing with the continuity of $D_p(\bar{x},\cdot)$ [23, theorem 2.12(c)] or [6]), we conclude

$$p^{-1}\|\bar{x}\|^p = D_p(\bar{x},0) = \lim_{n\in\hat{\Sigma}} D_p(\bar{x},x_n) \leqslant D_p(\bar{x},x_{n'+1}) < D_p(\bar{x},x_{n'}) \leqslant p^{-1}\|\bar{x}\|^p,$$

where $n' \in \mathbb{N}$ is an arbitrary element of $\hat{\Sigma}$ (note that (20) holds with strict inequality for all $n' \in \hat{\Sigma}$). This is clearly a contradiction. Thus, $\hat{n}$ must be finite.

Add (ii). Note that in the noise-free case we have $\delta_i = 0$, $i = 1, 2, \ldots, m$. In this particular case, (22) holds for all $n \in \mathbb{N}$. Consequently,

$$\sum_{n\in\mathbb{N}} \tau_n\|x_n\|^{p-1}\|R_n\| \leqslant (1-\beta)^{-1}(1-c)^{-1}D_p(\bar{x},x_0).$$

Assume the existence of $\varepsilon > 0$ such that the inequality $\|R_{n_k}\| > \varepsilon$ holds true for some subsequence and define $\hat{\Sigma} := \{n_k;\ k \in \mathbb{N}\}$. Using the same reasoning as in the proof of the second assertion, we arrive at a contradiction, concluding the proof. □

4. Convergence analysis

In this section, the main results of the manuscript are presented. A convergence analysis of the proposed method is given, and stability results are derived. We start the presentation discussing a result related to the existence of minimum-norm solutions.

Lemma 4.1. *Assume the continuous Fréchet differentiability of the operators F_i in D. Moreover, assume that (A2) is satisfied and also that problem (2) is solvable in $\mathcal{B}^2_p(x_0,\rho_0)$, where $x_0 \in X$ and $\rho_0 > 0$ is chosen as in (A2).*

(1) There exists a unique minimum-norm solution $x^\dagger$ of (2) in $\mathcal{B}^2_p(x_0,\rho_0)$.

(2) If $x^\dagger \in \mathrm{int}(\mathcal{B}^2_p(x_0,\rho_0))$, it can be characterized as the solution of (2) in $\mathcal{B}^2_p(x_0,\rho_0)$ satisfying the condition

$$J_p(x^\dagger) \in \mathcal{N}(F_i'(x^\dagger))^\perp, \qquad i = 1, 2, \ldots, m. \tag{23}$$

(Here, $A^\perp \subset X^*$ denotes the annihilator of $A \subset X$, while $\mathcal{N}(\cdot)$ represents the null-space of a linear operator.)

Proof. As an immediate consequence of (A2), we obtain [12, proposition 2.1]

$$F_i(z) = F_i(x) \iff z - x \in \mathcal{N}(F_i'(x)), \qquad i = 1, 2, \ldots m, \tag{24}$$

for $x, z \in \mathcal{B}^2_p(x_0,\rho_0)$. Next, we define for each $x \in \mathcal{B}^2_p(x_0,\rho_0)$ the set $M_x := \{z \in \mathcal{B}^2_p(x_0,\rho_0) : F_i(z) = F_i(x), i = 1, 2, \ldots, m\}$. Note that $M_x \neq \emptyset$, for all $x \in \mathcal{B}^2_p(x_0,\rho_0)$. Moreover, it follows from (24) that

$$M_x = \bigcap_{i=1}^{m} \left(x + \mathcal{N}(F_i'(x))\right) \cap \mathcal{B}^2_p(x_0,\rho_0). \tag{25}$$

Since $D_p(\cdot, x_0)$ is continuous (see corollary 2.6) and $\mathcal{B}^2_p(x_0,\rho_0)$ is convex (by definition), it follows from (25) that M_x is nonempty, closed and convex, for all $x \in \mathcal{B}^2_p(x_0,\rho_0)$. Therefore, there exists a unique $x^\dagger \in X$ corresponding to the projection of 0 on $M_{\bar{x}}$, where $\bar{x}$ is a solution of (2) in $\mathcal{B}^2_p(x_0, \rho_0)$ [6]. This proves the first assertion.

Add (ii). From the definition of $x^\dagger$ and $M_{\bar{x}} = M_{x^\dagger}$, we conclude that [23, theorem 2.5 (h)]

$$\langle J_p(x^\dagger), x^\dagger\rangle \leqslant \langle J_p(x^\dagger), y\rangle \quad \forall y \in M_{x^\dagger}. \tag{26}$$

From the assumption $x^\dagger \in \mathrm{int}\big(\mathcal{B}^2_p(x_0, \rho_0)\big)$, it follows that given $h \in \cap_{i=1}^m \mathcal{N}(F_i'(x^\dagger))$, there exists $\varepsilon_0 > 0$ such that

$$x^\dagger + \varepsilon h, x^\dagger - \varepsilon h \in M_{x^\dagger} \quad \forall \varepsilon \in [0, \varepsilon_0). \tag{27}$$

Thus, (23) follows from (26), (27) in a straightforward way. In order to prove uniqueness, let $\tilde{x}$ be any solution of (2) in $\mathcal{B}^2_p(x_0, \rho_0)$ satisfying

$$J_p(\tilde{x}) \in \mathcal{N}\big(F_i'(\tilde{x})\big)^\perp, \qquad i = 1, 2, \ldots, m. \tag{28}$$

Let $i \in \{1, 2, \ldots, m\}$. We claim that

$$\mathcal{N}\big(F_i'(x^\dagger)\big) \subset \mathcal{N}\big(F_i'(\tilde{x})\big). \tag{29}$$

Indeed, let $h \in \mathcal{N}(F_i'(x^\dagger))$ and set $x_\theta = (1-\theta)x^\dagger + \theta\tilde{x}$, with $\theta \in \mathbb{R}$. Since $x^\dagger \in \mathrm{int}\big(\mathcal{B}^2_p(x_0, \rho_0)\big)$, we obtain a $\theta_0 > 0$ such that $x_\theta \in \mathrm{int}\big(\mathcal{B}^2_p(x_0, \rho_0)\big)$, for all $\theta \in [0, \theta_0)$. Take $\theta \in (0, \theta_0)$ and define $x_{\theta,\mu} = x_\theta + \mu h$, for $\mu \in \mathbb{R}$. Using the same reasoning, we obtain $\mu_0 > 0$ such that $x_{\theta,\mu} \in B^2_p(x_0, \rho_0)$ $\forall \mu \in [0, \mu_0)$.

For a fixed $\mu \in (0, \mu_0)$, note that $x_{\theta,\mu} - x^\dagger = \theta(\tilde{x} - x^\dagger) + \mu h$. Using (24), we obtain $\tilde{x} - x^\dagger \in \mathcal{N}(F_i'(x^\dagger))$, and consequently, $x_{\theta,\mu} - x^\dagger \in \mathcal{N}(F_i'(x^\dagger))$. From (24), it follows that $F(x_{\theta,\mu}) = F(x^\dagger)$, and consequently, $F(x_{\theta,\mu}) = F(\tilde{x})$. Applying the same reasoning as above (based on (24)), we conclude that $x_{\theta,\mu} - \tilde{x} \in \mathcal{N}(F_i'(\tilde{x}))$.

Since $x_{\theta,\mu} - \tilde{x} = (1-\theta)(x^\dagger - \tilde{x}) + \mu h$ and $x^\dagger - \tilde{x} \in \mathcal{N}(F_i'(\tilde{x}))$, it follows $h \in \mathcal{N}(F_i'(\tilde{x}))$, completing the proof of our claim.

Combining (28) and (29), we obtain $J_p(\tilde{x}) \in \mathcal{N}(F_i'(x^\dagger))^\perp$. Consequently, $J_p(x^\dagger) - J_p(\tilde{x}) \in \mathcal{N}(F_i'(x^\dagger))^\perp$. Since $x^\dagger - \tilde{x} \in \mathcal{N}(F_i'(x^\dagger))$, we conclude that $\langle J_p(x^\dagger) - J_p(\tilde{x}), x^\dagger - \tilde{x}\rangle = 0$. Moreover, since J_p is strictly monotone [23, theorem 2.5(e)], we obtain $x^\dagger = \tilde{x}$. □

Theorem 4.2 (Convergence for exact data)**.** *Assume that $\delta_i = 0$, $i = 1, 2, \ldots, m$. Let the assumptions (A0), (A1), (A2) and (A3) be satisfied (for simplicity we assume $\bar{\rho} = \rho_0$). Then, any iteration (x_n) generated by algorithm 3.1 converges strongly to a solution of (2).*

Additionally, if $x^\dagger \in \mathrm{int}\big(\mathcal{B}^2_p(x_0, \rho_0)\big)$, $J_p(x_0) \in \mathcal{N}(F_i'(x^\dagger))^\perp$ and $\mathcal{N}(F_i'(x^\dagger)) \subset \mathcal{N}(F_i'(x))$, $x \in B^1_p(\bar{x}, \bar{\rho})$, $i = 1, 2, \ldots, m$, then (x_n) converges strongly to $x^\dagger$.

Proof. From lemma 3.2, it follows that $D_p(\bar{x}, x_n)$ is bounded and so $(\|x_n\|)$ is bounded. In particular, $(J_p(x_n))$ is also bounded. Define $\varepsilon_n = q^{-1}\|x_n\|^p - \langle \bar{x}, J_p(x_n)\rangle$, $n \in \mathbb{N}$. From lemma 3.2 and corollary 2.6, it follows that (ε_n) is bounded and monotone non-increasing. Thus, there exists $\varepsilon \in \mathbb{R}$, such that $\varepsilon_n \to \varepsilon$, as $n \to \infty$.

Let $m, n \in \mathbb{N}$, such that $m > n$. It follows from corollary 2.7 that

$$\begin{aligned} D_p(x_n, x_m) = q^{-1}\left(\|x_m\|^p - \|x_n\|^p\right) + \langle J_p(x_n) - J_p(x_m), x_n\rangle = (\varepsilon_m - \varepsilon_n) \\ + \langle J_p(x_n) - J_p(x_m), x_n - \bar{x}\rangle. \end{aligned}$$

The first term of the above identity converges to zero, as $m, n \to \infty$. Note that

$$
\begin{aligned}
|\langle J_p(x_n) - J_p(x_m), x_n - \bar{x}\rangle| &= \left| \left\langle \sum_{k=n}^{m-1} (J_p(x_{k+1}) - J_p(x_k)), x_k - \bar{x} \right\rangle \right| \\
&\overset{(11)}{=} \left| \left\langle \sum_{k=n}^{m-1} \mu_k F'_{i_k}(x_k)^* J_r(R_k), x_k - \bar{x} \right\rangle \right| \\
&\leqslant \sum_{k=n}^{m-1} \mu_k \|J_r(R_k)\| \|F'_{i_k}(x_k)(x_k - \bar{x})\|.
\end{aligned}
$$

Moreover, from (A1), we have

$$
\begin{aligned}
\|F'_{i_k}(x_k)(x_k - \bar{x})\| &\leqslant \|F_{i_k}(x_k) - F_{i_k}(\bar{x}) - F'_{i_k}(x_k)(x_k - \bar{x})\| + \|F_{i_k}(x_k) - F_{i_k}(\bar{x})\| \\
&\leqslant (1+\eta)\|R_k\|.
\end{aligned}
$$

Therefore, using (A3) and the definition of μ_k in algorithm 3.1, we can estimate

$$
\begin{aligned}
|\langle J_p(x_n) - J_p(x_m), x_n - \bar{x}\rangle| &\leqslant (1+\eta) \sum_{k=n}^{m-1} \mu_k \|R_k\|^{r-1} \|R_k\| \\
&= (1+\eta) \sum_{k=n}^{m-1} \frac{\tau_k \|x_k\|^{p-1} \|R_k\|^r}{(1 \vee \|F'_{i_k}(x_k)\|)\|R_k\|^{r-1}} \\
&\leqslant (1+\eta) \sum_{k=n}^{m-1} \tau_k \|x_k\|^{p-1} \|R_k\|.
\end{aligned}
$$

(Note that the last two sums are carried out only for the terms with $\mu_k \neq 0$.) Consequently, $\langle J_p(x_n) - J_p(x_m), x_n - \bar{x}\rangle$ converges to zero, from what follows $D_p(x_n, x_m) \to 0$, as m, $n \to \infty$. Therefore, we conclude that (x_n) is a Cauchy sequence, converging to some element $\tilde{x} \in X$ [23, theorem 2.12(b)]. Since $x_n \in B^1_p(\bar{x}, \bar{\rho}) \subset D$, for $n \in \mathbb{N}$, it follows that $\tilde{x} \in D$. Moreover, from the continuity of $D_p(\cdot, \tilde{x})$, we have $D_p(x_n, \tilde{x}) \to D_p(\tilde{x}, \tilde{x}) = 0$, proving that $\|x_n - \tilde{x}\| \to 0$.

Let $i \in \{1, 2, \ldots, m\}$ and $\varepsilon > 0$. Since F_i is continuous, we have $F_i(x_n) \to F_i(\tilde{x})$, $n \to \infty$. This fact, together with $R_n \to 0$, allows us to find $n_0 \in \mathbb{N}$, such that

$$
\|F_i(x_n) - F_i(\tilde{x})\| < \varepsilon/2, \quad \|F_{i_n}(x_n) - y_{i_n}\| < \varepsilon/2 \ \ \forall n \geqslant n_0.
$$

Let $\tilde{n} \geqslant n_0$ be such that $i_{\tilde{n}} = i$. Then, $\|F_i(\tilde{x}) - y_i\| \leqslant \|F_i(x_{\tilde{n}}) - F_i(\tilde{x})\| + \|F_{i_{\tilde{n}}}(x_{\tilde{n}}) - y_{i_{\tilde{n}}}\| < \varepsilon$. Thus, $F_i(\tilde{x}) = y_i$, proving that $\tilde{x}$ is a solution of (2).

For each $n \in \mathbb{N}$, it follows from (11) and the theorem assumption that

$$
J_p(x_n) - J_p(x_0) \in \bigcap_{k=0}^{n-1} \mathcal{N}\big(F'_{i_k}(x_k)\big)^\perp \subset \bigcap_{k=0}^{n-1} \mathcal{N}\big(F'_{i_k}(x^\dagger)\big)^\perp.
$$

Moreover, due to $J_p(x_0) \in \mathcal{N}(F'_i(x^\dagger))^\perp$, $i = 1, 2, \ldots, m$, we have $J_p(x_n) \in \bigcap_{j=1}^m \mathcal{N}(F'_j(x^\dagger))^\perp$, $n \geqslant m$. Then, $J_p(x_n) \in \mathcal{N}(F'_i(x^\dagger))^\perp$, for $n \geqslant m$. Since J_p is continuous and $x_n \to \tilde{x}$, we conclude that $J_p(\tilde{x}) \in \mathcal{N}(F'_i(x^\dagger))^\perp$. However, due to $\mathcal{N}(F'_i(\tilde{x})) = \mathcal{N}(F'_i(x^\dagger))$ (which follows from $F_i(\tilde{x}) = F_i(x^\dagger)$) we conclude that $J_p(\tilde{x}) \in \mathcal{N}(F'_i(\tilde{x}))^\perp$, proving that $\tilde{x} = x^\dagger$. □

In what follows, we prove a convergence result in the noisy data case. For simplicity of the presentation, we assume for the rest of this section that $\delta_1 = \delta_2 = \cdots = \delta_m = \delta > 0$. Moreover, we denote by (x_n) and (x_n^δ) the iterations generated by algorithm 3.1 with exact data and noisy data, respectively.

Theorem 4.3 (Semi-convergence). *Let Y be a uniformly smooth Banach space and assumptions (A0), (A1), (A2) and (A3) be satisfied (for simplicity we assume $\bar{\rho} = \rho_0$). Moreover, let $(\delta_k > 0)_{k \in \mathbb{N}}$ be a sequence satisfying $\delta_k \to 0$ and $y_i^k \in Y$ be corresponding noisy data satisfying $\|y_i^k - y_i\| \leqslant \delta_k$, $i = 1, \ldots, m$, and $k \in \mathbb{N}$.*

If (for each $k \in \mathbb{N}$) the iterations $(x_n^{\delta_k})$ are stopped according to the discrepancy principle (13) at $\hat{n}_k = \hat{n}(\delta_k)$, then $(x_{\hat{n}_k}^{\delta_k})$ converges (strongly) to a solution $\tilde{x} \in B_p^1(\bar{x}, \bar{\rho})$ of (2) as $k \to \infty$.

Additionally, if $x^\dagger \in \operatorname{int}(\mathcal{B}_p^2(x_0, \rho_0))$, $J_p(x_0) \in \mathcal{N}(F_i'(x^\dagger))^\perp$ and $\mathcal{N}(F_i'(x^\dagger)) \subset \mathcal{N}(F_i'(x))$, $x \in B_p^1(\bar{x}, \bar{\rho})$, $i = 1, 2, \ldots, m$, then $(x_{\hat{n}_k}^{\delta_k})$ converges (strongly) to $x^\dagger$ as $k \to \infty$.

Proof. For each $k \in \mathbb{N}$, we can write $\hat{n}_k$ in (13) in the form $\hat{\ell}_k m + (m-1)$. Thus, $x_{\hat{n}_k}^{\delta_k} = x_{\hat{n}_k - 1}^{\delta_k} = \cdots = x_{\hat{n}_k - (m-1)}^{\delta_k}$ and

$$\left\| F_{i_n}\left(x_n^{\delta_k}\right) - y_{i_n}^k \right\| \leqslant \tau \delta_k, \quad n = \hat{\ell}_k m, \ldots, \hat{\ell}_k m + (m-1).$$

Since $i_n = 1, 2, \ldots, m$ as $n = \hat{\ell}_k m, \ldots, \hat{\ell}_k m + (m-1)$, it follows that

$$\left\| F_i\left(x_{\hat{n}_k}^{\delta_k}\right) - y_i^k \right\| \leqslant \tau\, \delta_k, \quad i = 1, 2, \ldots, m. \tag{30}$$

At this point, we must consider two cases separately.

*Case.*1. The sequence $(\hat{n}_k) \in \mathbb{N}$ is bounded.

If this is the case, we can assume the existence of $\hat{n} \in \mathbb{N}$ such that $\hat{n}_k = \hat{n}$, for all $k \in \mathbb{N}$. Note that, for each $k \in \mathbb{N}$, the sequence element $x_{\hat{n}}^{\delta_k}$ depends continuously on the corresponding data $\left(y_i^k\right)_{i=1}^m$ (this is the point where the uniform smoothness of Y is required). Therefore, it follows that

$$x_{\hat{n}}^{\delta_k} \to x_{\hat{n}}, \qquad F_i\left(x_{\hat{n}}^{\delta_k}\right) \to F_i(x_{\hat{n}}), \quad k \to \infty, \tag{31}$$

for each $i = 1, 2, \ldots, m$. Since each operator F_i is continuous, taking limit as $k \to \infty$ in (30) gives $F_i(x_{\hat{n}}) = y_i$, $i = 1, 2, \ldots, m$, which proves that $\tilde{x} := x_{\hat{n}}$ is a solution of (2).

Case 2. The sequence $(\hat{n}_k) \in \mathbb{N}$ is unbounded.

We can assume that $\hat{n}_k \to \infty$, monotonically. Due to theorem 4.2, $(x_{\hat{n}_k})$ converges to some solution $\tilde{x} \in B_p^1(\bar{x}, \bar{\rho})$ of (2). Therefore, $D_p(\tilde{x}, x_{\hat{n}_k}) \to 0$. Thus, given $\varepsilon > 0$, there exists $N \in \mathbb{N}$, such that

$$D_p(\tilde{x}, x_{\hat{n}_k}) < \varepsilon/2 \qquad \forall \hat{n}_k \geqslant N.$$

Since $x_N^{\delta_k} \to x_N$ as $k \to \infty$, and $D_p(\tilde{x}, \cdot)$ is continuous, there exists $\tilde{k} \in \mathbb{N}$, such that

$$\left| D_p\left(\tilde{x}, x_N^{\delta_k}\right) - D_p(\tilde{x}, x_N) \right| < \varepsilon/2 \qquad \forall k \geqslant \tilde{k}.$$

Consequently,

$$D_p(\tilde{x}, x_N^{\delta_k}) = D_p(\tilde{x}, x_N) + D_p\left(\tilde{x}, x_N^{\delta_k}\right) - D_p(\tilde{x}, x_N) < \varepsilon \qquad \forall k \geqslant \tilde{k}.$$

Since $D_p(\tilde{x}, x_{\hat{n}_k}^{\delta_k}) \leqslant D_p(\tilde{x}, x_N)$, for all $\hat{n}_k > N$, it follows that $D_p(\tilde{x}, x_{\hat{n}_k}^{\delta_k}) < \varepsilon$ for k large enough. Therefore, due to [23, theorem 2.12(d)] or [6], we conclude that $(x_{\hat{n}_k}^{\delta_k})$ converges to $\tilde{x}$.

To prove the last assertion, it is enough to observe that, due to the extra assumption, $\tilde{x} = x^\dagger$ must hold. □

5. Conclusions and future work

In this manuscript, we proposed an LK-type iteration for regularizing systems of nonlinear ill-posed operator equations in Banach spaces. We extended the results in [23], which considered the case of a single linear operator equation and obtained convergence and stability results for the Landweber iteration. Our results also extend the one obtained in [18], where nonlinear operator equations are considered in Banach spaces, but under the stronger assumption that X is p-convex.

One future perspective is to perform numerical experiments for the LKB method applied to parameter identification problems related to elliptic equations as the ones described in the last section of [18]. Another possible research direction is to extend the convergence analysis in this paper (in the framework of Banach spaces) to the SDK iteration [7], the LMK iteration [3] and ITK iteration [2]

Acknowledgments

The work of AL was partially supported by the Brazilian National Research Council CNPq, grant 303098/2009-0. MMA acknowledges support from the Brazilian National Research Council CNPq, grants 305414/2011-9, 479729/2011-5 and PRONEX-Optimization.

References

[1] Bakushinsky A B and Kokurin M Y 2004 Iterative methods for approximate solution of inverse problems (*Mathematics and Its Applications* vol 577) (Dordrecht: Springer)
[2] Baumeister J, De Cezaro A and Leitão A 2011 Modified iterated Tikhonov methods for solving systems of nonlinear ill-posed equations *Inverse Problems Imaging* **5** 1–17
[3] Baumeister J, Kaltenbacher B and Leitão A 2010 On Levenberg–Marquardt–Kaczmarz methods for regularizing systems of nonlinear ill-posed equations *Inverse Problems Imaging* **4** 335–50
[4] Bonesky T, Kazimierski K S, Maass P, Schöpfer F and Schuster T 2008 Minimization of Tikhonov functionals in Banach spaces *Abstr. Appl. Anal.* 192679
[5] Borwein J M and Vanderwerff J D 2010 Convex functions: constructions, characterizations and counterexamples (*Encyclopedia of Mathematics and Its Applications* vol 109) (Cambridge: Cambridge University Press)
[6] Cioranescu I 1990 Geometry of Banach spaces, duality mappings and nonlinear problems (*Mathematics and Its Applications* vol 62) (Dordrecht: Kluwer)
[7] De Cezaro A, Haltmeier M, Leitão A and Scherzer O 2008 On steepest-descent-Kaczmarz methods for regularizing systems of nonlinear ill-posed equations *Appl. Math. Comput.* **202** 596–607
[8] Engl H W, Hanke M and Neubauer A 1996 *Regularization of Inverse Problems* (Dordrecht: Kluwer)
[9] Haltmeier M, Kowar R, Leitão A and Scherzer O 2007 Kaczmarz methods for regularizing nonlinear ill-posed equations: II. Applications *Inverse Problems Imaging* **1** 507–23
[10] Haltmeier M, Leitão A and Scherzer O 2007 Kaczmarz methods for regularizing nonlinear ill-posed equations: I. Convergence analysis *Inverse Problems Imaging* **1** 289–98
[11] Haltmeier M, Leitão A and Resmerita E 2009 On regularization methods of EM-Kaczmarz type *Inverse Problems* **25** 075008
[12] Hanke M, Neubauer A and Scherzer O 1995 A convergence analysis of Landweber iteration for nonlinear ill-posed problems *Numer. Math.* **72** 21–37
[13] Hein T and Kazimierski K S 2010 Accelerated Landweber iteration in Banach spaces *Inverse Problems* **26** 055002
[14] Hein T and Kazimierski K S 2010 Modified Landweber iteration in Banach spaces—convergence and convergence rates *Numer. Funct. Anal. Optim.* **31** 1158–84
[15] Hofmannn B, Kaltenbacher B, Pöschl C and Scherzer O 1998 A convergence rates result in Banach spaces with non-smooth operators *Inverse Problems* **23** 987–1010
[16] Kaczmarz S 1993 Approximate solution of systems of linear equations *Int. J. Control* **57** 1269–71
[17] Kaltenbacher B, Neubauer A and Scherzer O 2008 Iterative regularization methods for nonlinear ill-posed problems (*Radon Series on Computational and Applied Mathematics* vol 6) (Berlin: de Gruyter & Co)
[18] Kaltenbacher B, Schöpfer F and Schuster T 2009 Iterative methods for nonlinear ill-posed problems in Banach spaces: convergence and applications to parameter identification problems *Inverse Problems* **25** 065003
[19] McCormick S 1977 The methods of Kaczmarz and row orthogonalization for solving linear equations and least squares problems in Hilbert space *Indiana Univ. Math. J.* **26** 1137–50
[20] Morozov V A 1993 *Regularization Methods for Ill-Posed Problems* (Boca Raton, FL: CRC Press)
[21] Natterer F 1997 Algorithms in tomography (*State of the Art in Numerical Analysis* vol 63) pp 503–24

[22] Resmerita E 2005 Regularization of ill-posed problems in Banach spaces: convergence rates *Inverse Problems* **21** 1303–14

[23] Schöpfer F, Louis A K and Schuster T 2006 Nonlinear iterative methods for linear ill-posed problems in Banach spaces *Inverse Problems* **22** 311–29

[24] Schöpfer F, Schuster T and Louis A K 2008 An iterative regularization method for the solution of the split feasibility problem in Banach spaces *Inverse Problems* **24** 055008

[25] Tikhonov A N and Arsenin V Y 1977 *Solutions of Ill-Posed Problems* ed F John (Washington, DC: Wiley)

[26] Xu Z B and Roach G F 1991 Characteristic inequalities of uniformly convex and uniformly smooth Banach spaces *J. Math. Anal. Appl.* **157** 189–210