TECHNISCHE UNIVERSITÄT
CHEMNITZ

Accelerated Landweber iteration in Banach spaces

T. Hein, K.S. Kazimierski

Preprint 2009-17

Fakultät für Mathematik

Impressum:
Herausgeber:
Der Dekan der
Fakultät für Mathematik
an der Technischen Universität Chemnitz
Sitz:
Reichenhainer Straße 39
09126 Chemnitz
Postanschrift:
09107 Chemnitz
Telefon: (0371) 531-22000
Telefax: (0371) 531-22009
E-Mail: dekanat@mathematik.tu-chemnitz.de
Internet:
http://www.tu-chemnitz.de/mathematik/
ISSN 1614-8835 (Print)

Accelerated Landweber iteration in Banach spaces

Torsten Hein [*], Kamil S. Kazimierski [†]

February 26, 2010

Abstract

Abstract. We investigate a method of accelerated Landweber type for the iteratve regularization of nonlinear ill-posed operator equations in Banach spaces. Based on an auxiliary algorithm with simplified choice of the step size parameter we present a convergence and stability analysis of the algorithm under consideration. We will close our discussion with the presentation of a numerical example.

Key words: Iterative Regularization, Landweber iteration, Banach spaces, smooth of power type, convex of power type, Bregman distance

AMS subject classifications: 47A52, 65F10, 46E15, 46B20

1 Introduction

Let $\mathcal{X}$ and $\mathcal{Y}$ be both Banach spaces with dual spaces $\mathcal{X}^*$ and $\mathcal{Y}^*$ respectively. We consider the nonlinear ill-posed operator equation

$$F(x) = y, \qquad x \in \mathcal{D}(F), \tag{1}$$

where $F : \mathcal{D}(F) \subseteq \mathcal{X} \longrightarrow \mathcal{Y}$ describes a continuous nonlinear mapping from the domain $\mathcal{D}(F)$ into the space $\mathcal{Y}$. In many applications the ill-posedness arises from instability effects: even if a solution $x \in \mathcal{D}(F)$ satisfying $F(x) = y$ exists it does not depend continuously on the data y. On the other hand it can be assumed that only a perturbed version y^δ of y is available where only the estimate $\|y^\delta - y\| \le \delta$ is known. Therefore we have to apply regularization methods. In particular, Tikhonov regularization has been well-established theoretically in e.g. [19], [11], [8] and the references therein. On the other hand, the corresponding numerical treatment (for linear problems) was considered in e.g. [2] and [4].

[*]Technische Universität Chemnitz, Fakultät für Mathematik, 09107 Chemnitz, Deutschland. Email: `Torsten.Hein@mathematik.tu-chemnitz.de`

[†]Center for Industrial Mathematics (ZeTeM), University of Bremen, 28334 Bremen, Germany Email: `kamilk@math.uni-bremen.de`

Major drawback of Tikhonov regularization is the high numerical effort. For the proper determination of the regularization parameter α we usually have to solve several non-quadratic minimization problems (exactly up to a numerical error). Therefore the development and analysis of iterative regularization methods in Banach spaces are of high interest. The theory of iterative regularization methods in Hilbert spaces has been deeply studied in the recent years. For a short overview we refer to [5], for more detailed information to [1] and [13].

Here the focus is on gradient-type methods. We refer also to [14] for an iteratively regularized Gauss-Newton-type approach in Banach spaces. For given parameter $p > 1$ we reformulate equation (1) as minimization problem

$$\Omega_p(x) := \frac{1}{p}\|F(x) - y^\delta\|^p \to \min \quad \text{subject to} \quad x \in \mathcal{D}(F). \tag{2}$$

We generalize the results of [6] and [18] in a first step. Using a gradient method for solving the problem (2) we therefore deal with the following iteration:

$$\begin{aligned} x_0^\delta &:= x_0 = J_{s^*}^*(x_0^*) \in \mathcal{D}(F) \;\text{ with }\; x_0^* \in \mathcal{X}^*, \\ x_{n+1}^* &:= x_n^* - \mu_n F'(x_n^\delta)^\star J_p\left(F(x_n^\delta) - y^\delta\right), \\ x_{n+1}^\delta &:= J_{s^*}^*(x_{n+1}^*), \end{aligned}$$

together with a proper choice of the step size μ_n and an appropriate stopping criterion. The choice of the parameter $s^* \in (1, 2]$ is determined by the supposed smoothness of the dual space $\mathcal{X}^*$. Moreover, $J_p : \mathcal{Y} \longrightarrow \mathcal{Y}^*$ and $J_{s^*}^* : \mathcal{X}^* \longrightarrow \mathcal{X}$ denote corresponding duality mappings with gauge functions $t \mapsto t^{p-1}$ and $t \mapsto t^{s^*-1}$ respectively. The algorithm above was considered in Banach spaces for linear operators in [18] and generalized to nonlinear problems in [14]. There, similar nonlinearity restrictions to the operator F were applied as already supposed in [6] in the Hilbert space setting. We present here an analysis which is closely related to the one in [14] even the results are somewhat different. If $\mathcal{X}$ and $\mathcal{Y}$ are Hilbert spaces then the choices $p = 2$ and $\mu_n \equiv 1$ reduce the algorithm to classical Landweber iteration for nonlinear ill-posed problems which was originally considered in [6]. However, a constant step size leads usually to a slow convergence of gradient methods. Therefore an appropriate choice of the parameter μ_n in each iteration step is crucial for a satisfying speed of convergence of the iteration process. We also point out that the update of the iterates (i.e. the search of the optimal step size μ_n) in fact takes place in the dual space $\mathcal{X}^*$.

Therefore the paper is organized as follows: in Section 2 we introduce basic notations and assumptions. In Section 3 the main algorithm is derived and the existence of the suggested choice of the step size μ_n is proved. Furthermore an auxiliary algorithm based on an explicit calculation of the step size μ_n and its descent property is shown. Section 5 deals with convergence and stability of the algorithm under consideration. Two numerical examples in the last section illustrate these theoretical results.

2 Preliminaries

Throughout the paper let $1 < s, s^* < \infty$ denote conjugate exponents, i.e.

$$\frac{1}{s} + \frac{1}{s^*} = 1.$$

For $x \in \mathcal{X}$ and $x^* \in \mathcal{X}^*$ we denote by $\langle x, x^* \rangle$ or $\langle x^*, x \rangle$ the associated duality product. Norms will be denoted as usual by $\|\cdot\|$. We obmit indices indicating the underlying spaces since it will become clear out of the context.

For the convergence analysis we need the following assumptions:

(A1) The Banach space $\mathcal{X}$ is supposed to be s-convex for some $s \in [2, \infty)$ and $\mathcal{Y}$ is assumed to be smooth.

(A2) For $\delta = 0$ there exists a solution $x_* \in \mathcal{D}(F)$ of (1), i.e. $F(x_*) = y$ holds.

(A3) There exists a ball $\mathcal{B}_\varrho(x_*) \subseteq \mathcal{D}(F)$ around x_* with radius $\varrho > 0$ such that:

(i) For all $x \in \mathcal{B}_\varrho(x_*)$ the operator F is Fréchet-differentiable with Fréchet-derivative $F'(x) : \mathcal{X} \longrightarrow \mathcal{Y}$.

(ii) The operator F is of degree $(1, 0)$ of nonlinearity with uniform constant $0 \leq L < 1$ on $\mathcal{B}_\varrho(x_*)$, i.e.

$$\|F(\tilde{x}) - F(x) - F'(x)(\tilde{x} - x)\| \leq L \, \|F(\tilde{x}) - F(x)\| \tag{3}$$

holds for all $x, \tilde{x} \in \mathcal{B}_\varrho(x_*)$.

(iii) It holds $\|F'(x)\| \leq K$ uniformly for some constants $K > 0$ on $\mathcal{B}_\varrho(x_*)$.

We shortly discuss these conditions. We recall that the Banach space $\mathcal{X}$ is said to be convex of power-type $s \in [2, \infty)$ or s-convex if for the modulus of convexity $\sigma_X : [0, 2] \longrightarrow [0, 1]$

$$\delta_X(\varepsilon) := \inf \left\{ 1 - \frac{1}{2}\|x + \tilde{x}\| \ : \ \|x\| = \|\tilde{x}\| = 1, \ \|x - \tilde{x}\| \geq \varepsilon \right\} \geq C_X \varepsilon^s, \quad \varepsilon \in [0, 2],$$

holds for some constant $C_X > 0$ and smooth of power-type $r \in (1, 2]$ or r-smooth if for the modulus of smoothness $\rho_X : [0, \infty) \longrightarrow [0, \infty)$

$$\rho_X(\tau) := \frac{1}{2} \sup \left\{ \|x + \tilde{x}\| + \|x - \tilde{x}\| - 2 \ : \|x\| = 1, \ \|\tilde{x}\| \leq \tau \right\} \leq \hat{C}_X \tau^r, \quad \tau \in [0, \infty),$$

holds for another constant $\hat{C}_X > 0$. We refer e.g. to [3] and [16] for detailed information about geometric properties of Banach spaces.

Moreover, from the s-convexity of the (reflexive) space $\mathcal{X}$ we conclude the s^*-smoothness of the dual space $\mathcal{X}^*$. By the Xu/Roach inequalities [20] there exist constants $C_s > 0$ and $G^*_{s^*} > 0$ such that

$$\frac{1}{s}\|\tilde{x}\|^s - \frac{1}{s}\|x\|^s - \langle x^*, \tilde{x} - x \rangle \geq \frac{C_s}{s}\|x - \tilde{x}\|^s, \quad x^* \in J_s(x), \tag{4}$$

with duality mapping $J_s : \mathcal{X} \longrightarrow 2^{\mathcal{X}^*}$ where $2^{\mathcal{X}^*}$ is the power-set of $\mathcal{X}^*$ and

$$\frac{1}{s^*}\|\tilde{x}^*\|^{s^*} - \frac{1}{s^*}\|x^*\|^{s^*} - \langle J^*_{s^*}(x^*), \tilde{x}^* - x^*\rangle \leq \frac{G^*_{s^*}}{s^*}\|x^* - \tilde{x}^*\|^{s^*} \tag{5}$$

hold for all $x, \tilde{x} \in \mathcal{X}$ and $x^*, \tilde{x}^* \in \mathcal{X}^*$ respectively. Both constants we will need in our convergence analysis. We recall that the duality mapping $J_q : \mathcal{X} \longrightarrow 2^{\mathcal{X}^*}$ is defined as

$$J_q(x) := \left\{x^* \in \mathcal{X}^* \ : \ \langle x^*, x\rangle = \|x^*\|\,\|x\|,\ \|x^*\| = \|x\|^{q-1}\right\}$$

The properties of duality mappings are well-studied, see e.g. [3, Chapter 1 and 2] and [21, Proposition 47.17]. The smoothness of the space $\mathcal{Y}$ guarantees that duality mappings from $\mathcal{Y}$ into $\mathcal{Y}^*$ are always single valued.

Concerning assumption (A3) we remark the following. The nonlinearity restriction of type (3) with $L < \frac{1}{2}$ was applied in [6] for dealing with Landweber iteration for nonlinear ill-posed problems in Hilbert spaces. We emphasize that in our convergence analysis the weaker condition $L < 1$ is sufficient. In particular, we make use of the inequality

$$\frac{1}{1+L}\|F'(x)(\tilde{x} - x)\| \leq \|F(\tilde{x}) - F(x)\| \leq \frac{1}{1-L}\|F'(x)(\tilde{x} - x)\|$$

for $x, \tilde{x} \in \mathcal{B}_\varrho(x_*)$ which is an immediate consequence of (3). With $L = 0$ we also include the case of linear equations in our considerations. The concept of the degree of nonlinearity for nonlinear operators in Hilbert spaces was introduced in [12]. In [8] this approach was transfered to Banach spaces.

3 The algorithm

For $x^* \in \mathcal{X}^*$, $x = J^*_{s^*}(x^*)$ and arbitrary $\tilde{x} \in \mathcal{X}$ we introduce the notation

$$\Delta_s(\tilde{x}, x) := \frac{1}{s}\|\tilde{x}\|^s - \frac{1}{s}\|x\|^s - \langle x^*, \tilde{x} - x\rangle$$

for the Bregman distance of the functional $x \mapsto \frac{1}{s}\|x\|^s$. Here we applied that $x^* \in J_s(x)$ holds. Using the definition of duality mappings we have

$$\langle x^*, x\rangle = \|x\|^s = \|x^*\|^{s^*}$$

which allows us to reformulate the Bregman distance as

$$\Delta_s(\tilde{x}, x) = \frac{1}{s}\|\tilde{x}\|^s + \frac{1}{s^*}\|x\|^s - \langle x^*, \tilde{x}\rangle = \frac{1}{s}\|\tilde{x}\|^s + \frac{1}{s^*}\|x^*\|^{s^*} - \langle x^*, \tilde{x}\rangle.$$

We return to the minimization problem (2) with arbitrary parameter $p > 1$. Let the n-th iterates x_n^δ and x_n^* be given. Then – with $A_n^\star := F'(x_n^\delta)^\star$, $\psi_n^* = A_n^\star J_p(F(x_n^\delta) - y^\delta)$ and $\Delta_n := \Delta_s(x_*, x_n^\delta)$ – we derive

$$\begin{aligned}
\Delta_s(x_*, J^*_{s^*}(x_n^* - \mu\psi_n^*)) - \Delta_n \ &= \ \frac{1}{s^*}\|x_n^* - \mu\,\psi_n^*\|^{s^*} - \frac{1}{s^*}\|x_n^*\|^{s^*} + \mu\,\langle A_n^\star J_p(F(x_n^\delta) - y^\delta), x_*\rangle \\
&= \ \frac{1}{s^*}\|x_n^* - \mu\,\psi_n^*\|^{s^*} - \frac{1}{s^*}\|x_n^*\|^{s^*} + \mu\,\langle \psi_n^*, x_n^\delta\rangle \\
&\qquad + \mu\,\langle J_p(F(x_n^\delta) - y^\delta), A_n(x_* - x_n^\delta)\rangle.
\end{aligned}$$

Furthermore, with $T(x_*, x_n^\delta) := F(x_*) - F(x_n^\delta) - A_n(x_* - x_n^\delta)$ we have

$$\begin{aligned}
&\mu \, \langle J_p(F(x_n^\delta) - y^\delta), F(x_*) - F(x_n^\delta) + T(x_*, x_n^\delta)\rangle \\
&\quad = \mu_n \left(-\langle J_p(F(x_n^\delta) - y^\delta), F(x_n^\delta) - y^\delta\rangle - \langle J_p(F(x_n^\delta) - y^\delta), T(x_*, x_n^\delta)\rangle \right. \\
&\qquad\quad \left. + \langle J_p(F(x_n^\delta) - y^\delta), y - y^\delta\rangle \right) \\
&\quad \leq \mu_n \left(-\|F(x_n^\delta) - y^\delta\|^p + L\|F(x_n^\delta) - y\| \|F(x_n^\delta) - y^\delta\|^{p-1} + \delta \|F(x_n^\delta) - y^\delta\|^{p-1} \right) \\
&\quad \leq \mu_n \left(-(1-L)\|F(x_n^\delta) - y^\delta\|^p + (1+L)\delta \|F(x_n^\delta) - y^\delta\|^{p-1} \right).
\end{aligned}$$

We introduce the notation

$$c_n^\delta := (1-L)\|F(x_n^\delta) - y^\delta\|^p - (1+L)\delta\|F(x_n^\delta) - y^\delta\|^{p-1}$$

Then we derive

$$\Delta_s(x_*, J_{s^*}^*(x_n^* - \mu\psi_n^*)) - \Delta_n \leq \frac{1}{s^*}\|x_n^* - \mu\,\psi_n^*\|^{s^*} - \frac{1}{s^*}\|x_n^*\|^{s^*} + \mu\,\langle \psi_n^*, x_n^\delta\rangle - \mu\, c_n^\delta.$$

In the linear case $L = 0$ even equality holds for given exact data, i.e. $\delta = 0$. Therefore we suggest the following choice of the step size μ_n: choose the parameter $\mu = \mu_n$ in such a way, that the right hand side of the above estimate becomes minimal (with respect to μ). Collecting all terms on the right hand side depending on μ we define

$$f(\mu) := \frac{1}{s^*}\|x_n^* - \mu\,\psi_n^*\|^{s^*} + \mu\,\langle \psi_n^*, x_n^\delta\rangle - \mu\, c_n^\delta, \qquad \mu \geq 0.$$

Then we easily can prove the following lemma.

Lemma 3.1. *Assume (A1)-(A3), $x_n^\delta \in \mathcal{B}_\varrho(x_*) \cap \mathcal{D}(F)$ and $\psi_n^* \neq 0$. Then the minimization problem*

$$f(\mu) \to \min \qquad \textit{subject to} \quad \mu > 0 \tag{6}$$

has a unique solution $\mu^* > 0$ as long as $c_n^\delta > 0$.

Proof. Differentiating $f(\mu)$ we see

$$f'(\mu) = -\langle J_{s^*}^*(x_n^* - \mu\,\psi_n^*), \psi_n^*\rangle + \langle x_n^\delta, \psi_n^*\rangle - c_n^\delta.$$

By monotonicity of the duality mappings this function $f'(\mu)$ is strictly increasing. By assumption we have $c_n^\delta > 0$ which shows $f'(0) = -c_n^\delta < 0$. On the other hand we have $s^* > 1$. Hence the norm term in $f(\mu)$ dominates as $\mu \to \infty$ which shows $f(\mu) \to \infty$ as $\mu \to \infty$. From continuity of $f'(\mu)$ we can conclude the existence of a unique element $\mu = \mu^* > 0$ satisfying the necessary optimality condition $f'(\mu) = 0$. ■

We are now able to present the algorithm under consideration in detail:

Algorithm 3.1.

(S0) Init. Choose start point $x_0^* \in \mathcal{X}^*$, $x_0^\delta := J_{s^*}^*(x_0^*)$ with $\Delta_s(x_*, x_0^\delta) < \varrho^s \frac{C_s}{s}$ with constant C_2 from (4). Choose an upper bound $\overline{\mu} \in (0, \infty]$ for the step size and define the parameter $\tau > 1$ such that

$$\tau > \frac{1+L}{1-L}$$

holds. Set $n := 0$.

(S1) STOP, if for $\delta > 0$ the discrepancy criterion $\left\|F(x_n^\delta) - y^\delta\right\| \le \tau\,\delta$ is fulfilled or we have $F(x_n^\delta) = y$ for $\delta = 0$.

(S2) Calculate $\psi_n^* := F'(x_n^\delta)^\star J_p(F(x_n^\delta) - y^\delta)$ and find the solution μ^* of the equation

$$f'(\mu) = 0, \qquad \mu \ge 0. \tag{7}$$

Set $\mu_n := \min\left\{\mu^*, \overline{\mu}\,\|F(x_n^\delta) - y^\delta\|^{s-p}\right\}$.

(S3) Calculate the new iterate

$$\begin{aligned} x_{n+1}^* &:= x_n^* - \mu_n \psi_n^*, \qquad \textit{and} \\ x_{n+1}^\delta &:= J_{s^*}^*(x_{n+1}^*). \end{aligned}$$

Set $n := n+1$ and go to step (S1).

In the noiseless case we will write x_n instead of $x_n^\delta = x_n^0$ for the iterates. Let $N(\delta, y^\delta)$ denotes the index where the iteration stops. Then we have the relation

$$\left\|F(x_{N(\delta,y^\delta)}^\delta) - y^\delta\right\| \le \tau\,\delta < \left\|F(x_n^\delta) - y^\delta\right\|, \quad 0 \le n < N(\delta, y^\delta). \tag{8}$$

For the proof of convergence and stability we deal with an auxiliary problem.

Remark 3.1. *For linear operators in Hilbert spaces, noiseless data, i.e. $\delta = 0$, and the choice $p = 2$ this algorithm reduces to the classical method of minimal error which was already considered in [17] in the context of regularization methods.*

4 On an auxiliary problem

Main problem in the analysis of Algorithm 3.1 arises from the fact that we cannot state the step size μ_n explicitly. Therefore we now suggest the following modified algorithm with an explicit expression for the chosen step size. In turns out that convergence and stability results derived for the modified algorithm also remain valid for the original version.

For given x_n^δ and arbitrary $\mu \ge 0$ we recall the relation

$$\Delta_s(x_*, J_{s^*}^*(x_n^* - \mu\psi_n^*)) - \Delta_n = \frac{1}{s^*}\|x_n^* - \mu\,\psi_n^*\|^{s^*} - \frac{1}{s^*}\|x_n^*\|^{s^*} + \mu\,\langle A_n^\star J_p(F(x_n^\delta) - y^\delta), x_*\rangle.$$

From the s^*-smoothness of the space $\mathcal{X}^*$ we conclude

$$\begin{aligned} \frac{1}{s^*}\|x_n^* - \mu\psi_n^*\|^{s^*} - \frac{1}{s^*}\|x_n^*\|^{s^*} &\le -\mu\langle A_n^\star J_p(F(x_n^\delta) - y^\delta), x_n^\delta\rangle + \frac{G_{s^*}^*}{s^*}\|\mu A_n^\star J_p(F(x_n^\delta) - y^\delta)\|^{s^*} \\ &= -\mu\langle A_n^\star J_p(F(x_n^\delta) - y^\delta), x_n^\delta\rangle + \frac{G_{s^*}^*}{s^*}\mu^{s^*}\|\psi_n^*\|^{s^*}, \end{aligned}$$

where $G_{s^*}^*$ is the constant from inequality (5). Continuing as in the previous section we obtain

$$\Delta_s(x_*, J_{s^*}^*(x_n^* - \mu\psi_n^*)) \le \Delta_s(x_*, x_n^\delta) - \left(c_n^\delta\mu + \frac{G_{s^*}^*}{s^*}\|\psi_n^*\|^{s^*}\mu^{s^*}\right).$$

Minimizing the right hand side with respect to the step size μ we obtain the following modified algorithm:

Algorithm 4.1. *In Algorithm 3.1 we replace step (S2) by*

(S2') Calculate $\psi_n^* := F'(x_n^\delta)^\star J_p\left(F(x_n^\delta) - y^\delta\right)$,

$$\begin{aligned} c_n^\delta &:= \|F(x_n^\delta) - y^\delta\|^p(1-L) - (1+L)\delta\,\|F(x_n^\delta) - y^\delta\|^{p-1}, \\ \hat{c}_n^\delta &:= \max\left\{G_{s^*}^*\|\psi_n^*\|^{s^*}, c_n^\delta\left(\overline{\mu}\,\|F(x_n^\delta) - y^\delta\|^{s-p}\right)^{-\frac{1}{s-1}}\right\} \quad \textit{and} \\ \mu_n &:= \left(\frac{c_n^\delta}{\hat{c}_n^\delta}\right)^{s-1}. \end{aligned}$$

In particular, the choice $\overline{\mu} = \infty$ is allowed in this assertion. Then $\hat{c}_n^\delta = G_{s^*}^*\|\psi_n^*\|^{s^*}$ holds automatically. We further introduce the constants

$$\underline{\mu}_\tau = \min\left\{\frac{(1-L-(1+L)\tau^{-1})^{s-1}}{{G_{s^*}^*}^{s-1}K^s}, \overline{\mu}\right\} > 0 \quad \text{and} \quad \underline{\mu}_0 := \min\left\{\frac{(1-L)^{s-1}}{{G_{s^*}^*}^{s-1}K^s}, \overline{\mu}\right\} > 0,$$

as well as

$$\lambda_\tau := \frac{1-L-(1+L)\tau^{-1}}{s} > 0 \quad \text{and} \quad \lambda_0 := \frac{(1-L)}{s} > 0.$$

Then we can prove the following.

Lemma 4.1. *Assume (A1)-(A3) and all iterates* $\{x_n^\delta\}$ *generated by Algorithm 4.1 remain in* $\mathcal{B}_\varrho(x_*) \cap \mathcal{D}(F)$*. Then, for all* $n < N(\delta, y^\delta)$ *the following holds true:*

(i) The step size μ_n *is the (unique) maximizer of* $C(\mu) := c_n^\delta\mu - \frac{\hat{c}_n^\delta}{s^*}\mu^{s^*}$, $\mu \in [0,\infty)$.

(ii) For $\delta > 0$ *we have* $\mu_n \in [\underline{\mu}_\tau, \overline{\mu}]\|F(x_n^\delta) - y^\delta\|^{s-p}$ *and*

$$C(\mu_n) \geq \lambda_\tau\mu_n\|F(x_n^\delta) - y^\delta\|^p \geq \lambda_\tau\underline{\mu}_\tau\|F(x_n^\delta) - y^\delta\|^s > 0.$$

(iii) For $\delta = 0$ *we have* $\mu_n \in [\underline{\mu}_0, \overline{\mu}]\|F(x_n) - y\|^{s-p}$ *and*

$$C(\mu_n) \geq \lambda_0\mu_n\|F(x_n) - y\|^p \geq \lambda_0\underline{\mu}_\tau\|F(x_n) - y\|^s > 0.$$

PROOF. The first part follows immediately by the definition of μ_n. Moreover, we observe

$$C(\mu_n) = \frac{\left(c_n^\delta\right)^s}{\left(\hat{c}_n^\delta\right)^{s-1}} - \frac{1}{s^*}\frac{\left(c_n^\delta\right)^s}{\left(\hat{c}_n^\delta\right)^{s-1}} = \frac{1}{s}\frac{\left(c_n^\delta\right)^s}{\left(\hat{c}_n^\delta\right)^{s-1}} = \frac{1}{s}\mu_n c_n^\delta.$$

We now assume $\delta > 0$. Since the stopping criterion is not fulfilled we get $\tau\delta < \|F(x_n^\delta) - y^\delta\|$. Then we estimate

$$\begin{aligned} c_n^\delta &= \|F(x_n^\delta) - y^\delta\|^p\left[1 - L - (1+L)\frac{\delta}{\|F(x_n^\delta) - y^\delta\|}\right] \\ &\geq \|F(x_n^\delta) - y^\delta\|^p\left[1 - L - (1+L)\tau^{-1}\right] > 0. \end{aligned}$$

This automatically proves

$$C(\mu_n) = \frac{1}{s}\mu_n c_n^\delta \geq \lambda_\tau\mu_n\|F(x_n^\delta) - y^\delta\|^p.$$

Assume now $\overline{\mu} < \infty$ and $\hat{c}_n^\delta = \left(\overline{\mu}\, \|F(x_n^\delta) - y^\delta\|^{s-p}\right)^{-\frac{1}{s-1}} c_n^\delta$. Then $\mu_n = \overline{\mu}\, \|F(x_n^\delta) - y^\delta\|^{s-p}$ holds by construction. Hence we have

$$C(\mu_n) = \frac{1}{s}\mu_n c_n^\delta \geq \|F(x_n^\delta) - y^\delta\|^p \frac{1 - L - (1+L)\tau^{-1}}{s}\mu_n = \lambda_\tau \overline{\mu} \|F(x_n^\delta) - y^\delta\|^s.$$

On the other we now suppose $\hat{c}_n^\delta = G_{s^*}^* \|\psi_n^*\|^{s^*}$. Since $\hat{c}_n^\delta \geq \left(\overline{\mu}\, \|F(x_n^\delta) - y^\delta\|^{s-p}\right)^{-\frac{1}{s-1}} c_n^\delta$ we derive $\mu_n \leq \overline{\mu}\, \|F(x_n^\delta) - y^\delta\|^{s-p}$. With $\|F'(x_n^\delta)^\star\| = \|F'(x_n^\delta)\| \leq K$ we estimate

$$\hat{c}_n^\delta \leq G_{s^*}^* \|F'(x_n^\delta)^\star\|^{s^*} \|J_p(F(x_n^\delta) - y^\delta)\|^{s^*} \leq G_{s^*}^* K^{s^*} \|F(x_n^\delta) - y^\delta\|^{s^*(p-1)}.$$

Hence we obtain

$$\begin{aligned} \mu_n = \left(\frac{c_n^\delta}{\hat{c}_n^\delta}\right)^{s-1} &\geq \frac{\left(1 - L - (1+L)\tau^{-1}\right)^{s-1}}{{G_{s^*}^*}^{s-1} K^s} \|F(x_n^\delta) - y^\delta\|^{(p - s^*(p-1))(s-1)} \\ &= \frac{\left(1 - L - (1+L)\tau^{-1}\right)^{s-1}}{{G_{s^*}^*}^{s-1} K^s} \|F(x_n^\delta) - y^\delta\|^{s-p} \geq \underline{\mu}_\tau \|F(x_n^\delta) - y^\delta\|^{s-p} \end{aligned}$$

in that case. Consequently we can estimate

$$C(\mu_n) = \frac{1}{s} c_n^\delta \mu_n \geq \frac{1}{s} \|F(x_n^\delta) - y^\delta\|^s \left[1 - L - (1+L)\tau^{-1}\right] \underline{\mu}_\tau = \lambda_\tau \underline{\mu}_\tau \|F(x_n^\delta) - y^\delta\|^s.$$

This proves the second part. For $\delta = 0$ we have $c_n^\delta = \|F(x_n) - y\|^p (1 - L)$. Then an analogous calculation shows the third part of the assertions. ■

Remark 4.1. *Assume $F(x_n^\delta) \to y^\delta$ as $n \to \infty$. For $s > p$ we have $\|F(x_n^\delta) - y^\delta\|^{s-p} \to 0$ and hence $\mu_n \to 0$ as $n \to \infty$. On the other hand, for $s < p$ we conclude $\|F(x_n^\delta) - y^\delta\|^{s-p} \to \infty$ and consequently $\mu_n \to \infty$ as $n \to \infty$. However, in both cases we derive*

$$\|x_{n+1}^* - x_n^*\| = \mu_n \|\psi_n^*\| \leq \overline{\mu}\, \|F(x_n^\delta) - y^\delta\|^{s-p} K\, \|F(x_n^\delta) - y^\delta\|^{p-1} = \overline{\mu} K\, \|F(x_n^\delta) - y^\delta\|^{s-1}$$

and hence $\|x_{n+1}^* - x_n^*\| \to 0$ as $n \to \infty$ as long as we have chosen $\overline{\mu} < \infty$.

We now show that as long as the discrepancy principle is not fulfilled the algorithm generates a decreasing sequence $\{\Delta_s(x_*, x_n^\delta)\}$ of Bregman distances. Moreover, for $\delta > 0$ the algorithm terminates after a finite number $N(\delta, y^\delta)$ of iterations. We prove the following lemma.

Lemma 4.2. *Assume (A1)-(A3). Then, for all $0 \leq n < N(\delta, y^\delta)$, we have $x_n^\delta \in \mathcal{B}_\varrho(x_*)$ and the estimate*

$$\Delta_s(x_*, x_{n+1}^\delta) < \Delta_s(x_*, x_n^\delta)$$

is valid. . Moreover, the following hold:

(i) If $\delta > 0$ then the algorithm stops after a finite number $N(\delta, y^\delta)$ of iterations. Moreover we have the estimates

$$N(\delta, y^\delta) \leq C\, \delta^{-s}$$

for some constant $C > 0$ as well as

$$\sum_{n=0}^{N(\delta,y^\delta)-1} \mu_n \|F(x_n^\delta) - y^\delta\|^p \leq \lambda_\tau^{-1} \Delta_s(x_*, x_0)$$

and

$$\sum_{n=0}^{N(\delta,y^\delta)-1} \|F(x_n^\delta) - y^\delta\|^s \le \lambda_\tau^{-1} \underline{\mu}_\tau^{-1} \Delta_s(x_*, x_0).$$

(ii) For $\delta = 0$ we have

$$\sum_{n=0}^{\infty} \mu_n \|F(x_n) - y\|^p \le \lambda_0^{-1} \Delta_s(x_*, x_0) \quad \textit{and} \quad \sum_{n=0}^{\infty} \|F(x_n) - y\|^s \le \lambda_0^{-1} \underline{\mu}_\tau^{-1} \Delta_s(x_*, x_0).$$

Proof. With the introduced notation we derived

$$\Delta_{n+1} \le \Delta_n - \left(c_n^\delta \mu_n - \frac{\hat{c}_n^\delta}{s^*} \mu_n^{s^*} \right).$$

Hence, from Lemma 4.1 we get

$$\Delta_{n+1} \le \Delta_n - \lambda_0 \mu_n \|F(x_n) - y\|^p \le \Delta_n - \lambda_0 \underline{\mu}_0 \|F(x_n) - y\|^s$$

for $\delta = 0$ and

$$\Delta_{n+1} \le \Delta_n - \lambda_\tau \mu_n \|F(x_n^\delta) - y^\delta\|^p \le \Delta_n - \lambda_\tau \underline{\mu}_\tau \|F(x_n^\delta) - y^\delta\|^s$$

for $\delta > 0$. This proves the lemma by induction. We only have to remark that

$$\frac{C_s}{s} \|x_{n+1}^\delta - x_*\|^s \le \Delta_s(x_*, x_{n+1}^\delta) < \Delta_s(x_*, x_0^\delta) \le \varrho^s \frac{C_s}{s}$$

which implies $x_{n+1}^\delta \in \mathcal{B}_\varrho(x_*)$. We observe for $\delta > 0$ that

$$\Delta_0 \ge \Delta_0 - \Delta_{N(\delta,y^\delta)} \ge \lambda_\tau \underline{\mu}_\tau \sum_{n=0}^{N(\delta,y^\delta)-1} \|F(x_n^\delta) - y^\delta\|^s \ge \lambda_\tau \underline{\mu}_\tau N(\delta, y^\delta)(\tau\,\delta)^s$$

or equivalently $N(\delta, y^\delta) \le \Delta_0 \lambda_\tau^{-1} \underline{\mu}_\tau^{-1} \tau^{-s} \delta^{-s}$ which completes the proof. ■

Remark 4.2. *The use of Bregman distances here is opposite to their application in the analysis of Tihkonov functionals with $P(x) = \frac{1}{s}\|x\|^s$, see e.g. [10]. Whereas for Tikhonov regularization the Bregman distances of form $\Delta_s(x, x_*)$ are used we here deal with the Bregman distance $\Delta_s(x_*, x)$ with interchanged arguments x and x_*. In particular, the element $x^* \in \mathcal{X}^*$ with $J_{s^*}^*(x^*) = x$ depends here also on the iteration number n. Taking the non-symmetry of Bregman distances into account both approaches may vary.*

5 Convergence results

We now discuss the convergence properties of the algorithms. We start with Algorithm 4.1 and the noiseless case $\delta = 0$. Here we can present the following convergence result. We show that the main ideas of the proof in [6] for the convergence for classical Landweber iteration of nonlinear equations in Hilbert spaces can be applied also in a Banach space setting.

Theorem 5.1. *Assume (A1)-(A3) and $\delta = 0$. Then Algorithm 4.1 stops either after a finite number N of iterations with x_N satisfying $F(x_N) = y$ or the sequence $\{x_n\}$ convergences to a solution of equation (1).*

PROOF. Assume that the iteration process does not stop after a finite number of iterations. From Lemma 4.2 we conclude $F(x_n) \to y$ as $n \to \infty$. Moreover, the sequence $\{\Delta_n\}$ of Bregman distances is convergent since it is monotonically decreasing and bounded from below by zero. We now show that $\{x_n\}$ is a Cauchy sequence. For arbitrary chosen indices $k > l$ we can find an index $l \le j \le k$ such that

$$\|F(x_j) - y\| \le \|F(x_n) - y\|, \qquad \forall\, l \le n \le k.$$

Using the triangle inequality we obtain $\|x_k - x_l\| \le \|x_k - x_j\| + \|x_j - x_l\|$. On the other hand we have from inequality (4) that

$$\begin{aligned}
\frac{C_s}{s}\|x_k - x_j\|^s &\le \Delta_s(x_j, x_k) \\
&= \frac{1}{s}\|x_j\|^s - \frac{1}{s}\|x_k\|^s - \langle x_k^*, x_j - x_k\rangle \\
&= \frac{1}{s}\|x_*\|^s - \frac{1}{s}\|x_k\|^s - \langle x_k^*, x_* - x_k\rangle - \left(\frac{1}{s}\|x_*\|^s - \frac{1}{s}\|x_j\|^s - \langle x_j^*, x_* - x_j\rangle\right) \\
&\qquad -\langle x_k^* - x_j^*, x_* - x_j\rangle \\
&= \Delta_k - \Delta_j - \langle x_k^* - x_j^*, x_* - x_j\rangle.
\end{aligned}$$

We have $|\Delta_k - \Delta_j| \to 0$ as $j, k \to \infty$. Furthermore, we derive

$$\begin{aligned}
\left|\langle x_k^* - x_j^*, x_j - x_*\rangle\right| &= \left|\sum_{n=j}^{k-1}\langle x_{n+1}^* - x_n^*, x_j - x_*\rangle\right| \\
&\le \left|\sum_{n=j}^{k-1}\mu_n\langle J_p(F(x_n) - y), F'(x_n)(x_j - x_n + x_n - x_*)\rangle\right| \\
&\le \sum_{n=j}^{k-1}\mu_n\|F(x_n) - y\|^{p-1}(1+L)\left(\|F(x_n) - y\| + \|F(x_n) - F(x_j)\|\right) \\
&\le \sum_{n=j}^{k-1}\mu_n\|F(x_n) - y\|^{p-1}(1+L)\left(2\|F(x_n) - y\| + \|F(x_j) - y\|\right) \\
&\le 3\sum_{n=j}^{k-1}\mu_n\|F(x_n) - y\|^{p-1}(1+L)\|F(x_n) - y\| \\
&= 3\,(1+L)\sum_{n=j}^{k-1}\mu_n\|F(x_n) - y\|^p.
\end{aligned}$$

For $j, k \to \infty$ the right hand side goes to zero. This proves $\|x_j - x_k\| \to 0$ as $j, k \to \infty$. A similar argumentation shows $\|x_j - x_l\| \to 0$ as $j, l \to \infty$. Consequently, $\{x_n\}$ is a Cauchy sequence and hence $x_n \to \tilde{x}_* \in \mathcal{B}_\varrho(x_*)$. By the continuity of F and $F(x_n) \to y$ we have $F(\tilde{x}_*) = y$ which shows that the limit element $\tilde{x}_*$ is a solution of (1). ■

Under some additional assumptions we can characterize the limit element $\tilde{x}_* \in \mathcal{D}(F)$. Therefore we cite the following result, see [14, Proposition 1] and [13, Proposition 2.1].

Proposition 5.1. *Assume (A3). Then for all $x \in \mathcal{B}_\rho(x_0)$ with radius $\rho > 0$ chosen such that $\mathcal{B}_\rho(x_0) \subset \mathcal{B}_\varrho(x_*) \subseteq \mathcal{D}(F)$ we have*

$$M_x := \{\tilde{x} \in \mathcal{B}_\rho(x_0) \ : \ F(\tilde{x}) = F(x)\} = (x + \mathcal{N}(F'(x))) \cap \mathcal{B}_\rho(x_0) \tag{9}$$

and

$$\mathcal{N}(F'(\tilde{x})) = \mathcal{N}(F'(x)), \qquad \forall\, \tilde{x} \in M_x. \tag{10}$$

The property (9) was already observed in [6, Proposition 2.1] in Hilbert spaces. Moreover, we need an additional condition which was introduced in [6] for the characterization of the limit element in a Hilbert space setting. There, the condition

$$\mathcal{N}(F'(x_*)) \subseteq \mathcal{N}(F'(x)) \qquad \text{for all} \quad x \in \mathcal{B}_\varrho(x_*) \tag{11}$$

was supposed. Using this assumption we can prove the following result.

Theorem 5.2. *Assume (A1)-(A3) and $\mathcal{B}_\rho(x_0) \subset \mathcal{B}_\varrho(x_*)$ for some radius $\rho > 0$. If $\mathcal{X}$ is additionally supposed to be smooth then the following hold:*

(i) Assume the set $\{x \in \mathcal{D}(F) \ : \ F(x) = y\} \cap \mathcal{B}_\rho(x_0)$ to be non-empty. Then the minimization problem

$$\Delta_s(x, x_0) \to \min \quad \textit{subject to} \quad \{x \in \mathcal{D}(F) \ : \ F(x) = y\} \cap \mathcal{B}_\rho(x_0) \tag{12}$$

has a solution which is unique if this solution belongs to the interior of $\mathcal{B}_\rho(x_0)$.

(ii) Suppose $\delta = 0$. Assume $x^\dagger \in int\mathcal{B}_\rho(x_0)$ to be the (unique) solution of (12). Then $J_s(x^\dagger) - x_0^* \in \overline{\mathcal{R}(F'(x^\dagger)^\star)}$ holds. Moreover, if additionally (11) is valid, then the sequence $\{x_n\}$ generated by Algorithm 4.1 converges to $x^\dagger$.

Proof. Assume $x \in \mathcal{D}(F)$ to be arbitrary chosen and x_* is chosen such that $x_* \in \mathcal{B}_\rho(x_0)$. From (9) we can conclude that $F(x) = y$ if and only if $x - x_* \in \mathcal{N}(F'(x_*)) = \mathcal{N}(F'(x))$. We examine the Bregman distance $\Delta_s(x, x_0)$. Let be $x, \tilde{x}_n \in \mathcal{X}$ arbitrary chosen with $\tilde{x}_n \rightharpoonup x$. By the weak lower semi-continuity of the norm we conclude

$$\frac{1}{s}\|x\|^s \le \liminf_{n\to\infty} \frac{1}{s}\|\tilde{x}_n\|^s.$$

Furthermore, $\langle x_0^*, \tilde{x}_n\rangle \to \langle x_0^*, x\rangle$ as $n \to \infty$ holds by definition of the weak convergence. Hence we have shown that

$$\Delta_s(x, x_0) \le \liminf_{n\to\infty} \Delta_s(\tilde{x}_n, x_0),$$

i.e. the Bregman distance is sequentially weakly lower semi-continuous with respect to the first argument. Then the existence of a minimizer is clear by the coercivity of the Bregman distance $\Delta_s(x, x_0)$ and the weak closedness of the set $M := x_* + \mathcal{N}(F'(x_*)) \cap \mathcal{B}_\rho(x_0)$, see e.g. [21, Theorem 38.A and Corollary 38.14]. Let $x_\sharp$ denote a solution of (12). Considering the Bregman distance $\Delta_s(x, x_0)$ we have

$$\frac{\partial}{\partial x}\Delta_s(x, x_0) = \frac{\partial}{\partial x}\left(\frac{1}{s}\|x\|^s - \frac{1}{s}\|x_0\|^s - \langle x_0^*, x - x_0\rangle\right) = J_s(x) - x_0^* \in \mathcal{X}^*.$$

Assume $x_\sharp \in \operatorname{int}\mathcal{B}_\rho(x_0)$ and let $x \in \mathcal{N}(F'(x_*))$ be chosen arbitrary. Then $x_\sharp \pm \lambda x \in M$ for $\lambda > 0$ sufficiently small. From the optimality condition we conclude

$$0 \leq \langle J_s(x_\sharp) - x_0^*, x_\sharp \pm \lambda x - x_\sharp \rangle = \pm\lambda \langle J_s(x_\sharp) - x_0^*, x\rangle$$

which implies

$$\langle J_s(x_\sharp) - x_0^*, x\rangle = 0, \qquad \forall\, x \in \mathcal{N}(F'(x_*)).$$

Let $\mathcal{N}(F'(x))^\perp$ denote the annihilator of $\mathcal{N}(F'(x))$. Since $\overline{\mathcal{R}(F'(x_*)^\star)} = \mathcal{N}(F'(x_*))^\perp$ this proves $J_s(x_\sharp) - x_0^* \in \overline{\mathcal{R}(F'(x_*)^\star)}$. Let $\tilde{x}_\sharp$ denote another solution of (12). Then

$$\langle J_s(x_\sharp) - J_s(\tilde{x}_\sharp), x\rangle = 0, \qquad \forall\, x \in \mathcal{N}(F'(x_*)),$$

follows immediately. We set $x := x_\sharp - \tilde{x}_\sharp$. This and the strict monotonicity of the duality mapping J_p, see e.g. [21, Proposition 47.19], imply $x_\sharp = \tilde{x}_\sharp$.
We consider the second part with $\{x_n\}$ generated by Algorithm 4.1. From the theorem above we have $x_n \to \tilde{x}_*$ and $x_n^* \to J_s(\tilde{x}_*)$ as $n \to \infty$ with $F(\tilde{x}_*) = y$. We have to show $x^\dagger = \tilde{x}_*$. Since $\overline{\mathcal{R}(F'(x_n)^\star)} = \mathcal{N}(F'(x_n))^\perp \subseteq \mathcal{N}(F'(x^\dagger))^\perp$, we see from the iteration process that $J_s(\tilde{x}_*) - x_0^* \in \mathcal{N}(F'(x^\dagger))^\perp$. In particular, this implies

$$\langle J_s(\tilde{x}_*) - x_0^*, x\rangle = 0 \qquad \forall\, x \in \mathcal{N}(F'(x^\dagger)).$$

Since $x^\dagger$ is the minimizer of (12) we immediately observe that

$$\langle J_s(x^\dagger) - J_s(\tilde{x}_*), x\rangle = 0, \qquad \forall\, x \in \mathcal{N}(F'(x^\dagger)).$$

Setting $x := x^\dagger - \tilde{x}_* \in \mathcal{N}(F'(x^\dagger))$, this again implies $x^\dagger = \tilde{x}_*$. This proves the theorem. ■

Finally – under some additional assumptions – we present a result which proves that Algorithm 4.1 describes in fact a regularization method.

Theorem 5.3. *Assume (A1)-(A3), $x^\dagger \in int\mathcal{B}_\rho(x_0) \subset \mathcal{B}_\varrho(x_*)$ for some radius $\rho > 0$. Suppose furthermore, that F' depends continuously on x, $\mathcal{X}$ is smooth and $\mathcal{Y}$ is uniformly smooth. If (11) holds and $\{x_n^\delta\}$ is generated by Algorithm 4.1 then we have convergence $x^\delta_{N(\delta,y^\delta)} \to x^\dagger$ as $\delta \to 0$.*

Proof. Introducing a change in the notation we write $(x_n^\delta)^*$ for the iterates in the dual space $\mathcal{X}^*$ for noisy data and x_n^* for the case $\delta = 0$. Let n be a fixed index. Since $\mathcal{X}^*$ and $\mathcal{Y}$ are uniformly smooth the duality mappings $J^*_{s^*}$ and J_p are uniformly continuous on each bounded set, see e.g. [21, Proposition 47.19]. Hence, under the assumptions stated above the iterated x_n^δ and $(x_n^\delta)^*$ depend continuously on the given data y^δ. From Theorem 5.2(ii) we conclude $x_n \to x^\dagger$ as $n \to \infty$. Without loss of generality we can assume $N(\delta, y^\delta) \to \infty$ as $\delta \to 0$. Then, for $N(\delta, y^\delta) \geq n$ we obtain

$$\begin{aligned}
\Delta_s(x^\dagger, x^\delta_{N(\delta,y^\delta)}) &\leq \Delta_s(x^\dagger, x_n^\delta) \\
&= \frac{1}{s}\|x^\dagger\|^s - \frac{1}{s}\|x_n^\delta\|^s - \langle (x_n^\delta)^*, x^\dagger - x_n^\delta\rangle \\
&= \|x^\dagger\|^s - \frac{1}{s}\|x_n\|^s - \langle x_n^*, x^\dagger - x_n\rangle + \frac{1}{s}\|x_n\|^s - \frac{1}{s}\|x_n^\delta\|^s \\
&\qquad + \langle x_n^*, x^\dagger - x_n\rangle - \langle (x_n^\delta)^*, x^\dagger - x_n^\delta\rangle \\
&= \Delta_s(x^\dagger, x_n) + \frac{1}{s^*}\left(\|x_n^\delta\|^s - \|x_n\|^s\right) + \langle (x_n^\delta)^* - x_n^*, x^\dagger\rangle.
\end{aligned}$$

The right hand side vanishes for $\delta \to 0$ and $n \to \infty$. On the other hand, convergence $\Delta_s(x^\dagger, x^\delta_{N(\delta,y^\delta)}) \to 0$ as $\delta \to 0$ implies $x^\delta_{N(\delta,y^\delta)} \to x^\dagger$ as $\delta \to 0$ since $\mathcal{X}$ is supposed to be s-convex. This proves the theorem. ■

Remark 5.1. *The smoothness of the space $\mathcal{X}$ was applied only for the characterization of the limit element $x^\dagger$. The regularization property of Algorithm 3.1 remains valid without this assumption since the duality mapping $J_s : \mathcal{X} \longrightarrow \mathcal{X}^*$ is neither required to be single-valued nor continuous.*

We now return to our starting point and discuss convergence and regularization property of Algorithm 3.1. By a simple observation we can apply the results of Lemma 4.2, Theorem 5.1 and Theorem 5.3 to prove the following.

Theorem 5.4. *Assume (A1)-(A3) and all iterates $\{x_n^\delta\}$ generated by Algorithm 3.1 remain in $\mathcal{D}(F)$. Then the following hold:*

(i) We set $\mu_n := \mu^*$ where μ^* is the solution of the problem (6). Then all results of Lemma 4.2 remain true.

(ii) We set $\mu_n := \min\{\mu^*, \overline{\mu}\, \|F(x_n^\delta) - y^\delta\|^{s-p}\}$, $0 < \overline{\mu} < \infty$. Then all results of Theorem 5.1 and Theorem 5.3 remain true under the assumptions stated therein.

Proof. By Lemma 3.1 the parameter choices are well defined. Let $\tilde{\mu}_n$ denotes the parameter generated by Algorithm 4.1. Assume $\delta > 0$. Then we derive by definition of μ_n that

$$\begin{aligned}\Delta_{n+1} - \Delta_n \;\le\; f(\mu^*) - \frac{1}{s^*}\|x_n^*\|^{s^*} \;&\le\; f(\tilde{\mu}_n) - \frac{1}{s^*}\|x_n^*\|^{s^*} \\ &\le\; -\left(C_{n,1}\tilde{\mu}_n - \frac{C_{n,2}}{s^*}\tilde{\mu}_n^{s^*}\right) \\ &\le\; -\lambda_\tau \tilde{\mu}_n \|F(x_n^\delta) - y^\delta\|^p\end{aligned}$$

which was the essential property for proving Lemma 4.2. For convergence and regularization property we additionally have to apply an upper bound $\overline{\mu}\|F(x_n^\delta) - y^\delta\|^{s-p}$ on the suggested choice of the step sizes μ_n and $\tilde{\mu}_n$. The case $\delta = 0$ follows similarly. ■

Remark 5.2. *We observe the following:*

- *Finding the parameter μ^* we have to solve the nonlinear equation (7) numerically. It cannot be calculated explicitly in general. Hence the price of a lower iteration number $N(\delta, y^\delta)$ is a higher numerical effort in each iteration step. So it might depend on the specific problem which algorithm is numerically more efficient.*
- *Since $f'(\mu)$, $\mu \ge 0$, is strictly increasing such algorithms for solving equation (7) are easy to implement. Either one uses a secant method or – if the duality mapping $J^*_{s^*}$ is supposed to be differentiable – we even can apply Newton's methods. This is e.g. the case for $\mathcal{X}^* = L^r$ with $r \ge 2$.*

6 Some numerical results

Based on two sample functions we want to compare the numerical effort of the algorithms described above in two small numerical experiments. In the first example we additionally apply Landweber iteration with constant step size $\mu_n \equiv const.$ in order to see the acceleration effect of the control of the step size. It turns out that the choice $\mu_n \equiv 1$ is too large in that situation. Therefore we have set $\mu_n \equiv 0.1$ which was motivated by the observation of the calculated step sizes μ_n of Algorithm 3.1.

a) A linear benchmark example

Here we choose $\mathcal{X} := L^{1.1}(0,1)$ and $\mathcal{Y} := L^2(0,1)$ and consider the linear operator of integration, e.g. $A : \mathcal{X} \longrightarrow \mathcal{Y}$ is given as

$$[A\,x](t) := \int_0^t x(\tau)\,d\tau, \qquad t \in [0,1].$$

Because of its simple structure this operator has been well-established as benchmark example for numerical case studies dealing with inverse problems. Moreover we set $p = 2$ which coincides with the power of convexity of $\mathcal{X}$. For discretization we divide the interval $[0,1]$ into $k = 1000$ equidistant subintervals. We set $t_j := j/k$, $0 \le j \le k$, and approximate x by n piecewise constant ansatz functions, i.e.

$$x(t) \approx \sum_{j=1}^{k} x_j \varphi_j(t), \quad t \in [0,1], \quad \text{with} \quad \varphi_j = \chi_{(t_{j-1},t_j)}, \; 1 \le j \le k.$$

If x is supposed to be continuous on the intervals (t_{j-1}, t_j) we can set e.g. $x_j := x((t_{j-1} + t_j)/2)$. Then the specific discretization implies that we have no discretization error for discretizing the exact solution $x^\dagger$. For the discretization of the data $y \in \mathcal{Y}$ we can take the function values of $y \in \mathcal{Y}$ at the end points of the n subintervals, i.e. we approximate

$$y(t) \approx \sum_{j=1}^{k} y_j \varphi_j(t), \quad t \in [0,1], \qquad \text{with} \quad y_j := y(t_j), \;\; 1 \le j \le k.$$

The corresponding discretization of the norms and duality products is induced by the specific choice of the ansatz functions $\varphi_j(t)$, $1 \le j \le k$, $t \in [0,1]$. In the numerical example we perturb the exact data with random Gaussian noise. Here δ_{rel} denotes the relative size of noise error. We deal with the functions

$$x_1^\dagger(t) := 3\,(t-0.5)^2 + 0.2, \quad t \in (0,1), \qquad \text{and} \qquad x_2^\dagger(t) := \begin{cases} 5, & t \in [0.25, 0.27], \\ -3, & t \in [0.4, 0.45], \\ 4, & t \in [0.7, 0.73], \\ 0, & \text{else.} \end{cases}$$

For the discrepancy criterion we set $\tau := 1.2$ and $x_0 \equiv 0$ is chosen as initial guess. The number of iterations was limited by $n_{max} = 10^6$.

We now turn to the numerical results. The number of iterations as well as the calculation times are presented in Table 1 for $x_1^\dagger$ and in Table 2 for the second function $x_2^\dagger$. We summarize the results:

- Even not presented here the quality of the approximate solutions $x^\delta_{N(\delta,y^\delta)}$ do not depend on the specific algorithm. In all cases the achieved results were only little worse than the Tikhonov-regularized solutions with penalty functional $P(x) := \frac{1}{q}\|x\|^q_{L^q}$, $q = 1.1$, and the regularization parameter α chosen in an optimal way (using the knowledge of $x_i^\dagger$, $i = 1,2$).

	$\mu_n = const.$		Algorithm 4.1		Algorithm 3.1	
δ_{rel}	$N(\delta, y^\delta)$	time (sec.)	$N(\delta, y^\delta)$	time (sec.)	$N(\delta, y^\delta)$	time (sec.)
0.05	863	0.85	63	0.16	28	0.18
0.01	7530	6.56	335	0.40	93	0.33
10^{-3}	79120	69.01	2065	2.29	451	2.05
10^{-4}	$> 10^6$	–	24548	26.27	2068	8.69
10^{-5}	–	–	118823	126.81	12479	49.97

Table 1: Calculation times for sample function $x_1^\dagger$

	$\mu_n = const.$		Algorithm 4.1		Algorithm 3.1	
δ_{rel}	$N(\delta, y^\delta)$	time (sec.)	$N(\delta, y^\delta)$	time (sec.)	$N(\delta, y^\delta)$	time (sec.)
0.05	4023	3.52	253	0.35	104	0.29
0.01	36720	31.98	1520	1.35	358	1.28
10^{-3}	457270	391.40	11022	12.01	963	4.17
10^{-4}	$> 10^6$	–	94315	101.77	6729	27.10
10^{-5}	–	–	606582	653.37	50890	205.01

Table 2: Calculation times for sample function $x_2^\dagger$

- Choosing a constant step size $\mu_n \equiv const.$ the number $N(\delta, y^\delta)$ of necessary iterations grows rapidly when the noise level δ becomes smaller. For a relative noise level $\delta_{rel} = 10^{-4}$ the maximal number $n_{max} = 10^6$ of iterations was exceeded for both sample functions.
- Both accelerated versions lead to a strongly decrease of the iteration numbers. In particular, for moderate noise levels $\delta_{rel} = 10^{-3} \dots 0.05$ (which are the one occurring in practical applications) the calculation times shows the good performance of the algorithms under consideration.
- For very small noise levels (or $\delta = 0$) the calculation times are still quite high. Here additional numerical stopping criterions should be tested leading to a earlier termination (which was not done here).
- The application of the forward operator A and its adjoint $A^\star$ was implemented here in an efficient way needing only $\mathcal{O}(n)$ operations. Using a matrix-vector multiplication for the implementation the differences between Algorithm 3.1 and 4.1 in time will increase since the calculation of the step size μ_n in Algorithm 3.1 uses only vector-vector operations (of order $\mathcal{O}(n)$).

Finally, we present a graphical demonstration of the effect of using Banach spaces in our considerations. Therefore we choose the Hilbert space $\mathcal{X} = L^2(0,1)$ in an alternative calculation which leads back to classical (accelerated) Landweber iteration. The regularized solutions $x^\delta_{N(\delta,y^\delta)}$ for $x_2^\dagger$ and $\delta_{rel} = 0.01$ were plotted in Figure 1. We see that the choice $\mathcal{X} = L^q(0,1)$ has the same effect as the choice of the penalty functional $P(x) := \frac{1}{q}\|x\|^q_{L^q}$ with $q = 1.1$ or $q = 2$ in the Tikhonov functional. Consequently, the numerical effort as well as the obtained regularized solutions $x^\delta_{N(\delta,y^\delta)}$ of this numerical example show that accelerated Landweber methods are an interesting alternative to Tikhonov regularization

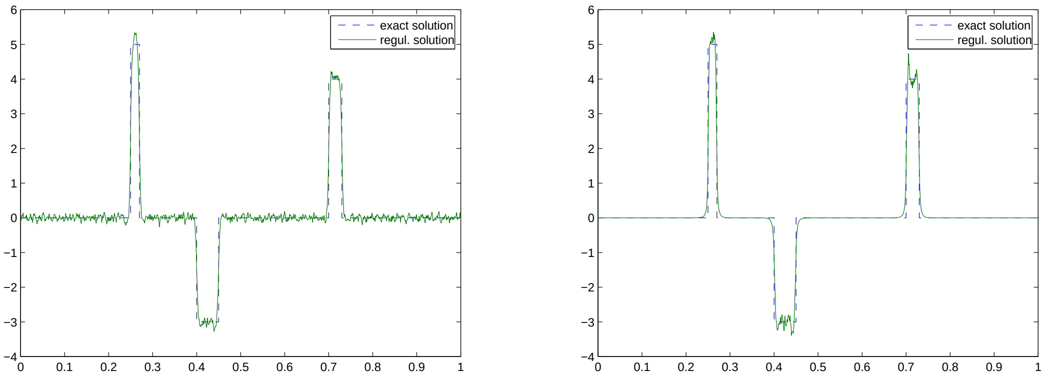

Figure 1: exact vs. regularized solution for $x_2^\dagger$ for $\mathcal{X} = L^2(0,1)$ (left plot) and $\mathcal{X} = L^{1.1}(0,1)$ (right plot)

with penalty terms based on Banach space norms.

b) A nonlinear application

We also want to demonstrate the applicability of our algorithm to nonlinear problems. We consider the following example which arises in option pricing theory, see e.g. [15]. The corresponding inverse problem was deeply studied in [7], see also the references therein for an overview about further aspects in the mathematical foundation of (inverse) option pricing. We also refer to [10] for some newer results.

Following the notation in [7] we we introduce the Black-Scholes function U_{BS} for the variables $X > 0$, $K > 0$, $r \geq 0$ and $s \geq 0$ as

$$U_{BS}(X,K,r,t,s) := \begin{cases} X\,\Phi(d_1) - K\,e^{-r\,t}\Phi(d_2), & s > 0, \\ \max\left\{X - K\,e^{-r\,t}, 0\right\}, & s = 0, \end{cases}$$

with

$$d_1 := \frac{\ln\left(\frac{X}{K}\right) + r\,t + \frac{s}{2}}{\sqrt{s}}, \quad d_2 := d_1 - \sqrt{s}$$

and $\Phi(\xi)$, $\xi \in \mathbb{R}$, denotes the cumulative density function of the standard normal distribution. We follow the generalization of the classical Black-Scholes analysis with time-dependent volatility function $\sigma(t)$, $t \geq 0$, and constant riskless short-term interest rate $r \geq 0$. Then the price $c(t)$ of a European call option on a traded asset with fixed strike price $K > 0$ as function of the maturity $t \in [0,T]$ is given by formula

$$c(t) := U_{BS}\left(X, K, r, t, \int_0^t \sigma^2(\tau)\,d\tau\right), \qquad t \in [0,T],$$

where $T > 0$ denotes the maximal time horizon of interest. Moreover, $X > 0$ denotes the actual price of the underlying asset at time $t_0 = 0$.

From the investigations in [7] we know that for $t \to 0$ some additional effects occurs which need a separate treatment. In order to keep the considerations here more simple we introduce a (small) time $t_\varepsilon > 0$ and assume the volatility to be known (and constant)

	Algorithm 4.1		Algorithm 3.1	
δ_{rel}	$N(\delta, y^\delta)$	$\frac{\|x_N^\delta - x_1^\dagger\|}{\|x_1^\dagger\|}$	$N(\delta, y^\delta)$	$\frac{\|x_N^\delta - x_1^\dagger\|}{\|x_1^\dagger\|}$
0.01	39	0.2274	14	0.2263
10^{-3}	691	0.0986	120	0.0995
10^{-4}	10098	0.0484	627	0.0497
10^{-5}	119982	0.0212	2692	0.0216

Table 3: Iteration numbers and regularization error for the nonlinear problem

on the interval $[0, t_\varepsilon]$, i.e. $\sigma(t) \equiv \sigma_0 > 0$, $t \in [0, t_\varepsilon]$. Then, for given $1 < a, b < \infty$ we define the nonlinear operator $F : \mathcal{D}(F) \subset L^a(t_\varepsilon, T) \longrightarrow L^b(t_\varepsilon, T)$ as

$$[F(x)](t) := U_{BS}\left(X, K, r, t, \sigma_0^2 t_\varepsilon + \int_{t_\varepsilon}^{t} x(\tau)\, d\tau\right), \qquad t \in [t_\varepsilon, T],$$

with domain $\mathcal{D}(F) := \{x \in L^a(t_\varepsilon, T) \ : \ x(t) \geq \underline{c} \ \text{ a.e. on } \ [t_\varepsilon, T]\}$. Then one can show that F is Fréchet differentiable for all $x \in \mathcal{D}(F)$ and the estimate

$$\|F(\tilde{x}) - F(x) - F'(x)(\tilde{x} - x)\| \leq C\, \|\tilde{x} - x\|_{L^a} \|F(\tilde{x}) - F(x)\|_{L^b}$$

holds for all $x, \tilde{x} \in \mathcal{D}(F)$ with uniform constant $C > 0$, see [9]. In particular, condition (3) holds with $L < 1$ whenever $\|\tilde{x} - x\|$ is sufficiently small.

In the specific experiment we choose the parameters

$$X = 1,\ K = 0.85,\ r = 0.05,\ t_\varepsilon = 0.1 \quad \text{and} \quad T = 1.$$

Moreover, we assume that the exact solution is given by $x_1^\dagger(t)$, $t \in [0.1, 1]$. Finally we set $\sigma_0 := x_1^\dagger(0.1) = 0.68$. As spaces we take $\mathcal{X} = L^{1.1}(t_\varepsilon, T)$ and $\mathcal{Y} = L^2(t_\varepsilon, T)$ in order to have a similar setting as in the linear example. Also the discretization is done in a similar manner dividing the interval $[0.1, 1]$ into $k = 1000$ equidistant subintervals. As initial guess we choose $x_0 \equiv 0.4$ such that we can set

$$L := \frac{1}{3} \qquad \text{and} \qquad \tau := 2.1 > \frac{1+L}{1-L} = 2.$$

The results for different (relative) noise levels are presented in Table 3 for Algorithm 4.1 and Algorithm 3.1. There the number $N(\delta, y^\delta)$ of necessary iterations as well as the relative errors of the regularized solution $x^\delta_{N(\delta, y^\delta)}$ are given. We remark the following:

- Both algorithms provide regularized solutions of similar quality.
- The number $N(\delta, y^\delta)$ of necessary iteration numbers can be dramatically reduced by applying Algorithm 3.1 also in the nonlinear case.
- The iteration number $N(\delta, y^\delta)$ is here somewhat lower than in the linear example for the sample function $x_1^\dagger$. This seems to be surprisingly on the first glance. However, the initial guess $x_0 \equiv 0.4$ is better than $x_0 \equiv 0$ which was the choice in the linear case. Moreover, we have chosen $\tau = 2.1$ instead of $\tau = 1.2$ which of course reduces the number of necessary iterations. As consequence, the reconstruction error in the linear case is smaller than in the nonlinear example.

Summarizing these results, this numerical experiment shows that we can apply this method successful also to nonlinear ill-posed problems.

7 Conclusions

We presented an accelerated Landweber iteration method for solving non-linear ill-posed operator equations in Banach spaces. Based on an auxiliary problem we proved convergence and stability of the algorithm under consideration. Even gradient-type methods are often regarded as too slow for practical applications we have demonstrated by a numerical example that including the search of an appropriate step size leads to acceptable number of necessary iterations and computational time. Because such algorithms are easy to implement we believe that accelerated Landweber approaches are a valuable tool in solving inverse problems also in Banach spaces.

References

[1] A.B. Bakushinskii and M.Y. Kokurin. *Iterative Methods for Approximate Solutions of Inverse Problems*. Springer, Dortrecht, 2004.

[2] T. Bonesky, K.S. Kazimierski, P. Maass, F. Schöpfer, and T. Schuster. Minimization of Tikhonov functionals in Banach spaces. *Abstract and Applied Analysis*, pages Article ID 192679, 19 pp. DOI:10.1155/2008/192679, 2008.

[3] I. Cioranescu. *Geometry of Banach spaces, Duality Mappings and Nonlinear Problems*. Kluwer Dortrecht, 1990.

[4] I. Daubechies, M. Defrise, and C. De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Comm. Pure Appl. Math.*, 57:1413–1457, 2004.

[5] H.W. Engl and O. Scherzer. Convergence rates results for iterative methods for solving nonlinear ill-posed problems. In D. Colton (ed.) et al, editor, *Surveys on solution methods for inverse problems*, pages 7–34. Springer, Wien, 2000.

[6] M. Hanke, A. Neubauer, and O. Scherzer. A convergence analysis of the Landweber iteration for nonlinear ill-posed problems. *Numer.Math.*, 72:21–27, 1995.

[7] T. Hein and B. Hofmann. On the nature of ill-posedness of an inverse problem arising in inverse option pricing. *Inverse Problems*, 19:1319–1338, 2003.

[8] T. Hein and B. Hofmann. Approximate source conditions for nonlinear ill-posed problems - chances and limitations. *Inverse Problems*, 25(3):Article ID 035003, 16 p., 2009.

[9] T. Hein and K.S. Kazimierski. Modified Landweber iteration in Banach spaces – convergence and convergence rates. *Preprint TU Chemnitz, Germany*, 2009-14, 2009.

[10] B. Hofmann, B. Kaltenbacher, C. Pöschl, and O. Scherzer. A convergence rates result in Banach spaces with non-smooth operators. *Inverse Problems*, 23:987–1010, 2007.

[11] B. Hofmann and P. Mathé. Analysis of profile functions for general linear regularization methods. *SIAM J.Numer.Anal.*, 45(3):1122–1141, 2007.

[12] B. Hofmann and O. Scherzer. Factors influencing the ill-posedness of nonlinear problems. *Inverse Problems*, 10:1277–1297, 1994.

[13] B. Kaltenbacher, A. Neubauer, and O. Scherzer. *Iterative Regularization Methods for Nonlinear Ill-Posed Problems*. Walter de Gruyter, Berlin, 2008.

[14] B. Kaltenbacher, Frank Schöpfer, and Thomas Schuster. Iterative methods for nonlinear ill-posed problems in Banach spaces: convergence and application to paramter identification problems. *Inverse Problems*, 25:Article ID 065003 (19pp), 2009.

[15] Y.K. Kwok. *Mathematical Models of Financial Derivatives*. Springer, Singarpore, 1998.

[16] J. Lindenstrauss and L. Tzafriri. *Classical Banach spaces II*. Springer-Verlag New York/Berlin, 1979.

[17] A. Neubauer and O. Scherzer. A convergent rate result for steepest descent method and a minimal error method for the solution of nonlinear ill-posed problems. *ZAA*, 14:369–377, 1995.

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

[19] T.I. Seidman and C.R. Vogel. Well posedness and convergence of some regularisation methods for non-linear ill-posed problems. *Inverse Problems*, 5:227–238, 1989.

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

[21] E. Zeidler. *Nonlinear Functional Analysis and its Applications III*. Springer-Verlag New York, 1985.