Modified Landweber iteration in Banach spaces – convergence and convergence rates

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

August 24, 2009

Abstract

Abstract. We introduce and discuss an iterative method of relaxed Landweber type for the regularization of the solution operator of the operator equation $F(x) = y$, where X and Y are Banach spaces and F is a non-linear, continuous operator mapping between them. We assume that the Banach space X is smooth and convex of power type. We will show that under the so-called approximate source conditions convergence rates may be achieved. 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

Running title: Regularized Landweber in Banach spaces

1 Introduction

Let X and Y be both Banach spaces. We consider the non-linear operator equation

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

where $F : \mathcal{D}(F) \subseteq X \longrightarrow Y$ describes a continuous non-linear mapping from the domain $\mathcal{D}(F)$ into the space Y. In many applications (1) turns out to be ill-posed. In particular a solution $x \in \mathcal{D}(F)$ satisfying $F(x) = y$ needs not to exist nor to be unique, and if it exists then it does not necessarily depend continuously on the data y. This means that small perturbations of the data may cause large perturbations of the solution. Usually F models some measurement system. Therefore it can be assumed that a noise cluttered

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

[†]Center for Industrial Mathematics (ZeTeM) in collaboration with SFB 747, University of Bremen, 28334 Bremen, Germany Email: `kamilk@math.uni-bremen.de;`

version y^δ of y is available where only the estimate $\|y^\delta - y\| \le \delta$ is known. This rises the need of the application of a regularization method.

In the case that both spaces X and Y are Hilbert many regularization methods have been established [11, 25, 23], see also the references therein. The most prominent classes of regularization strategies are

- regularization by minimizers of functionals, i.e. Tikhonov regularization and
- iterative regularization methods, like Landweber iteration etc.

Motivated by their successful application in image restoration, sparsity reconstruction the development and investigation of regularization methods for inverse problems in Banach spaces have become a field of modern research, cf. e.g. [8, 6, 20]. In particular, the Tikhonov regularization in Banach spaces has been established theoretically in e.g. [20] whereas their numerical treatment (for linear problems) was considered in e.g. [8, 5]. As described in [20] and references therein, many convergence rate results for Tikhonov regularization are available for linear and non-linear operators. We also refer to [15] for some newer results.

For the case that F in (1) is replaced by a linear operator A (with adjoint operator $A^\star$) an iterative regularization method which is given via

$$x_{n+1}^* \in x_n^* - \mu_n A^\star J_p(A\,x_n - y^\delta), \qquad x_{n+1} := J_q^*(x_{n+1}^*) \tag{2}$$

was recently introduced in [27]. Strong convergence of the iterates to the minimum-norm solution of (1) was proven under mild assumptions on the space X. However, from the theoretical point of view the lack of quantitative results, i.e. convergence rates, is a major drawback.

In this paper we introduce an alternative iterative method for regularization of (linear and) non-linear operator equations in Banach spaces. We assume F to be (Gâteaux-)differentiable in $\mathcal{D}(F)$ with derivative $F'(x)$. Then this method reads as

$$x_{n+1}^* := x_n^* - \mu F'(x_n)^\star J_p(F(x_n) - y^\delta) - \beta_n x_n^*, \qquad x_{n+1} := J_q^*(x_{n+1}^*), \tag{3}$$

where $J_q : Y \longrightarrow Y^*$ and $J_q^* : X^* \longrightarrow X$ denote duality mappings in the spaces under consideration. In Hilbert spaces this method is also known as modified Landweber iteration which was introduced and investigated in [26]. However, as opposite to the approach there we suggest here an a-posteriori choice of the parameter β_n. We will show that for properly chosen parameters μ and $\{\beta_n\}$ and under assumptions on the space X similar to those of [27] not only strong convergence but also convergence rates can be obtained. We point out, that to our best knowledge, this is the first time that convergence rates for Landweber-like algorithms could be proven in Banach spaces. Moreover, we can give an extension of the convergence rates results of [26] in Hilbert spaces.

The iteration of [27] can be understood as a steepest descent approach along the functional

$$\tfrac{1}{p}\|A\,x - y^\delta\|^p,$$

whereas the new iteration can be understood as the steepest descent along the Tikhonov functional

$$\tfrac{1}{p}\|F(x) - y^\delta\|^p + \beta_n \tfrac{1}{\mu\,p}\|x\|^p.$$

Therefore one can interpret our iteration as a regularized version of the iteration of [27] for nonlinear problems.

Since we also aim to prove convergence rates we need some (source) conditions describing the smoothness of a (minimum-norm) solution $x^\dagger \in \mathcal{D}(F)$ of equation (1). Here we deal with so-called approximative source condition of the form

$$J_p(x^\dagger) = F'(x^\dagger)^\star \omega + \upsilon, \qquad \|\omega\| \leq S,\ \|\upsilon\| \leq d(S),$$

for $S > 0$ where $d : [0, \infty) \longrightarrow [0, \infty)$ denotes the non-negative decreasing distance function of the element $J_p(x^\dagger)$ with respect to the set $\mathcal{R}(F'(x^\dagger)^\star)$. The idea of using the decay rate of this distance functions for proving convergence rates was developed by [19, 18] in Hilbert spaces. Since this method does not depend on Hilbert space techniques such as spectral calculus it can be applied in Banach spaces, see also [16].

Consequently the paper is organized as follows: In the next section, we introduce basic notions of geometry of Banach spaces. In Section 3 we recapitulate the idea and definitions of approximate source conditions and distance functions and derive a corresponding variational inequality which will be applied in the proofs of our main convergence rates result. In Section 4 the iteration (3) is introduced more precisely. As main result we will show that for this iteration together with approximate source condition convergence rate results can be obtained. In Section 5 we show that a slightly altered version of the iteration can be used if no approxiamte source conditions are available. We will close our discussion with theoretical and numerical analysis of an nonlinear example which arises in option pricing theory.

2 Preliminaries

Throughout the paper let $1 < p < \infty$ and $1 < q < \infty$ be conjugate exponents, i.e.

$$\tfrac{1}{p} + \tfrac{1}{q} = 1.$$

Let X and Y be Banach spaces with duals X^* and Y^*. Their norms will be denoted by $\|\cdot\|$. We omit indices indicating the space since it will become clear from the context which one is meant. For $x \in X$ and $z^* \in X^*$ we denote by $\langle x, z^* \rangle$ or $\langle z^*, x \rangle$ the duality paring, i.e. $\langle x, z^* \rangle = \langle z^*, x \rangle = z^*(x)$.

For $p > 1$ the (in general) set-valued mapping $J_p : X \rightrightarrows X^*$ defined by

$$J_p(x) = \{x^* \in X^* : \langle x^*, x \rangle = \|x\| \|x^*\|, \|x^*\| = \|x\|^{p-1}\}$$

is called a *duality mapping* of X with gauge function $t \mapsto t^{p-1}$. We remark that J_p is in general non-linear. One can show [7, Th.I.4.8] that J_2 is linear if and only if X is a Hilbert space. In the case of a Hilbert space J_2 is just the identity mapping. The space X is called smooth if $J_p(x)$ is single valued for all $x \in X$. Then for the sake of convenience we identify J_p with it's only single-valued selection.

In the convergence analysis we deal with Bregman distances. The definition is stated below.

Definition 2.1. *Let X be a smooth Banach space and $p > 1$. The Bregman distance $\Delta_p(x, \cdot)$ of the convex functional $x \mapsto \frac{1}{p}\|x\|^p$ at $x \in X$ is defined as*

$$\Delta_p(x, \tilde{x}) := \frac{1}{p}\|\tilde{x}\|^p - \frac{1}{p}\|x\|^p - \langle J_p(x), \tilde{x} - x \rangle, \qquad \tilde{x} \in X,$$

where $J_p : X \longrightarrow X^*$ denotes the duality mapping of X with gauge function $t \mapsto t^{p-1}$.

Due to the definition of duality mappings we can also rewrite the definition as

$$\begin{aligned} \Delta_p(x, \tilde{x}) &= \tfrac{1}{p}\|\tilde{x}\|^p - \tfrac{1}{p}\|x\|^p - \langle J_p(x), \tilde{x} \rangle + \|x\|^p \\ &= \tfrac{1}{p}\|\tilde{x}\|^p + \tfrac{1}{q}\|x\|^p - \langle J_p(x), \tilde{x} \rangle \end{aligned}$$

By the convexity of the functional $x \mapsto \frac{1}{p}\|x\|^p$ we have $\Delta_p(x, \tilde{x}) \geq 0$ for all $x, \tilde{x} \in X$. One can further show that the Bregman distance is continuous in both arguments [27, Theorem 2.12].

The presented convergence rate analysis is based essentially on the following assumptions.

(A1) There exists a solution $x^\dagger \in \mathcal{D}(F)$ of (1), i.e. $F(x^\dagger) = y$ holds.

(A2) The Banach space X is smooth and p-convex for some $p \geq 2$.

(A3) The exists a ball $\mathcal{B}_\varrho(x^\dagger)$ around $x^\dagger$ with radius $\varrho > 0$ and a constant $0 \leq \eta < 1$ such that for each $x \in \mathcal{D}(F) \cap \mathcal{B}_\varrho(x^\dagger)$ we can find a linear bounded operator $F'(x)$ with $\|F'(x)\| \leq K$ for some constant $K > 0$ and

$$\|F(\tilde{x}) - F(x) - F'(x)(\tilde{x} - x)\| \leq \eta \|F(\tilde{x}) - F(x)\| \tag{4}$$

for all $\tilde{x} \in \mathcal{D}(F) \cap \mathcal{B}_\varrho(x^\dagger)$.

We shortly discuss the assumptions (A2) and (A3). The space X is called convex of power type p or p-convex if there exists a constant $C > 0$ such that $\delta_X(\epsilon) \geq C\epsilon^p$, where the *modulus of convexity of X* $\delta_X : [0, 2] \to [0, 1]$ is defined via

$$\delta_X := \inf\{1 - \|\tfrac{1}{2}(x + \tilde{x})\| \ : \ \|x\| = \|\tilde{x}\| = 1, \ \|x - \tilde{x}\| \geq \epsilon\}.$$

Moreover, the space X is said to be smooth of power type q or *q-smooth* for $q > 1$ if there exists a constant $G > 0$ such that $\rho_X(\tau) \leq G\tau^q$, where $\rho_X : [0, \infty) \to [0, \infty)$ – called the *modulus of smoothness of X* – is defined via

$$\rho_X = \tfrac{1}{2} \sup\{\|x + \tilde{x}\| + \|x - \tilde{x}\| - 2 \ : \ \|x\| = 1, \ \|\tilde{x}\| \leq \tau\}.$$

As a consequence of the Lindenstrauss duality formula X is p-convex if and only if X^* is q-smooth and X is q-convex if and only if X^* is p-smooth, (cf. [9, IV.1.7 and IV.1.12]). Due to the polarization identity every Hilbert space is 2-smooth and 2-convex. Further for $1 < p < \infty$ the sequence spaces ℓ_p, Lebesgue spaces L_p and Sobolev spaces W_p^m are $\min\{2, p\}$-smooth and $\max\{2, p\}$-convex [13, 30, 7].

By (A2) and due to the Xu-Roach inequalities [30], we conclude from the p-convexity of X that there exists a constant $C_p > 0$ such that the estimate

$$\Delta_p(x, \tilde{x}) = \frac{1}{p}\|\tilde{x}\|^p - \frac{1}{p}\|x\|^p - \langle J_p(x), \tilde{x} - x\rangle \geq \frac{C_p}{p}\|x - \tilde{x}\|^p \tag{5}$$

holds for all $x, \tilde{x} \in X$. From the q-smoothness of the dual space X^* we derive the existence of a constant $G_q > 0$ such that

$$\frac{1}{q}\|\tilde{x}^*\|^q - \frac{1}{q}\|x^*\|^q - \langle J_q(x^*), \tilde{x}^* - x^*\rangle \leq \frac{G_q}{q}\|x^* - \tilde{x}^*\|^q \tag{6}$$

for all $x^*, \tilde{x}^* \in X^*$. We will use the relations (5) and (6) in our convergence analysis. In particular, the constants G_q and C_p will play the role of fixed parameters in the algorithm under consideration.

The condition (4) is also known as *η-condition* which is in fact a restriction of the nonlinearity of the operator F. It was originally introduced in [12] for proving convergence and convergence rates for Landweber iteration for nonlinear ill-posed problems in Hilbert spaces. There, the authors supposed $\eta < \frac{1}{2}$ for their analysis. For our considerations the weaker requirement $\eta < 1$ is sufficient. The same condition was applied in [26] for proving convergence of the modified Landweber iteration. In order to prove convergence rates for Landweber iteration the additional structural condition

$$F'(x) = R_x F'(x^\dagger) \quad \text{with} \quad \|R_x - I\| \leq C\,\|x - x^\dagger\|, \quad R_x : Y \longrightarrow Y$$

with $\|x - x^\dagger\|$ sufficiently small and constant $C > 0$ was introduced in [12]. On the other hand, in [26] only Litschitz continuity of the Fréchet-derivative was supposed which is a weaker restriction than (A3). In order to understand the necessity of the different strength of the applied assumptions we briefly recall convergence rates results for iteratively regularized Gauß-Newton methods, see e.g. [1] and [4] as well as [2] and [23] for a further discussion of conditions for iterative regularization methods of nonlinear problems in Hilbert spaces. There, for an a-priori guess $x_0 \in X$ the power-type source condition

$$x^\dagger - x_0 \in \mathcal{R}\left((F'(x^\dagger)^* F'(x^\dagger))^\nu\right)$$

for some exponent $0 < \nu \leq 1$ is considered. Then it is essential to distinguish between $\nu \geq \frac{1}{2}$, see [1], and $0 < \nu < \frac{1}{2}$, see [4]. For $\nu \geq \frac{1}{2}$ the Lipschitz continuity of the Fréchet-derivative $F'(x)$, $x \in \mathcal{D}(F)$, is sufficient to prove accordant convergence rates. But it has been well-established that this condition is not sufficient in the case $0 < \nu < \frac{1}{2}$. As consequence, in [4] was dealt with the (more general) structural condition

$$\begin{aligned} F'(\bar{x}) &= R(\bar{x}, x)F'(x) + Q(\bar{x}, x) \\ \|I - R(\bar{x}, x)\| &\leq C_R \\ \|Q(\bar{x}, x)\| &\leq C_Q\|F'(x^\dagger)(\bar{x} - x)\| \end{aligned}$$

with sufficiently small parameters C_R and C_Q and operators $R(\cdot,\cdot), Q(\cdot,\cdot)$ depending on $\bar{x}, x \in X$. In particular, $C_R + C_Q\|x - x^\dagger\| < \frac{1}{2}$ implies automatically the condition (A3). Since in [26] only the case $\nu = \frac{1}{2}$ was considered, the Lipschitz continuity of the Fréchet-derivative was sufficient in the corresponding convergence rates analysis. The presented convergence rates analysis here deals with the situation $0 < \nu \le \frac{1}{2}$ and show that condition (A3) is sufficient to prove accordant convergence results. Moreover, the presented convergence rates analysis includes also logarithmic source conditions which was treated in [22] separately for iteratively regularized Gauß-Newton methods.

For linear operators we can set $\eta = 0$. So linear inverse problems can be considered as special case of the subsequent analysis. With $F'(x)^\star : Y^* \longrightarrow X^*$ we denote the adjoint operator of $F'(x)$ which is defined by

$$\langle y^*, F'(x)\,\tilde{x}\rangle = \langle F'(x)^\star y^*, \tilde{x}\rangle, \qquad \tilde{x} \in X,\ y^* \in Y^*.$$

3 Approximate source conditions

Since our aim is also to prove convergence rates we need a source condition. Therefore, we introduce the notions of *approximate source conditions* and *distance functions*. Originally, approximate source conditions were established in [3] for proving error estimates for linear regularization methods in Hilbert spaces. In combination with distance functions in [18] a concept was developed for proving convergence rates based on approximate source conditions. Since no Hilbert space tools such as spectral calculus is needed the idea can be transferred to regularization approaches in Banach spaces, see e.g. [14] for some initial convergence rates results. We define distance functions.

Definition 3.1. *The distance function $d(\cdot;\xi) : [0,\infty) \longrightarrow \mathbb{R}$ of an element $\xi \in X^*$ with respect to the set $\mathcal{R}(F'(x^\dagger)^\star)$ is defined as*

$$d(S;\xi) := \inf\left\{\|\xi - F'(x^\dagger)^\star\omega\| \ :\ \|\omega\| \le S\right\}, \qquad S \ge 0. \tag{7}$$

Since the space X is supposed to be smooth of power type and hence X^* is convex of power type (and therefore in particular strictly convex) we can replace the infimum by the minimum. In particular, for each $S \ge 0$ there exist elements $\omega = \omega(S)$ and $\upsilon = \upsilon(S)$ such that

$$\xi = F'(x^\dagger)^\star\omega + \upsilon, \qquad \|\omega\| \le S,\ \|\upsilon\| = d(S;\xi), \tag{8}$$

holds. We refer to conditions of form (8) as *approximate source conditions*. We will emphasize the following properties of distance functions: distance functions $d(\cdot;\xi)$ are non-negative and non-increasing. They are strictly positive, if $\xi \notin \mathcal{R}(F'(x^\dagger)^\star)$. If $\xi \in \mathcal{R}(F'(x^\dagger)^\star)$, in particular if the 'classical' source condition

$$\xi = F'(x^\dagger)^\star\omega, \qquad \omega \in Y^*, \tag{9}$$

holds, then $d(S;\xi) = 0$ for all $S \ge \|\omega\|$. On the other hand, if $\xi \notin \mathcal{R}(F'(x^\dagger)^\star)$ but $\xi \in \overline{\mathcal{R}(F'(x^\dagger)^\star)}$ then $d(S;\xi)$ tends to zero as $S \to \infty$.

Of interest for our considerations is the following result.

Lemma 3.2. *Assume (A1)-(A3) and $J_p(x^\dagger)$ has distance function $d(S) = d(S; J_p(x^\dagger))$, $S \ge 0$. Then the estimate*

$$\begin{aligned} -\langle J_p(x^\dagger), x - x^\dagger\rangle \ \le\ & S(1+\eta)\|F(x) - y^\delta\| + S(1+\eta)\delta \\ & + \left(1 - \tfrac{1}{q}\right)\Delta_p(x, x^\dagger) + \tfrac{1}{q}\left(\tfrac{p}{C_p}\right)^{\frac{1}{p-1}} d(S)^q. \end{aligned} \tag{10}$$

holds for all $S \geq 0$ and all $x \in \mathcal{X}$. The same estimate holds if the Bregman distance $\Delta_p(x, x^\dagger)$ is replaced by $\Delta_p(x^\dagger, x)$.

Proof. We set $T(x, x^\dagger) := F(x) - F(x^\dagger) - F'(x^\dagger)(x - x^\dagger)$. By assumption the approximate source condition (8) holds for all $S \geq 0$ with $\xi = J_p(x^\dagger)$. We derive

$$\begin{aligned}
-\langle J_p(x^\dagger), x - x^\dagger \rangle &= -\langle w(S), F'(x^\dagger)(x - x^\dagger) \rangle - \langle \upsilon(S), x - x^\dagger \rangle \\
&= -\langle w(S), F(x) - y^\delta + y^\delta - y - T(x, x^\dagger) \rangle - \langle \upsilon(S), x - x^\dagger \rangle \\
&\leq S\|F(x) - y^\delta\| + S\,\delta + S\|T(x, x^\dagger)\| + d(S)\|x - x^\dagger\| \\
&\leq S(1+\eta)\|F(x) - y^\delta\| + S(1+\eta)\delta + d(S)\left(\tfrac{p}{C_p}\right)^{\frac{1}{p}} \Delta_p(x, x^\dagger)^{\frac{1}{p}}
\end{aligned}$$

and finally (10) by applying Young's inequality to the last term. □

The concept of introducing source conditions as variational inequalities was introduces in [20], see also the discussion of some consequences therein. The variational inequality (10) holds also in more general situations than (8). So, approximate source conditions can be considered as kind of motivation for formulating the condition (10).

In order to achieve convergence of the algorithm we further have to suppose that $d(S; J_p(x^\dagger))$ tends to zero as $S \to \infty$, i.e.

$$J_p(x^\dagger) \in \overline{\mathcal{R}(F'(x^\dagger)^\star)} \tag{11}$$

holds. On the first moment it looks like only a little extension of convergence rates results assuming the source condition (9). On the other hand, if $F'(x^\dagger)$ is injective then $\overline{\mathcal{R}(F'(x^\dagger)^\star)} = X$, see e.g. [29, Satz II.4.5]. Moreover, from [27, Lemma 2.10] we observe that at least for linear operator equations the minimum-norm-solution $x^\dagger$ always satisfies (11).

4 Convergence and Convergence Rates Results

We consider the following algorithm.

Algorithm 4.1. *Let y^δ be the given data with $\|y^\delta - y\| \leq \delta$ and $d_0 : [0, \infty) \longrightarrow [0, \infty)$ some non-negative non-increasing function with $d_0(S) \to 0$ as $S \to \infty$ and η, ϱ and K be constants of (A3). Let the iterative regularization algorithm be given via*

(S_0) *Init. Choose $\mu > 0$ such that*

$$(\eta - 1)\mu + 2^q \tfrac{G_q}{q} K^q \mu^q \leq 0,$$

start point x_0^*, $x_0 := J_q(x_0^*)$ such that $\Delta_p(x_0, x^\dagger) \leq R_0$ and $\Delta_p(x_0, x^\dagger) \leq \varrho^p \frac{C_p}{p}$. We choose $\bar{S} = \bar{S}(\delta)$ so big, that

$$d_0(\bar{S})^q \leq \delta \cdot \bar{S} \tag{12}$$

and we further choose $\tau > 0$. Set

$$C_1 := \tfrac{1}{q}\left(\tfrac{2(1+\eta)^p}{\mu(1-\eta)}\right)^{q-1} + \left(\tfrac{2(1+\eta)^q}{1-\eta}\right)^{p-1} \tfrac{\mu}{p}\tau^q + \left(1 + \eta + \tfrac{1}{q}\left(\tfrac{p}{C_p}\right)^{\frac{1}{p}}\right)\tau^{q-1}.$$

(S_1) *Compute*

$$\begin{array}{rcl}\gamma_n & := & 2^{q-1}G_q\|x_n^*\|^q + C_1\bar{S}^q \\ \beta_n & := & \min\left\{q, (R_n/(q\gamma_n))^{p-1}\right\} \\ R_{n+1} & := & \left(1-\frac{\beta_n}{q}\right)R_n + \frac{\gamma_n}{q}\beta_n^q.\end{array}$$

STOP, if $\tau\beta_n\bar{S} \le \delta^{p-1}$ *and* $\beta_n < q$.

(S_2) *Compute the new iterate via*

$$\begin{array}{l} x_{n+1}^* := x_n^* - \mu F'(x_n)^\star J_p(F(x_n)-y^\delta) - \beta_n x_n^* \\ x_{n+1} := J_q(x_{n+1}^*)\end{array}$$

Set $n \leftarrow (n+1)$ *and go to step* (S_1).

Then the following convergence results hold:

Theorem 4.2. *Assume (A1)-(A3),* $\delta > 0$ *and* $J_p(x^\dagger)$ *has distance function* $d(S; J_p(x^\dagger))$, $S \ge 0$. *We further assume that the iterates* $\{x_n\}$ *remain in* $\mathcal{D}(F)$. *If the function* d_0 *is chosen such that* $d(S; J_p(x^\dagger)) \le d_0(S)$ *for all* $S \ge 0$ *then the following hold:*

(i) The iterates $\{x_n\}$ *remain in* $\mathcal{B}_\varrho(x^\dagger)$.

(ii) The iteration terminates after a finite number of iterations, i.e. the stopping index $k(\delta) = k(\delta, y^\delta)$ *is well-defined.*

(iii) There exists a number $C = C(\delta_{\max}) > 0$ *such that for all* $\delta \le \delta_{\max}$ *the estimate*

$$\Delta_p(x_{k(\delta)}, x^\dagger) \le C\,\delta\,\bar{S}(\delta) \tag{13}$$

holds.

If in particular $\bar{S}(\delta)$ *is chosen such that* $\delta\,\bar{S}(\delta) \to 0$ *as* $\delta \to 0$ *then we have convergence* $x_{k(\delta)} \to x^\dagger$ *as* $\delta \to 0$.

Proof. To shorten the notation we set $\Delta_n := \Delta_p(x_n, x^\dagger)$, $A_n := F'(x_n)$ and $A := F'(x^\dagger)$. By definition of the Bregman distance we have

$$\Delta_{n+1} := \tfrac{1}{q}\|x_{n+1}^*\|^q - \langle x_{n+1}^*, x^\dagger\rangle + \tfrac{1}{p}\|x^\dagger\|^p.$$

By the q-smoothness of X^* we have

$$\begin{array}{rcl}\frac{1}{q}\|x_{n+1}^*\|^q & \le & \frac{1}{q}\|x_n^*\|^q + \langle x_n, -\mu\, A_n^\star J_p(F(x_n)-y^\delta) - \beta_n x_n^*\rangle \\ & & +\frac{G_q}{q}\|\mu\, A_n^\star J_p(F(x_n)-y^\delta) + \beta_n x_n^*\|^q.\end{array}$$

This gives

$$\begin{array}{rcl}\Delta_{n+1} & \le & \Delta_n + \langle x_n - x^\dagger, -\mu\, A_n^\star J_p(F(x_n)-y^\delta) - \beta_n x_n^*\rangle \\ & & +\frac{G_q}{q}\|\mu\, A_n^\star J_p(F(x_n)-y^\delta) + \beta_n x_n^*\|^q.\end{array}$$

Furthermore, we have with $T(x^\dagger, x_n) := F(x^\dagger) - F(x_n) - A_n(x^\dagger - x_n)$

$$
\begin{aligned}
&\langle x_n - x^\dagger, -\mu\, A_n^\star J_p(F(x_n) - y^\delta) - \beta_n x_n^* \rangle \\
&\quad = -\mu\langle F(x_n) - y^\delta, J_p(F(x_n) - y^\delta)\rangle + \mu\langle T(x^\dagger, x_n), J_p(F(x_n) - y^\delta)\rangle \\
&\qquad -\mu\langle y^\delta - y, J_p(F(x_n) - y^\delta)\rangle \\
&\qquad -\beta_n\langle x_n - x^\dagger, x_n^* - J_p(x^\dagger)\rangle - \beta_n\langle x_n - x^\dagger, J_p(x^\dagger)\rangle
\end{aligned}
$$

The first term generates a negative term since

$$
-\mu\langle F(x_n) - y^\delta, J_p(F(x_n) - y^\delta)\rangle = -\mu\|F(x_n) - y^\delta\|^p.
$$

The second term provides

$$
\begin{aligned}
\mu\left|\langle T(x^\dagger, x_n), J_p(F(x_n) - y^\delta)\rangle\right| &\le \mu\eta\|F(x_n) - y\|\|F(x_n) - y^\delta\|^{p-1} \\
&\le \mu\eta\|F(x_n) - y^\delta\|^p + \mu\eta\delta\|F(x_n) - y^\delta\|^{p-1}.
\end{aligned}
$$

The third term can be estimated as

$$
\mu\left|\langle y^\delta - y, J_p(F(x_n) - y^\delta)\rangle\right| \le \mu\delta\|F(x_n) - y^\delta\|^{p-1}.
$$

The forth term allows us to generate a second negative term, since

$$
-\beta_n\langle x_n - x^\dagger, x_n^* - J_p(x^\dagger)\rangle = -\beta_n\langle x_n - x^\dagger, J_p(x_n) - J_p(x^\dagger)\rangle = -\beta_n\Delta_n - \beta_n\Delta_p(x^\dagger, x_n) \le -\beta_n\Delta_n
$$

Finally, for the last term we use the source condition (10) to get

$$
\begin{aligned}
-\beta_n\langle x_n - x^\dagger, J_p(x^\dagger)\rangle \le\ & \bar{S}(1+\eta)\|F(x_n) - y^\delta\|\beta_n + \bar{S}(1+\eta)\delta\,\beta_n \\
& + \left(1 - \tfrac{1}{q}\right)\Delta_n\beta_n + \tfrac{1}{q}\left(\tfrac{p}{C_p}\right)^{\frac{1}{p-1}} d_0(\bar{S})^q\beta_n.
\end{aligned}
$$

With the estimations above we have

$$
\begin{aligned}
\Delta_{n+1} \le\ & \left(1 - \tfrac{\beta_n}{q}\right)\Delta_n - \mu(1-\eta)\|F(x_n) - y^\delta\|^p + \mu(1+\eta)\delta\|F(x_n) - y^\delta\|^{p-1} \\
& + \bar{S}(1+\eta)\|F(x_n) - y^\delta\|\beta_n + \bar{S}\,\delta(1+\eta)\beta_n + \tfrac{1}{q}\left(\tfrac{p}{C_p}\right)^{\frac{1}{p-1}} d_0(\bar{S})^q\beta_n \\
& + \tfrac{G_q}{q}\|\mu\, A_n^\star J_p(F(x_n) - y^\delta) + \beta_n x_n^*\|^q.
\end{aligned}
$$

Assume now, that the iteration did not stop in the last step. Since the stopping criterion is not fulfilled for x_n we have $\delta^{p-1} < \tau\beta_n\bar{S}$. Moreover we have chosen $\bar{S}$ such that $\bar{S}\delta \ge d_0(\bar{S})^q$, therefore we derive

$$
\begin{aligned}
\bar{S}(1+\eta)\delta\,\beta_n + \tfrac{1}{q}\left(\tfrac{p}{C_p}\right)^{\frac{1}{p}} d_0(\bar{S})^q\beta_n &= \bar{S}\,\beta_n\left((1+\eta)\delta + \tfrac{1}{q}\left(\tfrac{p}{C_p}\right)^{\frac{1}{p}}\tfrac{d_0(\bar{S})^q}{\bar{S}}\right) \\
&\le \left(1 + \eta + \tfrac{1}{q}\left(\tfrac{p}{C_p}\right)^{\frac{1}{p}}\right)\bar{S}\,\beta_n\delta \\
&\le \left(1 + \eta + \tfrac{1}{q}\left(\tfrac{p}{C_p}\right)^{\frac{1}{p}}\right)\tau^{\frac{1}{p-1}}(\bar{S}\,\beta_n)^{1+\frac{1}{p-1}} \\
&= \left(1 + \eta + \tfrac{1}{q}\left(\tfrac{p}{C_p}\right)^{\frac{1}{p}}\right)\tau^{q-1}(\bar{S}\,\beta_n)^q.
\end{aligned}
$$

Due to the Young inequality we get

$$\begin{aligned}\mu(1+\eta)\|F(x_n)-y^\delta\|^{p-1}\delta &\leq \mu(1+\eta)\|F(x_n)-y^\delta\|^{p-1}\left(\tau\,\beta_n\bar{S}\right)^{\frac{1}{p-1}} \\ &\leq \tfrac{\mu(1-\eta)}{2\,q}\|F(x_n)-y^\delta\|^p+\tfrac{\mu}{p}\left(\tfrac{2(1+\eta)^q}{1-\eta}\right)^{p-1}\left(\tau\,\beta_n\bar{S}\right)^q\end{aligned}$$

and

$$(1+\eta)\|F(x_n)-y^\delta\|\beta_n\bar{S}\leq\tfrac{\mu(1-\eta)}{2\,p}\|F(x_n)-y^\delta\|^p+\tfrac{1}{q}\left(\tfrac{2(1+\eta)^p}{\mu(1-\eta)}\right)^{\frac{1}{p-1}}\left(\beta_n\bar{S}\right)^q.$$

The last term we estimate by Jensen inequality

$$\begin{aligned}\tfrac{G_q}{q}\|\mu\,A_n^\star J_p(F(x_n)-y^\delta)+\beta_n x_n^*\|^q &\leq 2^q\tfrac{G_q}{q}\left(\tfrac{1}{2}\|\mu\,A_n^\star J_p(F(x_n)-y^\delta)\|+\tfrac{1}{2}\|\beta_n x_n^*\|\right)^q \\ &\leq 2^{q-1}\tfrac{G_q}{q}\mu^q\|A_n\|^q\|F(x_n)-y^\delta\|^p+2^{q-1}\tfrac{G_q}{q}\|x_n^*\|^q\beta_n^q.\end{aligned}$$

We arrive at

$$\begin{aligned}\Delta_{n+1}\leq\left(1-\tfrac{\beta_n}{q}\right)\Delta_n+\tfrac{\|F(x_n)-y^\delta\|^p}{2}\left(-\mu(1-\eta)+2^q\tfrac{G_q}{q}\mu^q\|A_n\|^q\right)+2^{q-1}\tfrac{G_q}{q}\|x_n^*\|^q\beta_n^q \\ +\left(\tfrac{1}{q}\left(\tfrac{2(1+\eta)^p}{\mu(1-\eta)}\right)^{\frac{1}{p-1}}+\left(\tfrac{2(1+\eta)^q}{1-\eta}\right)^{p-1}\tfrac{\mu}{p}\tau^q+\left(1+\eta+\tfrac{1}{q}\left(\tfrac{p}{C_p}\right)^{\frac{1}{p}}\right)\tau^{q-1}\right)\bar{S}^q\beta_n^q.\end{aligned}$$

Now the term in the first bracket is negative or zero by the initial assumption on μ. Therefore, the last estimation simplifies to

$$\Delta_{n+1}\leq\left(1-\tfrac{\beta_n}{q}\right)\Delta_n+2^{q-1}\tfrac{G_q}{q}\|x_n^*\|^q\beta_n^q+C_1\bar{S}^q\beta_n^q=\left(1-\tfrac{\beta_n}{q}\right)\Delta_n+\tfrac{\gamma_n}{q}\beta_n^q.$$

What we have proven may be seen as a part of an induction to show that $\Delta_n\leq R_n$ for all n. We have by assertion $\Delta_0\leq R_0$ and assuming $\Delta_n\leq R_n$ we have proven that $\Delta_{n+1}\leq R_{n+1}$. We also have $R_{n+1}\leq R_n$ by the choice of β_n. Hence the sequence $\{R_n\}$ is monotonically decreasing (and bounded, therefore convergent). We show next that $\{R_n\}$ is a zero sequence. The very same trick as used below to show convergence may also be found in e.g. [10, 6]. We define

$$\Gamma:=\sup\{2^{q-1}\tfrac{G_q}{q}\|J_p(x)\|^q+C_1\bar{S}^q\;:\;\Delta_p(x,x^\dagger)\leq R_0\}.$$

By the coercivity of Bregman distances we have $\Gamma<\infty$, therefore

$$R_{n+1}^{1-p}-R_n^{1-p}\geq\left(\tfrac{\beta_n}{q}-\tfrac{\gamma_n}{R_n}\beta_n^q\right)R_n^{1-p}\geq\tfrac{1}{q}\min\{R_0^{1-p},1/(q^q\Gamma)^{p-1}\}>0.$$

We have

$$R_n^{1-p}\geq\sum_{k=0}^{n-1}R_{k+1}^{1-p}-R_k^{1-p}\geq n\left(\tfrac{1}{q}\min\{R_0^{1-p},1/(q^q\Gamma)^{p-1}\}\right).$$

Therefore

$$R_n\leq q^{\frac{1}{p-1}}\max\{R_0,\Gamma q^q\}\cdot n^{-\frac{1}{p-1}}.$$

We define another auxiliary number

$$\gamma := C_1 \bar{S}^q.$$

It is clear that $0 < \gamma \le \gamma_n \le \Gamma < \infty$ for all n. We have then

$$\begin{aligned}\beta_n &\le 1/(q^q\gamma_n)^{p-1}\, R_n^{p-1} \le q/(q^q\gamma_n)^{p-1} \max\{R_0, \Gamma\, q^q\}^{p-1} n^{-1}\\ &\le q\, \max\{(R_0/(q\gamma)^{p-1}, (\Gamma/\gamma)^{p-1} q\} n^{-1}\end{aligned}$$

Hence $\{\beta_n\}$ is a zero sequence and the iteration is well defined. Let $k(\delta) = k(\delta, y^\delta)$ be the index, where the iteration terminates. Then we have that $\beta_{k(\delta)} = (R_{k(\delta)}/(q\gamma_{k(\delta)}))^{p-1}$ and $\tau\beta_{k(\delta)}\bar{S} \le \delta^{p-1}$. Further there exists a constant $c = c(\delta_{\max}) > 0$ such that $\bar{S}(\delta) \ge c$ for all $\delta \le \delta_{\max}$. Hence there exists a number $C_2 \ge C_1$ such that $\Gamma \le C_2 \bar{S}(\delta)^q$ for all $\delta \le \delta_{\max}$. Therefore

$$\Delta_{k(\delta)} \le R_{k(\delta)} = q\,\gamma_{k(\delta)} \beta_{k(\delta)}^{\frac{1}{p-1}} \le q\Gamma\, \tau^{-\frac{1}{p-1}}\, \bar{S}^{-\frac{1}{p-1}}\, \delta \le C_2 q\, \tau^{-\frac{1}{p-1}}\, \delta\, \bar{S}(\delta),$$

which proves the theorem. □

We assume now that the distance function is known. Based on the convexity estimate (5) we are now able to present the following conclusion which can be considered as an a-priori convergence rates result.

Corollary 4.3. *Under the condition of Theorem 4.2 we assume $d_0(S) = d(S; J_p(x^\dagger))$. Then the following holds:*

(i) If $J_p(x^\dagger) \in \mathcal{R}(F'(x^\dagger)^\star)$, i.e. $J_p(x^\dagger) = F'(x^\dagger)^\star\omega$ for some $\omega \in \mathcal{Y}^*$, then the choice of $\bar{S}$ as $\bar{S} \ge \|\omega\|$ independent of the noise level δ leads to the convergence rate

$$\|x_{k(\delta)} - x^\dagger\| \le C\,\sqrt{\delta}. \tag{14}$$

(ii) If $J_p(x^\dagger) \in \overline{\mathcal{R}(F'(x^\dagger)^\star)} \setminus \mathcal{R}(F'(x^\dagger)^\star)$ then the choice $\bar{S}(\delta) := \Phi^{-1}(\delta)$ with function $\Phi(S) := d_0(S)^q S^{-1}$ leads to the optimal convergence rate

$$\|x_{k(\delta)} - x^\dagger\| \le C\left(\delta\,\Phi^{-1}(\delta)\right)^{\frac{1}{p}} = C\, d_0\left(\Phi^{-1}(\delta)\right)^{\frac{1}{p-1}}. \tag{15}$$

Remark 4.4. *Notice that $\Phi^{-1}(\delta) \to \infty$ as $\delta \to 0$ if the source condition (9) is violated. Therefore the best rate that can be achieved with this iteration is $\mathcal{O}(\sqrt{\delta})$ and it is achieved in the case that the condition (9) is not violated.*

In order to understand these convergence rates we compare them with known results for Tikhonov regularization in Banach spaces.

Example 4.5. *Instead of the iterative procedure we consider Tikhonov regularization*

$$\tfrac{1}{p}\|F(x) - y^\delta\|^p + \tfrac{\alpha}{p}\|x\|^p \to \min \quad \textit{subject to } x \in \mathcal{D}(F).$$

We again assume that there exists an linear bounded operator $F'(x^\dagger)$ which satisfies the η-condition (4) for arbitrary $\eta > 0$. We have then the following convergence rates results

[16, Theorem 5.2 and 5.3]: If $J_p(x^\dagger) \in \mathcal{R}(F'(x^\dagger)^\star)$ then an a-priori parameter choice $\alpha \sim \delta^{p-1}$ leads to the convergence rate (14). On the other hand, for $J_p(x^\dagger) \in \overline{\mathcal{R}(F'(x^\dagger)^\star)} \setminus \mathcal{R}(F'(x^\dagger)^\star)$ with distance function $d(S) = d(S; J_p(x^\dagger))$ we have to introduce the functions $\Theta(S) := d(S)^p S^{-p}$, $\Psi(\alpha) := \alpha^{\frac{1}{p}} d\left(\Theta^{-1}(\alpha)\right)^{\frac{1}{p-1}}$ and $\Phi(S) := d(S)^{\frac{p}{p-1}} S^{-1} = d(S)^q S^{-1}$. Then an a-priori choice $\alpha := \Psi^{-1}(\delta)$ gives the convergence rate (15). We remark, that due to the monotonicity of the function $d(S)$ the inverse functions $\Theta^{-1}(\alpha)$, $\Psi^{-1}(\delta)$ and $\Phi^{-1}(\delta)$ are well-defined. In particular, Tikhonov regularization and the suggested iterative regularization approach provide the same convergence rates for proper chosen parameters.

Moreover, it also turns out, that these convergence rates can be considered as optimal if X and Y are assumed to be Hilbert spaces at least if we can suppose power-type distance functions.

Example 4.6. *Let X and Y be Hilbert spaces, i.e. we have $p = q = 2$. Moreover, we assume $x^\dagger \notin \mathcal{R}(F'(x^\dagger)^*) = \mathcal{R}\left((F'(x^\dagger)^* F'(x^\dagger))^{\frac{1}{2}}\right)$ but the element $x^\dagger$ satisfies a weaker source condition, i.e. $x^\dagger = f(F'(x^\dagger)^* F'(x^\dagger))\,\omega$ for some index function $f(t)$, $t \geq 0$, and some $\omega \in X$. Then we can rewrite this source condition also as approximate source condition with respect to the stronger (violated) source condition with corresponding distance function $d(S; x^\dagger)$, $S \geq 0$. Then – according [21, Theorem 5.9] – we can give an upper bound $\bar{d}(S)$, $S \geq 0$, for the distance function which allows us to apply the results of corollary 4.3(ii). We present two examples.*

(a) If $f(t) = t^\nu$ for some $0 < \nu < \frac{1}{2}$ then we have the estimate

$$d(S) \leq \kappa\, S^{\frac{2\nu}{2\nu-1}}, \qquad \kappa := \|\omega\|,$$

see also [19, Theorem 1] in the compact case. This and $p = 2$ imply the choice

$$\bar{S}^{\frac{4\nu}{2\nu-1}} \sim \delta\,\bar{S} \;\Leftrightarrow\; \delta \sim \bar{S}^{\frac{2\nu+1}{2\nu-1}} \;\Leftrightarrow\; \bar{S} \sim \delta^{\frac{2\nu-1}{2\nu+1}}$$

for the parameter $\bar{S}$. Then we get from (15) that

$$\|x_{k(\delta,y^\delta)} - x^\dagger\| \leq C\,\delta^{\frac{2\nu}{2\nu+1}}$$

which is optimal in this context.

(b) We assume a logarithmic source condition, i.e. $f(t) = (-\ln t)^{-\mu}$ for some $\mu > 0$ and $\|F'(x^\dagger)\| \leq 1$. Then we cannot state this upper bound $\bar{d}(S)$ for the distance function explicitly. However, from the consideration in [21] we know the following property: introducing a (regularization) parameter α then the (well-defined) choice $S = S(\alpha)$ such that $\sqrt{\alpha}S(\alpha) = \bar{d}(S(\alpha))$ gives the equality

$$\sqrt{\alpha}S(\alpha) = \bar{d}(S(\alpha)) = \kappa\,(-\ln\alpha)^{-\mu}.$$

For logarithmic source conditions the choice $\alpha \sim \delta$ for the regularization parameter is suggested. We set $\alpha := \delta$ Then we get the choice

$$(-\ln\delta)^{-2\mu} \sim \delta\,\bar{S} \;\Leftrightarrow\; \bar{S} \sim \delta^{-1}(-\ln\delta)^{-2\mu}$$

which ends up at the known convergence rate

$$\|x_\alpha^\delta - x^\dagger\| \leq C\,(-\ln\delta)^{-\mu}.$$

We also want to point out that the right choice of $d_0(S)$ and hence $\bar{S}$ is important for the convergence and the speed of convergence of the algorithm. In particular, if $\bar{S}$ is chosen too small then the algorithm might not work. Therefore we present the following indicator which helps to decide if the parameter $\bar{S}$ was chosen properly.

Lemma 4.7. *Under the assumptions of Theorem 4.2 we have*

$$\tfrac{C_p}{p}\|x_{n+1}-x_n\|^p + C_\mu\|F(x_n)-y^\delta\|^p \le 2^{p-1}\tfrac{2p-1}{p}R_n. \tag{16}$$

with $C_\mu := \frac{\mu(1-\eta)}{2} - 2^{q-1}\frac{G_q}{q}\mu^q K^q \ge 0.$

Proof. By assumption $\Delta_n \le R_n$ holds. Following the proof of Theorem 4.2 we start with

$$\begin{aligned}\tfrac{C_p}{p}\|x_{n+1}-x_n\|^p &\le \tfrac{C_p}{p}\left(\|x_{n+1}-x^\dagger\| + \|x_n - x^\dagger\|\right)^p \\ &\le 2^{p-1}\left(\Delta_{n+1}+\Delta_n\right) \\ &\le 2^{p-1}\left[\left(2-\tfrac{\beta_n}{q}\right)R_n + \beta_n\tfrac{\gamma_n}{q}\beta_n^{\frac{1}{p-1}} - C_\mu\|F(x_n)-y^\delta\|^p\right],\end{aligned}$$

since $q-1=\frac{1}{p-1}$. By construction we have $\beta_n^{\frac{1}{p-1}} \le \frac{R_n}{q\,\gamma_n}$ and $\beta_n \le q$. Hence, we continue

$$\begin{aligned}\tfrac{C_p}{p}\|x_{n+1}-x_n\|^p + C_\mu\|F(x_n)-y^\delta\|^p &\le 2^{p-1}\left[\left(2-\frac{\beta_n}{q}\right)R_n + \tfrac{\beta_n R_n}{q^2}\right] \\ &\le 2^{p-1}\left(2-\tfrac{\beta_n}{q\,p}\right)R_n \\ &\le 2^{p-1}\left(2-\tfrac{1}{p}\right)R_n = \tfrac{2^{p-1}(2p-1)}{p}R_n.\end{aligned}$$

This proves the estimate. □

The estimate (16) gives a necessary condition for the validity of the assumptions of Theorem 4.2 and hence on the proper choice of the function $d_0(S)$ and the parameter $\bar{S}$. A violation of (16) shows that the the decay rate of $d_0(S)$ is too high and hence $\bar{S}$ is chosen to small. On the other hand, if the right side of (16) is much smaller than the left hand side it might be an indication that the decay rate of $d_0(S)$ is too low and consequently $\bar{S}$ is chosen to large. This causes lower speed of convergence. So, checking this inequality (16) during the iteration and readjusting the parameter $\bar{S}$ in a proper way can help to ensure convergence and improve speed of convergence. It is still a topic on ongoing research of developing an a-posteriori choice of the function $d_0(S)$ leading to the optimal convergence rates (14) and (15), respectively.

5 A convergence analysis without source condition

The convergence rates results essentially base on the proper choice of the function $d_0(S)$, $S \ge 0$. We therefore present here a convergence analysis which does not depend on such assumption. Basically, it combines the convergence analysis of [26] and [27]. In order to prove convergence we need the following slight modifications of the algorithm:

(i) Choose μ such that

$$C_\mu := (1-\eta)\mu + 2^q \tfrac{G_q}{q} K^q \mu^q > 0.$$

(ii) Choose $C_s > 0$ sufficiently large such that $C_s \geq d_0(0) \geq \|J_p(x^\dagger)\|^q = \|x^\dagger\|^p$ and choose $\bar{S}$ as $\bar{S} := C_s\delta^{-1}$.

(iii) Choose a decreasing sequence $\{\bar{\beta}_j\}$ such that $\bar{\beta}_j > 0$ for all $j \geq 0$ and $\sum_{j=0}^\infty \bar{\beta}_j < \infty$. Calculate

$$\begin{aligned}
\gamma_n &:= \max\left\{2^{q-1}G_q\|x_n^*\| + C_1\bar{S}^q, R_n/(q\,\bar{\beta}_n^{\frac{1}{p-1}})\right\}, \\
\beta_n &:= \min\left\{q, (R_n/(q\gamma_n))^{p-1}\right\} \qquad \text{and} \\
R_{n+1} &:= \left(1 - \tfrac{\beta_n}{q}\right) R_n - \tfrac{\|F(x_n)-y^\delta\|^{p-1}}{2} C_\mu + \tfrac{\gamma_n}{q}\beta_n^q.
\end{aligned}$$

The suggested modifications do not annihilate the results of Theorem 4.2. Moreover, the modified calculation of the parameters has the following consequence: we have $\beta_n \leq \bar{\beta}_n$ for all $n \geq 0$ and $\beta_n = \bar{\beta}_n$ if and only if $\gamma_n = R_n/(q\,\bar{\beta}_n^{\frac{1}{p-1}})$. In particular, we bound the parameter β_n from above in each iteration step by $\bar{\beta}_n$.

Since by (ii) we have $\bar{S}(\delta) = C_s/\delta$ the estimate (13) is only a boundedness estimate. To establish convergence $x_{k(\delta)} \to x^\dagger$ as $\delta \to 0$ we have to carry out completely new analysis. First we prove the following:

Lemma 5.1. *Under the above modifications we have*

$$\sum_{n=0}^{k(\delta)-1} \|F(x_n) - y^\delta\|^p \leq \tfrac{2\,R_0}{C_\mu} \qquad \textit{and} \qquad k(\delta) \leq 1 + C\,\delta^{-p}$$

for some constant $C > 0$.

Proof. By construction we have

$$R_{n+1} := \left(1 - \tfrac{\beta_n}{q}\right) R_n - \tfrac{\|F(x_n)-y^\delta\|^{p-1}}{2} C_\mu + \tfrac{\gamma_n}{q}\beta_n^q \leq R_n - \tfrac{\|F(x_n)-y^\delta\|^{p}}{2} C_\mu$$

for all $n \leq k(\delta)$ which proves the first part. From the stopping criterion we derive for all $n < k(\delta)$ that $\beta_n \leq C\,n^{-1}$ and hence

$$\delta^{p-1} \leq \tau\,\beta_n\bar{S} \leq \tau\,C\,C_s\delta^{-1}n^{-1}$$

or equivalently $n \leq \tau\,C\,C_s\delta^{-p}$ which proves the second part by choosing $n = k(\delta) - 1$. □

We now discuss the noiseless case $\delta = 0$. For $\delta \to 0$ we have $\gamma_n \to \infty$ and hence $\beta_n \to 0$. Hence we set $\beta_n \equiv 0$ for given noiseless data.

Theorem 5.2. *Assume $\delta = 0$. Then the algorithm stops after a finite number k of iterations with $\tilde{x}^\dagger := x_k$ or we have convergence $x_n \to \tilde{x}^\dagger$ as $n \to \infty$. In both cases the element $\tilde{x}^\dagger$ denotes any solution of equation (1) in $\mathcal{B}_\varrho(x^\dagger)$, i.e. we have $F(\tilde{x}^\dagger) = y$.*

Proof. Let the iteration process does not stop after a finite number of steps. From the proof of theorem 4.2 with $\beta_n = 0$ we derive

$$\Delta_{n+1} \leq \Delta_n - \tfrac{\|F(x_n)-y\|^p}{2} C_\mu$$

for all $n \geq 0$. Hence we proved $\sum_{n=1}^{\infty} \|F(x_n) - y\|^p < \infty$ and consequently $F(x_n) \to y$ as $n \to \infty$. Now we can choose indices $k > l$ such that $\|F(x_n) - y\| \geq \|F(x_k) - y\|$ for all $l \leq n \leq k$. Then we derive

$$\begin{aligned}
\left|\langle J_p(x_k) - J_p(x_l), x_k - x^\dagger \rangle\right| &= \left| \sum_{n=l}^{k-1} \langle J_p(x_{n+1}) - J_p(x_n), x_k - x^\dagger \rangle \right| \\
&\leq \left| \sum_{n=l}^{k-1} \mu \langle J_p(F(x_n) - y), F'(x_n)(x_k - x_n + x_n - x^\dagger) \rangle \right| \\
&\leq \mu \sum_{n=l}^{k-1} \|F(x_n) - y\|^{p-1} (1+\eta) \left(\|F(x_n) - y\| + \|F(x_n) - F(x_k)\| \right) \\
&\leq \mu \sum_{n=l}^{k-1} \|F(x_n) - y\|^{p-1} (1+\eta) \left(2\|F(x_n) - y\| + \|F(x_k) - y\| \right) \\
&\leq 3\,\mu \sum_{n=l}^{k-1} \|F(x_n) - y\|^{p-1} (1+\eta) \|F(x_n) - y\| \\
&= 3\,\mu\,(1+\eta) \sum_{n=l}^{k-1} \|F(x_n) - y\|^p .
\end{aligned}$$

For $l \to \infty$ the right hand side goes to zero. Following the argumentation in [27] we have $\{x_n\}$ to be a Cauchy sequence and hence $x_n \to \tilde{x}^\dagger \in X$. By the continuity of F and $F(x_n) \to y$ we have $F(\tilde{x}^\dagger) = y$ which shows that $\tilde{x}^\dagger$ is a solution. □

We now can prove convergence.

Theorem 5.3. *We have $x_{k(\delta)} \to \tilde{x}^\dagger$ where $\tilde{x}^\dagger$ denotes any solution of equation (1) in $\mathcal{B}_\varrho(x^\dagger)$, i.e. we have $F(\tilde{x}^\dagger) = y$.*

Proof. Let $\{\delta_j\}$ be a sequence of noise levels with $\delta_j \to 0$ as $j \to \infty$. Without loss of generality we suppose $k(\delta_j) \to \infty$ as $j \to \infty$. Otherwise we see that the iterates depend continuously on y^{δ_j}. From lemma 5.1 we conclude $\|F(x_{k(\delta_j)}) - y^{\delta_j}\| \to 0$ as $j \to \infty$ and hence $F(x_{k(\delta_j)}) \to y$ as $j \to \infty$. Let be j fixed and $\delta = \delta_j$. Thus we can find $l < k \leq k(\delta)$

with $\|F(x_n) - y^\delta\| \geq \|F(x_k) - y^\delta\|$ for all $l \leq n \leq k$. This gives

$$\begin{aligned}
\left|\left\langle J_p(x_k) - J_p(x_l^\delta), x_k - x^\dagger \right\rangle\right| &= \left|\sum_{n=l}^{k-1} \left\langle J_p(x_{n+1}) - J_p(x_n), x_k - x^\dagger \right\rangle\right| \\
&= \left|\sum_{n=l}^{k-1} \left\langle \mu F'(x_n)^\star J_p(F(x_n) - y^\delta) + \beta_n\, J_p(x_n), x_k - x^\dagger) \right\rangle\right| \\
&\leq \mu \sum_{n=l}^{k-1} \left|\left\langle J_p(F(x_n) - y^\delta), F'(x_n)(x_k - x^\dagger) \right\rangle\right| \\
&\qquad + \sum_{n=l}^{k-1} \beta_n \left|\left\langle J_p(x_n), x_k - x^\dagger \right\rangle\right| .
\end{aligned}$$

We show that the right hand side goes to zero as $k(\delta) \to \infty$ and $l \to \infty$. Therefore we estimate

$$\begin{aligned}
&\sum_{n=l}^{k-1} \left|\left\langle J_p(F(x_n) - y^\delta), F'(x_n)(x_k - x^\dagger) \right\rangle\right| \\
&\quad \leq 3\,(1+\eta) \sum_{n=l}^{k-1} \|F(x_n) - y^\delta\|^{p-1} \|F(x_n) - y\| \\
&\quad \leq 3\,\mu\,(1+\eta) \sum_{n=l}^{k-1} \|F(x_n) - y^\delta\|^{p-1} \left(\|F(x_n) - y^\delta\| + \delta\right) \\
&\quad \leq 3\,\mu\,(1+\eta) \left(\sum_{n=l}^{k-1} \|F(x_n) - y^\delta\|^{p} + \sum_{n=l}^{k-1} \|F(x_n) - y^\delta\|^{p-1} \delta\right) .
\end{aligned}$$

The first sum tends to zeros as $l \to \infty$. For the second sum we apply Young's inequality to derive

$$\begin{aligned}
\sum_{n=l}^{k-1} \|F(x_n) - y^\delta\|^{p-1} \delta &\leq \tfrac{C_l}{q} \sum_{n=l}^{k-1} \|F(x_n) - y^\delta\|^{p} + \tfrac{1}{p} C_l^{-\frac{p}{q}} \sum_{n=l}^{k-1} \delta^p \\
&\leq \tfrac{C_l}{q} \sum_{n=l}^{k-1} \|F(x_n) - y^\delta\|^{p} + \tfrac{C}{p} C_l^{1-p}
\end{aligned}$$

with arbitrary constant $C_l > 0$. Here we additionally used that $k - 1 \leq k(\delta) - 1 \leq C\,\delta^{-p}$. We now choose

$$C_l := \left(\sum_{n=l}^{k-1} \|F(x_n) - y^\delta\|^p\right)^{-\frac{1}{2}} .$$

Then we see that both summands on the right hand side of the last inequality vanish as

$l \to \infty$. On the other hand we derive with some generic constant $\mathcal{C} > 0$ that

$$\begin{aligned}\sum_{n=l}^{k-1} \beta_n \left|\left\langle J_p(x_n), x_k - x^\dagger \right\rangle\right| &\leq \sum_{n=l}^{k-1} \beta_n \|J_p(x_n)\| \, \|x_k - x^\dagger\| \\ &\leq \mathcal{C} \left(\frac{p}{C_p}\right)^{\frac{1}{p}} \sum_{n=l}^{k-1} \beta_n \Delta_k^{\frac{1}{p}} \\ &\leq \mathcal{C} \sum_{n=l}^{k-1} \beta_n \leq \mathcal{C} \sum_{n=l}^{k-1} \bar{\beta}_n \to 0 \quad \text{as} \quad l \to \infty.\end{aligned}$$

We can again apply the argumentation of [27] to prove the assertion. $\square$

6 A nonlinear example

We start with the following example which arises in option pricing theory, see e.g. [24]. The corresponding inverse problem was deeply studied in [17], see also the references therein for an overview about further aspects in the mathematical foundation of (inverse) option pricing. We also refer to [20] for some newer results.

A European call option on a traded asset is a contract which gives the holder the right to buy the asset at time (maturity) $t > 0$ for a fixed strike price $K > 0$ independent on the actual asset price at time $t > 0$. For fixed current asset price $X > 0$ and time $t_0 = 0$ we denote with $c(t)$ the (fair) price of such call option with maturity $t \geq 0$. Following 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 $r \geq 0$ 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. Then the price of the option as function of the maturity $t \in [0, T]$ is given by the 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.

From the investigations in [17] 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) 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 $\mathcal{D}(F) := \{x \in L^a(t_\varepsilon, T) \;:\; x(t) \geq \underline{c}$ a.e. on $[t_\varepsilon, T]\}$. We further use the notation $k(t,s) := U_{BS}(X, K, r, t, s)$ and assume $x_0 \in \mathcal{D}(F)$. For some radius $\varrho > 0$ and arbitrary given $x \in \mathcal{D}(F) \cap \mathcal{B}_\varrho(x_0)$ we set $\Delta x := x - x_0$ and

$$\Delta s(t) := \int_{t_\varepsilon}^{t} \Delta x(\tau)\, d\tau, \qquad t \in [t_\varepsilon, T].$$

Then we have for $t \in [t_\varepsilon, T]$ that

$$|\Delta s(t)| \leq \|1\|_{L^{a/(a-1)}(t_\varepsilon,t)} \|\Delta x\|_{L^a} = (t - t_\varepsilon)^{\frac{a-1}{a}} \|\Delta x\|_{L^a} \leq (t - t_\varepsilon)^{\frac{a-1}{a}} \varrho.$$

Moreover we set

$$s_0(t) := \sigma_0^2 t_\varepsilon + \int_{t_\varepsilon}^{t} x_0(\tau)\, d\tau, \qquad t \in [t_\varepsilon, T].$$

Hence we observe that for $t \in [t_\varepsilon, T]$ we can estimate

$$\tilde{c}(t) := \max\left\{\sigma_o^2 t_\varepsilon + \underline{c}\,(t - t_\varepsilon), s_0(t) - \varrho\,(t - t_\varepsilon)^{\frac{a-1}{a}}\right\} \leq [s_0 + \tau\, \Delta s](t) \leq s_0(t) + \varrho\,(t - t_\varepsilon)^{\frac{a-1}{a}}$$

for all $\tau \in [0,1]$. Furthermore we see with $\nu := \ln\left(\frac{X}{K}\right)$ that

$$k_s(t,s) = \frac{X}{2\sqrt{2\,\pi\, s}} \exp\left(-\frac{(\nu + r\,t)^2}{2\,s} - \frac{\nu + r\,t}{2} - \frac{s}{8}\right) > 0$$

and

$$k_{ss}(t,s) = -\frac{X}{4\sqrt{2\,\pi\,s}} \left(-\frac{(\nu + r\,t)^2}{s^2} + \frac{1}{4} + \frac{1}{s}\right) \exp\left(-\frac{(\nu + r\,t)^2}{2\,s} - \frac{\nu + r\,t}{2} - \frac{s}{8}\right).$$

We can estimate

$$\begin{aligned}
|k_{ss}(t, [s_0 + \tau\,\Delta s](t))| \;&\leq\; \frac{X}{4\sqrt{2\,\pi\,\tilde{c}(t)}} \left(\frac{(\nu + r\,t)^2}{\tilde{c}(t)^2} + \frac{1}{4} + \frac{1}{\tilde{c}(t)}\right) \\
&\qquad \cdot \exp\left(-\frac{(\nu + r\,t)^2}{2\,(s_0(t) + \varrho\,(t - t_\varepsilon)^{\frac{a-1}{a}})} - \frac{\nu + r\,t}{2} \right. \\
&\qquad\qquad \left. - \frac{s_0(t) - \varrho\,(t - t_\varepsilon)^{\frac{a-1}{a}}}{8}\right) \\
&=: \; C_1(t)
\end{aligned}$$

and

$$\begin{aligned}
k_s(t, [s_0 + \tau\,\Delta s](t)) \;&\geq\; \frac{X}{2\sqrt{2\,\pi (s_0(t) + \varrho\,(t - t_\varepsilon)^{\frac{a-1}{a}})}} \\
&\qquad \cdot \exp\left(-\frac{(\nu + r\,t)^2}{2\,\tilde{c}(t)} - \frac{\nu + r\,t}{2} - \frac{s_0(t) + \varrho\,(t - t_\varepsilon)^{\frac{a-1}{a}}}{8}\right) \\
&=: \; C_2(t).
\end{aligned}$$

We consider the quotient $C_1(t)/C_2(t)$. Here we obtain

$$\frac{C_1(t)}{C_2(t)} = I_1 \exp\left(\frac{(\nu + r\,t)^2}{2} I_2 + \frac{\varrho\,(t-t_\varepsilon)^{\frac{a-1}{a}}}{4}\right)$$

with

$$\begin{aligned} I_1 &:= \frac{1}{2}\sqrt{\frac{s_0(t) + \varrho\,(t-t_\varepsilon)^{\frac{a-1}{a}}}{\tilde{c}(t)}}\left(\frac{(\nu + r\,t)^2}{\tilde{c}(t)^2} + \frac{1}{4} + \frac{1}{\tilde{c}(t)}\right) \\ &\leq \frac{1}{2}\sqrt{1 + \frac{2\,\varrho\,(t-t_\varepsilon)^{\frac{a-1}{a}}}{\tilde{c}(t)}}\left(\frac{(\nu + r\,t)^2}{\tilde{c}(t)^2} + \frac{1}{4} + \frac{1}{\tilde{c}(t)}\right) \\ &\leq \frac{1}{2}\sqrt{1 + \frac{2\,\varrho\,(T-t_\varepsilon)^{\frac{a-1}{a}}}{\sigma_0^2 t_\varepsilon}}\left(\frac{(\nu + r\,T)^2}{\sigma_0^4 t_\varepsilon^2} + \frac{1}{4} + \frac{1}{\sigma_0^2 t_\varepsilon}\right) \end{aligned}$$

and

$$I_2 := \frac{s_0(t) + \varrho\,(t-t_\varepsilon)^{\frac{a-1}{a}} - \tilde{c}(t)}{\tilde{c}(t)(s_0(t) + \varrho\,(t-t_\varepsilon)^{\frac{a-1}{a}})} \leq \frac{2\,\varrho\,(t-t_\varepsilon)^{\frac{a-1}{a}}}{\sigma_0^2 t_\varepsilon(s_0(t) + \varrho\,(t-t_\varepsilon)^{\frac{a-1}{a}})} \leq \frac{2\,\varrho\,(T-t_\varepsilon)^{\frac{a-1}{a}}}{\sigma_0^4 t_\varepsilon^2}.$$

This gives

$$\left\|\frac{C_1}{C_2}\right\|_{L^\infty} := \max_{t\in[t_\varepsilon,T]} I_1 \exp\left(\frac{(\nu + r\,t)^2}{2} I_2 + \frac{\varrho\,(t-t_\varepsilon)^{\frac{a-1}{a}}}{4}\right) < \infty.$$

Due to the mean value theorem and the positivity of k_s we have

$$\begin{aligned} |[F(x_0 + \Delta x) - F(x_0)](t)| &= |[k(t, [s_0 + \Delta s](t)) - k(t, s_0(t))| \\ &= k_s(t, [s_0 + \tau\,\Delta s](t))\,|\Delta s(t)| \\ &\geq C_2(t)|\Delta s(t)| \end{aligned}$$

a.e. on $[t_\varepsilon, T]$. This leads with $T(x_0) := F(x_0 + \Delta x) - F(x_0) - F'(x_0)\,\Delta x$ to

$$\begin{aligned} |[T(x_0)](t)| &= \left|\int_0^1 (k_s(t, [s_0 + \tau\,\Delta s](t)) - k_s(t, s_0(t)))\,\Delta s(t)\,d\tau\right| \\ &\leq C_1(t)\frac{|\Delta s(t)|^2}{2} \\ &\leq \frac{C_1(t)(T-t_\varepsilon)^{\frac{a-1}{a}}\|\Delta x\|_{L^a}}{2\,C_2(t)}\,|[F(x_0 + \Delta x) - F(x_0)](t)| \end{aligned}$$

a.e. on $[t_\varepsilon, T]$. This proves

$$\|T(x_0)\|_{L^b} \leq \frac{(T-t_\varepsilon)^{\frac{a-1}{a}}}{2}\left\|\frac{C_1}{C_2}\right\|_{L^\infty}\|\Delta x\|_{L^a}\|F(x_0 + \Delta x) - F(x_0)\|_{L^b}.$$

In particular, condition (A3) holds for $\varrho > 0$ sufficiently small.

7 Numerical results

In this section we present numerical results for the the nonlinear operator considered in the last section. 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

$$x^\dagger = (t - 0.5)^2 + 0.1, \qquad t \in [0.1, 1].$$

Finally we set $\sigma_0 := x^\dagger(0.1) = 0.26$. As spaces we take $X = L^{1.2}(t_\varepsilon, T)$ and $Y = L^2(t_\varepsilon, T)$, where we have chosen $X = L^{1.2}$ arbitrary as an example of a non-Hilbert space. For discretization we introduce $t_j := 0.1 + j/(0.9\,N)$, $0 \le j \le N = 500$, and use the values

$$y_j := [F(x^\dagger)](t_j), \qquad 1 \le j \le N,$$

as given discretized data $\underline{y} = (y_1, \ldots, y_N)^T \in \mathbb{R}^N$. The solutions x are discretized as piecewise constant functions on the the subintervals (t_{j-1}, t_j), $1 \le j \le N$. The initial guess $x_0 \in \mathcal{D}(F)$ was chosen such that we can set $\eta := 0.5$.

For our experiments we have chosen relative error in the range $10^{-5}...10^{-2}$. In a first variant we used the algorithm without assuming an approximate source condition. Hence we set $\bar{S} \sim \delta$. In order to ensure convergence we additionally introduced the bounds $\bar{\beta}_n := c^n$ with $c := 0.99$ for the parameters β_n, $n \ge 0$. For the second calculations we applied some knowledge about the approximate source condition and the corresponding distance function.

We set $\bar{S} \sim \delta^{-\frac{1}{3}}$ due to following arguments: The operator F is differentiable in $x^\dagger$ and the derivative $F'(x^\dagger) : X \longrightarrow Y$ is given via

$$[F'(x^\dagger)\,h](t) := k_s(t, [A\,x^\dagger](t))\,[A\,h](t), \qquad [A\,h](t) := \int_{t_\varepsilon}^{t} h(\tau)\,d\tau, \quad t \in [0.1, 1],$$

for all $h \in X$. We also observe that by the specific choice of the exact solution we have $x^\dagger$, $J_p(x^\dagger) \in L^2(0.1, 1)$ and the structure of the derivatives imply that all iterates x_n belongs to $L^\infty(0.1, 1)$. In particular, $x_n \in L^2(0.1, 1)$ holds. In that case we can apply lemma 3.2 also with the Hilbert space setting $X = Y = L^2(0.1, 1)$. In the Hilbert space setting $A : L^2(0.1, 1) \longrightarrow L^2(0.1, 1)$, we know from [19] that $x^\dagger \in \mathcal{R}((A^*A)^\nu)$ for all $\nu < \frac{1}{4}$. Since we have $k_s(\cdot, A\,x^\dagger) \in L^\infty(0.1, 1)$ and the function $k_s(\cdot, A\,x^\dagger)$ is bounded away from zero the mapping $z \mapsto k_s(\cdot, A\,x^\dagger)\,z$, $z \in L^2(0.1, 1)$ is continuously invertible. Therefore we can assume that the source condition is determined by the operator A. Keeping in mind Example 4.6(a) with $\nu = \frac{1}{4}$ we obtain the already mentioned choice $\bar{S} \sim \delta^{-\frac{1}{3}}$. Moreover, from the theoretical considerations we expect the convergence rate $\|x_{k(\delta)} - x^\dagger\| \sim \delta^{\frac{1}{3}}$.

The numerical results are presented in Figure 1 and Table 1. By comparing both calculations we see that the assumed approximate source condition in fact improves the achieved reconstruction of the exact solution $x^\dagger$. The convergence rate in that case is close to the expected rate $\delta^{\frac{1}{3}}$, too. On the other hand, by introducing the upper bounds $\bar{\beta}_n$, $n \ge 0$, we can expect from the results of [26] a similar convergence rate in that situation at least in a Hilbert space setting.

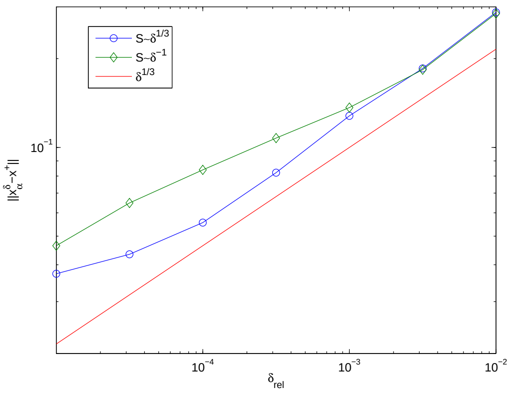

Figure 1: Relative reconstruction error depending on the noise level δ

	$\bar{S} \sim \delta^{-1}$	$\bar{S} \sim \delta^{-\frac{1}{3}}$
δ_{rel}	$\frac{\lVert x_{k(\delta)} - x^\dagger \rVert}{\lVert x^\dagger \rVert}$	$\frac{\lVert x_{k(\delta)} - x^\dagger \rVert}{\lVert x^\dagger \rVert}$
10^{-2}	0.2845	0.2873
10^{-3}	0.1365	0.1280
10^{-4}	0.0839	0.0556
10^{-5}	0.0464	0.0373

Table 1: Relative reconstruction error depending on the noise level δ

8 Conclusions

We point out, that the main results of this paper are the theoretical results contained in Section 6 iteration is computable.

If we have the information that exact source condition $J_q(x^\dagger) = F'(x^\dagger)^\star \omega$ for some $\omega \in Y^*$ is fulfilled and we can give some reasonable bound on $\|\omega\|$ then the method in this paper should be used to ensure that this information really translates into convergence rates.

If only the weaker generalized source conditions are fulfilled, but at the same time some more information (i.e. the information on $d(S)$) is given then again our method may be used for regularization, since again the information on the source condition is translated into convergence rates.

However, since the method of this paper can be considered as gradient method for minimizing Tikhonov functionals with fixed step width (but varying regularization parameter α) the convergence of the algorithm turns out to be rather slow in some examples. In particular the method introduced by [27] or the newly presented accelerated versions of [28] SESOP and RESOP promise a faster convergence. However, no convergence rates for $\delta \to 0$ are presented therein. So it part of ongoing research to improve the speed of convergence of the suggested algorithm.

References

[1] A. Bakushinskii, *The problem of the convergence of the iteratively regularized Gauss-Newton method*, Comput.Math.Math.Phys., 32 (1992), pp. 1353–1359.

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

[3] J. Baumeister, *Stable Solutions of Inverse Problems*, Vieweg, Braunschweig, 1987.

[4] B. Blaschke, A. Neubauer, and O. Scherzer, *On convergence rates for the iteratively regularized Gauss-Newton method*, IMA J.Numer.Anal., 17 (1997), pp. 421–436.

[5] T. Bonesky, K. Kazimierski, P. Maass, F. Schöpfer, and T. Schuster, *Minimization of Tikhonov functionals in Banach spaces*, to appear in Abstract and Applied Analysis, (2007).

[6] K. Bredies and D. A. Lorenz, *Iterated hard shrinkage for minimization problems with sparsity constraints.* To appear in *SIAM Journal on Scientific Computing*, 2007.

[7] I. Cioranescu, *Geometry of Banach spaces, duality mappings and nonlinear problems*, Kluwer, Dordrecht, 1990.

[8] I. Daubechies, M. Defrise, and C. D. Mol, *An iterative thresholding algorithm for linear inverse problems with a sparsity constraint*, Communications in Pure and Applied Mathematics, 57 (2004), pp. 1413–1457.

[9] R. Deville, G. Godefroy, and V. Zizler, *Smoothness and Renormings in Banach spaces*, Pitman Monographs and Surveys in Pure and Applied Mathematics, Longman Scientific & Technical, Harlow, UK, 1993.

[10] J. C. Dunn, *Global and asymptotic convergence rate estimates for a class of projected gradient processes*, SIAM Journal on Control and Optimization, 19 (1981), pp. 368–400.

[11] H. W. Engl, M. Hanke, and A. Neubauer, *Regularization of Inverse Problems*, vol. 375 of Mathematics and its Applications, Kluwer Academic Publishers Group, Dordrecht, 2000.

[12] M. Hanke, A. Neubauer, and O. Scherzer, *A convergence analysis of the Landweber iteration for nonlinear ill-posed problems*, Numerische Mathematik, 72 (1997), pp. 21–37.

[13] O. Hanner, *On the uniform convexity of L_p and ℓ_p*, Ark. Mat., 3 (1956), pp. 239–244.

[14] T. Hein, *Convergence rates for regularization of ill-posed problems in Banach spaces by approxinative source coniditions*, Inverse Problems, 24 (2008), pp. Article ID 045007, 10 pages.

[15] ——, *On Tihkonov regularization in Banach spaces - optimal convergence rates*, Applicable Analysis, 88 (2009). to appear.

[16] ——, *Regularization in Banach spaces – convergence rates by approxinative source conditions*, J.Inv.Ill-Posed Problems, 17 (2009), pp. 27–42.

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

[18] B. Hofmann, *Approximate source conditions in Tikhonov-Phillips regularization and consequences for inverse problems with multiplication operators*, Mathematical Methods in the Applied Sciences, 29 (2006), pp. 351–371.

[19] B. Hofmann, D. Düvelmeyer, and K. Krumbiegel, *Approximate source conditions in regularization – new analytical results and numerical studies*, Journal Mathematical Modelling and Analysis, 11 (2006), pp. 41–56.

[20] B. Hofmann, B. Kaltenbacher, C. Pöschl, and O. Scherzer, *A convergence rates result for Tikhonov regularization in Banach spaces with non-smooth operators*, Inverse Problems, 23 (2007), pp. 987–1010.

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

[22] T. Hohage, *Logarithmic convergence rates of the iteratively regularized Gauss-Newton method for an inverse potential and an inverse scattering problem*, Inverse Problems, 13 (1997), pp. 1279–1299.

[23] B. Kaltenbacher, A. Neubauer, and O. Scherzer, *Iterative Regularization Methods for Nonlinear Ill-posed problems*, de Gruyter, Berlin, 2008.

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

[25] A. K. Louis, *Inverse und schlecht gestellte Probleme*, B.G. Teubner, 1989.

[26] O. Scherzer, *A modified Landweber iteration for solving parameter estimation problems*, Appl.Math.Optim., 38 (1998), pp. 45–68.

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

[28] F. Schöpfer, T. Schuster, and A. Louis, *Metric and Bregman projections onto affine subspaces and their computation via sequential subspace optimization methods*, Journal of Inverse and Ill-Posed Problems, 16 (2008), pp. 479–506.

[29] D. Werner, *Funktionalanalysis* (in German), Berlin, Springer, 1997.

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