Research Article

Minimization of Tikhonov Functionals in Banach Spaces

Thomas Bonesky,[1] Kamil S. Kazimierski,[1] Peter Maass,[1] Frank Schöpfer,[2] and Thomas Schuster[2]

[1] *Center for Industrial Mathematics, University of Bremen, Bremen 28334, Germany*
[2] *Fakultät für Maschinenbau, Helmut-Schmidt-Universität, Universität der Bundeswehr Hamburg, Holstenhofweg 85, Hamburg 22043, Germany*

Correspondence should be addressed to Thomas Schuster, schuster@hsu-hh.de

Tikhonov functionals are known to be well suited for obtaining regularized solutions of linear operator equations. We analyze two iterative methods for finding the minimizer of norm-based Tikhonov functionals in Banach spaces. One is the steepest descent method, whereby the iterations are directly carried out in the underlying space, and the other one performs iterations in the dual space. We prove strong convergence of both methods.

1. Introduction

This article is concerned with the stable solution of operator equations of the first kind in Banach spaces. More precisely, we aim at computing a solution $x \in X$ of

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

for a linear, continuous mapping $A : X \to Y$, where X and Y are Banach spaces and $y \in Y$ denotes the measured data which are contaminated by some noise $\eta \in Y$. There exists a large variety of regularization methods for (1.1) in case that X and Y are Hilbert spaces such as the truncated singular value decomposition, the Tikhonov-Phillips regularization, or iterative solvers like the Landweber method and the method of conjugate gradients. We refer to the monographs of Louis [1], Rieder [2], Engl et al. [3] for a comprehensive study of solution methods for inverse problems in Hilbert spaces.

The development of explicit solvers for operator equations in Banach spaces is a current field of research which has great importance since the Banach space setting allows for dealing

with inverse problems in a mathematical framework which is often better adjusted to the requirements of a certain application. Alber [4] established an iterative regularization scheme in Banach spaces to solve (1.1) where particularly $A : X \to X^*$ is a monotone operator. In case that $X = Y$, Plato [5] applied a linear Landweber method together with the discrepancy principle in order to get a solution to (1.1) after a discretization. Osher et al. [6] developed an iterative algorithm for image restoration by minimizing the BV norm. Butnariu and Resmerita [7] used Bregman projections to obtain a weakly convergent algorithm for solving (1.1) in a Banach space setting. Schöpfer et al. [8] proved strong convergence and stability of a nonlinear Landweber method for solving (1.1) in connection with the discrepancy principle in a fairly general setting where X has to be smooth and uniformly convex.

The idea of this paper is to get a solver for (1.1) by minimizing a Tikhonov functional where we use Banach space norms in the data term as well as in the penalty term. Since we only consider the case of exact data we put $\eta = 0$ in (1.1). That means that we investigate the problem

$$\min_{x \in X} \Psi(x), \tag{1.2}$$

where the Tikhonov functional $\Psi : X \to \mathbb{R}$ is given by

$$\Psi(x) = \frac{1}{r} \|Ax - y\|_Y^r + \alpha \frac{1}{p} \|x\|_X^p, \tag{1.3}$$

with a continuous linear operator $A : X \to Y$ mapping between two Banach spaces X and Y.

If X and Y are Hilbert spaces, many results exist for problem (1.2) concerning solution methods, convergence, and stability of them and parameter choice rules for α can be found in the literature. In case that only Y is a Hilbert space, this problem has been thoroughly studied and many solvers have been established; see [9, 10]. A possibility to get an approximate solution for (1.2) is to use the steepest descent method. Assume for the moment that both X and Y are Hilbert spaces and $r = p = 2$. Then Ψ is Gâteaux differentiable and the steepest descent method applied to (1.2) coincides with the well-known Landweber method

$$x_{n+1} = x_n - \mu_n \nabla \Psi(x_n) = x_n - \mu_n A^*(Ax_n - y). \tag{1.4}$$

This iterative method converges to the unique minimizer of problem (1.2), if the stepsize μ_n is chosen properly.

In the present paper, we consider two generalizations of (1.4). First we notice that the natural extension of the gradient $\nabla \Psi$ for convex, but not necessarily smooth, functionals Ψ is the notion of the subdifferential $\partial \Psi$. We will elaborate the details later, but for the time being we note that $\partial \Psi$ is a set-valued mapping, that is, $\partial \Psi : X \rightrightarrows X^*$. Here we make use of the usual notation in the context of convex analysis, where $f : X \rightrightarrows Y$ means a mapping f from X to 2^Y. We then consider the formally defined iterative scheme

$$\begin{aligned} x_{n+1}^* &= x_n^* - \mu_n \psi_n \quad \text{with } \psi_n \in \partial \Psi(x_n), \\ x_{n+1} &= J_q^*(x_{n+1}^*), \end{aligned} \tag{1.5}$$

where $J_q^* : X^* \rightrightarrows X$ is a duality mapping of X^*. In the case of smooth Ψ we also consider a second generalization to (1.4)

$$x_{n+1} = x_n - \mu_n J_q^*(\nabla \Psi_n(x_n)). \tag{1.6}$$

We will show that both schemes converge strongly to the unique minimizer of problem (1.2), if μ_n is chosen properly.

Alber et al. presented in [11] an algorithm for the minimization of convex and not necessarily smooth functionals on uniformly smooth and uniformly convex Banach spaces which looks very similar to our first method in Section 3 and where the authors impose summation conditions on the stepsizes μ_n. However, only weak convergence of the proposed scheme is shown. Another interesting approach to obtain convergence results of descent methods in general Banach spaces can be found in the recent papers by Reich and Zaslavski [12, 13]. We want to emphasize that the most important novelties of the present paper are the strong convergence results.

In the next section, we give the necessary theoretical tools and apply them in Sections 3 and 4 to describe the methods and prove their convergence properties.

2. Preliminaries

Throughout the paper, 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 $x^* \in X^*$, we write

$$\langle x, x^* \rangle = \langle x^*, x \rangle = x^*(x). \tag{2.1}$$

Let $p, q \in (1, \infty)$ be conjugate exponents such that

$$\frac{1}{p} + \frac{1}{q} = 1. \tag{2.2}$$

2.1. Convexity and smoothness of Banach spaces

We introduce some definitions and preliminary results about the geometry of Banach spaces, which can be found in [14, 15].

The functions $\delta_X : [0, 2] \to [0, 1]$ and $\rho_X : [0, \infty) \to [0, \infty)$ defined by

$$\delta_X(\epsilon) = \inf \left\{ 1 - \left\| \frac{1}{2}(x + y) \right\| : \|x\| = \|y\| = 1, \|x - y\| \geq \epsilon \right\},$$
$$\rho_X(\tau) = \frac{1}{2} \sup\{ \|x + y\| + \|x - y\| - 2 : \|x\| = 1, \|y\| \leq \tau \} \tag{2.3}$$

are referred to as *the modulus of convexity of X* and *the modulus of smoothness of X*.

Definition 2.1. A Banach space X is said to be

(1) uniformly convex if $\delta_X(\epsilon) > 0$ for all $\epsilon \in (0, 2]$,

(2) p-convex or convex of power type if for some $p > 1$ and $C > 0$,

$$\delta_X(\epsilon) \geq C\epsilon^p, \tag{2.4}$$

(3) smooth if for every $x \neq 0$, there is a unique $x^* \in X^*$ such that $\|x^*\| = 1$ and $\langle x^*, x \rangle = \|x\|$,

(4) uniformly smooth if $\lim_{\tau\to 0}(\rho_X(\tau)/\tau) = 0$,

(5) q-smooth or smooth of power type if for some $q > 1$ and $C > 0$,

$$\rho_X(\tau) \leq C\tau^q. \tag{2.5}$$

There is a tight connection between the modulus of convexity and the modulus of smoothness. The Lindenstrauss duality formula implies that

$$X \text{ is } p\text{-convex iff } X^* \text{ is } q\text{-smooth,}$$
$$X \text{ is } q\text{-smooth iff } X^* \text{ is } p\text{-convex,} \tag{2.6}$$

(cf. [16], chapter II, Thereom 2.12). From Dvoretzky's theorem [17], it follows that $p \geq 2$ and $q \leq 2$. For Hilbert spaces the polarization identity

$$\|x - y\|^2 = \|x\|^2 - 2\langle x, y\rangle + \|y\|^2 \tag{2.7}$$

asserts that every Hilbert space is 2-convex and 2-smooth. For the sequence spaces ℓ_p, Lebesgue spaces L_p, and Sobolev spaces W_p^m it is also known [18, 19] that

$$\ell_p, L_p, W_p^m \quad \text{with } 1 < p \leq 2 \text{ are 2-convex, } p\text{-smooth,}$$
$$\ell_q, L_q, W_q^m \quad \text{with } 2 \leq q < \infty \text{ are } q\text{-convex, 2-smooth.} \tag{2.8}$$

2.2. Duality mapping

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

$$J_p(x) = \left\{x^* \in X^* : \langle x^*, x\rangle = \|x\|\|x^*\|, \|x^*\| = \|x\|^{p-1}\right\} \tag{2.9}$$

is called the *duality mapping* of X (with weight function $t \mapsto t^{p-1}$). By j_p we denote a single-valued selection of J_p.

One can show [15, Theorem I.4.4] that J_p is monotone, that is,

$$\langle x^* - y^*, x - y\rangle \geq 0 \quad \forall x^* \in J_p(x), \, y^* \in J_p(y). \tag{2.10}$$

If X is smooth, the duality mapping J_p is single valued, that is, one can identify it as $J_p : X \to X^*$ [15, Theorem I.4.5] .

If X is uniformly convex or uniformly smooth, then X is reflexive [15, Theorems II.2.9 and II.2.15]. By J_p^*, we then denote the duality mapping from X^* into $X^{**} = X$.

Let $\partial f : X \rightrightarrows X^*$ be the subdifferential of the convex functional $f : X \to \mathbb{R}$. At $x \in X$ it is defined by

$$\overline{x} \in \partial f(x) \Longleftrightarrow f(y) \geq f(x) + \langle \overline{x}, y - x\rangle \quad \forall y \in X. \tag{2.11}$$

Another important property of J_p is due to the theorem of Asplund [15, Theorem I.4.4]

$$J_p = \partial\left\{\frac{1}{p}\|\cdot\|^p\right\}. \tag{2.12}$$

This equality is also valid in the case of set valued duality mappings.

Example 2.2. In L_r spaces with $1 < r < \infty$, we have

$$\langle J_p(f), g \rangle = \int \left(\frac{1}{\|f\|_r^{r-p}} |f(x)|^{r-1} \text{sign}(f(x)) \right) \cdot g(x) dx. \tag{2.13}$$

In the sequence spaces ℓ_r with $1 < r < \infty$, we have

$$\langle J_p(x), y \rangle = \sum_i \left(\frac{1}{\|x\|_r^{r-p}} |x_i|^{r-1} \text{sign}(x_i) \right) \cdot y_i. \tag{2.14}$$

We also refer the interested reader to [20] where additional information on duality mappings may be found.

2.3. Xu-Roach inequalities

The next theorem (see [19]) provides us with inequalities which will be of great relevance for proving the convergence of our methods.

Theorem 2.3. (1) *Let X be a p-smooth Banach space. Then there exists a positive constant G_p such that*

$$\frac{1}{p}\|x - y\|^p \leq \frac{1}{p}\|x\|^p - \langle J_p(x), y \rangle + \frac{G_p}{p}\|y\|^p \quad \forall x, y \in X. \tag{2.15}$$

(2) *Let X be a q-convex Banach space. Then there exists a positive constant C_q such that*

$$\frac{1}{q}\|x - y\|^q \geq \frac{1}{q}\|x\|^q - \langle J_q(x), y \rangle + \frac{C_q}{q}\|y\|^q \quad \forall x, y \in X. \tag{2.16}$$

We remark that in a real Hilbert space these inequalities reduce to the well-known polarization identity (2.7). Further, we refer to [19] for the exact values of the constants G_p and C_q. For special cases like ℓ_p-spaces these constants have a simple form, see [8].

2.4. Bregman distances

It turns out that due to the geometrical characteristics of Banach spaces other than Hilbert spaces, it is often more appropriate to use Bregman distances instead of conventional-norm-based functionals $\|x - y\|$ or $\|J_p(x) - J_p(y)\|$ for convergence analysis. The idea to use such distances to design and analyze optimization algorithms goes back to Bregman [21] and since then his ideas have been successfully applied in various ways [4, 8, 22–26].

Definition 2.4. Let X be smooth and convex of power type. Then the Bregman distances $\Delta_p(x, y)$ are defined as

$$\Delta_p(x, y) := \frac{1}{q}\|J_p(x)\|^q - \langle J_p(x), y \rangle + \frac{1}{p}\|y\|^p. \tag{2.17}$$

We summarize a few facts concerning Bregman distances and their relationship to the norm in X (see also [8, Theorem 2.12]).

Theorem 2.5. *Let X be smooth and convex of power type. Then for all $p > 1$, $x, y \in X$, and sequences $(x_n)_n$ in X the following holds:*

(1) $\Delta_p(x, y) \geq 0$,

(2) $\lim_{n \to \infty} \|x_n - x\| = 0 \Longleftrightarrow \lim_{n \to \infty} \Delta_p(x_n, x) = 0$,

(3) $\Delta_p(\cdot, y)$ *is coercive, that is, the sequence (x_n) remains bounded if the sequence $(\Delta_p(x_n, y))$ is bounded.*

Remark 2.6. $\Delta_p(\cdot, \cdot)$ is in general not metric. In a real Hilbert space $\Delta_2(x, y) = (1/2)\|x - y\|^2$.

To shorten the proof in Chapter 3, we formulate and prove the following.

Lemma 2.7. *Let X be a p-convex Banach space, then there exists a positive constant c, such that*

$$c \cdot \|x - y\|^p \leq \Delta_p(x, y). \tag{2.18}$$

Proof. We have $(1/q)\|J_p(x)\|^q = (1/q)\|x\|^p$ and $\langle J_p(x), x \rangle = \|x\|^p$, hence

$$
\begin{aligned}
\Delta_p(x, y) &= \frac{1}{q}\|J_p(x)\|^q - \langle J_p(x), y \rangle + \frac{1}{p}\|y\|^p \\
&= \left(1 - \frac{1}{p}\right)\|x\|^p - \langle J_p(x), y \rangle + \frac{1}{p}\|y\|^p \\
&= \frac{1}{p}\|x - (x - y)\|^p - \frac{1}{p}\|x\|^p + \langle J_p(x), x - y \rangle.
\end{aligned}
\tag{2.19}
$$

By Theorem 2.3, we obtain

$$\Delta_p(x, y) \geq \frac{C_p}{p}\|x - y\|^p. \tag{2.20}$$

This completes the proof. $\qquad\square$

3. The dual method

This section deals with an iterative method for minimizing functionals of Tikhonov type. In contrast to the algorithm described in the next section, we iterate directly in the dual space X^*.

Due to simplicity, we restrict ourselves to the Tikhonov functional

$$\Psi(x) = \frac{1}{r}\|Ax - y\|_Y^r + \alpha\frac{1}{2}\|x\|_X^2 \quad \text{with } r > 1, \tag{3.1}$$

where X is a 2-convex and smooth Banach space, Y is an arbitrary Banach space and $A : X \to Y$ is a linear, continuous operator. For minimizing the functional, we choose an arbitrary starting point $x_0^* \in X^*$ and consider the following scheme

$$
\begin{aligned}
x_{n+1}^* &= x_n^* - \mu_n \psi_n \quad \text{with } \psi_n \in \partial\Psi(x_n), \\
x_{n+1} &= J_2^*(x_{n+1}^*).
\end{aligned}
\tag{3.2}
$$

We show the convergence of this method in a constructive way. This will be done via the following steps.

(1) We show the inequality

$$\Delta_2(x_{n+1}, x^\dagger) \leq \Delta_2(x_n, x^\dagger) - \mu_n \alpha \cdot \Delta_2(x_n, x^\dagger) + \mu_n^2 \frac{G_2}{2} \cdot \|\psi_n\|^2, \tag{3.3}$$

where $x^\dagger$ is the unique minimizer of the Tikhonov functional (3.1).

(2) We choose admissible stepsizes μ_n and show that the iterates approach $x^\dagger$ in the Bregman sense, if we assume

$$\Delta_2(x_n, x^\dagger) \geq \epsilon. \tag{3.4}$$

We suppose $\epsilon > 0$ to be small and specified later.

(3) We establish an upper estimate for $\Delta_2(x_{n+1}, x^\dagger)$ in the case that the condition $\Delta_2(x_n, x^\dagger) \geq \epsilon$ is violated.

(4) We choose ϵ such that in the case $\Delta_2(x_n, x^\dagger) < \epsilon$ the iterates stay in a certain Bregman ball, that is, $\Delta_2(x_{n+1}, x^\dagger) < \varepsilon_{\text{aim}}$, where ε_{aim} is some a priori chosen precision we want to achieve.

(5) Finally, we state the iterative minimization scheme.

(i) First, we calculate the estimate for Δ_{n+1}, where

$$\Delta_n := \Delta_2(x_n, x^\dagger). \tag{3.5}$$

Under our assumptions on X, we know that Ψ has a unique minimizer $x^\dagger$. Using (3.2) we get

$$\begin{aligned} \Delta_{n+1} &= \frac{1}{2} \|x_{n+1}^*\|^2 - \langle x_{n+1}^*, x^\dagger \rangle + \frac{1}{2} \|x^\dagger\|^2 \\ &= \frac{1}{2} \|x_n^* - \mu_n \psi_n\|^2 - \langle x_n^* - \mu_n \psi_n, x^\dagger \rangle + \frac{1}{2} \|x^\dagger\|^2. \end{aligned} \tag{3.6}$$

We remember that X is 2-convex, hence X^* is 2-smooth; see Section 2.1. By Theorem 2.3 applied to X^*, we get

$$\frac{1}{2} \|x_n^* - \mu_n \psi_n\|^2 \leq \frac{1}{2} \|x_n^*\|^2 - \mu_n \langle x_n, \psi_n \rangle + \frac{G_2}{2} \cdot \mu_n^2 \|\psi_n\|^2. \tag{3.7}$$

Therefore,

$$\begin{aligned} \Delta_{n+1} &\leq \frac{1}{2} \|x_n^*\|^2 - \mu_n \langle \psi_n, x_n \rangle + \frac{G_2}{2} \cdot \mu_n^2 \|\psi_n\|^2 - \langle x_n^*, x^\dagger \rangle + \mu_n \langle \psi_n, x^\dagger \rangle + \frac{1}{2} \|x^\dagger\|^2 \\ &= \Delta_n + \mu_n \langle \psi_n, x^\dagger - x_n \rangle + \mu_n^2 \cdot \frac{G_2}{2} \cdot \|\psi_n\|^2. \end{aligned} \tag{3.8}$$

We have

$$\partial\Psi(x) = A^* J_r(Ax - y) + \alpha J_2(x), \tag{3.9}$$

(cf. [27], Chapter I; Propositons 5.6, 5.7). By definition, $x^\dagger$ is the minimizer of Ψ, hence $\psi^\dagger :=$ $0 \in \partial\Psi(x^\dagger)$. Therefore, with the monotonicity of J_r, we get

$$
\begin{aligned}
\langle \psi_n, x^\dagger - x_n \rangle \\
&= \langle \psi_n - \psi^\dagger, x^\dagger - x_n \rangle \\
&= \alpha \langle J_2(x_n) - J_2(x^\dagger), x^\dagger - x_n \rangle + \langle A^* j_r(Ax_n - y) - A^* j_r(Ax^\dagger - y), x^\dagger - x_n \rangle \\
&= -\alpha \langle J_2(x_n) - J_2(x^\dagger), x_n - x^\dagger \rangle - \langle j_r(Ax_n - y) - j_r(Ax^\dagger - y), (Ax_n - y) - (Ax^\dagger - y) \rangle \\
&\leq -\alpha \langle J_2(x_n) - J_2(x^\dagger), x_n - x^\dagger \rangle.
\end{aligned}
\tag{3.10}
$$

Consider

$$
\begin{aligned}
\langle \psi_n, x^\dagger - x_n \rangle &\leq -\alpha \langle J_2(x_n) - J_2(x^\dagger), x_n - x^\dagger \rangle \\
&= -\alpha \left[\Delta_2(x_n, x^\dagger) + \Delta_2(x^\dagger, x_n) \right] \leq -\alpha \cdot \Delta_n.
\end{aligned}
\tag{3.11}
$$

Finally, we arrive at the desired inequality

$$\Delta_{n+1} \leq \Delta_n - \mu_n \alpha \cdot \Delta_n + \mu_n^2 \frac{G_2}{2} \cdot \|\psi_n\|^2. \tag{3.12}$$

(ii) Next, we choose admissible stepsizes. Assume that

$$\Delta_2(x_0, x^\dagger) = \Delta_0 \leq R. \tag{3.13}$$

We see that the choice

$$\mu_n = \frac{\alpha}{G_2 \|\psi_n\|^2} \cdot \Delta_n \tag{3.14}$$

minimizes the right-hand side of (3.12). We do not know the distance Δ_n, therefore, we set

$$\mu_n := \frac{\alpha}{G_2 P} \cdot \epsilon. \tag{3.15}$$

We will impose additional conditions on ϵ later. For the time being, assume that ϵ is small. The number P is defined by

$$P = P(R) = \sup \left\{ \|\psi\|^2 : \psi \in \partial\Psi(x) \text{ with } \Delta_2(x, x^\dagger) \leq R \right\}. \tag{3.16}$$

The Tikhonov functional Ψ is bounded on norm bounded sets, thus also $\partial\Psi$ is bounded on

norm-bounded sets. By Lemma 2.7, we know then that

$$\left\| x_0 - x^\dagger \right\| \leq \sqrt{\frac{R}{c}}. \tag{3.17}$$

Hence, P is finite for finite R.

Remark 3.1. If we assume $\|x^\dagger\| \leq \rho$ and with the help of Lemma 2.7, the definition of P, and the duality mapping J_2, we get an estimate for P. We have

$$\left\| x - x^\dagger \right\| \leq \sqrt{\frac{R}{c}},$$
$$\|x\| \leq \left\| x - x^\dagger \right\| + \left\| x^\dagger \right\| \leq \sqrt{\frac{R}{c}} + \rho. \tag{3.18}$$

We calculate an estimate for $\|\psi\|$:

$$\begin{aligned}
\|\psi\| &= \left\| A^* j_r (Ax - y) + \alpha J_2(x) \right\| \\
&\leq \|A^*\| \left\| j_r (Ax - y) \right\| + \alpha \|J_2(x)\| \\
&\leq \|A^*\| \|Ax - y\|^{r-1} + \alpha \|x\| \\
&\leq \|A\| \left(\|A\| \left[\sqrt{\frac{R}{c}} + \rho \right] + \|y\| \right)^{r-1} + \alpha \left(\sqrt{\frac{R}{c}} + \rho \right).
\end{aligned} \tag{3.19}$$

This calculation gives us an estimate for P. In practice, we will not determine this estimate exactly, but choose P in a sense big enough.

For $\Delta_n \geq \epsilon$ we approach the minimizer $x^\dagger$ in the Bregman sense, that is,

$$\begin{aligned}
\Delta_{n+1} &\leq \Delta_n - \frac{\alpha^2}{G_2 P} \epsilon^2 + \frac{\alpha^2}{2 G_2 P} \epsilon^2 \\
&= \Delta_n - \frac{\alpha^2}{2 G_2 P} \epsilon^2 =: \Delta_n - D \epsilon^2,
\end{aligned} \tag{3.20}$$

where

$$D := D(R) = \frac{\alpha^2}{2 G_2 P}. \tag{3.21}$$

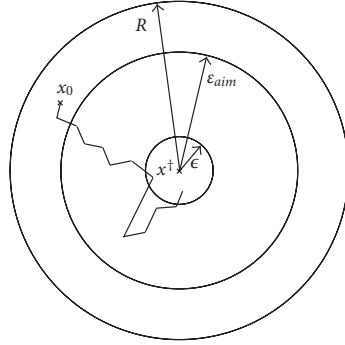This ensures

$$\Delta_{n+1} < \Delta_n < \cdots < \Delta_0 \tag{3.22}$$

as long as $\Delta_n \geq \epsilon$ is fulfilled.

(iii) We know the behavior of the Bregman distances, if $\Delta_n \geq \epsilon$ holds. Next, we need to know what happens if $\Delta_n < \epsilon$. By (3.12), we then have

$$\Delta_{n+1} \leq \Delta_n + D \epsilon^2 < \epsilon + D \epsilon^2. \tag{3.23}$$

Figure 1: Geometry of the problem. The iterates x_n approach $x^\dagger$ as long as $\Delta_2(x_n, x^\dagger) \geq \epsilon$. The auxiliary number ϵ is chosen such that, if the iterates enter the Bregman ball with radius $\varepsilon_{\mathrm{aim}}$ around $x^\dagger$, the following iterates stay in that ball.

(iv) We choose

$$\epsilon := \frac{-1 + \sqrt{1 + 4D \cdot \varepsilon_{\mathrm{aim}}}}{2D}, \tag{3.24}$$

where $\varepsilon_{\mathrm{aim}} > 0$ is the accuracy we aim at. For the case $\Delta_n < \epsilon$ this choice of ϵ assures that

$$\Delta_{n+1} < \epsilon + D\epsilon^2 = \varepsilon_{\mathrm{aim}}. \tag{3.25}$$

Note that the choice of ϵ implies $\epsilon \leq \varepsilon_{\mathrm{aim}}$.

Next, we calculate an index N, which ensures that the iterates x_n with $n \geq N$ are located in a Bregman ball with radius $\varepsilon_{\mathrm{aim}}$ around $x^\dagger$. We know that if x_n fulfills $\Delta_n \leq \varepsilon_{\mathrm{aim}}$, then all following iterates fulfill this condition as well.

Hence, the opposite case is $\Delta_{n+1} \geq \varepsilon_{\mathrm{aim}} \geq \epsilon$. By (3.20), we know that this is only the case if

$$\varepsilon_{\mathrm{aim}} \leq \Delta_{n+1} \leq R - nD\epsilon^2. \tag{3.26}$$

By choosing N such that

$$N > \frac{R - \varepsilon_{\mathrm{aim}}}{D\epsilon^2} = \frac{R - \varepsilon_{\mathrm{aim}}}{\left(1 + (1 - \sqrt{1 + 4D\varepsilon_{\mathrm{aim}}})/2D\varepsilon_{\mathrm{aim}}\right)\varepsilon_{\mathrm{aim}}}, \tag{3.27}$$

we get

$$\Delta_N \leq \varepsilon_{\mathrm{aim}}. \tag{3.28}$$

Figure 1 illustrates the behavior of the iterates.

(v) We are now in the same situation as described in (2). If we replace R by $\varepsilon_{\mathrm{aim}}$, x_0 by x_N and $\varepsilon_{\mathrm{aim}}$ by some $\varepsilon_{\mathrm{aim},2} < \varepsilon_{\mathrm{aim}}$ and repeat the argumentation in (2)–(4), we obtain a contracting sequence of Bregman balls.

If the sequence $(\varepsilon_{\mathrm{aim},k})_k$ is a null sequence, then by Lemma 2.7 the iterates x_n converge strongly to $x^\dagger$. This proves the following.

Theorem 3.2. *The iterative method, defined by*

(S$_0$) *choose an arbitrary x_0 and a decreasing positive sequence $(\varepsilon_k)_k$ with*

$$\lim_{k \to \infty} \varepsilon_k = 0$$

$$\Delta_2(x_0, x^\dagger) < \varepsilon_1,$$

$$(3.29)$$

set $k = 1$;

(S$_1$) *compute P, D, ϵ, and μ as*

$$P = \sup \{ \|\psi\|^2 : \psi \in \partial\Psi(x) \ \text{with} \ \Delta_2(x, x^\dagger) \le \varepsilon_k \},$$

$$D = \frac{\alpha^2}{2, G_2 P}$$

$$\epsilon = \frac{-1 + \sqrt{1 + 4D \cdot \varepsilon_{k+1}}}{2, D}$$

$$(3.30)$$

$$\mu = \frac{\alpha}{G_2 P} \epsilon;$$

(S$_2$) *iterate x_n by*

$$x_{n+1}^* = x_n^* - \mu \cdot \psi_n \quad \text{with} \ \psi_n \in \partial\Psi(x_n),$$

$$x_{n+1} = J_2^*(x_{n+1}^*),$$

$$(3.31)$$

for at least N iterations, where

$$N > \frac{\varepsilon_k - \varepsilon_{k+1}}{\left(1 + (1 - \sqrt{1 + 4D\varepsilon_{k+1}})/2D\varepsilon_{k+1}\right)\varepsilon_{k+1}};$$

$$(3.32)$$

(S$_3$) *let $k \leftarrow (k + 1)$, reset P, D, ϵ, μ, N and go to step (S$_1$), defines an iterative minimization method for the Tikhonov functional Ψ, defined in (3.1) and the iterates converge strongly to the unique minimizer $x^\dagger$.*

Remark 3.3. A similar construction can be carried out for any p-convex and smooth Banach space.

4. Steepest descent method

Let X be uniformly convex and uniformly smooth and let Y be uniformly smooth. Then the Tikhonov functional

$$\Psi(x) := \frac{1}{r} \|Ax - y\|^r + \frac{\alpha}{p} \|x\|^p$$

$$(4.1)$$

is strictly convex, weakly lower semicontinuous, coercive, and Gâteaux differentiable with derivative

$$\nabla\Psi(x) = A^* J_r(Ax - y) + \alpha J_p(x).$$

$$(4.2)$$

Hence, there exists the unique minimizer $x^\dagger$ of Ψ, which is characterized by

$$\Psi(x^\dagger) = \min_{x \in X} \Psi(x) \iff \nabla\Psi(x^\dagger) = 0. \tag{4.3}$$

In this section, we consider the steepest descent method to find $x^\dagger$. In [28, 29], it has already been proven that for a general continuously differentiable functional Ψ every cluster point of such steepest descent method is a stationary point. Recently, Canuto and Urban [30] have shown strong convergence under the additional assumption of ellipticity, which our Ψ in (4.1) would fulfill if we required X to be p-convex. Here we prove strong convergence without this additional assumption. To make the proof of convergence more transparent, we confine ourselves here to the case of r-smooth Y and p-smooth X (with then $r, p \in (1, 2]$ being the ones appearing in the definition of the Tikhonov functional (4.1)) and refer the interested reader to the appendix, where we prove the general case.

Theorem 4.1. *The sequence $(x_n)_n$, generated by*

(S_0) *choose an arbitrary starting point $x_0 \in X$ and set $n = 0$;*

(S_1) *if $\nabla\Psi(x_n) = 0$, then STOP else do a line search to find $\mu_n > 0$ such that*

$$\Psi(x_n - \mu_n J_q^*(\nabla\Psi(x_n))) = \min_{\mu \in \mathbb{R}} \Psi(x_n - \mu J_q^*(\nabla\Psi(x_n))); \tag{4.4}$$

(S_2) *set $x_{n+1} := x_n - \mu_n J_q^*(\nabla\Psi(x_n))$, $n \leftarrow (n+1)$ and go to step (S_1), converges strongly to the unique minimizer $x^\dagger$ of Ψ.*

Remark 4.2. (a) If the stopping criterion $\nabla\Psi(x_n) = 0$ is fulfilled for some $n \in \mathbb{N}$, then by (4.3), we already have $x_n = x^\dagger$ and we can stop iterating.

(b) Due to the properties of Ψ, the function $f_n : \mathbb{R} \to [0, \infty)$ defined by

$$f_n(\mu) := \Psi(x_n - \mu J_q^*(\nabla\Psi(x_n))) \tag{4.5}$$

appearing in the line search of step (S_1) is strictly convex and differentiable with continuous derivative

$$f_n'(\mu) = -\langle \nabla\Psi(x_n - \mu J_q^*(\nabla\Psi(x_n))), J_q^*(\nabla\Psi(x_n)) \rangle. \tag{4.6}$$

Since $f_n'(0) = -\|\nabla\Psi(x_n)\|^q < 0$ and f_n' is increasing by the monotonicity of the duality mappings, we know that μ_n must in fact be positive.

Proof of Theorem 4.1. By the above remark it suffices to prove convergence in case $\nabla\Psi(x_n) \neq 0$ for all $n \in \mathbb{N}$. We fix $\gamma \in (0, 1)$ and show that there exists positive $\tilde{\mu}_n$ such that

$$\Psi(x_{n+1}) \leq \Psi(x_n) - \tilde{\mu}_n \|\nabla\Psi(x_n)\|^q (1 - \gamma), \tag{4.7}$$

which will finally assure convergence. To establish this relation, we use the characteristic inequalities in Theorem 2.3 to estimate, for all $\mu > 0$,

$$
\begin{aligned}
\Psi(x_{n+1}) &\leq \Psi(x_n - \mu J_q^*(\nabla\Psi(x_n))) \\
&= \frac{1}{r}\|(Ax_n - y) - \mu AJ_q^*(\nabla\Psi(x_n))\|^r + \frac{\alpha}{p}\|x_n - \mu J_q^*(\nabla\Psi(x_n))\|^p \\
&\leq \frac{1}{r}\|Ax_n - y\|^r - \langle J_r(Ax_n - y), \mu AJ_q^*(\nabla\Psi(x_n))\rangle, + \frac{G_r}{r}\|\mu AJ_q^*(\nabla\Psi(x_n))\|^r \\
&\quad + \frac{\alpha}{p}\|x_n\|^p - \alpha\langle J_p(x_n), \mu J_q^*(\nabla\Psi(x_n))\rangle + \alpha\frac{G_p}{p}\|\mu J_q^*(\nabla\Psi(x_n))\|^p.
\end{aligned}
\tag{4.8}
$$

By (4.1) and (4.2) for $x = x_n$ and

$$
\langle \nabla\Psi(x_n), J_q^*(\nabla\Psi(x_n))\rangle = \|\nabla\Psi(x_n)\|^q = \|J_q^*(\nabla\Psi(x_n))\|^p,
\tag{4.9}
$$

we can further estimate

$$
\begin{aligned}
\Psi(x_{n+1}) &\leq \Psi(x_n) - \mu\|\nabla\Psi(x_n)\|^q + \frac{G_r}{r}\|AJ_q^*(\nabla\Psi(x_n))\|^r\mu^r + \alpha\frac{G_p}{p}\|\nabla\Psi(x_n)\|^q\mu^p \\
&= \Psi(x_n) - \mu\|\nabla\Psi(x_n)\|^q(1 - \phi_n(\mu)),
\end{aligned}
\tag{4.10}
$$

whereby we set

$$
\phi_n(\mu) := \frac{G_r}{r}\frac{\|AJ_q^*(\nabla\Psi(x_n))\|^r}{\|\nabla\Psi(x_n)\|^q}\mu^{r-1} + \alpha\frac{G_p}{p}\mu^{p-1}.
\tag{4.11}
$$

The function $\phi_n : (0, \infty) \to (0, \infty)$ is continuous and increasing with $\lim_{\mu\to 0}\phi_n(\mu) = 0$ and $\lim_{\mu\to\infty}\phi_n(\mu) = \infty$. Hence, there exists a $\tilde{\mu}_n > 0$ such that

$$
\phi_n(\tilde{\mu}_n) = \gamma
\tag{4.12}
$$

and we get

$$
\Psi(x_{n+1}) \leq \Psi(x_n) - \tilde{\mu}_n\|\nabla\Psi(x_n)\|^q(1 - \gamma).
\tag{4.13}
$$

We show that $\lim_{n\to\infty}\|\nabla\Psi(x_n)\| = 0$. From (4.13), we infer that the sequence $(\Psi(x_n))_n$ is decreasing and especially bounded and that

$$
\lim_{n\to\infty}\tilde{\mu}_n\|\nabla\Psi(x_n)\|^q = 0.
\tag{4.14}
$$

Since Ψ is coercive, the sequence $(x_n)_n$ remains bounded and (4.2) then implies that the sequence $(\nabla\Psi(x_n))_n$ is bounded as well. Suppose $\limsup_{n\to\infty}\|\nabla\Psi(x_n)\| = \epsilon > 0$ and let $\|\nabla\Psi(x_{n_k})\| \to \epsilon$ for $k \to \infty$. Then we must have $\lim_{k\to\infty}\tilde{\mu}_{n_k} = 0$ by (4.14). But by

the definition of ϕ_n (4.11) and the choice of $\tilde{\mu}_n$ (4.12), we get for some constant $C > 0$ with $\|AJ_q^*(\nabla\Psi(x_n))\|^r \leq C$,

$$0 < \gamma = \phi_{n_k}(\mu_{n_k}) \leq \frac{G_r}{r} \frac{C}{\|\nabla\Psi(x_{n_k})\|^q} \tilde{\mu}_{n_k}^{r-1} + \alpha \frac{G_p}{p} \tilde{\mu}_{n_k}^{p-1}. \tag{4.15}$$

Since the right-hand side converges to zero for $k \to \infty$, this leads to a contradiction. So we have $\limsup_{n\to\infty}\|\nabla\Psi(x_n)\| = 0$ and thus $\lim_{n\to\infty}\|\nabla\Psi(x_n)\| = 0$. We finally show that $(x_n)_n$ converges strongly to $x^\dagger$. By (4.3) and the monotonicity of the duality mapping J_r, we get

$$\begin{aligned}
\|\nabla\Psi(x_n)\| \, \|x_n - x^\dagger\| &\geq \langle \nabla\Psi(x_n), x_n - x^\dagger \rangle \\
&= \langle \nabla\Psi(x_n) - \nabla\Psi(x^\dagger), x_n - x^\dagger \rangle \\
&= \langle J_r(Ax_n - y) - J_r(Ax^\dagger - y), (Ax_n - y) - (Ax^\dagger - y) \rangle \\
&\quad + \alpha \langle J_p(x_n) - J_p(x^\dagger), x_n - x^\dagger \rangle \\
&\geq \alpha \langle J_p(x_n) - J_p(x^\dagger), x_n - x^\dagger \rangle.
\end{aligned} \tag{4.16}$$

Since $(x_n)_n$ is bounded and $\lim_{n\to\infty}\|\nabla\Psi(x_n)\| = 0$, this yields

$$\lim_{n\to\infty} \langle J_p(x_n) - J_p(x^\dagger), x_n - x^\dagger \rangle = 0, \tag{4.17}$$

from which we infer that $(x_n)_n$ converges strongly to $x^\dagger$ in a uniformly convex X [15, Theorem II.2.17.] □

5. Conclusions

We have analyzed two conceptionally quite different nonlinear iterative methods for finding the minimizer of norm-based Tikhonov functionals in Banach spaces. One is the steepest descent method, where the iterations are directly carried out in the X-space by pulling the gradient of the Tikhonov functional back to X via duality mappings. The method is shown to be strongly convergent in case the involved spaces are nice enough. In the other one, the iterations are performed in the dual space X^*. Though this method seems to be inherently slow, strong convergence can be shown without restrictions on the Y-space.

Appendix

Steepest descent method in uniformly smooth spaces

As already pointed out in Section 4, we prove here Theorem 4.1 for the general case of X being uniformly convex and uniformly smooth and Y being uniformly smooth, and with $r, p \geq 2$ in the definition of the Tikhonov functional (4.1). To do so, we need some additional results based on the paper of Xu and Roach [19].

In what follows $C, L > 0$ are always supposed to be (generic) constants and we write

$$a \vee b = \max\{a, b\}, \qquad a \wedge b = \min\{a, b\}. \tag{A.1}$$

Let $\overline{\rho}_X : (0, \infty) \to (0, 1]$ be the function

$$\overline{\rho}_X(\tau) := \frac{\rho_X(\tau)}{\tau}, \tag{A.2}$$

where ρ_X is the modulus of smoothness of a Banach space X. The function $\overline{\rho}_X$ is known to be continuous and nondecreasing [14, 31].

The next lemma allows us to estimate $\|J_p(x) - J_p(y)\|$ via $\overline{\rho}_X(\|x - y\|)$, which in turn will be used to derive a version of the characteristic inequality that is more convenient for our purpose.

Lemma A.1. *Let X be a uniformly smooth Banach space with duality mapping J_p with weight $p \geq 2$. Then for all $x, y \in X$ the following inequalities are valid:*

$$\|J_p(x) - J_p(y)\| \leq C \max\{1, (\|x\| \vee \|y\|)^{p-1}\}\overline{\rho}_X(\|x - y\|) \tag{A.3}$$

(hence, J_p is uniformly continuous on bounded sets) and

$$\|x - y\|^p \leq \|x\|^p - p\langle J_p(x), y\rangle + C(1 \vee (\|x\| + \|y\|)^{p-1})\rho_X(\|y\|). \tag{A.4}$$

Proof. We at first prove (A.3). By [19, formula (3.1)], we have

$$\|J_p(x) - J_p(y)\| \leq C(\|x\| \vee \|y\|)^{p-1}\overline{\rho}_X\left(\frac{\|x - y\|}{\|x\| \vee \|y\|}\right). \tag{A.5}$$

We estimate similarly as after inequality (3.5) in the same paper. If $1/(\|x\| \vee \|y\|) \leq 1$, then we get by the monotonicity of $\overline{\rho}_X$

$$\overline{\rho}_X\left(\frac{\|x - y\|}{\|x\| \vee \|y\|}\right) \leq \overline{\rho}_X(\|x - y\|) \tag{A.6}$$

and therefore (A.3) is valid. In case $1/(\|x\| \vee \|y\|) \geq 1$ ($\Leftrightarrow \|x\| \vee \|y\| \leq 1$), we use the fact that ρ_X is equivalent to a decreasing function (i.e. $\rho_X(\eta)/\eta^2 \leq L(\rho_X(\tau)/\tau^2)$ for $\eta \geq \tau > 0$ [14]) and get

$$\rho_X\left(\frac{\|x - y\|}{\|x\| \vee \|y\|}\right) \leq \frac{L}{(\|x\| \vee \|y\|)^2}\rho_X(\|x - y\|) \tag{A.7}$$

and therefore

$$\overline{\rho}_X\left(\frac{\|x - y\|}{\|x\| \vee \|y\|}\right) \leq \frac{L}{\|x\| \vee \|y\|}\overline{\rho}_X(\|x - y\|). \tag{A.8}$$

For $p \geq 2$, we thus arrive at

$$\|J_p(x) - J_p(y)\| \leq CL(\|x\| \vee \|y\|)^{p-2}\overline{\rho}_X(\|x - y\|)$$
$$\leq CL\overline{\rho}_X(\|x - y\|) \tag{A.9}$$

and also in this case (A.3) is valid.

Let us prove (A.4). As in [19], we consider the continuously differentiable function $f :$ $[0, 1] \to \mathbb{R}$ with

$$
\begin{aligned}
f(t) &:= \|x - ty\|^p, \qquad f'(t) = -p\langle J_p(x - ty), y\rangle, \\
f(0) &= \|x\|^p, \qquad f(1) = \|x - y\|^p, \qquad f'(0) = -p\langle J_p(x), y\rangle
\end{aligned}
\tag{A.10}
$$

and get

$$
\begin{aligned}
\|x - y\|^p - \|x\|^p + p\langle J_p(x), y\rangle &= f(1) - f(0) - f'(0) \\
&= \int_0^1 f'(t) - f'(0)\,dt \\
&= p\int_0^1 \langle J_p(x) - J_p(x - ty), y\rangle\,dt \\
&\leq p\int_0^1 \|J_p(x) - J_p(x - ty)\|\|y\|\,dt.
\end{aligned}
\tag{A.11}
$$

For $t \in [0, 1]$, we set $\tilde{y} := x - ty$ and get $x - \tilde{y} = ty$, $\|\tilde{y}\| \leq \|x\| + \|y\|$ and thus $\|x\| \vee \|\tilde{y}\| \leq \|x\| + \|y\|$. By the monotonicity of $\overline{\rho}_X$, we have

$$
\overline{\rho}_X(t\|y\|)\|y\| \leq \overline{\rho}_X(\|y\|)\|y\| = \rho_X(\|y\|)
\tag{A.12}
$$

and by (A.3), we thus obtain

$$
\begin{aligned}
\|x - y\|^p - \|x\|^p + p\langle J_p(x), y\rangle &\leq p\int_0^1 C\max\left\{1, (\|x\| + \|y\|)^{p-1}\right\}\overline{\rho}_X(t\|y\|)\|y\|\,dt \\
&\leq C\max\left\{1, (\|x\| + \|y\|)^{p-1}\right\}\rho_X(\|y\|).
\end{aligned}
\tag{A.13}
$$

$\square$

The proof of Theorem 4.1 is now quite similar to the case of smoothness of power type, though it is more technical, and we only give the main modifications.

Proof of Theorem 4.1 (for uniformly smooth spaces). We fix $\gamma \in (0, 1)$, $\overline{\mu} > 0$ and for $n \in \mathbb{N}$, we choose $\tilde{\mu}_n \in (0, \overline{\mu}]$ such that

$$
\phi_n(\tilde{\mu}_n) = \phi_n(\overline{\mu}) \wedge \gamma.
\tag{A.14}
$$

Here the function $\phi_n : (0, \infty) \to (0, \infty)$ is defined by

$$
\begin{aligned}
\phi_n(\mu) :=\ & \frac{C_Y}{r}\left(1 \vee \left(\|Ax_n - y\| + \overline{\mu}\|AJ_q^*(\nabla\Psi(x_n))\|\right)^{r-1}\right) \\
&\times \frac{\|AJ_q^*(\nabla\Psi(x_n))\|}{\|\nabla\Psi(x_n)\|^q}\overline{\rho}_Y(\mu\|AJ_q^*(\nabla\Psi(x_n))\|) \\
&+ \alpha\frac{C_X}{p}\left(1 \vee \left(\|x_n\| + \overline{\mu}\|\nabla\Psi(x_n)\|^{q-1}\right)^{p-1}\right) \\
&\times \frac{1}{\|\nabla\Psi(x_n)\|}\overline{\rho}_X\left(\mu\|\nabla\Psi(x_n)\|^{q-1}\right)
\end{aligned}
\tag{A.15}
$$

with the constants C_X, C_Y being the ones appearing in the respective characteristic inequalities (A.4). This choice of $\tilde{\mu}_n$ is possible since by the properties of $\overline{\rho}_Y$ and $\overline{\rho}_X$, the function ϕ_n is continuous, increasing and $\lim_{\mu \to 0} \phi_n(\mu) = 0$. We again aim at an inequality of the form

$$\Psi(x_{n+1}) \le \Psi(x_n) - \tilde{\mu}_n \|\nabla\Psi(x_n)\|^q (1 - \gamma), \tag{A.16}$$

which will finally assure convergence. Here we use the characteristic inequalities (A.4) to estimate

$$\begin{aligned}
\Psi(x_n + 1) &\le \Psi(x_n) - \tilde{\mu}_n \|\nabla\Psi(x_n)\|^q \\
&\quad + \frac{C_Y}{r}\Big(1 \vee \big(\|Ax_n - y\| + \|\tilde{\mu}_n A J_q^*(\nabla\Psi(x_n))\|\big)^{r-1}\Big)\rho_Y\big(\|\tilde{\mu}_n A J_q^*(\nabla\Psi(x_n))\|\big) \\
&\quad + \alpha\frac{C_X}{p}\Big(1 \vee \big(\|x_n\| + \|\tilde{\mu}_n J_q^*(\nabla\Psi(x_n))\|\big)^{p-1}\Big)\rho_X\big(\|\tilde{\mu}_n J_q^*(\nabla\Psi(x_n))\|\big).
\end{aligned}$$

$$\tag{A.17}$$

Since $\tilde{\mu}_n \le \overline{\mu}$ and by the definition of ϕ_n (A.15), we can further estimate

$$\begin{aligned}
\Psi(x_n + 1) &\le \Psi(x_n) - \tilde{\mu}_n \|\nabla\Psi(x_n)\|^q \\
&\quad + \frac{C_Y}{r}\Big(1 \vee \big(\|Ax_n - y\| + \overline{\mu}\|A J_q^*(\nabla\Psi(x_n))\|\big)^{r-1}\Big)\rho_Y\big(\|\tilde{\mu}_n A J_q^*(\nabla\Psi(x_n))\|\big) \\
&\quad + \alpha\frac{C_X}{p}\Big(1 \vee \big(\|x_n\| + \overline{\mu}\|J_q^*(\nabla\Psi(x_n))\|\big)^{p-1}\Big)\rho_X\big(\tilde{\mu}_n\|J_q^*(\nabla\Psi(x_n))\|\big). \\
&= \Psi(x_n) - \tilde{\mu}_n \|\nabla\Psi(x_n)\|^q \big(1 - \phi_n(\tilde{\mu}_n)\big)
\end{aligned}$$

$$\tag{A.18}$$

The choice of $\tilde{\mu}_n$ (A.14) finally yields

$$\Psi(x_{n+1}) \le \Psi(x_n) - \tilde{\mu}_n \|\nabla\Psi(x_n)\|^q (1 - \gamma). \tag{A.19}$$

It remains to show that this implies $\lim_{n\to\infty}\|\nabla\Psi(x_n)\| = 0$. The rest then follows analogously as in the proof of Theorem 4.1. From (A.19), we infer that

$$\lim_{n\to\infty}\tilde{\mu}_n \|\nabla\Psi(x_n)\|^q = 0 \tag{A.20}$$

and that the sequences $(x_n)_n$ and $(\nabla\Psi(x_n))_n$ are bounded.

Suppose $\limsup_{n\to\infty}\|\nabla\Psi(x_n)\| = \epsilon > 0$ and let $\|\nabla\Psi(x_{n_k})\| \to \epsilon$ for $k \to \infty$. Then we must have $\lim_{k\to\infty}\tilde{\mu}_{n_k} = 0$ by (A.20). We show that this leads to a contradiction. On the one hand by (A.15), we get

$$\phi_{n_k}(\tilde{\mu}_{n_k}) \le \frac{L_1}{\|\nabla\Psi(x_{n_k})\|^q}\overline{\rho}_Y(\tilde{\mu}_{n_k}L_2) + \frac{C_1}{\|\nabla\Psi(x_{n_k})\|}\overline{\rho}_X(\tilde{\mu}_{n_k}C_2). \tag{A.21}$$

Since the right-hand side converges to zero for $k \to \infty$, so does $\phi_{n_k}(\tilde{\mu}_{n_k})$. On the other hand,

by (A.14), we have

$$\phi_{n_k}(\widetilde{\mu}_{n_k}) = \phi_{n_k}(\overline{\mu}) \wedge \gamma,$$

$$\phi_{n_k}(\overline{\mu}) \geq 0 + C\,\overline{\rho}_X\left(\overline{\mu}\|\nabla\Psi(x_{n_k})\|^{q-1}\right). \tag{A.22}$$

Hence, $\phi_{n_k}(\widetilde{\mu}_{n_k}) \geq L > 0$ for all k big enough which contradicts $\lim_{k\to\infty}\phi_{n_k}(\widetilde{\mu}_{n_k}) = 0$. So we have $\limsup_{n\to\infty}\|\nabla\Psi(x_n)\| = 0$ and thus $\lim_{n\to\infty}\|\nabla\Psi(x_n)\| = 0$. □

Acknowledgment

References

[1] A. K. Louis, *Inverse und schlecht gestellte Probleme*, Teubner Studienbücher Mathematik, B. G. Teubner, Stuttgart, Germany, 1989.

[2] A. Rieder, *No Problems with Inverse Problems*, Vieweg & Sohn, Braunschweig, Germany, 2003.

[3] H. Engl, M. Hanke, and A. Neubauer, *Regularization of Inverse Problems*, Kluwer Academic, Dordrecht, The Netherlands, 2000.

[4] Y. I. Alber, "Iterative regularization in Banach spaces," *Soviet Mathematics*, vol. 30, no. 4, pp. 1–8, 1986.

[5] R. Plato, "On the discrepancy principle for iterative and parametric methods to solve linear ill-posed equations," *Numerische Mathematik*, vol. 75, no. 1, pp. 99–120, 1996.

[6] S. Osher, M. Burger, D. Goldfarb, J. Xu, and W. Yin, "An iterative regularization method for total variation-based image restoration," *Multiscale Modeling & Simulation*, vol. 4, no. 2, pp. 460–489, 2005.

[7] D. Butnariu and E. Resmerita, "Bregman distances, totally convex functions, and a method for solving operator equations in Banach spaces," *Abstract and Applied Analysis*, vol. 2006, Article ID 84919, 39 pages, 2006.

[8] F. Schöpfer, A. K. Louis, and T. Schuster, "Nonlinear iterative methods for linear ill-posed problems in Banach spaces," *Inverse Problems*, vol. 22, no. 1, pp. 311–329, 2006.

[9] I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Communications on Pure and Applied Mathematics*, vol. 57, no. 11, pp. 1413–1457, 2004.

[10] K. Bredies, D. Lorenz, and P. Maass, "A generalized conditional gradient method and its connection to an iterative shrinkage method," to appear in *Computational Optimization and Applications*.

[11] Y. I. Alber, A. N. Iusem, and M. V. Solodov, "Minimization of nonsmooth convex functionals in Banach spaces," *Journal of Convex Analysis*, vol. 4, no. 2, pp. 235–255, 1997.

[12] S. Reich and A. J. Zaslavski, "Generic convergence of descent methods in Banach spaces," *Mathematics of Operations Research*, vol. 25, no. 2, pp. 231–242, 2000.

[13] S. Reich and A. J. Zaslavski, "The set of divergent descent methods in a Banach space is σ-porous," *SIAM Journal on Optimization*, vol. 11, no. 4, pp. 1003–1018, 2001.

[14] J. Lindenstrauss and L. Tzafriri, *Classical Banach Spaces. II*, vol. 97 of *Results in Mathematics and Related Areas*, Springer, Berlin, Germany, 1979.

[15] I. Cioranescu, *Geometry of Banach Spaces, Duality Mappings and Nonlinear Problems*, vol. 62 of *Mathematics and Its Applications*, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1990.

[16] R. Deville, G. Godefroy, and V. Zizler, *Smoothness and Renormings in Banach Spaces*, vol. 64 of *Pitman Monographs and Surveys in Pure and Applied Mathematics*, Longman Scientific & Technical, Harlow, UK, 1993.

[17] A. Dvoretzky, "Some results on convex bodies and Banach spaces," in *Proceedings of the International Symposium on Linear Spaces*, pp. 123–160, Jerusalem Academic Press, Jerusalem, Israel, 1961.

[18] O. Hanner, "On the uniform convexity of L^p and l^p," *Arkiv för Matematik*, vol. 3, no. 3, pp. 239–244, 1956.

[19] Z. B. Xu and G. F. Roach, "Characteristic inequalities of uniformly convex and uniformly smooth Banach spaces," *Journal of Mathematical Analysis and Applications*, vol. 157, no. 1, pp. 189–210, 1991.

[20] S. Reich, "Review of I. Cioranescu "Geometry of Banach spaces, duality mappings and nonlinear problems"," *Bulletin of the American Mathematical Society*, vol. 26, no. 2, pp. 367–370, 1992.

[21] L. M. Bregman, "The relaxation method for finding common points of convex sets and its application to the solution of problems in convex programming," *USSR Computational Mathematics and Mathematical Physics*, vol. 7, pp. 200–217, 1967.

[22] C. Byrne and Y. Censor, "Proximity function minimization using multiple Bregman projections, with applications to split feasibility and Kullback-Leibler distance minimization," *Annals of Operations Research*, vol. 105, no. 1–4, pp. 77–98, 2001.

[23] Y. I. Alber and D. Butnariu, "Convergence of Bregman projection methods for solving consistent convex feasibility problems in reflexive Banach spaces," *Journal of Optimization Theory and Applications*, vol. 92, no. 1, pp. 33–61, 1997.

[24] H. H. Bauschke, J. M. Borwein, and P. L. Combettes, "Bregman monotone optimization algorithms," *SIAM Journal on Control and Optimization*, vol. 42, no. 2, pp. 596–636, 2003.

[25] H. H. Bauschke and A. S. Lewis, "Dykstra's algorithm with Bregman projections: a convergence proof," *Optimization*, vol. 48, no. 4, pp. 409–427, 2000.

[26] J. D. Lafferty, S. D. Pietra, and V. D. Pietra, "Statistical learning algorithms based on Bregman distances," in *Proceedings of the 5th Canadian Workshop on Information Theory*, Toronto, Ontario, Canada, June 1997.

[27] I. Ekeland and R. Temam, *Convex Analysis and Variational Problems*, North-Holland, Amsterdam, The Netherlands, 1976.

[28] R. R. Phelps, "Metric projections and the gradient projection method in Banach spaces," *SIAM Journal on Control and Optimization*, vol. 23, no. 6, pp. 973–977, 1985.

[29] R. H. Byrd and R. A. Tapia, "An extension of Curry's theorem to steepest descent in normed linear spaces," *Mathematical Programming*, vol. 9, no. 1, pp. 247–254, 1975.

[30] C. Canuto and K. Urban, "Adaptive optimization of convex functionals in Banach spaces," *SIAM Journal on Numerical Analysis*, vol. 42, no. 5, pp. 2043–2075, 2005.

[31] T. Figiel, "On the moduli of convexity and smoothness," *Studia Mathematica*, vol. 56, no. 2, pp. 121–155, 1976.