Noname manuscript No.
(will be inserted by the editor)

An analysis of a multi-level projected steepest descent iteration for nonlinear inverse problems in Banach spaces subject to stability constraints

Maarten V. de Hoop · Lingyun Qiu · Otmar Scherzer

the date of receipt and acceptance should be inserted later

Abstract We consider nonlinear inverse problems described by operator equations in Banach spaces. Assuming conditional stability of the inverse problem, that is, assuming that stability holds on a compact, convex subset of the domain of the operator, we introduce a novel nonlinear projected steepest descent iteration and analyze its convergence to an approximate solution given limited accuracy data. We proceed with developing a multi-level algorithm based on a nested family of compact, convex subsets on which stability holds and the stability constants are ordered. Growth of the stability constants is coupled to the increase in accuracy of approximation between neighboring levels to ensure that the algorithm can continue from level to level until the iterate satisfies a desired discrepancy criterion, after a finite number of steps.

Keywords inverse problems · projected steepest descent iteration · stability

This research was supported in part National Science Foundation grant CMG DMS-1025318 and in part by the members of the Geo-Mathematical Imaging Group at Purdue University. The work of OS has been supported by the Austrian Science Fund (FWF) within the national research networks Photoacoustic Imaging in Biology and Medicine, project S10505 and Geometry and Simulation S11704.

Maarten V. de Hoop
Center for Computational and Applied Mathemematics, Purdue University, West Lafayette, IN 47907
E-mail: mdehoop@purdue.edu

Lingyun Qiu
Institute for Mathematics and its Applications, University of Minnesota, Minneapolis, MN 55455
E-mail: qiu.lingyun@ima.umn.edu

Otmar Scherzer
Computational Science Center, University of Vienna, Nordbergstr. 15, A-1090 Vienna, Austria
E-mail: otmar.scherzer@univie.ac.at

1 Introduction

We consider nonlinear inverse problems described by operator equations in Banach spaces. Assuming conditional stability of the inverse problem, we introduce a nonlinear projected steepest descent iteration and analyze its convergence. We take the point of view of reconstructing an approximation of the solution to the inverse problem in a compact, convex subset of the domain on which the operator is defined and where the stability holds. Assuming that we can identify a nested sequence of compact, convex subsets on which the stability holds such that the stability constant grows in a controlled way, we then extend our analysis to a multi-level approach which mitigates this growth via successive approximation. We account also for the possibility that a parameter in the operator which defines the inverse problems, and changes the data, affects for a given compact, convex subset the accuracy of approximation, and the stability constant. In the sequel of the paper we show with Example 4.6 that the radius of convergence can be enlarged by multi-level techniques. In fact in this example the convergence radius is exponentially decreasing as a function of the levels, and the conditions there show that multi-level iteration stay within these sets, starting from a large initial set. In our analysis, we incorporate inaccuracy of the data. Our analysis applies, for example, to electrical impedance tomography (EIT) and inverse boundary value problems for the Helmholtz equation using multiple frequencies.

Initially, we consider a class of inverse problems defined by a nonlinear map from parameter or model functions to the data. The parameter functions and data are contained in certain Banach spaces. This situation can be modeled mathematically by the operator equation

$$F(x) = y, \quad x \in \mathcal{D}(F), \; y \in Y, \tag{1.1}$$

with domain $\mathcal{D}(F) \subset X$, where X is a p-convex and q-smooth Banach space, with $p, q > 1$ and Y is an arbitrary Banach space. We assume that F is continuous, and that F is locally Fréchet differentiable. We do *not* assume that the data are attainable, that is, y may not belong to the range of F. We assume that there exists a compact, convex subset $Z \subset X$ such that

$$\Delta_p(x, \tilde{x}) \leq \mathfrak{C}^p \|F(x) - F(\tilde{x})\|^p, \quad \forall x, \tilde{x} \in Z. \tag{1.2}$$

Here Δ_p denotes the Bregman distance (defined below) and $p > 1$. This states conditional Lipschitz stability of the inverse problem. Motivated by [12], we employ a steepest descent iteration, here, to give an approximation to the solution of (1.1). More precisely, we construct a sequence of parameter functions by a projected gradient descent iteration with posterior stepsize.

In many inverse problems, logarithmic type stability is the optimal stability obtained with minimal assumptions on the domain or pre-image space; see, for example, [19]. By constraining the pre-image space, however, Lipschitz stability can be obtained; for the case of EIT, see [3,6] and for the case of inverse boundary value problems for the Helmholtz equation, see [5,4]. This is reflected by conditional stability given in (1.2). The mentioned projected gradient descent iteration can then be viewed as a projection regularization method, which is natural and avoids possibly artificial regularization techniques [16].

Our first main result concerns restricted convergence of the projected steepest descent iteration with a certain Hölder or Lipschitz type stability condition on a compact, convex subset. Moreover, we prove monotonicity of the residuals defined by the sequence induced by the iteration. This result is related to two areas of iterative regularization, which are steepest descent algorithms for solving nonlinear inverse problems [20,21,18,23] and projected iteration regularization techniques for the solution of inverse problems with convexity constraints. While the first three mentioned papers [20,21,18] provide a convergence analysis in a Hilbert space setting, the fourth one [23] provides results in a Banach space setting. However, these results are for Landweber iterations and iteratively regularized Gauss-Newton methods. As we show in Remark 3.4 our method is a generalization of the classical steepest descent method, as analyzed for instance by [14] for linear problems and [20,21] for nonlinear problems in a Hilbert space setting, i.e., with an adaptive damping parameter. Moreover, the developed steepest descent method, in the nonlinear setting, takes into account previous iterates. In this sense it is similar to conjugate gradient type methods. Steepest descent methods have been analyzed mostly in the context of *linear* inverse problems (see, for example, [13]) and later as accelerated methods in [11]. Accelerated methods have been modified for nonlinear problems by [24]. The main differences of our work to the above mentioned papers are the conditions under which we prove convergence rate. In fact, instead of source and nonlinearity conditions (as in [20,21]), we assume certain Hölder or Lipschitz stability of the inverse problem. This is a novel view point, which has been raised in [12].

In this paper, based on our first main result, we then introduce a multilevel algorithm. The motivation is to find progressively more accurate approximations to the model function, as the 'stable subset', Z, is incrementally enlarged. Interest in designing a hierarchy algorithm comes from two sources. In the beginning of the iterations, a coarse stable subset is preferred to ensure a stabilized problem, and hence a large convergence radius. Near the end of the iterations, a fine stable subset is expected to make a high accuracy approximation available. A fine stable subset usually does not fit the first several iterations since it results in a decay of the convergence radius. We assume that there are compact, convex subsets $\{Z_\alpha\}_{\alpha\in\mathbb{R}}$ of X, on which the restricted operator $F_\alpha = F\,|_{Z_\alpha}$ exhibits a certain Hölder or Lipschitz type stability estimate with stability constant $\mathfrak{C}_\alpha$, that is,

$$\Delta_p(x,\tilde{x}) \le \mathfrak{C}_\alpha^p \|F_\alpha(x) - F_\alpha(\tilde{x})\|^p, \quad \forall x,\tilde{x} \in Z_\alpha. \tag{1.3}$$

In fact, F_α need not be a restriction of F only, but can also account for a varying parameter in F which does affect the data. Here, we assume that $Z_{\alpha_1} \subset Z_{\alpha_2}$ and $\mathfrak{C}_{\alpha_1} \le \mathfrak{C}_{\alpha_2}$ if $\alpha_1 < \alpha_2$. In the context of discretization methods, Z_α stands for a finite-dimensional subspace of X and the number of basis vectors increases as α increases, while the projection can be an orthogonal projection on Z_α. In our second main result, we introduce a condition on the stability constants and on the approximation errors between neighboring levels. These conditions between levels are coupled and guarantee that the result from the previous level is a proper starting point for the present level. Thus, the algorithm can continue from level to level until the desired discrepancy criterion is satisfied.

2 Preliminaries

Let X and Y be Banach spaces. The duals of X and Y are denoted by X^* and Y^*, respectively. Their norms are denoted uniformly by $\|\cdot\|$. Throughout this paper, we assume that X is p-convex and q-smooth with $p, q > 1$, and hence it is uniformly smooth and uniformly convex. Furthermore, X is reflexive and its dual X^* has the same properties with the roles of p and q interchanged. Y is allowed to be an arbitrary Banach space; j_p will be a single-valued selection of the possibly set-valued duality mapping of Y with gauge function $t \mapsto t^{p-1}$, $p > 1$.

Several constants appear in the analysis. For the readers convenience we have grouped them notational wise:

1. $\mathfrak{C}$ denotes a constant for the Lipschitz stability of the inverse mapping of F (cf. (1.2), (1.3)),
2. $\mathfrak{L}$ and $\hat{\mathfrak{L}}$ are properties of the the operator F (cf. (3.1) and (3.2)).
3. C and G with and without subscripts denote properties of the Banach space (cf. (2.6), (2.7)).

2.1 Duality mappings

We denote the space of continuous linear operators $X \to Y$ by $\mathcal{L}(X, Y)$. Let $F : \mathcal{D}(F) \subset X \to Y$ be continuous. Here $\mathcal{D}(F)$ denotes the domain of definition of the nonlinear operator F. Let $h \in \mathcal{D}(F)$ and $k \in X$ and assume that $h + t(k - h) \in \mathcal{D}(F)$ for all $t \in (0, t_0)$ for some $t_0 > 0$, then we denote by $DF(h)(k)$ the directional derivative of F at $h \in \mathcal{D}(F)$ in direction $k \in \mathcal{D}(F)$, that is,

$$DF(h)(k) := \lim_{t \to 0^+} \frac{F(h + tk) - F(h)}{t}.$$

If $DF(h) \in \mathcal{L}(X, Y)$, then F is called Gâteaux differentiable at h. If, in addition, the limit is uniform for all k belonging a neighborhood of 0, F is called Fréhet differentiable at h. For $x \in X$ and $x^* \in X^*$, we write the dual pair as $\langle x, x^* \rangle = x^*(x)$. For a linear operator $A \in \mathcal{L}(X, Y)$, we write A^* for the dual operator $A^* \in \mathcal{L}(Y^*, X^*)$ and $\|A\| = \|A^*\|$ for the operator norm of F. We let $1 < p, q < \infty$ be conjugate exponents, that is,

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

For $p > 1$, the subdifferential mapping $J_p = \partial f_p : X \to 2^{X^*}$ of the convex functional $f_p : x \mapsto \frac{1}{p}\|x\|^p$ defined by

$$J_p(x) = \{x^* \in X^* \mid \langle x, x^* \rangle = \|x\| \cdot \|x^*\| \text{ and } \|x^*\| = \|x\|^{p-1}\} \tag{2.1}$$

is called the duality mapping of X with gauge function $t \mapsto t^{p-1}$. Generally, the duality mapping is set-valued. In order to let J_p be single valued, we need to introduce the notion of convexity and smoothness of Banach spaces.

One defines the convexity modulus δ_X of X by

$$\delta_X(\epsilon) = \inf_{x,\tilde{x} \in X} \{1 - \|\tfrac{1}{2}(x + \tilde{x})\| \mid \|x\| = \|\tilde{x}\| = 1 \text{ and } \|x - \tilde{x}\| \geq \epsilon\} \tag{2.2}$$

and the smoothness modulus ρ_X of X by

$$\rho_X(\tau) = \sup_{x,\tilde{x} \in X} \{\tfrac{1}{2}(\|x + \tau\tilde{x}\| + \|x - \tau\tilde{x}\| - 2) \mid \|x\| = \|\tilde{x}\| = 1\}. \tag{2.3}$$

Definition 2.1. *A Banach space X is said to be*

(a) uniformly convex if there exists an $\epsilon \in (0, 2]$ such that $\delta_X(\epsilon) > 0$,
(b) uniformly smooth if $\lim_{\tau \to 0} \frac{\rho_X(\tau)}{\tau} = 0$,
(c) convex of power type p or p-convex if there exists a constant $C > 0$ such that $\delta_X(\epsilon) \geq C\epsilon^p$,
(d) smooth of power type q or q-smooth if there exists a constant $C > 0$ such that $\rho_X(\tau) \leq C\tau^q$.

Let $p > 1$. In the following, we list some properties of the duality mapping and convex and smooth Banach spaces. For a detailed introduction to this topic, we refer to [10].

(a) For every $x \in X$, the set $J_p(x)$ is not empty and it is convex and weakly closed in X^*.
(b) Theorem of Milman-Pettis: If a Banach space is uniformly convex, it is reflexive.
(c) A Banach space X is uniformly convex (resp. uniformly smooth) if and only if X^* is uniformly smooth (resp. uniformly convex).
(d) If a Banach space X is uniformly smooth, $J_p(x)$ is single valued for all $x \in X$.
(e) If a Banach space X is uniformly smooth and uniformly convex, $J_p(x)$ is bijective and the inverse $J_p^{-1} : X^* \to X$ is given by $J_p^{-1} = J_q^*$ with J_q^* being the duality mapping of X^* with gauge function $t \mapsto t^{q-1}$, where $1 < p, q < \infty$ are conjugate exponents.

2.2 Bregman distances

Because the geometrical characteristics of Banach spaces are different from those of Hilbert spaces, it is often more appropriate to use the Bregman distance instead of the conventional norm-based functionals $\|x-\tilde{x}\|^p$ or $\|J_p(x)-J_p(\tilde{x})\|^p$ for convergence analysis. This idea goes back to Bregman [7].

Definition 2.2. *Let X be a uniformly 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*

$$\Delta_p(x,\tilde{x}) = \frac{1}{p}\|\tilde{x}\|^p - \frac{1}{p}\|x\|^p - \langle J_p(x), \tilde{x}-x\rangle, \quad \tilde{x} \in X, \tag{2.4}$$

where J_p denotes the duality mapping of X with gauge function $t \mapsto t^{p-1}$. Note, that under the general assumptions of this paper the duality mapping J_p is single valued.

In the following proposition, we summarize some facts concerning the Bregman distance and the relationship between the Bregman distance and the norm [1,2,8,25].

Proposition 2.3. *Let X be a uniformly smooth and uniformly convex Banach space. Then, for all $x,\tilde{x} \in X$, the following holds:*

(a)

$$\begin{aligned}\Delta_p(x,\tilde{x}) &= \frac{1}{p}\|\tilde{x}\|^p - \frac{1}{p}\|x\|^p - \langle J_p(x), \tilde{x}\rangle + \|x\|^p \\ &= \frac{1}{p}\|\tilde{x}\|^p + \frac{1}{q}\|x\|^p - \langle J_p(x), \tilde{x}\rangle.\end{aligned} \tag{2.5}$$

(b) $\Delta_p(x,\tilde{x}) \geq 0$ *and* $\Delta_p(x,\tilde{x}) = 0 \Leftrightarrow x = \tilde{x}$.

(c) Δ_p *is continuous in both arguments.*

(d) The following statements are equivalent

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

  (ii) $\lim_{n\to\infty}\Delta_p(x_n,x) = 0$,

  (iii) $\lim_{n\to\infty}\|x_n\| = \|x\|$ *and* $\lim_{n\to\infty}\langle J_p(x_n), x\rangle = \langle J_p(x), x\rangle$.

(e) If X is p-convex, there exists a constant $C_p > 0$ such that

$$\Delta_p(x,\tilde{x}) \geq \frac{C_p}{p}\|x-\tilde{x}\|^p. \tag{2.6}$$

(f) If X^* is q-smooth, there exists a constant $G_q > 0$ such that

$$\Delta_q(x^*,\tilde{x}^*) \leq \frac{G_q}{q}\|x^*-\tilde{x}^*\|^q, \tag{2.7}$$

for all $x^*,\tilde{x}^* \in X^*$.

The Bregman distance Δ_p is similar to a metric, but, in general, does not satisfy the triangle inequality nor symmetry. In a Hilbert space, $\Delta_2(x, \tilde{x}) = \frac{1}{2}\|x - \tilde{x}\|^2$.

2.3 Bregman Projection

In this subsection, we briefly introduce the Bregman projection and its properties, especially, the non-expansiveness. A comprehensive introduction to this topic, including a proof of Lemma 2.6, can be found in [8].

Definition 2.4. *Let X be a uniformly smooth Banach space and $p > 1$. Given a compact convex set $Z \subset X$ and Bregman distance Δ_p, which is defined in Definition 2.4, the Bregman projection of a point $x \in X$ onto Z is the point*

$$P_Z(x) = \arg\min\{\Delta_p(y, x) \mid y \in Z\}. \tag{2.8}$$

Definition 2.5. *Let $T : X \to X$ be an operator. The point $z \in X$ is called a non-expansivity pole of T if, for every $x \in X$,*

$$\Delta_p(T(x), T(z)) + \Delta_p(x, T(x)) \leq \Delta_p(x, z).$$

A operator T, which has at least one non-expansivity pole, is called totally non-expansive.

Lemma 2.6. *Let X be a uniformly smooth Banach space and $p > 1$ and $Z \subset X$ be a compact convex subset. The following statements hold:*

(a) The Bregman projection P_Z is well defined;
(b) P_Z is totally non-expansive and every point in Z is a non-expansivity pole of P_Z;
(c) For every $z \in Z$,

$$\Delta_p(P_Z(x), z) \leq \Delta_p(x, z), \quad \forall x \in X. \tag{2.9}$$

3 Convergence rate of a projected steepest descent iteration with a priori stopping rule

Here, we assume conditional stability, that is stability if the operator F is restricted to a compact, convex subset, Z, of X(see (3.3)). We introduce a projected steepest descent iteration and analyze its convergence. In this section, we keep Z fixed. We are concerned with an approximate solution, in Z, of the inverse problem subject to a discrepancy principle. In the analysis, we note that, if the iterate is close enough to the Z-best approximation, which gives the least fidelity term, then the steepest descent direction is nearly orthogonal to Z. After the projection P_Z applied (see (3.10)), the iterate is pulled back to Z and the contribution of this iteration to the decrease of the fidelity term is small. This usually causes a slow convergence rate. Hence, we use a discrepancy principle rather than study the convergence to the Z-best approximation. This promotes a uniform monotonicity estimate (see (3.13)).

Assumption 3.1. *Let*

$$\mathcal{B} = \mathcal{B}_\rho^{\Delta}(z^\dagger) = \{x \in X \mid \Delta_p(x, z^\dagger) \le \rho\} \subset \mathcal{D}(F)$$

for some $\rho > 0$, where ρ here will come into play as a convergence radius (see (3.9)*) and $z^\dagger$ is defined below.*

(a) The Fréchet derative, DF, of F is Lipschitz continuous on $\mathcal{B}$ and

$$\|DF(x)\| \le \hat{\mathfrak{L}} \quad \forall x \in \mathcal{B}, \tag{3.1}$$

$$\|DF(x) - DF(\tilde{x})\| \le \mathfrak{L}\|x - \tilde{x}\| \quad \forall x, \tilde{x} \in \mathcal{B}. \tag{3.2}$$

(b) F is weakly sequentially closed, i.e.,

$$\left.\begin{array}{l} x_n \rightharpoonup x, \\ F(x_n) \rightharpoonup y \end{array}\right\} \Rightarrow \left\{\begin{array}{l} x \in \mathcal{D}(F), \\ F(x) = y. \end{array}\right.$$

(c) Let Z denote a compact, convex subset of X. The inversion has the uniform Lipschitz type stability for elements in Z, i.e., there exists a constant $\mathfrak{C} > 0$ such that

$$\Delta_p(x, \tilde{x}) \le \mathfrak{C}^p \|F(x) - F(\tilde{x})\|^p \quad \forall x, \tilde{x} \in \mathcal{B} \cap Z. \tag{3.3}$$

For given data $y \in Y$, we assume that

$$\operatorname{dist}(y, F(Z)) \le \eta, \tag{3.4}$$

for some $\eta > 0$. Note that F is continuous and Z is compact. Hence there must exist a $z^\dagger \in Z$ such that

$$\|F(z^\dagger) - y\| = \operatorname{dist}(y, F(Z)). \tag{3.5}$$

Note, that this condition also accounts for data errors. In the case of noisy data, even if the stable subset Z equals to the space X, η can be a positive number associated with the noise level.

We introduce the following algorithm:

Algorithm 3.2. *We fix some abbreviations first: For x_k, $k = 0, 1, 2, \ldots$, fixed denote*

$$R_k = F(x_k) - y\,, \quad T_k = DF(x_k)^* j_p(F(x_k) - y)\,, \quad r_k = \|R_k\|\,, \quad t_k = \|T_k\|\,. \tag{3.6}$$

Moreover, we define

$$\tilde{\mathfrak{C}} := \frac{1}{2}\left(\frac{C_p}{p}\right)^{-2/p} \mathfrak{L}\mathfrak{C}^2\,, \tag{3.7}$$

and for $k = 0, 1, \ldots$

$$\begin{aligned} \hat{t}_k &:= G_q t_k^q\,, \\ u_k &:= -\tilde{\mathfrak{C}} r_k^2 + (1 - 2\tilde{\mathfrak{C}}\eta) r_k - \eta - \tilde{\mathfrak{C}}\eta^2\,, \\ v_k &:= \hat{t}_k^{-\frac{1}{q-1}} u_k^{\frac{1}{q-1}} r_k^{p^2-p}(r_k - \eta) - \frac{1}{q}\hat{t}_k^{-\frac{1}{q-1}} u_k^p r_k^{p^2-p}\,, \\ w_k &:= \frac{\mathfrak{L}}{2}\left(\frac{C_p}{p}\right)^{-2/p} \hat{t}_k^{-\frac{1}{q-1}} u_k^{\frac{1}{q-1}} r_k^{p^2-p}\,, \\ \mu_k &:= \hat{t}_k^{-\frac{1}{q-1}} u_k^{\frac{1}{q-1}} r_k^{\frac{p-1}{q-1}}\,. \end{aligned} \tag{3.8}$$

Now, the main steps of the algorithm:

($S0$) *Choose a starting point $x_0 \in Z$ such that*

$$\Delta_p(x_0, z^\dagger) < \rho := \frac{C_p}{p}(2\tilde{\mathfrak{C}}\hat{\mathfrak{L}})^{-p}\left(1 + \sqrt{1 - 8\tilde{\mathfrak{C}}\eta} - 4\eta\tilde{\mathfrak{C}}\right)^p\,, \tag{3.9}$$

where $z^\dagger$ is specified in Theorem 3.3 below.

($S1$) *Compute the new iterate via*

$$\begin{aligned} \tilde{x}_{k+1} &= J_q^*(J_p(x_k) - \mu_k T_k) \\ x_{k+1} &= \mathcal{P}_Z(\tilde{x}_{k+1}). \end{aligned} \tag{3.10}$$

($S2$) *Stop, if the discrepancy criterion*

$$\|F(x_k) - y\| \leq \hat{\eta},$$

is satisfied.

($S3$) *Set $k \leftarrow k + 1$ and go to step ($S1$).*

With the Hilbert space setting and a twice Fréchet-differentiable operator F the essence of the constants is more transparent. Let D^2F stand for the second-order Fréchet derivative of F and we have,

$$\tilde{\mathfrak{C}} := \frac{1}{2}\mathfrak{L}\mathfrak{C}^2 \approx \frac{1}{2}\frac{\|D^2F(x)\|}{\|DF(x)\|^2}\,, \tag{3.11}$$

which shows that the constant $\tilde{\mathfrak{C}}$ is a curvature to gradient condition, which degenerates for linear operators to zero. Thus any bound on $\tilde{\mathfrak{C}}$ restricts the nonconvexity of F. For aspects of curvature-size conditions see [9].

Fig. 1 Projected steepest descent iteration

Theorem 3.3. *Let X be a p-convex and q-smooth Banach space with $p, q > 1$ and Y be an arbitrary Banach space. Assume that Assumption 3.1 holds and the estimate (3.4) holds for some positive constant $\eta \in (0, (8\tilde{\mathfrak{C}})^{-1})$ and $z^\dagger \in Z$. Moreover, choose the tolerance $\hat{\eta} \geq 3\eta$.*

Algorithm 3.2 stops after a finite number

$$K(\hat{\eta}) := \min\{k \in \mathbb{N} \mid r_k := \|F(x_k) - y\| \leq \hat{\eta}\}. \tag{3.12}$$

That is when the discrepancy principle is satisfied. Moreover, strict monotonicity of the Bregman distance

$$\Delta_p(x_{k+1}, z^\dagger) \leq \Delta_p(x_k, z^\dagger) + w_k \Delta_p(x_k, z^\dagger)^{2/p} - v_k, \tag{3.13}$$

holds with

$$w_k \Delta_p(x_k, z^\dagger)^{2/p} - v_k < 0,$$

for all $k \leq K(\hat{\eta}) - 1$.

Proof. We use the same abbreviations for r_k and t_k as in Algorithm 3.2.

We start with a collection of elementary estimates that will be used frequently afterwards. With the abbreviations defined in (3.8), (3.7), inequalities

(2.6) and (3.3) yield

$$\begin{aligned}\frac{\mathfrak{L}}{2}\|x_k - z^\dagger\|^2 \leq& \frac{\mathfrak{L}}{2}\left(\Delta_p(x_k, z^\dagger)\frac{p}{C_p}\right)^{2/p} \\ \leq& \frac{\mathfrak{L}}{2}\left(\frac{C_p}{p}\right)^{-2/p}\mathfrak{C}^2\|F(x_k) - F(z^\dagger)\|^2 \\ \leq& \tilde{\mathfrak{C}}(r_k + \|F(z^\dagger) - y\|)^2 \\ \leq& \tilde{\mathfrak{C}}r_k^2 + 2\tilde{\mathfrak{C}}\eta r_k + \tilde{\mathfrak{C}}\eta^2 \\ =& r_k - u_k - \eta\ .\end{aligned} \tag{3.14}$$

With the mean value inequality and (2.6), it follows that

$$r_k \leq \|F(x_k) - F(z^\dagger)\| + \eta \leq \hat{\mathfrak{L}}\left(\Delta_p(x_k, z^\dagger)\frac{p}{C_p}\right)^{1/p} + \eta. \tag{3.15}$$

Using the definition of μ_k it follows that for $k = 0, 1, \ldots$,

$$\mu_k r_k^{p-1} = \hat{t}_k^{-\frac{1}{q-1}} u_k^{\frac{1}{q-1}} r_k^{p^2-p}\,, \quad \frac{G_q}{q}\mu_k^q t_k^q = \frac{1}{q}\hat{t}_k^{-\frac{1}{q-1}} u_k^p r_k^{p^2-p}\ . \tag{3.16}$$

Now, we start with the main body of the proof: We claim that

$$\Delta_p(x_m, z^\dagger) < \rho, \quad m = 0, 1, \ldots, K,$$

which we prove by induction. Note that (3.9) gives the base case. Assume the induction hypothesis that

$$\Delta_p(x_k, z^\dagger) < \rho.$$

With (3.15), we have that

$$r_k < \hat{\mathfrak{L}}(\rho\frac{p}{C_p})^{1/p} + \eta = \frac{1 + \sqrt{1 - 8\tilde{\mathfrak{C}}\eta}}{2\tilde{\mathfrak{C}}} - \eta. \tag{3.17}$$

Note that we can rewrite

$$u_k = -\tilde{\mathfrak{C}}\left(r_k - \frac{1 - \sqrt{1 - 8\tilde{\mathfrak{C}}\eta}}{2\tilde{\mathfrak{C}}} + \eta\right)\left(r_k - \frac{1 + \sqrt{1 - 8\tilde{\mathfrak{C}}\eta}}{2\tilde{\mathfrak{C}}} + \eta\right).$$

Then, (3.17), combined with the fact that

$$r_k > \hat{\eta} \geq 3\eta > \frac{1 - \sqrt{1 - 8\tilde{\mathfrak{C}}\eta}}{2\tilde{\mathfrak{C}}} - \eta$$

gives the positiveness of u_k. Note that this leads to the positiveness of v_k as following

$$\begin{aligned} v_k =&\hat{t}_k^{-\frac{1}{q-1}} u_k^{\frac{1}{q-1}} r_k^{p^2-p}(r_k - \eta - \frac{1}{q}u_k) \\ \geq&\hat{t}_k^{-\frac{1}{q-1}} u_k^{\frac{1}{q-1}} r_k^{p^2-p}(r_k - \eta - u_k) \\ =&\tilde{\mathfrak{C}}\hat{t}_k^{-\frac{1}{q-1}} u_k^{\frac{1}{q-1}} r_k^{p^2-p}(r_k + \eta)^2 > 0. \end{aligned}$$

Using (2.5) and (2.1) we obtain, for the sequence of residues,

$$\begin{aligned} &\Delta_p(\tilde{x}_{k+1}, z^\dagger) \\ =& \Delta_p(x_k, z^\dagger) + \frac{1}{q}\left(\|\tilde{x}_{k+1}\|^p - \|x_k\|^p\right) - \langle J_p(\tilde{x}_{k+1}) - J_p(x_k), z^\dagger\rangle \\ =& \Delta_p(x_k, z^\dagger) + \frac{1}{q}\left(\|J_p(\tilde{x}_{k+1})\|^q - \|J_p(x_k)\|^q\right) - \langle J_p(\tilde{x}_{k+1}) - J_p(x_k), z^\dagger\rangle. \end{aligned} \tag{3.18}$$

Applying (2.5) and (f) of Proposition 2.3 with $x^* = J_p(\tilde{x}_{k+1})$ and $\tilde{x}^* = J_p(x_k)$, we get

$$\begin{aligned} &\frac{1}{q}\left(\|J_p(\tilde{x}_{k+1})\|^q - \|J_p(x_k)\|^q\right) \\ &\qquad \leq \frac{G_q}{q}\|J_p(\tilde{x}_{k+1}) - J_p(x_k)\|^q + \langle J_p(\tilde{x}_{k+1}) - J_p(x_k), x_k\rangle. \end{aligned}$$

Substituting (3.10) and using this inequality in (3.18) yields

$$\begin{aligned} &\Delta_p(\tilde{x}_{k+1}, z^\dagger) - \Delta_p(x_k, z^\dagger) \\ =&\frac{G_q}{q}\|J_p(\tilde{x}_{k+1}) - J_p(x_k)\|^q + \langle J_p(\tilde{x}_{k+1}) - J_p(x_k), x_k - z^\dagger\rangle \\ =&\mu_k\left(\frac{G_q}{q}\mu_k^{q-1}t_k^q - \langle T_k, x_k - z^\dagger\rangle\right). \end{aligned} \tag{3.19}$$

We estimate the second term in (3.19). Using (2.6) and the Lipschitz type stability (3.3), and (3.4), we find that

$$\begin{aligned} &-\langle T_k, x_k - z^\dagger\rangle \\ =&-\langle j_p(R_k), DF(x_k)(x_k - z^\dagger)\rangle \\ =&-\langle j_p(R_k), R_k\rangle + \langle j_p(R_k), F(z^\dagger) - y\rangle) \\ &\qquad + \langle j_p(R_k), F(x_k) - F(z^\dagger) - DF(x_k)(x_k - z^\dagger)\rangle \\ \leq&-r_k^{p-1}\left(r_k - \eta - \frac{\mathfrak{L}}{2}\|x_k - z^\dagger\|^2\right). \end{aligned} \tag{3.20}$$

In the last step of above inequalities, we use the properties of the duality mapping, $\langle j_p(R_k), R_k\rangle = \|R_k\|^p$ and $\|j_p(R_k)\| = \|R_k\|^{p-1}$ (see (2.1)). From

(3.19) and (3.20), it follows that, for $k = 0, 1, 2, \ldots$,

$$\begin{aligned}&\Delta_p(\tilde{x}_{k+1}, z^\dagger) - \Delta_p(x_k, z^\dagger)\\ \le&\mu_k r_k^{p-1}\left(\frac{G_q}{q}\frac{\mu_k^{q-1}t_k^q}{r_k^{p-1}} - r_k + \eta + \frac{\mathfrak{L}}{2}\|x_k - z^\dagger\|^2\right),\end{aligned}\tag{3.21}$$

and hence, by (3.21), (2.6) and the non-expansiveness of the Bregman projection (2.9), we arrive at

$$\begin{aligned}&\Delta_p(x_{k+1}, z^\dagger) - \Delta_p(x_k, z^\dagger)\\ \le&\Delta_p(\tilde{x}_{k+1}, z^\dagger) - \Delta_p(x_k, z^\dagger)\\ \le&\mu_k r_k^{p-1}\left(\frac{G_q}{q}\frac{\mu_k^{q-1}t_k^q}{r_k^{p-1}} - r_k + \eta + \frac{\mathfrak{L}}{2}\left(\Delta_p(x_k, z^\dagger)\frac{p}{C_p}\right)^{2/p}\right).\end{aligned}\tag{3.22}$$

Using the identities in (3.16) and abbreviations (3.8), (3.7), from the above inequality, we derive that

$$\begin{aligned}&\Delta_p(x_{k+1}, z^\dagger) - \Delta_p(x_k, z^\dagger)\\ \le&\frac{1}{q}\hat{t}_k^{-\frac{1}{q-1}}u_k^p r_k^{p^2-p} - \hat{t}_k^{-\frac{1}{q-1}}u_k^{\frac{1}{q-1}}r_k^{p^2-p}(r_k - \eta)\\ &+\frac{\mathfrak{L}}{2}\left(\frac{C_p}{p}\right)^{-2/p}\hat{t}_k^{-\frac{1}{q-1}}u_k^{\frac{1}{q-1}}r_k^{p^2-p}\Delta_p(x_k, z^\dagger)^{2/p}\\ =&-v_k + w_k\Delta_p(x_k, z^\dagger)^{2/p}.\end{aligned}\tag{3.23}$$

We finish the proof of the monotonicity of $\Delta_p(x_k, z^\dagger)$ by showing that

$$-v_k + w_k\Delta_p(x_k, z^\dagger)^{2/p} < 0.$$

In fact,

$$\begin{aligned}&w_k\Delta_p(x_k, z^\dagger)^{2/p}\\ \le&w_k\mathfrak{C}^2\|F(x_k) - F(z^\dagger)\|^2\\ \le&\frac{\mathfrak{L}}{2}\left(\frac{C_p}{p}\right)^{-2/p}\mathfrak{C}^2\,(r_k + \eta)^2\hat{t}_k^{-\frac{1}{q-1}}u_k^{\frac{1}{q-1}}r_k^{p^2-p}\\ =&(-u_k + r_k - \eta)\hat{t}_k^{-\frac{1}{q-1}}u_k^{\frac{1}{q-1}}r_k^{p^2-p}.\end{aligned}\tag{3.24}$$

Hence

$$\begin{aligned}&-v_k + w_k\Delta_p(x_k, z^\dagger)^{2/p}\\ \le&-v_k - \hat{t}_k^{-\frac{1}{q-1}}u_k^{\frac{1}{q-1}+1}r_k^{p^2-p} + (r_k - \eta)\hat{t}_k^{-\frac{1}{q-1}}u_k^{\frac{1}{q-1}}r_k^{p^2-p}\\ =&-\frac{1}{p}\hat{t}_k^{-\frac{1}{q-1}}u_k^p r_k^{p^2-p} < 0.\end{aligned}\tag{3.25}$$

The above monotonicity of $\Delta_p(x_k, z^\dagger)$ with the induction hypothesis completes the induction.

It is left to show that Algorithm 3.2 stops after a finite number of iterations(i.e. $K(\hat{\eta})$ iterations). We prove this by contradiction. Let us assume that the number of iterations of Algorithm 3.2 is infinite, and hence,

$$r_k > \hat{\eta}, \quad \forall k \geq 0. \tag{3.26}$$

Then, from the monotonicity of the Bregman distances (3.13) and (3.25), we have that

$$0 \leq \Delta_p(x_k, z^\dagger) \leq \Delta_p(x_0, z^\dagger) - \frac{1}{p}\sum_{n=0}^{k-1} \hat{t}_n^{-\frac{1}{q-1}} u_n^p r_n^{p^2-p}, \quad \forall k > 0.$$

It follows that

$$\sum_{n=0}^{\infty} \hat{t}_n^{-\frac{1}{q-1}} u_n^p r_n^{p^2-p} < \infty.$$

From the lower boundedness of r_k, (3.26), and the uniform boundedness of the Frechét derivative $DF(x_k)$, (3.1), we conclude that both $\hat{t}_n^{-\frac{1}{q-1}}$ and $r_n^{p^2-p}$ are uniformly positively bounded from below. Hence u_k converges to 0 as k goes to infinity. By writing

$$u_k = -\tilde{\mathfrak{C}}\left(r_k - \frac{1-\sqrt{1-8\tilde{\mathfrak{C}}\eta}}{2\tilde{\mathfrak{C}}} + \eta\right)\left(r_k - \frac{1+\sqrt{1-8\tilde{\mathfrak{C}}\eta}}{2\tilde{\mathfrak{C}}} + \eta\right)$$

we have that

$$\lim_{k\to\infty} r_k = \frac{1-\sqrt{1-8\tilde{\mathfrak{C}}\eta}}{2\tilde{\mathfrak{C}}} - \eta < 3\eta \leq \hat{\eta},$$

which is a contradiction.

□

Remark 3.4. *We refer to Algorithm 3.2 as a steepest descent algorithm in the sense that it is a generalization of the steepest descent algorithm for linear inverse problems. Indeed, let F be linear and assume that we have an unconstrained problem. Then both $\mathfrak{L}$ and η can be chosen to be equal to zero. Then we have*

$$\mu_k = \left(\frac{r_k^p}{t_k^q G_q}\right)^{1/(q-1)}, \quad k = 0, 1, 2, \ldots,$$

with $r_k = \|Fx_k - y\|$ and $t_k = \|F^* j_p(Fx_k - y)\|$. In particular, for a Hilbert space setting, where

$$p = q = 2, \quad C_p = G_q = 1\,, J_p = J_q = Id,$$

we get

$$\mu_k = \frac{r_k^2}{t_k^2}\,, \quad k = 0, 1, 2, \ldots,$$

which is the standard parameter choice of the steepest descent method [14]. See also [15] for efficient adaptations of the Landweber iteration.

In the Hilbert space setting, moreover, the condition (3.9) requires that $\tilde{\mathfrak{C}} < \frac{1}{8\eta}$*, which in some sense restricts the* curvature. *Note that for* $p = 2$ *we have* $\frac{1}{\|DF(x)\|} \approx \|x - \tilde{x}\|^2 / \|F(x) - F(\tilde{x})\|^2 \leq \tilde{\mathfrak{C}}$ *and therefore* $\frac{\|D^2F(x)\|}{\|DF(x)\|} \leq \tilde{\mathfrak{C}}\mathfrak{L}$*, where* $\|DF(x)\|$ *denotes the operator norm of a directional derivative in direction* $x - \tilde{x}$*, and* D^2F *is the second derivative in the same direction. Thus condition (3.9) can be interpreted as a curvature to size condition (see [9] for the curvature to size concept for variational regularization).*

Remark 3.5. *We refer to (3.9) as a generalized radius of convergence from the nonlinear Landweber iteration to a steepest descent algorithm in Banach spaces. Indeed, let* η *be equal to zero. Then (3.9) can be reduced to*

$$\Delta_p(x_0, z^\dagger) < \rho = \hat{\mathfrak{L}}^{-p} \frac{C_p}{p} \tilde{\mathfrak{C}}^{-p} = \left(\frac{C_p}{p}\right)^3 \left(\frac{\hat{\mathfrak{L}}\mathfrak{L}\mathfrak{C}^2}{2}\right)^{-p},$$

which coincides the convergence radius for the nonlinear Landweber iteration in Banach spaces[12].

4 Extension to a multi-level algorithm

In this section, based on the results of the previous section, we introduce a multi-level algorithm. The basic idea is to design an algorithm which incorporates both high accuracy and large convergence radius. This can be done by using a varying stable subset and balancing the decay of the approximation error and the blow-up of the stability constant.

We consider a set, $\{Z_\alpha\}_{\alpha \geq 0}$, of compact and convex subsets of X, and an operator family $\{F_\alpha\}_{\alpha \geq 0}$, where F_α is obtained as the restriction of F on Z_α, $F_\alpha = F|_{Z_\alpha}$. We let

$$\mathcal{B} = \mathcal{B}^{\Delta}_{\rho_0}(x^\dagger) = \{x \in X \mid \Delta_p(x, x^\dagger) \leq \rho_0\} \subset \mathcal{D}(F)$$

for some $\rho_0 > 0$, which is specified in Theorem 4.4 and invoke

Assumption 4.1. *(a)* F *is weakly sequentially closed, that is,*

$$\left.\begin{array}{l} x_n \rightharpoonup x, \\ F(x_n) \rightharpoonup y \end{array}\right\} \Rightarrow \left\{\begin{array}{l} x \in \mathcal{D}(F), \\ F(x) = y. \end{array}\right.$$

(b) *The Fréchet derivative, DF_α, of F_α is Lipschitz continuous on $\mathcal{B}\cap Z_\alpha$ and*

$$\|DF_\alpha(x)\| \leq \hat{\mathfrak{L}}_\alpha \quad \forall x \in \mathcal{B}\cap Z_\alpha, \tag{4.1}$$

$$\|DF_\alpha(x) - DF_\alpha(\tilde{x})\| \leq \mathfrak{L}_\alpha \|x-\tilde{x}\| \quad \forall x, \tilde{x} \in \mathcal{B}\cap Z_\alpha. \tag{4.2}$$

(c) *The inversion has the uniform Lipschitz type stability for elements in Z_α, that is, there exists a constant $\mathfrak{C}_\alpha > 0$ such that*

$$\varDelta_p(x,\tilde{x}) \leq \mathfrak{C}_\alpha^p \|F_\alpha(x) - F_\alpha(\tilde{x})\|^p \quad \forall x, \tilde{x} \in \mathcal{B}\cap Z_\alpha. \tag{4.3}$$

For the stability constants, $\{\mathfrak{C}_\alpha\}$, and the approximation error, $\{\eta_\alpha\}$, we introduce

Assumption 4.2. *(a) Let $\eta_\alpha = \eta_\alpha(y)$ be defined by*

$$\eta_\alpha = \operatorname{dist}(y, F_\alpha(Z_\alpha)), \quad y \in Y;$$

Moreover, we assume that η_α is non-negative and monotonically decreasing with respect to α for every fixed $y \in Y$.

(b) If $Z_{\alpha_1} \subset Z_{\alpha_2}$ then $\mathfrak{C}_{\alpha_1} \leq \mathfrak{C}_{\alpha_2}$.

(c) If $\alpha_1 < \alpha_2$ then $Z_{\alpha_1} \subset Z_{\alpha_2}$ and therefore also $\eta_{\alpha_1} \geq \eta_{\alpha_2}$.

Typically, the subsets Z_α are finite dimensional and the stability constant for the inversion grows with the dimension of these subsets. The nature of our multi-level algorithm is intimately connected to finding sparse, albeit approximate, representations of the solution to the inverse problem, mitigating the mentioned growth of the stability constants. Indeed, the objective is very similar to *multi-level* techniques for solving inverse problems [22, 16, 17], where one exploits that the finite-dimensional problems are stable and that the outcome of an iteration on a coarse level gives a good initial guess on a finer level. In this section, we combine any known controllable factors to an abstract index α of the operator family and design a progressive iteration method with the aid of the result from the previous section.

In the following algorithm, we refer to the parameter α as an index and only nonnegative integer valued α is considered.

Algorithm 4.3. ($S0$) *Use $x_{0,0}$ as the starting point. Set $\alpha = 0$.*

($S1$) *Iteration. Use F_α and Z_α as the modelling operator and convex subset to run Algorithm 3.2 with the discrepancy criterion given by*

$$K_\alpha = \min\{k \in \mathbb{N} \mid \|F_\alpha(x_{\alpha,k}) - y\| \leq (3+\varepsilon)\eta_\alpha\}, \tag{4.4}$$

where $\varepsilon > 0$ is a given uniform relaxation constant.

(N-level iteration)^{K_N}

(α-level iteration)$^{K_\alpha}$

prolongation

Fig. 2 A illustration of Algorithm 4.3

($S2$) *Stop, if the discrepancy criterion*

$$\|F_\alpha(x_{\alpha,K_\alpha}) - y\| \le \tilde{\eta},$$

is satisfied, where the tolerance $\tilde{\eta}$ *is given.*

($S3$) *Set* $x_{\alpha+1,0} = x_{\alpha,K_\alpha}$, $\alpha = \alpha + 1$ *and go to step* ($S1$).

This algorithm is illustrated in Figure 2.

Theorem 4.4. *Let* X *be a* p*-convex and* q*-smooth Banach space with* $p, q > 1$. *Moreover, let* Y *be an arbitrary Banach space. Assume that Assumptions 4.1 and 4.2 hold. Assume that there exists a subset of operators,* $\{F_\alpha\}_{\alpha=0}^N$ *family* $\{F_\alpha\}$ *such that*

(a) The starting point $x_{0,0}$ *is within the first convergence radius, that is,*

$$\Delta_p(x_{0,0}, z_0^\dagger) < \rho_0, \tag{4.5}$$

where $z_0^\dagger$ *denotes the* Z_0 *best approximating solution, i.e.,*

$$\|F_0(z_0^\dagger) - y\| = \mathrm{dist}(y, F_0(Z_0)),$$

and the Z_0 *convergence radius* ρ_0 *is defined by*

$$\rho_0 := \frac{C_p}{p}\hat{\mathfrak{L}}_0^{-p}\left(\frac{1+\sqrt{1-8\tilde{\mathfrak{C}}_0\eta_0}}{2\tilde{\mathfrak{C}}_0} - 2\eta_0\right)^p,$$

with $\tilde{\mathfrak{C}}_0 = \frac{1}{2}\left(\frac{C_p}{p}\right)^{-2/p}\mathfrak{L}_0\mathfrak{C}_0^2$. *Assume that* $8\tilde{\mathfrak{C}}_0\eta_0 < 1$;

(b) For every two neighbor levels Z_α and $Z_{\alpha+1}$, $\alpha = 0, \dots, N-1$, the constants η_α and $\eta_{\alpha+1}$, $\hat{\mathfrak{L}}_{\alpha+1}$, $\mathfrak{L}_{\alpha+1}$, $\mathfrak{C}_{\alpha+1}$ satisfy the following inequalities

$$8\tilde{\mathfrak{C}}_{\alpha+1}\eta_{\alpha+1} < 1, \tag{4.6}$$

and

$$(3+\varepsilon)\eta_\alpha < \left(\frac{C_p}{p}\right)^{1/p} (\hat{\mathfrak{L}}_{\alpha+1}\mathfrak{C}_{\alpha+1})^{-1} \left(\frac{1+\sqrt{1-8\tilde{\mathfrak{C}}_{\alpha+1}\eta_{\alpha+1}}}{2\tilde{\mathfrak{C}}_{\alpha+1}} - 2\eta_{\alpha+1}\right) - \eta_{\alpha+1}, \tag{4.7}$$

where $\tilde{\mathfrak{C}}_{\alpha+1} = \frac{1}{2}\left(\frac{C_p}{p}\right)^{-2/p} \mathfrak{L}_{\alpha+1}\mathfrak{C}^2_{\alpha+1}$.

(c) N is the first positive integer such that $\eta_N \le (3+\varepsilon)^{-1}\tilde{\eta}$, that is,

$$(3+\varepsilon)\eta_\alpha > \tilde{\eta} \quad \forall \alpha < N$$

and

$$(3+\varepsilon)\eta_N \le \tilde{\eta}.$$

Then, Algorithm 4.3 has the property that it stops after a finite number of iterations when the discrepancy criterion

$$\|F_N(x_{N,K_N}) - y\| \le \tilde{\eta} \tag{4.8}$$

is satisfied.

In the above theorem, we focus on the condition of continuing the iteration between levels. Note that, with a varying stable subset and hence a sequence of approximation errors $\{\eta_\alpha\}$, the tolerance $\tilde{\eta}$ is not required to be bounded from below as in Section 3. In Theorem 3.3, we stop the iteration with a tolerance $\hat{\eta}$, three times the approximation error η, rather than the optimal Z-best approximation $z^\dagger$. In this multi-level scheme, if the approximation error η_α tends to zero as α goes to infinity, one can conclude the proper convergence of $\{x_{\alpha,K_\alpha}\}$ to the true solution by applying Theorem 4.4 with arbitrary small tolerance $\tilde{\eta}$.

The strategy of the proof is to estimate the decreasing objective function $\|F_\alpha(x_{\alpha,K_\alpha}) - y\|$ level by level. That is, one applies Theorem 3.3 to guarantee that the discrepancy criterion (4.4) is attained with a finite number of iterations on each level. Then, with (4.7) and (4.4), we show that the initial point $x_{\alpha+1,0}$ on $(\alpha+1)$-level, which coincides with the iteration result x_{α,K_α} on α-level, is within the convergence radius $\rho_{\alpha+1}$. Therefore, the procedure continues until (4.8) is satisfied.

Proof. We first adapt the convergence radius, ρ, in Theorem 3.3 to a α-level convergence radius ρ_α. For any α-level, $\alpha = 0, 1, 2, \ldots, N$, one can use Algorithm 3.2 to obtain an approximate solution to the operator equation

$$F_\alpha(x) = y, \quad x \in Z_\alpha,$$

with a given starting point $x_{\alpha,0}$ and the discrepancy criterion given in (4.4). If the starting point $x_{\alpha,0}$ satisfy

$$\Delta(x_{\alpha,0}, z_\alpha^\dagger) < \rho_\alpha := \frac{C_p}{p}\hat{\mathfrak{L}}_\alpha^{-p}\left(\frac{1+\sqrt{1-8\tilde{\mathfrak{C}}_\alpha\eta_\alpha}}{2\tilde{\mathfrak{C}}_\alpha} - 2\eta_\alpha\right)^p, \tag{4.9}$$

where $z_\alpha^\dagger$ denotes the best Z_α-approximation, then Theorem 3.3 can be applied to show that Algorithm 3.2 stops after a finite number of iterations with

$$\|F_\alpha(x_{\alpha,k}) - y\| \le (3+\varepsilon)\eta_\alpha$$

satisfied.

Next, we show that, in particular with condition (4.7), if the starting point for the present level, $x_{\alpha,0}$, is within the convergence radius, then the starting point for the next level, $x_{\alpha+1,0}$, which is equal to x_{α,K_α}, is within the convergence radius for the next level. That is to say,

$$\Delta_p(x_{\alpha,0}, z_\alpha^\dagger) \le \rho_\alpha$$

implies

$$\Delta_p(x_{\alpha+1,0}, z_{\alpha+1}^\dagger) \le \rho_{\alpha+1},$$

for all $\alpha < N$. Indeed, for any $\alpha < N$, according to (4.9) and Theorem 3.3, after K_α steps, the α-level discrepancy criterion,

$$\|F_\alpha(x_{\alpha,K_\alpha}) - y\| \le (3+\varepsilon)\eta_\alpha, \tag{4.10}$$

is satisfied. Note that we set the $(\alpha+1)$-level starting point $x_{\alpha+1,0}$ to be the iteration result of α-level, x_{α,K_α}. Moreover, F_α is the restriction of F on Z_α. Hence, $F_{\alpha+1}(x_{\alpha+1,0}) = F_\alpha(x_{\alpha,K_\alpha})$. Then, with the above inequality (4.10) and (4.3), we estimate

$$\begin{aligned}
&\Delta_p(x_{\alpha+1,0}\,,\, z_{\alpha+1}^\dagger)^{1/p}\\
\le&\mathfrak{C}_{\alpha+1}\|F_{\alpha+1}(x_{\alpha+1,0}) - F_{\alpha+1}(z_{\alpha+1}^\dagger)\|\\
\le&\mathfrak{C}_{\alpha+1}(\|F_{\alpha+1}(x_{\alpha+1,0}) - y\| + \|F_{\alpha+1}(z_{\alpha+1}^\dagger) - y\|)\\
\le&\mathfrak{C}_{\alpha+1}((3+\varepsilon)\eta_\alpha + \eta_{\alpha+1}).
\end{aligned} \tag{4.11}$$

Note that (4.7) leads to the inequality

$$\mathfrak{C}_{\alpha+1}((3+\varepsilon)\eta_\alpha + \eta_{\alpha+1}) \le \left(\frac{C_p}{p}\right)^{1/p}\hat{\mathfrak{L}}_{\alpha+1}^{-1}\left(\frac{1+\sqrt{1-8\tilde{\mathfrak{C}}_{\alpha+1}\eta_{\alpha+1}}}{2\tilde{\mathfrak{C}}_{\alpha+1}} - 2\eta_{\alpha+1}\right)$$

Substituting this into (4.11), we have that

$$\Delta_p(x_{\alpha+1,0}\,,\, z^\dagger_{\alpha+1})^{1/p} \leq \left(\frac{C_p}{p}\right)^{1/p} \hat{\mathfrak{L}}^{-1}_{\alpha+1} \left(\frac{1+\sqrt{1-8\tilde{\mathfrak{C}}_{\alpha+1}\eta_{\alpha+1}}}{2\tilde{\mathfrak{C}}_{\alpha+1}} - 2\eta_{\alpha+1} \right) = \rho^{1/p}_{\alpha+1}. \quad (4.12)$$

For the last N-level, we apply Theorem 3.3 again to find that

$$\|F_N(x_{N,K_N}) - y\| \leq (3+\varepsilon)\eta_N \leq \tilde{\eta}.$$

□

Remark 4.5. *We interpret that Algorithm 4.3 is designed to achieve the optimal (or nearly optimal) accuracy for a feasible starting point. Usually, the finest level bears both the smallest approximation error, which corresponds to the optimal accuracy, and the largest stability constant. Note that the definition of the convergence radius (3.9) shows its algebraically decaying property with respect to the stability constant. There are cases when only a rough starting point is available. For these cases, one may fail to obtain a reasonable result using Algorithm 3.2 directly on the finest level but Algorithm 4.3 leads to a good approximation solution. The condition (4.7) can be interpreted as a strategy for selecting next finer level, which is characterized by its stability constant constants $\mathfrak{C}_{\alpha+1}$, approximation error $\eta_{\alpha+1}$ and $\hat{\mathfrak{L}}_{\alpha+1}$, $\mathfrak{L}_{\alpha+1}$.*

Theorem 4.4, especially (4.7), indicates that a sufficient condition for the existence of such a selection of operators is that the tolerated best-Z_α-approximation is within the convergence radius according to $Z_{\alpha+1}$. In fact, this condition comes from a bootstrap type competition between η_α and ρ_α.

We give an example of how conditions in Theorem 4.4 can be satisfied. The prototype for this example is the inverse boundary value problem for the Helmholtz equation with multi-frequency data. The following dynamic models for the constants can be examined. The detail will be discussed in another paper.

Example 4.6. *Assume that X and Y are Banach spaces and that we can reindex the convex subsets $\{Z_\alpha\}$ such that Assumptions 4.1 and 4.2 hold. Moreover, for a given tolerance $\tilde{\eta} > 0$, the following conditions hold:*

(i) Given starting point $x_{0,0}$ is within the first convergence radius ρ_0, i.e.,

$$\Delta_p(x_{0,0}, z^\dagger_0) < \rho_0 := \frac{C_p}{p} \hat{\mathfrak{L}}^{-p}_0 \left(\frac{1+\sqrt{1-8\tilde{\mathfrak{C}}_0\eta_0}}{2\tilde{\mathfrak{C}}_0} - 2\eta_0 \right)^p.$$

(ii) The approximation error $\eta_\alpha = \lambda e^{-3\alpha}$ for some constant $\lambda >> \tilde{\eta}$.

(iii) The stability constant $\mathfrak{C}_\alpha = 2e^\alpha$,

(iv) The dynamic models of the constants $\hat{\mathfrak{L}}_\alpha$ and $\mathfrak{L}_\alpha$, which are related to the Lipschitz continuity of the Fréchet derivative DF_α, are given by

$$\hat{\mathfrak{L}}_\alpha = \alpha + 1 \quad \textit{and} \quad \mathfrak{L}_\alpha = \tau(\alpha+1)^2,$$

for some constant τ such that

$$0 < \tau < \frac{1}{16\lambda}\left(\frac{C_p}{p}\right)^{2/p}.$$

Now, we can choose the operators $\{F_\alpha\}_{\alpha=0}^N$ defined by $F_\alpha = F\mid_{Z_\alpha}$ and set the uniform relaxation constant $\varepsilon = 1$ (see (4.4)) to run Algorithm 4.3, where N is the first integer such that $4\eta_N \le \tilde{\eta}$ is satisfied. Applying Theorem 4.4, we conclude that

$$\|F_N(x_{N,K_N}) - y\| \le \tilde{\eta}$$

is satisfied after a finite number of iterations.

In this example, we can quantify the helpful constant $\tilde{\mathfrak{C}}_\alpha$ and the convergence radius ρ_α by

$$\tilde{\mathfrak{C}}_\alpha = 2\tau\left(\frac{C_p}{p}\right)^{-2/p}(\alpha+1)e^{2\alpha}$$

and

$$\rho_\alpha = \frac{C_p}{p}\hat{\mathfrak{L}}_\alpha^{-p}\left(\frac{1+\sqrt{1-8\tilde{\mathfrak{C}}_\alpha\eta_\alpha}}{2\tilde{\mathfrak{C}}_\alpha} - 2\eta_\alpha\right)^p.$$

Noting that

$$\frac{1}{2} < 1 - 4\tilde{\mathfrak{C}}_\alpha\eta_\alpha < 1 + \sqrt{1-8\tilde{\mathfrak{C}}_\alpha\eta_\alpha} - 4\tilde{\mathfrak{C}}_\alpha\eta_\alpha < 2 - 4\tilde{\mathfrak{C}}_\alpha\eta_\alpha < 2,$$

for $\alpha = 0, 1, \ldots, N$, we conclude that, for the convergence radius ρ_α, the dynamic model is

$$(8\tau)^{-p}\left(\frac{C_p}{p}\right)^3(\alpha+1)^{-3p}e^{-2\alpha p} < \rho_\alpha < (2\tau)^{-p}\left(\frac{C_p}{p}\right)^3(\alpha+1)^{-3p}e^{-2\alpha p}.$$

The convergence radius decays exponentially as the level number increases. This helps us to understand how to enlarge the convergence radius by involving this multi-level scheme. Let us assume that we are in a situation where only a rough starting point $\tilde{x}$ is available such that

$$\Delta_p(\tilde{x}, z_0^\dagger) < (8\tau)^{-p}\left(\frac{C_p}{p}\right)^3 < \rho_0 \tag{4.13}$$

but

$$\Delta(\tilde{x}, z_N^\dagger) > (2\tau)^{-p}\left(\frac{C_p}{p}\right)^3(N+1)^{-3p}e^{-2Np} > \rho_N. \tag{4.14}$$

Without applying a multi-level scheme, if we run Algorithm 3.2 for single 0-level, by (4.13), Theorem 3.3 can be applied but the optimal residue estimate we can expect can not be smaller than the 0-level approximation error $\eta_0 = \lambda >> \tilde{\eta}$; if we run Algorithm 3.2 for single N-level, according to (4.14), there is no guarantee that Algorithm 3.2 will stop after a finite number of iterations nor yield a reasonable result. Hence a multilevel approach, as Algorithm 4.3, is proposed to obtain a high-accuracy arroximation x_{N,K_N} satisfying

$$\|F(x_{N,K_N}) - y\| \leq \tilde{\eta},$$

with a relatively larger convergence radius $\rho_0 >> \rho_N$.

5 Discussion

We discuss a steepest descent iteration method for solving nonlinear operator equations in Banach spaces. Provided that the nonlinearity of the forward operator obeys a Lipschitz type stability in a convex and compact subset of the preimage space, we could prove a restricted convergence result and provide an estimate of the error decease. Based on the analysis of the radius of convergence, we introduce a multilevel method and obtain a sufficient condition on the choices of the parameters, mainly on the approximation errors and stability constants.

Acknowledgments

The research was initiated at the Isaac Newton Institute for Mathematical Sciences (Cambridge, England) during a programme on Inverse Problems in Fall 2011. The authors would like to thank the anonymous referees for their valuable comments and suggestions to improve the quality of the paper.

References

1. Alber, Y. I., and Butnariu, D. Convergence of Bregman projection methods for solving consistent convex feasibility problems in reflexive Banach spaces. *J. Optim. Theory Appl. 92* (January 1997), 33–61.
2. Alber, Y. I., Kartsatos, A. G., and Litsyn, E. Iterative solution of unstable variational inequalities on approximately given sets. *Abstr. Appl. Anal. 1*, 1 (1996), 45–64.
3. Alessandrini, G., and Vessella, S. Lipschitz stability for the inverse conductivity problem. *Adv. in Appl. Math. 35*, 2 (2005), 207–241.
4. Ammari, H., Bahouri, H., Dos Santos Ferreira, D., and Gallagher, I. Stability estimates for an inverse scattering problem at high frequencies. *ArXiv e-prints* (May 2012).
5. Beretta, E., de Hoop, M. V., and Qiu, L. Lipschitz stability of an inverse boundary value problem for a Schrödinger-type equation. *SIAM J. Math. Anal. 45*, 2 (2013), 679–699.

6. Beretta, E., and Francini, E. Lipschitz stability for the electrical impedance tomography problem: the complex case. *Comm. Partial Differential Equations 36*, 10 (2011), 1723–1749.
7. Bregman, L. M. The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. *USSR Computational Mathematics and Mathematical Physics 7*, 3 (1967), 200 – 217.
8. Butnariu, D., Iusem, A. N., and Resmerita, E. Total convexity for powers of the norm in uniformly convex Banach spaces. *J. Convex Anal. 7*, 2 (2000), 319–334.
9. Chavent, G., and Kunisch, K. On weakly nonlinear inverse problems. *SIAM. J. Appl. Math. 56*, 2 (1996), 542–572.
10. Cioranescu, I. *Geometry of Banach spaces, duality mappings and nonlinear problems*, vol. 62 of *Mathematics and its Applications*. Kluwer Academic Publishers Group, Dordrecht, 1990.
11. Daubechies, I., Fornasier, M., and Loris, I. Accelerated projected gradient method for linear inverse problems with sparsity constraints. *J. Fourier Anal. Appl. 14*, 5-6 (2008), 764–792.
12. de Hoop, M. V., Qiu, L., and Scherzer, O. Local analysis of inverse problems: Hölder stability and iterative reconstruction. *Inverse Problems 28*, 4 (2012), 045001.
13. Eicke, B. Iteration methods for convexly constrained ill-posed problems in Hilbert space. *Numer. Funct. Anal. Optim. 13*, 5-6 (1992), 413–429.
14. Gilyazov, S. F. Iterative solution methods for inconsistent linear equations with nonself-adjoint operator. *Moscow Univ. Comput. Math. Cybernet. 13* (1977), 8–13.
15. Hanke, M. Accelerated Landweber iterations for the solution of ill-posed equations. *Numer. Math. 60*, 3 (1991), 341–373.
16. Kaltenbacher, B. Toward global convergence for strongly nonlinear ill-posed problems via a regularizing multilevel approach. *Numer. Funct. Anal. Optim. 27*, 5-6 (2006), 637–665.
17. Kaltenbacher, B. Convergence rates of a multilevel method for the regularization of nonlinear ill-posed problems. *J. Integral Equations Appl. 20*, 2 (2008), 201–228.
18. Kaltenbacher, B., Neubauer, A., and Scherzer, O. *Iterative regularization methods for nonlinear ill-posed problems*, vol. 6 of *Radon Series on Computational and Applied Mathematics*. Walter de Gruyter GmbH & Co. KG, Berlin, 2008.
19. Mandache, N. Exponential instability in an inverse problem for the Schrödinger equation. *Inverse Problems 17*, 5 (2001), 1435–1444.
20. Neubauer, A., and Scherzer, O. A convergence rate result for a steepest descent method and a minimal error method for the solution of nonlinear ill-posed problems. *Z. Anal. Anwendungen 14*, 2 (1995), 369–377.
21. Scherzer, O. A convergence analysis of a method of steepest descent and a two-step algorithm for nonlinear ill-posed problems. *Numer. Funct. Anal. Optim. 17*, 1-2 (1996), 197–214.
22. Scherzer, O. An iterative multi-level algorithm for solving nonlinear ill-posed problems. *Numer. Math. 80*, 4 (1998), 579–600.
23. Schuster, T., Kaltenbacher, B., Hofmann, B., and Kazimierski, K. S. *Regularization methods in Banach spaces*, vol. 10 of *Radon Series on Computational and Applied Mathematics*. Walter de Gruyter GmbH & Co. KG, Berlin, 2012.
24. Teschke, G., and Borries, C. Accelerated projected steepest descent method for nonlinear inverse problems with sparsity constraints. *Inverse Problems 26*, 2 (2010), 025007, 23.
25. Xu, Z. B., and Roach, G. F. Characteristic inequalities of uniformly convex and uniformly smooth Banach spaces. *J. Math. Anal. Appl. 157*, 1 (1991), 189–210.