

## A CONVERGENCE ANALYSIS OF A MULTI-LEVEL PROJECTED STEEPEST DESCENT ITERATION FOR NONLINEAR INVERSE PROBLEMS IN BANACH SPACES SUBJECT TO STABILITY CONSTRAINTS \*

MAARTEN V. DE HOOP<sup>†</sup>, LINGYUN QIU<sup>‡</sup>, AND OTMAR SCHERZER<sup>§</sup>

**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 closed, 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 closed, 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.

**Key words.** inverse problems, projected steepest descent iteration, stability

**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 closed, 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 closed, 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 the a parameter in the operator which defines the inverse problems, and changes the data, affects for a given closed, convex subset the accuracy of approximation, and the stability constant, as well. Our approach results in a significantly larger radius of convergence. 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

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

with domain  $\mathcal{D}(F) \subset X$ , where  $X$  and  $Y$  are Banach spaces. 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 closed, convex subset  $Z \subset X$  such that

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

Here  $\Delta_p$  denotes the Bregman distance (defined below) and  $p > 1$ . This states conditional Lipschitz stability of the inverse problem. Motivated by [13], we employ a steepest descent iteration, here, to

---

\*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.

<sup>†</sup>Center for Computational and Applied Mathematics, Purdue University, West Lafayette, IN 47907 (mdehoop@purdue.edu).

<sup>‡</sup>Center for Computational and Applied Mathematics, Purdue University, West Lafayette, IN 47907 (qiu@purdue.edu).

<sup>§</sup>Computational Science Center, University of Vienna, Nordbergstr. 15, A-1090 Vienna, Austria (otmar.scherzer@univie.ac.at).

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, [20]. By constraining the pre-image space, however, Lipschitz stability can be obtained; for the case of EIT, see [3, 7] and for the case of inverse boundary value problems for the Helmholtz equation, see [7]. 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 [17].

Our first main result concerns restricted convergence of the projected steepest descent iteration with a certain Hölder or Lipschitz type stability condition on a closed, 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 [21, 22, 19] and projected iteration regularization techniques for the solution of inverse problems with convexity constraints. The latter have been analyzed mostly in the context of *linear* inverse problems (see, for example, [14]) and later as accelerated methods in [12]. The later has been modified to nonlinear problems by [26]. The main differences of our work to the above mentioned papers are the conditions under which we prove convergence. In fact, instead of source and nonlinearity conditions (as in [21, 22]), we assume certain Hölder or Lipschitz stability of the inverse problem. This is a novel view point, which has been raised in [13]. The steepest descent method proposed here is a generalization of the steepest descent method for unconstrained linear problems (see for example [15]). It is however different to the generalization for nonlinear problems proposed in [21, 22], even for unconstrained problems. The steepest descent method proposed here for nonlinear problems, in fact, shares common features with the conjugate gradient (CG) method, in the sense that it takes into account all previous iterations. For a survey on CG methods for solving linear inverse problems in Hilbert spaces we refer to [16], and for recent developments of CG in Banach spaces we refer to [25].

Based on our first main result, we then introduce a multilevel algorithm. We assume that there are closed, 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  $C_\alpha$ , that is,

$$\Delta_p(x, \tilde{x}) \leq C_\alpha^p \|F_\alpha(x) - F_\alpha(\tilde{x})\|^p, \quad \forall x, \tilde{x} \in Z_\alpha.$$

In fact,  $F_\alpha$  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  $C_{\alpha_1} \leq C_{\alpha_2}$  if  $\alpha_1 < \alpha_2$ . In the context of discretization methods,  $Z_\alpha$  stands for a finite-dimensional subspace of  $X$  and the number of basis vectors increases as  $\alpha$  increases, while the projection can be an orthogonal projection on  $Z_\alpha$ . 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. This condition guarantees 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.

**2.1. Duality mappings.** 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  $\|\cdot\|$ . We denote the space of continuous linear operators  $X \rightarrow Y$  by  $\mathcal{L}(X, Y)$ . Let  $F : \mathcal{D}(F) \subset X \rightarrow Y$  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,

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

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 to a neighborhood of 0,  $F$  is called Fréchet 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  $J_p = \partial f_p : X \rightarrow 2^{X^*}$  of the convex functional  $f_p : x \mapsto \frac{1}{p}\|x\|^p$  defined by

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

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  $\delta_X$  of  $X$  by

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

and the smoothness modulus  $\rho_X$  of  $X$  by

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

**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  $\lim_{\tau \rightarrow 0} \frac{\rho_X(\tau)}{\tau} = 0$ ,*
- (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  $\rho_X(\tau) \leq C\tau^q$ .*

For a detailed introduction to the geometry of Banach spaces and the duality mapping, we refer to [11, 24]. We list the properties we need here in the following theorem.

**THEOREM 2.2.** *Let  $p > 1$ . The following statements hold true:*

- (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  $J_p^{-1} : X^* \rightarrow X$  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 [8].

**DEFINITION 2.3.** *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*

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

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 theorem, we summarize some facts concerning the Bregman distance and the relationship between the Bregman distance and the norm [1, 2, 9, 27].

**THEOREM 2.4.** *Let  $X$  be a uniformly smooth and uniformly convex Banach space. Then, for all  $x, \tilde{x} \in X$ , the following holds:*

(a)

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

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

(c)  $\Delta_p$  is continuous in both arguments.

(d) The following statements are equivalent

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

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

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

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

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

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

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

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

The Bregman distance  $\Delta_p$  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 2.7, can be found in [9].

**DEFINITION 2.5.** *Let  $X$  be a uniformly smooth Banach space and  $p > 1$ . Given a closed convex set  $Z \subset X$  and Bregman distance  $\Delta_p$ , which is defined in Definition 2.4, the Bregman projection of a point  $x \in X$  onto  $Z$  is the point*

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

**DEFINITION 2.6.** *Let  $T : X \rightarrow X$  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 2.7. *Let  $X$  be a uniformly smooth Banach space and  $p > 1$  and  $Z \subset X$  be a closed convex subset. The following statements hold:*

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

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

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.  $Y$  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  $t \mapsto t^{p-1}$ ,  $p > 1$ . Further restrictions on  $X$  and  $Y$  will be indicated in the respective theorems below.

**3. Convergence of a projected steepest descent iteration.** Here, we assume conditional stability, that is stability if operator  $F$  is restricted to closed, 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.

ASSUMPTION 3.1. *Let*

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

for some  $\rho > 0$ .  $\rho$  here will come into play as a convergence radius and  $z^\dagger$  is defined below.

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

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

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

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

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

- (c) *Let  $Z$  denote a closed, convex subset of  $X$ . The inversion has the uniform Lipschitz type stability for elements in  $Z$ , i.e., there exists a constant  $C_F > 0$  such that*

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

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

$$(3.4) \quad \text{dist}(y, F(Z)) \leq \eta,$$

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

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

Note, that this condition also accounts for data errors.

The stopping index  $K = K(\eta)$  of the following iteration is determined by a discrepancy principle

$$(3.6) \quad K(\eta) = \min\{k \in \mathbb{N} \mid \|F(x_k) - y\| \leq \hat{\eta}\}$$

with a fixed

$$(3.7) \quad \hat{\eta} > \eta .$$

We introduce the following algorithm:

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

$$(3.8) \quad \boxed{R_k = F(x_k) - y, \quad T_k = DF(x_k)^* j_p(F(x_k) - y), \quad r_k = \|R_k\|, \quad t_k = \|T_k\| .}$$

Moreover, we define

$$(3.9) \quad \boxed{\begin{aligned} C_k &= G_q t_k^q, \quad k = 0, 1, 2, \dots, \\ D_0 &= r_0 - \hat{\eta}, D_k = r_k - \eta - \frac{L}{2} E_k^{2/p}, \quad k = 1, 2, \dots \\ E_k &= \left( \frac{2(\hat{\eta} - \eta)}{L} \right)^{p/2} - \frac{p}{C_p} \sum_{n=0}^{k-1} e_n, \quad k = 1, 2, \dots \\ e_k &= \frac{1}{p} C_k^{-\frac{1}{q-1}} r_k^{p^2-p} D_k^p, \quad k = 0, 1, 2, \dots \\ \mu_k &= C_k^{-\frac{1}{q-1}} r_k^{\frac{p-1}{q-1}} D_k^{\frac{1}{q-1}}, \quad k = 0, 1, 2, \dots \end{aligned}}$$

Now, the main steps of the algorithm:

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

$$(3.10) \quad \|F(x_0) - F(z^\dagger)\| \leq \rho = \left( \frac{C_p}{p} \right)^{1/p} \left( \frac{2(\hat{\eta} - \eta)}{LC_F^2} \right)^{1/2},$$

where  $z^\dagger$  and  $\rho$  are specified in Theorem 3.3 below.

(S1) Compute the new iterate via

$$(3.11) \quad \begin{aligned} \tilde{x}_{k+1} &= J_q^*(J_p(x_k) - \mu_k T_k) \\ x_{k+1} &= \mathcal{P}_Z(\tilde{x}_{k+1}). \end{aligned}$$

Set  $k \leftarrow k + 1$  and repeat step (S1).

THEOREM 3.3. *Assume that Assumption 3.1 holds. Moreover, assume that the estimate (3.4) holds for some positive constant  $\eta$  and  $z^\dagger \in Z$ .*

*Then Algorithm 3.2 stops after a finite number  $K = K(\eta)$  of iterations with the discrepancy criterion*

$$r_K = \|F(x_K) - y\| \leq \hat{\eta},$$

being satisfied.

$$(3.12) \quad \Delta_p(x_{k+1}, z^\dagger) - \Delta_p(x_k, z^\dagger) \leq -e_k < 0,$$

holds for all  $k \leq K(\eta) - 1$ .

*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. We have

$$(3.13) \quad \|x_k - z^\dagger\| \underbrace{\leq}_{(2.6)} \left( \Delta_p(x_k, z^\dagger) \frac{p}{C_p} \right)^{1/p} \underbrace{\leq}_{(3.3)} C_F \|F(x_k) - F(z^\dagger)\| \left( \frac{p}{C_p} \right)^{1/p} .$$

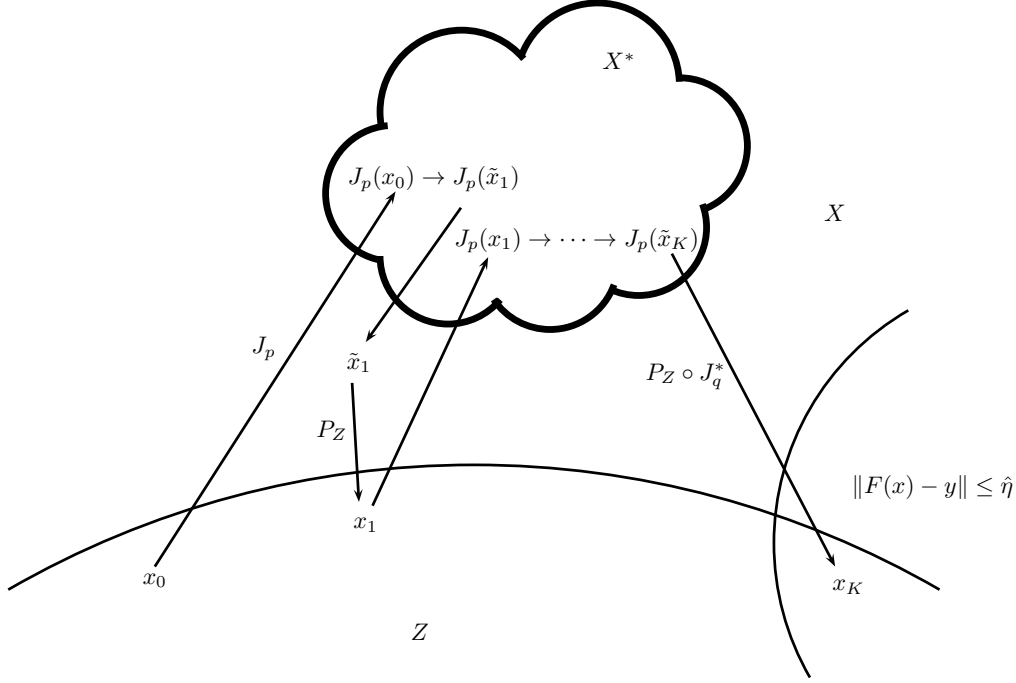


FIG. 1. Projected steepest descent iteration

Using this and (3.10) it follows that

$$(3.14) \quad \|x_0 - z^\dagger\|^2 \leq C_F^2 \|F(x_0) - F(z^\dagger)\|^2 \left(\frac{p}{C_p}\right)^{2/p} \leq \frac{2(\hat{\eta} - \eta)}{L}.$$

Moreover, from (3.3) and (3.10) we obtain

$$(3.15) \quad \Delta_p(x_0, z^\dagger) \leq \frac{C_p}{p} \left(\frac{2(\hat{\eta} - \eta)}{L}\right)^{p/2}.$$

Using the definition of  $r_k$  it follows that for  $k = 0, 1, \dots$

$$(3.16) \quad \mu_k r_k^{p-1} = G_q^{-1/(q-1)} t_k^{-p} r_k^{p^2-p} D_k^{1/(q-1)}, \quad \frac{G_q}{q} \mu_k^q t_k^q = \frac{G_q^{-1/(q-1)}}{q} t_k^{-p} r_k^{p^2-p} D_k^p.$$

Now, we start with the main body of the proof: Using (2.5) and (2.1) we obtain, for the sequence of residues,

$$(3.17) \quad \begin{aligned} & \Delta_p(\tilde{x}_{k+1}, z^\dagger) \\ &= \Delta_p(x_k, z^\dagger) + \frac{1}{q} (\|\tilde{x}_{k+1}\|^p - \|x_k\|^p) - \langle J_p(\tilde{x}_{k+1}) - J_p(x_k), z^\dagger \rangle \\ &= \Delta_p(x_k, z^\dagger) + \frac{1}{q} (\|J_p(\tilde{x}_{k+1})\|^q - \|J_p(x_k)\|^q) - \langle J_p(\tilde{x}_{k+1}) - J_p(x_k), z^\dagger \rangle. \end{aligned}$$

Applying (2.5) and (f) of Theorem 2.4 with  $x^* = J_p(\tilde{x}_{k+1})$  and  $\tilde{x}^* = J_p(x_k)$ , we get

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

Substituting (3.11) and using this inequality in (3.17) yields

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

We estimate the second term in (3.18). Using (2.6) and the Lipschitz type stability (3.3), and (3.4), we find that

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

From (3.18), (3.19) and the definition of the steepest descent algorithm (3.11) it follows that, for  $k = 0, 1, 2, \dots$ ,

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

We prove by induction that

$$\Delta_p(x_{k+1}, z^\dagger) - \Delta_p(x_k, z^\dagger) \leq \Delta_p(\tilde{x}_{k+1}, z^\dagger) - \Delta_p(x_k, z^\dagger) \leq -e_k, \quad \forall k \geq 0.$$

For  $k = 0$ , we find that

$$\begin{aligned}
(3.21) \quad & \Delta_p(x_1, z^\dagger) - \Delta_p(x_0, z^\dagger) \\
&\stackrel{(2.9)}{\leq} \underbrace{\Delta_p(\tilde{x}_1, z^\dagger) - \Delta_p(x_0, z^\dagger)}_{(2.9)} \\
&\stackrel{(3.20)}{\leq} \underbrace{\frac{G_q}{q} \mu_0^q t_0^q - \mu_0 r_0^p + \mu_0 r_0^{p-1} \eta + \frac{L}{2} \mu_0 r_0^{p-1} \|x_0 - z^\dagger\|^2}_{(3.20)} \\
&\stackrel{(3.14)}{\leq} \underbrace{\frac{G_q}{q} \mu_0^q t_0^q + \mu_0 r_0^{p-1} (\hat{\eta} - r_0)}_{(3.14)} \\
&\stackrel{(3.9)}{\leq} \underbrace{-\frac{1}{p} C_0^{-\frac{1}{q-1}} r_0^{p^2-p} D_0^p}_{(3.9)} \\
&\stackrel{(3.9)}{\leq} \underbrace{-e_0}_{(3.9)}.
\end{aligned}$$



For  $k = 1, 2, \dots$ , we first note that

$$\begin{aligned}
 (3.22) \quad \|x_k - z^\dagger\|^2 &\stackrel{(2.6)}{\leq} \left( \frac{p}{C_p} \Delta_p(x_k, z^\dagger) \right)^{2/p} \\
 &= \left( \frac{p}{C_p} \left( \sum_{n=1}^k [\Delta_p(x_n, z^\dagger) - \Delta_p(x_{n-1}, z^\dagger)] + \Delta_p(x_0, z^\dagger) \right) \right)^{2/p} \\
 &\stackrel{\text{induction assumption}}{\leq} \left( \frac{p}{C_p} \left( \Delta_p(x_0, z^\dagger) - \sum_{n=0}^{k-1} e_n \right) \right)^{2/p}.
 \end{aligned}$$

Then we derive the following chain of inequalities

$$\begin{aligned}
 &\Delta_p(\tilde{x}_{k+1}, z^\dagger) - \Delta_p(x_k, z^\dagger) \\
 &\stackrel{(2.9)}{\leq} \underbrace{\Delta_p(\tilde{x}_{k+1}, z^\dagger) - \Delta_p(x_k, z^\dagger)} \\
 &\stackrel{(3.20), (3.22)}{\leq} \underbrace{\frac{G_q}{q} \mu_k^q t_k^q - \mu_k r_k^p + \mu_k \eta r_k^{p-1} + \frac{L}{2} \mu_k r_k^{p-1} \left( \frac{p}{C_p} \left( \Delta_p(x_0, z^\dagger) - \sum_{n=0}^{k-1} e_n \right) \right)^{2/p}}.
 \end{aligned}$$

Then we have that

$$\begin{aligned}
 &\Delta_p(\tilde{x}_{k+1}, z^\dagger) - \Delta_p(x_k, z^\dagger) \\
 &\stackrel{(3.10), (3.3)}{\leq} \underbrace{\frac{G_q}{q} \mu_k^q t_k^q - \mu_k r_k^p + \mu_k \eta r_k^{p-1} + \frac{L}{2} \mu_k r_k^{p-1} \left( \left( \frac{2}{L} (\hat{\eta} - \eta) \right)^{p/2} - \frac{p}{C_p} \sum_{n=0}^{k-1} e_n \right)^{2/p}} \\
 &\stackrel{(3.9)}{=} \underbrace{\frac{G_q}{q} \mu_k^q t_k^q - \mu_k r_k^{p-1} \left( r_k - \eta - \frac{L}{2} E_k^{2/p} \right)} \\
 &\stackrel{(3.9)}{=} \underbrace{\frac{G_q}{q} \mu_k^q t_k^q - \mu_k r_k^{p-1} D_k} \\
 &\stackrel{(3.16)}{=} \underbrace{-\frac{1}{p} C_k^{-\frac{1}{q-1}} r_k^{p^2-p} D_k^p} \\
 &\stackrel{(3.9)}{=} \underbrace{-e_k}.
 \end{aligned}$$

This completes the induction.

It is left to show that Algorithm 3.2 stops after a finite number of iterations (i.e.  $K(\eta)$  iterations). We prove this by contradiction. Suppose that Algorithm 3.2 does not stop within a finite number of iterations and, hence,

$$(3.23) \quad r_k > \hat{\eta}, \quad \forall k \geq 0.$$

Then, from the monotonicity of the Bregman distances (3.12), we have that

$$0 \leq \Delta_p(x_k, z^\dagger) \leq \Delta_p(x_0, z^\dagger) - \sum_{n=0}^{k-1} e_n, \quad \forall k > 0.$$

It follows that

$$\sum_{n=0}^{\infty} e_n \leq \Delta_p(x_0, z^\dagger),$$

and  $e_k$  converges to 0 as  $k$  goes to infinity. Using (3.23) and the definitions of  $e_k$  and  $E_k$  in (3.9)

$$\begin{aligned} 0 &= \lim_{k \rightarrow \infty} e_k \\ &\stackrel{(3.9)}{=} \lim_{k \rightarrow \infty} \left( r_k - \eta - \frac{L}{2} E_k^{2/p} \right) \\ &= \lim_{k \rightarrow 0} \left( r_k - \eta - \frac{L}{2} \left( \left( \frac{2(\hat{\eta} - \eta)}{L} \right)^{p/2} - \frac{p}{C_p} \sum_{n=0}^{k-1} e_n \right)^{2/p} \right) \\ &\geq \lim_{k \rightarrow 0} \left( \hat{\eta} - \eta - \frac{L}{2} \left( \left( \frac{2(\hat{\eta} - \eta)}{L} \right)^{p/2} - \frac{p}{C_p} \sum_{n=0}^{k-1} e_n \right)^{2/p} \right) \\ &> \hat{\eta} - \eta - \frac{L}{2} \left( \frac{2(\hat{\eta} - \eta)}{L} \right) \\ &= 0, \end{aligned}$$

which is a contradiction.  $\square$

**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  $L$  and  $\eta$  can be chosen to be equal to zero. Then we have

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

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

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

we get

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

which is the standard parameter choice of the steepest descent method [15].

In the nonlinear setting however, via the  $E_k$  (3.9), the iterate  $x_{k+1}$  depends explicitly on all previous iterates. The nonlinear method shares the common aspect with the CG method for solving linear problems that it accounts for previous iterates.

**4. Extension to a multi-level algorithm.** In this section, we consider a set,  $\{Z_\alpha\}_{\alpha \in \mathbb{R}}$ , of closed and convex subsets of  $X$ , and an operator family  $\{F_\alpha\}_{\alpha \in \mathbb{R}}$ , where  $F_\alpha$  is obtained as  $F_\alpha = F|_{Z_\alpha}$ . We let

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

for some  $\rho > 0$  and invoke

ASSUMPTION 4.1.

(a)  $F$  is weakly sequentially closed, that is,

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

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

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

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

(c) The inversion has the uniform Lipschitz type stability for elements in  $Z_\alpha$ , that is, there exists a constant  $C_\alpha > 0$  such that

$$(4.3) \quad \Delta_p(x, \tilde{x}) \leq C_\alpha^p \|F_\alpha(x) - F_\alpha(\tilde{x})\|^p \quad \forall x, \tilde{x} \in \mathcal{B} \cap Z_\alpha.$$

For the stability constants,  $\{C_\alpha\}$ , we introduce

ASSUMPTION 4.2.

(a) Let  $\eta_\alpha = \eta_\alpha(y)$  be defined by

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

$\eta_\alpha$  is continuous with respect to  $\alpha$  for every fixed  $y \in Y$ .

(b) If  $Z_{\alpha_1} \subset Z_{\alpha_2}$  then  $C_{\alpha_1} \leq C_{\alpha_2}$ .

(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  $Z_\alpha$  are finite-dimensional and the stability constant 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 [23, 17, 18], 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  $\alpha$  of the operator family and design a progressive iteration method with the aid of the result from previous section.

REMARK 4.3. *As an example, we mention inverse boundary value problems for the Helmholtz equation. Indeed, stability estimates satisfying Assumption 4.2 have been obtained [7] where  $Z_\alpha$  represents a space spanned by a finite linear combination of piecewise constant functions. Using our multi-level algorithm, we arrive at a convergence result by successive approximation. This result can be further improved, using the same algorithm, by combining different frequencies and exploiting the frequency dependence of the stability constants. The idea of using multiple frequencies was proposed by Chen [10], who introduced an algorithm based on recursive linearization. The algorithm starts with an initial guess at the lowest frequency, which typically captures the coarse scale variations in the wavepeed. Then the Born approximation is invoked [10, 4, 5, 6]. The iteration is based on a linearization of the inverse problem at the present frequency. By progressively increasing the frequency and carrying out the iterations, increasingly finer details are added to the wavespeed model until a sufficiently accurate result is obtained. In [6], the convergence of this algorithm was established under certain conditions. As a direct application of Theorem 4.5, the convergence of this algorithm can be revisited. Especially, (4.6) offers a strategy for picking the frequencies and regularization parameters.*

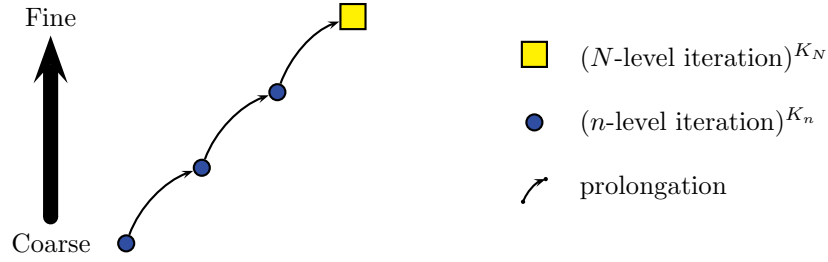


FIG. 2. A illustration of Algorithm 4.4

ALGORITHM 4.4.

(S0) Use  $x_{0,0}$  as the starting point. Set  $n = 0$ .(S1) Iteration. Use  $F_n$  and  $Z_n$  as the modelling operator and convex subset to run Algorithm 3.2 with the discrepancy criterion given by

$$(4.4) \quad K_n = \min\{k \in \mathbb{N} \mid \|F_n(x_{n,k}) - y\| \leq (1 + \varepsilon)\eta_n\},$$

where  $\varepsilon > 0$  is the uniform tolerance constant.STOP, if  $n = N$ , a given number.(S2) Set  $x_{n+1,0} = x_{n,K_n}$ ,  $n = n + 1$  and go to step (S1).

This algorithm is illustrated in Figure 2.

THEOREM 4.5. Assume that Assumptions 4.1 and 4.2 hold. Assume that there exists a finite subset of operators,  $\{F_n\}_{n=1}^N$  say, from the operator family  $\{F_\alpha\}$  such that(i) The starting point  $x_{0,0}$  is within the first convergence radius, that is,

$$(4.5) \quad \|F_0(x_{0,0}) - F(z_0^\dagger)\| \leq \rho_0,$$

where  $z_0^\dagger$  and  $\rho_0$  are defined in Theorem 3.3;

(ii)

$$(4.6) \quad (1 + \varepsilon)\eta_n \leq \left(\frac{C_p}{p}\right)^{1/p} \left(\frac{2}{L_{n+1}C_{n+1}^2}\right)^{1/2} \varepsilon^{1/2}\eta_{n+1}^{1/2} - \eta_{n+1}, \quad n = 1, \dots, N - 1.$$

(iii)  $N$  is the first positive integer such that  $\eta_N \leq (1 + \varepsilon)^{-1}\hat{\eta}$ , that is,

$$(1 + \varepsilon)\eta_n > \hat{\eta} \quad \forall n < N$$

and

$$(1 + \varepsilon)\eta_N \leq \hat{\eta}.$$

Then, Algorithm 4.4 has the property that the number of iterations is finite with the discrepancy criterion

$$(4.7) \quad \|F_N(x_{N,K_N}) - y\| \leq \hat{\eta}$$

satisfied.

The strategy of the proof is to estimate the decreasing objective function  $\|F(x_{n,K_n}) - y\|$  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  $n$ . Then, with (4.6) and (4.4), we show that the initial point  $x_{n+1,0}$  on level  $n + 1$ , which coincides with the iteration result  $x_{n,K_n}$  on level  $n$ , is within the convergence radius  $\rho_{n+1}$ . Therefore, the procedure continues until (4.7) is satisfied.

*Proof.* We adapt the convergence radius,  $\rho$ , in Theorem 3.3 to an  $n$ -level convergence radius  $\rho_n$  according to

$$(4.8) \quad \|F_n(x_{n,0}) - F_n(z_n^\dagger)\| \leq \rho_n,$$

where  $z_n^\dagger$  denotes the best  $Z_n$ -approximation as defined in Theorem 3.3. The same argument with discrepancy criterion (3.6) replaced by (4.4) leads to

$$(4.9) \quad \rho_n = \left(\frac{C_p}{p}\right)^{1/p} \left(\frac{2\varepsilon\eta_n}{L_n C_n^2}\right)^{1/2}.$$

Next, we prove that

$$\|F_n(x_{n,0}) - F_n(z_n^\dagger)\| \leq \rho_n \Rightarrow \|F_{n+1}(x_{n+1,0}) - F_{n+1}(z_{n+1}^\dagger)\| \leq \rho_{n+1}, \quad \forall n < N.$$

Indeed, for  $n < N$ , according to (4.8) and Theorem 3.3, after  $K_n$  steps, the  $n$ -level discrepancy criterion,

$$\|F(x_{n,K_n}) - y\| \leq (1 + \varepsilon)\eta_n,$$

is satisfied. Then, with (4.6), we estimate

$$\begin{aligned} & \|F_{n+1}(x_{n+1,0}) - F_{n+1}(z_{n+1}^\dagger)\| \\ &= \|F_n(x_{n,K_n}) - F_{n+1}(z_{n+1}^\dagger)\| \\ &\leq \|F_n(x_{n,K_n}) - y\| + \|F_{n+1}(z_{n+1}^\dagger) - y\| \\ &\leq (1 + \varepsilon)\eta_n + \eta_{n+1} \\ &\leq \left(\frac{C_p}{p}\right)^{1/p} \left(\frac{2\varepsilon\eta_{n+1}}{L_n C_n^2}\right)^{1/2} \\ &= \rho_{n+1}. \end{aligned}$$

For the final level  $N$ , we apply Theorem 3.3 again to find that

$$\|F_N(x_{N,K_N}) - y\| \leq (1 + \varepsilon)\eta_N \leq \hat{\eta}.$$

□

Theorem 4.5, especially (ii), indicates that a sufficient condition for the existence of such a selection of operators is that the tolerated best- $Z_n$ -approximation is within the convergence radius of  $Z_{n+1}$ . In fact, this condition comes from a bootstrap type competition between  $\eta_n$  and  $\rho_n$ .

We give an example of how conditions (ii) and (iii) in Theorem 4.5 can be satisfied.

**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, the following conditions hold:

- (a) Given starting point  $x_{0,0}$  is within the first convergence radius  $\rho_0$ ,
- (b)  $\eta_\alpha = C_1(\alpha + 1)^{-1}$  for some constant  $C_1 > 2\hat{\eta}$ ,
- (c) the stability constant  $C_\alpha = \alpha + 1$ ,

$$(d) L_\alpha = \frac{2}{25C_1} \left( \frac{C_p}{p} \right)^{2/p} (\alpha + 1)^{-1}.$$

Now the operators  $\{F_n\}_{n=0}^N$  defined by  $F_n = F|_{Z_n}$ . If we run Algorithm 4.4 with this set and the uniform tolerance constant  $\varepsilon = 1$ , where  $N$  is the first integer such that  $2\eta_N \leq \hat{\eta}$  is satisfied, then, we have that

$$\|F_n(x_{n,K_n}) - y\| \leq 2\eta_n, \quad n = 1, 2, \dots, N - 1,$$

and

$$\|F_N(x_{N,K_N}) - y\| \leq \hat{\eta}.$$

**Acknowledgments.** The research was initiated at the Isaac Newton Institute for Mathematical Sciences (Cambridge, England) during a programme on Inverse Problems in Fall 2011.

#### REFERENCES

- [1] YAKOV I. ALBER AND DAN BUTNARIU, *Convergence of Bregman projection methods for solving consistent convex feasibility problems in reflexive Banach spaces*, J. Optim. Theory Appl., 92 (1997), pp. 33–61.
- [2] YAKOV I. ALBER, ATHANASSIOS G. KARTSATOS, AND ELENA LITSYN, *Iterative solution of unstable variational inequalities on approximately given sets*, Abstr. Appl. Anal., 1 (1996), pp. 45–64.
- [3] GIOVANNI ALESSANDRINI AND SERGIO VESSELLA, *Lipschitz stability for the inverse conductivity problem*, Adv. in Appl. Math., 35 (2005), pp. 207–241.
- [4] GANG BAO AND PEIJUN LI, *Inverse medium scattering for the Helmholtz equation at fixed frequency*, Inverse Problems, 21 (2005), pp. 1621–1641.
- [5] ———, *Numerical solution of an inverse medium scattering problem for Maxwell’s equations at fixed frequency*, J. Comput. Phys., 228 (2009), pp. 4638–4648.
- [6] GANG BAO AND FAOUZI TRIKI, *Error estimates for the recursive linearization of inverse medium problems*, J. Comput. Math., 28 (2010), pp. 725–744.
- [7] ELENA BERETTA, MAARTEN DE HOOP, AND LINGYUN QIU, *Lipschitz stability of an inverse boundary value problem for a Schrödinger type equation*, ArXiv e-prints, (2012).
- [8] LEV M. BREGMAN, *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 (1967), pp. 200 – 217.
- [9] DAN BUTNARIU, ALFREDO N. IUSEM, AND ELENA RESMERITA, *Total convexity for powers of the norm in uniformly convex Banach spaces*, J. Convex Anal., 7 (2000), pp. 319–334.
- [10] YU CHEN, *Inverse scattering via Heisenberg’s uncertainty principle*, Inverse Problems, 13 (1997), pp. 253–282.
- [11] IOANA CIORANESCU, *Geometry of Banach spaces, duality mappings and nonlinear problems*, vol. 62 of Mathematics and its Applications, Kluwer Academic Publishers Group, Dordrecht, 1990.
- [12] INGRID DAUBECHIES, MASSIMO FORNASIER, AND IGNACE LORIS, *Accelerated projected gradient method for linear inverse problems with sparsity constraints*, J. Fourier Anal. Appl., 14 (2008), pp. 764–792.
- [13] MAARTEN V DE HOOP, LINGYUN QIU, AND OTMAR SCHERZER, *Local analysis of inverse problems: Hölder stability and iterative reconstruction*, Inverse Problems, 28 (2012), p. 045001.
- [14] BERTOLT EICKE, *Iteration methods for convexly constrained ill-posed problems in Hilbert space*, Numer. Funct. Anal. Optim., 13 (1992), pp. 413–429.
- [15] SERGEI F. GILYAZOV, *Iterative solution methods for inconsistent linear equations with nonself-adjoint operator*, Moscow Univ. Comput. Math. Cybernet., 13 (1977), pp. 8–13.
- [16] MARTIN HANKE, *Conjugate gradient type methods for ill-posed problems*, vol. 327 of Pitman Research Notes in Mathematics Series, Longman Scientific & Technical, Harlow, 1995.
- [17] BARBARA KALTENBACHER, *Toward global convergence for strongly nonlinear ill-posed problems via a regularizing multilevel approach*, Numer. Funct. Anal. Optim., 27 (2006), pp. 637–665.
- [18] ———, *Convergence rates of a multilevel method for the regularization of nonlinear ill-posed problems*, J. Integral Equations Appl., 20 (2008), pp. 201–228.
- [19] BARBARA KALTENBACHER, ANDREAS NEUBAUER, AND OTMAR SCHERZER, *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.
- [20] NICULAE MANDACHE, *Exponential instability in an inverse problem for the Schrödinger equation*, Inverse Problems, 17 (2001), pp. 1435–1444.
- [21] A. NEUBAUER AND O. SCHERZER, *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 (1995), pp. 369–377.

- [22] OTMAR SCHERZER, *A convergence analysis of a method of steepest descent and a two-step algorithm for nonlinear ill-posed problems*, Numer. Funct. Anal. Optim., 17 (1996), pp. 197–214.
- [23] ———, *An iterative multi-level algorithm for solving nonlinear ill-posed problems*, Numer. Math., 80 (1998), pp. 579–600.
- [24] F. SCHÖPFER, A. K. LOUIS, AND T. SCHUSTER, *Nonlinear iterative methods for linear ill-posed problems in Banach spaces*, Inverse Problems, 22 (2006), pp. 311–329.
- [25] F. SCHÖPFER AND T. SCHUSTER, *Fast regularizing sequential subspace optimization in Banach spaces*, Inverse Problems, 25 (2009), pp. 015013, 22.
- [26] GERD TESCHKE AND CLAUDIA BORRIES, *Accelerated projected steepest descent method for nonlinear inverse problems with sparsity constraints*, Inverse Problems, 26 (2010), pp. 025007, 23.
- [27] ZONG BEN XU AND GARY F. ROACH, *Characteristic inequalities of uniformly convex and uniformly smooth Banach spaces*, J. Math. Anal. Appl., 157 (1991), pp. 189–210.