Convergence rates for the iteratively regularized Gauss–Newton method in Banach spaces

This article has been downloaded from IOPscience. Please scroll down to see the full text article.

2010 Inverse Problems 26 035007


The Table of Contents and more related content is available

Download details:
IP Address: 134.109.41.19
The article was downloaded on 10/03/2010 at 12:03

Please note that terms and conditions apply.
Convergence rates for the iteratively regularized Gauss–Newton method in Banach spaces

Barbara Kaltenbacher$^1$ and Bernd Hofmann$^2$

$^1$ Institute of Mathematics and Scientific Computing, University of Graz, Heinrichstraße 36, 8010 Graz, Austria
$^2$ Department of Mathematics, Chemnitz University of Technology, 09107 Chemnitz, Germany

E-mail: barbara.kaltenbacher@uni-graz.at and hofmannb@mathematik.tu-chemnitz.de

Received 3 October 2009, in final form 15 January 2010
Published 19 February 2010
Online at stacks.iop.org/IP/26/035007

Abstract
In this paper we consider the iteratively regularized Gauss–Newton method (IRGNM) in a Banach space setting and prove optimal convergence rates under approximate source conditions. These are related to the classical concept of source conditions that is available only in Hilbert space. We provide results in the framework of general index functions, which include, e.g. Hölder and logarithmic rates. Concerning the regularization parameters in each Newton step as well as the stopping index, we provide both a priori and a posteriori strategies, the latter being based on the discrepancy principle.

1. Introduction

We are going to consider a nonlinear ill-posed operator equation

$$F(x) = y$$  \hspace{1cm} (1)

where the possibly nonlinear operator $F : \mathcal{D}(F) \subseteq X \to Y$ with domain $\mathcal{D}(F)$ maps between real Banach spaces $X$ and $Y$. For simplicity, let the symbol $\| \cdot \|$ designate the norm for both spaces. Specifically, we assume $X$ to be reflexive and uniformly smooth. For some of our results we will assume that $X$ is $q$-convex with some $q > 1$.

Since we are interested in the ill-posed situation, i.e. $F$ fails to be continuously invertible, and the data are contaminated with noise, regularization has to be applied (see, e.g., [4, 25], and references therein).

* Research has been partly conducted during the Mini Special Semester on Inverse Problems, 18 May–15 July 2009, organized by RICAM (Austrian Academy of Sciences), Linz, Austria, and moreover supported by Deutsche Forschungsgemeinschaft (DFG) under Grant HO1454/7-2 as well as within the Cluster of Excellence SimTech, University of Stuttgart.
Throughout this paper we will assume that an exact solution $x^1 \in D(F)$ of (1) exists, i.e. $F(x^1) = y$, and that the (deterministic) noise level $\delta$ in an upper estimate

$$\|y - y^\delta\| \leq \delta$$

(2)
of the difference between exact right-hand side $y$ and noisy data $y^\delta$ is known.

Tikhonov-type variational regularization in Banach spaces has been studied recently with error estimates measured by Bregman distances, e.g. in [3] for linear ill-posed problems, and in [9, 12, 13, 19–21, 23] for nonlinear ill-posed problems (1).

Iterative regularization approaches in Hilbert spaces pose an attractive alternative to variational regularization methods. These approaches were comprehensively analyzed in the monographs [1, 17] (see also the references therein). So far, to the authors’ best knowledge, iterative solvers for nonlinear ill-posed problems in Banach spaces have only been formulated in [1, section 4.3] and [18]. In [1], the case $X = Y$ was considered and convergence including rates under sufficiently strong source conditions was proven for generalized Gauss–Newton methods. On the other hand, in [18] convergence of the iteratively regularized Gauss–Newton method and the nonlinear Landweber iteration has been proven in the general situation of possibly different Banach spaces $X$ and $Y$ without imposing any source condition. For an analysis of Landweber-type methods in Banach space we refer to [10] and [24].

The aim of this paper is to provide rate results for the iteratively regularized Gauss–Newton method in a complementary situation, i.e. under weaker source conditions than those assumed in [1], and for not necessarily equal preimage and image space. The obtained rates will be called optimal referring to corresponding optimal rate results in Hilbert space settings.

For Hilbert spaces $X$ by spectral theory one can define at a point $x^1$, where $F$ is Gâteaux differentiable with derivative $F'(x^1)$, linear operators $f(F'(x^1)^*F'(x^1)) : X \to X$ for any index function $f$. We call a function $f : (0, \infty) \to (0, \infty)$ (or its restriction to a right neighborhood of zero) the index function if $f$ is continuous and strictly increasing with $\lim_{t \to 0^+} f(t) = 0$. The properties of non-negativity and self-adjointness of the operator $F'(x^1)^*F'(x^1) : X \to X$ carry over to the new operators. This allows expressing the smoothness of the solution $x^1$ to (1) with respect to the linearization $F'(x^1)$ of the forward operator $F$ in that point. Depending on the specific character of such occurring smoothness Hölder source conditions and general source conditions (see below (8) and (12), respectively) leads to corresponding convergence rates for various regularization methods. For Banach spaces, however, we have $F'(x^1)^* : Y^* \to X^*$ and hence $f(F'(x^1)^*F'(x^1))$ is not well defined. Since general source conditions measuring the solution’s smoothness are not available, additional ideas and concepts have to be exploited. Originally developed in [11] for linear ill-posed problems, the concept of approximate source conditions can help to bridge this gap also in the nonlinear case (see, e.g., [9]). In this context, the degree of violation of a benchmark source condition is expressed by so-called distance functions $d(R)$.

The iteratively regularized Gauss–Newton method can be generalized to a Banach space setting by calculating iterates $x_{k+1}^\delta = x_k^\delta(\alpha_k)$ in a variational form as

$$x_{k+1}^\delta(\alpha_k) \in \arg\min_{x \in D(F)} \|T_k(x - x_k^\delta) + g_k\|_p^p + \alpha \|x - x_0\|_p^p,$$  \hspace{0.5cm} k = 0, 1, \ldots , \hspace{0.5cm} (3)

where $p, r \in (1, \infty)$, $(\alpha_k)_{k \in \mathbb{N}}$ is a sequence of regularization parameters, $x_0$ is some a priori guess and we abbreviate $T_k = F'(x_k^\delta)$. $g_k = F(x_k^\delta) - y^\delta$.

Under the assumptions on $X$ the functional $x \mapsto \frac{1}{p}\|x\|_p^p$ is strictly convex and Fréchet-differentiable for all $p > 1$. Hence, the subdifferential $J_p(x) := \partial \left(\frac{1}{p}\|x\|_p^p\right)$ is single valued and the corresponding duality mapping $J_p$ with the gauge function $t \mapsto t^{p-1}$ is continuous and
bijective from $X$ to its dual space $X^*$. This in general nonlinear mapping $J_p$ is characterized by
$$x^* \in J_p(x) \iff \langle x^*, x \rangle = \|x\|^p \quad \text{and} \quad \|x^*\| = \|x\|^{p-1},$$
where $\langle x^*, x \rangle$ with $x \in X$ and $x^* \in X^*$ is the dual pairing of $X$ and $X^*$. To analyze convergence rates we employ the Bregman distance $\Delta_p(x, y)$ between $x, y \in X$ defined as
$$\Delta_p(x, y) = \frac{1}{p} \|x\|^p - \frac{1}{p} \|y\|^p - \langle J_p(x), x - y \rangle.$$
If $X$ is $q$-convex, then there is a constant $c > 0$ depending on $q$ such that
$$\Delta_p(x, y) \geq c \|x - y\|^q \quad \text{for all} \quad x, y \in X$$
(see, e.g., [2, lemma 2.7]).

2. Approximate source conditions and variational inequalities

In order to overcome the absence of Hölder and general source conditions we first extend the Hilbert space standard source condition [4, p 277, formula (11.2)] to the Banach space setting as
$$\exists w \in X^* : \quad J_p(x - x_0) = F'(x_0)^* w.$$ \hspace{1cm} (5)
Under condition (5) we can estimate
$$\|J_p(x - x_0), x - x_0\| = |\langle w, F'(x_0)^*(x - x_0) \rangle| \leq \|w\| \|F'(x_0)(x - x_0)\|,$$
which implies the variational inequality
$$\exists \beta > 0 \ \forall x \in D(F) : \quad |\langle J_p(x - x_0), x - x_0 \rangle| \leq \beta \|F'(x_0)(x - x_0)\|,$$ \hspace{1cm} (6)
where in contrast to the ideas of [12] we only use $\|F'(x_0)(x - x_0)\|$ instead of $\|F(x) - F(x_0)\|$ on the right-hand side.

Usually (see [12] and [13]) variational inequalities for proving convergence rates for the Tikhonov-type regularization in Banach spaces have to hold for appropriate $x \in D(F)$ in an additive form
$$\exists \beta_1, \beta_2 > 0 : \quad |\langle J_p(x - x_0), x - x_0 \rangle| \leq \beta_1 \Delta_p(x, x_0) + \beta_2 \|F(x) - F(x_0)\|,$$
rather than in the product form (6). Note, however, that the additive form under the assumption
$$\exists K > 0 \ \forall x \in D(F) : \quad \|F(x) - F(x_0)\| \leq K \Delta_p(x, x_0)$$ \hspace{1cm} (7)
immediately follows from the product form by the triangle inequality.

By avoiding $\|F(x) - F(x_0)\|$ on the right-hand side we are up to some extent independent of the tangential cone condition (7). In particular, we will, e.g., prove optimal rates under a mere Lipschitz condition on $F'$ provided (5) holds.

Moreover, for the Banach space setting the form (6) allows us to use as a substitute for the Hölder-type Hilbert space source condition
$$\exists w \in X : \quad J_2(x - x_0) = F'(x_0)^* F'(x_0)^{1/2} w,$$ \hspace{1cm} (8)
for $0 < \nu < 1$, the following variational inequality:
$$\exists \beta > 0 \ \forall x \in B : \quad |\langle J_p(x - x_0), x - x_0 \rangle| \leq \beta D_p^{\nu}(x_0, x)^{\frac{\nu}{1-\nu}} \|F'(x_0)(x - x_0)\|^\nu.$$ \hspace{1cm} (9)
Here
$$B = D(F) \cup B_p(x_0)$$
with \( B_\rho(x_0) \) being a closed ball with radius \( \rho > 0 \) around \( x_0 \), and we use the notation
\[
D_p^{\rho}(\delta, x) := \Delta_p(\delta - x_0, x - x_0).
\]

Precisely, the intermediate source condition \( (9) \) can be motivated from the Hilbert space case, since the usual source condition \( (8) \) implies \( (9) \), cf. \( \cite{9} \). This includes, e.g., logarithmic source conditions as appropriate for exponentially ill-posed problems, cf. \( \cite{14} \).

Now we will show that variational inequalities like \( (9) \) and \( (11) \) can also be concluded from the Hilbert space case. Namely, if \( (10) \) holds, by Jensen’s inequality, the general Hilbert space source condition
\[
\forall x \in D(F), \, x \neq x^*: \quad |(J_p(x^i - x_0), x - x^i)| \leq D_p^{\rho}(x^i, x) \frac{\|f(F(x^i))Q(x^i - x_0)\|^2}{D_p^{\rho}(x^i, x)}
\]

(11)

implies
\[
|(J_p(x^i - x_0), x - x^i)| \leq \|w\| \|x - x^i\|^\nu \|f(F(x^i))Q(x^i - x_0)\|^\nu / \|x - x^i\|^\nu.
\]

(12)

This includes, e.g., logarithmic source conditions as appropriate for exponentially ill-posed problems, cf. \( \cite{14} \).

The distance function is well defined as a non-negative and non-increasing continuous function for all \( R \geq 0 \). Since by Alaoglu’s theorem the unit ball in \( Y^\ast \) is weak* compact and the dual norm function is weak* lower semicontinuous, the infimum in \( (14) \) is a minimum and assumed in some \( w_R \in Y^\ast \). Under the condition
\[
J_p(x^i - x_0) \in \mathcal{R}(f(F(x^i))Q) \quad \forall x^i \in K
\]

(15)

it is evident that \( d(R) \) is strictly positive for all \( R \geq 0 \) and tends to zero as \( R \to \infty \), cf \( \cite{9} \), lemma 4.1 and remark 4.2. In such a case the decay rate of the distance function
$d(R)$ to zero as $R \to \infty$ measures the degree of violation of $J_p(x^1 - x_0)$ with respect to the benchmark source condition (5). As the following proposition will show, this degree of violation determines the function $f$ in variational inequalities like (11).

**Proposition 1.** Let $X$ be $q$-convex. Under conditions (4) and (15) let $\overline{d}$ be a continuous and strictly decreasing majorant of the distance function $d$ from (14) in the sense that the inequality $0 < d(R) \leq \overline{d}(R) \text{ holds for all } R > 0$ and that we have the limit condition $\lim_{R \to \infty} \overline{d}(R) = 0$. Then a variational inequality

$$|J_q(x^1 - x_0), x - x^1| \leq D_q^{\psi_1}(x^1, x)^{1/q} f \left( \frac{\|F(x^1)\|}{D_q^{\psi_1}(x^1, x)} \right)$$

(16)

holds with the index function

$$f(t) = 2 \max[1, \zeta^{-1/q}] \overline{d}(\Psi^{-1}(t)) \quad t > 0,$$

with

$$\Psi(R) = \left( \frac{\overline{d}(R)}{R} \right)^q \quad R > 0,$$

(17)

for all $x \in \mathcal{D}(F)$ such that $x - x^1 \notin \mathcal{N}(F(x^1))$.

**Proof.** Since the infimum in (14) is a minimum, we have for all $R \geq 0$ an additive decomposition (13) with $\|r_R\|_{x^\ast} = d(R)$. Then the following equations and estimates can be stated for $0 < R < \infty$:

$$|\langle J_q(x^1 - x_0), x - x^1 \rangle| = |\langle F(x^1) \delta w_R + r_R, x - x^1 \rangle|$$

$$= |\langle w_R, F(x^1)(x - x^1) \rangle + \langle r_R, x - x^1 \rangle|$$

$$\leq R\|F(x^1)(x - x^1)\| + d(R)\|x - x^1\|.$$ 

Taking into account the $q$-convexity of $X$ this yields

$$|\langle J_q(x^1 - x_0), x - x^1 \rangle| \leq R\|F(x^1)(x - x^1)\| + \frac{d(R)}{\zeta^{1/q}} D_q^{\psi_1}(x^1, x)^{1/q}$$

$$\leq R\|F(x^1)(x - x^1)\| + \frac{\overline{d}(R)}{\zeta^{1/q}} D_q^{\psi_1}(x^1, x)^{1/q}$$

$$\leq \max\{1, \zeta^{-1/q}\} [R\|F(x^1)(x - x^1)\| + \overline{d}(R)D_q^{\psi_1}(x^1, x)^{1/q}].$$

Since $\Psi(R)$ is strictly decreasing and continuous for $0 < R < \infty$ with limits $\lim_{R \to 0} \Psi(R) = \infty$ and $\lim_{R \to \infty} \Psi(R) = 0$, the equation $\Psi(R) = (\frac{\overline{d}(x^1, x^0)}{D_q^{\psi_1}(x^1, x^0)})^{1/q}$ has a unique solution $\zeta_0 > 0$ for all $x \in \mathcal{D}(F)$ such that $x - x^1 \notin \mathcal{N}(F(x^1))$. For that $\zeta_0 > 0$ the two terms in the last sum above coincide and we obtain the estimate (16). As $\Psi^{-1}(t)$ is strictly decreasing for all $0 < t < \infty$ with limits $\lim_{t \to 0} \Psi^{-1}(t) = \infty$ and $\lim_{t \to \infty} \Psi^{-1}(t) = 0$, under the assumption on $\overline{d}$ stated in the proposition the composite function $\overline{d} \circ \Psi^{-1}$ is an index function. This completes the proof.

**Remark 1.** The function $f$ from (17) has the following property: by using the monotonically inverting substitution $R := \Psi^{-1}(t)$, the quotient function

$$\zeta(t) := \frac{t^{1/q}}{d(\Psi^{-1}(t))} = \frac{\Psi(R)^{1/q}}{d(R)} \leq \frac{\overline{d}(R)}{Rd(R)} = \frac{1}{R}$$

is strictly increasing for $0 < t < \infty$, and tends to zero as $t \to 0$ and $R \to \infty$, respectively. Hence, the quotient $\frac{t^{1/q}}{d(\Psi^{-1}(t))}$ is strictly decreasing for all $t > 0$. Moreover, we should note here that for 2-convex Banach spaces $X$, i.e. for $q = 2$, the variational inequality (16) obtained by
3. Convergence rates with a priori parameter choice

To prove convergence rates we make the following assumption on the nonlinearity of $F$:

$$\sup_{v, \bar{\nu} \in X, x^\dagger + \nu \in B} \frac{\| (F'(x^\dagger + \bar{\nu}) - F'(x^\dagger))v \|}{\| F'(x^\dagger)v \|^2 + \| F'(x^\dagger)\bar{\nu} \|^2} \leq K$$  \hspace{1cm} (18)

with

$$\tilde{c}_1 + \tilde{c}_2 r \geq \frac{1}{2} \quad \text{and} \quad \tilde{c}_3 + \tilde{c}_4 r \geq \frac{1}{2}$$

and

$$(\tilde{c}_1 + \tilde{c}_2 r > \frac{1}{2} \wedge \tilde{c}_3 + \tilde{c}_4 r > \frac{1}{2}) \text{ or } K \text{ sufficiently small}.$$  \hspace{1cm} (19)

The latter, for $v = 1$, follows from the usual Lipschitz condition on $F'$ in terms of the Bregman distance in $X$:

$$\sup_{v, \bar{\nu} \in X, x^\dagger + \nu \in B} \frac{\| (F'(x^\dagger + \bar{\nu}) - F'(x^\dagger))v \|^2}{D^\lambda_{\nu}(x^\dagger, v + x^\dagger)D^\lambda_{\bar{\nu}}(x^\dagger, \bar{\nu} + x^\dagger)} \leq L^2.$$  \hspace{1cm} (20)

Note the relation to the concept of degree of nonlinearity, see, e.g., [9], with (18) implying (2.5) in [9, definition 2.5] for $c_1 = \tilde{c}_1 + \tilde{c}_3$, $c_2 = \tilde{c}_2 + \tilde{c}_4$. The necessity of using a slightly stronger condition here comes from the need for estimating the difference between the derivatives of $F$ in the proof of theorem 1, see (38) below.

An a priori choice of $\alpha_k$ and $k_\delta$ satisfying

$$\alpha_0 \leq 1, \quad \alpha_k \to 0 \quad \text{as} \quad k \to \infty, \quad 1 \leq \frac{\alpha_k}{\alpha_{k+1}} \leq \tilde{C} \quad \text{for all} \quad k$$  \hspace{1cm} (21)

and

$$k_\delta(\delta) = \min \left\{ k \in \mathbb{N} : \frac{\alpha_k^{\frac{1}{\nu}}}{\alpha_{k+1}^{\frac{1}{\nu}}} \leq \tau \delta \right\}, \quad \text{in the case of (9)}$$  \hspace{1cm} (22)

$$k_\delta(\delta) = \min \left\{ k \in \mathbb{N} : \alpha_k \leq \varphi_{\nu}(\delta) \right\}, \quad \text{in the case of (11)}$$  \hspace{1cm} (23)

with

$$\varphi_{\nu}(t) = t^{\nu-2}\Theta^{-1}(t), \quad \Theta(\lambda) := f(\lambda)\sqrt{\lambda}$$  \hspace{1cm} (24)

yields the following rate result.

**Proposition 2.** Assume that a solution $x^\dagger$ to (1) exists, and that $F$ satisfies (18) with (19), (20). Moreover, let $p, r \in (1, \infty)$, let $\tau$ be chosen sufficiently large and let $x_0$ be close enough to $x^\dagger$ so that $D^\lambda_{\nu}(x^\dagger, x_0)$ is sufficiently small. Additionally, assume that $B^\lambda_{\nu}(x^\dagger)$ is $B$ for some $\hat{\rho} > 0$, where $B^\lambda_{\nu}(x^\dagger)$ is a ball with respect to the Bregman distance.

Then for all $k \leq k_\delta(\delta)$ the iterates $x^\dagger_{k+1}$ with $x^\dagger_k(\delta)$ according to (22), (23), the iterates $x^\dagger_{k+1}$ with $\alpha_k$ according to (21) are well defined.
Proof. The assertion follows from results in [18]. □

Theorem 1. Let the assumptions of proposition 2 be satisfied.

(i) Let a variational inequality (9) with $\beta$ sufficiently small hold.

Then, with the a priori choice (22) we obtain optimal convergence rates

$$D^\delta_p(x^1, x_{k^*}) = O(\delta^2/\Theta^\alpha_k), \text{ as } \delta \to 0$$

as well as in the noise free case $\delta = 0$

$$\|T(x_{k+1}^\delta - x^1)\| = O(\alpha_k^{2/\theta^\alpha_k}),$$

$$D^\delta_p(x^1, x_{k+1}^\delta) = O(\alpha_k^{2/\theta^\alpha_k})$$

for all $k \in \mathbb{N}$.

(ii) Let a variational inequality (11) with

$$t \mapsto \frac{f(t)}{\sqrt{t}} \text{ monotonically decreasing,}$$

and

$$\forall 0 < t \leq \hat{t} : f(\hat{C}_t t) \leq \hat{C}_t f(t)$$

$$\forall 0 < t \leq \tilde{t} : f(\tilde{C}_t t) \leq \tilde{C}_t f(t)$$

with

$$\hat{C}_t = \left(\hat{C} \hat{C}^2_t \right)^{2/\theta_t}, \quad \hat{C}_t = \left(\hat{C} \hat{C}^2_t \right)^{1/r},$$

$$1 \leq \hat{C}_0 := \left(\hat{C} \hat{C}^2_0 \right)^{1/r} \leq \frac{1}{(2C_0)^{1/\theta_0} K}, \quad 1 \leq \hat{C}_0 := \left(\hat{C} \hat{C}^2_0 \right)^{1/r},$$

$$\hat{C} = (2M)^{2-r} \hat{C}, \quad \hat{C}_t, C_0, M \text{ as in (21), (50), (51)}$$

$$\hat{t} = \Theta^{-1}\left(\hat{C}_t \psi^{-1}(\alpha_0)\right)/\hat{C}_t, \quad \tilde{t} = \Theta^{-1}\left(\tilde{C}_t \psi^{-1}(2M)^{2-r} \alpha_0\right)/\tilde{C}_t,$$

hold and assume

$$\tilde{c}_1 = \tilde{c}_3 = \frac{1}{2}, \quad \tilde{c}_2 = \tilde{c}_4 = 0,$$

as well as $K$ sufficiently small in (18).

Then with the a priori choice (23), we obtain optimal convergence rates

$$D^\delta_p(x^1, x_{k^*}) = O(f^2(\Theta^{-1}(\delta))) = O\left(\frac{\delta^2}{\Theta^{-1}(\delta)}\right) \text{ as } \delta \to 0$$

with $\Theta$ as in (24), as well as in the noise free case $\delta = 0$

$$\|T(x_{k+1}^\delta - x^1)\| = O(\psi^{-1}(\alpha_k)), $$

$$D^\delta_p(x^1, x_{k+1}^\delta) = O(\psi^{-1}(\alpha_k))$$

for all $k \in \mathbb{N}$.

Remark 2. Condition (27) implies for all $C > 0$ the inequality

$$f(\Theta^{-1}(C t)) \leq \max\{\sqrt{C}, 1\} f(\Theta^{-1}(t)) \quad (t \geq 0).$$

Because of the monotonicity of the index functions $f$ and $\Theta^{-1}$, we have $f(\Theta^{-1}(C t)) \leq f(\Theta^{-1}(t))$ for $0 < C \leq 1$. On the other hand, by substituting $u := \Theta(t)$ we have that

$$\frac{f(\Theta^{-1}(C t))}{\sqrt{C}} = \frac{f(u)}{\sqrt{\Theta}(u)} = \sqrt{\frac{f(u)}{\Theta(u)}}$$

showing in view of (27) that these quotient functions with positive
arguments $\tau$ and $u$, respectively, are both monotonically increasing. Consequently, we have $\frac{f(\Theta^{-1}(\tau))}{\sqrt{\tau^2}} \leq \frac{f(\Theta^{-1}(u))}{\sqrt{u^2}}$ for $C > 1$. Both facts imply together (33).

Moreover, condition (27) means that the variational inequality condition determined by the index function $f$ is not too strong, i.e. the decay rate of $f(t) \to 0$ as $t \to 0$ is not faster than the corresponding decay rate of $\sqrt{t}$. A sufficient condition for that is the concavity of $f^2$ which is equivalent to condition (10). From remark 1 we learned that condition (27) is satisfied for the function $f$ from proposition 1 whenever $q = 2$. By the same arguments it follows that this remains true for all $2 \leq q < \infty$.

We wish to point out that (18), (19) and (20) get weaker for a larger smoothness index $\nu$, which corresponds to results in Hilbert space (see, e.g., [5]), where—as here—in the case $\nu = 1$ a Lipschitz condition suffices to prove optimal convergence rates. In the case of a general index function $f$, we have to restrict ourselves to the strongest case in (18), (19) and (20) corresponding to $\nu = 0$.

Note that $q$-convexity of $X$ is not required for the results of theorem 1. If $X$ is $q$-convex, then inequality (4) implies
\[
\|\tilde{x} - x\|^q = O\left(\frac{\delta^{2}}{\Theta^{-1}(\delta)}\right), \quad \text{as} \ \delta \to 0
\]
in case (i) of theorem 1 and
\[
\|\tilde{x} - x\|^q = O\left(f^2(\Theta^{-1}(\delta))\right) = O\left(\frac{\delta^2}{\Theta^{-1}(\delta)}\right), \quad \text{as} \ \delta \to 0
\]
in case (ii) of theorem 1.

Proof. To show (i), observe that under the assumption (9) we get, with the notation $T = F'(x^\dagger)$,
\[
\|x^{\delta}_{k+1} - x_0\|^p - \|x^\dagger - x_0\|^p = p\Delta_p(x^\dagger - x_0, x^{\delta}_{k+1} - x_0) + p\left[\mathcal{F}_p(x^\dagger - x_0, x^{\delta}_{k+1} - x^\dagger)\right]
\geq pD_p^\mu(x^\dagger, x^{\delta}_{k+1}) - p\beta D_p^\mu(x^\dagger, x^{\delta}_{k+1})^{(1-\nu)/2} \|T(x^{\delta}_{k+1} - x^\dagger)\|^\nu
\geq pD_p^\mu(x^\dagger, x^{\delta}_{k+1}) - p\beta\left(\epsilon D_p^\mu(x^\dagger, x^{\delta}_{k+1}) + C\left(\epsilon, \frac{\nu + 1}{2}\right) \|T(x^{\delta}_{k+1} - x^\dagger)\|^{2\nu/(\nu+1)}\right)
\]
with $\epsilon > 0$ to be chosen sufficiently small later on,
\[
C(\epsilon, 1) = 1,
\]
and
\[
\begin{align*}
C(\epsilon, \mu) &= \max \left\{ 1, \frac{\mu}{(1-\mu)} \right\} \\
&= \max \left\{ 1, \frac{\mu}{(1-\mu)(\nu+1)/\nu} \right\}
\end{align*}
\]
for $\mu \in (0, 1)$, where $\phi(\lambda) = \frac{\nu + 1}{\nu}$ so that
\[
\lambda^\mu \leq \epsilon + C(\epsilon, \mu)\lambda \quad \text{for all} \ \lambda > 0.
\]

By minimality in (3) we have for any solution $x^\dagger \in B_\rho(x_0)$ of (1)
\[
\|T_k(x^{\delta}_{k+1} - x_0) + g_k\|^p + \alpha_k \|x^{\delta}_{k+1} - x_0\|^p \\
\leq \|T_k(x^\dagger - x_0) + g_k\|^p + \alpha_k \|x^\dagger - x_0\|^p.
\]
Combining (34) and (36) we get by the simple inequality \((a - b') + b' \geq \frac{1}{2}a'\)

\[
\frac{1}{2^{r-1}} \left\| T(x_{k+1} - x_1) \right\|^r + \alpha_k p(1 - \beta e)D_p^{\gamma_0}(x_1, x_{k+1}) \\
\leq \left\| T_k(x_1 - x_k^2) + g_k \right\|^r + \left\| (T_k - T)(x_k^2 - x_1) \right\|^r + \left\| T_k(x_k^2 - x_1) + g_k \right\|^r \\
+ \alpha_k p\beta C \left( \epsilon, \frac{\nu + 1}{2} \right) \left\| T(x_{k+1}^2 - x_1) \right\|^{2^{r/(r+1)}}
\]

The terms on the right-hand side can be estimated by means of (18),

\[
\left\| T_k(x_1 - x_k^2) + g_k \right\| \leq \left\| (T_k - T)(x_k^2 - x_1) \right\| + \left\| F(x_k^2) - F(x_1) \right\| + \left\| T(x_k^2 - x_1) \right\| + \delta \\
\leq 2K \left\| T(x_k^2 - x_1) \right\|^{2^{r/(r+1)} + \delta} + \delta
\]

\[
\left\| (T_k - T)(x_k^2 - x_1) \right\| \leq K \left\| T(x_k^2 - x_1) \right\|^{2^{r/(r+1)}} \left\| T(x_k^2 - x_1) \right\|^{\frac{\nu}{r}D_p^{\gamma_0} (x_1, x_k^2)^{\frac{1}{r}D_p^{\gamma_0} (x_1, x_k^2)}}
\]

which, together with the simple inequality \((a + b') \leq 2^{r-1}(a' + b')\), yields

\[
\frac{1}{2^{r-1}} \left\| T(x_{k+1} - x_1) \right\|^r + \alpha_k p(1 - \beta e)D_p^{\gamma_0}(x_1, x_{k+1}) \\
\leq (1 + 2^{r-1})(2K \left\| T(x_k^2 - x_1) \right\|^{2^{r/(r+1)} + \delta} + \delta) \\
+ 2^{r-1}(K \left\| T(x_k^2 - x_1) \right\|^{2^{r/(r+1)} + \delta})^{\frac{\nu}{r}D_p^{\gamma_0} (x_1, x_k^2)^{\frac{1}{r}D_p^{\gamma_0} (x_1, x_k^2)}} \\
+ \alpha_k p\beta C \left( \epsilon, \frac{\nu + 1}{2} \right) \left\| T(x_{k+1}^2 - x_1) \right\|^{2^{r/(r+1)}}
\]

Applying the estimate

\[
a^\beta b \leq \hat{a}a + C(\hat{a}, 1 - \zeta)b^{\lambda(1-\zeta)}
\]

for \(\zeta \in (0, 1]\), that follows from (35) with \(\lambda \leq \frac{\beta(1-\zeta)}{\alpha}\) and \(\mu = 1 - \zeta\) to the last term, and \(ab \leq \frac{1}{2}a^2 + \frac{1}{2}b^2\) to the second term on the right-hand side, we get

\[
\left( \frac{1}{2^{r-1}} - \hat{\epsilon} \right) \left\| T(x_{k+1} - x_1) \right\|^r + \alpha_k p(1 - \beta e)D_p^{\gamma_0}(x_1, x_{k+1}) \\
\leq (1 + 2^{r-1})(2K \left\| T(x_k^2 - x_1) \right\|^{2^{r/(r+1)} + \delta} + \delta) \\
+ \frac{2^{r-1}K'}{2} \left\| T(x_k^2 - x_1) \right\|^{2^{r/(r+1)} + \delta} \\
+ \left\| T(x_k^2 - x_1) \right\|^{2^{r/(r+1)} + \delta} \\
+ C \left( \hat{\epsilon}, \frac{\nu(1 + 2^{r-1})}{r(1 + 2^{r-1})} \right) \left( \alpha_k p\beta C \left( \epsilon, \frac{\nu + 1}{2} \right) \right)^{\frac{\nu(1 + 2^{r-1})}{r(1 + 2^{r-1})}}
\]

where we choose \(\hat{\epsilon} = \frac{1}{2^{r-1}}\). Considering (40) and (44) and neglecting the rest (which is just an estimate of the nonlinearity error) for a moment, we expect that (26) can be obtained, which we prove as follows: dividing (40)–(44) by \(\alpha_k^{\frac{\nu(1 + 2^{r-1})}{r(1 + 2^{r-1})}}\), using (21), (19) and (22), and defining

\[
\gamma_k := \max \left\{ \frac{\left\| T(x_{k+1} - x_1) \right\|}{\alpha_k^{\frac{\nu(1 + 2^{r-1})}{r(1 + 2^{r-1})}}}, \frac{D_p^{\gamma_0}(x_1, x_{k+1})}{\alpha_k^{\frac{\nu(1 + 2^{r-1})}{r(1 + 2^{r-1})}}} \right\}
\]
we get the following estimate:

\[
\min \left\{ \frac{1}{2\gamma^r - 1} - \tilde{c}, \ p(1 - \beta \epsilon) \right\} \gamma_{k+1} \leq \left(1 + 2\gamma^r - 1\right) 2^{\gamma^r - 1} (2K)^r \hat{C}_{\gamma_k} \gamma_k \gamma_{k+1}^{2(\gamma_k^r + \gamma_k^r)} + \gamma_k \gamma_{k+1}^{2(\gamma_k^r + \gamma_k^r)} + \frac{2^{\gamma^r - 1} K^r}{2} \gamma_{k+1}^{2(\gamma_k^r + \gamma_k^r)} + \frac{2^{\gamma^r - 1} K^r}{2} \gamma_{k+1}^{2(\gamma_k^r + \gamma_k^r)} + \frac{C(\varepsilon, r(\nu + 1) - 2\nu)}{r(\nu + 1)} \left( p\beta C \left( \varepsilon, \frac{\nu}{2} \right) \right) \gamma_k \gamma_{k+1}^{2(\gamma_k^r + \gamma_k^r)} + \gamma_k \gamma_{k+1}^{2(\gamma_k^r + \gamma_k^r)} + \frac{1 + 2\gamma^r - 1}{\tau^r}.
\]

Firstly we get a recursive estimate of the form

\[
\left(1 - A \gamma_k \gamma_{k+1}^{2(\gamma_k^r + \gamma_k^r)}\right) \gamma_{k+1} \leq B \left( \gamma_k \gamma_{k+1}^{2(\gamma_k^r + \gamma_k^r)} + \gamma_k \gamma_{k+1}^{2(\gamma_k^r + \gamma_k^r)} \right) \gamma_k + c,
\]

(45)

where \( c \) can be made small by making \( \beta \) small and \( \tau \) large.

From this we can now derive an induction step of the form

\[
\gamma_k \leq \tilde{\gamma} \Rightarrow \gamma_{k+1} \leq \tilde{\gamma}
\]

(46)

as follows: using (20) and the fact that \( A \) and \( B \) will be small if \( K \) is small, we can first of all conclude that for \( \tilde{\gamma}, \tilde{\xi} \) sufficiently small, the function

\[
h(\gamma) : \begin{align*}
(0, \tilde{\gamma}) & \quad \rightarrow (0, \tilde{\xi}) \\
\gamma & \quad \mapsto (1 - A \gamma) \gamma \gamma_{k+1}^{2(\gamma_k^r + \gamma_k^r)}
\end{align*}
\]

is strictly monotonically increasing and invertible with

\[
h^{-1}(\tilde{\xi}) \leq 2\tilde{\xi}.
\]

By using the induction hypothesis \( \gamma_k \leq \tilde{\gamma} \) with a possibly reduced value of \( \tilde{\gamma} \), we can achieve that the right-hand side of (45) is smaller than \( \tilde{\gamma} \) so that by applying \( h^{-1} \) to both sides of (45), we can conclude

\[
\gamma_k \leq 2B(\gamma_k \gamma_{k+1}^{2(\gamma_k^r + \gamma_k^r)} + \gamma_k \gamma_{k+1}^{2(\gamma_k^r + \gamma_k^r)}) \gamma_k + 2c
\]

\[
\leq 2B(\gamma_k \gamma_{k+1}^{2(\gamma_k^r + \gamma_k^r)} + \gamma_k \gamma_{k+1}^{2(\gamma_k^r + \gamma_k^r)}) \gamma_k + \tilde{c}.
\]

(47)

where we use the fact that we can make \( \beta \) small and \( \tau \) large so that \( c < \frac{\tilde{c}}{2} \). Now we use (20) again to achieve

\[
2B(\gamma_k \gamma_{k+1}^{2(\gamma_k^r + \gamma_k^r)} + \gamma_k \gamma_{k+1}^{2(\gamma_k^r + \gamma_k^r)}) \gamma_k + \tilde{c} \leq \frac{1}{2}
\]

by possibly decreasing \( \tilde{\gamma} \). Inserting this into (47) yields \( \gamma_{k+1} \leq \tilde{\gamma} \).

Applying (46) as an induction step we can conclude that

\[
\gamma_k \leq \tilde{\gamma} \quad \text{for all } k \leq k_0
\]

and therewith, by possibly decreasing \( \tilde{\gamma} \) to below \( \tilde{\gamma}^2 \),

\[
D_{p,0}(x_1, x_k) \leq \gamma_k a_k \frac{\tilde{\gamma}}{\tau \delta} \leq \tilde{\gamma}^2 \quad \text{for all } k \leq k_0
\]

provided \( \gamma_0 \) and \( D_{p,0}(x_1, x_0) \) are sufficiently small. By the assumption \( B_{p,0}(x_1) \subseteq B \), this yields well definedness of the iterates. Moreover,

\[
D_{p,0}(x_1, x_k) \leq \gamma_k a_k \frac{\tilde{\gamma}}{\tau \delta} \leq \tilde{\gamma}^2
\]

In the general case (ii) i.e. with the variational inequality (11), we have to apply somewhat different techniques as compared to the special case (9). We get, in place of (34), the estimate

\[
\| x_{k+1} - x_0 \|_p \leq \| x_1 - x_0 \|_p
\]

\[
= p \Delta_p(x_1 - x_0, x_{k+1} - x_0) + p(J_p(x_1 - x_0), x_{k+1} - x_0) \]

\[
\geq p D_{p,0}(x_1, x_{k+1}) - p D_{p,0}(x_1, x_{k+1})^{1/2} f \left( \frac{\| F(x_1)(x_{k+1} - x_1) \|_2}{D_{p,0}(x_1, x_{k+1})} \right),
\]

(48)
which together with (36)–(38) implies
\[
\frac{1}{2r-1} \left\| T(x^k - x^1) \right\|^r + \alpha_k p D_p^{\text{loc}}(x^1, x^k) \leq (1 + 2r^{-1}) \left( 2K \left\| T(x^k - x^1) \right\| + \delta \right)^r
\]
\[
+ \frac{2r^{-1} K^r}{2} \left\| T(x^k - x^1) \right\|^r + \frac{2r^{-1} K^r}{2} \left\| T(x^k - x^1) \right\|^r
\]
\[
+ \alpha_k p D_p^{\text{loc}}(x^1, x^k) \leq \left( 1 + 2r^{-1} \right)^r \left\| T(x^k - x^1) \right\|^r
\]
in place of (40)–(44), which by moving the second term on the right-hand side to the left-hand side, using \( K^r < \frac{2}{2r-1} \) and (23), yields an inequality of the form
\[
t^r_{k+1} + \alpha_k \tilde{d}^2_{k+1} \leq \kappa t^r_k + m(\psi^{-1}(\alpha_k))' + M\alpha_k d_{k+1} f \left( \frac{t^r_{k+1}}{d^2_{k+1}} \right)
\]
for all \( k \leq k_* - 1 \), where we use the abbreviations
\[
d_k = D_p^{\text{loc}}(x^1, x^k)^{1/2},
\]
\[
t_k = \left\| T(x^k - x^1) \right\|,
\]
\[
k = \frac{2r(1 + 2r^{-1})2r^{-1} + 2r^{-1}/2}{\tilde{c}} K' = C_* K',
\]
\[
m = \frac{(1 + 2r^{-1})2r^{-1}}{\tau r \tilde{c}},
\]
\[
M = \frac{p}{\tilde{c}},
\]
(51)
\[
\tilde{c} = \min \left\{ \frac{1}{2r-1} - \frac{2r^{-1} K^r}{2}, p \right\}.
\]

Now we prove by induction that for all \( k \leq k_* \) (or in the case \( \delta = 0 \) for all \( k \in \mathbb{N} \))
\[
d_k \leq C_1 f \left( \Theta^{-1}(\psi^{-1}(\alpha_k)) \right)
\]
\[
t_k \leq C_2 \psi^{-1}(\alpha_k)
\]
and \( C_2 \) is sufficiently large so that (cf (30))
\[
\hat{C}_\psi \leq (2(\kappa + m/C_2))^{-1/r}, \quad \tilde{C}_\psi \leq \frac{C_2}{2M}
\]
(54)

For this purpose, observe that (49) together with the induction hypothesis implies
\[
t^r_{k+1} + \alpha_k \tilde{d}^2_{k+1} \leq (\kappa C_* + m) \left( \psi^{-1}(\alpha_k) \right)' + M\alpha_k d_{k+1} f \left( \frac{t^r_{k+1}}{d^2_{k+1}} \right).
\]
(56)

We distinguish between two cases:
\[
\text{if } \left( \kappa C_* + m \right) \left( \psi^{-1}(\alpha_k) \right)' \leq M\alpha_k d_{k+1} f \left( \frac{t^r_{k+1}}{d^2_{k+1}} \right), \text{ we get from (56)}
\]
\[
t^r_{k+1} + \alpha_k \tilde{d}^2_{k+1} \leq 2M\alpha_k d_{k+1} f \left( \frac{t^r_{k+1}}{d^2_{k+1}} \right).
\]
(57)
Since in the case \( d_{k+1} = 0 \) (and therewith \( t_{k+1} = 0 \)) and in the case \( t_{k+1} = 0 \) (and therewith \( d_{k+1} = 0 \) by \( d_{k+1}^2 \leq 2M d_{k+1} f(\frac{\beta_k}{\alpha_k t_{k+1}}) \)), the assertions (52) and (53) trivially hold for \( k \) replaced by \( k + 1 \), we may assume w.l.o.g. that \( d_{k+1} \neq 0 \) and \( t_{k+1} \neq 0 \). Multiplying (57) with \( t_{k+1} \) and dividing by \( d_{k+1}^2 \) we get

\[
\frac{t_{k+1}^2}{d_{k+1}^2} t_{k+1}' - \alpha_k t_{k+1} \leq 2M \alpha_k \Theta \left( \frac{t_{k+1}^2}{d_{k+1}^2} \right),
\]

which implies

\[
\Phi \left( \frac{t_{k+1}^2}{d_{k+1}^2} \right) t_{k+1}' - \alpha_k t_{k+1} \leq 2M \alpha_k
\]

with, according to (27), the monotonically increasing function

\[
\Phi : u \mapsto \frac{\sqrt{u}}{f(u)} = \frac{u}{\Theta(u)}
\]

and

\[
t_{k+1} \leq 2M \Theta^{-1} \left( \frac{t_{k+1}^2}{2M} \right) \quad \text{i.e.} \quad \Theta^{-1} \left( \frac{t_{k+1}^2}{2M} \right) \leq \frac{t_{k+1}^2}{d_{k+1}^2}. \tag{58}
\]

consequently

\[
\Phi \left( \Theta^{-1} \left( \frac{t_{k+1}^2}{2M} \right) \right) t_{k+1}' - \alpha_k \leq 2M \alpha_k.
\]

Since \( \Phi \left( \Theta^{-1} \left( \frac{1}{t} \right) \right) t^{-1} = C \Theta^{-1} \left( \frac{1}{t} \right) t^{-2} = \phi_t \left( \frac{1}{t} \right) C^{-1} \), this implies

\[
t_{k+1} \leq 2M \phi_t^{-1}((2M)^{-2} \alpha_k) \tag{59}
\]

from which by (58) we get

\[
d_{k+1}^2 \leq \frac{t_{k+1}^2}{\Theta^{-1} \left( \frac{t_{k+1}^2}{2M} \right)} = (2M)^2 \left( f \left( \Theta^{-1} \left( \frac{t_{k+1}^2}{2M} \right) \right) \right)^2 \leq (2M)^2 \left( f \left( \Theta^{-1} \left( \frac{1}{(2M)^{-2} \alpha_k} \right) \right) \right)^2. \tag{60}
\]

Otherwise, if \((\kappa C^*_z + m) (\phi_t^{-1}(\alpha_k)') \geq M \alpha_t d_{k+1} f(\frac{\beta_k}{\alpha_k t_{k+1}}) \), we get from (56)

\[
t_{k+1}' + \alpha d_{k+1}^{-1} \leq 2(\kappa C^*_z + m) (\phi_t^{-1}(\alpha_k)'), \tag{61}
\]

From (59)–(61), using the identity

\[
f(\Theta^{-1}(\phi_t^{-1}(\alpha))) = \frac{z}{\sqrt{\Theta^{-1}(z)}} = \frac{1}{\sqrt{\theta^{-2} \Theta^{-1}(z)}} = \frac{1}{\sqrt{\psi_t(z)}} \theta^{1/2} = \frac{1}{\sqrt{\alpha}} (\phi_t^{-1}(\alpha))^{1/2},
\]

and (21), we see that in order to complete the induction proof of (52), (53), it suffices to show

\[
\phi_t^{-1}(\alpha) \leq \hat{C}_t \phi_t^{-1}(\alpha/\hat{C}) \quad \forall 0 < \alpha \leq \alpha_0, \tag{62}
\]

\[
\phi_r^{-1}(\alpha) \leq \hat{C}_r \phi_r^{-1}(\alpha/\hat{C}) \quad \forall 0 < \alpha \leq (2M)^{-2} \alpha_0. \tag{63}
\]

and use (54), (55). By the definition of \( \phi_t \), (62) can be concluded from (28) as follows: with \( \hat{C}_t = \sqrt{\hat{C}_r \hat{C} t}, \hat{C}_r = \hat{C} \bar{C}^{2-r} \) (cf (30)), \( \lambda = \hat{C}_t \phi_t^{-1}(\alpha/\hat{C}) \), \( t = \Theta^{-1}(\lambda)/\hat{C}_t \), we have for any
\( \alpha \in (0, \alpha_0] \):

\[
\begin{align*}
f(\hat{C}_r t) \leq \hat{C}_f f(t) & \iff \Theta(\hat{C}_r t) \leq \sqrt{\hat{C}_r \hat{C}_f} \Theta(t) \\
& \iff \Theta^{-1}(\lambda/\hat{C}_r) \leq \frac{1}{\hat{C}_r} \Theta^{-1}(\lambda)
\end{align*}
\]

\[
\begin{align*}
\Theta^{-1}(\lambda/\hat{C}_r) \leq t = \frac{1}{\hat{C}_r} \Theta^{-1}(\lambda)
\end{align*}
\]

\[
\begin{align*}
\Theta^{-1}(\lambda/\hat{C}_r) \leq \frac{1}{\hat{C}_r} \Theta^{-1}(\lambda)
\end{align*}
\]

\[
\begin{align*}
\Theta^{-1}(\lambda/\hat{C}_r) \leq \frac{1}{\hat{C}_r} \Theta^{-1}(\lambda)
\end{align*}
\]

\[
\begin{align*}
\hat{C}_r \Theta^{-2} \varphi_r(\lambda/\hat{C}_r) \leq \varphi_r(\lambda)
\end{align*}
\]

\[
\begin{align*}
\hat{C}_r \Theta^{-2} \varphi_r(\lambda/\hat{C}_r) \leq \varphi_r(\lambda)
\end{align*}
\]

where we have used the fact that the functions \( \varphi_r, \Theta \) as well as their inverses are strictly monotonically increasing. Analogously, (63) follows from (29). Therewith, the induction proof of (52), (53) is finished.

The estimates (52), (53) immediately yield (32).

Inserting (23) into (52) for \( k = k_n \) directly yields with (33)

\[
\begin{align*}
d_{k_n} \leq C_1 f(\Theta^{-1}(\varphi^{-1}_r(\alpha_{k_n}))) \leq C_1 f(\Theta^{-1}(\tau \delta)) \leq C_1 \max\{\sqrt{\tau}, 1\} f(\Theta^{-1}(\delta))
\end{align*}
\]

\[
\begin{align*}
= C_1 \max\{\sqrt{\tau}, 1\} \Theta(\Theta^{-1}(\delta)) = C_1 \max\{\sqrt{\tau}, 1\} \frac{\delta}{\sqrt{\Theta^{-1}(\delta)}}.
\end{align*}
\]

This provides us with the convergence rate assertion (31) and completes the proof of (ii).

Corollary 1. Let the assumptions of propositions 2 and 1 with

\[
\begin{align*}
q = p = 2,
\end{align*}
\]

and

\[
\begin{align*}
\forall R \geq \hat{R} : \quad \mathcal{J}(\hat{C}_f \hat{C}_r^{-1/p} R) & \leq \hat{C}_f \mathcal{J}(R), \\
\forall R \geq \tilde{R} : \quad \mathcal{J}(\tilde{C}_f \tilde{C}_r^{-1/p} R) & \leq \tilde{C}_f \mathcal{J}(R),
\end{align*}
\]

(64)

(65)

with (30) hold, where

\[
\begin{align*}
\hat{R} = \Psi^{-1}(\Theta^{-1}(\varphi^{-1}_r(\alpha_0))) / \hat{C}_r,
\end{align*}
\]

\[
\begin{align*}
\tilde{R} = \Psi^{-1}(\Theta^{-1}(\varphi^{-1}_r((2M)^{-2} \alpha_0)) / \tilde{C}_r).
\end{align*}
\]

Moreover, assume that

\[
\begin{align*}
\tilde{c}_1 = \tilde{c}_3 = \frac{1}{2}, \quad \tilde{c}_2 = \tilde{c}_4 = 0,
\end{align*}
\]

and \( K \) is sufficiently small in (18).

Then, with the a priori choice (23), we obtain convergence rates (31), (32) with \( f \) as in (17).

Proof. The assertion follows by a combination of part (ii) of theorem 1, proposition 1 and the fact that (28), (29) can be concluded from (64), (65): with \( R = \Psi^{-1}(t) \) we get for any \( t \in (0, T] \):

\[
\begin{align*}
f(\hat{C}_r t) / f(t) & = \frac{\mathcal{J}(\Psi^{-1}(\hat{C}_r t))}{\mathcal{J}(R)} = \left( \frac{(\mathcal{J}(\Psi^{-1}(\hat{C}_r t)) / \Psi^{-1}(\hat{C}_r t))^p}{(\mathcal{J}(R) / R)^p} \right)^{1/p} \Psi^{-1}(\hat{C}_r t) / R \\
& = \left( \hat{C}_r t / t \right)^{1/p} \Psi^{-1}(\hat{C}_r t) / \Psi^{-1}(t) = \hat{C}_r^{1/p} \Psi^{-1}(\hat{C}_r t) / \Psi^{-1}(t) \leq \hat{C}_f,
\end{align*}
\]
since we have the equivalences
\[ \Psi^{-1}(\hat{C}_r t) \leq \hat{C}_f \hat{C}_r^{-1/p} \Psi^{-1}(t) \iff \hat{C}_f \Psi(R) \geq \Psi(\hat{C}_f \hat{C}_r^{-1/p} R) \]
\[ \iff \hat{C}_r (\hat{d}(R)/R)^p \geq (\hat{d}(\hat{C}_f \hat{C}_r^{-1/p} R)/(\hat{C}_f \hat{C}_r^{-1/p} R))^p \]
\[ \iff \hat{C}_r \hat{d}(R) \geq \hat{d}(\hat{C}_f \hat{C}_r^{-1/p} R)^p , \]
where we have used the fact that $\Psi^{-1}$ is strictly decreasing. Analogously we get (29). Note that (27) is automatically satisfied for $f$ defined by (17), see remark 1. Moreover, in the case $\hat{c}_3 = \hat{c}_4 = 0$ as well as $K$ sufficiently small in (18) that $x^\delta_{k+1}$ solves (1).

4. Convergence rates with a posteriori parameter choice

If the exponent $\nu$ in the source condition is not known, we require a nonlinearity assumption that corresponds to the strongest case $\nu = 0$ in (18)–(20), namely the tangential cone condition
\[ \| F(x) - F(\bar{x}) - F'(x)(x - \bar{x}) \| \leq c_{tc} \| F(x) - F(\bar{x}) \| \quad \forall x, \bar{x} \in B \]
for some $0 < c_{tc} < 1, \rho > 0$. Note that (18) for $\nu = 0$ with $K$ sufficiently small becomes (66) at $x = x^1$ with $c_{tc} = \frac{K}{\kappa}$. Therewith, we can prove convergence rates with a posteriori choices of the regularization parameters $\alpha_k$
\[ \sigma \| g_k \| \leq \| T_k(x_{k+1}^\delta(\alpha_k) - x_0^\delta) + g_k \| \leq \sigma \| g_k \| \leq \sigma \| g_k \| (67) \]
(cf [6]), and of the stopping index $k_*$ by the discrepancy principle:
\[ k_*(\delta) = \min \{ k \in \mathbb{N} : \| F(x_k^\delta) - y^\delta \| \leq \tau \delta \}. \]

Proposition 3. Assume that a solution $x^1$ to (1) exists, that $F$ is weakly sequentially closed (see, e.g., (11), (12) in [18]), and satisfies (64) with $c_{tc}$ sufficiently small
\[ c_{tc} < \sigma < \sigma < 1. \]
Moreover, let $\tau$ be chosen sufficiently large so that
\[ c_{tc} + \frac{1 + c_{tc}}{\tau} \leq \sigma \quad \text{and} \quad c_{tc} < \frac{1 - \sigma}{2} , \]
and let $x_0$ be close enough to $x^1$ so that $D^0_F(x^1, x_0)$ is sufficiently small. Additionally, assume that either

(a) $F'(x) : X \to Y$ is weakly closed for all $x \in D(F)$ and $Y$ reflexive
or
(b) $D(F)$ is weakly closed
and
\[ \delta \leq \frac{\| F(x_0) - y^\delta \|}{\tau} . \]
Then for all $k \leq k_*(\delta)$ with $k_*(\delta)$ according to (68), the iterates
\[ x_{k+1}^\delta := \begin{cases} x_{k+1}^\delta(\alpha_k), \text{ with } \alpha_k \text{ as in (67)} & \text{if } \| T_k(x_0 - x_k^\delta) + g_k \| \geq \sigma \| g_k \| \\ x_0 \text{ else} & \end{cases} \]
are well defined.
**Theorem 2.** Let the assumptions of proposition 3 be satisfied.

(i) Under a variational inequality (9) we obtain optimal convergence rates

\[ D_{\alpha_k}^{\delta}(x^1, x_{k+1}) = O(\delta^{\frac{2}{2\pi}}) \quad \text{as} \quad \delta \to 0. \tag{70} \]

(ii) Under a variational inequality (11) we obtain optimal convergence rates

\[ D_{\beta_k}^{\delta}(x^1, x_{k+1}) = O(f^2(\Theta^{-1}(\delta))) = O\left(\frac{\delta^2}{\Theta^{-1}(\delta)}\right) \quad \text{as} \quad \delta \to 0 \tag{71} \]

with \( \Theta \) as in (24).

**Proof.** The stopping index \( k_*(\delta) \) according to (68) is finite, since on one hand, the case that \( \|T_k(x_0 - x^\delta_k) + g_k\| < \sigma \|g_k\| \) and therewith \( x^\delta_{k+1} := x_0 \) can happen at most every second step:

\[ x^\delta_{k+1} = x_0 \Rightarrow \|T_k(x_0 - x^\delta_{k+1}) + g_{k+1}\| = \|g_{k+1}\| \geq \sigma \|g_{k+1}\|, \]

so \( \alpha_k \) can be chosen as in (67) (with \( k \) replaced by \( k + 1 \)). On the other hand, in steps where \( \alpha_k \) is chosen as in (67), the residual norm decreases by a factor of \( \frac{\sigma + c_{\epsilon \delta}}{1-c_{\epsilon \delta}} \), which is smaller than 1 by (69):

\[ \|g_{k+1}\| = \|T_k(x^\delta_{k+1} - x^\delta_k) + g_k + F(x^\delta_{k+1}) - F(x^\delta_k) - T_k(x^\delta_{k+1} - x^\delta_k)\| \]
\[ \leq \sigma \|g_k\| + c_{\epsilon \delta} \|F(x^\delta_{k+1}) - F(x^\delta_k)\| \]
\[ \leq (\sigma + c_{\epsilon \delta}) \|g_k\| + c_{\epsilon \delta} \|g_{k+1}\|, \]

Hence,

\[ \|g_k\| \leq \left(\frac{\sigma + c_{\epsilon \delta}}{1-c_{\epsilon \delta}}\right)^{\lfloor k/2 \rfloor} \leq \tau \delta \]

for \( k \) sufficiently large.

Estimates (34), (36), together with (66), (2), (67), (68), yield

\[ \alpha_k \|x^\delta_{k+1} - x_0\|^p \leq \left(c_{\epsilon \delta} + \frac{1 + c_{\epsilon \delta}}{\tau}\right)^r \|g_k\|^r + \alpha_k \|x^\delta - x_0\|^p \]

for all \( k \leq k_*(\delta) - 1 \), provided \( x_k \in B_{\rho}(x_0) \).

Inserting (34) into (72) and taking into account (69), (66), we get

\[ (1 - r\epsilon)D_{\beta_k}^{\delta}(x^1, x^\delta_{k+1}) \leq \beta C \left(\frac{\nu + 1}{2}\right)((1 + c_{\epsilon \delta})(x^\delta_{k+1}) + g_k)\|F(x^\delta_{k+1}) - F(x^\delta_k)\|^{2\nu/(\nu+1)} \]

in the case \( \alpha_k \) is chosen as in (67). Hence, with \( \epsilon < \beta^{-1} \), for \( k = k_* - 1 \) the discrepancy principle (68) yields the optimal rate

\[ D_{\alpha_k}^{\delta}(x^1, x^\delta_{k+1}) \leq \frac{\beta C (\epsilon^{-1})}{\left(1 - \beta \epsilon\right)^r}\left((1 + c_{\epsilon \delta})(1 + r)\|g_k\|^r + c_{\epsilon \delta} \|x^\delta - x_0\|^p\right) \]

since by the signal to noise ratio assumption \( \delta < \|F(x_0) - y^\delta\|/\tau \) we can exclude the case \( x^\delta_{k+1} = x_0 \), i.e. the case that \( \alpha_{k+1} \) is not chosen according to (67).

In the general case (11) we get, in place of (34), (73), the estimates (48) and

\[ D_{\beta_k}^{\delta}(x^1, x^\delta_{k+1})^{1/2} \leq f \left(\frac{(1 + c_{\epsilon \delta})^2 \|F(x^\delta_{k+1}) - F(x^\delta_k)\|^2}{D_{\beta_k}^{\delta}(x^1, x^\delta_{k+1})}\right) \]
respectively. Hence, with $k = k_* - 1$, using (66) and (68) we get

$$C\delta = \frac{C\delta}{D_p^0(x^1, x^2_\delta)} = D_p^0(x^1, x^2_\delta)^{1/2}$$

$$\leq \frac{C\delta}{D_p^0(x^1, x^2_\delta)} f\left(\frac{C^2\delta^2}{\Theta^{-1}(C\delta)}\right) = \Theta\left(\frac{(C\delta)^2}{D_p^0(x^1, x^2_\delta)}\right)$$

with $C := (1 + c_\omega)(1 + \tau)$ so taking the inverse of $\Theta$ on both sides, we get

$$D_p^0(x^1, x^2_\delta) \leq \frac{C^2\delta^2}{\Theta^{-1}(C\delta)} \leq C^2 \frac{\delta^2}{\Theta^{-1}(\delta)},$$

since $C > 1$ and $\Theta^{-1}$ is strictly monotonically increasing.

\[\square\]

**Corollary 2.** Under the assumptions of propositions 3, 1 with $q = p = 2$, we obtain convergence rates (71), with $f$ as in (17).

**Remark 3.** Note that proposition 1 together with corollaries 1, 2 for $p = q = 2$ gives a relation between logarithmic decay of the distance function and logarithmic convergence rates (see, e.g., [14, 15]), which are particularly important for exponentially ill-posed problems.

For

$$d(R) = \ln(R) - N (R > e),$$

with some $N > 0$, we get $\Psi(\lambda) = \frac{1}{\ln(R)^{\lambda}} \leq \frac{1}{R}$, hence with $\tilde{C} = 2 \max\{1, c^{-1/2}\}$, we obtain $f(\lambda) = \tilde{C} \ln(\Psi^{-1}(\lambda))^{-\lambda} \leq \tilde{C} \ln(\frac{1}{\lambda})^{-\lambda}$, so $\Theta(\lambda) = f(\lambda) \sqrt{\lambda} \leq \tilde{C} \ln(\frac{1}{\lambda})^{-\lambda}$, which implies for the quotient terms occurring in the convergence rates of corollaries 1 and 2

$$\frac{\delta^2}{\Theta^{-1}(\delta)} = [\Theta(\Theta^{-1}(\delta))]^2 \leq [f(\Theta^{-1}(\delta))]^2 \leq \tilde{C}_2 \ln\left(\frac{1}{\Theta^{-1}(\delta)}\right)^{-2N} \leq \tilde{C}_N \ln\left(\frac{1}{\delta}\right)^{-2N}$$

for some $\tilde{C}_N > 0$. Here we have considered only the case of sufficiently large $R > 0$ which corresponds with sufficiently small noise levels $\delta > 0$.

5. Two parameter identification examples

In this section, we consider two model problems that have previously been studied in the Hilbert space setting, e.g., in [5–7, 16, 22], and in the Banach space setting in [18]. Since in both examples, $X$ and $Y$ will be defined by Lebesgue or Sobolev–Slobodeckij spaces, we first of all quote some facts on these spaces, see, e.g., [2, 8, 24, 26].

**Lemma 1.** Let $\Omega \subseteq \mathbb{R}^{\text{dim}}$ be a smooth domain.

(a) $L^p(\Omega), W^{m,p}(\Omega)$ are $\{2\text{-convex and } P\text{-smooth for } 1 < P \leq 2 \}
\{P\text{-convex and } 2\text{-smooth for } 2 \leq P < \infty \}$.

(b) The duality mapping $J_p$ is given by

$$J_p(x) = \|x\|_{W^{m,p}}^{-P} |x|^{P-1} \text{sgn}(x) \text{ in } X = L^p(\Omega),$$

$$J_p(x) = \|x\|_{W^{1,p}}^{-P} (-\nabla(|\nabla x|^{P-2}\nabla x) + |x|^{P-1} \text{sgn}(x))$$

in $X = W^{1,p}(\Omega)$ if $\frac{\partial x}{\partial n} = 0$ on $\partial \Omega,$

(74)

(75)
\[ J_p(x) = \|x\|_X^{p-\frac{1}{p}} (|\Delta x|^{p-2} \Delta x) - \nabla(|\nabla x|^{p-2} \nabla x) + |x|^{p-1} \operatorname{sgn}(x) \]

in \( X = W^{2,p}(\Omega) \) if \( \frac{\partial x}{\partial n} = \Delta x = 0 \) on \( \partial \Omega \),

(76)

provided that \( W^{2,p}(\Omega) \) is equipped with the norm

\[ \|x\|_{W^{2,p}(\Omega)} = \left( \int_{\Omega} (|\Delta x|^p + |\nabla x|^p + |x|^p) \, dx \right)^{1/p}. \]

**Proof.** Referring to e.g. [2, 8, 24, 26] for (a), and (74), we only show (75), here. If \( \frac{\partial x}{\partial n} = 0 \) on \( \partial \Omega \), then with \( x^* := J_p(x) \) as claimed in (75), we indeed have

\[
\langle x^*, x \rangle_X = \int_{\Omega} \|x\|_X^{p-\frac{1}{p}} (-\nabla(|\nabla x|^{p-2} \nabla x) + |x|^{p-1} \operatorname{sgn}(x)) x \, dx
\]

\[ = \|x\|_X^{p-\frac{1}{p}} \int_{\Omega} (|\nabla x|^p + |x|^p) \, dx = \|x\|_X^p, \]

where we have used integration by parts. Assertion (76) can be shown analogously. \( \square \)

As a first example, we consider identification of the space-dependent coefficient \( c \) in the elliptic boundary value problem

\[
\begin{align*}
-\Delta u + cu &= f \quad \text{in } \Omega \\
u &= 0 \quad \text{on } \partial \Omega
\end{align*}
\]

(77)

(78)

from measurements of \( u \) in \( \Omega \) (note that inhomogeneous Dirichlet boundary conditions can be easily incorporated into the right-hand side \( f \) if necessary). Here \( \Omega \subseteq \mathbb{R}^{\dim} \), \( \dim \in \{1, 2, 3\} \) is assumed to be a smooth bounded domain. The forward operator

\[ F : D(F) \subseteq X \to Y \]

(79)

and its derivative as well as the Banach space adjoint can be written as

\[
\begin{align*}
F(c) &= A(c)^{-1} f, \\
F'(c)h &= -A(c)^{-1} (h \cdot F(c)), \\
F'(c)^* w &= -F(c) \cdot (A(c)^{-1} w),
\end{align*}
\]

with

\[
A(c) : H^2(\Omega) \cap H^1_0(\Omega) \to L^2(\Omega) \text{g} \\
u \to -\Delta u + cu.
\]

It was shown in [18] that for

\[
X = L^P(\Omega), \quad Y = L^R(\Omega)
\]

(80)

with

\[
P \in (1, \infty), \quad P \geq \frac{\dim}{2}, \quad R > \frac{P}{P-1}, \quad R \geq \frac{2\dim P}{\dim P + 2P - 2\dim}
\]

(81)

the assumptions on \( F \) in theorem 1 and with

\[
P \in (1, \infty), \quad R \in [2, \infty], \quad \frac{2R}{R-2} \leq P
\]

(82)

the assumptions on \( F \) in theorem 2 are satisfied. Here, the domain of \( F \) is set to

\[
D(F) = \{ c \in L^P(\Omega) \mid \exists \tilde{c} \in L^\infty(\Omega), \tilde{c} \geq 0 \text{ a.e.: } \|c - \tilde{c}\|_{L^P(\Omega)} \leq \gamma \},
\]

(83)
where \( \tilde{\gamma} \leq \min\{1/\|\hat{\gamma}\|_{H^1_0(\Omega)} \rightarrow L^{2\gamma/(\gamma-1)}(\Omega), 1/\|\hat{\gamma}\|_{H^2(\Omega) \cap H^1_0(\Omega)} \rightarrow L^P/(P-1)(\Omega) \} \) for some
\[
k \in [\tilde{a}, \tilde{b}] \cap (1, \infty)
\]
with
\[
\tilde{a} = \max\{2\dim/(\dim + 2), \dim P/(\dim + 2P)\},
\]
\[
\tilde{b} = \min\{P, 2\dim/\max\{0, \dim - 2\}, R, PR/(P + R)\}
\]
and \((k < P \land R < \infty)\) or \(k > \dim/2\)
\[(84)\]
in the first case (81), and to
\[
D(F) = \{c \in L^\infty(\Omega) \mid \hat{\gamma} \geq c \geq 0 \ \text{a.e.}\}
\]
\[(85)\]
for some \(\hat{\gamma} > 0\) in the second case (82).

Therewith the benchmark source condition (5) is equivalent to
\[
w = -\|\hat{\gamma}\|_{H^1_0(\Omega)} \rightarrow L^{2\gamma/(\gamma-1)}(\Omega).
\]
Choosing \(P\) as small as possible and \(R\) as large as possible corresponds to formulating the inverse problem as weakly ill-posed as possible and therewith obviously also to making the source condition (86) as weak as possible. Note that indeed the noise level is in practice often given in the \(L^\infty\) norm. Under conditions (81), we might, e.g., set
\[
R = \infty, \quad P = \dim/2 + \varepsilon, \quad k := \max\{2\dim/(\dim + 2), \dim\}\]
for \(\varepsilon > 0\) arbitrarily small, and under conditions (82)
\[
R = \infty, \quad P = 2.
\]
This allows for a relaxation as compared to the Hilbert space case \(P = R = 2\).

In the second example we deal with the identification of the space-dependent coefficient \(a\) in
\[
-\nabla(a \nabla u) = f \quad \text{in } \Omega
\]
\[
u = 0 \quad \text{on } \partial \Omega
\]
\[(87)\]
\[(88)\]
from measurements of \(u\), where again \(\Omega \subseteq \mathbb{R}^{\dim}\), \(\dim \in \{1, 2, 3\}\) is assumed to be a smooth bounded domain. Using the differential operator
\[
A(a) : H^2(\Omega) \cap H^1_0(\Omega) \rightarrow L^2(\Omega)
\]
\[
u \rightarrow -\nabla(a \nabla \nu)
\]
we can write the forward operator, its derivative, as well as the Banach space adjoint as
\[
F(a) = A(a)^{-1} f, \quad F'(a)h = A(a)^{-1}(\nabla(h \nabla F(a))), \quad F'(a)^* w = -\nabla F(a) \cdot \nabla(A(a)^{-1} w).
\]
It has been shown in [18] that with
\[
D(F) = \{a \in X \mid a \geq \alpha\}
\]
\[(89)\]
with \(\alpha > 0\),
\[
X = W^{1,Q}(\Omega), \quad Y = L^R(\Omega)
\]
\[(90)\]
under conditions
\[
Q > \dim, \quad Q \in (1, \infty), \quad Q \geq \frac{R}{R-1},
\]
\[
R \leq \frac{2\dim}{\max\{0, \dim - 2\}} \quad \text{and} \quad (R < \infty \lor \dim < 2),
\]
the assumptions on \(F\) in propositions 2, 3, theorems 1, 2, and corollaries 1, 2 are satisfied.
For this example, the benchmark source condition (5) is equivalent to
\[ \exists w \in Y^* = L^{R/(R-1)}(\Omega) : -\nabla F(a^\dagger) \cdot \nabla (A(a^\dagger)^{-1} w) = \| e^0 \|_{W^{1,R}}^p - \Omega \cdot (\nabla e^0)^{Q-2} \nabla e^0 + |e^0|^{Q-1} \operatorname{sgn}(e^0), \]
(91)
as well as
\[ \frac{\partial e^0}{\partial n} = 0 \quad \text{on } \partial \Omega, \]
(92)
for
\[ e^0 = a^\dagger - a_0, \]
which amounts to a transport equation for \( A(a^\dagger)^{-1} w \). In the 1D case \( \Omega = (0, L) \), condition (91) becomes
\[ w = -\| e^0 \|_{W^{1,2}}^2 A(a^\dagger) \left( \int_0^1 \frac{-(e^0_1)^{(Q-2)\epsilon_1} + |e^0|^Q \operatorname{sgn}(e^0)}{F(a^\dagger)} \, dx \right) \]
\[ \in Y^* = L^{R/(R-1)}(\Omega); \]
(93)
hence, the benchmark source condition is satisfied if \( F(a^\dagger) \) is bounded away from zero as well as
\[ e^0(0) = e^0(L) = 0 \quad \text{and} \quad e^0 \in W^{3,R/(R-1)}(\Omega). \]
(94)
Here we may, e.g. for arbitrarily small \( \varepsilon > 0 \), set
\[ R = \infty, \quad Q = 1 + \varepsilon \quad \text{if dim} = 1, \]
\[ R = \frac{1}{\bar{\varepsilon}}, \quad Q = 2 + \varepsilon \quad \text{if dim} = 2, \]
(95)
(96)
with \( \bar{\varepsilon} \in (0, 1 - 1/(2 + \varepsilon)] \) arbitrarily small
\[ R = 6, \quad Q = 3 + \varepsilon \quad \text{if dim} = 3. \]
(97)
In the case dim = 1, (94) can be directly compared to the Hilbert space situation \( Q = R = 2 \), see, e.g., [5], and with (95) yields an obvious relaxation. Note that in the higher dimensional case, the Hilbert space setting requires a higher order Sobolev space, namely \( H^s(\Omega) \) with \( s \geq 1 + \text{dim}/2 - \text{dim}/Q \) so that \( H^s(\Omega) \) is continuously embedded in \( W^{1,Q}(\Omega) \). The Hilbert space benchmark source condition with \( s = 2 \) therefore becomes
\[ -\nabla F(a^\dagger) \cdot \nabla (A(a^\dagger)^{-1} w) = (\Delta^2 e^0 - \Delta e^0 + e^0) \quad \text{and} \quad \frac{\partial e^0}{\partial n} = \Delta e^0 = 0 \quad \text{on } \partial \Omega, \]
(98)
(99)
(100)
(101)
(102)
(103)
(104)
where we have used (76) with \( p = P = 2 \), which is obviously stronger than (91), (92) with (96) or (97), since it requires more knowledge on the boundary values of \( a^\dagger \) as well as a higher order of differentiability.

Implementation of the IRGNM in Banach space requires numerical solution of the minimization problem (3) with a linear operator \( T_k \) in each step. If we do so, e.g., by one of the gradient-type methods devised in [2], we have to apply \( T_k \) as well as its Banach space adjoint (which amounts to solving a linear PDE in our parameter identification examples) and the duality mappings \( J_p, J_r, J_{p/(p-1)} = J_p^{-1} \) in each inner iteration. While \( J_p, J_r \) only involve multiplication (and in example (87), (88) by (75) also differentiation), application of \( J_p^{-1} \) in example (87), (88) amounts to solving a PDE with the differential operator given by (75), which is even nonlinear unless \( P = p = 2 \).
6. Conclusions and remarks

In this paper, we provide convergence rate results for the IRGNM under approximate source conditions with general index functions including Hölder and logarithmic rates. Both a priori and a posteriori parameter choice strategies are studied.

Possible future research will be on the case of enhanced source conditions corresponding to \( v \in (1, 2] \) (cf \([19, 20]\) for Tikhonov regularization in Banach space). Moreover, different regularization terms in place of \( \|x - x_0\|^p \) are of interest. Especially sparsity enhancing terms like the \( L^1 \) norm are not covered by the theory of this paper, since \( L^1(\Omega) \) is not a uniformly convex space. For this purpose, new ideas will have to be developed and first of all well definedness and convergence without rates will have to be proven (see \([18]\) for the case of uniformly convex spaces). Like, e.g., in \([1]\) and \([17]\), one might also think of using a general regularization method (in place of Tikhonov) in each Newton step (e.g. the Landweber iteration from \([24]\)).

Acknowledgments

The authors would like to thank Austrian Academy of Sciences and the German Research Foundation (DFG) for financial support under grant HO1454/7-2 as well as within the Cluster of Excellence in Simulation Technology (EXC 310/1) at the University of Stuttgart. Moreover, useful comments by the referees are gratefully acknowledged.

References

[8] Hanner O 1956 On the uniform convexity of \( L^p \) and \( F^p \) Ark. Mat. 3 239–44


