www.oeaw.ac.at

Johann Radon Institute for
Computational and Applied Mathematics
Austrian Academy of Sciences (ÖAW)

Convergence of projected iterative regularization methods for nonlinear problems with smooth solutions

B. Kaltenbacher, A. Neubauer

RICAM-Report 2006-06

www.ricam.oeaw.ac.at

Convergence of projected iterative regularization methods for nonlinear problems with smooth solutions

B. Kaltenbacher‡

Department of Sensor Technology, University of Erlangen, Germany

E-mail: `barbara.kaltenbacher@lse.eei.uni-erlangen.de`

A. Neubauer

Industrial Mathematics Institute, University of Linz, Austria

E-mail: `neubauer@indmath.uni-linz.ac.at`

Abstract. This paper is concerned with two aspects in the convergence analysis of regularization methods for nonlinear problems.

Firstly, if a solution of the inverse problem (or its difference to an inital guess) is sufficiently smooth, the conditions on the nonlinearity of the forward operator become rather weak. Here we prove a convergence result for general regularized Newton methods that – among others – includes regularization by discretization.

Secondly, constraints given by a closed and convex set can be realized by projection onto this set within each step. It is shown that also for this projected version of the regularized Newton method, convergence and optimal convergence rates are obtained.

‡ Supported by the Austrian Academy of Sciences within the Radon Institute for Computational and Applied Mathematics, as well as the German Science Foundation DFG under grant Ka 1778/1

1. Introduction

Consider the nonlinear ill-posed operator equation

$$F(x) = y\,, \tag{1}$$

where $F : \mathcal{D}(F) \to \mathcal{Y}$ with domain $\mathcal{D}(F) \subset \mathcal{X}$. Here, $\mathcal{D}(F)$ is not necessarily the maximal domain of definition, but rather a set of interest whose definition might include certain features of the solution one is interested in. Throughout this paper we will assume that $\mathcal{D}(F)$ is closed and convex.

We restrict our attention to Hilbert spaces $\mathcal{X}$ and $\mathcal{Y}$ with inner products $\langle\,\cdot\,,\cdot\,\rangle$ and norms $\|\cdot\|$, respectively; they can always be identified from the context in which they appear. Moreover, we assume that we have noisy data y^δ with

$$\|y^\delta - y\| \leq \delta\,, \tag{2}$$

and that for exact data y, a solution $x^\dagger$ of (1) exists in $\mathcal{D}(F)$.

In this paper, we will especially focus on the case that an exact solution $x^\dagger \in \mathcal{D}(F)$ of (1) is sufficiently smooth, i.e., a source condition

$$x^\dagger - x_0 = (F'(x^\dagger)^* F'(x^\dagger))^\mu v\,, \qquad \text{for some } v \in \mathcal{N}(F'(x^\dagger))^\perp\,, \tag{3}$$

with $\mu \geq 1/2$ is satisfied for an initial guess $x_0 \in \mathcal{D}(F) \cap \mathcal{B}_\rho(x^\dagger)$. In the absence of such a source condition, strong structural asumptions on the operator F have to be made that restrict its nonlinearity in order to obtain convergence results for regularization methods applied to (1), (cf. [4, 5] as well as [15] and the references therein). However, it has been shown in several recent publications (cf., e.g., [2, 3, 7, 15, 17]), that such a sufficiently strong source condition enables a convergence analysis under quite general conditions on F. More precisely, if (3) with $\mu \geq 1/2$ holds, usually a Lipschitz condition on the Fréchet derivative of F, i.e.,

$$\|F'(\tilde{x}) - F'(x)\| \leq L\|\tilde{x} - x\|\,, \qquad x, \tilde{x} \in \mathcal{B}_{2\rho}(x_0) \cap \mathcal{D}(F)\,, \tag{4}$$

suffices. In this context, we first of all wish to refer to results for convergence rates of Tikhonov regularization [7] where optimal convergence rates

$$\|x_\alpha^\delta - x^\dagger\| = O(\delta^{\frac{2\mu}{2\mu+1}}) \tag{5}$$

are established for the Tikhonov regularized solution x_α^δ under a source condition (3) with $\mu = 1/2$. Under the same conditions, optimal convergence rates have been established for modified Landweber [15] or Tikhonov gradient [17] methods. Moreover, we wish to especially refer to the results on regularized Newton methods in [2], which we will discuss in more detail in the last section.

We do not intend to palliate the fact that source conditions (3) with $\mu \geq 1/2$, depending on the problem under consideration, might impose a severe restriction on the exact solution. This is especially the case for exponentially ill-posed problems, where Hölder type source conditions (3) with $\mu > 0$ often imply that the initial error is analytic. Still, there is a considerable range of practically relevant inverse problems, where on the one hand any of the usual nonlinearity restrictions made in case of $\mu < 1/2$ are not met and which on the other hand are only moderately ill-posed, so that (3) with $\mu \geq 1/2$ is satisfied under finite differentialbility assumptions on the initial error. Among these, we wish to mention some parameter identification problems for nonlinear PDEs from boundary measurements arising in the context of material characterization (cf., e.g., [13, 14, 16]). For example, we refer to [16] for a detailed and complete verification of all convergence conditions (especially also (4)) and an interpretation of the source condition (3) with $\mu = 1/2$, that amounts to fourth order differentiability and boundary conditions for the searched for parameter.

The central task of this paper is to provide a convergence proof for general Newton type regularization methods under the conditions (3) with $\mu \geq 1/2$ and (4).

Additionally to that, we consider constraints described by a closed and convex set and prove convergence of regularized Newton iterates, when feasibility is preserved by projection onto this closed and convex set.

This enables to weaken the topology in $\mathcal{X}$ and therewith the source condition (3), as illustrated in Section 3 by means of an example, for which we also show some numerical experiments.

2. Regularized Newton methods

Newton's method

$$x_{k+1} = x_k + F'(x_k)^{-1}(y^\delta - F(x_k))$$

or Newton type methods are known to provide fast local convergence for well-posed problems. In the ill-posed situation considered here, computation of the Newton step is in general unstable, though, and therefore has to be regularized. We do so by making the replacement

$$R_\alpha(F'(x)) \approx F'(x)^\dagger \,,$$

where R_α is a regularizing operator. More precisely, $\alpha > 0$ is a small regularization parameter, and R_α satisfies

$$R_\alpha(K)y \to K^\dagger y \quad \text{as} \quad \alpha \to 0 \quad \text{for all} \quad y \in \mathcal{R}(K) \,, \tag{6}$$

$$\|R_\alpha(K)\| \leq \frac{c_1}{\sqrt{\alpha}} \,, \tag{7}$$

$$\|R_\alpha(K)K\| \leq c_2 \tag{8}$$

$$\text{for all} \quad K \in \mathcal{L}(\mathcal{X},\mathcal{Y}) \quad \text{with} \quad \|K\| \leq c_s \,,$$

for some positive constants c_1, c_2, and c_s. A possibility of defining regularization methods satisfying (6), (7) is given via spectral theory (cf., e.g., [8]) by choosing a piecewise continuous real function $g_\alpha : [0,\overline{\lambda}] \to \mathbb{R}$, $\overline{\lambda} > 0$, satisfying

$$\begin{aligned} &g_\alpha(\lambda) \to \lambda^{-1} \quad \text{as} \quad \alpha \to 0 \quad \text{for all} \quad \lambda \in (0,\overline{\lambda}] \,, \\ &\sup_{\lambda \in [0,\overline{\lambda}]} |\lambda g_\alpha(\lambda)| \leq c_g \,, \qquad \text{and} \qquad \sup_{\lambda \in [0,\overline{\lambda}]} |g_\alpha(\lambda)| \leq c(\alpha) \,, \end{aligned} \tag{9}$$

for some positive constant c_g and some positive function $c(\alpha)$, and by setting

$$R_\alpha(K) := g_\alpha(K^*K)K^* \,. \tag{10}$$

Within the so-defined class, many well-known regularization methods such as Tikhonov regularization, Landweber iteration, the ν-methods and Lardy's method can be found. Note, however, that the slightly more general concept (6) – (8) additionally includes other methods such as regularization by discretization, which we also want to consider below.

Using a monotonically decreasing sequence $\alpha_k \searrow 0$ of regularization parameters, one arrives at a class of regularized Newton methods

$$x_{k+1}^\delta = x_0 + R_{\alpha_k}(F'(x_k^\delta))(y^\delta - F(x_k^\delta) - F'(x_k^\delta)(x_0 - x_k^\delta)) \,, \tag{11}$$

see [2, 15]. Here, y^δ are noisy data satisfying the estimate (2) and the superscript δ is omitted in the noise free case. Note that in the limiting case $\alpha_k \to 0$, i.e., $R_{\alpha_k}(F'(x)) \to F'(x)^\dagger$, this formulation is equivalent to the usual Newton method. The special choice

$$R_{\alpha_k}(F'(x_k^\delta)) = (F'(x_k^\delta)^* F'(x_k^\delta) + \alpha_k I)^{-1} F'(x_k^\delta)^* \tag{12}$$

corresponds to the iteratively regularized Gauss-Newton method (IRGNM) proposed and analyzed by Bakushinskii, see [1] as well as [15]. To maintain stability, in [1] and also here a growth restriction

$$\alpha_k > 0 \,, \qquad 1 \leq \frac{\alpha_k}{\alpha_{k+1}} \leq r \,, \qquad \lim_{k\to\infty} \alpha_k = 0 \,, \tag{13}$$

for some $r > 1$ on the regularization parameters is imposed, which, e.g., allows for a geomentrically decaying sequence $\alpha_k = \alpha_0 q^k$ for some $q \in (0,1)$. Note that this method is related to the well-known Levenberg-Marquardt method, see [9], where the regularization parameter choice as well as the convergence analysis differ essentially, though. As a matter of fact, we do not expect that the assertions we make in this paper can be carried over to the Levenberg-Marquardt method in a straightforward manner.

As opposed to previous publications on regularized Newton methods, we do not assume that the domain $\mathcal{D}(F)$ has nonempty interior containing the exact solution (so that one can replace $\mathcal{D}(F)$ in the convergence analysis by a ball of sufficiently small radius around the solution). Rather, we consider the situation that closeness of some iterate x_k^δ to the exact solution is not sufficient for well-definedness of $F(x_k^\delta)$, or that we search for a solution satisfying additional constraints. In this situation we eventually have to project our iterates into the domain of F in order to be able to carry out the next step or, more generally, stay in the feasible set. For this purpose, we use the (possibly nonlinear) metric projector $P_{\mathcal{D}(F)}$ onto $\mathcal{D}(F)$ to arrive at

$$\begin{aligned} \tilde{x}_{k+1}^\delta &= x_0 + R_{\alpha_k}(F'(x_k^\delta))(y^\delta - F(x_k^\delta) - F'(x_k^\delta)(x_0 - x_k^\delta)) \\ x_{k+1}^\delta &= P_{\mathcal{D}(F)}\left(\tilde{x}_{k+1}^\delta\right) . \end{aligned} \tag{14}$$

Note that if $\tilde{x}_{k+1}^\delta$ lies within $\mathcal{D}(F)$ – which will be always the case if $\mathcal{D}(F)$ has nonempty interior containing the solution $x^\dagger$ and one concentrates on a sufficiently small ball around $x^\dagger$ – then the projection can be omitted and the respective step is equivalent to (11).

Since $\mathcal{D}(F)$ no longer has to have a nonempty interior, we also relax the notion of differentiability: we assume in the following that for all $x \in \mathcal{B}_{2\rho}(x_0) \cap \mathcal{D}(F)$ there exists $F'(x) \in \mathcal{L}(\mathcal{X}, \mathcal{Y})$ with

$$\lim_{t \to 0} \frac{F(x+th) - F(x)}{t} = F'(x)h$$

for all directions $h \in \mathcal{X}$ where $x + th \in \mathcal{B}_{2\rho}(x_0) \cap \mathcal{D}(F)$ for some $t > 0$. This together with (4) guarantees that

$$\|F(x+h) - F(x) - F'(x)h\| \leq \frac{L}{2} \|h\|^2 . \tag{15}$$

This means F has to be Fréchet-differentiable on the subset $\mathcal{B}_{2\rho}(x_0) \cap \mathcal{D}(F)$.

For proving convergence rates for generalized Newton type methods (11) under source conditions (3), usually an assumption of the type

$$\begin{aligned} &\|(I - R_\alpha(K)K)(K^*K)^\mu\| \leq c_\mu \alpha^\mu \\ &\text{for all} \quad K \in \mathcal{L}(\mathcal{X}, \mathcal{Y}) \quad \text{with} \quad \|K\| \leq c_s \end{aligned} \tag{16}$$

is made on the regularization operators R_α, in analogy to the linear case.

To make all these methods well defined, we assume that the derivatives F' are uniformly bounded in a neighbourhood of x_0. This uniform bound has to be such that applicability of the respective regularization method can be guaranteed. Therefore, we assume that

$$\|F'(x)\| \leq c_s \qquad \text{for all } x \in \mathcal{B}_{2\rho}(x_0) \cap \mathcal{D}(F) , \tag{17}$$

with c_s as in (6). This can always be achieved by proper scaling.

To show convergence in the situation of noisy data, one has to use an appropriate stopping rule. In the general case, convergence can be achieved if the stopping index $k_* = k_*(\delta)$ is chosen such that

$$k_* \to \infty \qquad \text{and} \qquad \eta \geq \delta \alpha_{k_*}^{-\frac{1}{2}} \to 0 \text{ as } \delta \to 0 .$$

If additional source conditions hold, an appropriate a priori choice is

$$\eta \alpha_{k_*}^{\mu + \frac{1}{2}} \leq \delta < \eta \alpha_k^{\mu + \frac{1}{2}} , \qquad 0 \leq k < k_* , \tag{18}$$

In classical convergence proofs of Newton's method for well-posed problems, a local Lipschitz condition (4) on F' is used implying a Taylor remainder estimate that is quadratic in terms of the difference between x and $\tilde{x}$. It can be shown (cf. [15]) that, unless a source condition (3) with $\mu \geq 1/2$ is satisfied, stronger assumptions on F are required in a convergence analysis. On the other hand, we will here prove that if (3) holds with $\mu \geq 1/2$, then convergence and convergence rates can be established under a Lipschitz condition (4) without any further assumptions on the operator F. For this purpose, we make heavy use of the following lemma, which might be of interest on its own.

Lemma 1 *Let $\mathcal{X}$ and $\mathcal{Y}$ be Hilbert spaces and let $A, B \in \mathcal{L}(\mathcal{X}, \mathcal{Y})$. Then the estimate*

$$\|(A^*A)^\mu - (B^*B)^\mu\| \le 2\mu\tilde{c}_\mu \|A - B\| \max\{\|A\|, \|B\|\}^{2\mu-1}$$

holds for all $\mu > 1/2$, where $\tilde{c}_\mu$ is a constant satisfying $1 \le \tilde{c}_\mu \le 2$ for $\mu \ge 1$, $\lim_{\mu\to\infty} \tilde{c}_\mu = 1$, and $\lim_{\mu\to\frac{1}{2}} \tilde{c}_\mu = \infty$.

Proof.
Proceeding as in the proof of Lemma 5.7 in [8] we use the formula

$$(A^*A)^\mu - (B^*B)^\mu = \frac{\sin\mu\pi}{\pi} \int_0^\infty t^\mu [(B^*B + tI)^{-1} - (A^*A + tI)^{-1}]\, dt \quad (19)$$

which is valid for $0 < \mu < 1$, and the estimate

$$\begin{aligned}
&\|(B^*B + tI)^{-1} - (A^*A + tI)^{-1}\| \\
&\quad = \|(B^*B + tI)^{-1}(A^*A - B^*B)(A^*A + tI)^{-1}\| \\
&\quad = \|(B^*B + tI)^{-1}(B^*(A - B) + (A^* - B^*)A)(A^*A + tI)^{-1}\| \\
&\quad \le \begin{cases} t^{-\frac{3}{2}} \|A - B\| \\ t^{-2} \|A^*A - B^*B\| \end{cases}
\end{aligned}$$

to obtain for $1/2 < \mu < 1$ that

$$\begin{aligned}
&\|(A^*A)^\mu - (B^*B)^\mu\| \\
&\le \frac{\sin\mu\pi}{\pi}\Big[\int_0^a t^{\mu-\frac{3}{2}} \|A - B\|\, dt + \int_a^\infty t^{\mu-2} \|A^*A - B^*B\|\, dt\Big] \\
&\le \frac{\sin\mu\pi}{\pi}\Big[\frac{2}{2\mu - 1} a^{\mu-\frac{1}{2}} \|A - B\| + \frac{1}{1-\mu} a^{\mu-1} \|A^*A - B^*B\|\Big]
\end{aligned}$$

and further with the choice

$$a = \left(\frac{\|A^*A - B^*B\|}{\|A - B\|}\right)^2$$

that

$$\|(A^*A)^\mu - (B^*B)^\mu\| \le \frac{\sin\mu\pi}{\pi(2\mu-1)(1-\mu)} \|A - B\|^{2(1-\mu)} \|A^*A - B^*B\|^{2\mu-1}. \quad (20)$$

Obviously for $n \in \mathbb{N}$ we obtain that

$$\begin{aligned}
\|(A^*A)^n - (B^*B)^n\| \le\ & \|B^*B\|\, \|(A^*A)^{n-1} - (B^*B)^{n-1}\| \\
&+ \|A^*A - B^*B\|\, \|(A^*A)^{n-1}\|
\end{aligned}$$

and hence by induction that

$$\|(A^*A)^n - (B^*B)^n\| \le \|A^*A - B^*B\| \sum_{i=0}^{n-1} \|A\|^{2i} \|B\|^{2(n-1-i)}. \quad (21)$$

Now let us assume that $\mu = n + \rho$ for some $n \in \mathbb{N}$ and $0 < \rho < 1$. Then we obtain similarly as above that

$$\begin{aligned}
\|(A^*A)^\mu - (B^*B)^\mu\| \le\ & \|B\|^{2\rho} \|(A^*A)^n - (B^*B)^n\| \\
&+ \|((A^*A)^\rho - (B^*B)^\rho)A^*A\|\, \|A\|^{2(n-1)}.
\end{aligned}$$

The second term is estimated using again formula (19), i.e.,

$$\begin{aligned}
&\|((A^*A)^\rho - (B^*B)^\rho)A^*A\| \\
&\quad\le \frac{\sin\rho\pi}{\pi}\int_0^\infty t^\rho\,\|(B^*B+tI)^{-1}(A^*A-B^*B)(A^*A+tI)^{-1}A^*A\|\,dt \\
&\quad\le \frac{\sin\rho\pi}{\pi}\Big[\int_0^{\|A\|^2} t^{\rho-1}\,\|A^*A-B^*B\|\,dt \\
&\qquad\qquad + \int_{\|A\|^2}^\infty t^{\rho-2}\,\|A^*A-B^*B\|\,\|A\|^2\,dt\Big] \\
&\quad\le \frac{\sin\rho\pi}{\pi\rho(1-\rho)}\,\|A^*A-B^*B\|\,\|A\|^{2\rho}\,.
\end{aligned}$$

Combining this together with (21) we obtain for $\mu \ge 1$ the estimate

$$\|(A^*A)^\mu - (B^*B)^\mu\| \le \mu\,\tilde c_\mu\,\|A^*A-B^*B\|\,\max\{\,\|A\|,\|B\|\,\}^{2(\mu-1)}\,, \tag{22}$$

where $1 \le \tilde c_\mu \le 2$ and $\lim_{\mu\to\infty}\tilde c_\mu = 1$.

Since $\|A^*A-B^*B\| \le \|A-B\|\,(\,\|A\|+\|B\|\,)$ and $2^{2\mu-1} \le 2\mu$ for $1/2 \le \mu \le 1$ the assertions now follow with (20) and (22). $\diamondsuit$

Remark 2 It is an immediate consequence of the proof of the lemma above that

$$\|(A^*A)^\mu - (B^*B)^\mu\| \le \mu\tilde c_\mu\,\|A-B\|^{\min\{2,2\mu\}}\max\{\,\|A\|,\|B\|\,\}^{\max\{0,2(\mu-1)\}}$$

if $\|A^*A-B^*B\| = \|A-B\|^2$, which for instance holds if $B = QA$ and Q is an orthogonal projector. It follows as in the proof of Lemma 5.7 in [8] that $\tilde c_\mu$ can be then chosen such that it is also bounded for $\mu \to 1/2$.

The following lemma is useful for concluding convergence rates from error estimates to be derived in the proof of Theorem 4 below.

Lemma 3 *Let $\{a_k^\delta\}_{k\in\mathbb{N}_0}$, $\delta \ge 0$, be a family of sequences satisfying*

$$0 \le a_k^\delta \le a \qquad \textit{and} \qquad \limsup_{\delta\to 0,\,k\to\infty} a_k^\delta \le a^0$$

for some $a, a^0 \in \mathbb{R}_0^+$. Moreover, let γ_k^δ be such that

$$0 \le \gamma_{k+1}^\delta \le a_k^\delta + b\gamma_k^\delta + c(\gamma_k^\delta)^2\,, \qquad 0 \le k < k_*(\delta)\,, \qquad \gamma_0^\delta := \gamma_0 \tag{23}$$

holds, where b, c, γ_0 are nonnegative constants, $k_*(\delta) \in \mathbb{N}_0$ for any $\delta > 0$, and $k_*(\delta) \to k_*(0) := \infty$ as $\delta \to 0$.

If $c > 0$, $b + 2\sqrt{ac} \le 1$, and $\gamma_0 \le \overline{\gamma}(a)$, then

$$\gamma_k^\delta \le \max\{\gamma_0, \underline{\gamma}(a)\}\,, \qquad 0 \le k \le k_*(\delta)\,.$$

If in addition $a^0 < a$, then

$$\limsup_{\delta\to 0}\gamma_{k_*(\delta)}^\delta \le \underline{\gamma}(a^0)\,, \qquad \limsup_{k\to\infty}\gamma_k^0 \le \underline{\gamma}(a^0)\,. \tag{24}$$

Here $\underline{\gamma}(p)$ and $\overline{\gamma}(p)$, where $p \in [0,a]$, denote the fixpoints of the equation $p+b\gamma+c\gamma^2 = \gamma$, i.e.,

$$\underline{\gamma}(p) := \frac{2p}{1-b+\sqrt{(1-b)^2-4pc}}\,, \qquad \overline{\gamma}(p) := \frac{1-b+\sqrt{(1-b)^2-4pc}}{2c}\,. \tag{25}$$

Proof.
See [15, Lemma 4.11]. ♢

Theorem 4 *Let* (2), (4), (6) – (8), (13), *and* (17) *hold and let* x_k^δ *be defined by the sequence* (14). *Moreover, let the source condition* (3) *and condition* (16) *hold with* $\mu \geq 1/2$.

If $k_* = k_*(\delta)$ *is chosen according to the stopping rule* (18) *and if* $\|v\|$ *and* η *are sufficiently small, then we obtain the convergence rates*

$$\|x_{k_*}^\delta - x^\dagger\| = O\Big(\delta^{\frac{2\mu}{2\mu+1}}\Big)$$

in case of noisy data and

$$\|x_k - x^\dagger\| = O(\alpha_k^\mu)$$

in the noise free case ($\delta = 0, \eta = 0$).

Proof.
To derive a recursive error estimate, we assume that the current iterate x_k^δ is in $\mathcal{B}_\rho(x^\dagger)$ and that $k < k_*$ (note that $k_* = \infty$ if $\delta = 0$). This guarantees that $x^\dagger, x_k^\delta \in \mathcal{B}_{2\rho}(x_0)$. Denoting $K_k := F'(x_k^\delta)$, $K := F'(x^\dagger)$, and using (3), we can decompose the error before projection as follows:

$$\begin{aligned}\tilde{x}_{k+1}^\delta - x^\dagger = {} & (I - R_{\alpha_k}(K_k)K_k)(K_k^*K_k)^\mu v \\ & + (I - R_{\alpha_k}(K_k)K_k)\Big((K^*K)^\mu - (K_k^*K_k)^\mu\Big)v \\ & - R_{\alpha_k}(K_k)(F(x_k^\delta) - F(x^\dagger) - K_k(x_k^\delta - x^\dagger)) \\ & - R_{\alpha_k}(K_k)(y - y^\delta)\end{aligned} \tag{26}$$

By Lemma 1, (2), (4), (7), (8), (15) – (17), and the fact that the metric projection is Lipschitz contiouous with Lipschitz constant 1, we get

$$\begin{aligned}\|x_{k+1}^\delta - x^\dagger\| = {} & \|P_{\mathcal{D}(F)}(\tilde{x}_{k+1}^\delta) - P_{\mathcal{D}(F)}(x^\dagger)\| \;\leq\; \|\tilde{x}_{k+1}^\delta - x^\dagger\| \\ & \leq c_\mu \alpha_k^\mu \|v\| + 2(1+c_2)\mu\tilde{c}_\mu c_s^{2\mu-1} L \|x_k^\delta - x^\dagger\|\,\|v\| \\ & \quad + \frac{c_1}{\sqrt{\alpha_k}}\left(\frac{L}{2}\|x_k^\delta - x^\dagger\|^2 + \delta\right)\end{aligned} \tag{27}$$

in case $\mu > 1/2$. If $\mu = 1/2$ in (3), then this is equivalent (see [8, Proposition 2.18]) to the existence of an element $w \in \mathcal{Y}$ with $\|w\| = \|v\|$ such that

$$x^\dagger - x_0 = F'(x^\dagger)^* w\,,$$

so that we can replace the terms $(K_k^*K_k)^\mu v$ and $\Big((K^*K)^\mu - (K_k^*K_k)^\mu\Big)v$ in the error decomposition (26) by $K_k^* w$ and $(K^* - K_k^*)w$, respectively, to obtain an estimate similar to (27).

Setting $\gamma_k^\delta := \alpha_k^{-\mu}\|x_k^\delta - x^\dagger\|$ we obtain together with (13), (18), and (27) that

$$\gamma_{k+1}^\delta \leq a + b\gamma_k^\delta + c(\gamma_k^\delta)^2$$

with

$$a := r^\mu (c_\mu \|v\| + c_1 \eta)\,,$$
$$b := 2r^\mu (1+c_2)\mu \tilde{c}_\mu c_s^{2\mu-1} L \|v\|\,,$$
$$c := r^\mu c_1 \frac{L}{2} \alpha_0^{\mu-\frac{1}{2}}\,.$$

If $\|v\|$ and η are sufficiently small, then Lemma 3 is applicable and we obtain that γ_k^δ is bounded and that $\|x_{k+1}^\delta - x^\dagger\| \le \alpha_0^\mu \max\{\gamma_0, \underline{\gamma}(a)\} \le \rho$. This together with (18) yields the assertion. ♢

Remark 5 The result above is a considerable improvement of Theorem 4.16 in [15] (in case Assumption 4.15 (iv) (c) holds) not only because of the fact that we use projected iterative regularization methods to allow $\mathcal{D}(F)$ with nonempty interior but since we do not need a rather strong assumption on the regularization operator R_α (see [15, (4.85)]).

It is obvious from the proof of Theorem 4.16 in [15] that all the other convergence and convergence rates results remain valid for projected iterative regularization methods if only weaker source conditions are satisfied, i.e., $\mu < 1/2$ in (3). Then, however, stronger assumptions on the operator F and on the regularization operator R_α are needed, e.g.,

$$\begin{aligned}
&F'(\tilde{x}) = F'(x) R(\tilde{x}, x) \quad \text{and} \quad \|I - R(\tilde{x}, x)\| \le c_R \|\tilde{x} - x\| \\
&x, \tilde{x} \in \mathcal{B}_{2\rho}(x_0) \cap \mathcal{D}(F)\,, \quad c_R > 0\,, \\
&\|R_\alpha(KR)KR - R_\alpha(K)K\| \le c_3 \|I - R\| \\
&\text{for all} \quad K \in \mathcal{L}(\mathcal{X}, \mathcal{Y}) \quad \text{with} \quad \|K\| \le c_s \quad \text{and} \\
&R \in \mathcal{L}(\mathcal{X}, \mathcal{X}) \quad \text{with} \quad \|KR\| \le c_s \quad \text{and} \quad \|I - R\| \le c_I < 1\,.
\end{aligned}$$

The latter condition can be verified, e.g., for Tikhonov regularization, iterated Tikhonov regularization, or Landweber iteration, cf. [15].

As already mentioned, the methodology proposed here can also be applied to regularization methods outside the class of methods defined via (10). Among those is regularization by projection, where the infinite-dimensional linear operator equation is projected to a finite-dimensional subspace $\mathcal{Y}_k$ of the data space $\mathcal{Y}$, where we assume that

$$\mathcal{Y}_1 \subset \mathcal{Y}_2 \subset \mathcal{Y}_3 \subset \ldots \subset \overline{\mathcal{R}(K)}\,, \qquad \overline{\bigcup_{k \in \mathbb{N}} \mathcal{Y}_k} = \overline{\mathcal{R}(K)}\,, \tag{28}$$

and is solved in a best approximate sense, so that with the superscript † denoting the generalized inverse,

$$R_\alpha(K) = (Q_k K)^\dagger Q_k = (Q_k K)^\dagger = P_k K^\dagger\,, \tag{29}$$

where Q_k and P_k are the orthogonal projectors onto $\mathcal{Y}_k$ and $\mathcal{X}_k := K^* \mathcal{Y}_k$, respectively. An advantage of regularization by discretization is that it enables the use of multigrid methods as optimal preconditioners for the fast convergence of iterative solutions of the linear system in each Newton step (cf. [11, 12, 14]).

Note that $\|R_\alpha(K)K\| = \|P_k\| = 1$ and that $P_k K^\dagger y \to K^\dagger y$ as $k \to \infty$ (cf. [8, Theorem 3.24]). Hence, (6) and (8) hold. Moreover, it holds that (cf. [8, Lemma 5.10])

$$\|(I - R_\alpha(K)K)(K^*K)^\mu\| = \|(I - P_k)(K^*K)^\mu\| \le \tfrac{4}{\pi}\|(I - Q_k)K\|^{2\mu} \quad (30)$$

for $\mu \in (0, 1]$.

Let us assume that the spaces $\mathcal{Y}_k$ have the property that

$$\|(I - Q_k)y\| \le \tilde{c}_1 h_k^p \|y\|_{\mathcal{Y}^p} \quad (31)$$

for all $y \in \mathcal{Y}^p \subset \mathcal{Y}$, where $p, \tilde{c}_1 > 0$. Such properties are for instance fulfilled for finite element spaces $\mathcal{Y}_k$ and Sobolev spaces $\mathcal{Y}^p$ (cf., e.g., Ciarlet [6]). The parameter h_k usually plays the role of a mesh size of the discretization, with $h_k \to 0$ as $k \to \infty$. If K has the smoothing property that

$$\mathcal{R}(K) \subset \mathcal{Y}^p\,, \quad (32)$$

then we obtain an estimate

$$\|(I - Q_k)K\| \le \tilde{c}_1 \|K\|_{\mathcal{X},\mathcal{Y}^p} h_k^p\,. \quad (33)$$

On the other hand, inverse inequalities (cf. [6] in the context of finite elements) yield an estimate for $(Q_kK)^\dagger$ in terms of the mesh size, i.e.,

$$\|(Q_kK)^\dagger\| \le \tilde{c}_2 h_k^{-\tilde{p}} \quad (34)$$

for some $\tilde{p}, \tilde{c}_2 > 0$ and all linear operators K satisfying

$$K \in \mathcal{L}(\mathcal{X}, \mathcal{Y}^p)\,, \qquad \|K\|_{\mathcal{X},\mathcal{Y}^p} \le c_s\,, \qquad \overline{\mathcal{R}(K)} = \mathcal{Y}\,. \quad (35)$$

With the correspondence

$$\alpha = \alpha_k := h_k^{2p} \quad (36)$$

for the regularization parameter, (30), (33), and (35) imply that (16) holds for $\mu \in (0, 1]$, where we additionallly restrict the opertors K to those satisfying (35). An inspection of the proof of Theorem 4 shows that the results remain valid, as long as (35) holds for all operators $K = F'(x)$ with $x \in \mathcal{D}(F) \cap \mathcal{B}_{2\rho}(x_0)$.

If the additional condition

$$p = \tilde{p} \quad (37)$$

holds, (34) and (36) imply that R_α defined by (29) satisfies (7).

By means of application examples it has been illustrated that (31), (32), (34), (35) and (37) are natural conditions in the context of parameter identification in PDEs and discretization by finite elements or splines (cf. [10, 11, 14]).

With these preliminaries we can apply Theorem 4 to obtain:

Corollary 6 *Let* (2), (4), (13), *and* (17) *hold and let* x_k^δ *be defined by the sequence* (14), *where the regularization method* R_α *is defined by* (29). *Moreover, we assume that* (33), (34), *and* (37) *hold for all operators as in* (35), *which we assume to hold for all* $K = F'(x)$ *with* $x \in \mathcal{D}(F) \cap \mathcal{B}_{2\rho}(x_0)$, *and that the source condition* (3) *holds with* $1/2 \le \mu \le 1$. *Then, the assertions of Theorem 4 hold.*

This improves the results of Corollary 4.18 in [15], since there rates were only shown for $\mu \le 1/2$. Note that by using smoother discretization spaces $\mathcal{Y}_k$ also higher convergence rates can be obtained, see [10, Section 4].

3. An Example

We consider a standard test example for parameter identification, namely the problem of identifying the space dependent diffusivity $a = a(x)$ in

$$\begin{aligned} \nabla \cdot (a \nabla u) &= f \quad \text{in } \Omega \\ u &= 0 \quad \text{on } \partial\Omega \end{aligned} \tag{38}$$

from measurements of u in the domain $\Omega \subseteq \mathbb{R}^n$. A solution to this problem exists if $a \in L^\infty(\Omega)$ with $a \ge \underline{\gamma} > 0$ a.e. and if $f \in H^{-1}(\Omega)$. For this purpose, usually the domain of the forward operator

$$\begin{aligned} F : \mathcal{X} &\to \mathcal{Y} = L^2(\Omega) \\ a &\mapsto u \quad \text{(solution of (38))} \end{aligned}$$

is chosen as the set of functions $\ge \underline{\gamma}$ in the Hilbert space

$$\mathcal{X} = H^{n/2+\varepsilon}$$

with $\varepsilon > 0$. Then one can not only verify Lipschitz continuity of the derivative (cf. (4)) but also more restrictive assumptions on the nonlinearity of the forward operator F (see, e.g., [15]).

The choice of $\mathcal{X}$ above restricts the set of parameters found by the regularization method to continuous ones. Unfortunately, the higher the smoothness order of the preimage space $\mathcal{X}$ is chosen, the more ill-posed the inverse problem becomes. Thus, by this choice instability is introduced in a somewhat artificial manner.

The approach proposed in this paper allows to work on the weak preimage space

$$\mathcal{X} = L^2(\Omega)$$

and to guarantee well-posedness of the forward problem by projecting onto the set

$$\mathcal{D}(F) := \{a \in L^\infty(\Omega) \, : \, \underline{\gamma} \le a \le \overline{\gamma} \text{ a.e. in } \Omega\}$$

for some constants $0 < \underline{\gamma} < \overline{\gamma}$. This means, we look for approximate solutions a_k^δ in L^∞ but consider their convergence to the exact solution $a^\dagger$ in L^2.

We will now verify all assumptions on the forward operator F made in the previous section for the one-dimensional case $\Omega = [0, 1]$: one can show that $F(a)$ is given by

$$\begin{aligned} F(a)(t) = {}& G(a)(t) - G(a)(1)(B(a)(1))^{-1} B(a)(t) \quad \text{with} \\ B(a)(t) := {}& \int_0^t a(s)^{-1}\, ds \qquad \text{and} \qquad G(a)(t) := \int_0^t a(s)^{-1} \int_0^s f(\sigma)\, d\sigma\, ds \end{aligned}$$

and that F is differentiable in the sense of the last section, namely

$$\begin{aligned} (F'(a)h)(t) = {}& (G'(a)h)(t) - G(a)(1)(B(a)(1))^{-1}(B'(a)(h))(t) \\ & - \; (B(a)(1))^{-1} B(a)(t) \Big((G'(a)h)(1) \end{aligned}$$

$$- G(a)(1)(B(a)(1))^{-1}(B'(a)(h))(1)\Big)$$

$$(G'(a)h)(t) = -\int_0^t a(s)^{-2}h(s)\int_0^s f(\sigma)\,d\sigma\,ds$$

$$(B'(a)h)(t) = -\int_0^t a(s)^{-2}h(s)\,ds$$

for all $a \in \mathcal{D}(F)$ and $h \in \mathcal{L}^\infty[0,1]$ such that $a + th \in \mathcal{D}(F)$ for some $t > 0$. Obviously $F'(a) \in \mathcal{L}(L^2[0,1], L^2[0,1])$. If $f \in L^1[0,1]$, then F' satisfies (4): for $a, b \in \mathcal{D}(F)$ it holds that

$$\|F'(a) - F'(b)\| \le c(\underline{\gamma}, \overline{\gamma})\,\|f\|_{L^1}\,\|a-b\|_{L^2}\,.$$

The adjoint of F' is given by

$$(F'(a)^*w)(s) = -a(s)^{-1}\dot{u}(a)(s)\Big(\int_s^1 w(t)\,dt - (B(a)(1))^{-1}\int_0^1 B(a)(t)w(t)\,dt\Big)\,,$$

where $u(a)$ is the solution of (38) and the dot denotes derivative with respect to s. The source condition (3) with $\mu = 1/2$, i.e.,

$$a_0 - a^\dagger = F'(a^\dagger)^*w \quad \text{for some} \quad w \in \mathcal{Y}$$

is then equivalent to the conditions

$$a^\dagger \frac{a_0 - a^\dagger}{\dot{u}(a^\dagger)} \in H^1(0,1) \quad \wedge \quad \int_0^1 \left(\frac{a_0 - a^\dagger}{\dot{u}(a^\dagger)}\right)(s)\,ds = 0$$

and w is given by

$$w = \left(a^\dagger \frac{a_0 - a^\dagger}{\dot{u}(a^\dagger)}\right)'\,.$$

Note that these conditions are weaker than the ones implied by the setting $\mathcal{X} = H^1(0,1)$, where the source condition for $\mu = 1/2$ implies at least that $a^\dagger - a_0 \in H^3$ and has to satisfy some boundary conditions (see, e.g., [8, Example 10.17] for details). In our setting, however, we only conclude L^2 convergence rates instead of H^1 ones.

To numerically test the approach proposed here, we applied it to the test problem:

$$\begin{aligned} a^\dagger(s) &= 100 - 99.9(1-2s)^2 \\ u(s) &= s(1-s) \\ f(s) &= -200 + 599.4(1-2s)^2 \\ \underline{\gamma} &= 0.1 \quad \overline{\gamma} = 200 \\ a_0 &\equiv 100 \end{aligned} \tag{39}$$

Note that $a^\dagger$ and a_0 satisfy the source conditions for $\mu = 1/2$ with $w = 99.9f$.

The solution of the forward problem (38) was based on linear finite elements on a uniform grid with mesh size $h = 1/100$. The same finite element space was used to approximate a, i.e.,

$$F(a) \approx F_n(a) = \sum_{i=1}^{n-1} u_{n,i}\varphi_i\,, \quad u_n := [u_{n,i}]\,,$$

Figure 1. IRGNM with projection: exact solution (dashed line) and computed solution (solid line); 5-th (left) and 12-th (right) iteration

where u_n solves the linear equation

$$A_n(a)u_n = -[\langle f, \varphi_j \rangle], \quad A_n(a) := [\langle a\varphi_i', \varphi_j' \rangle].$$

The Fréchet-derivative $F_n'(a)$ is given by

$$F_n(\gamma)'h = \sum_{i=1}^{n-1} du_{n,i}\varphi_i, \quad du_n := [du_{n,i}],$$

where du_n solves the linear equation

$$A_n(a)du_n = -A_n(h)u_n,$$

u_n is as above, and $A_n(h)$ is similarly defined as $A_n(a)$.

The diffusivity is approximated by

$$a = \sum_{i=0}^{n} a_{n,i}\varphi_i, \qquad a_n := [a_{n,i}] \quad \text{with} \quad \underline{\gamma} \le a_{n,i} \le \overline{\gamma}.$$

The projected IRGNM in this setting then reads as follows:

$$\begin{aligned} a_{n,k+1}^\delta &= P_n(a_{n,k}^\delta + (B_{n,k} + \alpha_k M_n)^{-1}(r_{n,k}^\delta + \alpha_k M_n(a_{n,0} - a_{n,k}^\delta))), \\ B_{n,k} &:= [\langle F'(a_{n,k}^\delta)\varphi_i, F'(a_{n,k}^\delta)\varphi_j \rangle], \\ M_n &:= [\langle \varphi_i, \varphi_j \rangle], \\ r_{n,k}^\delta &:= [\langle y^\delta - F(a_{n,k}^\delta), F'(a_{n,k}^\delta)\varphi_j \rangle], \end{aligned}$$

where P_n is the metric projector onto the set

$$\{a_n \in \mathbb{R}^{n+1} : \underline{\gamma} \le a_{n,i} \le \overline{\gamma} \quad \text{for all} \quad i = 0, \ldots, n\}.$$

Figure 1 shows the results obtained by this method for exact data ($\delta = 0$) with $\alpha_k = 2^{-k}10^{-9}$. Without projection we do not get convergence.

Experiments with noisy data are displayed in Table 1. Here we used uniformly distributed random noise. For each noise level, we carried out three test runs and

$\delta-\%$	k_*	$\frac{\Vert a^\delta_{n,k_*}-a^\dagger\Vert}{\Vert a^\dagger\Vert}$
1.00E00	9	0.337E00
0.50E00	10	0.417E00
0.25E00	11	0.291E00
0.13E00	12	0.194E00
0.63E-1	13	0.719E-1
0.31E-1	14	0.323E-2
0.16E-2	15	0.256E-2
0.78E-3	16	0.117E-2

$\delta-\%$	k_*	$\frac{\Vert a^\delta_{n,k_*}-a^\dagger\Vert}{\Vert a^\dagger\Vert}$
1.00E00	9	0.594E00
0.50E00	11	0.277E00
0.25E00	13	0.222E00
0.13E00	15	0.195E00
0.63E-1	17	0.168E00
0.31E-1	18	0.128E00
0.16E-2	20	0.112E00
0.78E-3	22	0.829E-1

Table 1. Convergence as $\delta \to 0$ with smooth (left) and nonsmooth (right) initial error

computed the average relative error $\Vert a^\delta_{n,k_*} - a^\dagger\Vert \,/\, \Vert a^\dagger\Vert$. The left three columns in Table 1 show the results for the original example (39) with $\alpha_k = 2^{-k}10^{-6}$; the iteration was stopped according to (18) with $\eta = 4.10^5$ and $\mu = 1/2$. In the right three colums we display the results with a different starting parameter, namely

$$a_0 \equiv 0.1\,,$$

so that the source condition with $\mu = 1/2$ is no longer satisfied. In this case α_k is as above, but the iteration is stopped according to (18) with $\eta = 40$ and $\mu = 1/38$; this choice guarantees that $\delta\alpha_{k_*}^{-\frac{1}{2}} \to 0$. The iterations still converge but much slower than for the smooth initial error. Note that in all cases an error is also introduced due to the finite dimensional approximation.

4. Conclusions and remarks

The aim of this paper was to emphasize that for mildly ill-posed problems, convergence (rates) can be established under moderate smoothness assumptions on the solutions as well as the forward operator F without requiring restrictions on the nonlinearity of F. Moreover, constraints defined by closed convex sets can be handled by projection.

It can be seen in a straightforward manner that F' appearing in (14) need not necessarily be the Fréchet derivative of F. More precisely, an inspection of the proof shows that $F'(x^\delta_k)$ can be replaced by any linear operator K_k satisfying

$$\Vert K_k - F'(x^\delta_k)\Vert \le c_1\delta^{\frac{2\mu}{2\mu+1}} + c_2\alpha_k^\mu$$

with constants $c_1, c_2 > 0$, where c_2 is sufficiently small.

For the unconstrained case, in [2] a similar result was obtained by different methods of proof and under different assumptions on the regularization methods. There, the difference $R_\alpha(F'(x^\delta_k)) - R_\alpha(F'(x^\dagger))$ is estimated via the Riesz-Dunford formula. For this purpose, the regularization method has to lie within the class defined via spectral theory (9), (10), with an analytic function g_α. The analysis there – although we wish to stress that it is more general than previously presented proofs under the given assumptions –

therefore does not include regularization methods according to (9), (10) with nonsmooth g_α such as truncated SVD, as well as regularization methods that are not contained in the class defined by (9), (10), such as regularization by discretization. Note that here we instead estimate the difference $(F'(x_k^\delta)^*F'(x_k^\delta))^\mu - (F'(x^\dagger)^*F'(x^\dagger))^\mu$, which enables more flexibility to include also the mentioned exceptions in the analysis.

We here have only considered the a priori stopping rule (18). Concerning an a posteriori stopping rule, we wish to refer to [3], where (11) with R_α being defined by (iterated) Tikhonov regularization or Landweber iteration was analyzed together with a Lepskii type stopping rule. Note that by the error estimates derived here, the results from [3] can be directly carried over to the setting of Theorems 4 and Corollary 6 to provide an optimally convergent a posteriori strategy in place of (18).

Acknowledgment

Part of this paper has been completed during the first author's stay at the Radon Institute for Computational and Applied Mathematics (RICAM) during the Special Semester on Computational Mechanics 2005. Support by the Austrian Academy of Sciences is gratefully acknowledged.

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

[2] A.B. Bakushinsky and M.Yu. Kokurin, *Iterative Methods for Approximate Solution of Inverse Problems*, Springer, Dordrecht, 2004.

[3] F. Bauer and T. Hohage, *A Lepskij-type stopping rule for regularized Newton methods*, Inverse Problems 21 (2005), 1975–1991.

[4] B.Hofmann and O. Scherzer, *Factors influencing the ill-posedness of nonlinear problems*, Inverse Problems 10 (1994), 1277–1298.

[5] G. Chavent and K. Kunisch, *On weakly nonlinear inverse problems*, SIAM J. Appl. Math. 56 (1996), 542–572.

[6] P.C. Ciarlet, *The Finite Element Method for Elliptic Problems*, North Holland, Amsterdam, 1978.

[7] H.W. Engl, K. Kunisch, and A. Neubauer, *Convergence rates for tikhonov regularisation of non-linear ill-posed problems*, Inverse Problems 5 (1989), 523–540.

[8] H.W. Engl, M. Hanke, and A. Neubauer, *Regularization of Inverse Problems*, Kluwer, Dordrecht, 1996.

[9] M. Hanke, *A regularization Levenberg-Marquardt scheme, with applications to inverse groundwater filtration problems*, Inverse Problems 13 (1997), 79–95.

[10] B. Kaltenbacher, *Regularization by projection with a posteriori discretization level choice for linear and nonlinear ill-posed problems*, Inverse Problems 16 (2000), 1523–1539.

[11] ——, *On the regularizing properties of a full multigrid method for ill-posed problems*, Inverse Problems 17 (2001), 767–788.

[12] ——, *V-cycle convergence of some multigrid methods for ill-posed problems*, Math. Comp. 72 (2003), 1711–1730.

[13] ——, *Determination of parameters in nonlinear hyperbolic pdes via a multiharmonic formulation, used in piezoelectric material characterization*, Math. Meth. Mod. Appl. Sci. (M3AS) (2006).

[14] B. Kaltenbacher, M. Kaltenbacher, and S. Reitzinger, *Identification of nonlinear* $B-H$

curves based on magnetic field computations and multigrid methods for ill-posed problems, Eur. J. Appl. Math. 14 (2003), 13–38.

[15] B. Kaltenbacher, A. Neubauer, and O. Scherzer, *Iterative Regularization Methods for Nonlinear Problems*, Springer, Dordrecht, to appear.

[16] P. Kügler, *Identification of a temperature dependent heat conductivity by Tikhonov regularization*, J. Inv. Ill-Posed Probl. 10 (2002), 67–90.

[17] R. Ramlau, *TIGRA – an iterative algorithm for regularizing nonlinear ill-posed problems*, Inverse Problems 19 (2003), 433–465.