JOURNAL OF INTEGRAL EQUATIONS
AND APPLICATIONS
Volume 20, Number 2, Summer 2008

CONVERGENCE RATES OF A MULTILEVEL METHOD FOR THE REGULARIZATION OF NONLINEAR ILL-POSED PROBLEMS

BARBARA KALTENBACHER

Communicated by Charles Groetsch

Dedicated to M. Zuhair Nashed

ABSTRACT. In this paper, we prove convergence rates for a previously [22] proposed multilevel method for solving nonlinear ill-posed operator equations

$$F(x) = y.$$

By minimizing the distance to some initial guess under the constraint of a discretized version of the operator equation for different levels of discretization, we define a sequence of regularized approximations to the exact solution, that in [22] had been shown to be stable and convergent for arbitrary initial guess, and can be computed via a multilevel procedure that altogether yields a globally convergent method. In the present paper we prove optimal logarithmic and Hölder type convergence rates under respective source conditions. Moreover we provide a tool for possible numerical solution strategies for the minimization problem on each level of discretization by providing an exact penalty function derived via an augmented Lagrangian approach.

1. Introduction.

Consider a nonlinear operator equation

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

with a continuous operator $F : \mathcal{D}(F) \subseteq X \to Y$ between Hilbert spaces X, Y, that is ill-posed in the sense of unstable dependence of a solution

2000 AMS *Mathematics subject classification.* 65J15, 65J20, 47J06, 65R30.
Received by the editors on July 27, 2007, and in revised form on September 21, 2007.
DOI:10.1216/JIE-2008-20-2-201

on the data. Since in practice we are given only noisy data y^δ (with here presumably known noise level δ) according to

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

it is necessary to apply regularization (see, e.g., [6, 12, 24, 25, 31, 36, 44, 46], and especially [1, 14, 15, 16, 23, 34, 40] on iterative methods for nonlinear problems).

Combining the idea of regularization by discretization (cf., e.g., [4, 8, 13, 21, 24, 27, 28, 35, 37, 38, 45]) with an iterative solution aspect, in [22] we proposed a multilevel regularization method as follows: Considering a sequence of finite dimensional subspaces $(Y_l)_{l\in\mathbb{N}}$ of data space Y with $k_l := dim(Y_l) < \infty$ for all $l \in \mathbb{N}$, but $k_l \to \infty$ as $l \to \infty$, we restrict the original equation (1) to Y_l by a mapping $Q_l : Y \to Y_l$

$$Q_l F(x) = Q_l y^\delta \,. \tag{3}$$

A solution x in the infinite dimensional space X of the finitely many equations defined by (3) is highly nonunique, so we adopt the concept of a best approximate solution known from the linear case (see, e.g., [6, 33]), by minimizing the distance to some point x_0

$$\min \|x - x_0\| \quad \text{s.t.} \quad Q_l F(x) = Q_l y^\delta \,.$$

Rather than exactly stipulating the equation (3) containing noisy data, it makes sense to consider an inexact solution, so that we arrive at minimization problems of the form

$$\min \|x - {x_0}^2\| \quad \text{s.t.} \quad \|Q_l(F(x) - y^\delta)\|^2 \leq \eta_l^2 \,. \tag{PI_l}$$

where $\eta_l \geq \delta$, with δ being the noise level in (2). More precisely, we will consider an a priori fixed monotonically decreasing sequence of tolerances $(\eta_l)_{l\in\mathbb{N}}$ with

$$\eta_l \searrow 0 \quad \text{as} \;\; l \to \infty \tag{4}$$

as well as a sequence of mappings $(Q_l)_{l\in\mathbb{N}}$, consider solutions ${x^\delta}_l$ of problem (PI_l) for all l such that

$$\eta_l \geq \delta \,,$$

and choose the discretization level $l_* = l_*(\delta)$, which here plays the role of a regularization parameter, such that

$$\eta_{l_*(\delta)} \sim \delta \,.$$

This corresponds to the well-known discrepancy principle (cf. Morozov [31]).

Using a solution of (PI_l) as a starting guess for the iterative minimization of (PI_{l+1}), one arrives at a multilevel method that by the use of information on coarser grids is highly efficient and has some nice global convergence properties, cf. [22]. Additionally, the restrictions on the nonlinearity of the forward operator that are usually required in convergence proofs for regularization methods for nonlinear ill-posed problems cf., e.g., [5, 14, 15, 16, 17, 19, 20, 34, 39, 40]), can be considerably relaxed (cf. Remark 1 in [22]): In place of the often used tangential cone condition

$$\begin{aligned}(5)\qquad &\|F(x) - F(\bar{x}) - F'(\bar{x})(x-\bar{x})\| \le C\|x-\bar{x}\|\|F'(\bar{x})(x-\bar{x})\| \\ &\qquad \forall\, x,\ \ \bar{x} \in \mathcal{D}(F)\end{aligned}$$

in [22] the assumption

$$\begin{aligned}&\|Q_l(F(x+w) - F(x) - F'(x)w)\| \le \frac{1}{\eta_l}\|Q_l(F(x+w) - F(x))\|^2 \\ &\qquad \forall\, x,\ \ x+w \in \mathcal{D}(F)\,,\end{aligned}$$

turned out to suffice, which is similar to (and actually motivated by) the assumption in the multilevel approach in [41].

In the present paper we extend the results of [22] by proving optimal convergence rates

$$(6)\quad \|x_{l_*}^\delta - x^\dagger\| = O\left(\delta^{2\nu/(2\nu+1)}\right) \qquad \text{or} \qquad \|x_{l_*}^\delta - x^\dagger\| = O((-\ln\delta)^{-p})\,,$$

under respective a priori regularity conditions of Hölder or logarithmic type (cf., e.g., [17, 18] for the latter)

$$\begin{aligned}(7)\qquad x_0 - x^\dagger &= (F'(x^\dagger)^* F'(x^\dagger))^\nu v \qquad \text{or} \\ x_0 - x^\dagger &= (-\ln(F'(x^\dagger)^* F'(x^\dagger)))^{-p} v\,,\end{aligned}$$

for some $v \in X$, where $f(F'(x^\dagger)^*F'(x^\dagger))$ for some real function f is defined in the sense of the functional calculus resulting from spectral theory (see, e.g., [6]), $x^\dagger$ is a solution of (1), and x_0 an initial guess that need not necessarily be close to $x^\dagger$, though.

Also, we will show that we do not need global minimizers of (PI_l) but it suffices to consider KKT (Karush Kuhn Tucker) points i.e., points ${x^\delta}_l$ that together with Lagrange multipliers ${\lambda^\delta}_l$ satisfy

$$\left.\begin{array}{l} {x^\delta}_l - x_0 + {\lambda^\delta}_l (Q_l F'({x^\delta}_l))^* Q_l (F({x^\delta}_l) - y^\delta) = 0 \\ {\lambda^\delta}_l \geq 0\,,\ \|Q_l(F({x^\delta}_l) - y^\delta)\| \leq \eta_l\,, \\ \lambda^\delta(\|Q_l(F({x^\delta}_l) - y^\delta)\| - \eta_l) = 0 \end{array}\right\} \ (KKT_l)\,.$$

Moreover, we will consider the constrained minimization problem (PI_l) on fixed level in more detail and take advantage of its similarity to the trust region subproblem in unconstrained nonlinear programming to define a differentiable exact penalty function Φ_α. The latter allows to either replace (PI_l) by unconstrained minimization of Φ_α or to use Φ_α to define a merit function for the stepsize control of an SQP type iteration applied to (PI_l), see, e.g., [9, 10, 11].

2. Preliminaries and assumptions. First of all, we assume that a solution $x^\dagger$ of (1) exists, but need not necessarily be unique. The initial guess x_0 used in (PI_l) need not be close to $x^\dagger$ — this fact is essential when speaking of global convergence. We will assume that x_0 is not in the feasible set of (PI_l), i.e.

$$\|Q_l(F(x_0) - y^\delta)\| > \eta_l\,, \tag{8}$$

which just excludes the trivial case that x_0 itself solves (PI_l) and therewith is a quite natural assumption.

The operator $F : \mathcal{D}(F) \subseteq X \to Y$ (X, Y Hilbert spaces) is assumed to be continuous, compact and (weakly) sequentially closed. i.e.,

$$\Big(x_k \rightharpoonup x \wedge F(x_k) \to f\Big) \quad \Rightarrow \quad \Big(x \in \mathcal{D}(F) \wedge F(x) = f\Big)\,.$$

Additionally, we assume the nonlinearity condition

$$\forall\, x \in \mathcal{D}(F)\ \exists F'(x) \in L(X,Y) \quad \forall\, w \in X \ \text{ s.t. } \ x + w \in \mathcal{D}(F) : \tag{9}$$

$$\|Q_l(F(x+w) - F(x) - F'(x)w)\| \leq \frac{c}{\eta_l}\|Q_l(F(x+w) - F(x))\|^2 + \tilde{c}\gamma_l\|w\|^2$$

to hold for some $0 < c < 1$, $\tilde{c} > 0$. In here, $F'(x)$ is not necessarily the Fréchet derivative but only denotes some linearization of F satisfying the first order Taylor remainder estimate (9) on some possibly non-open set $\mathcal{D}(F)$.

The sequence $(Q_l)_{l\in\mathbb{N}}$ consists of linear operators mapping Y into finite dimensional subsets

$$Y_l = \mathcal{R}(Q_l) \subseteq Y$$

and is assumed to pointwise converge to the identity

$$\forall\, f \in Y \;:\quad Q_l f \to f \quad \text{as } l \to \infty\,. \tag{10}$$

The latter implies uniform boundedness of the operators Q_l,

$$\|Q_l\| \le C_Q \quad \forall\, l \in \mathbb{N}\,. \tag{11}$$

Additionally, we assume that

$$\forall\, f \in Y \;:\quad \|Q_l^2 f\| \ge c_l^Q \|Q_l f\| \tag{12}$$

for some $c_l^Q > 0$, and that

$$\mathcal{N}(Q_l)^\perp \subseteq \overline{\mathcal{R}(F'(x))} \quad \forall\, x \in \mathcal{D}(F)\,. \tag{13}$$

To prove convergence rates we make an assumption quantifying the rate of convergence of the approximation error on $\mathcal{R}(F'(x^\dagger))$, which by the smoothing property of $F'(x)$ is typically a space of smoother functions than general elements of Y:

$$\|(I - Q_l)F'(x^\dagger)\| \le \overline{\gamma}_l\,, \tag{14}$$

with $\overline{\gamma}_l \to 0$ as $l \to \infty$. Assumptions (10, 11, 12) are satisfied, e.g., if the Q_l are projections onto subspaces Y_l of the data space, with

$$Y_1 \subseteq Y_2 \subseteq \cdots \subseteq Y \quad \wedge \quad \overline{\bigcup_{n\in\mathbb{N}} Y_n} = Y$$

for all $x \in \mathcal{D}(F)$. In this case $c_l^Q = 1$ and, if the Q_l are orthogonal projections, $C_Q = 1$.

From compactness of $F'(x)$ for $x \in \mathcal{D}(F)$ it follows that for the generalized inverses (cf [33]) of the projected operators there holds

$$\|(Q_l F'(x))^\dagger\| \to \infty \quad \text{as } l \to \infty \,. \tag{15}$$

Conversely, each Y_l is finite dimensional, so that by (13) we can assume

$$\inf_{z \in Q_l^* Y} \frac{\|F'(x)^* z\|}{\|z\|} \geq \hat{\gamma}_l \tag{16}$$

to hold for a sequence $\hat{\gamma}_l$ of strictly positive numbers, that tends to zero due to (15). In general $\hat{\gamma}_l$ will depend on x; an assumption we make here is that a uniform constant $\hat{\gamma}_l > 0$ for all $x \in \mathcal{D}(F)$ exists for each level l. The lower bound (16) together with (12) and (13) implies the estimate

$$\|(Q_l F'(x))^\dagger\|_{Y_l \to Y_l} \leq \frac{1}{\gamma_l}\,, \tag{17}$$

with $\gamma_l = \hat{\gamma}_l c_l^Q$, for the inverse of the linearized forward operator on level l, as well as

$$\mathcal{R}(Q_l F'(x)) = \mathcal{R}(Q_l) \quad \forall\, x \in \mathcal{D}(F)\,, \tag{18}$$

cf. [22], which implies

$$\begin{aligned} Q_l(F(x) - y^\delta) \in \mathcal{R}(Q_l) &= \mathcal{R}(Q_l F'(x)) \\ &= \mathcal{N}((Q_l F'(x))^*)^\perp, \quad \forall\, x \in \mathcal{D}(F)\,. \end{aligned} \tag{19}$$

For the sequence of tolerances η_l,besides monotone decrease (4) we will later on make some assumption on the decay rate, see (46) below.

Finally, we wish to comment on condition (9). As a matter of fact, it is weaker than the usual tangential cone condition (5), since if the latter holds and if we can restrict attention to a sufficiently small neighborhood of the initial guess, i.e., consider

$$\begin{aligned} &\|F(x+w) - F(x) - F'(x)w\| \leq C\|w\|\, \|F'(x)w\|, \\ &\qquad \forall\, x,\ \ x+w \in \mathcal{D}(F) \cap \mathcal{B}_\rho(x_0) \end{aligned} \tag{20}$$

extend (14) to all points $x \in \mathcal{D}(F) \cap \mathcal{B}_\rho(x_0)$ in place of $x^\dagger$, and assume that $\overline{\gamma}_l/\gamma_l$ is uniformly bounded by some constant C_γ, as it is often done, we can conclude, for all $x,\ x+w \in \mathcal{D}(F) \cap \mathcal{B}_\rho(x_0)$

$$\begin{aligned}
&\|Q_l(F(x+w) - F(x) - F'(x)w)\| \\
&\qquad\qquad \le CC_Q\|w\|\Big(\|Q_lF'(x)w\| + \|(I-Q_l)F'(x)w\|\Big) \\
&\qquad\qquad \le CC_Q\|w\|\Big(\|Q_lF'(x)w\| + \overline{\gamma}_l\|w\|\Big)
\end{aligned}$$

which by the second triangle inequality and provided ρ is small enough so that $2\rho\, CC_Q < 1$, yields

$$\|Q_lF'(x)w\| \le \frac{1}{1-2\rho CC_Q}\Big(\|Q_l(F(x+w)-F(x))\| + 2\rho CC_Q\overline{\gamma}_l\|w\|\Big)$$

Inserting this into the inequality above yields

$$\begin{aligned}
&\|Q_l(F(x+w) - F(x) - F'(x)w)\| \\
&\quad \le \frac{CC_Q}{1-2\rho CC_Q}\left(\|w\|\|Q_l(F(x+w)-F(x))\| + \overline{\gamma}_l\|w\|^2\right) \\
&\quad \le \frac{CC_Q}{1-2\rho CC_Q}\left(\frac{1}{2\epsilon}\|Q_l(F(x+w)-F(x))\|^2 + \left(\frac{\varepsilon}{2}+\overline{\gamma}_l\right)\|w\|^2\right) \\
&\quad \le \frac{c}{\eta_l}\|Q_l(F(x+w)-F(x))\|^2 + \left(\frac{1-2\rho CC_Q}{4cCC_Q}\frac{\eta_l}{\gamma_l} + C_\gamma\right)\gamma_l\|w\|^2
\end{aligned}$$

by setting $\varepsilon := (1-2\rho CC_Q)/(2cCC_Q)\,\eta_l$. Hence choosing $\eta_l \sim \gamma_l^\beta$ for any $\beta \in [1,2]$ (to be compatible with assumption (46) that will we make on (η_l) later in this paper), we arrive at (9). Note, that while for (5) the full infinite dimensional Taylor remainder has to be estimated, (9) only requires upper bounds on finite dimensional subspaces Y_l. Moreover, assuming (9) only for the fastest possible decaying sequence $\eta_l \sim \gamma_l^2$ (instead of $\eta_l \sim \gamma_l^\beta$ for all $\beta \in [1,2]$) gives a looser upper bound for the first term on the right hand side.

3. Solution of the subproblem on level l. Having in mind (PI_l), in this section we consider problems of the form

$$\text{(PI)} \qquad \min \|x-x_0\|^2 \quad \text{s.t.} \quad \|G(x)\|^2 \le \eta^2\,.$$

with a continuous, compact and (weakly) sequentially closed operator $G : \mathcal{D}(G) \subset X \to \tilde{Y}$ satisfying the nonlinearity condition

$$\text{(21)}\ \forall\, x \in \mathcal{D}(G) \exists\, G'(x) \in L(X,Y) \quad \forall\, w \in X \ \text{ s.t. } \ x+w \in \mathcal{D}(G) :$$
$$\|G(x+w) - G(x) - G'(x)w\| \leq \frac{c}{\eta}\|G(x+w) - G(x)\|^2 + \tilde{c}\gamma\|w\|^2$$

with $0 < c < 1$, $\tilde{c} > 0$, cf. (9), as well as

$$\|G'(x)^\dagger\| \leq \frac{1}{\gamma}\,, \tag{22}$$

cf. (17),

$$G(x) \in \mathcal{N}(G(x)^*)^\perp\,, \tag{23}$$

cf. (19), and

$$\|G(x_0)\| > \eta\,, \tag{24}$$

cf. (8).

The latter two assumptions and $\eta > 0$ imply that at global minimizers the linear independence constraint qualification

$$G'(x)^* G(x) \neq 0 \tag{25}$$

is satisfied for (PI), since global minimizers of (PI) by (24) automatically satisfy the constraint with equality:

Lemma 1. (*Lemma 1 in* [22])

If there exists an $x^\dagger$ such that $\|G(x^\dagger)\| \leq \eta$, then Problem (PI) has a global minimizer.

If (24) *holds, then any minimizer of* (PI) *lies on the relative boundary of the feasible set, i.e., the inequality constrained problem* (PI) *is equivalent to the equality constrained problem*

$$\|x - x_0\|^2 = \min! \quad \|G(x)\|^2 = \eta\,. \tag{PE}$$

For sufficiently small $c, \tilde{c}$ in the nonlinearity condition (21), the primal parts of KKT points of (PI), i.e., of solutions (x_*, λ_*) to

$$\text{(KKT)} \qquad \left.\begin{aligned} x_* - x_0 + \lambda_* G'(x_*)^* G(x_*) = 0 \\ \lambda_* \geq 0\,, \quad \|G(x_*)\| \leq \eta\,, \quad \lambda_*(\|G(x_*)\| - \eta) = 0 \end{aligned}\right\} .$$

are already global minimizers:

Lemma 2. *If (21) with $c \leq 1/2$, holds, then for any solution (x_*, λ_*) of (KKT) satisfying $\tilde{c}\|x_* - x_0\| < 1/2$, the primal part x_* is a strict (hence unique) global minimizer of (PI).*

Proof. Let $x \neq x_*$ be an arbitrary feasible point for (PI), i.e., satisfying $\|G(x)\| \leq \eta$. We consider the representation

$$\begin{aligned} \|x - x_0\|^2 - \|x_* - x_0\|^2 &= \|x - x_*\|^2 + 2\langle x - x_*, x_* - x_0\rangle \\ &= \|x - x_*\|^2 - 2\lambda_* \langle G'(x_*)(x - x_*), G(x_*)\rangle \end{aligned}$$

(where we have used the first line of (KKT)) of the difference between the cost function values. According to the complementarity condition $\lambda_*(\|G(x_*)\| - \eta) = 0$, we distinguish between two cases:

If $\lambda_* = 0$, then obviously (26) implies

$$\|x - x_0\|^2 > \|x_* - x_0\|^2 \,. \tag{27}$$

Otherwise, if $\lambda_* > 0$ then $\|G(x_*)\| = \eta$, hence, from (26) we deduce

$$\begin{aligned} &\|x - x_0\|^2 - \|x_* - x_0\|^2 \\ &\quad = \|x - x_*\|^2 \\ &\qquad + \lambda_* \Big(\underbrace{\|G(x_*)\|^2 - \|G(x)\|^2}_{=\eta_l^2 - \|G(x)\|^2 \geq 0} + \|G(x) - G(x_*)\|^2 \\ &\qquad + 2\langle G(x) - G(x_*) - G'(x_*)(x - x_*), G(x_*)\rangle \Big)\,, \end{aligned} \tag{28}$$

Forming the inner product with $(G'(x_*))^\dagger G(x_*)$ of the first line of (KKT), and using (23), we obtain an estimate on the Lagrange multiplier λ_*:

$$|\lambda_*| = \left|\frac{\langle x_* - x_0, G'(x_*)^\dagger G(x_*)\rangle}{\|G(x_*)\|^2}\right| \le \|x_* - x_0\| \frac{1}{\|G(x_*)\|\gamma}\,. \tag{29}$$

Inserting this into (28) and using (21) with $c \le 1/2$, $\tilde{c}\|x_* - x_0\| < 1/2$ yields (27). □

The following lemma states existence of a KKT point, provided $\mathcal{D}(G)$ is convex and there exists a multiple of the steepest descent direction that yields a point in $\mathcal{D}(G)$ (condition (30)), which is, e.g., the case if $\mathcal{D}(G)$ is open.

Lemma 3. *Let $\mathcal{D}(G)$ be convex, $x_0 \in \mathcal{D}(G)$, G Hölder continuous with exponent $\alpha > 1/2$, and let $x_* \neq x_0$ be a local minimizer of (PI) satisfying*

$$\exists\, \bar{\varepsilon} > 0 \ : \quad x_* - \bar{\varepsilon} G'(x_*) G(x_*) \in \mathcal{D}(G)\,. \tag{30}$$

Then there exists a λ_* such that (x_*, λ_*) solves (KKT).

Proof. We will see that following the usual existence proof relying on the Farkas Lemma which in its turn is a consequence of a strong separation theorem (see, e.g., [11]) for our special problem (PI) naturally leads us to the use of assumption (30):

Denote

$$r(x_*) := G'(x_*) G(x_*)$$

and consider the set

$$\mathcal{M} := \{d \in X \mid \exists\, \bar{\tau} > 0 \ : \quad x_* + \bar{\tau} d \in \mathcal{D}(G)\}$$

Assume that no solution to (KKT) exists, i.e.,

$$x_* - x_0 \notin \mathcal{C} := \{-\lambda r(x_*) \mid \lambda \ge 0\}\,. \tag{31}$$

Due to convexity and closedness of $\mathcal{C}$, this set can be strictly separated from the point $x_* - x_0$ in the sense that there exists an element $a \in \mathcal{M}$ such that

$$\forall\, y \in \mathcal{C}: \quad \langle a, y\rangle \geq 0 > \langle a, x_* - x_0\rangle \tag{32}$$

e.g.,

$$a := -\bar{\lambda} r(x_*) - (x_* - x_0) \quad \text{with} \quad \bar{\lambda} := \max\left\{0, -\frac{\langle x_* - x_0, r(x_*)\rangle}{\|r(x_*)\|^2}\right\}.$$

(The separation property (32) is readily checked, using the fact that by our assumption (31) $\langle x_* - x_0, r(x_*)\rangle^2 < \|x_* - x_0\|^2 \|r(x_*)\|^2$. The fact that $a \in \mathcal{M}$ follows from convexity of $\mathcal{D}(G)$ by setting $\overline{\tau} := \overline{\varepsilon}/(\overline{\varepsilon} + \bar{\lambda})$.) On the other hand,

$$\forall\, d \in \mathcal{M} \;:\; \langle r(x_*), d\rangle \leq 0 \;\Rightarrow\; \langle d, x_* - x_0\rangle \geq 0\,, \tag{33}$$

which can be seen as follows: For arbitrary $d \in \mathcal{M}$ with $\langle r(x_*), d\rangle \leq 0$, and any $\tau \in (0, \overline{\tau}/2]$, $\varepsilon \in (0, \overline{\varepsilon}/2]$, consider

$$d^{\tau,\varepsilon} := \tau d - \varepsilon r\,,$$

which by convexity of $\mathcal{D}(G)$ satisfies

$$\begin{aligned} x_* + d^{\tau,\varepsilon} = {} & \frac{1}{2}\left(\frac{2\tau}{\overline{\tau}}(x_* + \overline{\tau} d) + \left(1 - \frac{2\tau}{\overline{\tau}}\right) x_*\right) \\ & + \frac{1}{2}\left(\frac{2\varepsilon}{\overline{\varepsilon}}(x_* - \overline{\varepsilon} r(x_*)) + \left(1 - \frac{2\varepsilon}{\overline{\varepsilon}}\right) x_*\right) \in \mathcal{D}(G)\,. \end{aligned}$$

Therewith, we get

$$\begin{aligned} & \|G(x_* + d^{\tau,\varepsilon})\|^2 - \eta^2 \\ & = \|G(x_*)\|^2 - \eta^2 + 2\langle r(x_*), d^{\tau,\varepsilon}\rangle \\ & \quad + 2\langle G(x_*), G(x_* + d^{\tau,\varepsilon}) - G(x_*) - G'(x_*) d^{\tau,\varepsilon}\rangle \\ & \quad + \|G(x_* + d^{\tau,\varepsilon}) - G(x_*)\|^2 \\ & \leq -2\varepsilon \|r(x_*)\|^2 + \left(1 + \frac{2c\|G(x_*)\|}{\eta}\right) L^2 \|d^{\tau,\varepsilon}\|^{2\alpha} + 2\tilde{c}\gamma \|G(x_*)\| \|d^{\tau,\varepsilon}\|^2 \end{aligned}$$

where L denotes the constant in the Hölder estimate of G. With a choice $\varepsilon = \varepsilon(\tau) = (\overline{\varepsilon}/\overline{\tau})\tau$, this implies

$$\|G(x_* + d^{\tau,\varepsilon(\tau)})\|^2 - \eta^2 \leq 0$$

i.e., feasibility of $d^{\tau,\varepsilon(\tau)}$ for all τ sufficiently small. Hence, optimality of x_* yields

$$0 \leq \|x_* + d^{\tau,\varepsilon(\tau)} - x_0\|^2 - \|x_* - x_0\|^2 = 2\langle d^{\tau,\varepsilon(\tau)}, x_* - x_0\rangle + \|d^{\tau,\varepsilon(\tau)}\|^2$$

from which, by letting τ tend to zero, we can conclude (33). However, (33) is a contradiction to our previous construction of $a \in \mathcal{M}$ satisfying (32) . $\square$

Note the similarity of (PI) to the trust region subproblem

$$\min f(x^k) + f'(x^k)d + f''(x^k)[d,d] \quad \text{s.t.} \quad \|d\|^2 \leq \Delta$$

for determining the step d within the kth iteration of the unconstrained minimization of a functional f by a trust region method. With this analogy as well the approach proposed by Lucidi, Palagi, and Roma [26] (see also Section 14.3 in [10]) in mind, we show some properties of (PI), that enable to derive an exact penalty function Φ_α, i.e., such that (PI) is equivalent to the unconstrained minimization of Φ_α for sufficiently large α. Equivalence is shown to hold true both in the sense of global minimizers and in the sense of critical points, see Proposition 1 and Corollary 1 below.

An exact penalty function for (PI) can be derived analogously to the trust region subproblem (cf., e.g., [10]) as follows: It is obvious, that

$$x \mapsto \lambda(x) = -\frac{\langle x - x_0, G'(x)^* G(x)\rangle}{\|G'(x)^* G(x)\|^2}$$

is a Lagrange multiplier function in the sense that if (x_*, λ_*) is a KKT point and in $\lambda(x_*) \geq 0$, then $(x_*, \lambda(x_*))$ is also a KKT point. (Note that by the linear independence constraint qualification (25) the Lagrange multiplier is unique.) Inserting this into the augmented Lagrangian

$$\|x - x_0\|^2 + \frac{\alpha}{4}\left(\max{}^2\left\{0, \lambda + \frac{2}{\alpha}(\|G(x)\|^2 - \eta^2)\right\} - \lambda^2\right)$$

with $\alpha > 0$, we arrive at

$$\Phi_\alpha(x) = \|x - x_0\|^2 + \frac{\alpha}{4}\left(\max{}^2\left\{0, \lambda(x) + \frac{2}{\alpha}(\|G(x)\|^2 - \eta^2)\right\} - \lambda(x)^2\right) .$$

For sufficiently small $\alpha > 0$, this is an exact penalty functional in the following sense:

Proposition 1. *Let G be twice Fréchet differentiable.*

(i) *If $(x_*, \lambda(x_*))$ is a KKT point of (PI) then for any $\alpha > 0$, the primal part x_* is a stationary point of Φ_α and $\Phi_\alpha(x_*) = \|x_* - x_0\|$.*

(ii) *If for some $R > 0$,*

$$0 < \alpha < \frac{4(\gamma\underline{\eta})^2}{1 + RM/(\gamma\underline{\eta})} \tag{34}$$

with

$$M := M_1^2 + M_0 M_2 \,, \quad M_j := \sup_{x \in \mathcal{B}_R(x_0)} \|G^{(j)}(x)\| \,,\ j \in \{0, 1, 2\} \,,$$

then for any stationary point $x_* \in \mathcal{B}_R(x_0)$ of Φ_α with

$$\|G'(x_*)^* G(x_*)\| \geq \gamma\underline{\eta} \,, \tag{35}$$

the pair $(x_*, \lambda(x_*))$ *is a KKT point of (PI).*

Remark 1. The notation in (35) is supposed to remind of the fact that for $G = Q_l F$ as in the previous section, this condition is typically satisfied with $\gamma = \gamma_l$ and $\underline{\eta} = \|Q_l(F(x_*) - y^\delta)\|$. Note that by Lemma 1 the global minimizer of (PI) indeed satisfies $\|G(x)\| = \eta$ (unless it is trivial) and that in our original problem it is certainly easier to find points with larger residuals $\|Q_l(F(x) - y^\delta)\|$ than points violating (35) with $\underline{\eta} = \eta_l$ especially for larger l, since $\eta_l \to 0$ as $l \to \infty$.

Proof. We will use the fact the derivative of Φ_α into a direction $h \in X$ exists and is given by

$$\Phi'_\alpha(x)[h] = 2\langle x - x_0, h\rangle + 2\lambda(x)\langle G'(x)^* G(x), h\rangle \tag{36}$$

$$+ \max\left\{-\frac{\alpha}{2}\lambda(x), \|G(x)\|^2 - \eta^2\right\}\left(\lambda'(x)[h] + \frac{4}{\alpha}\langle G'(x)^* G(x), h\rangle\right) \tag{37}$$

with

$$\lambda'(x)[h] = -\frac{1}{\|r(x)\|^2}\Big(\langle r(x), h\rangle + \langle x - x_0, (I - 2Proj_{r(x)})r'(x)h\rangle\Big) \tag{38}$$

where

$$r(x) = G'(x)^* G(x)\,.$$

Moreover, solutions of the second line in (KKT) can be characterized as zeros of the max function appearing in (37)

$$\begin{aligned}(39)\quad &\Big(\lambda_* \geq 0\,,\ \|G(x_*)\| \leq \eta\,,\ \lambda_*(\|G(x_*)\| - \eta) = 0\Big)\\ &\qquad\Longleftrightarrow\quad \max\Big\{-\frac{\alpha}{2}\lambda_*, \|G(x)\|^2 - \eta^2\Big\} = 0\,.\end{aligned}$$

To show sufficiency (i) in the assertion of the proposition, assume that $(x_*, \lambda(x_*))$ is a KKT point. Then by the first line in (KKT), the expression in (36) vanishes, and by the second line in (KKT) and (39), so does the term in (37), i.e. $\Phi'_\alpha(x) = 0$.

For α satisfying (34), necessity (ii) follows from

$$\begin{aligned}(40)\qquad 0 &= \Phi'_\alpha(x_*)[G'(x_*)G(x_*)]\\ &= \max\Big\{-\frac{\alpha}{2}\lambda(x_*), \|G(x_*)\|^2 - \eta^2\Big\}\,(\lambda'(x_*)G'(x_*)^*G(x_*)\\ &\qquad\qquad + \frac{4}{\alpha}\|G'(x_*)^*G(x_*)\|^2)\,,\end{aligned}$$

where the second equation follows from the definition of $\lambda(x)$. The term multiplied with $\max\{-\frac{\alpha}{2}\lambda, \|G(x)\|^2 - \eta^2\}$ in (40) is strictly positive, since

$$\begin{aligned}&\lambda'(x_*)G'(x_*)^*G(x_*) + \frac{4}{\alpha}\|G'(x_*)^*G(x_*)\|^2\\ &= -1 - \frac{1}{\|r(x_*)\|^2}\langle x_* - x_0, (I - 2Proj_{r(x_*)})r'(x_*)r(x_*)\rangle + \frac{4}{\alpha}\|r(x_*)\|^2\\ &\geq -1 - \frac{RM}{\gamma\underline{\eta}} + \frac{4(\gamma\underline{\eta})^2}{\alpha} > 0\,.\end{aligned}$$

Hence, by (39), we can conclude the second line of (KKT) from (40). Now the first line of (KKT) directly follows from $\Phi'_\alpha(x_*) = 0$ and (36, 37). □

Note that for the proof of Proposition 1 the nonlinearity condition on G was not required.

Corollary 1. *Let G be twice Fréchet differentiable, let the assumptions of Lemma* 2 *hold, and assume that α satisfies* (34)*.*

Then x_* is a global minimizer of (PI) if and only if it is a global minimizer of Φ_α.

Proof . Using Lemma 2 and Proposition 1, the proof can be carried out exactly along the lines of the proof of Satz 14.14 in [10]. □

4. Convergence Rates. In [22] we have shown that for weakly sequentially closed forward operators F, a minimizer of (PI_l) exists, converges to a solution for exact data and with an appropriately chosen level $l_*(\delta)$ also for noisy data as $\delta \to 0$. Under additional nonlinearity conditions (9) with $c \leq 1$, $\tilde{c} = 0$, these assertions could be carried over to KKT points ${x^\delta}_l$ of (PI_l).

Since assumption (9) with $c < 1$ as made here allows to simplify the convergence proof considerably and this new proof contains an estimate that is the starting point also for our convergence rates proof, we provide it in the following proposition. The stopping rule we will use here is the generalized discrepancy principle

$$\underline{\tau}\delta \leq \eta_{l_*(\delta)} \leq \overline{\tau}\delta\,, \quad \underline{\underline{\tau}}\delta \leq \|Q_{l_*(\delta)}(F({x^\delta}_{l_*(\delta)}) - y^\delta)\| \tag{41}$$

with constants $0 < \underline{\tau} < \overline{\tau}$, $\underline{\underline{\tau}} > 0$ satisfying

$$\underline{\underline{\tau}} \geq \frac{1}{1-c}\left\{1 + 2c + \frac{c}{\underline{\tau}} C_Q\right\} C_Q\,. \tag{42}$$

Well-definedness of $l_*(\delta)$ according to (41) immediately follows from Lemmas 1, 2, if $c \leq 1/2$.

Proposition 2. *Fix $R > \|x^\dagger - x_0\|$ and let, for each l, $({x^\delta}_l, {\lambda^\delta}_l)$ be a KKT point of (PI_l) with $\|{x^\delta}_l - x_0\| \leq R$. Assume that* (9) *holds with $c < 1$, $\tilde{c}R < 1$.*

(i) *Let $y^\delta = y$ i.e., $\delta = 0$ in* (ii). *Then $(x_l^0)_{l\in\mathbb{N}} =: (x_l)_{l\in\mathbb{N}}$ has a convergent subsequence and the limit of each convergent subsequence of $(x_l)_{l\in\mathbb{N}}$ is a solution to* (1)*. If $x^\dagger$ is the unique solution to* (1)*, then x_l converges to $x^\dagger$.*

(ii) *Let $(y^n)_{n\in\mathbb{N}}$ be a sequence of data such that $\|y^n - y\| \leq \delta^n$ with noise levels $(\delta^n)_{n\in\mathbb{N}}$ tending to zero. Then the sequence $(x^n)_{n\in\mathbb{N}}$ defined by $x^n := x^{\delta^n}_{l_*(\delta^n)}$ with l_* chosen according to* (41) *with* (42) *has a convergent subsequence and the limit of each convergent subsequence of $(x^n)_{n\in\mathbb{N}}$ is a solution to* (1)*. Again, if $x^\dagger$ is the unique solution to* (1)*, then x^n converges to $x^\dagger$.*

Remark 2. Note that on one hand the case $R = \infty$, $\tilde{c} = 0$ (and $\tilde{c}R := 0$ by definition) is included in this proposition. On the other hand, it is readily checked that since $x^\dagger$ is feasible for (PI_l), $l \geq l_*(\delta)$, a global minimizer $\overline{x}_l^\delta$ of (PI_l) satisfies $\|\overline{x}_l^\delta - x_0\| \leq \|x^\dagger - x_0\|$, hence it makes sense to restrict attention to KKT points satisfying $\|{x^\delta}_l - x_0\| \leq R$ for some $R > \|x^\dagger - x_0\|$.

Proof. First of all, consider the case ${x^\delta}_l \neq x_0$, which by the first line in (KKT_l) and (19) implies ${\lambda^\delta}_l \neq 0$, $Q_l(F({x^\delta}_l) - y^\delta) \neq 0$. Analogously to (29) we get

$$|{\lambda^\delta}_l| = \left| \frac{\langle {x^\delta}_l - x_0, (Q_l F'({x^\delta}_l))^\dagger Q_l(F({x^\delta}_l) - y^\delta)\rangle}{\|Q_l(F({x^\delta}_l) - y^\delta)\|^2} \right| \leq \|{x^\delta}_l - x_0\| \frac{1}{\|Q_l(F({x^\delta}_l) - y^\delta)\|\gamma_l} \;. \tag{43}$$

Taking the inner product of the first line in (KKT_l) with ${x^\delta}_l - x^\dagger$, we get

$$
\begin{aligned}
&\|x^\delta{}_l - x^\dagger\|^2 \\
&= \langle x_0 - x^\dagger, x^\delta{}_l - x^\dagger\rangle - \lambda^\delta{}_l \langle Q_l(F(x^\delta{}_l) - y^\delta), Q_l F'(x^\delta{}_l)(x^\delta{}_l - x^\dagger)\rangle \\
&\leq \langle x_0 - x^\dagger, x^\delta{}_l - x^\dagger\rangle \\
&\quad - \lambda^\delta{}_l \|Q_l(F(x^\delta{}_l) - y^\delta)\| \Big\{ \|Q_l(F(x^\delta{}_l) - y^\delta)\| \\
&\qquad\quad - \Big(\frac{c}{\eta_l} (\|Q_l(F(x^\delta{}_l) - y^\delta)\| + C_Q\delta)^2 + C_Q\delta \Big) - \tilde{c}\gamma_l \|x^\delta{}_l - x^\dagger\|^2 \Big\} \\
&\leq \langle x_0 - x^\dagger, x^\delta{}_l - x^\dagger\rangle \\
&\quad - \lambda^\delta{}_l \|Q_l(F(x^\delta{}_l) - y^\delta)\| \Big\{ (1-c) \|Q_l(F(x^\delta{}_l) - y^\delta)\| \\
&\qquad\qquad\qquad\qquad - \Big(1 + 2c + \frac{c\delta}{\eta_l} C_Q \Big) C_Q \delta \Big\} + \tilde{c}R \|x^\delta{}_l - x^\dagger\|^2
\end{aligned}
$$

which for $l = l_*$ by (41, 42) implies

$$
(1 - \tilde{c}R)\|x^\delta{}_{l_*(\delta)} - x^\dagger\|^2 \leq \langle x_0 - x^\dagger, x^\delta{}_{l_*(\delta)} - x^\dagger\rangle\,, \tag{44}
$$

and in the noise free case

$$
(1 - \tilde{c}R)\|x_l - x^\dagger\|^2 \leq \langle x_0 - x^\dagger, x_l - x^\dagger\rangle\,, \tag{45}
$$

for all $l \in \mathbb{N}$. In case $x^\delta{}_l = x_0$ these inequalities trivially hold. By the Cauchy-Schwarz inequality, this implies boundedness of $x^\delta{}_{l_*(\delta)}$ or x_l, respectively. Since therewith the rest of the proof is exactly the same as the one for Theorem 1 in [22], (which, in its turn is analogous to the convergence proof for Tikhonov regularization in [42], see also [7]), we omit it here. □

Now we proceed to our main result, the proof of optimal convergence rates.

Theorem 1. *Fix $R > \|x^\dagger - x_0\|$ and let $(x^\delta{}_{l_*(\delta)}, \lambda^\delta{}_{l_*(\delta)})$ be a KKT point of $(PI_{l_*(\delta)})$ with $\|x^\delta{}_{l_*(\delta)} - x_0\| \leq R$ and $l_*(\delta)$ chosen according to* (41) *with* (42)*. Assume that* (9) *holds with $c < 1$, $\tilde{c}R < 1$ and that and (η_l) is chosen so that*

$$
\eta_l \geq c_\gamma(\overline{\gamma}_l + \tilde{c}R\gamma_l)^2 \tag{46}
$$

for some $c_\gamma > 0$.

Then the following assertions hold.

(a) *If* $\|F'(x^\dagger)\|^2 \leq e^{-(2p+1)}$, *and the initial error satisfies the source condition*

$$x_0 - x^\dagger = (-\ln(F'(x^\dagger)^* F'(x^\dagger)))^{-p} v \tag{47}$$

for some $v \in X$, $p > 0$, *then*

$$\|x_{l_*}^\delta - x^\dagger\| = O((-\ln\delta)^{-p}) \,, \tag{48}$$

(b) *If the initial error satisfies the source condition*

$$x_0 - x^\dagger = (F'(x^\dagger)^* F'(x^\dagger))^\nu v \tag{49}$$

for some $v \in X$, $\nu \in [0, 1/2]$, *then*

$$\|x_{l_*}^\delta - x^\dagger\| = O\left(\delta^{2\nu/(2\nu+1)}\right) \,. \tag{50}$$

Remark 3. The restriction $\nu \leq 1/2$ corresponds to the well known saturation phenomenon of Tikhonov regularization at $\nu = 1$ which is shifted to $\nu = 1/2$ if the discrepancy principle is used for regularization parameter choice.

Proof. We make use of Jensen's inequality

$$\phi\left(\frac{\int \chi \, d\mu}{\int d\mu}\right) \leq \frac{\int \phi \circ \chi \, d\mu}{\int d\mu} \,, \tag{51}$$

that holds for a convex function $\phi \in C^2(\alpha, \beta)$ with $\alpha, \beta \in \mathbb{R} \cup \{\pm\infty\}$, a finite measure μ on some measure space Ω and $\chi \in L^1(\Omega, d\mu)$ satisfying $\alpha \leq \chi \leq \beta$ almost everywhere $d\mu$. Also, we invoke the spectral theorem for bounded self-adjoint operators, that implies existence of a locally compact space Ω, a positive Borel measure $\overline{\mu}$ on Ω, a unitary map

$$W : L^2(\Omega, d\overline{\mu}) \longrightarrow X,$$

and a real-valued function $\lambda \in C(\Omega)$, such that

$$W^{-1}F'(x^\dagger)^*F'(x^\dagger)W = M_\lambda,$$

where $M_\lambda \in L(L^2(\Omega, d\overline{\mu}))$ is the multiplication operator defined by $(M_\lambda\psi)(\omega) := \lambda(\omega)\psi(\omega)$ for $\psi \in L^2(\Omega, d\overline{\mu})$ and $\omega \in \Omega$, cf., e.g. Section 8.1 in [43].

For $\chi = \phi^{-1} = f^2$ with

$$f(\lambda) = (-\ln\lambda)^{-p} \text{ in case (a)} \quad \text{and} \quad f(\lambda) = \lambda^\nu \text{ in case (b)},$$

it is readily checked that ϕ is convex and strictly monotonically increasing, since

$$\begin{gathered} f(\lambda) > 0\,, \quad f'(\lambda) > 0\,, \quad f(\lambda)f''(\lambda) + (f'(\lambda))^2 \leq 0\,, \\ \text{for all } \lambda \in (0, \|F'(x^\dagger)\|^2]\,, \end{gathered}$$

and

$$\begin{aligned} \phi'(\xi) &= \frac{1}{(f^2)'(\lambda)} = \frac{1}{2f(\lambda)f'(\lambda)} > 0\,, \\ \phi''(\xi) &= -\frac{(f^2)''(\lambda)}{((f^2)'(\lambda))^2} = -\frac{2f(\lambda)f''(\lambda) + 2f'(\lambda)^2}{((f^2)'(\lambda))^2} \geq 0 \\ &\qquad \text{with } \lambda = (f^2)^{-1}(\xi)\,,\ \text{for all } \lambda \in (0, \|F'(x^\dagger)\|^2]\,. \end{aligned}$$

Moreover, $\chi \in L^1(\Omega, d\mu)$, where we set $d\mu = W^{-1}(x^\delta{}_{l_*(\delta)} - x^\dagger)d\overline{\mu}$. Therewith, and with the notation

$$K = F'(x^\dagger)\,, \quad e = x^\delta{}_{l_*(\delta)} - x^\dagger$$

we get from (44) and (47) or (49), respectively,

$$\begin{aligned} \|e\|^2 &\leq \frac{1}{1-\tilde{c}R}\langle x_0 - x^\dagger, e\rangle = \frac{1}{1-\tilde{c}R}\langle v, f(K^*K)e\rangle \\ &\leq \frac{1}{1-\tilde{c}R}\|v\|\sqrt{\int\Big(f(\lambda)W^{-1}e\Big)^2 d\overline{\mu}} = \frac{1}{1-\tilde{c}R}\|v\|\,\|e\|\sqrt{\frac{\int f^2(\lambda)\,d\mu}{\int\,d\mu}} \\ &\leq \frac{1}{1-\tilde{c}R}\|v\|\,\|e\|\sqrt{f^2\left(\frac{\int\lambda\,d\mu}{\int\,d\mu}\right)} = \frac{1}{1-\tilde{c}R}\|v\|\,\|e\|f\left(\frac{\|Ke\|^2}{\|e\|^2}\right) \end{aligned}$$

This together with the estimate

$$\begin{aligned}\|Ke\| &\le \|(I-Q_{l_*(\delta)})Ke\| + \|Q_{l_*(\delta)}(F({x^\delta}_{l_*(\delta)})-y^\delta)\| \\ &\quad + \Big(\frac{c}{\eta_{l_*(\delta)}}(\|Q_{l_*(\delta)}(F({x^\delta}_{l_*(\delta)})-y^\delta)\|+C_Q\delta)^2+\tilde{c}\gamma_{l_*(\delta)}\|e\|^2+C_Q\delta\Big) \\ &\le \underbrace{(\overline{\gamma}_{l_*(\delta)}+\tilde{c}R\gamma_{l_*(\delta)})}_{=:\tilde{\gamma}_{l_*(\delta)}}\|e\| + \underbrace{\left\{\overline{\tau}+C_Q+\frac{c}{\underline{\tau}}(\overline{\tau}+C_Q)^2\right\}}_{=:\tilde{C}}\delta\end{aligned}$$

yields

$$\|e\| \le \frac{1}{1-\tilde{c}R}\|v\|f\left(\frac{(\tilde{\gamma}_{l_*(\delta)}\|e\|+\tilde{C}\delta)^2}{\|e\|^2}\right) \tag{52}$$

In case (a), we choose some constant $\hat{C} > \overline{\tau}/c_\gamma$ and distinguish between two subcases: If $(\tilde{\gamma}_{l_*(\delta)}\|e\|+\tilde{C}\delta)^2/\|e\|^2 \ge \hat{C}\delta$, then $\|e\|$ has to lie between the two roots of the quadratic polynomial $\zeta \mapsto \left(\hat{C}-\left(\tilde{\gamma}^2_{l_*(\delta)}\right)/\delta\right)\zeta^2 - 2\tilde{C}\tilde{\gamma}_{l_*(\delta)}\zeta - \tilde{C}^2\delta$,

$$\zeta_{1,2} = \frac{\tilde{C}\left(\tilde{\gamma}_{l_*(\delta)} \pm \sqrt{\hat{C}\delta}\right)}{\hat{C}-\tilde{\gamma}^2_{l_*(\delta)}/\delta}\,,$$

hence by (41, 46)

$$\|e\| \le \frac{\tilde{C}\left(\sqrt{\overline{\tau}/c_\gamma}+\sqrt{\hat{C}}\right)}{\hat{C}-\overline{\tau}/c_\gamma}\sqrt{\delta} \le C(-\ln\delta)^{-p}$$

for all δ sufficiently small. If $(\tilde{\gamma}_{l_*(\delta)}\|e\|+\tilde{C}\delta)^2/\|e\|^2 \le \hat{C}\delta$ we can make use of the monotonicity of f in (52) to conclude

$$\|e\| \le \frac{1}{1-\tilde{c}R}\|v\|f\left(\hat{C}\delta\right) \le C\|v\|f\left(\delta\right)\,.$$

for some constant $C>0$ independent of δ.

In case (b), (52) directly implies

$$\|e\| \le \frac{1}{1-\tilde{c}R}\|v\| \left(\frac{(\tilde{\gamma}_{l_*(\delta)}\|e\| + \tilde{C}\delta)^2}{\|e\|^2} \right)^\nu ,$$

hence if $\tilde{\gamma}_{l_*(\delta)} \le \delta/\|e\|$, we already arrive at (50). Otherwise, i.e., in case $\tilde{\gamma}_{l_*(\delta)} > \delta/\|e\|$, we get

$$\|e\| \le \frac{1}{1-\tilde{c}R}\|v\|(1+\tilde{C})^{2\nu}\tilde{\gamma}_{l_*(\delta)}^{2\nu} \le \frac{1}{1-\tilde{c}R}\|v\|(1+\tilde{C})^{2\nu}\left(\frac{\overline{\tau}}{c_\gamma}\delta\right)^\nu ,$$

where we have used (41, 46) in the last estimate. Hence, by $\nu \le 1/2$ we can conclude (50). □

5. Numerical Experiments. To illustrate the convergence rates result as well as the assertions on the exactness of Φ_α, we implemented a gradient method with an Armijo stepsize choice for the unconstrained minimization of Φ_α and carried out computational tests for the test example from [22] using the same starting function x_0 and three different exact solutions $x^\dagger$ such that

(a) only a very weak source condition

(b) a logarithmic source condition

(c) a Hölder type source condition

is satisfied. In the context of (a), we wish to mention that it can be shown that there is always an index function $\varphi : (0,a] \to (0,\infty)$ such that a general source condition

$$x_0 - x^\dagger = \varphi(F'(x^\dagger)^* F'(x^\dagger))v$$

holds for some $v \in X$, and refer, e.g., to [3, 29, 30, 32] for convergence rates results under such general source conditions.

The details on the test example are the following: Consider the nonlinear integral equation

$$\int_0^1 \frac{1}{\sqrt{1+(t-s)^2+x(s)^2}}\, ds = y(t) \quad t \in [0,1] , \tag{53}$$

δ (per cent)	l_*	$\frac{\Vert x_{l_*}^\delta - x^\dagger\Vert}{\Vert x^\dagger\Vert}$
0.5	7	0.3174
1	6	0.3211
2	5	0.3249
4	4	0.3285
8	3	0.3482

TABLE 1: Mean relative errors for $\delta = 0.5, \dots, 8$ per cent noise for example (a).

on the spaces $X = Y = L^2(0,1)$ with

$$\text{(a) } x^\dagger(s) = \begin{cases} 2 & s \in \left[0, \frac{1}{2}\right] \\ 1 & s \in \left(\frac{1}{2}, 1\right] \end{cases},$$

$$y(t) = \ln\left(\frac{\left(\frac{1}{2} - t + \sqrt{\left(\frac{1}{2} - t\right)^2 + 5}\right)\left(1 - t + \sqrt{(1-t)^2 + 2}\right)}{\left(-t + \sqrt{t^2 + 5}\right)\left(\frac{1}{2} - t + \sqrt{\left(\frac{1}{2} - t\right)^2 + 2}\right)}\right)$$

$$\text{(b) } x^\dagger(s) = \begin{cases} 2(1-s) & s \in \left[0, \frac{1}{2}\right] \\ 1 & s \in \left(\frac{1}{2}, 1\right] \end{cases},$$

$$y(t) = \ln\left(\frac{\left(\frac{1}{2} - \tilde{t} + \sqrt{\left(\frac{1}{2} - \tilde{t}\right)^2 + a^2}\right)\left(1 - t + \sqrt{(1-t)^2 + 2}\right)}{\left(-\tilde{t} + \sqrt{\tilde{t}^2 + a^2}\right)\left(\frac{1}{2} - t + \sqrt{\left(\frac{1}{2} - t\right)^2 + 2}\right)}\right)$$

with $a^2 := 1 + \frac{4}{5}(t-1)^2$, $\tilde{t} = \frac{t+4}{\sqrt{5}}$

(c) $x^\dagger(s) \equiv 1$,

$$y(t) = \ln\left(\frac{\left(1 - t + \sqrt{(1-t)^2 + 2}\right)}{\left(-t + \sqrt{t^2 + 2}\right)}\right)$$

Note that these explicit formulas for the solutions and the data allow us to avoid an inverse crime. The forward operator F defined by the integral operator in (53) is continuous and compact from $L^2(0,1)$ into itself. Since the kernel is analytic, the inverse problem is exponentially ill-posed. Thus we expect that

(a) since $x_0 - x^\dagger$ is not even continuous, only a very weak source condition

(b) since $x_0 - x^\dagger$ is continuous but not differentiable a logarithmic source condition

(c) since $x_0 - x^\dagger$ is analytic, a Hölder type source condition

is satisfied. The optimal exponents p and ν in (7) for (b) and (c) are not known analytically, however, by the numerical results (see Figure 1) we conjecture that $p = 1/4$ and $\nu \geq 1/2$.

Note that since we partially consider nondifferentiable solutions $x^\dagger$ here, F is not differentiable at these points. Nevertheless, due to the practical relevance of discontinuous solutions in view of the relation of (53) to geophysical applications (cf., e.g., [47]) we study the numerical behaviour of the proposed method for this kind of examples as well.

The operators Q_l are defined by L^2 projection to the piecewise linear functions with breakpoints at the equidistant nodes $t_1, \ldots, t_{k_l}$, where $k_l = 2^{l-1}$, the tolerances η_l were chosen as $\eta_l = 0.01 * 2^{-(l-5)}$. Discretization in preimage space was done with piecewise linear splines on an equidistant grid of size 0.05 that turned out to be sufficiently fine in all of our computations. To simulate perturbed data, we added random noise at the levels given in the tables to the exact data y. Each test was carried out at least five times and the tables display the mean values of the resulting relative errors, while the plots show the results of the actual realizations. As an initial guess x_0 we used the constant function with value 5. This function was also the starting point of the gradient method with Armijo linesearch. Figure 1 displays the relative errors plotted over noise levels in per cent and Tables 1-3 show the mean relative errors with different noise levels for the test cases (a), (b), (c), respectively. The convergence behavior appears to correspond to (a) a very slow rate, (b) an $O((-\ln\delta)^{-p})$ rate, and (c) an $O\left(\delta^{2\nu/(2\nu+1)}\right)$ rate. In the latter two cases the rates $O((-\ln\delta)^{-0.25})$, $O(\sqrt{\delta})$ show up as dashed lines in the respective plots for comparison.

6. Conclusions and Remarks. In this paper we show convergence rates for a regularization method based on a sequence of finite dimensional constrained minimization problems. Moreover, we provide an exact penalty function that enables one to treat each of these constrained problems via unconstrained minimization.

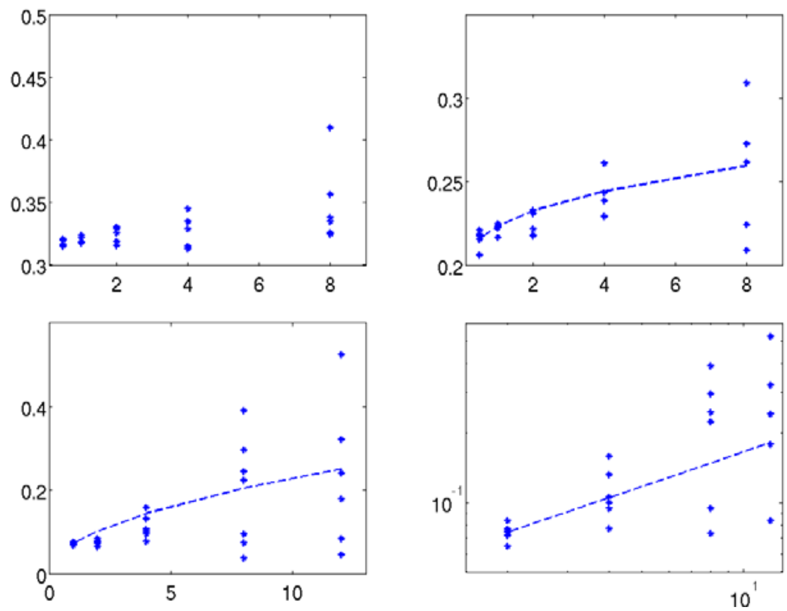

FIGURE 1: Relative errors for different moise levels δ for examples (a) (top left), (b) (top right), (c) (bottom left), (c) in a doubly logarithmic plot (bottom right) and expected rates (dashed).

δ (per cent)	l_*	$\frac{\lVert x_{l_*}^{\delta} - x^\dagger \rVert}{\lVert x^\dagger \rVert}$
0.5	7	0.2157
1	6	0.2223
2	5	0.2242
4	4	0.2403
8	3	0.3002

TABLE 2: Mean relative errors for $\delta = 0.5, \dots, 8$ per cent noise for example (b).

δ (per cent)	l_*	$\frac{\Vert x^\delta_{l_*} - x^\dagger \Vert}{\Vert x^\dagger \Vert}$
2	5	0.0743
4	4	0.1114
8	3	0.1944
12	2	0.2327

TABLE 3: Mean relative errors for $\delta = 2, \dots, 12$ per cent noise for example (c).

The present approach allows one to improve the multilevel method proposed in [22] in the sense that the globally convergent gradient method with Armijo stepsize choice for the exact penalty function Φ_α allows for larger distances between subsequent levels of discretization. Note that especially the closeness assumptions made to guarantee nonnegativity of the Lagrange multiplier in [22] were quite strict. This can now be completely avoided.

Of course we are aware of the fact that carrying out a gradient method for an exact penalty function is usually not the method of choice in numerical optimization, since the small (but finite!) parameter α deteriorates the condition of the problem, which leads to a slow convergence of the gradient method. Computational efficiency of the multilevel method can be considerably improved by using an SQP method. Here again the exact penalty function Φ_α can be made use of to define a merit function in the stepsize control.

Acknowledgment. The author wishes to express her thanks to the referee whose detailed report has led to a considerably improved presentation of the paper.

REFERENCES

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

2. D. Bertsekas, *Constrained Optimization and Lagrange Multiplier Methods*, Academic Press, (1982); republished by Athena Scientific, (1996).

3. A. Böttcher, B. Hofmann, U. Tautenhahn, M. Yamamoto, *Convergence rates* *for Tikhonov regularization from different kinds of smoothness conditions*, Applicable Analysis **85** (2006), 555–578.

4. G. Bruckner, S. Prößdorf, G. Vainikko, *Error bounds of discretization methods for boundary integral equations with noisy data*, Applicable Analysis **63** (1996), 25–37.

5. G. Chavent, K. Kunisch, *On weakly nonlinear inverse problems*, SIAM J. on Applied Mathematics **56** (1996), 542–572.

6. H.W. Engl, M. Hanke, A. Neubauer, *Regularization of Inverse Problems*, Kluwer, Dordrecht, (1996).

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

8. H.W. Engl, A. Neubauer, *On projection methods for solving linear ill–posed problems*, in: A. Vogel, ed., Model Optimization in Exploration Geophysics, Vieweg, Braunschweig, (1987), 73–92.

9. F. Facchinei, S. Lucidi, *Quadratically and superlinearly convergent algorithms for the solution of inequality constrained minimization problems*, J. Opt. Theo. Appl. **85** (1994), 265–289.

10. C. Geiger, C. Kanzow, *Numerische Verfahren zur Lösung unrestringierter Optimierungsaufgaben*, Springer, New York, (1999).

11. C. Geiger, C. Kanzow, *Theorie und Numerik restringierter Optimierungsaufgaben*, Springer, New York, (2002).

12. C.W. Groetsch, *Inverse Problems in Mathematical Sciences*, Vieweg, Braunschweig, (1993).

13. C.W. Groetsch, A. Neubauer, *Convergence of a general projection method for an operator equation of the first kind*, Houston J. Mathem. **14** (1988), 201–207.

14. M. Hanke, *Regularizing properties of a truncated Newton–CG algorithm for nonlinear inverse problems*, Numer. Funct. Anal. Optim. **18** (1997), 971–993.

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

16. M. Hanke, A. Neubauer, O. Scherzer, *A convergence analysis of the Landweber iteration for nonlinear ill-posed problems*, Numer. Math. **72** (1995), 21–37.

17. T. Hohage, *Logarithmic convergence rates of the iteratively regularized Gauß-Newton method for an inverse potential and an inverse scattering problem*, Inverse Problems **13** (1997), 1279–299.

18. T. Hohage, *Regularization of exponentially ill-posed problems*, Numer. Funct. Anal. Optim. **21** (2000) 439–464.

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

20. B. Blaschke(-Kaltenbacher), A. Neubauer, O. Scherzer, *On convergence rates for the iteratively regularized Gauß–Newton method*, IMA J. Numer. Anal. **17** (1997), 421–436.

21. B. Kaltenbacher, *A projection-regularized Newton method for nonlinear ill-posed problems with application to parameter identification problems with finite element discretization*, SIAM J.Numer.Anal. **37** (2000) 1885–1908.

22. B. Kaltenbacher, *Towards global convergence for strongly nonlinear ill-posed problems via a regularizing multilevel approach* Numer. Func. Anal.. Optim. **27** (2006), 637–665.

23. B. Kaltenbacher, A. Neubauer, O. Scherzer, *Iterative Regularization Methods for Nonlinear Ill-Posed Problems*, to appear.

24. A. Kirsch, *An Introduction to the Mathematical Theory of Inverse Problems*, Springer, New York, (1996).

25. A.K. Louis, *Inverse und schlecht gestellte Probleme*, Teubner, Stuttgart, (1989).

26. S. Lucidi, L. Palagi, M. Roma, *On some properties of quadratic programs with a convex quadratic constraint*, SIAM J. Opt. **8** (1998), 105–122.

27. G.R. Luecke, K.R. Hickey, *Convergence of approximate solutions of an operator equation*, Houston J. Math. **11** (1985), 343–353.

28. P. Mathé, S.V. Pereverzev, *Optimal discretization of inverse problems in Hilbert scales: Regularization and self-regularization of projection methods*, SIAM J. Numer. Anal. **38** (2001), 1999–2021.

29. P. Mathé, S.V. Pereverzev, *Geometry of linear ill-posed problems in variable Hilbert scales*, Inverse Problems **19** (2003), 789–803.

30. M.T. Nair, S. Pereverzev, U. Tautenhahn, *Regularization in Hilbert scales under general smoothing conditions*, Inverse Problems **21** (2005), 1851–1869.

31. V.A. Morozov, *Regularization Methods for Ill-Posed Problems*, CRC Press, Boca Raton, (1993).

32. M.T. Nair, E. Schock, U. Tautenhahn, *Morozov's discrepancy principle under general source conditions*, Zeitschr. Anal. Anw. **22** (2003), 199–214.

33. M.Z. Nashed *Generalized Inverses and Applications*, Academic Press, New York, (1976).

34. M.Z. Nashed, X. Chen, *Convergence of Newton-like methods for singular operator equations using outer inverses*, Numer. Math., **66** (1993), 235–257.

35. F. Natterer, *Regularisierung schecht gestellter Probleme durch Projektionsverfahren*, Numer. Math. **28** (1977), 329–341.

36. F. Natterer, *The Mathematics of Computerized Tomography*, Teubner, Stuttgart, (1986).

37. .V. Pereverzev, S. Prössdorf, *On the characterization of self-regularization properties of a fully discrete projection method for Symm's integral equation* J. Integral Equations Appl. **12** (2000), no. 2, 113–130.

38. R. Plato, G. Vainikko, *On the regularization of projection methods for solving ill-posed problems*, Numer. Math. **28** (1990), 63–79.

39. A.G. Ramm, A.B. Smirnova, *A numerical method for solving nonlinear ill-posed problems*, Nonlinear Funct. Anal. and Optimiz. **20** (1999), 317–332.

40. A. Rieder, *On the regularization of nonlinear ill-posed problems via inexact Newton methods*, Inverse Problems **15** (1999), 309–327.

41. O. Scherzer, *An iterative multi level algorithm for solving nonlinear ill-posed problems*, Numer. Math. **80** (1998), 579–600.

42. T.I. Seidman, C.R. Vogel, *Well-posedness and convergence of some regu-*

larization methods for nonlinear ill-posed problems, Inverse Problems **5** (1989), 227–238.

43. M. Taylor, *Partial Differential Equations*, Vol. **2**, Springer, New York, (1996).

44. A.N. Tikhonov, V.A. Arsenin, *Methods for Solving Ill-Posed Problems*, Nauka, Moskau, (1979).

45. G. Vainikko, U. Hämarik, *Projection methods and self-regularization in ill-posed problems*, Sov. Math. **29** (1985), 1–20. In Russian.

46. G. Vainikko, A.Y. Veterennikov, *Iteration Procedures in Ill–Posed Problems*, Nauka, Moscow, (1986). In Russian.

47. V.V. Vasin, A.L. Ageev, *Ill-Posed Problems with A Priori Information*, VSP, Zeist, (1995).

Department of Mathematics, University of Stuttgart, Germany
Email Address: Barbara.Kaltenbacher@mathematik.uni-stuttgart.de