Regularization by Projection with A Posteriori Discretization Level Choice for Linear and Nonlinear Ill-Posed Problems [1]

Barbara Kaltenbacher

Abstract. In this paper we study the regularization of linear and nonlinear ill-posed operator equations by projection onto finte dimensional spaces with a posteriori chosen space dimension. We show that this results in a regularization method, i.e., a stable solution method for the ill-posed problem, that converges to an exact solution as the data noise level goes to zero, with optimal rates under additional regularity conditions.

Key words. Ill-posed problems, regularization methods, discrepancy principle, convergence rates

AMS(MOS) Subject Classifications: 65J20, 65R30

1 Introduction

The mathematical formulation of inverse problems, as they appear in many different applications in science and industry, often takes the form of a (presumably solvable) operator equation

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

with an, in general, nonlinear continuous operator $F : \mathcal{X} \to \mathcal{Y}$ between Hilbert spaces $\mathcal{X}$, $\mathcal{Y}$, that is ill-posed in the sense that noisy data y^δ with arbitrarily small noise level δ in

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

can lead to large deviations in the solution. Therefore, for their numerical solution, so-called regularization methods (cf. [3, 5, 12, 13, 15, 17, 18]) have to be applied, i.e., approximation strategies that are stable with respect to perturbations in the data and converge to a solution $x^\dagger$ of (1) as the noise level δ in (2) goes to zero.

It is a typical feature of ill-posed problems that approximations $\tilde{x}$ to $x^\dagger$ converge, if at all, only arbitrarily slowly as $\delta \to 0$, and that convergence rates can only be obtained under additional so-called regularity or source wise representation conditions of the type

$$x_0 - x^\dagger \in \mathcal{R}((F'(x^\dagger)^* F'(x^\dagger))^\mu) \ , \tag{3}$$

where x_0 is an initial guess, (in the linear case in (1) one may set $x_0 := 0$,) μ is a real positive exponent, and the operator $(F'(x^\dagger)^* F'(x^\dagger))^\mu$ is defined in the sense of functional calculus (cf., e.g., Section 2.3 in [3]). Since in ill-posed problems the (linearized) forward operator

[1] Supported by the Fonds zur Förderung der wissenschaftlichen Forschung under grant T 7-TEC and project F1308 within Spezialforschungsbereich 13

$F'(x)$ is typically smoothing, (3) can usually be interpreted as some a priori smoothness assumption on the initial difference $x_0 - x^\dagger$ and is the stronger, the higher μ is. The under condition (3) optimal convergence rates are

$$\|\tilde{x} - x^\dagger\| = O(\delta^{\frac{2\mu}{2\mu+1}}) \ , \tag{4}$$

cf. e.g., [3].

In the nonlinear case our convergence analysis will be a local one, since we do not touch the wide field of globalization strategies here, i.e., we assume that our initial guess x_0 is sufficiently close to a solution $x^\dagger$ of (1) and work in a neighborhood $\mathcal{B}_\rho(x_0)$ of x_0. Moreover, the usual nonlinearity conditions on F, applied in convergence proofs for solution methods of well-posed operator equations, such as a Lipschitz condition on its Fréchet derivative, implying an estimate

$$\|F(x) - F(\bar{x}) - F'(x)(x - \bar{x})\| \leq \tfrac{L}{2} \|x - \bar{x}\|^2$$

of the first order Taylor remainder, are not appropriate for ill-posed problems, though. Instead, one makes use of Taylor remainder estimates like

$$\|F(x) - F(\bar{x}) - T(x - \bar{x})\| \leq c_{tc} \|T(x - \bar{x})\| \quad \forall x, \bar{x} \in \mathcal{B}_\rho(x_0) \tag{5}$$

(the so-called tangential cone condition, cf. [16]) or, alternatively, range invariance conditions of the type

$$F(x) - F(\bar{x}) = TR(x, \bar{x}) \qquad \|R(x, \bar{x}) - I\| \leq c_{ri} \ , \quad \forall x, \bar{x} \in \mathcal{B}_\rho(x_0) \ , \tag{6}$$

which means that the range of divided difference operators $DF(x, \bar{x})$ in

$$F(x) - F(\bar{x}) = DF(x, \bar{x})(x - \bar{x}) \tag{7}$$

remains unchanged in $\mathcal{B}_\rho(x_0)$. Here T in (5) or (6) is a known bounded linear operator that can be seen as some approximation to $F'(x)$ in $\mathcal{B}_\rho(x_0)$, which we must be able to make use of in our computations; note that in contrast to this, the linear operators $R(x, \bar{x})$ in (6) need not be known explicitely. The divided difference operators $DF(x, \bar{x})$ are not uniquely defined by (7); if F is Fréchet differentiable in $\mathcal{B}_\rho(x_0)$, then one may set $DF(x, \bar{x}) = \int_0^1 F'(\bar{x} + \theta(x - \bar{x}))d\theta$. Conditions like (5), (6) have been based upon by convergence proofs of several different regularization methods for nonlinear ill-posed problems, and have been verified for many applications; see, e.g., [1], [7], [8], and note that (6) is closely related to the affine covariant Lipschitz condition in [2] (cf. [9]).

Regularization methods for ill-posed operator equations (1) such as the well-known Tikhonov regularization, where $\tilde{x}$ is a global minimizer of

$$\|F(\tilde{x}) - y^\delta\|^2 + \alpha \|\tilde{x} - x_0\|^2 = \min! \ ,$$

are usually defined in an infinite-dimensional setting and have to be discretized for an implementation. Now, since finite dimensional problems are always well-posed in the sense of

stable data dependence, (as soon as existence and uniqueness of a solution are guaranteed), one could also think of stabilizing an ill-posed problem *by* discretization (cf. [4], [6], [14]).

A motivation for considering regularization solely by discretization, is that it seems to be a natural way of treating inverse problems, avoiding the introduction of possibly artificial regularization terms; by the choice of appropriate projection bases it can be expected to be adaptable to a variety of different applications. After all, since in practice many ill-posed problem (partially without being recognized as such) are just discretized and the discretized version is solved, it also seems to be an interesting task to investigate to what extent this approach can be mathematically justified.

An analysis of discretization methods for linear ill-posed problems (cf. e.g., Section 3.3 in [3], see also Section 2 below,) shows that, while discretization in pre-image space does not generically converge, finite dimensional projection in image space yields a regularization method. Therefore we here prefer the latter approach. After having considered the application of linear regularization by projection in a Newton method for nonlinear ill-posed problems in [10], we here propose a direct generalization of regularization by image space projection from the known linear theory to nonlinear problems satisfying (5) or (6), and show its well-definedness and, with a priori chosen discretization level, its convergence, that takes optimal rates (4) under respective source conditions (3) (see Section 3). In practice the exponent μ in the source condition (3), that contains the unknown solution $x^\dagger$, can in general not be expected to be known explicitely, though. Therefore it is of major interest also to have an optimal *a posteriori* discretization strategy, that yields convergence and optimal rates (4) without using knowledge on μ. It is supposed to be the main new contribution of the present paper, to provide such an a posteriori discretization choice, namely a discrepancy type principle, and to show its convergence as well as optimal rates (4) both in the linear (see Section 2) and in the nonlinear (see Section 3) case.

In the following we will use the notation c or C, for positive constants that are typically "small" or "large" (in the sense of "not necessarily small") respectively; they can have different values whenever they appear in the text, but they are introduced in such a way that no confusion should occur.

2 Linear Problems

In order to define regularization by projection in image space for a linear ill-posed operator equation

$$Tx = y \tag{8}$$

where $T : \mathcal{X} \mapsto \mathcal{Y}$ is a compact linear operator between Hilbert spaces $\mathcal{X}$, $\mathcal{Y}$, and y lies in the (non-closed) range of T, we consider, for a sequence of finite dimensional nested subspaces $\mathcal{Y}_n$ of $\mathcal{Y}$, whose union is dense in $\mathcal{Y}$:

$$\mathcal{Y}_1 \subseteq \mathcal{Y}_2 \subseteq \ldots \subseteq \mathcal{Y} \quad \wedge \quad \overline{\bigcup_{n\in\mathbb{N}} \mathcal{Y}_n} = \mathcal{Y} \ , \tag{9}$$

the sequence of projected equations

$$\mathcal{Q}_n T x = \mathcal{Q}_n y^\delta \tag{10}$$

where $\mathcal{Q}_n$ denotes the $\mathcal{Y}$-orthogonal projection onto $\mathcal{Y}_n$

$$\mathcal{Q}_n := Proj^{\mathcal{Y}}_{\mathcal{Y}_n}$$

and y^δ are the given noisy data with noise level δ in (2). A regularized approximation of the exact solution $x^\dagger = T^\dagger y \in \mathcal{N}(T)^\perp$ is then given by the best-approximate solution of (10)

$$x_n := (\mathcal{Q}_n T)^\dagger \mathcal{Q}_n y^\delta \ . \tag{11}$$

For exact data $y^\delta = y$, ($\delta = 0$,) we have

$$x_n = \mathcal{P}_n x^\dagger \ ,$$

where

$$\mathcal{P}_n := Proj^{\mathcal{X}}_{\mathcal{X}_n} \ ,$$
$$\mathcal{X}_n := T^* \mathcal{Y}_n,$$

i.e., the x_n are just the projections of $x^\dagger$ onto the finite dimensional spaces $\mathcal{X}_n$, whose union is dense in $\mathcal{N}(T)^\perp$, and therefore converge to $x^\dagger$ as n goes to infinity, with the speed of convergence under a source condition

$$x^\dagger \in \mathcal{R}((T^*T)^\mu) \tag{12}$$

for some $\mu > 0$, being determined by the operator norms

$$\overline{\gamma}_n(\mu) := \|(I - \mathcal{P}_n)(T^*T)^\mu\| \ .$$

Convergence of the images Tx_n to y is proportional to

$$\overline{\overline{\gamma}}_n := \|(I - \mathcal{Q}_n)T\|$$

and, under a source condition (12), to $\overline{\overline{\gamma}}_n \overline{\gamma}_n(\mu)$. By the interpolation inequality and functional calculus (cf. e.g., Section 2.3 in [3]) we get

$$\begin{aligned} \overline{\gamma}_n(\mu) &= \|(T^*T)^\mu [I - \mathcal{P}_n]\| \le (\|(T^*T)^{\frac{1}{2}} [I - \mathcal{P}_n]\|)^{2\mu} \\ &= (\|T[I - \mathcal{P}_n]\|)^{2\mu} = (\|[I - \mathcal{Q}_n]T[I - \mathcal{P}_n]\|)^{2\mu} \le \overline{\overline{\gamma}}_n^{2\mu} \end{aligned} \tag{13}$$

for

$$\mu \le \frac{1}{2} \ , \tag{14}$$

where we have used the identity

$$\mathcal{Q}_n T [I - \mathcal{P}_n] = 0 \ . \tag{15}$$

Non-vanishing data noise is in the worst case amplified by a factor of

$$\frac{1}{\underline{\gamma}_n} := \|(\mathcal{Q}_n T)^\dagger \mathcal{Q}_n\| \ ,$$

which is finite for fixed n, so that x_n stably depends on the data, but, by the unboundedness of $T^\dagger$, goes to infinity as $n \to \infty$.

The total error, that consists of the sum of the approximation error and the propagated data noise, can be estimated by

$$\|x_n - x^\dagger\| \leq \|[I - \mathcal{P}_n]x^\dagger\| + \frac{\delta}{\underline{\gamma}_n} \tag{16}$$

whence an order optimal a priori choice $N = N(\delta)$ of the discretization parameter n is obviously given by

$$N(\delta) \to \infty \ \wedge \ \frac{\delta}{\underline{\gamma}_{N(\delta)}} \to 0 \text{ as } \delta \to 0 \tag{17}$$

without source condition and

$$\underline{\gamma}_{N(\delta)} \overline{\gamma}_{N(\delta)}(\mu) \sim \delta$$

if (12) holds, and yields convergence of $x_{N(\delta)}$ to $x^\dagger$ as $\delta \to 0$ with a rate $\|x_{N(\delta)} - x^\dagger\| = O(\overline{\gamma}_{N(\delta)}(\mu))$ under a source condition.

We just mention in passing (cf. [3], where also the results on regularization by projection quoted so far can be found) that truncated SVD, i.e., the choice $\mathcal{Y}_n := \text{span}\{v_1, \ldots, v_n\}$, where v_j are eigenfunctions of TT^* corresponding to eigenvalues in decreasing order, is optimal both with respect to convergence speed and stability. Since eigenfunctions are often hard to obtain explicitely, we are here more interested in the usual piecewise polynomial projection bases such as finite elements, though.

The fact that $\underline{\gamma}_n$ goes to zero as n tends to infinity, reflects the ill-posedness of (8) or, in other words, the smoothing property of the forward operator and its adjoint. The latter, on the other hand, in combination with an approximation property of the finite dimensional spaces $\mathcal{Y}_n$, also makes the quantities $\overline{\gamma}_n(\mu)$ and $\overline{\overline{\gamma}}_n$ go to zero at respective rates. The following assumptions quantify these properties in the form in which we will need it for our convergence analysis.

For this purpose, we denote by h_n some discretization parameter (think, e.g., of a mesh size) that geometrically goes to zero as $n \to \infty$:

$$h_n = C\sigma^n$$

for some $\sigma \in (0,1)$, and, by H^{t-p} a Hilbert space that contains $\mathcal{Y}$:

$$H^s := \mathcal{X} \quad H^{t-p} \supseteq H^t := \mathcal{Y} \ .$$

Assumption 1. (Smoothing property of T^*:)
Let, for some $p > 0$,

$$\|T^*\|_{H^{t-p} \to H^s} \leq C \ ,$$

Assumption 2. (Approximation property of $\mathcal{Y}_n$)
Let, for p as in Assumption 1

$$\|I - \mathcal{Q}_n\|_{H^t \to H^{t-p}} \leq C h_n^p \ ,$$

By these Assumptions (whose notation deliberately suggests Sobolev spaces and piecewise polynomial approximation) we get the following asymptotics for the real sequences $\overline{\overline{\gamma}}_n$, $\overline{\gamma}_n(\mu)$:

$$\overline{\overline{\gamma}}_n = \|T^*[I - \mathcal{Q}_n]\|_{H^t \to H^s} \leq \|T^*\|_{H^{t-p} \to H^s} \|I - \mathcal{Q}_n\|_{H^t \to H^{t-p}} \leq \overline{\overline{C}} h_n^p \ , \tag{18}$$

for some constant $\overline{\overline{C}}$ and, by (13)

$$\overline{\gamma}_n(\mu) = O(h_n^{2p\mu}) \tag{19}$$

Observe that

$$\begin{array}{rcl} \underline{\gamma}_n & = & \inf_{w \in \mathcal{Y}_n} \frac{\|w\|}{\|(Q_n T)^\dagger Q_n w\|} = \inf_{w \in \mathcal{Y}_n} \frac{\|Q_n T (Q_n T)^\dagger Q_n w\|}{\|(Q_n T)^\dagger Q_n w\|} \\ & = & \inf_{\tilde{w} \in (Q_n T)^\dagger \mathcal{Y}_n = \mathcal{X}_n} \frac{\|Q_n T \tilde{w}\|}{\|\tilde{w}\|} = \inf_{v \in \mathcal{Y}_n} \frac{\|Q_n T T^* v\|}{\|T^* v\|} \\ & \geq & \inf_{v \in \mathcal{Y}_n} \frac{\langle Q_n T T^* v, v \rangle}{\|T^* v\| \, \|v\|} = \inf_{v \in \mathcal{Y}_n} \frac{\|T^* v\|}{\|v\|} =: \underline{\underline{\gamma}}_n \ , \end{array}$$

where we have used Cauchy-Schwarz inequality at the beginning of the last line, so that, in order to estimate $\underline{\gamma}_n$ from below, we can make use of some kind of inverse inequality on $\mathcal{Y}_n$:

Assumption 3. (Inverse inequality)
Let, for some $\tilde{p} > 0$,

$$\sup_{v \in \mathcal{Y}_n} \frac{\|v\|_Y}{\|T^* v\|_X} \leq C h_n^{-\tilde{p}}$$

which yields

$$\underline{\gamma}_n \geq \underline{\underline{\gamma}}_n \geq \underline{\underline{c}} h_n^{\tilde{p}} \tag{20}$$

for some $\underline{\underline{c}} > 0$.

Note that, by

$$\begin{array}{rcl} \underline{\underline{\gamma}}_{n+1} & \leq & \inf_{v \in \mathcal{Y}_{n+1} : T^* v \in \mathcal{X}_n^\perp} \frac{\|T^* v\|}{\|v\|} = \inf_{v \in \mathcal{Y}_{n+1} : T^* v \in \mathcal{X}_n^\perp} \frac{\|[I - \mathcal{P}_n] T^* v\|}{\|v\|} \\ & = & \inf_{v \in \mathcal{Y}_{n+1} : T^* v \in \mathcal{X}_n^\perp} \frac{\|[I - \mathcal{P}_n] T^* [I - \mathcal{Q}_n] v\|}{\|v\|} \leq \|T^*[I - \mathcal{Q}_n]\| = \overline{\overline{\gamma}}_n \end{array} \tag{21}$$

where we have used (15), one always gets

$$p \leq \tilde{p} \ .$$

If we have

$$p = \tilde{p} \tag{22}$$

which we assume to hold in the following, and which, when thinking of $\mathcal{Y}_n$ e.g. as finite element spaces, seems to be realistic, then the quotient $\frac{\overline{\overline{\gamma}}_n}{\underline{\underline{\gamma}}_n}$ and therefore $\frac{\overline{\overline{\gamma}}_n}{\underline{\gamma}_n}$ is uniformly bounded

$$\frac{\overline{\overline{\gamma}}_n}{\underline{\gamma}_n} \leq \frac{\overline{\overline{\gamma}}_n}{\underline{\underline{\gamma}}_n} \leq \frac{\overline{\overline{C}}}{\underline{\underline{c}}} \,, \tag{23}$$

which is crucial for guaranteeing optimal convergence rates as $\delta \to 0$. Moreover it yields a uniform bound for the operator $T(\mathcal{Q}_n T)^\dagger \mathcal{Q}_n$

$$\|T(\mathcal{Q}_n T)^\dagger \mathcal{Q}_n\|^2 = \|[I - \mathcal{Q}_n]T(\mathcal{Q}_n T)^\dagger \mathcal{Q}_n\|^2 + \|\mathcal{Q}_n T(\mathcal{Q}_n T)^\dagger \mathcal{Q}_n\|^2 \leq (\frac{\overline{\overline{\gamma}}_n}{\underline{\gamma}_n})^2 + 1 \tag{24}$$

since $\mathcal{Q}_n T(\mathcal{Q}_n T)^\dagger = \mathcal{Q}_n$ i.e., the quite useful fact that the bad condition of the approximation $(\mathcal{Q}_n T)^\dagger \mathcal{Q}_n$ to $T^\dagger$ can be compensated by an application of the smoothing operator T; note that this is trivial when reversing the order of application of $(\mathcal{Q}_n T)^\dagger \mathcal{Q}_n$ and T in (24):

$$\|(\mathcal{Q}_n T)^\dagger \mathcal{Q}_n T\| = \|\mathcal{P}_n\| = 1$$

and that also in the original infinite-dimensional setting one has $\|TT^\dagger\| = \|Proj^{\mathcal{Y}}_{\overline{\mathcal{R}(T)}}|_{\mathcal{D}(T^\dagger)}\| = 1$. Similarly one obtains

$$\begin{aligned} \|T(\mathcal{Q}_n T)^\dagger \mathcal{Q}_n - I\|^2 &= \|[I - \mathcal{Q}_n]T(\mathcal{Q}_n T)^\dagger \mathcal{Q}_n - [I - \mathcal{Q}_n]\|^2 + \|\mathcal{Q}_n T(\mathcal{Q}_n T)^\dagger \mathcal{Q}_n - \mathcal{Q}_n\|^2 \\ &\leq (\frac{\overline{\overline{\gamma}}_n}{\underline{\gamma}_n} + 1)^2 \end{aligned} \tag{25}$$

The total error estimate under a source condition then becomes

$$\|x_n - x^\dagger\| \leq C(h_n^{2\mu p} + h_n^{-p}\delta)$$

and therefore, together with the a priori choice

$$h_{N(\delta)}^{(1+2\mu)p} \sim \delta \tag{26}$$

yields the optimal convergence rates (4).

Since in practice the exponent μ is rarely known exactly, it is of interest to have a posteriori strategies for the choice of N, that do not need explicit knowlegde of μ and nevertheless yield optimal rates (4). As the following Theorem shows, this can be achieved by the well known discrepancy principle, that selects the regularization parameter in such a way, that the regularized equation is as stable as possible, under the restriction that the residual is of the order of magnitude of the data noise level at most:

$$N(\delta) := \min\{n \in \mathbb{N}_0 \mid \|Tx_n - y^\delta\| \leq \tau\delta\} \in \mathbb{N}_0 \cup \{\infty\} \tag{27}$$

for some fixed constant $\tau > 1$.

Theorem 1. *Let Assumptions 1, 2, 3 with (22) hold and fix* $\tau > \frac{\overline{\overline{C}}}{\underline{\underline{c}}} + 1$, *where* $\overline{\overline{C}}$, $\underline{\underline{c}}$ *are the constants in (18), (20).*

Then for any $\delta \le \frac{\|Tx_0 - y\|}{\tau+1}$ *and for any data* y^δ *with (2), a finite* $N(\delta)$, *with (27) exists, where* x_n *is defined by (11).*

For any family of data y^δ *with (2),* $x_{N(\delta)}$ *converges to* $x^\dagger$ *as* $\delta \to 0$.

Under an additional source condition (12) for $\mu \le \frac{1}{2}$, *the optimal convergence rate (4) is achieved.*

Proof. By (10), the definition (11) of x_n and the bound (25) we obtain, for all $n \in \mathbb{N}$:

$$\begin{aligned} \|Tx_n - y^\delta\| &= \|[I - \mathcal{Q}_n](Tx_n - y^\delta)\| \\ &= \|[I - \mathcal{Q}_n](T(\mathcal{Q}_n T)^\dagger \mathcal{Q}_n - I)y^\delta\| \\ &\le \|[I - \mathcal{Q}_n](T(\mathcal{Q}_n T)^\dagger \mathcal{Q}_n - I)(y^\delta - y)\| + \|[I - \mathcal{Q}_n]T[\mathcal{P}_n - I]x^\dagger\| \\ &\le (\frac{\overline{\overline{C}}}{\underline{\underline{c}}} + 1)\delta + \overline{\overline{\gamma}}_n \|[\mathcal{P}_n - I]x^\dagger\| \ , \end{aligned} \tag{28}$$

where the last term on the right hand side goes to zero as $n \to \infty$ and is hence smaller than $(\tau - \frac{\overline{\overline{C}}}{\underline{\underline{c}}} - 1)\delta$ for some sufficiently large n_0, that is therefore, by estimate (28), a finite upper bound for the $N(\delta)$ defined in (27). Moreover we have, for $n := N(\delta) - 1$

$$\tau\delta < \|Tx_{N(\delta)-1} - y^\delta\|$$

and therefore by (28)

$$\overline{\overline{\gamma}}_{N(\delta)-1} \|[\mathcal{P}_{N(\delta)-1} - I]x^\dagger\| > (\tau - \frac{\overline{\overline{C}}}{\underline{\underline{c}}} - 1)\delta \ .$$

Hence $N_{apr}(\delta) < \infty$ defined a priori by

$$N_{apr}(\delta) := \min\{n \mid \overline{\overline{\gamma}}_n \|[\mathcal{P}_n - I]x^\dagger\| \le \tilde{\tau}\delta\} \tag{29}$$

with

$$\tilde{\tau} := \tau - \frac{\overline{\overline{C}}}{\underline{\underline{c}}} - 1 \tag{30}$$

is always larger or equal $N(\delta)$, so that we can use an inverse inequality on $\mathcal{Y}_{N_{apr}(\delta)} \supseteq \mathcal{Y}_{N(\delta)}$ to obtain, by (10) with $n := N_{apr}(\delta)$, by the definition of $N(\delta)$, and by the error estimate (16),

$$\begin{aligned} \|x_{N(\delta)} - x^\dagger\| &\le \|(\mathcal{Q}_{N_{apr}(\delta)}T)^\dagger \mathcal{Q}_{N_{apr}(\delta)} T(x_{N(\delta)} - x_{N_{apr}(\delta)})\| + \|x_{N_{apr}(\delta)} - x^\dagger\| \\ &\le \underline{\gamma}^{-1}_{N_{apr}(\delta)} \|\mathcal{Q}_{N_{apr}(\delta)} T(x_{N(\delta)} - x_{N_{apr}(\delta)})\| + \|x_{N_{apr}(\delta)} - x^\dagger\| \qquad (31) \\ &= \underline{\gamma}^{-1}_{N_{apr}(\delta)} \|\mathcal{Q}_{N_{apr}(\delta)}(Tx_{N(\delta)} - y^\delta)\| + \|x_{N_{apr}(\delta)} - x^\dagger\| \\ &\le \underline{\gamma}^{-1}_{N_{apr}(\delta)} \tau\delta + \|x_{N_{apr}(\delta)} - x^\dagger\| \\ &\le \|[\mathcal{P}_{N_{apr}(\delta)} - I]x^\dagger\| + (\tau+1)\underline{\gamma}^{-1}_{N_{apr}(\delta)}\delta \\ &\le \tilde{\tau}\overline{\overline{\gamma}}^{-1}_{N_{apr}(\delta)}\delta + (\tau+1)\underline{\gamma}^{-1}_{N_{apr}(\delta)}\delta \ , \qquad (32) \end{aligned}$$

where in the last line we have estimated $\|[\mathcal{P}_{N_{apr}(\delta)} - I]x^\dagger\|$ by inserting into the definition of $N_{apr}(\delta)$. Now, by using the estimates

$$\overline{\overline{\gamma}}^{-1}_{N_{apr}(\delta)} \leq \underline{\gamma}^{-1}_{N_{apr}(\delta)+1} \leq \frac{\overline{\overline{C}}}{\underline{\underline{c}}} \sigma^{-2p} \overline{\overline{\gamma}}^{-1}_{N_{apr}(\delta)-1} \ , \tag{33}$$

$$\underline{\gamma}^{-1}_{N_{apr}(\delta)} \leq \frac{\overline{\overline{C}}}{\underline{\underline{c}}} \sigma^{-p} \overline{\overline{\gamma}}^{-1}_{N_{apr}(\delta)-1} \ , \tag{34}$$

where $\overline{\overline{C}}$ and $\underline{\underline{c}}$ are the constants in (18) and (20), respectively, we have used (21), as well as the fact that the quotient $\frac{h_{n-1}}{h_n}$ is bounded from above by σ^{-1}, and by taking into account minimality in the definition of $N_{apr}(\delta)$, that yields

$$\overline{\overline{\gamma}}_{N_{apr}(\delta)-1} \|[\mathcal{P}_{N_{apr}(\delta)-1} - I]x^\dagger\| > \tilde{\tau}\delta \ , \tag{35}$$

we arrive at

$$\|x_{N(\delta)} - x^\dagger\| \leq C \|[\mathcal{P}_{N_{apr}(\delta)-1} - I]x^\dagger\| \tag{36}$$

for some constant C.

To prove convergence in the general case without source condition, we use an argument similar to the one in [7], consider an arbitrary sequence of noisy data $(y_i^\delta)_{i\in\mathbb{N}}$ satisfying (2) with $\delta := \delta_i \to 0$ as $i \to \infty$, and set $x^i := x_{N(\delta_i)}$ according to (11), (27). Then, for any subsequence $(x^l)_{l\in\mathbb{N}}$ of $(x^i)_{i\in\mathbb{N}}$, we distinguish between two cases:

If $N_{apr}(\delta_l) \to \infty$ then, by (36), x^l converges to $x^\dagger$.

If $N_{apr}(\delta_l)$ does not converge to infinity, then there exists a finite $\bar{N} \in \mathbb{N}$ and a subsequence $(\delta^m)_{m\in\mathbb{N}}$ of $(\delta^l)_{l\in\mathbb{N}}$, such that for all $m \in \mathbb{N}$: $N_{apr}(\delta_m) = \bar{N}$. By (32)

$$\|x^m - x^\dagger\| \leq (\tilde{\tau}\overline{\overline{\gamma}}^{-1}_{\bar{N}} + (\tau+1)\underline{\gamma}^{-1}_{\bar{N}})\delta_m \ \to 0 \text{ as } m \to \infty \ . \tag{37}$$

Combination of these two cases to a subsequence-subsequence argument, yields convergence of all of x^i to $x^\dagger$.

Under a source condition (12), we can combine the estimates (18), (19), (35) to

$$h^{(1+2\mu)p}_{N_{apr}(\delta)-1} > c\delta$$

for some constant $c > 0$, so that (32) together with (20), (21), (33), (34), and $\frac{h_{n-1}}{h_n} \leq \sigma^{-1}$, yields

$$\|x_{N(\delta)} - x^\dagger\| \leq C\delta^{-\frac{1}{(1+2\mu)}}\delta \ .$$

□

3 Nonlinear Problems

A straightforward generalization of regularization by projection to nonlinear problems satisfying (5) or (6), is to approximate an exact solution $x^\dagger$ of the operator equation (1) by a solution x_n of the finite dimensional problem

$$\mathcal{Q}_n F(x_n) = \mathcal{Q}_n y^\delta \ , \quad x_n \in x_0 + \mathcal{X}_n := x_0 + T^* \mathcal{Y}_n \ , \tag{38}$$

where T is the linear operator appearing in the respective nonlinearity condition (5) or (6), that also enables us to carry over the definitions of the quantities $\overline{\overline{\gamma}}_n$, $\overline{\gamma}_n(\mu)$, $\underline{\gamma}_n$ as well as Assumptions 1, 2, 3 from the previous subsection. The initial guess x_0, that takes over the special role that zero plays in the linear situation, is in the nonlinear case supposed to be sufficiently close to $x^\dagger$, and the difference $x_0 - x^\dagger$ is assumed to be orthogonal to the nullspace of T,

$$x_0 - x^\dagger \in \mathcal{N}(T)^\perp \tag{39}$$

so that it can be expected to be well approximated by functions in $\mathcal{X}_n = T^* \mathcal{Y}_n$. This does not mean a smoothness restriction, since it is just equivalent to the case $\mu \to 0^+$ in the source condition, that here takes the form

$$x_0 - x^\dagger \in \mathcal{R}((T^*T)^\mu) \ . \tag{40}$$

Under the nonlinearity condition (5) or (6), assumption (39) uniquely selects a solution of (1) (for the case of (5), see also Proposition 2.1 in [7]):

Lemma 1. *Let (6) with $c_{ri} < \frac{1}{2}$ or (5) with $c_{tc} < 1$ hold and assume that for $x, \bar{x} \in \mathcal{X} \cap \mathcal{B}_\rho(x_0)$: $F(x) = F(\bar{x})$ and $x - \bar{x} \in \mathcal{N}(T)^\perp$. Then $x = \bar{x}$.*

Proof. If (6) holds then

$$0 = F(x) - F(\bar{x}) = TR(x, \bar{x})(x - \bar{x})$$

implies that $R(x, \bar{x})(x - \bar{x})$ is in $\mathcal{N}(T)$ and therefore, since $x - \bar{x} \in \mathcal{N}(T)^\perp$,

$$R(x, \bar{x})(x - \bar{x}) = Proj_{\mathcal{N}(T)} R(x, \bar{x})(x - \bar{x}) = Proj_{\mathcal{N}(T)} (R(x, \bar{x}) - I)(x - \bar{x}) \ ,$$

from which we conclude that

$$\|x - \bar{x}\| = \|R(x, \bar{x})^{-1} Proj_{\mathcal{N}(T)} (R(x, \bar{x}) - I)(x - \bar{x})\| \le \frac{c_{ri}}{1 - c_{ri}} \|x - \bar{x}\| \ ,$$

hence, since $\frac{c_{ri}}{1 - c_{ri}} < 1$, $\|x - \bar{x}\| = 0$.

In case of (5), (for completeness of exposition repeating the argument of [7],) we have

$$0 = \|F(x) - F(\bar{x})\| \ge \frac{1}{1 - c_{tc}} \|T(x - \bar{x})\|$$

so that $x - \bar{x} \in \mathcal{N}(T) \cap \mathcal{N}(T)^\perp = \{0\}$. □

The regularized approximation x_n according to (38) is in fact well-defined and yields error estimates similar to those obtained in the linear case:

Theorem 2. *Let (2), (23),*

$$\bar{n} := \min\{n \in \mathbb{N} \mid (C_1 \underline{\gamma}_n^{-1}\delta + C_2 \|[I - \mathcal{P}_n](x_0 - x^\dagger)\|) + \|\mathcal{P}_n(x_0 - x^\dagger)\| > \rho\} - 1 \tag{41}$$

and either (6) with

$$c_{ri} < 1 \ ,$$

$C_1 := \frac{1}{1-c_{ri}}$, $C_2 := \frac{1+c_{ri}}{1-c_{ri}}$, *or (5) with*

$$(\frac{\overline{\overline{C}}}{\underline{\underline{c}}} + 2)c_{tc} < 1 \tag{42}$$

and $C_1 := \frac{\sqrt{(\overline{\overline{C}}/\underline{\underline{c}})^2+1}}{1-\sqrt{(\overline{\overline{C}}/\underline{\underline{c}})^2+1}\ c_{tc}}$, $C_2 := \frac{\sqrt{(\overline{\overline{C}}/\underline{\underline{c}})^2+1}(\overline{\overline{C}}/\underline{\underline{c}})(1+c_{tc})}{1-\sqrt{(\overline{\overline{C}}/\underline{\underline{c}})^2+1}\ c_{tc}}$ *hold. Then for all* $n \le \bar{n}$ *a solution to (38) exists in* $\mathcal{B}_\rho(x_0)$ *and*

$$\|x_n - x^\dagger\| \le C(\|[I - \mathcal{P}_n](x_0 - x^\dagger)\| + \frac{\delta}{\underline{\gamma}_n}) \ . \tag{43}$$

Proof. To show existence of a solution to (38) we consider the fixed point iteration

$$x^{l+1} := x^l + (\mathcal{Q}_n T)^\dagger \mathcal{Q}_n(y^\delta - F(x^l)) \ , \quad x^0 := x_0 + \mathcal{P}_n(x^\dagger - x_0) \ . \tag{44}$$

Obviously $x^0 - x^\dagger = [I - \mathcal{P}_n](x_0 - x^\dagger)$, all the iterates are elements of $x_0 + \mathcal{X}_n$, and

$$x^{l+1} - x^l = (\mathcal{Q}_n T)^\dagger \mathcal{Q}_n(F(x^l) - F(x^{l-1}) - T(x^l - x^{l-1})) \ . \tag{45}$$

If, for fixed $l \in \mathbb{N}$, we assume that for all $i \le l$, x^i is in $\mathcal{B}_{\rho^i}(x^0) \subseteq \mathcal{B}_\rho(x_0)$ with $\rho^i := \frac{1-c_{ri}^i}{1-c_{ri}}(\underline{\gamma}_n^{-1}\delta + (1 + c_{ri})\|[I - \mathcal{P}_n](x_0 - x^\dagger)\|)$, and (6) holds, then

$$\begin{aligned}
\|x^{l+1} - x^l\| &= \|(\mathcal{Q}_n T)^\dagger \mathcal{Q}_n T(R(x^l, x^{l-1}) - I)(x^l - x^{l-1})\| \\
&\le c_{ri}\|x^l - x^{l-1}\| \\
&\le c_{ri}^l\|(\mathcal{Q}_n T)^\dagger \mathcal{Q}_n(y^\delta - F(x^0))\| \\
&\le c_{ri}^l(\underline{\gamma}_n^{-1}\delta + \|(\mathcal{Q}_n T)^\dagger \mathcal{Q}_n T R(x^\dagger, x^0)(x^\dagger - x^0)\|) \\
&\le c_{ri}^l(\underline{\gamma}_n^{-1}\delta + (1 + c_{ri})\|[I - \mathcal{P}_n](x_0 - x^\dagger)\|) \ ,
\end{aligned}$$

whence, by an inductive argument together with (41), all the iterates remain in $\mathcal{B}_\rho(x_0)$. Moreover, for any two different $l \le \tilde{l} \in \mathbb{N}$ we have, by the triangle inequality,

$$\|x^l - x^{\tilde{l}}\| \le \sum_{i=l}^{\tilde{l}-1} c_{ri}^{i-1}(\underline{\gamma}_n^{-1}\delta + (1 + c_{ri})\|[I - \mathcal{P}_n](x_0 - x^\dagger)\|) \to 0 \text{ as } l \to \infty \ ,$$

so that $(x^l)_{l\in\mathbb{N}} \subseteq x_0 + \mathcal{X}_n \cap \mathcal{B}_\rho(x_0)$ is a Cauchy sequence and therefore convergent to an $\bar{x}_n \in x_0 + \mathcal{X}_n \cap \mathcal{B}_\rho(x_0)$. Taking limits on both sides of (44), together with the continuity of F yields

$$(\mathcal{Q}_n T)^\dagger \mathcal{Q}_n (y^\delta - F(\bar{x}_n)) = 0 \ ,$$

from which, applying $\mathcal{Q}_n T$ and taking into account that $\mathcal{Q}_n T (\mathcal{Q}_n T)^\dagger = \mathcal{Q}_n$ we conclude that $\bar{x}_n$ solves (38).

Under condition (5), considering the images under T of differences between subsequent iterates yields, by (24), (45), and if $x^{l-1}, x^l \in \mathcal{B}_\rho(x_0)$,

$$\|T(x^{l+1} - x^l)\| \leq \sqrt{(\tfrac{\overline{\overline{C}}}{\underline{\underline{c}}})^2 + 1}\ c_{tc} \|T(x^l - x^{l-1})\| \leq (\sqrt{(\tfrac{\overline{\overline{C}}}{\underline{\underline{c}}})^2 + 1}\ c_{tc})^l \sqrt{(\tfrac{\overline{\overline{C}}}{\underline{\underline{c}}})^2 + 1} \|y^\delta - F(x^0)\| \ ,$$

so that by

$$\|y^\delta - F(x^0)\| \leq \delta + (1 + c_{tc}) \|T(x^0 - x^\dagger)\| = \delta + (1 + c_{tc}) \|[I - \mathcal{Q}_n] T [I - \mathcal{P}_n](x_0 - x^\dagger)\|$$

where we have used (15), we get

$$\begin{aligned} \|x^{l+1} - x^l\| &= \|(\mathcal{Q}_n T)^\dagger \mathcal{Q}_n T (x^{l+1} - x^l)\| \leq \underline{\gamma}_n^{-1} \|T(x^{l+1} - x^l)\| \\ &\leq \underline{\gamma}_n^{-1} (\sqrt{(\tfrac{\overline{\overline{C}}}{\underline{\underline{c}}})^2 + 1} c_{tc})^l \sqrt{(\tfrac{\overline{\overline{C}}}{\underline{\underline{c}}})^2 + 1} (\delta + (1 + c_{tc}) \overline{\overline{\gamma}}_n \|[I - \mathcal{P}_n](x_0 - x^\dagger)\|) \ . \end{aligned}$$

Now, analogously to the previous case, we can show that the sequence $(x^l)_{l\in\mathbb{N}}$ remains in $\mathcal{B}_\rho(x_0)$ and converges to an $\bar{x}_n$ that solves (38).

The estimates of the total error follow from the identities

$$\begin{aligned} \mathcal{P}_n(x_n - x^\dagger) &= (\mathcal{Q}_n T)^\dagger \mathcal{Q}_n (y^\delta - y \ + \ F(x^\dagger) - F(x_n) - T(x^\dagger - x_n)) \\ [I - \mathcal{P}_n](x_n - x^\dagger) &= [I - \mathcal{P}_n](x_0 - x^\dagger) \ . \end{aligned}$$

They imply, in case of condition (6),

$$\begin{aligned} \|x_n - x^\dagger\| &= \|(\mathcal{Q}_n T)^\dagger \mathcal{Q}_n (y^\delta - y) - \mathcal{P}_n (R(x_n, x^\dagger) - I)(x_n - x^\dagger) + [I - \mathcal{P}_n](x_0 - x^\dagger)\| \\ &\leq \underline{\gamma}_n^{-1} \delta + c_{ri} \|x_n - x^\dagger\| + \|[I - \mathcal{P}_n](x_0 - x^\dagger)\| \ , \end{aligned}$$

and therefore

$$\|x_n - x^\dagger\| \leq \tfrac{1}{1 - c_{ri}} (\underline{\gamma}_n^{-1} \delta + \|[I - \mathcal{P}_n](x_0 - x^\dagger)\|) \ . \tag{46}$$

If (5) holds then

$$\begin{aligned} \|x_n - x^\dagger\| &= \|(\mathcal{Q}_n T)^\dagger \mathcal{Q}_n \big(y^\delta - y + F(x^\dagger) - F(x_n) - T(x^\dagger - x_n)\big) + [I - \mathcal{P}_n](x_0 - x^\dagger)\| \\ &\leq \underline{\gamma}_n^{-1} (\delta + c_{tc} \|T(x^\dagger - x_n)\|) + \|[I - \mathcal{P}_n](x_0 - x^\dagger)\| \ , \end{aligned}$$

where

$$\begin{aligned} \|T(x^\dagger - x_n)\| &\leq \tfrac{1}{1 - c_{tc}} \|F(x^\dagger) - F(x_n)\| \\ &\leq \tfrac{1}{1 - c_{tc}} (\delta + \| \underbrace{y^\delta - F(x_n)}_{= [I - \mathcal{Q}_n](y^\delta - F(x_n))} \|) \\ &\leq \tfrac{1}{1 - c_{tc}} (2\delta + \|[I - \mathcal{Q}_n](F(x^\dagger) - F(x_n)\|) \\ &\leq \tfrac{1}{1 - c_{tc}} (2\delta + \|[I - \mathcal{Q}_n] T (x^\dagger - x_n)\| + c_{tc} \|T(x^\dagger - x_n)\|) \ , \end{aligned}$$

so that

$$\|T(x^\dagger - x_n)\| \leq \frac{1}{1-2c_{tc}}(2\delta + \overline{\overline{\gamma}}_n \|x^\dagger - x_n\|)$$

and therefore, by (23),

$$\|x_n - x^\dagger\| \leq \tfrac{1}{1-2c_{tc}}\underline{\gamma}_n^{-1}\delta + \tfrac{\overline{\overline{C}}/\underline{\underline{c}}\ c_{tc}}{1-2c_{tc}}\|x_n - x^\dagger\| + \|[I-\mathcal{P}_n](x_0 - x^\dagger)\| \ ,$$

which by (42) yields

$$\|x_n - x^\dagger\| \leq \tfrac{1}{1-(\overline{\overline{C}}/\underline{\underline{c}}+2)c_{tc}}\underline{\gamma}_n^{-1}\delta + \tfrac{1-2c_{tc}}{1-(\overline{\overline{C}}/\underline{\underline{c}}+2)c_{tc}}\|[I-\mathcal{P}_n](x_0 - x^\dagger)\| \ . \tag{47}$$

□

Remark 1. Analogously to the estimates in the proof of Theorem 2, one sees that for fixed n, x_n is stable with respect to data perturbations.

By the estimate (43), the a priori choice (17), and, under a source condition, (26), yields convergence and optimal rates. Again we can make use of a discrepancy principle to obtain an optimal a posteriori chosen $N(\delta)$:

$$N(\delta) := \min\{n \in \mathbb{N}_0 \mid \|F(x_n) - y^\delta\| \leq \tau\delta\} \in \mathbb{N}_0 \cup \{\infty\} \tag{48}$$

for fixed $\tau > 1$.

Theorem 3. *Let Assumptions 1, 2, 3 with (22) hold as well as either (6) with $c_{ri} < \frac{1}{2}$ or (5) with $c_{tc} < \frac{1}{\overline{\overline{C}}/\underline{\underline{c}}+2}$ hold and fix $\tau > \frac{\overline{\overline{C}}}{\underline{\underline{c}}(1-c_{ri})} + 1$ in the first or $\tau > \frac{1+\overline{\overline{C}}/\underline{\underline{c}}}{1-(\overline{\overline{C}}/\underline{\underline{c}}+2)c_{tc}}$ in the second case.*

Then for any $\delta \leq \frac{\|F(x_0)-y\|}{\tau+1}$ and for any data y^δ with (2), a finite $N(\delta) \leq \bar{n}$, with (48) exists, where x_n is defined by (38), and $\bar{n}$ by (41).

For any family of data y^δ with (2), $x_{N(\delta)}$ converges to $x^\dagger$ as $\delta \to 0$.

Under an additional source condition (40) for $\mu \leq \frac{1}{2}$, the optimal convergence rate (4) is achieved.

Proof. Since the proof goes along the lines of the proof of Theorem 1, it suffices to replace $[I-\mathcal{P}_n]x^\dagger$ by $[I-\mathcal{P}_n](x^\dagger - x_0)$ whenever it appears and to clarify the points that differ due to the nonlinearity:

Consider first the case of (6). Then we replace (28) in the proof of Theorem 1, by

$$\begin{aligned} \|F(x_n) - y^\delta\| &= \|[I-\mathcal{Q}_n](F(x_n) - y^\delta)\| \\ &\leq \delta + \|[I-\mathcal{Q}_n]TR(x_n, x^\dagger)(x_n - x^\dagger)\| \\ &\leq \delta + \overline{\overline{\gamma}}_n(1+c_{ri})\|x_n - x^\dagger\| \\ &\leq (\tfrac{\overline{\overline{C}}}{\underline{\underline{c}}(1-c_{ri})} + 1)\delta + \tfrac{1+c_{ri}}{1-c_{ri}}\overline{\overline{\gamma}}_n\|[I-\mathcal{P}_n](x_0 - x^\dagger)\| \ , \end{aligned} \tag{49}$$

for all $n \le \bar{n}$, where $\bar{n}$ is as in Theorem 2 and we have used (46). In the nonlinear case we have to make sure that not only $N(\delta) < \infty$ but also $N(\delta) \le \bar{n}$. For all $n \le N_{apr}(\delta)$ defined by (29) with $\tilde{\tau} := (\tau - \frac{\overline{\overline{C}}}{\underline{\underline{c}}(1-c_{ri})} - 1)\frac{1-c_{ri}}{1+c_{ri}}$ (and $x^\dagger$ replaced by $x^\dagger - x_0$) we get, similarly to (34) and using (21)

$$\underline{\gamma}_n^{-1} \le \frac{\overline{\overline{C}}}{\underline{\underline{c}}} \sigma^{-2p} \overline{\overline{\gamma}}_{n-2}^{-1} \le \frac{\overline{\overline{C}}}{\underline{\underline{c}}} \sigma^{-2p} \underline{\gamma}_{n-1}^{-1}$$

and therefore, with C_1, C_2 as in Theorem 2,

$$\begin{aligned} & C_1 \underline{\gamma}_n^{-1} \delta + C_2 \, \|[I - \mathcal{P}_n](x_0 - x^\dagger)\| + \|\mathcal{P}_n(x_0 - x^\dagger)\| \\ & \le C_1 \frac{\overline{\overline{C}}}{\underline{\underline{c}}} \sigma^{-2p} \tilde{\tau}^{-1} \, \|[I - \mathcal{P}_{n-1}](x_0 - x^\dagger)\| + C_2 \, \|[I - \mathcal{P}_n](x_0 - x^\dagger)\| + \|\mathcal{P}_n(x_0 - x^\dagger)\| \ , \end{aligned}$$

which is smaller than ρ for sufficiently small $\|x_0 - x^\dagger\|$, so that we can conclude $N_{apr}(\delta) \le \bar{n}$. Like in the proof of Theorem 1, it follows from the definitions of $N(\delta)$ and $N_{apr}(\delta)$, together with estimate (49) , that

$$N(\delta) \le N_{apr}(\delta) \le \bar{n} \ .$$

For proceeding further analogously to the proof of Theorem 1, we have to be able to replace (32) by

$$\|x_{N(\delta)} - x^\dagger\| \le C \overline{\overline{\gamma}}_{N_{apr}(\delta)}^{-1} \delta \ , \tag{50}$$

with some constant C. For this purpose, observe that under condition (6)

$$\begin{aligned} \|\mathcal{Q}_{N_{apr}(\delta)}(F(x_{N(\delta)}) - F(x_{N_{apr}(\delta)}))\| &= \|\mathcal{Q}_{N_{apr}(\delta)} TR(x_{N(\delta)}, x_{N_{apr}(\delta)})(x_{N(\delta)} - x_{N_{apr}(\delta)})\| \\ &\ge \underline{\tilde{\gamma}}_{N_{apr}(\delta)} \|x_{N(\delta)} - x_{N_{apr}(\delta)}\| \ , \end{aligned}$$

where

$$\begin{aligned} \underline{\tilde{\gamma}}_{N_{apr}(\delta)} &:= \inf_{w \in \mathcal{Y}_{N_{apr}(\delta)}} \frac{\|\mathcal{Q}_{N_{apr}(\delta)} TR(x_{N(\delta)}, x_{N_{apr}(\delta)}) T^* w\|}{\|T^* w\|} \\ &\ge \inf_{w \in \mathcal{Y}_{N_{apr}(\delta)}} \frac{\langle \mathcal{Q}_{N_{apr}(\delta)} TR(x_{N(\delta)}, x_{N_{apr}(\delta)}) T^* w, w \rangle}{\|w\| \, \|T^* w\|} \\ &= \inf_{w \in \mathcal{Y}_{N_{apr}(\delta)}} \frac{\|T^* w\|^2 - \langle [I - R(x_{N(\delta)}, x_{N_{apr}(\delta)})] T^* w, T^* w \rangle}{\|w\| \, \|T^* w\|} \\ &\ge (1 - c_{ri}) \underline{\underline{\gamma}}_{N_{apr}(\delta)} \ . \end{aligned}$$

Due to this fact, and by (38) with $n := N_{apr}(\delta)$, (48), (43) and (29) we get

$$\begin{aligned} \|x_{N(\delta)} - x^\dagger\| &\le \|x_{N(\delta)} - x_{N_{apr}(\delta)}\| + \|x_{N_{apr}(\delta)} - x^\dagger\| \\ &\le \tfrac{1}{1-c_{ri}} \underline{\underline{\gamma}}_{N_{apr}(\delta)}^{-1} \tau \delta + C \big(\underbrace{\|[I - \mathcal{P}_{N_{apr}(\delta)}](x_0 - x^\dagger\|}_{\le \tilde{\tau} \overline{\overline{\gamma}}_{N_{apr}(\delta)}^{-1} \delta} + \underline{\gamma}_{N_{apr}(\delta)}^{-1} \delta \big) \ , \end{aligned}$$

which, by $\underline{\gamma}_{N_{apr}(\delta)}^{-1} \le \frac{\overline{\overline{C}}}{\underline{\underline{c}}} \overline{\overline{\gamma}}_{N_{apr}(\delta)}^{-1}$, yields (50).

If (5) holds, then, instead of (28) we estimate

$$\begin{aligned} \|F(x_n) - y^\delta\| &= \|[I - \mathcal{Q}_n]\big(y - y^\delta \ + \ T(x_n - x^\dagger) \ + \ F(x_n) - F(x^\dagger) - T(x_n - x^\dagger)\big)\| \\ &\le \delta + \overline{\overline{\gamma}}_n \, \|x_n - x^\dagger\| + \tfrac{c_{tc}}{1 - c_{tc}} \, \|F(x_n) - F(x^\dagger)\| \end{aligned}$$

and therefore, by (47),

$$\begin{aligned}\|F(x_n)-y^\delta\| &\leq \tfrac{1-c_{tc}}{1-2c_{tc}}\Big(\tfrac{1}{1-c_{tc}}\delta+\overline{\overline{\gamma}}_n\|x_n-x^\dagger\|\Big)\\ &\leq \tfrac{1+\overline{\overline{C}}/\underline{\underline{c}}}{1-(\overline{\overline{C}}/\underline{\underline{c}}+2)c_{tc}}\delta+\tfrac{1-c_{tc}}{1-(\overline{\overline{C}}/\underline{\underline{c}}+2)c_{tc}}\overline{\overline{\gamma}}_n\|[I-\mathcal{P}_n](x_0-x^\dagger)\|\ .\end{aligned}$$

Analogously to the previous case we can show that, if $\|x_0-x^\dagger\|$ is sufficiently small, $N(\delta)\leq N_{apr}(\delta)\leq\bar{n}$, where we set $\tilde{\tau}:=\tau-\frac{1+\overline{\overline{C}}/\underline{\underline{c}}}{1-(\overline{\overline{C}}/\underline{\underline{c}}+2)c_{tc}}$. To obtain (50) under condition (5), observe that

$$\|x_{N(\delta)}-x_{N_{apr}(\delta)})\|\leq\underline{\gamma}^{-1}_{N_{apr}(\delta)}\|\mathcal{Q}_{N_{apr}(\delta)}T(x_{N(\delta)}-x_{N_{apr}(\delta)})\| \tag{51}$$

with

$$\begin{aligned}&\|\mathcal{Q}_{N_{apr}(\delta)}T(x_{N(\delta)}-x_{N_{apr}(\delta)})\|\\ &\leq\|\mathcal{Q}_{N_{apr}(\delta)}(F(x_{N(\delta)})-F(x_{N_{apr}(\delta)}))\|+c_{tc}\|T(x_{N(\delta)}-x_{N_{apr}(\delta)})\|\\ &\leq\|\mathcal{Q}_{N_{apr}(\delta)}(F(x_{N(\delta)})-F(x_{N_{apr}(\delta)}))\|\\ &\quad+c_{tc}(\overline{\overline{\gamma}}_{N_{apr}(\delta)}\|x_{N(\delta)}-x_{N_{apr}(\delta)})\|+\|\mathcal{Q}_{N_{apr}(\delta)}T(x_{N(\delta)}-x_{N_{apr}(\delta)})\|\end{aligned}$$

and therefore

$$\begin{aligned}&\|\mathcal{Q}_{N_{apr}(\delta)}T(x_{N(\delta)}-x_{N_{apr}(\delta)})\|\\ &\leq\tfrac{1}{1-c_{tc}}(\|\mathcal{Q}_{N_{apr}(\delta)}(F(x_{N(\delta)})-F(x_{N_{apr}(\delta)}))\|+c_{tc}\overline{\overline{\gamma}}_{N_{apr}(\delta)}\|x_{N(\delta)}-x_{N_{apr}(\delta)})\|)\ ;\end{aligned}$$

inserting this into (51) and using the triangle inequality yields

$$\begin{aligned}\|x_{N(\delta)}-x^\dagger\| &\leq \|x_{N(\delta)}-x_{N_{apr}(\delta)}\|+\|x_{N_{apr}(\delta)}-x^\dagger\|\\ &\leq \tfrac{1}{1-(\frac{\overline{\overline{C}}}{\underline{\underline{c}}}+1)c_{tc}}\underline{\gamma}^{-1}_{N_{apr}(\delta)}\|\mathcal{Q}_{N_{apr}(\delta)}(F(x_{N(\delta)})-F(x_{N_{apr}(\delta)}))\|+\|x_{N_{apr}(\delta)}-x^\dagger\|\ .\end{aligned}$$

The rest follows analogously to the case of (6). □

4 Conclusions

For a numerical implementation of regularization by projection, a finite-dimensional system of linear (10) or nonlinear (38) equations has to be solved. The hierarchical structure of these systems with nested projection spaces $\mathcal{Y}_n$ suggests to use multi-grid methods, that are extremely efficient in the numerical solution of well-posed problems such as elliptic PDEs or second kind integral equations. In systems of equations arising from the discretization of ill-posed problems, as it is the case here, the behavior of multi-grid methods is different, though, and special care has to be taken in the use of appropriate smoothers. We therefore refer to [11], where we propose a nested multi-grid iteration for the approximate solution of discretized ill-posed problems and study its convergence with a priori and posteriori chosen finest grid level.

References

[1] P. Deuflhard, H.W. Engl, O. Scherzer, *A convergence analysis of iterative methods for the solution of nonlinear ill-posed problems under affinely invariant conditions*, Inverse Problems 14 (1998), 1081-1106.

[2] P. Deuflhard, G. Heindl, *Affine invariant convergence theorems for Newton's method and extensions to related methods*, SIAM J. Numer. Anal. 16 (1979), 1-10.

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

[4] 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.

[5] C.W. Groetsch, *Inverse Problems in Mathematical Sciences*, Vieweg, Braunschweig, 1993.

[6] 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.

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

[8] sc B. Blaschke–Kaltenbacher, H.W. Engl, *Regularization methods for nonlinear ill–posed problems with applications to phase reconstruction*, H.W. Engl, A.K. Louis, W.Rundell, eds., Inverse Problems in Medical Imaging and Nondestructive Testing, Springer, Vienna–New York, 1997.

[9] B. Kaltenbacher, *On Broyden's method for nonlinear ill–posed problems*, Numerical Functional Analysis and Optimization 19 (1998), 807-833.

[10] B. Kaltenbacher, *A projection-regularized Newton method for nonlinear ill-posed problems with application to parameter identification problems with finite element discretization*, to appear in SIAM J.Numer.Anal.

[11] B. Kaltenbacher, A multi-grid method with a priori and a posteriori level choice for the regularization of nonlinear ill-posed problems, SFB013-report No.00-??, University of Linz, 2000.

[12] A.K. Louis, *Inverse und schlecht gestellte Probleme*, Teubner, Stuttgart, 1989.

[13] V.A. Morozov, *Regularization Methods for Ill-Posed Problems*, CRC Press, Boca Raton, 1993.

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

[15] F. Natterer, *The Mathematics of Computerized Tomography*, Teubner, Stuttgart, 1986.

[16] O. Scherzer, *Convergence criteria of iterative methods based on Landweber iteration for nonlinear problems*, J. Math. Anal. Appl. 194 (1995), 911-933.

[17] A.N. Tikhonov, V.A. Arsenin, *Methods for Solving Ill-Posed Problems*, Nauka, Moskau, 1979.

[18] G.M. Vainikko, A.Y. Veterennikov, *Iteration Procedures in Ill–Posed Problems*, Nauka, Moscow, 1986. In Russian.