REGULARIZATION BY PROJECTION: APPROXIMATION THEORETIC ASPECTS AND DISTANCE FUNCTIONS

BERND HOFMANN, PETER MATHÉ, AND SERGEI V. PEREVERZEV

Abstract. The authors study regularization of non-linear inverse problems by projection methods. Non-linearity is controlled by some range invariance assumption. Emphasis is on approximation theoretic properties of the discretization which determine the convergence rates. Instead of using source conditions for the true solution to represent the error, the authors show how distance functions with respect to some benchmark smoothness are able to replace this. Some examples indicate how the results can be applied.

Dedicated to Academician M. M. Lavrentiev
on the occasion of his 75^{th} *birthday*

1. Introduction

The focus of the present study is on regularization of non-linear problems between Hilbert spaces X and Y. Such problems are given by a (non-linear) mapping $F\colon \mathcal{D}(F) \to Y$, where $\mathcal{D}(F) \subset X$ is the domain of definition of the mapping F. Specifically we assume that the equation is noisy, i.e.,

$$y^\delta = F(x) + \delta\xi, \tag{1}$$

where $\xi \in Y$ is bounded noise with $\|\xi\| \le 1$, and $\delta > 0$ is the known noise level. In general this equation is numerically not feasible, and discretization is required to treat it. Here we focus on *one-sided* discretization, given through a subspace $Y_n \subset Y$ and respective orthogonal projection $Q_n\colon Y \to Y$, leaving Y_n invariant. Thus, instead of (1) we

Date: April 25, 2007.
B. Hofmann acknowledges support by Deutsche Forschungsgemeinschaft (DFG) under grant HO1454/7-1.

are actually given data

$$Q_n y^\delta = (Q_n F)(x) + \delta Q_n \xi. \tag{2}$$

As typical for non-linear problems, our a priori knowledge about the unknown true solution $x^\dagger \in X$ is given through an *initial guess* x_0 as

$$x^\dagger \in B_r(x_0) \subset \mathcal{D}(F). \tag{3}$$

where $B_r(x_0)$ denotes the ball around x_0 with known radius $r > 0$.

In this study the approximate computation of $x^\dagger$ from noisy data $Q_n y^\delta$ is carried out by discretization only, i.e., by using some projection method (see also [3, Chap. 7], [23, Chap. 3]). Its numerical behavior is determined by several effects, and these will be discussed. Firstly, certain approximation theoretic properties of the design spaces $Y_n \subset Y$ are important and discussed in Section 2. If these are known, then the error of the classical linear projection method can be bounded, and this is briefly sketched in Section 3.

Then, under some *range invariance assumption* of the non-linear mapping F we discuss a modification of the projection method, a fixed point iteration, to be presented and analyzed in Section 4. Error bounds for this non-linear projection method are obtained in Section 5 under some new perspective. In a first step, in § 5.1, such bounds are obtained for some benchmark smoothness, given in terms of a source condition. The main result of the paper refers to the important case that the solution smoothness fails to satisfy the benchmark smoothness. This degree of failure is expressed by some distance function. In that general case § 5.2 presents error bounds based on the corresponding distance function. As shown in the concluding paragraph § 5.3 such distance functions can be verified whenever range inclusions express the interplay between the solution smoothness and smoothing properties of the operator F governing the equation (1).

2. Design spaces

Assumption A.3 below assigns the non-linear mapping F a corresponding compact linear mapping A, and we shall use this linear mapping to design a projection method. Throughout this paper $\mathcal{N}(A)$ and $\mathcal{R}(A)$ denote null-space and range of the linear operator A, respectively. Several approximation theoretic requirements on the chosen subspaces Y_n are important, we refer to [30] for a compound treatment. Such are the *degree of approximation*, i.e., the validity of some Jackson inequality, and the *modulus of injectivity*, i.e., the validity of some Bernstein inequality.

Degree of approximation. It measures the approximation power of some finite dimensional subspace, say $M \subset Y$, with respect to the operator A as

$$\eta(M) := \sup_{\|x\| \leq 1} \operatorname{dist}(Ax, M). \tag{4}$$

In general this may be poor and we need to compare it with its best possible performance. If $\dim(M) \leq n$ then this is given by

$$d_n(A) := \inf_{\dim(M) \leq n} \eta(M), \tag{5}$$

where the infimum runs over all at most n-dimensional subspaces $M \subset Y$. The latter quantity is known to be the n-th *Kolmogorov width* and it equals the $(n+1)$-st singular number s_{n+1} of A.

To establish optimality of the projection method, introduced below we shall require that the design spaces Y_n are order optimal in the following sense.

Assumption A.1 (Jackson inequality)**.** There is a constant $1 \leq C_Q < \infty$ such that

$$\eta(Y_n) \leq C_Q s_{n+1}, \quad n \in \mathbb{N}. \tag{6}$$

For later use we mention that in Hilbert space Y the distance $\operatorname{dist}(y, M)$ of some element $y \in Y$ with respect to a given subspace $M \subset Y$ is given by its orthogonal projection onto M. Hence we have

$$\eta_n := \eta(Y_n) = \|(I - Q_n)A\|, \quad n \in \mathbb{N}. \tag{7}$$

Modulus of injectivity. Given some finite dimensional subspace $M \subset Y$ this is given as

$$\beta(M) := \inf \left\{ \|A^* u\| / \|u\|,\ 0 \neq u \in M \right\}. \tag{8}$$

In order that $\beta(M) > 0$ we need to assume that $\mathcal{N}(A^*) \cap M = \{0\}$, which in turn is equivalent to $M \subset \overline{\mathcal{R}(A)}$, the closure of the range of A in Y. It describes the quality of inversion of the mapping A restricted to the subspace M in the image space. It is clear from this definition that

$$\beta(M) \leq \sup_{\dim(Z) \geq n} \beta(Z) =: b_{n-1}(A^*), \tag{9}$$

where the supremum runs over all at least n-dimensional subspaces $Z \subset Y$. The latter quantity $b_{n-1}(A^*)$ is well known in approximation theory as the $(n-1)$*-st Bernstein width* of A^*. For operators in Hilbert space this coincides with the n-th singular number of A^*, hence $b_{n-1}(A^*) = s_n$. For the design spaces Y_n this amounts to

$$\beta(Y_n) \leq s_n, \quad n \in \mathbb{N}. \tag{10}$$

We shall require that the spaces Y_n are order optimal with respect to the modulus of injectivity.

Assumption A.2 (Bernstein inequality)**.** The design spaces Y_n obey $Y_n \subset \overline{\mathcal{R}(A)}$, and there is a constant $C_B < \infty$ for which

$$s_n \le C_B \beta(Y_n), \quad n \in \mathbb{N}. \tag{11}$$

Remark 1. Although within the present context the Kolmogorov widths and Bernstein widths coincide, the requirements made in Assumptions A.1 and A.2 reflect different aspects of approximation. Typically the validity of some Bernstein inequality reflects smoothness properties of the elements in Y_n, rather than its approximation power.

For many specific approximation methods their degree of approximation and modulus of injectivity (with respect to specific operators) are known up to constants. In the following example we shall see that for regular finite elements with inverse property both the assumptions A.1 and A.2 are fulfilled, provided the operator A^* obeys a step condition.

Example 1 (Finite element approximation)**.** Suppose that the target space Y consists of functions on a (sufficiently smooth) bounded domain $\Omega \subset \mathbb{R}^d$. For such classes of functions we denote by $H^l(\Omega)$, $-\infty < l < \infty$ the usual Sobolev spaces of bounded smoothness l for $l > 0$ and their duals for $l < 0$. We denote the respective norms by $\|y\|_l := \|y\|_{H^l(\Omega)}$. Then we may choose finite element approximations as a design space. We shall not give many details, instead we allow any spaces $S_h^{t,k}(\Omega) \subset H^k(\Omega)$ of finite elements, which constitute *regular (t,k)-systems with the inverse property* in the sense of [1, Chapt. 4], see also [29, § 4]. Specifically we assume the following inequalities to hold for the projection $Q_h \colon L_2(\Omega) \to S_h^{t,k}(\Omega)$:

$$\|(I - Q_h)y\|_{-k} \le C h^k \|y\|, \quad y \in L_2(\Omega), \tag{12}$$

$$\|y\| \le C h^{-k} \|y\|_{-k}, \quad y \in S_h^{t,k}(\Omega), \tag{13}$$

with some $t \ge k > 0$ and a common constant $C < \infty$ which is independent of the mesh size h. For recent treatment of this topic we refer to [12, Chapt. 1, §1.5 and §1.7].

To control the quality of the spaces $S_h^{t,k}(\Omega)$ with respect to the operator A we require a *step condition* to hold for A^*, precisely we require that there is $l > 0$ and there are constants $0 < m \le M < \infty$ for which

$$m\|y\|_{-l} \le \|A^* y\| \le M \|y\|_{-l}, \quad y \in L_2(\Omega). \tag{14}$$

Notice that this implies $\mathcal{N}(A^*) \cap S_h^{t,k}(\Omega) = \{0\}$. If $l \le k$ then it follows from interpolation, that the bounds (12) and (13) extend as

$$\|(I - Q_h)y\|_{-l} \le Ch^l \|y\|, \quad y \in L_2(\Omega),$$
$$\|y\| \le Ch^{-l}\|y\|_{-l}, \quad y \in S_h^{t,k}(\Omega).$$

Consequently, using the step condition (14) this implies

$$\|A^*(I - Q_h)y\| \le M\|(I - Q_h)y\|_{-l} \le MCh^l\|y\|, \quad y \in L_2(\Omega),$$

and

$$\|y\| \le Ch^{-l}\|y\|_{-l} \le \frac{C}{m}h^{-l}\|A^*y\|. \quad y \in S_h^{t,k}(\Omega).$$

In terms of the degree of approximation and the modulus of injectivity this rewrites as

$$\eta(S_h^{t,k}(\Omega)) := \|(I - Q_h)A\| \le MCh^l \tag{15}$$

and respectively

$$\beta(S_h^{t,k}(\Omega)) \ge \frac{m}{C}h^l. \tag{16}$$

For regular meshes the dimensions $\#(S_h^{t,k}(\Omega))$ of the finite element spaces obey $\#(S_h^{t,k}(\Omega)) \asymp h^{-d}$, which follows from volume considerations. Next we observe that as a consequence we obtain for the singular numbers of A that

$$s_{n+1} \le \eta(S_h^{t,k}(\Omega)) \le MCh^l \asymp n^{-l/d}, \quad s_n \ge \beta(S_h^{t,k}(\Omega)) \asymp n^{-l/d}$$

Thus, we finally obtain $s_n \asymp \eta(S_h^{t,k}(\Omega)) \asymp \beta(S_h^{t,k}(\Omega)) \asymp n^{-l/d}$. It means that for sufficiently good finite element spaces and if the operator A^* obeys a step condition both the assumptions A.1 and A.2 are fulfilled.

3. Linear projection methods

Projection methods for linear ill-posed problems are well known and their analysis is well understood, starting from the original research in [29], and recently analyzed in [21]. If we have chosen some design spaces Y_n, then we may construct a projection method as follows. Given the corresponding projection Q_n as in (2) we let

$$B_n := Q_nA\colon X \to Y, \tag{17}$$

and $B_n^+\colon Y \to X_n$ the Moore-Penrose inverse of B_n. We associate each $Y_n \subset Y$ the space $X_n := A^*Y_n$ $(= \mathcal{R}(B_n^*))$ and the corresponding orthogonal projection $P_n\colon X \to X$. Under the present assumptions we actually have $B_n^+ = (B_n|_{X_n})^{-1}$. Indeed, if $x \in X_n$ and $B_nx = 0$ then

$x \in \mathcal{N}(B_n) \cap X_n = \mathcal{R}(B_n^*) \cap X_n = X_n^\perp \cap X_n = \{0\}$, which proves the injectivity of the operator B_n restricted to X_n.

Now we are going to mention some elementary properties.

Lemma 1. *The following identities hold true.*

$$B_n(I - P_n) = 0 \tag{18}$$

$$B_n B_n^+ = Q_n, \tag{19}$$

and

$$B_n^+ B_n = P_n. \tag{20}$$

Proof. By construction of X_n we have $B_n^* y = P_n B_n^* y,\ y \in Y$, hence by duality we obtain (18). The last assertions follow from the Moore-Penrose identities, see e.g. [11, Eq. (2.12/13)], in particular we obtain

$$B_n B_n^+ = Q_{\mathcal{R}(B_n)} = Q_n$$

and finally

$$B_n^+ B_n = P_{\mathcal{N}^\perp(B_n)} = P_{\mathcal{R}(B_n^*)} = P_n,$$

where we used $Y_n \subset \overline{\mathcal{R}(A)}$ to establish that $\mathcal{R}(B_n) = Y_n$. Indeed, obviously we have $\mathcal{R}(Q_n A) \subset Y_n$, and equality holds true since

$$Y_n = Q_n Y_n \subset Q_n \overline{\mathcal{R}(A)} = Q_n \mathcal{R}(A) = \mathcal{R}(Q_n A).$$

□

For the linear equation $y^\delta = Ax + \delta\xi$, the corresponding projection method (dual least squares method) would now be defined as

$$x_n^\delta = x_0 + B_n^+ Q_n (y^\delta - A x_0). \tag{21}$$

Remark 2. Notice, that in contrast to most studies we allow for an initial guess x_0. On the one hand, this makes a comparison to the non-linear case, as studied below more transparent. On the other hand side this might be useful if the true solution is smooth up to some specific and known feature, as captured in x_0.

The following Lemma relates the norm of $B_n^+ Q_n$ to the modulus of injectivity.

Lemma 2. *[see* [33, Lemma 1.2], *or* [21, p. 1528]*] If $Y_n \subset \overline{\mathcal{R}(A)}$ then*

$$\|B_n^+ Q_n\| \le 1/\beta(Y_n), \quad n \in \mathbb{N}.$$

Proof. The proof is based on the use of Hölder's inequality as follows.

$$\|B_n^* u\|^2 = \langle B_n^* u, B_n^* u\rangle = \langle B_n B_n^* u, u\rangle \leq \|B_n B_n^* u\| \|u\|, \quad u \in Y_n. \tag{22}$$

Consequently we deduce, using (19), that

$$\begin{aligned}
\|B_n^+ Q_n\| &= \sup_{u\in Y_n} \frac{\|B_n^+ u\|}{\|u\|} = \sup_{u\in Y_n} \frac{\|B_n^+ u\|}{\|B_n B_n^+ u\|} \\
&\leq \sup_{v\in X_n} \frac{\|x\|}{\|B_n x\|} = \sup_{u\in Y_n} \frac{\|B_n^* u\|}{\|B_n B_n^* u\|} \\
&\leq \sup_{u\in Y_n} \frac{\|u\|}{\|B_n^* u\|} = \frac{1}{\beta(Y_n)},
\end{aligned}$$

and the proof is complete. ☐

This provides us with the following important bound, if the Bernstein inequality holds true.

Corollary 1. *Under Assumption A.2 we have*

$$\|B_n^+ Q_n\| \leq C_B / s_n, \quad n \in \mathbb{N}. \tag{23}$$

From the previous discussion we can derive the following error bound at any element $x^\dagger \in B_r(x_0)$ such that $y^\delta = Ax^\dagger + \delta\xi$.

Proposition 1. *Under assumption A.2 we have*

$$\|x^\dagger - B_n^+ Q_n y^\delta\| \leq \|(I - P_n)(x^\dagger - x_0)\| + C_B \delta / s_n. \tag{24}$$

The above error bound will be generalized to the non-linear situation in Theorem 2. In particular, the error analysis in Section 5 for the non-linear problem extends to the linear case and yields corresponding error bounds.

4. Projection methods for non-linear problems

As in the linear case also for non-linear ill-posed inverse problems, which can be written as operator equation (1), regularization methods are required in order to obtain stable approximate solutions (see, e.g., [7, 11, 24, 32]). Besides variants of the Tikhonov regularization method iterative procedures are frequently applied, where the number of performed iterations plays the role of the regularization parameter. In this study we focus on projection methods, and the regularization parameter corresponds to the dimension of the design space. Precisely, we consider the natural extension of the projection method from (21) to non-linear problems, and complement in some crucial points the results of [21].

4.1. **Range invariance.** To obtain convergence rates, in the literature on regularization methods for non-linear ill-posed problems (see, e.g., the monographs [11], [2], [22] and the papers [4], [6], [8], [16], [18], [20], [31]) there were prescribed very different conditions concerning the smoothness, regularity and structure of the operator F. The method suggested in the present study will be based on the *range invariance condition* (see, e.g., [21]), which assumes that there is a linear operator A approximating the nonlinearity of F from (1) sufficiently well in the sense that throughout the ball $B_r(x_0)$ the operator F maps into the range $\mathcal{R}(A)$ of A.

Assumption A.3. There are a compact linear operator $A\colon X \to Y$ and a constant $0 \le k < 1$ such that

$$F(x) - F(\bar{x}) = A(x - \bar{x}) + AR(x, \bar{x}), \quad x, \bar{x} \in B_r(x_0), \tag{25}$$

where

$$R\colon B_r(x_0) \times B_r(x_0) \to \mathcal{N}^{\perp}(A) \tag{26}$$

is a mapping with

$$\|R(x, \bar{x})\| \le k\|x - \bar{x}\|, \quad x, \bar{x} \in B_r(x_0). \tag{27}$$

Remark 3. Plainly, if the mapping F is bounded linear, then we may take $A := F$ and Assumption A.3 is trivially fulfilled with $k = 0$. On the other hand, we note that the mapping R in (26) is unique: Due to the injectivity of A on the subspace $\mathcal{N}^{\perp}(A)$ for given $x, \bar{x} \in B_r(x_0)$ with $F(x) - F(\bar{x}) \in \mathcal{R}(A)$ there is a uniquely determined element $R(x, \bar{x}) \in \mathcal{N}^{\perp}(A)$ satisfying the equation (25).

The author of the papers [13, 14] recently studied (ill-posed) Hammerstein equations. Often such non-linear equations obey range invariance properties as in Assumption A.3. The convergence analysis of the iterative schemes as considered in that papers is also based on fixed point arguments, thus similar to ours. However, the assumptions from there are slightly different (due to the different methods under consideration), and the issue of discretization is not touched at all.

4.2. **The iterative procedure.** The following bound will be important, and it shows that the imposed range invariance assumption A.3 is useful for the analysis of projection methods.

Lemma 3. *Under Assumption A.3 we have*

$$\|B_n^+ Q_n(F(x) - F(\bar{x}) - A(x - \bar{x}))\| \le k\|x - \bar{x}\|, \quad x, \bar{x} \in B_r(x_0). \tag{28}$$

Proof. We use the function R from (26), which was introduced in Assumption A.3, to conclude that

$$B_n^+ Q_n(F(x) - F(\bar{x}) - A(x - \bar{x})) = B_n^+ Q_n AR(x, \bar{x}) = P_n R(x, \bar{x}).$$

Taking into account (27) this proves the lemma. □

We introduce the mapping

$$G_n(x) := x_0 + B_n^+ Q_n(y^\delta - F(x) - A(x_0 - x)), \quad x \in B_r(x_0). \tag{29}$$

Remark 4. In the linear case, thus if $A := F$, then we obtain the typical form (21), of the projection method in Hilbert space, see e.g. [21, 28].

For specific parameters n and x_0 this mapping has a fixed point in the set $x_0 + X_n \cap B_r(x_0)$. For a similar result we refer to [21, Lemma 2].

First we make the following assumption on the initial guess.

Assumption A.4. For k from Assumption A.3 we assume that the initial guess x_0 obeys

$$\|x^\dagger - x_0\| < r\frac{1-k}{1+k}.$$

Given $\delta > 0$, and under Assumption A.2 we let

$$M(\delta, x_0) := \left\{n, \quad C_B\delta/s_n + (1+k)\|x^\dagger - x_0\| < r(1-k)\right\}. \tag{30}$$

We stress that under Assumptions A.2, A.4 and for δ small enough the set $M(\delta, x_0)$ is not empty. Let $\delta_0 := \delta(x_0)$ be such noise level and set

$$\bar{n}(\delta) := \max\left\{n, \quad n \in M(\delta, x_0)\right\}, \quad 0 < \delta \le \delta_0. \tag{31}$$

For $n \le \bar{n}(\delta)$ we consider the following iterative procedure.

ITERATE(x)

INIT: $x^0 := x$;
IF: $x^l \in x_0 + X_n \cap B_r(x_0)$
THEN: $x^{l+1} := G_n(x^l)$
ELSE: STOP.

FIGURE 1. The iterative procedure

Theorem 1. *Assume A.2, A.3 and A.4 to hold and let $\bar{n}(\delta)$ from (31). For $n \le \bar{n}(\delta)$ the procedure* **ITERATE**(x_0) *converges to a fixed point, say $x_n^0 \in x_0 + X_n \cap B_r(x_0)$. Furthermore it holds true that*

$$\|x_n^0 - x^l\| \le k^l r, \quad l = 1, 2, \ldots \tag{32}$$

Proof. We shall apply the variant of the *Banach fixed point theorem*, as e.g. formulated in [9, 10.1.2], to the mapping $x \to G_n(x) - x_0$, which, by construction, maps X_n into itself. We need to show that

$$\|G_n(x) - G_n(\bar{x})\| \le k\|x - \bar{x}\|, \quad x, \bar{x} \in x_0 + X_n \cap B_r(x_0), \tag{33}$$

as well as

$$\|G_n(x_0) - x_0\| < r(1-k). \tag{34}$$

The proof of (33) is immediate, since

$$G_n(x) - G_n(\bar{x}) = B_n^+ Q_n(F(x) - F(\bar{x}) - A(x - \bar{x})),$$

such that Lemma 3 applies. Furthermore, by adding and subtracting $B_n^+ Q_n F(x^\dagger)$ in the definition of G_n we obtain for any $x \in B_r(x_0)$ that

$$\begin{aligned} G_n(x) = x_0 &+ B_n^+ Q_n(y^\delta - F(x^\dagger)) \\ &+ B_n^+ Q_n(F(x^\dagger) - F(x) - A(x^\dagger - x)) + B_n^+ B_n(x^\dagger - x_0). \end{aligned} \tag{35}$$

We use this representation for $x := x_0$ and Lemma 3 to conclude that

$$\|G(x_0) - x_0\| \le \delta\|B_n^+ Q_n\| + k\|x^\dagger - x_0\| + \|x^\dagger - x_0\|.$$

By the choice of $n \le \bar{n}(\delta)$ this allows to complete the proof of (34). The proof the fixed point theorem also provides us with the estimate

$$\|x^{l+m} - x^l\| \le \frac{k^l}{1-k}\|G(x_0) - x_0\| \le k^l r, \quad l \in \mathbb{N}. \tag{36}$$

Letting $m \to \infty$ we can complete the proof. □

Remark 5. In the linear case, i.e., for $k = 0$ the first iterate $x^1 = x_n^0$ is already the fixed point and this coincides with x_n^δ from (21). There is no iteration necessary in this case.

Let, as in Theorem 1, the element x_n^0 denote the unique fixed point of the mapping G_n. The following error bound, similar to [21, Lemma 3], holds true.

Lemma 4. *Under Assumptions A.2, A.3 and A.4 we have*

$$\|x^\dagger - x_n^0\| \le \frac{1}{1-k}\left(C_B\delta/s_n + \|(I - P_n)(x^\dagger - x_0)\|\right). \tag{37}$$

Proof. We use the representation (35) for $x := x_n^0 = G_n(x_n^0)$. The norm bound from Lemma 3 implies

$$\begin{aligned} \|x^\dagger - x_n^0\| &= \|x^\dagger - x_0 - B_n^+ Q_n(y^\delta - F(x_n^0) - A(x_0 - x_n^0))\| \\ &\le \delta\|B_n^+ Q_n\| + \|(I - P_n)(x^\dagger - x_0)\| + k\|x^\dagger - x_n^0\|, \end{aligned}$$

from which the proof can easily be completed. □

This bound gives us a hint how long to carry out the iteration, based solely on the discretization level n and the noise level δ. Under Assumptions A.2 and A.4, and given $n \le \bar{n}(\delta)$ from (31), we let

$$\bar{l} = l(n,\delta) := \min\left\{l, \quad k^l r(1-k) \le C_B\,\delta/s_n\right\}. \tag{38}$$

The following error bound holds true for the iterate $x^{l(n,\delta)} \in X_n$ at the true solution $x^\dagger$ and we let

$$e_n(x^\dagger,\delta) := \|x^\dagger - x^{l(n,\delta)}\|. \tag{39}$$

Theorem 2. *Under Assumptions A.2, A.3 and A.4 for $n \le \bar{n}(\delta)$ and $\bar{l}$ chosen according to (38) we have*

$$e_n(x^\dagger,\delta) \le \frac{1}{1-k}\left(2C_B\delta/s_n + \|(I-P_n)(x^\dagger - x_0)\|\right) \tag{40}$$

The error bound in (40) corresponds to the error decomposition in the linear case. It is a sum, for which the first term expresses the noise propagation (stability), whereas the second term characterizes the discretization error in the case of noiseless data (approximation). As we shall see in Section 5 bounds for the discretization error will result in overall error bounds by balancing both terms. However, the following a posteriori parameter choice will allow for error bounds based solely on the size of the noise propagation term.

4.3. **A posteriori parameter choice.** Based on the error bound (40) one might want to choose n which minimizes this bound, thus we let

$$n_{opt} = \arg\min\left\{e_n(x^\dagger,\delta), \quad n = 1,2,\dots,\bar{n}(\delta)\right\}. \tag{41}$$

Under assumptions A.2 – A.4, the resulting quantity $e_{n_{opt}}(x^\dagger,\delta)$ is the best possible accuracy that can be guaranteed by using the iterative procedure. Of course, to find the discretization level n_{opt} that realizes this best possible accuracy one needs a reliable estimate for the quantity $\|(I-P_n)(x^\dagger - x_0)\|$, which depends on the (usually unknown) smoothness of the unknown solution $x^\dagger$. We will now present a rule for the adaptive choice of the discretization level n_+ that allows us to reach the best possible accuracy up to the multiplier $6D$, where

$$D = \max\{s_{n-1}/s_n, \quad n = 1,2,\dots,\bar{n}(\delta)\}.$$

As we will see, such n_+ can be chosen without any *a priori* information about $\|(I-P_n)(x^\dagger - x_0)\|$. The idea of this adaptive choice rule is the same as for the *balancing principle* that was introduced in the context of ill-posed problems in [15] and attracts some interest recently (see, for example, [25] and the references therein).

Let $N(\delta)$ be the set of discretization levels $n < \bar{n}(\delta)$ such that

$$N(\delta) := \left\{ n,\ \|x^{l(n,\delta)} - x^{l(m,\delta)}\| \leq \frac{8C_B\delta}{s_m(1-k)},\ \text{for all } n < m \leq \bar{n}(\delta) \right\}.$$

The discretization level n_+ we are interested in is now defined as

$$n_+ = \min\{n, \quad n \in N(\delta)\}. \tag{42}$$

We stress that no information about $\|(I - P_n)(x^\dagger - x_0)\|$ is involved for this choice of n_+. The same argument as in [25] gives the following result, called *oracle inequality*, sometimes.

Theorem 3. *Under the Assumptions of Theorem 2*

$$e_{n_+}(x^\dagger, \delta) \leq 6De_{n_{opt}}(x^\dagger, \delta).$$

5. Error bounds

Throughout the error analysis all assumptions as made in A.1 – A.4 are assumed to hold. We shall not mention this explicitly.

For linear ill-posed problems, i.e., when (1) holds for a linear operator $A := F$, it is reasonable to assume the true solution $x^\dagger$ obeys a source representation

$$x^\dagger = x_0 + \psi(A^*A)v, \tag{43}$$

for some known element x_0 and some index function ψ, where we say that $\psi(t)$ defined for $0 \leq t \leq \|A\|^2$ is an index function if ψ is continuous and strictly increasing with $\psi(0) = 0$. This theory (with $x_0 = 0$) is well established, and we only mention the papers [27, 5]. For non-linear mappings F one might instead assume, as in [21], that (43) holds true for the neighboring linear mapping A from Assumption A.3. Here, in order to obtain convergence rates, we adopt an alternative approach, for example suggested in [19] and [17]. We are going to use the method of *approximate source conditions*, where the element $x^\dagger - x_0$ fulfills a *benchmark* source condition, only approximately. Error bounds will be obtained in terms of *distance functions*, which measure the degree of violation.

5.1. Error bound under benchmark smoothness.

We shall briefly discuss the case of the benchmark smoothness, and we give an explicit error bound under an a priori parameter choice. Recall that the degree of approximation of the subspaces Y_n was measured by $\eta_n := \|(I - Q_n)A\|$, see (7). How does this transfer to approximation properties of the projections P_n, which proves to be important in Theorem 2 and which was introduced in Section 3? If this can be answered in terms of smoothness properties of $x^\dagger - x_0$, then it is reasonable to choose an appropriate limit situation as *benchmark smoothness*. The following result points in this direction.

Lemma 5. *Let Q_n be as in (2) and P_n the corresponding projection onto X_n with degree of approximation (7). Then we have the estimate*

$$\|(I - P_n)A^*A\| \le \eta_n^2. \tag{44}$$

Proof. We start with Lemma 1 to infer that $(I - P_n)B_n^* = 0$. Hence

$$(I - P_n)A^*A = (I - P_n)(A^*A - B_n^*B_n),$$

which implies

$$\|(I - P_n)A^*A\| = \|(I - P_n)(A^*A - B_n^*B_n)\| \le \|A^*A - B_n^*B_n\|,$$

and with

$$\|A^*A - B_n^*B_n\| = \|(I - Q_n)A\|^2 = \eta_n^2$$

the assertion of the lemma. □

For the prescription of an appropriate benchmark smoothness for a wide class of projection operators Q_n in our approach we would be interested in finding index functions $\varphi(t)$ with maximal decay rate to zero as $t \to 0$ such that an inequality

$$\|(I - P_n)\varphi(A^*A)\| \le C_\varphi\, \varphi(\eta_n^2) \tag{45}$$

for some constant $1 \le C_\varphi < \infty$ gets valid. By the inequality (45) a qualification concept for regularization by discretization with regularization parameter $n \in \mathbb{N}$ can be introduced complementing the qualification concept presented in [27, Definition 1] for regularization methods with continous regularization parameters $\alpha > 0$. Evidently from Lemma 5 we have $\varphi(t) = t$ with $C_\varphi = 1$ as such function. On the other hand, from [26, Proposition 2] we see that an inequality

$$\|(I - P_n)\zeta((A^*A)^2)\| \le \zeta(\|(I - P_n)A^*A\|^2)$$

holds true for functions ζ with ζ^2 concave, and hence setting $\varphi = \zeta^2$ the monomials $\varphi(t) = t^\kappa$ satisfy (45) with $C_\varphi = 1$ if $0 < \kappa \le 1$.

This gives some motivation to choose $x^\dagger - x_0 \in \mathcal{R}(A^*A)$ as benchmark smoothness, i.e., we assume that there is a unique $v \in \mathcal{N}^\perp(A^*A)$ for which $x^\dagger - x_0 = A^*Av$. Precisely, let

$$x^\dagger = x_0 + A^*Av, \qquad \|v\| \le L. \tag{46}$$

Then norm bounds for $(I - P_n)(x^\dagger - x_0)$ will be given in terms of η_n and L. Precisely, Lemma 5 implies that

$$\|(I - P_n)(x^\dagger - x_0)\| \le L\eta_n^2. \tag{47}$$

For simplicity we shall restrict our subsequent analysis to $L := 1$, and we let

$$x^\dagger \in H_1 := \{x,\ x = x_0 + A^*Av,\ \|v\| \le 1\}\,. \tag{48}$$

Both the bounds from Theorem 2 and Lemma 5 provide us with the main error estimate. For its formulation we let

$$\underline{n} = \underline{n}(\delta) := \max\left\{ n, \quad s_n \ge \left(\frac{2C_B}{C_Q^2}\delta\right)^{1/3} \right\}. \tag{49}$$

We first show the following

Lemma 6. *There is $\delta_1 > 0$ such that*

$$\underline{n}(\delta) \le \bar{n}(\delta), \quad 0 < \delta \le \delta_1. \tag{50}$$

Proof. To this end we need to show that there is $\delta_1 > 0$ with the property: If n is such that $s_n \ge \left(2C_B\delta/C_Q^2\right)^{1/3}$ then $n \in M(\delta_0, x_0)$, the set from (30). Indeed, the assumption implies that $1/s_n \le \left(C_Q^2/(2C_B\delta)\right)^{1/3}$, and hence

$$\begin{aligned} C_B\delta/s_n + (1+k)\|x^\dagger - x_0\| &\le \frac{C_Q^{1/3} C_B \delta}{2^{1/3} C_B^{1/3} \delta^{1/3}} + (1+k)\|x^\dagger - x_0\| \\ &\le \left(\frac{C_Q}{2}\right)^{1/3} (C_B\delta)^{2/3} + (1+k)\|x^\dagger - x_0\|. \end{aligned}$$

Under A.4 the right hand side can be made smaller than $r(1-k)$ if δ is small enough, from which the proof can be completed. ☐

Now we are ready to formulate convergence rates results under benchmark smoothness.

Theorem 4. *Let $\underline{n}$ from (49), $\bar{l}$ be as in (38) and δ_1 from Lemma 6. Then we have for $x^\dagger$ satisfying (48)*

$$\|x^\dagger - x^{\bar{l}(\underline{n},\delta)}\| \le \frac{2}{1-k}\left(2C_BC_Q\delta\right)^{2/3}, \quad 0 < \delta \le \delta_1. \tag{51}$$

Proof. The proof directly follows from the error estimate (40) taking into account $1/s_{\underline{n}} \le \left(C_Q^2/(2C_B\delta)\right)^{1/3}$ for the noise propagation term and the upper bound (44) for the discretization error term. The latter contains the expression $\eta_{\underline{n}}^2$, which can further be estimated from above by the assumed Jackson inequality (6) as

$$\eta_{\underline{n}}^2 \le C_Q^2 s_{\underline{n}+1}^2 \le C_Q^2 \left(\frac{2C_B}{C_Q^2}\right)^{2/3} \delta^{2/3},$$

since $s_{\underline{n}+1} \leq \left(\frac{2C_B}{C_Q^2}\right)^{1/3} \delta^{1/3}$. This yields (51) and completes the proof. $\square$

Corollary 2. *The a posteriori parameter choice n_+ from (42) provides for $x^\dagger$ satisfying (48) an error estimate*

$$e_{n_+}(x^\dagger, \delta) \leq \frac{12D}{1-k} \left(2C_B C_Q \delta\right)^{2/3}, \quad 0 < \delta \leq \delta_1.$$

We finish this paragraph with the remark that an analogue to Theorem 4 can also be formulated if a condition (46) is satisfied instead of (48). Then by (47) the estimate (51) attains the form

$$\|x^\dagger - x^{\bar{l}(\underline{n},\delta)}\| \leq \frac{2L}{1-k} \left(2C_B C_Q \left(\frac{\delta}{L}\right)\right)^{2/3}, \quad 0 < \delta \leq \delta_1 \tag{52}$$

(see also [28]).

5.2. **Distance functions for non-linear problems.** If the element $x^\dagger - x_0$ is not smooth enough, i.e.,

$$x^\dagger - x_0 \notin \mathcal{R}(A^*A), \tag{53}$$

then we propose to measure its lack of smoothness by a decay rate of a certain distance function with respect to the benchmark smoothness H_1. Here, we generalize the concept of distance functions introduced in [17] to the present case and consider as relevant distance function

$$\rho_{x^\dagger}(t) := \operatorname{dist}(t(x^\dagger - x_0), H_1 - x_0), \quad t \geq 0. \tag{54}$$

This definition extends the previous one to non-centralized sets, as H_1, which however are centered around some element, say x_0. The result from [17, Lemma 5.3] extends to the present situation and we can state Lemma 7.

Lemma 7. *Under the assumption (53) the functions $\rho_{x^\dagger}(t)$ $(0 \leq t < \infty)$ and $\rho_{x^\dagger}(t)/t$ $(0 < t < \infty)$ are both index functions. Moreover, we have $\lim_{t\to\infty} \rho_{x^\dagger}(t) = \infty$, and the inverse $\rho_{x^\dagger}^{-1}(t)$ $(0 \leq t < \infty)$ exists and is also an index function.*

We may restrict our consideration to the set H_1. If some L-fold multiple of the ball would be used then a simple calculation reveals that

$$\operatorname{dist}(t(x^\dagger - x_0), L(H_1 - x_0)) = L\rho_{x^\dagger}(t/L),$$

such that the general case follows from this one in a simple manner. An analog situation occurs in the context of formula (52).

For obtaining convergence rates again the discretization error has to be estimated from above.

Lemma 8. *Let η_n be as in (7) and let $x^\dagger$ satisfy (53). Then for the corresponding distance function $\rho_{x^\dagger}(t)$ from (54) we can estimate*

$$\|(I-P_n)(x^\dagger - x_0)\| \le 2\frac{\eta_n^2}{\rho_{x^\dagger}^{-1}(\eta_n^2)}, \quad n \in \mathbb{N}. \tag{55}$$

Proof. Let $\bar{x}$ be the element of best approximation to $t(x-x_0)$ from $H_1 - x_0$. Using

$$t(x^\dagger - x_0) = t(x^\dagger - x_0) - (\bar{x} - x_0) + (\bar{x} - x_0), \quad t > 0,$$

we arrive at

$$\|(I-P_n)(x^\dagger - x_0)\| \le \frac{1}{t}\left(\|(I-P_n)(\bar{x} - x_0)\| + \rho_{x^\dagger}(t)\right), \quad t > 0. \tag{56}$$

Using the bound from Lemma 5 (for the element $\bar{x}$ instead of $x^\dagger$) and letting $t := \rho_{x^\dagger}^{-1}(\eta_n^2)$ the proof is complete. □

Given $\rho_{x^\dagger}$ we can introduce the two auxiliary functions

$$f(t) := \frac{t}{\rho_{x^\dagger}^{-1}(t)}, \qquad \Theta(t) := \sqrt{t}\, f(t), \qquad t > 0.$$

As a consequence of Lemma 7 both f and Θ are index functions on the non-negative half axis. Similarly to (49) we let

$$\underline{n} = \underline{n}(\delta) := \max\{n, \quad s_n^2 \ge \Theta^{-1}(C_B\delta/C_Q^2)\}. \tag{57}$$

As in Lemma 6 there is $\delta_2 > 0$ such that

$$\underline{n}(\delta) \le \bar{n}(\delta), \quad 0 < \delta \le \delta_2. \tag{58}$$

The main error bound using distance functions is as follows.

Theorem 5. *Let $\underline{n}$ from (57), $\bar{l}$ be as in (38) and δ_2 from above. Then for given distance function $\rho_{x^\dagger}(t)$ from (54) and $x^\dagger$ satisfying (53) we obtain the bound*

$$\|x^\dagger - x^{\bar{l}(\underline{n},\delta)}\| \le \frac{4C_Q^2}{1-k} f(\Theta^{-1}(C_B\delta/C_Q^2)), \quad 0 < \delta \le \delta_2. \tag{59}$$

Proof. The proof is again based on the error estimate (40) taking into account the Jackson inequality (6) as well as the bounds (55) and (57). This yields because of the identity

$$\frac{t}{\sqrt{\Theta^{-1}(t)}} = f\left(\Theta^{-1}(t)\right), \qquad t > 0,$$

and because of the inequality

$$f(C_B^2 t) \le C_B^2 f(t) \ (t > 0) \qquad \text{for} \quad C_B \ge 1,$$

being a consequence of the decreasing character of the function $f(t)/t = 1/\rho_{x^\dagger}(t)$ $(t > 0)$, the estimates

$$\begin{aligned} \|x^\dagger - x^{\bar{l}(\underline{n},\delta)}\| &\le \frac{1}{1-k}\left(\frac{2C_B\delta}{s_{\underline{n}}} + 2f(\eta_{\underline{n}}^2)\right) \le \frac{1}{1-k}\left(\frac{2C_B\delta}{s_{\underline{n}}} + 2C_Q^2\, f(s_{\underline{n}+1}^2)\right) \\ &\le \frac{2}{1-k}\left(\frac{C_B\delta}{\sqrt{\Theta^{-1}(C_B\delta/C_Q^2)}} + C_Q^2\, f(\Theta^{-1}(C_B\delta/C_Q^2))\right) \\ &\le \frac{4C_Q^2}{1-k}\, f(\Theta^{-1}(C_B\delta/C_Q^2)). \end{aligned}$$

This yields (59) and hence completes the proof. □

Example 2 (The monomial case)**.** We indicate the error bounds obtained from Theorem 5 for distance functions of monomial type

$$\rho_{x^\dagger}(t) = t^{1/\kappa}, \quad t \ge 0,$$

where $0 < \kappa < 1$ must hold in order to have the properties of Lemma 7. Consequently, there occur the functions

$$f(t) = t^{1-\kappa}, \qquad \Theta(t) = t^{\frac{3}{2}-\kappa}, \qquad \Theta^{-1}(t) = t^{\frac{2}{3-2\kappa}}, \qquad t \ge 0.$$

This yields a convergence rate of order

$$\|x^\dagger - x^{\bar{l}(\underline{n},\delta)}\| = \mathcal{O}\left(\delta^{\frac{2-2\kappa}{3-2\kappa}}\right) \qquad \text{as} \qquad \delta \to 0. \tag{60}$$

As κ varies in $(0,1)$ any rate exponent from the open interval $(0, \frac{2}{3})$ can be obtained.

5.3. **Range inclusions yield distance functions.** In the last paragraph we have shown that distance functions $\rho_{x^\dagger}$ can be utilized to obtain convergence rates for regularization by discretization applied to equation (1) when the benchmark smoothness $x^\dagger - x_0 \in H_1$ is violated. Such functions $\rho_{x^\dagger}$ can be verified if the a priori smoothness is characterized as

$$x^\dagger - x_0 \in \mathcal{R}(G) \tag{61}$$

with some self-adjoint bounded linear operator $G : X \to X$ having non-closed range. Since G can, in principle, be independent of A, this assumption is rather general. On the other hand, we need some *link condition* combining G and A. For a discussion of the universe of such link conditions we refer to [5] and [17]. Here we focus on *range inclusions* recently studied in [5, 10, 17, 19]. Due to the benchmark smoothness characterized by H_1 we shall assume that

$$\mathcal{R}(\gamma(G)) \subseteq \mathcal{R}(A^*A) \tag{62}$$

holds for some index functions $\gamma(t)$ $(0 \le t \le \|G\|)$. In order to use the range inclusion from (62) for bounding the decay of the distance function $\rho_{x^\dagger}$ for $x^\dagger$ from (61) we need to assume that $\mathcal{R}(\gamma(G)) \subset \mathcal{R}(G)$. From [17, Proof of Theorem 6.7], and in somewhat different notation from [10, Lemma 4.2], we derive the following proposition.

Proposition 2. *Under the conditions (61) and (62) and if there is some* $0 < \varepsilon \le \|G\|$ *such that*

$$q(0) := 0, \qquad q(t) := \frac{\gamma(t)}{t} \qquad (0 < t \le \varepsilon)$$

is an index function on $[0, \varepsilon]$*, then we have for sufficiently small* $t > 0$ *an upper bound*

$$\rho_{x^\dagger}(t) \le K_1\, \gamma(q^{-1}(K_2\, t)) \tag{63}$$

for the distance function from (54) with some constants $K_1, K_2 > 0$*.*

The bound (63) thus obtained allows for error estimates (59) whenever $\delta > 0$ is sufficiently small.

In case of Example 2 we have for

$$\gamma(t) = t^{\frac{1}{1-\kappa}}, \quad t \ge 0,$$

where $0 < \kappa < 1$ must hold,

$$q(t) = t^{\frac{\kappa}{1-\kappa}}, \qquad q^{-1}(t) = t^{\frac{1-\kappa}{\kappa}}$$

and thus

$$\rho_{x^\dagger}(t) \le K_3\, t^{\frac{1}{\kappa}}$$

for some constant $K_3 > 0$ and sufficiently small $t > 0$. This reduces the present analysis to the situation of Example 2.

References

[1] Ivo Babuška and A. K. Aziz. Survey lectures on the mathematical foundations of the finite element method. In *The Mathematical Foundations of the Finite Element Method with Applications to Partial Differential Equations (Proc. Sympos., Univ. Maryland, Baltimore, Md., 1972)*, pages 1–359. Academic Press, New York, 1972.

[2] A. B. Bakushinsky and M. Yu. Kokurin. *Iterative Methods for Approximate Solution of Inverse Problems*, volume 577 of *Mathematics and Its Applications (New York)*. Springer, Dordrecht, 2004.

[3] Johann Baumeister. *Stable Solution of Inverse Problems.* Advanced Lectures in Mathematics. Friedr. Vieweg & Sohn, Braunschweig, 1987.

[4] Barbara Blaschke-Kaltenbacher and Heinz W. Engl. Regularization methods for nonlinear ill-posed problems with applications to phase reconstruction. In *Inverse Problems in Medical Imaging and Nondestructive Testing (Oberwolfach, 1996)*, pages 17–35. Springer, Vienna, 1997.

[5] Albrecht Böttcher, Bernd Hofmann, Ulrich Tautenhahn, and Masahiro Yamamoto. Convergence rates for Tikhonov regularization from different kinds of smoothness conditions. *Appl. Anal.*, 85(5):555–578, 2006.
[6] G. Chavent and K. Kunisch. On weakly nonlinear inverse problems. *SIAM J. Appl. Math.*, 56(2):542–572, 1996.
[7] David Colton and Rainer Kress. *Inverse Acoustic and Electromagnetic Scattering Theory*, volume 93 of *Applied Mathematical Sciences*. Springer-Verlag, Berlin, 1992.
[8] Peter Deuflhard, Heinz W. Engl, and Otmar Scherzer. A convergence analysis of iterative methods for the solution of nonlinear ill-posed problems under affinely invariant conditions. *Inverse Problems*, 14(5):1081–1106, 1998.
[9] J. Dieudonné. *Éléments d'Analyse. Tome I.* Cahiers Scientifiques [Scientific Reports], XXVIII. Gauthier-Villars, Paris, third edition, 1981. Fondements de l'analyse moderne.
[10] Dana Düvelmeyer, Bernd Hofmann, and Masahiro Yamamoto. Range inclusions and approximate source conditions with general benchmark functions. 2007 (Paper submitted).
[11] Heinz W. Engl, Martin Hanke, and Andreas Neubauer. *Regularization of Inverse Problems*. Kluwer Academic Publishers Group, Dordrecht, 1996.
[12] Alexandre Ern and Jean-Luc Guermond. *Theory and Practice of Finite Elements*, volume 159 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 2004.
[13] S. George. Newton-Lavrentiev regularization of ill-posed Hammerstein type operator equation. *J. Inverse Ill-Posed Probl.*, 14(6):573–582, 2006.
[14] S. George. Newton-Tikhonov regularization of ill-posed Hammerstein operator equation. *J. Inverse Ill-Posed Probl.*, 14(2):135–145, 2006.
[15] Alexander Goldenshluger and Sergei V. Pereverzev. Adaptive estimation of linear functionals in Hilbert scales from indirect white noise observations. *Probab. Theory Related Fields*, 118(2):169–186, 2000.
[16] Martin Hanke, Andreas Neubauer, and Otmar Scherzer. A convergence analysis of the Landweber iteration for nonlinear ill-posed problems. *Numer. Math.*, 72(1):21–37, 1995.
[17] Bernd Hofmann and Peter Mathé. Analysis of profile functions for general regularization methods. *SIAM J. Numer. Anal.*, 2007 (Paper to appear).
[18] Bernd Hofmann and Otmar Scherzer. Factors influencing the ill-posedness of nonlinear problems. *Inverse Problems*, 10(6):1277–1297, 1994.
[19] Bernd Hofmann and Masahiro Yamamoto. Convergence rates for Tikhonov regularization based on range inclusions. *Inverse Problems*, 21(3):805–820, 2005.
[20] B. Kaltenbacher and A. Neubauer. Convergence of projected iterative regularization methods for nonlinear problems with smooth solutions. *Inverse Problems*, 22(3):1105–1119, 2006.
[21] Barbara Kaltenbacher. Regularization by projection with a posteriori discretization level choice for linear and nonlinear ill-posed problems. *Inverse Problems*, 16(5):1523–1539, 2000.
[22] Barbara Kaltenbacher, Andreas Neubauer, and Otmar Scherzer. *Iterative Regularization Methods for Nonlinear Problems*. 2007.

[23] Andreas Kirsch. *An Introduction to the Mathematical Theory of Inverse Problems*, volume 120 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 1996.

[24] M. M. Lavrentiev, A. V. Avdeev, M. M. Lavrentiev, Jr., and V. I. Priimenko. *Inverse Problems of Mathematical Physics*. Inverse and Ill-posed Problems Series. VSP, Utrecht, 2003.

[25] Peter Mathé. The Lepskiĭ principle revisited. *Inverse Problems*, 22(3):L11–L15, 2006.

[26] Peter Mathé and Sergei V. Pereverzev. Discretization strategy for linear ill-posed problems in variable Hilbert scales. *Inverse Problems*, 19(6):1263–1277, 2003.

[27] Peter Mathé and Sergei V. Pereverzev. Geometry of linear ill-posed problems in variable Hilbert scales. *Inverse Problems*, 19(3):789–803, 2003.

[28] Peter Mathé and Nadine Schöne. Regularization by projection in variable Hilbert scales. 2007 (Paper submitted).

[29] Frank Natterer. Regularisierung schlecht gestellter Probleme durch Projektionsverfahren (in German). *Numer. Math.*, 28(3):329–341, 1977.

[30] A. Pinkus. *n-Widths in Approximation Theory*. Springer, Berlin, Heidelberg, New York, 1985.

[31] O. Scherzer, H. W. Engl, and K. Kunisch. Optimal a posteriori parameter choice for Tikhonov regularization for solving nonlinear ill-posed problems. *SIAM J. Numer. Anal.*, 30(6):1796–1838, 1993.

[32] A. N. Tikhonov and V. Ya. Arsenin. *Metody resheniya nekorrektnykh zadach* (in Russian). "Nauka", Moscow, third edition, 1986.

[33] G. M. Vaĭnikko and U. A. Khyamarik. Projection methods and self-regularization in ill-posed problems. *Izv. Vyssh. Uchebn. Zaved. Mat.*, 84(10):3–17, 1985.

Department of Mathematics, Chemnitz University of Technology, 09107 Chemnitz, Germany
E-mail address: hofmannb@mathematik.tu-chemnitz.de

Weierstrass Institute for Applied Analysis and Stochastics, Mohrenstrasse 39, 10117 Berlin, Germany
E-mail address: mathe@wias-berlin.de

Johann-Radon-Institute (RICAM), Altenberger Strasse 69, A-4040 Linz, Austria
E-mail address: sergei.pereverzyev@oeaw.ac.at