Electronic Transactions on Numerical Analysis.
Volume 41, pp. 81-92, 2014.

ISSN 1068-9613.

ETNA
Kent State University
http://etna.math.kent.edu

ON CONVERGENCE RATES FOR QUASI-SOLUTIONS OF ILL-POSED PROBLEMS*

ANDREAS NEUBAUER† AND RONNY RAMLAU†

Abstract. Usually, one needs information about the noise level to find proper regularized solutions when solving ill-posed problems. However, for several inverse problems, it is not easy to obtain an accurate estimation of the noise level. If one has information about bounds of the solution in some stronger norm, quasi-solutions are an excellent alternative. Besides existence, stability, and convergence results, it is the major emphasis of this paper to prove convergence rates for quasi-solutions in Hilbert scales.

Key words. quasi-solutions, regularization in Hilbert scales, convergence rates

AMS subject classifications. 47A52, 65J20

1. Introduction.

We consider (nonlinear) ill-posed problems

$$F(x) = y\,, \tag{1.1}$$

where $F : \mathcal{D}(F) \subset X \to Y$ is a linear or nonlinear bounded operator between Hilbert spaces. Here, y are the exact data, and we assume that there exists a best-approximate solution of (1.1). In practice, however, only noisy data y^δ are available, where δ denotes the noise level, i.e., $\left\|y - y^\delta\right\| \le \delta$.

Due to the ill-posedness, one has to use regularization methods to obtain stable approximations of a solution of equation (1.1). Almost all regularization methods have in common that they depend on a so-called regularization parameter. A prominent example is Tikhonov regularization, where an approximation to a solution is obtained as a minimizer of the functional

$$\left\|F(x) - y^\delta\right\|^2 + \alpha R(x)\,, \qquad \alpha > 0\,.$$

Here, $R(x)$ denotes the chosen penalty term, and α is the regularization parameter. Popular choices for penalty terms can be found, e.g., in [1], where X and Y are Hilbert spaces and $R(x) = \|x - x_*\|^2$, or in [12, 13], where a general Banach space setting is considered. In iterative regularization methods (see, e.g., [5, 13]), the iteration index plays the role of the regularization parameter.

All convergence and especially convergence rates results for regularized solutions show that the choice of this parameter depends on the knowledge of δ. However, for several examples in practice, it is not at all trivial to get precise estimates of the noise level.

In such cases, one often uses so-called heuristic parameter choice rules. Unfortunately, it was shown by Bakushinskii (see, e.g., [1, Theorem 3.3]) that such methods can never yield convergence if one is interested in worst case error estimates. In contrast, these parameter choice rules often display good results in practice. The reason is that measured data are usually not always as bad as in the worst case situation. Under some restrictions on the noise, one can even prove convergence rates; see, e.g., [6, 11].

For some problems one has additional a priori information about the solution, e.g., that it lies in a smoother space and that an upper bound on the norm in this space is known. It turns

*Received December 12, 2013. Accepted March 25, 2014. Published online on May 16, 2014. Recommended by L. Reichel.
†Industrial Mathematics Institute, Johannes Kepler University Linz, A-4040 Linz, Austria (neubauer@indmath.uni-linz.ac.at, ronny.ramlau@jku.at).

ETNA
Kent State University
http://etna.math.kent.edu

out that this knowledge can be used instead of the knowledge of δ to prove convergence and in some cases even rates for so called quasi-solutions.

The concept of quasi-solutions was developed by Ivanov (see [3, 4] and also [14]) for injective completely continuous operators F. They are solutions of the problem

$$\inf_{x\in M} \left\|F(x) - y^\delta\right\| ,$$

where M is a compact subset of $\mathcal{D}(F)$. We will generalize this setting to non-compact and non-injective operators F and allow weakly compact sets M.

In the next section, we deal with existence, stability, and convergence of quasi-solutions. In Section 3, we derive convergence rates in Hilbert scales. An analysis of numerical approximations for linear problems as well as a fast algorithm and numerical results are presented in the last section.

2. Convergence of quasi-solutions. As mentioned in the introduction, we assume that equation (1.1) has a least squares solution in a smoother space, say $X_s \subset X$ and that a norm bound ρ in X_s is known. We therefore approximate solutions of (1.1) by the following quasi-solutions, namely solutions of the problem

$$\inf_{x\in M_\rho} \left\|F(x) - y^\delta\right\| , \qquad M_\rho := \{x \in \mathcal{D}(F) \cap X_s : \|x - x_*\|_s \le \rho\} . \tag{2.1}$$

Here, $\|\cdot\|_s$ obviously denotes the norm in X_s, and $\rho > 0$. The element $x_* \in \mathcal{D}(F) \cap X_s$ is either an initial guess for a solution or, in case of multiple solutions, it plays the role of a selection criterion (see [1, Chapter 10]). If F is a linear operator, usually $x_* = 0$.

For the following existence and convergence proofs, we require several conditions but not all of them are needed for every result. Nevertheless, we state all conditions that may be needed at once:

(A1) $X_s \subset X$ and Y are Hilbert spaces.

(A2) $\mathcal{D} := \mathcal{D}(F) \cap X_s$ and $F : \mathcal{D} \subset X_s \to Y$ is weakly (sequentially) closed, i.e., if (x_k) is a sequence in $\mathcal{D}$ such that $x_k \rightharpoonup x$ in X_s and $F(x_k) \rightharpoonup y$ in Y, then $x \in \mathcal{D}$ and $F(x) = y$.

(A3) The equation $F(x) = y^\delta$ has a least squares solution in $M_\rho \subset \mathcal{D}$, i.e., $\overline{x} \in M_\rho$ exists with $\left\|F(\overline{x}) - y^\delta\right\| = \inf_{x\in\mathcal{D}} \left\|F(x) - y^\delta\right\|$.

(A4) $\mathcal{D}$ is convex and $x \mapsto \left\|F(x) - y^\delta\right\|$ is convex in $\mathcal{D}$.

REMARK 2.1. Let us shortly discuss the meaning of the conditions above: first of all, (A2) and (A4) trivially hold if $F \in L(X_s, Y)$, i.e., if F is a linear bounded operator from $X_s \to Y$.

Note that, in general, for noisy data y^δ, least squares solutions do not exist and, therefore, condition (A3) will not hold in this case. However, we will assume that (A3) holds for exact data ($\delta = 0$). Due to the weak lower semi-continuity of the norms in X_s and Y, it is an immediate consequence of (A1) and (A2) that then also an x_*-minimum norm least squares solution, also called best-approximate solution, exists in M_ρ. These solutions are denoted by $x^\dagger$, i.e.,

$$\left\|x^\dagger - x_*\right\|_s = \min\{\|\overline{x} - x_*\|_s : \overline{x} \text{ is a least squares solution of } F(x) = y\} .$$

For nonlinear operators F, $x^\dagger$ is not necessarily unique. However, if condition (A4) holds for $\delta = 0$, then $x^\dagger$ will be unique.

Of course, as for least squares solutions, one can always choose quasi-solutions that minimize $\left\|x_\rho^\delta - x_*\right\|_s$ among all possible quasi-solutions, called x_*-minimum norm quasi-solutions.

ETNA
Kent State University
http://etna.math.kent.edu

PROPOSITION 2.2. *Let the conditions* (A1) *and* (A2) *be satisfied. Then the following assertions hold:*

(i) *Problem* (2.1) *has a solution* x_ρ^δ.
(ii) *If* (A3) *holds, then a least squares solution also solves problem* (2.1). *Moreover, an* x_**-minimum norm least squares solutions solve* (2.1).
(iii) *If* (A3) *does not hold but* (A4) *holds, then* x_ρ^δ *is unique with* $\left\|x_\rho^\delta - x_*\right\|_s = \rho$.

Proof. Assertion (i) follows immediately from (A1), (A2), and the weak lower semi-continuity of the norms. Assertion (ii) is obvious.

Let us now assume that (A4) holds and that equation (1.1) has no least squares solution in M_ρ. Then there exists an element $\overline{x} \in \mathcal{D}\backslash M_\rho$ such that

$$\left\|F(\overline{x}) - y^\delta\right\| < \left\|F(x_\rho^\delta) - y^\delta\right\| . \tag{2.2}$$

Let us assume that $\left\|x_\rho^\delta - x_*\right\|_s < \rho$, then $x(t) = t\overline{x} + (1-t)x_\rho^\delta \in M_\rho$ for some $t > 0$ and (2.2) implies that

$$\left\|F(x(t)) - y^\delta\right\| \le t\left\|F(\overline{x}) - y^\delta\right\| + (1-t)\left\|F(x_\rho^\delta) - y^\delta\right\| < \left\|F(x_\rho^\delta) - y^\delta\right\|$$

in contradiction to x_ρ^δ minimizing the residual over M_ρ. The uniqueness of x_ρ^δ follows from the fact that, if there were two quasi-solutions, then any convex combination would be also a quasi-solution but with a distance to x_* less than ρ. This proves assertion (iii). □

Assertion (iii) above means that, if (A3) does not hold, then (A4) guarantees that quasi-solutions are at the boundary of the set M_ρ.

Next, we are interested in stability and convergence of quasi-solutions.

PROPOSITION 2.3. *Let conditions* (A1) *and* (A2) *be satisfied and let* $\delta_k \to 0$ *as* $k \to \infty$. *Then the following assertions hold:*

(i) *The sequence* $x_\rho^{\delta_k}$ *has a weakly convergent subsequence. The limit of every weakly convergent subsequence is a quasi-solution. If, in addition,* x_ρ^0 *is unique, then*

$$x_\rho^{\delta_k} \rightharpoonup x_\rho^0 .$$

(ii) *If* x_ρ^0 *is unique with* $\left\|x_\rho^0 - x_*\right\|_s = \rho$*, then we obtain strong convergence, i.e.,*

$$x_\rho^{\delta_k} \to x_\rho^0 .$$

Proof. Obviously, $\left\|x_\rho^{\delta_k} - x_*\right\|_s \le \rho$, and $\left\|F(x_\rho^{\delta_k}) - y^{\delta_k}\right\|$ is bounded. Therefore, (A1) and (A2) imply that a subsequence (again denoted by $x_\rho^{\delta_k}$) and an element $\overline{x} \in M_\rho$ exist such that

$$x_\rho^{\delta_k} \rightharpoonup \overline{x} \qquad \text{and} \qquad F(x_\rho^{\delta_k}) \rightharpoonup F(\overline{x}) .$$

The weak lower semi-continuity of the norms now yields that

$$\|F(\overline{x}) - y\| \le \liminf_{k\to\infty} \left\|F(x_\rho^{\delta_k}) - y^{\delta_k}\right\| \le \lim_{k\to\infty} \left\|F(x) - y^{\delta_k}\right\| = \|F(x) - y\|$$

for all $x \in M_\rho$. Thus, $\overline{x}$ is a quasi-solution. If x_ρ^0 is unique, then the weak convergence of $x_\rho^{\delta_k}$ follows from a standard subsequence argument with the same limit.

If the conditions of (ii) hold, then

$$\rho = \left\|x_\rho^0 - x_*\right\|_s \le \liminf_{k\to\infty} \left\|x_\rho^{\delta_k} - x_*\right\|_s \le \limsup_{k\to\infty} \left\|x_\rho^{\delta_k} - x_*\right\|_s \le \rho .$$

Thus,

$$\lim_{k\to\infty} \left\| x_\rho^{\delta_k} - x_* \right\|_s = \rho\,.$$

Now weak convergence and convergence of the norms imply strong convergence. ∎

Therefore, in general, stability and convergence can only be guaranteed in the weak topology. However, note that if X_s is compactly embedded in X, then we get at least convergence in the norm-topology of X.

If $F(x) = y$ has a best-approximate solution $x^\dagger$ that is not an element of M_ρ and if condition (A4) holds for $\delta = 0$, then Proposition 2.2 (iii) and Proposition 2.3 (ii) imply that the quasi-solutions x_ρ^δ converge strongly towards the unique quasi-solution $x_\rho^0 \neq x^\dagger$. This means it is allowed to overestimate the bound ρ of the best-approximate solution but not to underestimate it.

REMARK 2.4. It is obvious from the proofs that the results of Propositions 2.2 and 2.3 remain valid if X_s and Y are reflexive Banach spaces. For the proof of assertion (ii) in Proposition 2.3 one needs in addition that X_s has the Radon-Riesz property, i.e., if $x_k \rightharpoonup x$ and $\|x_k\|_s \to \|x\|_s$, then $x_k \to x$ in X_s. This condition is valid in every Hilbert space, but also in some Banach spaces, e.g., in the Sobolev spaces $W^{s,p}$ with $1 < p < \infty$.

We also want to mention that there are cross-connections of quasi-solutions and Tikhonov regularized solutions: in [2], it was shown that the concept of quasi-solutions can be used to analyze the modulus of continuity to derive error bounds for regularized solutions. In case (A4) holds, one can characterize quasi-solutions as minimizers of a Tikhonov functional with a special regularization parameter (for compact linear operators this was already shown in [14]).

PROPOSITION 2.5. *Let the conditions* (A1), (A2), *and* (A4) *hold. If* (A3) *does not hold, then x_ρ^δ is the unique minimizer x_α^δ of the Tikhonov functional*

$$\left\| F(x) - y^\delta \right\|^2 + \alpha \left\| x - x_* \right\|_s^2$$

over $\mathcal{D}$, where the regularization parameter $\alpha > 0$ is chosen such that

$$\left\| x_\alpha^\delta - x_* \right\|_s = \rho\,.$$

Proof. Proposition 2.2 already implies that x_ρ^δ exists and is either a least squares solution of $F(x) = y^\delta$ in M_ρ or it is unique with $\left\| x_\rho^\delta - x_* \right\|_s = \rho$.

It follows from the Karush-Kuhn-Tucker theory (note that the Slater condition holds, since $x_* \in M_\rho$) that the solutions x_ρ^δ are characterized as follows: there exists an $\alpha \in \mathbb{R}$ such that x_ρ^δ minimizes

$$\left\| F(x) - y^\delta \right\|^2 + \alpha \left(\left\| x - x_* \right\|_s^2 - \rho^2 \right)$$

over $\mathcal{D}$ and

$$\alpha \left(\left\| x_\rho^\delta - x_* \right\|_s^2 - \rho^2 \right) = 0\,.$$

If (A3) does not hold, then $\alpha > 0$ and x_ρ^δ is as stated. ∎

If $F \in L(X_s, Y)$ and $x_* = 0$, then it is well-known that

$$x_\alpha^\delta = (F^\# F + \alpha I)^{-1} F^\# y^\delta\,,$$

ETNA
Kent State University
http://etna.math.kent.edu

where $F^\# : Y \to X_s$ denotes the adjoint of F. (We use the notation $F^\#$ instead of F^* since the latter symbol is used as the adjoint between the original spaces, i.e., $F^* : Y \to X$.) Assuming that (A3) does not hold, the problem of calculating α such that

$$\left\| x_\alpha^\delta \right\|_s = \rho \tag{2.3}$$

can be solved approximately via Newton's method. Let

$$f(\alpha) := \left\| x_\alpha^\delta \right\|_s^2 = \int_0^\infty \frac{\lambda}{(\alpha+\lambda)^2} \, d \left\| F_\lambda Q y^\delta \right\|^2 .$$

Here $\{F_\lambda\}$ denotes the spectral family of $FF^\#$. Since

$$f'(\alpha) = -2 \left\langle x_\alpha^\delta, (F^\# F + \alpha I)^{-1} x_\alpha^\delta \right\rangle_s = -2 \int_0^\infty \frac{\lambda}{(\alpha+\lambda)^3} \, d \left\| F_\lambda Q y^\delta \right\|^2 < 0 ,$$
$$f''(\alpha) = 6 \int_0^\infty \frac{\lambda}{(\alpha+\lambda)^4} \, d \left\| F_\lambda Q y^\delta \right\|^2 > 0 ,$$

Newton's method

$$\alpha_{k+1} = \alpha_k - \frac{f(\alpha_k) - \rho^2}{f'(\alpha_k)}$$

is globally convergent and locally quadratically convergent if $\alpha > 0$ exists such that (2.3) holds and if $\alpha_0 < \alpha$.

Noting that f is a monotonically decreasing function with

$$\lim_{\alpha\to\infty} f(\alpha) = 0 ,$$
$$\lim_{\alpha\to 0} f(\alpha) = \begin{cases} \left\| F^\dagger y^\delta \right\|_s & \text{if } y^\delta \in \mathcal{D}(F^\dagger) , \\ \infty & \text{otherwise} , \end{cases}$$

it is obvious that (2.3) will never have a solution if $\left\| F^\dagger y^\delta \right\|_s < \rho$. However, in that case, x_ρ^δ is a least squares solution and $F^\dagger y^\delta$ is the unique minimum norm quasi-solution.

The following modified algorithm will always converge towards the unique minimum norm quasi-solution which is always an element of $\mathcal{N}(F)^\perp$.

ALGORITHM 2.6.

(i) Choose $\alpha > 0$, $q \in (0,1)$.

(ii) Solve the equations $(F^\# F + \alpha I) x_\alpha^\delta = F^\# y^\delta$ and $(F^\# F + \alpha I) z_\alpha^\delta = x_\alpha^\delta$.

(iii) Calculate $d\alpha := \dfrac{\rho^2 - \left\| x_\alpha^\delta \right\|_s^2}{2 \left\langle x_\alpha^\delta, z_\alpha^\delta \right\rangle_s}$.

(iv) Set $\alpha := \begin{cases} \alpha - d\alpha & \text{if } d\alpha < \alpha , \\ q\,\alpha & \text{otherwise} . \end{cases}$

(v) Goto (ii).

Thus, if $\left\| F^\dagger y^\delta \right\|_s \le \rho$, then the sequence α_k obtained by this algorithm will monotonically decrease towards 0. Otherwise, it will converge from below towards the solution $\alpha > 0$ of equation (2.3). In practice, the algorithm will be stopped whenever the relative error of two consecutive α values is below a certain bound.

ETNA
Kent State University
http://etna.math.kent.edu

3. Convergence rates. As already mentioned in the last section, usually we can not expect strong convergence of quasi-solutions in X_s if δ goes to 0, but only in the space X if X_s is compactly embedded in X. If X and X_s are part of a Hilbert scale, then we can even prove convergence rates in X.

Let L be a densely defined unbounded selfadjoint strictly positive operator in X. Then $(X_s)_{s\in\mathbb{R}}$ denotes the Hilbert scale induced by L if X_s is the completion of $\bigcap_{k=0}^{\infty} D(L^k)$ with respect to the Hilbert space norm $\|x\|_s := \|L^s x\|_X$. Obviously $\|x\|_0 = \|x\|_X$; see [7] or [1, Section 8.4] for details.

The conditions needed to prove convergence rates are different for linear and nonlinear operators. We first consider the nonlinear case:

- (N1) $X = X_0$ and X_s, $s > 0$, are part of a Hilbert scale, and Y is a Hilbert space. Moreover, X_s is compactly embedded into X_t whenever $t < s$.
- (N2) $F(x) = y$ has a unique solution $x^\dagger \in M_\rho$, i.e., $F(x^\dagger) = y$. (Thus, for nonlinear problems we restrict ourselves to the attainable case.)
- (N3) $F : \mathcal{D}(= \mathcal{D}(F) \cap X_s) \to Y$ is Fréchet-differentiable in X_s.
- (N4) $\left\|F'(x^\dagger)x\right\| \ge m\,\|x\|_{-a}$ for all $x \in X$, some $a > 0$ and $m > 0$.
- (N5) There exist $c \ge 0$, $r \in [-a, s)$, $\beta \in (0, 1]$, and $\varepsilon > 0$ such that $M_\rho \cap B^r_\varepsilon(x^\dagger)$ is convex and $\left\|(F'(x) - F'(x^\dagger))(x - x^\dagger)\right\| \le c\left\|F'(x^\dagger)(x - x^\dagger)\right\|\left\|x - x^\dagger\right\|_r^\beta$ for all $x \in M_\rho \cap B^r_\varepsilon(x^\dagger)$, where $B^r_\varepsilon(x^\dagger) := \{x \in X_r : \left\|x - x^\dagger\right\|_r \le \varepsilon\}$.

REMARK 3.1. The Fréchet-differentiability of F in (N3) has to be understood in the following sense: F may be extended from $\mathcal{D}$ to an open subset $U \supset \mathcal{D}$, and this extension is Fréchet-differentiable.

Usually, for the analysis of regularization methods in Hilbert scales, a stronger condition than (N4) is used, namely (see, e.g., [8, 9])

$$\left\|F'(x^\dagger)x\right\| \sim \|x\|_{-a} \qquad \text{for all} \quad x \in X\,.$$

Condition (N5) means that the smoothing property of $F'(x)$ is similar to that of $F'(x^\dagger)$ in a neighbourhood of $x^\dagger$. Such conditions are common for proving convergence rates for nonlinear regularization in Hilbert scales and for iterative regularization methods; see [5, 9].

To prove convergence rates, we need the so-called interpolation inequality that holds in Hilbert scales (see, e.g., [1, Proposition 8.19]), i.e., for some $c > 0$

$$\|x\|_t \le c\,\|x\|_{-a}^{\frac{s-t}{a+s}}\,\|x\|_s^{\frac{a+t}{a+s}}\,, \quad -a \le t \le s\,, \quad x \in X_s\,. \tag{3.1}$$

THEOREM 3.2. *Let* (A2) *and conditions* (N1)–(N5) *hold. Then*

$$\left\|x^\delta_\rho - x^\dagger\right\|_t = O\left(\delta^{\frac{s-t}{a+s}}\right)$$

for any $-a \le t < s$.

Proof. First of all note that, due to Proposition 2.2, x^δ_ρ exists for all $y^\delta \in Y$ and $x^0_\rho = x^\dagger$ is unique due to (N2). Thus, Proposition 2.3 and (N1) imply that $x^\delta_\rho \to x^\dagger$ in X_t for $t < s$. Therefore, we may assume that δ is sufficiently small so that $\left\|x^\delta_\rho - x^\dagger\right\|_r \le \varepsilon$ and that we may apply (N5).

Now we prove rates: (N2) and $\left\|y - y^\delta\right\| \le \delta$ imply that

$$\left\|x^\delta_\rho - x^\dagger\right\|_s \le 2\rho\,, \tag{3.2}$$

$$\left\|F(x^\delta_\rho) - y\right\| \le \delta + \left\|F(x^\delta_\rho) - y^\delta\right\| \le \delta + \left\|F(x^\dagger) - y^\delta\right\| \le 2\delta\,. \tag{3.3}$$

ETNA
Kent State University
http://etna.math.kent.edu

Using (N2), (N3), and (N5), we obtain that

$$\begin{aligned} \left\|F'(x^\dagger)(x_\rho^\delta - x^\dagger)\right\| &\le \left\|F(x_\rho^\delta) - y\right\| \\ &\quad + \int_0^1 \left\|(F'(x^\dagger + \xi(x_\rho^\delta - x^\dagger)) - F'(x^\dagger))(x_\rho^\delta - x^\dagger)\right\| \, d\xi \\ &\le \left\|F(x_\rho^\delta) - y\right\| + \frac{c}{\beta+1} \left\|F'(x^\dagger)(x_\rho^\delta - x^\dagger)\right\| \left\|x_\rho^\delta - x^\dagger\right\|_r^\beta . \end{aligned} \tag{3.4}$$

Since $x_\rho^\delta \to x^\dagger$ in X_r, we may as well assume that δ is so small that

$$\left\|x_\rho^\delta - x^\dagger\right\|_r \le \left(\frac{\beta+1}{2c}\right)^{\frac{1}{\beta}} .$$

This together with (3.4) implies that

$$\left\|F'(x^\dagger)(x_\rho^\delta - x^\dagger)\right\| \le 2 \left\|F(x_\rho^\delta) - y\right\| .$$

Now, (N4), (3.1), (3.2), and (3.3) yield the assertion. ∎

For the special case $t = 0$, we obtain the convergence rate $O\left(\delta^{\frac{s}{a+s}}\right)$. If $x^\dagger \in X_u$ with $u > s$, then one could even obtain a better rate with Tikhonov regularization in Hilbert scales provided that δ is known. However, if the guess of s is the best possible, i.e., $x^\dagger \notin X_u$ for $u > s$, then the quasi-solutions converge order optimal without the knowledge of δ.

Let us now consider the linear case: an inspection of the proof above shows that the conditions (N1)–(N5) may be relaxed for linear operators F. First of all, we do not need the compact embedding condition of (N1) since (N5) is trivially satisfied with $c = 0$. Moreover, we may restrict ourselves to the unique minimum norm quasi-solutions (see the end of the section above) and note that

$$\left\|Fx - y^\delta\right\|^2 = \left\|Fx - Qy^\delta\right\|^2 + \left\|(I-Q)y^\delta\right\|^2 ,$$

where Q is the orthogonal projector of Y onto $\overline{\mathcal{R}(F)}$. Thus, a quasi-solution for y^δ is also a quasi-solution for Qy^δ. Therefore, the proof above remains valid if we replace y^δ by Qy^δ, meaning that there is no need to restrict ourselves to the attainable case for linear problems.

These arguments show that for linear operators we need the following conditions:

(L1) $X = X_0$ and X_s, $s > 0$, are part of a Hilbert scale, and Y is a Hilbert space.
(L2) $F \in L(X_s, Y)$ and $x^\dagger = F^\dagger y \in M_\rho$.
(L3) $\|Fx\| \ge m \|x\|_{-a}$ for all $x \in X$, some $a > 0$ and $m > 0$.

Then we obtain the following result:

THEOREM 3.3. *Let the conditions (L1)–(L3) hold and let x_ρ^δ be minimum norm quasi-solutions. Then*

$$\left\|x_\rho^\delta - x^\dagger\right\|_t = O\left(\delta^{\frac{s-t}{a+s}}\right)$$

for any $-a \le t < s$.

If $\left\|x^\dagger\right\|_s = \rho$, then it follows as in Proposition 2.3 (ii) that the minimum norm quasi-solutions will converge strongly towards $x^\dagger$ in X_s. Thus, $O(\cdot)$ in Theorem 3.3 may then even be replaced by $o(\cdot)$.

ETNA
Kent State University
http://etna.math.kent.edu

We now present two applications of the previous theorems, a nonlinear and a linear one.

EXAMPLE 3.4. In the first example, F is a nonlinear Hammerstein integral operator, defined as

$$\begin{aligned} F : H^1[0,1] &\to L^2[0,1] \\ x &\mapsto \int_0^t \phi(x(s))\,ds\,. \end{aligned}$$

We assume that ϕ is in $C^{2,\beta}(I)$ for all intervals $I \subset \mathbb{R}$ and some $\beta \in (0,1]$. Then one can show (see [10, Section 4] for details) that F is weakly closed and Fréchet-differentiable with

$$\begin{aligned} (F'(x)h)(t) &= \int_0^t \phi'(x(s))h(s)\,ds\,, \\ (F'(x)^* w) &= B^{-1}\left[\phi'(x(\bullet))\int_\bullet^1 w(t)\,dt\right]\,, \end{aligned}$$

where

$$\begin{aligned} B : \mathcal{D}(B) := \{\psi \in H^2[0,1]\,:\,\psi'(0) = \psi'(1) = 0\} &\to L^2[0,1] \\ \psi &\mapsto B\psi := -\psi'' + \psi\,; \end{aligned}$$

note that B^{-1} is the adjoint of the embedding operator from $H^1[0,1]$ in $L^2[0,1]$.

Let us assume that $F(x^\dagger) = y$ with $|\phi'(x^\dagger(s))| \ge \gamma > 0$ for all $s \in [0,1]$. Then

$$\mathcal{R}(F'(x^\dagger)^*) = \{\psi \in H^3[0,1]\,:\,\psi'(0) = \psi'(1) = 0, \psi(1) = \psi''(1)\}\,.$$

If we set $X = X_0 = H^1[0,1]$ (with the usual norm $\|x\|_0 := (\|x\|_{L^2}^2 + \|x'\|_{L^2}^2)^{\frac{1}{2}}$), then the operator L defined via

$$\begin{aligned} &\mathcal{D}(L^4) = \{\psi \in H^5[0,1]\,:\,\psi'(0) = \psi'(1) = \psi'''(0) = 0, \psi(1) = \psi''(1)\}\,, \\ &L^4\psi := \psi - 2\psi'' + \psi^{(iv)}\,, \end{aligned}$$

induces a Hilbert scale such that

$$X_2 = \mathcal{R}(F'(x^\dagger)^*) \quad \text{and} \quad \left\|F'(x^\dagger)^* w\right\|_2 = \left\|BF'(x^\dagger)^* w\right\|_0 \sim \|w\|_{L^2}\,.$$

Due to [1, Corollary 8.22] (compare [10, Remark 2.2]), this is equivalent to

$$\left\|F'(x^\dagger)x\right\| \sim \|x\|_{-2} \qquad \text{for all} \quad x \in X\,.$$

Moreover,

$$\begin{aligned} \left\|(F'(x) - F'(x^\dagger))^* w\right\|_2 &= \left\|(\phi'(x(\bullet)) - \phi'(x^\dagger(\bullet)))\int_\bullet^1 w(t)\,dt\right\|_0 \\ &\le c\left\|x - x^\dagger\right\|_0^\beta \|w\|_{L^2} \end{aligned}$$

for all $x \in X_0$ with $\left\|x - x^\dagger\right\|_0 \le \varepsilon$. The constant c depends on $x^\dagger$ and ε. Thus,

$$\begin{aligned} \left\|(F'(x) - F'(x^\dagger))h\right\|_{L^2} &= \sup_{\|w\|_{L^2}=1} \left\langle (F'(x) - F'(x^\dagger))h, w \right\rangle_{L^2} \\ &= \sup_{\|w\|_{L^2}=1} \left\langle h, (F'(x) - F'(x^\dagger))^* w \right\rangle_0 \\ &\le \sup_{\|w\|_{L^2}=1} \|h\|_{-2} \left\|(F'(x) - F'(x^\dagger))^* w\right\|_2\,. \end{aligned}$$

ETNA
Kent State University
http://etna.math.kent.edu

Hence,

$$\left\|(F'(x) - F'(x^\dagger))h\right\|_{L^2} \le c\,\|h\|_{-2}\left\|x - x^\dagger\right\|_0^\beta .$$

This shows that the conditions (A2), (N1), (N3)–(N5) hold with $a = 2$ and $r = 0$. If we assume that $x^\dagger$ satisfies (N2) (for some $s > 0$), then Theorem 3.2 is applicable.

EXAMPLE 3.5. In the second example, $F : L^2[0,1] \to L^2[0,1]$ is a linear Fredholm integral operator, defined as

$$(Fx)(t) = \int_0^1 k(s,t)x(s)ds$$

with kernel

$$k(s,t) := \begin{cases} s(1-t) & s < t\,, \\ t(1-s) & t \le s\,. \end{cases}$$

The operator $F = F^*$ is compact and

$$Fx = y \qquad \Longleftrightarrow \qquad y \in H^2[0,1] \cap H_0^1[0,1] \quad \text{and} \quad y'' = -x \ \text{ a.e.}$$

If we set $X = X_0 = L^2[0,1]$, then the operator L defined via

$$L^2 x := -x'' \quad \text{on} \quad \mathcal{D}(L^2) = H^2[0,1] \cap H_0^1[0,1]$$

induces a Hilbert scale such that

$$X_2 = \mathcal{R}(F^*) \quad \text{and} \quad \|F^* w\|_2 = \|w\|_{L^2}\,.$$

If we assume that the unique solution $x^\dagger = F^\dagger y$ exists and is an element of M_ρ (for some $s > 0$), then the conditions (L1)–(L3) are satisfied. Hence, Theorem 3.3 is applicable.

4. Finite dimensional approximation and numerical results. For numerical calculations, one has to approximate the infinite-dimensional spaces and the operator F as, e.g., in [10, Section 3]. In this section, we restrict ourselves to the case of linear compact operators F. For our convergence rates analysis we assume that the conditions (L1)–(L3) hold.

The spaces X_s and Y are approximated by finite-dimensional subspaces $\{X_m\}_{m\in\mathbb{N}}$ and $\{Y_m\}_{m\in\mathbb{N}}$. We assume that $Y_m \subset \overline{\mathcal{R}(F)}$. This guarantees that $Q_m y = Q_m Q y$, where Q_m is the orthogonal projector of Y onto Y_m and Q is (as above) the orthogonal projector of Y onto $\overline{\mathcal{R}(F)}$.

We then look for the unique minimum norm quasi-solution $x_\rho^{m,\delta} \in X_m$ of the problem

$$\inf_{x\in M_\rho^m} \left\|Q_m F x - Q_m y^\delta\right\|\,, \qquad M_\rho^m := \{x \in X_m : \|x\|_s \le \rho\}$$

or, equivalently,

$$\inf_{x\in M_\rho} \left\|Q_m F P_m x - Q_m y^\delta\right\|\,, \qquad M_\rho := \{x \in X_s : \|x\|_s \le \rho\}\,, \tag{4.1}$$

where P_m is the orthogonal projector of X_s onto X_m.

Since we want to prove convergence (rates) as $\delta \to 0$ and $m \to \infty$, we have to assume that $\|(I - P_m)x\| \to 0$ for all $x \in X_s$ and $\|(I - Q_m)y\| \to 0$ for all $y \in \overline{\mathcal{R}(F)}$. This is, for instance, guaranteed if $X_m \subset X_{m+1}$ and $\bigcup_{m\in\mathbb{N}} X_m$ is dense in X_s and if $Y_m \subset Y_{m+1}$ and $\bigcup_{m\in\mathbb{N}} Y_m$ is dense in $\overline{\mathcal{R}(F)}$.

ETNA
Kent State University
http://etna.math.kent.edu

Setting $F_m := Q_m F P_m$, we obtain according to the results above that $x_\rho^{m,\delta} = F_m^\dagger y^\delta$ if $\left\|F_m^\dagger y^\delta\right\|_s \le \rho$ and that $x_\rho^{m,\delta} = (F_m^\# F_m + \alpha I)^{-1} F_m^\# y^\delta$ with $\alpha > 0$ such that $\left\|x_\rho^{m,\delta}\right\|_s = \rho$ if $\left\|F_m^\dagger y^\delta\right\|_s > \rho$.

The estimates

$$\begin{aligned}\left\|F x_\rho^{m,\delta} - Qy\right\| &\le \left\|F x_\rho^{m,\delta} - F_m x_\rho^{m,\delta}\right\| + \left\|F_m x_\rho^{m,\delta} - Q_m y^\delta\right\| + \left\|Q_m y^\delta - Qy\right\| \\ &\le \left\|F x_\rho^{m,\delta} - F_m x_\rho^{m,\delta}\right\| + \left\|F_m x^\dagger - Q_m y^\delta\right\| + \left\|Q_m y^\delta - Qy\right\| , \\ \left\|Q_m y^\delta - Qy\right\| &\le \left\|Q_m(y - y^\delta)\right\| + \left\|(I - Q_m) F x^\dagger\right\| , \\ \|F - F_m\|_{X_s,Y} &\le \|Q_m F (I - P_m)\|_{X_s,Y} + \|(I - Q_m) F\|_{X_s,Y}\end{aligned}$$

imply that

$$\left\|F x_\rho^{m,\delta} - Qy\right\| \le 2\rho \|Q_m F (I - P_m)\|_{X_s,Y} + 4\rho \|(I - Q_m) F\|_{X_s,Y} + 2 \left\|Q_m(y - y^\delta)\right\| .$$

Thus, we obtain the following convergence rates result (compare the proof of Theorem 3.2).

THEOREM 4.1. *Let conditions (L1)–(L3) hold. Moreover, let F be compact and let the data y^δ be such that $\left\|Q_m(y - y^\delta)\right\| \le \delta$. Then we obtain the following estimate for the minimum norm quasi-solutions $x_\rho^{m,\delta}$*

$$\left\|x_\rho^{m,\delta} - x^\dagger\right\|_t = O\left((\gamma_m + \delta)^{\frac{s-t}{a+s}}\right)$$

for any $-a \le t < s$, where

$$\gamma_m := \|Q_m F (I - P_m)\|_{X_s,Y} + \|(I - Q_m) F\|_{X_s,Y} .$$

Note that the compactness of F implies that $\gamma_m \to 0$ as $m \to \infty$. There are two interesting special cases: in the first one $X_m := F^\# Y_m$. Then $Q_m F (I - P_m) = 0$ (see, e.g., [1, (3.49)]), and hence $\gamma_m = \|(I - Q_m) F\|_{X_s,Y}$. In the second case, we assume that $Y_m = \overline{\mathcal{R}(F)}$. Then $Q_m = Q$, $(I - Q_m) F = 0$, and $\gamma_m = \|F(I - P_m)\|_{X_s,Y}$.

As already mentioned above, the correct value of α so that $\left\|x_\rho^{m,\delta}\right\|_s = \rho$ may be computed approximately via Newton's method. Assuming that

$$Y_m = \operatorname{span}\{\psi_1, \dots, \psi_{d(m)}\} ,$$

where $d(m)$ is the dimension of Y_m, for the first special case Algorithm 2.6 turns into

ALGORITHM 4.2.

(i) Calculate $M := \left[\left\langle F^\# \psi_i, F^\# \psi_j \right\rangle_s\right]$, $H := [\langle \psi_i, \psi_j \rangle]$, $\overline{y} := \left[\left\langle y^\delta, \psi_j \right\rangle\right]$. Moreover, choose an initial value $\alpha > 0$, small numbers $\varepsilon_1, \varepsilon_2 > 0$, and $q \in (0, 1)$.

(ii) Solve the equations $(M + \alpha H)\overline{x} = \overline{y}$ and $(M + \alpha H)\overline{z} = H\overline{x}$.

(iii) Calculate $d\alpha := \dfrac{\rho^2 - \overline{x}^T M \overline{x}}{2\overline{x}^T M \overline{z}}$.

(iv) If $|d\alpha| < \varepsilon_1 \alpha$ or $\alpha < \varepsilon_2$, then stop;
else: set $\alpha := \begin{cases} \alpha - d\alpha , & \text{if } d\alpha < \alpha - \varepsilon_2 , \\ q\,\alpha , & \text{otherwise} . \end{cases}$

(v) Goto (ii).

Using the output vector $\overline{x}$ of this algorithm, the minimum norm quasi-solution is given by

$$x_\rho^{m,\delta} = \sum_{i=1}^{d(m)} \overline{x}_i F^\# \psi_i .$$

ETNA
Kent State University
http://etna.math.kent.edu

TABLE 4.1
Results of Example 4.3.

m	δ	err_2	$\text{err}_2 * m$	err_3	$\text{err}_3 * m$	$\text{err}_{1.5}$
4	$1.85 * 10^{-2}$	0.18444	0.74	0.18953	0.76	0.18116
8	$4.62 * 10^{-3}$	0.04620	0.37	0.07134	0.57	0.03887
16	$1.16 * 10^{-3}$	0.02460	0.39	0.03581	0.57	0.03127
32	$2.89 * 10^{-4}$	0.01370	0.44	0.01772	0.57	0.03124
64	$7.23 * 10^{-5}$	0.00436	0.28	0.00797	0.51	0.03124
128	$1.81 * 10^{-5}$	0.00184	0.24	0.00354	0.45	0.03124
256	$4.52 * 10^{-6}$	0.00067	0.17	0.00166	0.42	0.03124
512	$1.13 * 10^{-6}$	0.00028	0.14	0.00067	0.34	0.03124

We apply this algorithm to the operator F of Example 3.5.

EXAMPLE 4.3. Let F and the Hilbert scale be defined as in Example 3.5. In addition, we assume that $x^\dagger \in X_2$ with $\left\|x^\dagger\right\|_2 = \left\|(x^\dagger)''\right\|_{L^2} \le \rho$, i.e., $s = 2$. As finite-dimensional spaces Y_m, we choose linear splines with equidistant knots of mesh size $h = 1/m$, i.e., $d(m) = m + 1$. As basis functions ψ_i we use the usual hat functions.

The spaces X_m are chosen as $F^\# Y_m$. Noting that $F^\# = L^{-2s}F^*$, that $s = 2$, and that $F = F^* = L^{-2}$, we obtain $F^\# = L^{-6} = F^3$ and that

$$\left\langle F^\# \psi_i, F^\# \psi_j \right\rangle_s = \left\langle F^2 \psi_i, F^2 \psi_j \right\rangle_{L^2} .$$

The operator F^2 is again an integral operator with kernel

$$k_2(s,t) = \frac{1}{6} \begin{cases} s(1-t)(2t - t^2 - s^2) & s < t\,, \\ t(1-s)(2s - s^2 - t^2) & t \le s\,. \end{cases}$$

For the right-hand side $y(t) = (t - 2t^3 + t^4)/12$, the exact solution of $Fx = y$ is given by $x^\dagger(s) = s - s^2$. Obviously, $x^\dagger \in X_2$ with $\left\|x^\dagger\right\|_2 = 2$. Since for this example we find $\gamma_m = O(m^{-2})$, Theorem 4.1 yields the convergence rate

$$\left\|x_\rho^{m,\delta} - x^\dagger\right\|_{L^2} = O\left((m^{-2} + \delta)^{\frac{1}{2}}\right) . \tag{4.2}$$

Here $x_\rho^{m,\delta}$ are the minimum norm quasi-solutions (cf. (4.1)) for $s = 2$. (Note that X_m is not only a subspace of $X_s = X_2$ but also a subspace of $X_{a+2s} = X_6$.)

Since $x^\dagger \in X_u$ for all $u < 5/2$, the best possible rate with respect to δ that is obtainable by Tikhonov regularization combined with an a-posteriori parameter choice is $O\left(\delta^{\frac{5}{9}}\right)$ (see [1]) if δ is known. Thus, the rate in (4.2) is not optimal. However, we do not need the knowledge of δ !

Finally, we present some numerical results: we have calculated the solutions with Algorithm 4.2 ($\varepsilon_1 = 10^{-6}, \varepsilon_2 = 10^{-16}, q = 0.1$) for $m = 4, 8, 16, 32, 64, 128, 256, 512$. Uniformly distributed noise was added to the data with the noise level chosen such that $\delta_m \sim m^{-2}$ and δ_4 was equal to 10% of $\|y\|$. For each m, 100 different perturbations were chosen and the worst L^2-case was selected: $\text{err}_\rho = \sup \left\|x_\rho^{m,\delta} - x^\dagger\right\|_{L^2}$.

The results for $\rho = 2, 3, 1.5$ can be found in Table 4.1: as expected, the error does not go to 0 in the X_2-norm for $\rho = 3$ and $\rho = 1.5$, however, it does for $\rho = 2$; note that $\left\|x^\dagger\right\|_2 = 2$. The solutions converge in L^2 for $\rho = 2$ and $\rho = 3$. Again, as expected, the results are better for the case $\rho = 2$. Of course, no convergence is obtained for $\rho = 1.5$ since $x^\dagger \notin M_\rho$. In this case α does not go to 0 as m goes to ∞.

ETNA
Kent State University
http://etna.math.kent.edu

REFERENCES

[1] H. W. ENGL, M. HANKE, AND A. NEUBAUER, *Regularization of Inverse Problems*, Kluwer, Dordrecht, 1996.
[2] B. HOFMANN, P. MATHÉ, AND M. SCHIECK, *Modulus of continuity for conditionally stable ill-posed problems in Hilbert space*, J. Inv. Ill-posed Probl., 16 (2008), pp. 567–585.
[3] V. K. IVANOV, *On linear problems which are not well-posed*, Dokl. Akad. Nauk SSSR, 145 (1962), pp. 270–272.
[4] ———, *On ill-posed problems*, Mat. Sb., 61 (1963), pp. 211–223.
[5] B. KALTENBACHER, A. NEUBAUER, AND O. SCHERZER, *Iterative Regularization Methods for Nonlinear Ill-Posed Problems*, vol. 6 of Radon Series on Computational and Applied Mathematics, de Gruyter, Berlin, 2008.
[6] S. KINDERMANN AND A. NEUBAUER, *On the convergence of the quasioptimality criterion for (iterated) Tikhonov regularization*, Inverse Probl. Imaging, 2 (2008), pp. 291–299.
[7] S. G. KREIN AND J. I. PETUNIN, *Scales of Banach spaces*, Russian Math. Surveys, 21 (1966), pp. 85–159.
[8] F. NATTERER, *The Mathematics of Computerized Tomography*, Teubner, Stuttgart, 1986.
[9] A. NEUBAUER, *Tikhonov regularization of nonlinear ill-posed problems in Hilbert scales*, Appl. Anal., 46 (1992), pp. 59–72.
[10] ———, *On Landweber iteration for nonlinear ill-posed problems in Hilbert scales*, Numer. Math., 85 (2000), pp. 309–328.
[11] ———, *The convergence of a new heuristic parameter selection criterion for general regularization methods*, Inverse Problems, 24 (2008), 055005 (10 pages).
[12] O. SCHERZER, M. GRASMAIR, H. GROSSAUER, M. HALTMEIER, AND F. LENZEN, *Variational Methods in Imaging*, vol. 167 of Applied Mathematical Sciences, Springer, New York, 2009.
[13] T. SCHUSTER, B. KALTENBACHER, B. HOFMANN, AND K. S. KAZIMIERSKI, *Regularization Methods in Banach Spaces*, vol. 10 of Radon Series on Computational and Applied Mathematics, de Gruyter, Berlin, 2012.
[14] A. N. TIKHONOV AND V. ARSENIN, *Solutions of Ill-Posed Problems*, Wiley, New York, 1977.