MATHEMATICS OF COMPUTATION
Volume 66, Number 217, January 1997, Pages 193–206
S 0025-5718(97)00811-9

THE TRADE–OFF BETWEEN REGULARITY AND STABILITY IN TIKHONOV REGULARIZATION

M. THAMBAN NAIR, MARKUS HEGLAND, AND ROBERT S. ANDERSSEN

Abstract. When deriving rates of convergence for the approximations generated by the application of Tikhonov regularization to ill–posed operator equations, assumptions must be made about the nature of the stabilization (i.e., the choice of the seminorm in the Tikhonov regularization) and the regularity of the least squares solutions which one looks for. In fact, it is clear from works of Hegland, Engl and Neubauer and Natterer that, in terms of the rate of convergence, there is a trade–off between stabilization and regularity. It is this matter which is examined in this paper by means of the best–possible worst–error estimates. The results of this paper provide better estimates than those of Engl and Neubauer, and also include and extend the best possible rate derived by Natterer. The paper concludes with an application of these results to first–kind integral equations with smooth kernels.

1. Introduction

In the solution of ill–posed operator equations, Tikhonov regularization with a suitable regularizing operator has played a seminal role (cf. [6] and [18]). The reasons are two–fold. On the one hand, it has had a considerable success in generating stable approximations to the solutions of practical inverse problems [7, 2]. On the other hand, it has useful equivalent mathematical representations, such as the Euler–Lagrange equations and its minimum–norm least squares interpretation, which allow a rigorous study of its mathematical and numerical properties (cf. Locker and Prenter [14, 15], Engl and Neubauer [3], Natterer [20], and Nair [19]). In fact, the motivation for detailed and careful investigations of the finer theoretical properties of Tikhonov regularization is the success of its application in the solution of practical inverse problems.

For example, such information is required in optimizing the choice of the regularization parameter. In particular, within this framework, a key consideration is the rate of convergence of Tikhonov approximations as a function of the regularization parameter. Because of both its practical as well as its theoretical importance, it is a topic which has already been examined from different points of view by various authors including Natterer [20], Engl and Neubauer [3], Neubauer [21], Hegland [10], Schock [23], Nair [19], and George and Nair [4, 5].

What is clear from the earlier works is that, in terms of the rate of convergence, there is a trade–off between the degree of stabilization built into the Tikhonov regularization and the regularity imposed on the solution. One sees this clearly in

Received by the editor April 13, 1994 and, in revised form, November 13, 1995.
1991 *Mathematics Subject Classification.* Primary 65R30; Secondary 65J20, 45B05.

the work of Hegland [10], as well as Engl and Neubauer [3], though the nature of the trade–off, which is the focus of this paper, is not explicitly explored.

2. Preliminaries

In this section, we introduce basic concepts and notation used throughout the paper.

We always let $K : X \to Y$ denote a bounded linear operator between Hilbert spaces X and Y with its range $R(K)$ not necessarily closed in Y. As the canonical ill–posed problem to which Tikhonov regularization will be applied, we consider the operator equation

$$Kx = y, \quad y \in Y. \tag{1}$$

Furthermore, we let L denote a densely defined closed linear operator with domain $D(L) \subset X$ and range $R(L) \subset Z$, where Z is also a Hilbert space. *Tikhonov regularization*, with *regularizing operator* L, applied to (1), is defined variationally as the problem of minimizing, for $\alpha > 0$ (cf. [25]), the Tikhonov functional G_α defined by

$$G_\alpha(x) := \|Kx - y\|^2 + \alpha \|Lx\|^2, \quad x \in D(L).$$

The Euler–Lagrange equation or *regularized equation* formulation of Tikhonov regularization is

$$(K^*K + \alpha L^*L)x_\alpha = K^*y, \tag{2}$$

where x_α denotes, for a fixed value of the *regularization parameter* α and the regularizing operator L, the Tikhonov regularization solution of (1). Any minimizer of G_α also solves the Euler-Lagrange equation, and is, thus, an element of $D(L^*L)$, and conversely, any solution of the Euler-Lagrange equation minimizes G_α (cf. Locker and Prenter [14]).

For $L = I$, equation (2) is uniquely solvable, and, if $y \in D(K^\dagger)$, the domain of the Moore–Penrose inverse $K^\dagger$, then $x_\alpha \to \hat{x}$ as $\alpha \to 0$, where $\hat{x} := K^\dagger y$ is the minimum–norm least squares solution of (1) (cf. Groetsch [6]).

From a practical point of view, however, it is more appropriate to seek the least squares solution x_0 which minimizes the seminorm $\|Lx\|$ for $L \neq I$ (cf. Varah [27]); i.e., one looks for x_0 such that

$$x_0 \in S_y := \{x \in D(L) \mid \|Kx - y\| \le \|Ku - y\|, \forall u \in D(L)\}$$

and

$$\|Lx_0\| \le \|Lx\| \quad \text{for all } x \in S_y.$$

In applications, K is often a compact integral operator with a nondegenerate kernel, and L a differential operator (cf. [15]).

Choices for the regularizing operator L, different from I, were suggested in the earliest papers discussing regularization [22]. Practical computations show that this approach can lead to smaller errors in some cases [26]. For special choices of L, theoretical improvements in convergence were established by Natterer [20]. A parameter choice strategy for such situations was proposed by Neubauer [21]; but, such deliberations do not cover the use of general differential regularizers L for the solution of first–kind integral equations with very smooth kernels (such as Fujita's equation [16]). Nevertheless, this situation has been examined by Hegland [10].

If the data y is only known approximately as y^δ, with $\|y - y^\delta\| \le \delta$, $\delta > 0$, then one must solve

$$(K^*K + \alpha L^*L)x_\alpha^\delta = K^*y^\delta \tag{3}$$

instead of (2). Here, the choice of the regularization parameter α (depending on δ and possibly y_δ) is important, for, in general, the family $\{x_\alpha^\delta\}_{\alpha>0}$ need not be bounded. As an example, consider a K such that $R(K)$ is not closed and $L = I$. In this case, the family $\{(K^*K + \alpha I)^{-1}K^*\}_{\alpha>0}$ is not uniformly bounded. It can then be seen from the Uniform Boundedness Principle, that, for each $\delta > 0$, there exists $y^\delta \in Y$ with $\|y - y^\delta\| \le \delta$ such that $\{x_\alpha^\delta\}_{\alpha>0}$ is not bounded in X.

Practical considerations suggest that it is desirable to choose the regularization parameter α during the computation of x_α^δ, using a so-called a posteriori method, rather than an a priori method based on δ only (cf. [1]). In fact, we will use a modified form of a method suggested by Schock [23] for the case $L = I$, where α is computed to satisfy

$$\|Kx_\alpha^\delta - y^\delta\| = \frac{\delta^p}{\alpha^q}, \quad p > 0,\ q > 0. \tag{4}$$

Convergence for this method has been further investigated by Nair [19], and by George and Nair [4, 5]. Below, this analysis is extended to a more general class of L.

In this paper, the error in a Tikhonov regularization approximation is compared with the *best possible worst error*

$$E_M(\rho,\delta) := \inf_R \sup\{\|x - R(y^\delta,\delta)\| \mid x \in D(M),\ y^\delta \in Y; \\ \|Mx\| \le \rho,\ \|Kx - y^\delta\| \le \delta\}.$$

The infimum runs over all *reconstruction algorithms* $R : Y \times (0, \delta_0] \to X$ for some $\delta_0 > 0$, where M is a densely defined linear operator related to L with domain $D(M) \subset X$ which specifies the 'smoothness' of the solution in some sense.

No regularization method and parameter choice is able to get errors of order less than $E_M(\rho,\delta)$ for a particular M and ρ.

It is proved in Micchelli and Rivlin [17] that

$$e_M(\rho,\delta) \le E_M(\rho,\delta) \le 2\ e_M(\rho,\delta), \tag{5}$$

where

$$e_M(\rho,\delta) := \sup\{\|x\| \mid x \in D(M),\ \|Mx\| \le \rho,\ \|Kx\| \le \delta\}.$$

A method for obtaining approximations $x^\delta := R(\delta, y^\delta)$ to x_0 corresponding to a reconstruction algorithm R, is said to be an *optimal–order regularization method*, with respect to an operator M, if one has

$$\|x_0 - x^\delta\| = O(e_M(\rho,\delta))$$

for all $x_0 \in D(M)$, with $\|Mx_0\| \le \rho$.

The above quantity $e_M(\rho,\delta)$ should be compared with the worst error that a regularization method, using a specific regularization operator, can generate under the most favorable smoothness conditions. Estimates of this kind are well known for the case $L = I$ and are representative of the *saturation phenomena* discussed in [8].

3. Solvability of the regularized equation

In order to guarantee the unique solvability of the regularized equations (2) and (3), we assume that the following *completion condition* (cf. Morozov [18, p.3])

$$\|Kx\|^2 + \|Lx\|^2 \geq \gamma \|x\|^2, \quad x \in D(L), \tag{6}$$

holds for some $\gamma > 0$.

In fact, we derive the following result, which is similar to that obtained by Locker and Prenter [14].

Proposition 3.1. *If the completion condition* (6) *holds, then*

(i) $(u, v)_* := (Ku, Kv) + (Lu, Lv)$, *with* $u, v \in D(L)$, *defines a complete inner product on* $D(L)$, *and*
(ii) *for every* $\alpha > 0$, *the operator* $K^*K + \alpha L^*L$ *is a closed, self-adjoint and bijective linear operator with domain* $D(L^*L)$, *and its inverse is a bounded linear operator.*

Proof. (i) The completion condition (6) implies the positive definiteness of $(\cdot, \cdot)_*$, and the remaining axioms for $(\cdot, \cdot)_*$ to be an inner product follow from its definition.

It remains to show that any Cauchy sequence with respect to the norm $\|\cdot\|_*$, defined by

$$\|x\|_* := \sqrt{(x, x)_*}, \quad x \in D(L),$$

converges. If (x_n) is a Cauchy sequence with respect to the norm $\|\cdot\|_*$, then it follows that (Kx_n) and (Lx_n) are both Cauchy sequences, in the Hilbert spaces Y and Z, respectively. By the completion condition (6), (x_n) is a Cauchy sequence in X. Thus, there exist limits of these three sequences, which can be defined formally as

$$x_n \to x, \quad Kx_n \to y \quad \text{and} \quad Lx_n \to z, \quad x \in X,\ y \in Y,\ z \in Z.$$

Since K is bounded, $Kx = y$ and, as L is closed, $x \in D(L)$ and $Lx = z$. Thus, (x_n) converges to x with respect to the norm $\|\cdot\|_*$.

(ii) Now, we observe (cf. [12]) that the operator L^*L is closed and self-adjoint, and K^*K is bounded and self-adjoint, and so the operator $K^*K + \alpha L^*L : D(L^*L) \to X$ is also closed and self-adjoint.

Then, from the completion condition (6) and the Schwarz inequality we have

$$\begin{aligned} \|(K^*K + \alpha L^*L)x\| \|x\| &\geq ((K^*K + \alpha L^*L)x, x) \\ &= \|Kx\|^2 + \alpha \|Lx\|^2 \\ &\geq \gamma \ \min\{1, \alpha\} \|x\|^2. \end{aligned}$$

Thus, the self-adjoint operator $K^*K + \alpha L^*L$ is bounded below, so that, by standard arguments, it follows that it is bijective, and its inverse is a bounded linear operator from X onto $D(L^*L)$. ☐

Locker and Prenter [14] derive their results under the assumption that

(i) $N(K) \cap N(L) = \{0\}$;
(ii) $R(L)$ is closed;
(iii) there exists a $\gamma_0 > 0$ such that

$$\|Kx\| \geq \gamma_0 \|x\| \quad \text{for all} \quad x \in N(L).$$

In the sequel, we will call these conditions the Locker-Prenter conditions. The next proposition explores the interrelationship between the completion condition and the Locker-Prenter conditions.

Proposition 3.2. *If the Locker-Prenter conditions* (i), (ii), (iii) *hold, then the completion condition* (6) *holds as well for some* $\gamma > 0$. *On the other hand, if the completion condition* (6) *holds, then the Locker-Prenter conditions* (i) *and* (iii) *hold.*

Proof. The operator F with $Fx := (Kx, Ly) \in Y \times Z$ maps $D(L)$ into the product space $Y \times Z$, which is a Hilbert space with respect to the induced inner product.

Under the conditions (i),(ii), (iii), it can be seen (as in [14, Lemma 5.1]) that F is an injective closed linear operator with its range $R(F)$ closed in $Y \times Z$. Therefore, the inverse operator $F^{-1} : R(F) \to X$ is a closed linear operator between the Hilbert spaces $R(F)$ and X, so that, by the Closed Graph Theorem, F^{-1} is a bounded linear operator; i.e., there exists a constant $c > 0$ such that $\|F^{-1}(Fx)\| \leq c\|Fx\|, x \in D(L)$; i.e., $\gamma\|x\|^2 \leq \|Kx\|^2 + \|Lx\|^2, x \in D(L)$, where $\gamma = 1/c^2$.

Starting from (6), the Locker-Prenter conditions (i) and (iii) are obvious consequences. $\square$

Remark. The Locker-Prenter condition (ii) is not a natural consequence of (6).

Clearly, (6) is satisfied if L is bounded below, which occurs when L is a strictly positive definite and self-adjoint operator with $X = Z$.

In view of Proposition 3.1, let $(\cdot,\cdot)_*$ be the inner product and $\|\cdot\|_*$ be the corresponding norm on $D(L)$ defined by

$$(u, v)_* := (Ku, Kv) + (Lu, Lv), \quad u, v \in D(L),$$

and

$$\|x\|_* := (\|Kx\|^2 + \|Lx\|^2)^{\frac{1}{2}}, \quad x \in D(L),$$

respectively. Let X_* be the Hilbert space $D(L)$ with the inner product $(\cdot,\cdot)_*$ and let

$$A : X_* \to Y \quad \text{be defined by} \quad Ax = Kx, \ x \in X_*;$$

i.e., A is the restriction of K to the space $D(L)$ with inner product $(\cdot,\cdot)_*$. Let $A^\sharp$ and $L^\sharp$ be the adjoints of $A : X_* \to Y$ and $L : X_* \to Z$, respectively. It is shown in Locker and Prenter [14] that, on their respective domains,

(7)
$$A^\sharp = (K^*K + L^*L)^{-1}K^*, \quad L^\sharp = (K^*K + L^*L)^{-1}L^* \quad \text{and} \quad A^\sharp A + L^\sharp L = I.$$

As usual, the Moore–Penrose inverse of $A : X_* \to Y$ will be denoted by $A^\dagger$. Note that, in general, $A^\dagger \neq K^\dagger$.

4. Earlier work

In this section, we revisit some recent results on optimal–order methods. The choice of the regularizing operator imposes restrictions on the convergence rate. It is shown in [6, 24] that, for $L = I$, the best possible convergence rate is $\|x_\alpha^\delta - x_0\| = O(\delta^{2/3})$, and that this can be obtained if $x_0 \in R(K^*K)$ and $\alpha = c\delta^{2/3}$ for some constant $c > 0$. If $x_0 \in R((K^*K)^\nu)$ with $\nu > 1$, the same rate holds, while, for $\nu < 1$, it is smaller (cf. [6, 24]).

For compact, nondegenerate K, Tikhonov regularization, with $L = I$ and $x_0 \in R(K^*K)$, is optimal with respect to $M = (K^*K)^\dagger$, since

$$e_M(\rho, \delta) = O(\delta^{2/3}).$$

However, when $x_0 \in R((K^*K)^\nu)$, $M = [(K^*K)^\dagger]^\nu$, and $\nu \geq 1$, one obtains

$$e_M(\rho, \delta) = O(\delta^{2\nu/(2\nu+1)}),$$

while, for other M, one can achieve even higher rates (cf. Hegland [10]). In such situations, Tikhonov regularization, with $L = I$, does not have optimal convergence rates. Nevertheless, optimality can be achieved through an appropriate choice of the regularizing operators (cf. [9, 10]).

An example of Tikhonov regularization having optimal convergence rates was given by Natterer [20]. He considered the case $Z = X$ and $L = T^k$, $k > 0$, where $T : D(T) \to X$ is a densely defined strictly positive definite and self-adjoint operator. Furthermore, he assumed that there exist positive reals γ_1 and γ_2 such that

$$\gamma_1 \|x\|_{-a} \leq \|Kx\| \leq \gamma_2 \|x\|_{-a}, \quad x \in X, \tag{8}$$

for some positive real a. Here, $\|x\|_r := \|T^r x\|$, $x \in D(T^r)$, for real r. Taking H_r to be the Hilbert space obtained through the completion of $\bigcap_{i=1}^\infty D(T^i)$ with respect to the norm $x \mapsto \|x\|_r$, Natterer proved the following result.

Theorem 4.1 (Natterer [20]). *If $x_0 \in H_s$ with $\|x_0\|_s \leq \rho$, $s \leq 2k + a$ and $\alpha = c(\frac{\delta}{\rho})^{\frac{2(a+k)}{a+s}}$, then*

$$\|x_0 - x_\alpha^\delta\| = O(\delta^{\frac{s}{a+s}}).$$

In particular, if $s = 2k + a$, then

$$\|x_0 - x_\alpha^\delta\| = O(\delta^{\frac{a+2k}{2a+2k}}).$$

The above estimate of Natterer is optimal with respect to the choice $M = T^s$, as it is known, for this case, that $e_M(\rho, \delta) = O(\delta^{\frac{s}{a+s}})$. Note that, in the above result, an a priori parameter choice was used. Neubauer [21] suggested an a posteriori method which leads to the above result of Natterer. A similar result using an a posteriori choice can be found in [9]. Thus, for $x \in R((K^*K)^\nu)$ (e.g., $T = [(K^*K)^{\frac{1}{2}}]^\dagger$), $s = 2\nu$ and $a = 1$, the order is $O(\delta^{2\nu/(2\nu+1)})$ if $k \geq \nu - 0.5$. This is achieved by imposing minimal 'smoothness' on the regularizing operator L. However, by choosing a smoother operator, the convergence rate governed by the smoothness of the data is maintained. This idea was further pursued by Hegland [10] and has led to a method which gives optimal convergence rates for all $\nu > 0$.

Theorem 4.2 (Hegland [10]). *Let K be compact with singular value decomposition $K = \sum_{i=0}^\infty \sigma_i u_i \otimes v_i$. Furthermore, let*

$$L = \sum_{i=0}^\infty (\sigma_0/\sigma_i)^{\log(i)} v_i \otimes v_i$$

and α be chosen such that $\|Kx_\alpha^\delta - y^\delta\| = 2\delta$. Then, for any $\nu > 0$ and $x \in R((K^*K)^\nu)$, one has optimal convergence, i.e.,

$$\|x_0 - x_\alpha^\delta\| = O(\delta^{2\nu/(2\nu+1)}).$$

In his convergence analysis for $L = I$, Schock [23] proved, for his parameter choice (4) with $p = 4q(q+1)/(6q+1)$, that

$$\|x_0 - x_\alpha^\delta\| = O(\delta^{2/(3+0.5q^{-1})})$$

for $x_0 \in R(K^*K)$. This is asymptotically optimal in q.

Recently, it was found by Nair [19] and George and Nair [4, 5] that Schock's parameter choice can actually also give rise to optimal convergence rates.

Theorem 4.3 (George and Nair [4, 5]). *For a fixed pair of positive reals p and q, there exists a unique $\alpha = \alpha(\delta)$ satisfying* (4), *and we have the following*:

(i) $\alpha = O(\delta^{\frac{p}{q+1}})$;

(ii) *if $y \in R(K)$, $\hat{x} := K^\dagger y \in R((K^*K)^\nu)$, $\nu \geq 0$, and*

$$\frac{p}{q+1} \leq \frac{1}{\omega} \quad \textit{with} \quad \omega = \min\{1, \nu + \frac{1}{2}\},$$

then

$$\frac{\delta^p}{\alpha^q} = O(\delta^t), \quad t = \frac{p\omega}{q+1};$$

(iii) *if $p < 2q + \frac{2q}{2q+1}$, then*

$$\|\hat{x} - x_\alpha^\delta\| \to 0 \quad \textit{as} \quad \delta \to 0;$$

(iv) *if $y \in R(K)$, $\hat{x} \in R((K^*K)^\nu)$, $0 \leq \nu \leq 1$, and*

$$\frac{p}{q+1} \leq \min\{\frac{1}{\omega},\ \frac{2}{1 + (\frac{1-\omega}{q})}\},$$

then

$$\|\hat{x} - \hat{x}_\alpha^\delta\| = O(\delta^\eta),$$

where $\eta = \min\{\frac{p\nu}{q+1}, 1 - \frac{p}{2(q+1)}(1 + \frac{1-\omega}{q})\}$.

In particular, when $x_0 \in R(K^*K)$ and $p = 2(q+1)$, these authors establish $\|x_\alpha^\delta - x_0\| = O(\delta^{2/3})$. Similar results, for other L, are derived below.

For K and L satisfying the Locker-Prenter conditions (i), (ii), (iii), Engl and Neubauer [3] considered the discrepancy principle

$$\|(1-\alpha)A^\sharp A x_\alpha^\delta - A^\sharp y^\delta\| = \delta^p \left(\frac{1-\alpha}{\alpha}\right)^q, \quad 0 < \alpha < 1,\ p > 0,\ q > 0, \tag{9}$$

where $A^\sharp$ is as in (7), and proved the following result.

Theorem 4.4 (Engl and Neubauer [3]). *If $2q \geq p > 0$ and $y \in D(A^\dagger)$, then for $\delta > 0$ small enough, there exists $\alpha := \alpha(\delta) \in (0,1)$ such that* (9) *is satisfied and*

$$\|x_0 - x_\alpha^\delta\|_* \to 0 \quad \textit{as} \quad \delta \to 0,$$

where $x_0 = A^\dagger y$.

Moreover,

(a) $p = \frac{2}{3}(q+1)$, $\quad x_0 \in D(L^*L) \quad$ *and* $\quad L^*Lx_0 \in R(K^*K)$ *imply*

$$\|x_0 - x_\alpha^\delta\|_* = O(\delta^{\frac{2}{3}}); \tag{10}$$

(b) $p = q+1$, $\quad x_0 \in D(L^*L) \quad$ *and* $\quad L^*Lx_0 \in R(K^*)$ *imply*

$$\|x_0 - x_\alpha^\delta\|_* = O(\delta^{\frac{1}{2}}). \tag{11}$$

Note, in particular, that for $p = 2$ and $q = 1$, and $L = I$, these authors obtain the optimal error rate $O(\delta^{2/3})$ provided $x_0 \in R(K^*K)$.

Since, by (6), $\|x_0 - x_\alpha^\delta\|_* \geq \|x_0 - x_\alpha^\delta\|$, it follows that (10) and (11) also yield estimates for the error with respect to the original norm on X. A natural question which arises is whether one can obtain better estimates for $\|x_0 - x_\alpha^\delta\|$ than the ones provided in (10) and (11). Clearly, convergence in the original norm does not have to be faster than in the norm $\|\cdot\|_*$. For example, if $u_n = \lambda_n u$ for some fixed $u \in D(L)$, then this sequence is of order $O(\lambda_n)$ with respect to both norms.

5. Characterization of the trade–off

In this section, it is shown how, in conjunction with a modified form of the parameter choice strategy (4), convergence rates can be improved through the choice of appropriate regularizing operators.

Initially, we observe (cf. [14, Remark 4.6]) that the least squares solution $x_0 \in S_y$, minimizing the norm $x \mapsto \|x\|_*$, is exactly the same as the one which minimizes the seminorm $x \mapsto \|Lx\|$. In fact, since $Kx = Kx_0$ for every least squares solution $x \in S_y$, it follows that

$$\|Kx_0\|^2 + \|Lx_0\|^2 = \|x_0\|_*^2 \leq \|Kx\|^2 + \|Lx\|^2,$$

and hence,

$$\|Lx_0\|^2 \leq \|Lx\|^2.$$

Thus, the least squares solution x_0 minimizes the seminorm $x \mapsto \|Lx\|$. The converse is also true.

The key to understanding the general case is the following argument of Locker and Prenter [14], which reduces the situation to the case $L = I$. With $0 < \alpha < 1$, let x_α and x_α^δ be the solutions of the equations (2) and (3), respectively. Applying the operator $(K^*K + L^*L)^{-1}$ to both sides of the equations (2) and (3) yields

$$(A^\sharp A + \beta I)x_\alpha = (1+\beta)A^\sharp y$$

and

$$(A^\sharp A + \beta I)x_\alpha^\delta = (1+\beta)A^\sharp y^\delta, \tag{12}$$

respectively, where $\beta = \frac{\alpha}{1-\alpha}$, $\quad 0 < \alpha < 1$. These equations are the regularized equations for the problem $Ax = (1+\beta)y$, $A : X_* \to Y$, with regularizing operator $L = I$, and regularization parameter β. Since $(1+\beta)y \to y$ as $\beta \to 0$, it follows that (using standard arguments as in [6])

$$\|x_0 - x_\alpha\|_* \to 0 \quad \text{as} \quad \alpha \to 0$$

provided $y \in D(A^\dagger)$, and furthermore

$$\|x_\alpha - x_\alpha^\delta\|_* \leq (1+\beta)\frac{\delta}{\sqrt{\beta}}.$$

By the triangle inequality,

$$\|x_0 - x_\alpha^\delta\|_* \leq \|x_0 - x_\alpha\|_* + (1+\beta)\frac{\delta}{\sqrt{\beta}},$$

where x_0 is the least squares solution of (1) minimizing the norm $x \mapsto \|x\|_*$.

We propose that the regularization parameter α in (3) (or equivalently, β in (12)) be chosen so that the equality

$$\|Kx_\alpha^\delta - (1+\beta)y^\delta\| = \frac{\delta^p}{\beta^q} \tag{13}$$

holds for some preassigned $p > 0$ and $q > 0$. The above parameter strategy has properties similar to (9), as established by Engl and Neubauer [3]. Here, however, it is simpler to evaluate.

The following theorem is a generalization of Theorem 4.3 to the case of a general L.

Theorem 5.1. *For a fixed pair (p, q) of positive reals and for each δ sufficiently small, say $0 < \delta < \delta_0$, there exists a unique $\alpha := \alpha(\delta)$ satisfying* (13). *Moreover,*

(i) $\beta = O(\delta^{\frac{p}{q+1}})$;

(ii) *if $y \in R(A)$, $x_0 \in R((A^\sharp A)^\nu)$, $\nu \geq 0$, and*

$$\frac{p}{q+1} \leq \frac{1}{\omega} \quad \textit{with} \quad \omega = \min\{1, \nu + \frac{1}{2}\},$$

then

$$\frac{\delta^p}{\beta^q} = O(\delta^t), \quad t = \frac{p\omega}{q+1};$$

(iii) *if $p < 2q + \frac{2q}{2q+1}$, then*

$$\|x_0 - x_\alpha^\delta\|_* \to 0 \quad \textit{as} \quad \delta \to 0;$$

(iv) *if $y \in R(A)$, $x_0 \in R((A^\sharp A)^\nu)$, $0 \leq \nu \leq 1$, and*

$$\frac{p}{q+1} \leq \min\{\frac{1}{\omega},\ \frac{2}{1 + (\frac{1-\omega}{q})}\},$$

then

$$\|x_0 - x_\alpha^\delta\|_* = O(\delta^\eta),$$

with $\eta = \min\{\frac{p\nu}{q+1}, 1 - \frac{p}{2(q+1)}\left(1 + \frac{1-\omega}{q}\right)\}$ and $\omega = \min\{1, \nu + \frac{1}{2}\}$.

Proof. The proof proceeds exactly along the same lines as for the case $L = I$ (cf. Theorem 4.3). □

Remark. From (7) it follows that

$$R(A^\sharp A) = \{x \in D(L^*L) : L^*Lx \in R(K^*K)\}$$

and

$$R(A^\sharp) = \{x \in D(L^*L) : L^*Lx \in R(K^*)\}, \tag{14}$$

so that Theorem 5.1 includes the conclusions of Theorem 4.4 on taking $\frac{p}{q+1} = \frac{2}{3}, \nu = 1$, to obtain (10), and $\frac{p}{q+1} = 1$, $\nu = \frac{1}{2}$, to obtain (11) .

As a consequence, we can now derive the following theorem about the convergence rate in the original norm.

Theorem 5.2. *If* $y \in R(A)$, $x_0 \in R((A^\sharp A)^\nu)$, $0 \le \nu \le 1$,

$$\frac{p}{q+1} \le \min\{\frac{1}{\omega}, \frac{2}{1+\frac{1-\omega}{q}}\} \quad \text{with} \quad \omega = \min\{1, \nu + \frac{1}{2}\},$$

and $M : D(L) \to Y \times Z$ *defined by* $Mx = (Kx, Lx)$, $x \in D(L)$, *then*

$$\|x_0 - x_\alpha^\delta\| \le c\ e_M(\delta^\eta, \delta^t),$$

where $\eta = \min\{\frac{p\nu}{q+1}, 1 - \frac{p}{2(q+1)}\left(1 + \frac{1-\omega}{q}\right)\}$, $t = \frac{p\omega}{q+1}$, *and* $c > 0$ *is a constant.*

In particular, if $p = q + 1$, *then*

$$\|x_0 - x_\alpha^\delta\| \le c\ e_M(\delta^{1/2}, \delta),$$

for every $\nu \ge \frac{1}{2}$; *and if* $p = \frac{2}{3}(q+1)$, *then*

$$\|x_0 - x_\alpha^\delta\| \le c\ e_M(\delta^{2/3}, \delta^{2/3})$$

for $\nu = 1$.

Proof. By the definition of the norm on the product space $Y \times Z$, we have $\|Mx\| = \|x\|_*$, $x \in D(L)$. Now, Theorem 5.1 (iv) implies

$$\|M(x_0 - x_\alpha^\delta)\| \le c\ \delta^\eta,$$

and Theorem 5.1 (i), (ii) imply

$$\begin{aligned} \|K(x_\alpha^\delta - x_0)\| &= \|Kx_\alpha^\delta - y\| \\ &\le \|Kx_\alpha^\delta - (1+\beta)y^\delta\| + \|(1+\beta)y^\delta - y\| \\ &\le \frac{\delta^p}{\alpha^q} + \beta\|y\| + (1+\beta)\delta\| \\ &= O(\delta^t). \end{aligned} \tag{15}$$

From these observations, the results follow. ☐

In order to discuss this result further, we need to make some assumptions about the nature of our operators K and L. We will apply the interpolation inequality in the framework of Hilbert scales (cf. [13]).

Let $T : D(L) \to X$ be a densely defined, strictly positive definite and self-adjoint operator on X. The Hilbert scale $\{X_a\}_{a \in \mathbf{R}}$ is the family of Hilbert spaces, where X_a is the completion of $\bigcap_{k=1}^\infty D(T^k)$ with respect to the norm $x \mapsto \|x\|_a := \|T^a x\|$, and T^a is defined using the spectral representation. We note that $\|x\| = \|x\|_0$ for all $x \in X$.

If $a \le b$, then there is a continuous embedding $X_b \hookrightarrow X_a$, and therefore the norm $\|\cdot\|_a$ is also defined on X_b. Further, for $a \le b \le c$, the *interpolation inequality*

$$\|x\|_b \le \|x\|_a^\theta \|x\|_c^{1-\theta}, \qquad \theta = \frac{c-b}{c-a}, \tag{16}$$

holds for all $x \in X_c$ (cf. [13]).

Hilbert scales were used by Natterer [20] in proving Theorem 4.1, and the following result can be viewed as a generalization of his result.

Theorem 5.3. *Let* K *and* L *be such that for some* $a > 0$, $r > 0$, $c_1 > 0$ *and* $c_2 > 0$, *the conditions*

$$D(L) \subseteq X_r,$$

$$\|x\|_{-a} \leq c_1 \|Kx\|, \quad x \in X,$$

and

$$\|x\|_r \leq c_2 \|x\|_*, \quad x \in D(L),$$

are satisfied. Then, under the assumptions and notations of Theorem 5.2*,*

$$\|x_0 - x_\alpha^\delta\| \leq c_3 \delta^\ell, \quad \ell = \frac{rt + a\eta}{r + a},$$

for some constant $c_3 > 0$*.*

Proof. From the interpolation inequality (16) and the conditions on K and L, we get

$$\begin{aligned} \|x\| &\leq \|x\|_r^{\frac{a}{r+a}} \|x\|_{-a}^{\frac{r}{r+a}} \\ &\leq c_3 \|x\|_*^{\frac{a}{r+a}} \|Kx\|^{\frac{r}{r+a}}. \end{aligned}$$

Now, Theorem 5.1 (iv) and the relation (15) imply the result. □

Corollary 5.4. *Let* $L = T^k$ *for some* $k > 0$ *and*

$$\|x\|_{-a} \leq \|Kx\|, \quad x \in X,$$

for some $a > 0$*. Then, under the assumptions and notations of Theorem* 5.2*,*

$$\|x_0 - x_\alpha^\delta\| = O(\delta^\ell), \quad \ell = \frac{tk + a\eta}{k + a}.$$

Proof. This follows from Theorem 5.3 on taking $r = k$. □

6. Concluding remarks

6.1. From the definition of x_α^δ, it follows that $x_\alpha^\delta \in D(L^*L)$ and $L^*Lx_\alpha^\delta \in R(K^*)$, and thus, using (14), that $x_\alpha^\delta \in R(A^\sharp) = R((A^\sharp A)^{\frac{1}{2}})$. Therefore, on taking

$$\tilde{M} : D(\tilde{M}) := R(A^\sharp) \to Y \times Z \quad \text{defined by} \quad \tilde{M}x = (Kx, Lx),\ x \in R(A^\sharp),$$

instead of M, and $\nu = \frac{1}{2}$ in Theorem 5.2, it follows that

$$\|x_0 - x_\alpha^\delta\| = O(e_{\tilde{M}}(\delta_0, \delta))$$

for all $x_0 \in \tilde{M}$. This shows that the above method, defined by $R(y^\delta, \delta) = x_\alpha^\delta$, is optimal with respect to the operator $\tilde{M}$.

6.2. If $\nu \geq \frac{1}{2}$ in Theorems 5.1 and 5.2, then the maximum value of η is attained when $\frac{p}{q+1} = \frac{2}{2\nu+1}$, for which η equals $\frac{2\nu}{2\nu+1}$. Consequently, the maximum value of ℓ in Theorem 5.3 is given by

$$\ell(\nu) = \frac{2}{2\nu + 1} \left(\frac{r + a\nu}{r + a} \right),\ \frac{1}{2} \leq \nu \leq 1.$$

It is easily seen that, if $r \geq \frac{a}{2}$, then the function $\nu \mapsto \ell(\nu)$ is decreasing with respect to $\nu \in [\frac{1}{2}, 1]$, and

$$\ell(1/2) = \frac{2r + a}{2(r + a)}, \quad \ell(1) = 2/3.$$

Thus, for $\nu \geq \frac{1}{2}$ and $r \geq \frac{a}{2}$, the best rate is obtained when $\frac{p}{q+1} = 1$, i.e.,

$$\|x_0 - x_\alpha^\delta\| = O(\delta^\ell), \quad \ell = \frac{2r+a}{2(r+a)}.$$

Note that the above convergence rate is independent of ν, i.e., large values of ν have no effect on the convergence rate unless a different regularizing operator L is introduced. However, by requiring L to have appropriate smoothing properties, we get a convergence rate close to $O(\delta)$. This is clear from Corollary 5.4 with $\nu = \frac{1}{2}$, for then the error estimate is

$$\|x_0 - x_\alpha^\delta\| = O(\delta^{\frac{2k+a}{2k+2a}}) \approx O(\delta)$$

for large values of k.

The above rate is the best possible rate of Natterer [20] (cf. Theorem 3.2) obtained using similar assumptions.

If we choose a weaker regularizing operator with $r \leq \frac{a}{2}$, we see that the function $\nu \mapsto \ell(\nu)$ is increasing on $[\frac{1}{2}, 1]$, and, in this case, we have $\ell(\nu) \geq \frac{2r+a}{2(r+a)}$, $\nu \geq \frac{1}{2}$, and the best rate possible is $O(\delta^{2/3})$, which is attained for $\nu = 1$ and $\frac{p}{q+1} = \frac{2}{3}$. Note that this rate is already obtained for $r = 0$! Thus, for less smooth data, we can end up with lower convergence rates than the ones possible for the case $L = I$.

The case of $\nu \leq \frac{1}{2}$ can also be analyzed in a similar manner by using the function

$$\nu \mapsto \ell(\nu) := \frac{2}{2\nu + 1 + \frac{1-\omega}{q}} \left(\frac{r\omega + a\nu}{r+a} \right)$$

with $\omega = \nu + \frac{1}{2}$, $\nu \in [0, \frac{1}{2}]$.

6.3. If $y \in R(A)$ and no additional regularity is assumed for x_0 (i.e., $\nu = 0$), then $\omega = \frac{1}{2}$, so that the requirement on p and q in Theorems 5.2 and 5.3 is $\frac{p}{q+1} \leq \frac{4q}{2q+1}$. In this case, taking $\frac{p}{q+1} = \frac{4q}{2q+1}$, we have $\eta = 0$, and $t = \frac{2q}{2q+1}$, so that the results in Theorems 5.2 and 5.3 yield

$$\|x_0 - x_\alpha^\delta\| \leq c\ e_M(1, \delta^t), \quad t = \frac{2q}{2q+1},$$

and

$$\|x_0 - x_\alpha^\delta\| = O(\delta^\ell), \quad \ell = \frac{2qr}{(2q+1)(r+a)},$$

respectively, whereas the results in Theorems 4.3 and 4.4 yield only

$$\|x_0 - x\| \leq \|x_0 - x_\alpha^\delta\|_* = O(1).$$

6.4. **Integral equations.** The previous theory is applied to the practical problem of reconstructing solutions of integral equations from measured data. In this case, $X = Y = L_2[0,1]$, while K is a Fredholm integral operator of the first kind,

$$Kx(s) = \int_0^1 k(s,t)x(t)dt, \quad s \in [0,1].$$

Examples include the Laplace transform where $k(s,t) = e^{-st}$.

An easily computed regularizer is

$$L = d^k/dt^k.$$

For this choice, the theory of Natterer [20] is not applicable to integral operators K where the kernels are smooth like those for the Laplace transform. This is so because the condition $\gamma\|x\|_{-a} \leq \|Kx\|$ (cf. equation (8)) cannot hold for all x in such situations.

Theorem 4.4 is more widely applicable than Natterer's [20] but it gives convergence rates independent of the regularizer L. On the other hand, Theorem 5.2 is also widely applicable, and, in particular, it shows how the choice of the regularizer affects the convergence rates. The errors are expressed in terms of the "best worst-case error" e_M of the operator $M = (K, L)$. It remains to compute e_M for particular operators. This has been done in [10, 11] for operators satisfying the conditions of Theorem 3.1 in [11].

In [11], the method is applied to three examples: Numerical differentiation, the numerical solution of Abel's integral equation, and the solution of Fredholm integral equations of the first kind with smooth kernels. It is shown how convergence rates very close to $O(\delta)$ can be achieved. However, in the case of very smooth kernels $k(s,t)$, it is also observed in [11] that the choice of differential operators as regularizers can only have a limited effect on convergence. Typically, in such situations, convergence is only improved by a factor $\log(1/\delta)$, so that other regularizers are required if enhanced performance is required.

Acknowledgement

This work was carried out during the visit of M.T. Nair to the Centre for Mathematics and its Applications, Australian National University, Canberra (June–December, 1993). The support received is gratefully acknowledged.

References

1. F. de Hoog, Review of Fredholm equations of the first kind, In *Application and Numerical Solution of Integral Equations*, (R.S.Anderssen, F.de Hoog and M.A.Lucas, eds.), Sijthoff & Noordhoff, Alphen aan den Rijn, Germantown, 1980, pp. 119–134. MR **82a:**45002
2. H. W. Engl and C.W. Groetsch, eds., *Inverse and Ill–Posed Problems*, Academic Press, London, 1987. MR **90f:**00004
3. H.W. Engl and A. Neubauer, Optimal discrepancy principles for the Tikhonov regularization of integral equations of the first kind, In *Constructive Methods for the Practical Treatment of Integral Equations*, (G.Hämmerlin and K.-H.Hoffmann, eds.), Birkhäuser, Basel, Boston, Stuttgart, 1985, pp. 120–141. CMP 19:10
4. S. George and M.T. Nair, Parameter choice by discrepancy principles for ill–posed problems leading to optimal convergence rates, *J. Optim. Theory Appl.*, **83** (1994), 217–222. MR **95i:**65094
5. S. George and M. T. Nair, On a generalized Arcangeli's method for Tikhonov regularization with inexact data, *Research Report, CMA–MR 43–93; SMS–88–93, Australian National University*, 1993.
6. C. W. Groetsch, *The Theory of Tikhonov Regularization for Fredholm Equations of the first Kind*, Pitman, Boston, 1984. MR **85k:**45020
7. C. W. Groetsch, *Inverse Problems in the Mathematical Sciences*, Vieweg Publishing, Wiesbaden, 1993. MR **94m:**00008
8. C. W. Groetsch and J. T. King The saturation phenomena for Tikhonov regularization. *J. Austral. Math. Soc.* (*Series A*) **35** (1983), 254–262. MR **84k:**47011
9. M. Hegland, *Numerische Lösung von Fredholmschen Integralgleichungen erster Art bei ungenauen Daten*. PhD thesis, ETHZ, 1988.
10. M. Hegland, An optimal order regularization method which does not use additional smoothness assumptions, *SIAM J. Numer. Anal.* **29** (1992), 1446–1461. MR **93j:**65090
11. M. Hegland, Variable Hilbert Scales and their Interpolation Inequalities with Applications to Tikhonov Regularization, *Applicable Anal.* **59** (1995), 207–223. CMP 96:09
12. T. Kato, *Perturbation Theory for Linear Operators*, 2nd Edition, Springer–Verlag, Berlin, Heidelberg, New York, 1976. MR **53:**11389

13. S. G. Krein and J. I. Petunin, Scales of Banach spaces, *Russian Math. Surveys* **21** (1966), 85–160. MR **33:**1719
14. J. Locker and P. M. Prenter, Regularization with differential operators. I: General theory, *J. Math. Anal. and Appl.* **74** (1980), 504–529. MR **83j:**65062a
15. J. Locker and P. M. Prenter, Regularization with differential operators. II: Weak least squares finite element solutions to first kind integral equations, *SIAM J. Numer. Anal.* **17** (1980), 247–267. MR **83j:**65062b
16. J. T. Marti, Numerical solution of Fujita's equation, In *Improperly Posed Problems and Their Numerical Treatment*, Intern. Ser. Numer. Math. **63** (G.Hämmerlin and K.-H. Hoffmann, eds.), Birkhäuser, Basel, 1983, pp. 179–187. MR **86a:**65134
17. C. A. Micchelli and T. J. Rivlin, A survey of optimal recovery, In *Optimal Estimation in Approximation Theory* (C.A.Micchelli and T.J.Rivlin, eds.), Plenum Press, New York, London, 1977, pp. 1–53. MR **58:**29718
18. V. A. Morozov, *Methods of Solving Incorrectly Posed Problems*, Springer–Verlag, New York, Berlin Heidelberg, 1984. MR **86d:**65005
19. M. T. Nair, A generalization of Arcangeli's method for ill–posed problems leading to optimal rates, *Integral Equations Operator Theory* **15** (1992), 1042–1046. MR **93j:**65091
20. F. Natterer, Error bounds for Tikhonov regularization in Hilbert scales, *Applicable Anal.* **18** (1984), 29–37. MR **86e:**65081
21. A. Neubauer, An a posteriori parameter choice for Tikhonov regularization in Hilbert scales leading to optimal convergence rates. *SIAM J. Numer. Anal.* **25** (1988), 1313–1326. MR **90b:**65108
22. D. L. Phillips, A technique for the numerical solution of certain integral equations of the first kind. *J. Assoc. Comput. Mach.* **9** (1962), 84–97. MR **24:**B534
23. E. Schock, Parameter choice by discrepancy principles for the approximate solution of ill–posed problems, *Integral Equations Operator Theory* **7** (1984), 895–898. MR **86m:**65165
24. E. Schock, On the asymptotic order of accuracy of Tikhonov regularizations, *J. Optim. Theory Appl.* **44** (1984), 95–104. MR **86c:**47012
25. A. N. Tikhonov and V. Y. Arsenin, *Solutions of Ill–Posed Problems*, Wiley, New York, 1977. MR **56:**13604
26. M. R. Trummer, A method for solving ill-posed linear operator equations, *SIAM J. Numer. Anal.* **21** (1984), 729–737. MR **85m:**65051
27. J. M. Varah, Pitfalls in the numerical solution of linear ill–posed problems, *SIAM J. Sci. Stat. Comput.* **4** (1983), 164–176. MR **84g:**65052

Centre for Mathematics and Its Applications, Australian National University, Canberra ACT 0200, Australia

Current address: Department of Mathematics, Indian Institute of Technology, Madras - 600 036, India

E-mail address: `mtnair@acer.iitm.ernet.in`

Centre for Mathematics and Its Applications, Australian National University, Canberra ACT 0200, Australia

Current address: Computer Sciences Laboratory, RSISE, Australian National University, Canberra ACT 0200, Australia

E-mail address: `Markus.Hegland@anu.edu.au`

Centre for Mathematics and Its Applications, Australian National University, Canberra ACT 0200, Australia

Current address: CSIRO Division of Mathematics and Statistics, GPO Box 1965, Canberra ACT 2601, Australia

E-mail address: `boba@cbr.dms.csiro.au`