

THE SATURATION PHENOMENA FOR TIKHONOV REGULARIZATION

C. W. GROETSCH and J. T. KING

(Received 22 April 1982)

Communicated by R. O. Vyborny

Abstract

This paper is concerned with a characterization of the optimal order of convergence of Tikhonov regularization for first kind operator equations in terms of the “smoothness” of the data.

1980 *Mathematics subject classification* (*Amer. Math. Soc.*): 47 B 30, 65 J 05, 45 E 99.

Keywords and phrases: regularization, first kind operation equation, order of convergence.

1. Introduction

This paper deals with some aspects of the relationship between the “smoothness” of a vector b and the rate of convergence of certain regularization methods as applied to an ill-posed operator equation

$$(1) \quad Ku = b.$$

We assume throughout that K is a bounded linear operator from a Hilbert space H_1 into a Hilbert space H_2 and our aim is to approximate the minimal norm least squares solution of equation (1). The Moore-Penrose generalized inverse of K is the closed linear operator K^\dagger , with domain $\mathcal{D}(K^\dagger) = R(K) + R(K)^\perp$, which associates with each vector $b \in \mathcal{D}(K^\dagger)$ the unique least squares solution, with minimal norm, of equation (1). That is, $K^\dagger b$ is the vector of minimal norm which satisfies the equation

$$(2) \quad Ku = Qb$$

The work of the first author was partially supported by the United Kingdom Science Research Council while he was a Senior Visiting Fellow at the University of Manchester.

© 1983 Australian Mathematical Society 0263-6115/83 \$A2.00 + 0.00

where Q is the orthogonal projection of H_2 onto $\overline{R(K)}$, the closure of the range of K (see for example [3] for equivalent characterizations).

It is easy to see that (2) is equivalent to the so-called normal equation

$$(3) \quad K^*Ku = K^*b$$

where K^* denotes the adjoint of K . Note that $\tilde{K} := K^*K$ is self-adjoint with spectrum $\sigma(\tilde{K}) \subseteq [0, \|K\|^2]$. If $0 \notin \sigma(\tilde{K})$, then from (3) we have $K^\dagger = \tilde{K}^{-1}K^*$. However, in general $0 \in \sigma(\tilde{K})$ but nevertheless a class of approximations to K^\dagger have been investigated in [3] and take the form $S_\lambda(\tilde{K})K^*$, where S_λ is a real-valued continuous function on $[0, \|K\|^2]$ which approximates t^{-1} in some sense. In [4] certain asymptotic rates of convergence for methods of this form are established.

To illustrate the type of result we obtain, consider the method of Tikhonov regularization. Here $S_\lambda(t) = (t + \lambda)^{-1}$, where $\lambda > 0$ is called the regularization parameter. For $b \in \mathcal{D}(K^\dagger)$ it is well known that

$$(4) \quad K^\dagger b = \lim_{\lambda \rightarrow 0^+} S_\lambda(\tilde{K})K^*b.$$

Moreover, we also have

$$(5) \quad \|K^\dagger b - S_\lambda(\tilde{K})K^*b\| = O(\lambda)$$

if $Qb \in R(K\tilde{K})$ (see for example [8]). Both (4) and (5) are valid for bounded operators K . In this note we show that the rate of convergence (5) is optimal for compact K in the sense that a rate of $o(\lambda)$ implies that $K^\dagger b = S_\lambda(\tilde{K})K^*b = 0$. In addition we establish the converse of (5). These results may be viewed as relatives of the ‘‘saturation’’ theorems in approximation theory (see for example [2]).

2. Results

We consider a family of real-valued continuous functions $\{S_\lambda: \lambda \in I\}$, indexed by a subset I of positive real numbers with $0 \in \bar{I}$. Each function S_λ is defined on $[0, \|K\|^2]$ and satisfies

$$(6a) \quad S_\lambda(t) \rightarrow t^{-1} \quad \text{as } \lambda \rightarrow 0^+ \text{ for } t \neq 0,$$

$$(6b) \quad t^\nu |1 - tS_\lambda(t)| \leq \omega(\lambda, \nu) \quad \text{for } \nu > 0,$$

where $\omega(\lambda, \nu) \rightarrow 0$ as $\lambda \rightarrow 0^+$ and

$$(6c) \quad \infty > \liminf_{\lambda \rightarrow 0^+} \frac{t^{\nu_0} |1 - tS_\lambda(t)|}{\omega(\lambda, \nu_0)} \geq \kappa_{\nu_0} > 0$$

for some $\nu_0 > 0$ and each $t > 0$. We will find it convenient to use the notation $x := K^\dagger b$, $x_\lambda = S_\lambda(\tilde{K})K^*b$ and $e_\lambda = x - x_\lambda$. The following result is proved in [4].

THEOREM 1. *If S_λ satisfies (6a) and (6b) and $Qb = K\tilde{K}^\nu w$ for some $\nu > 0$, then*

$$\|e_\lambda\| \leq \omega(\lambda, \nu)\|w\|.$$

Under the additional hypotheses that (6c) is valid and K is compact we shall establish a converse of Theorem 1. Assume then that $K: H_1 \rightarrow H_2$ is compact with singular system $\{u_n, v_n; \mu_n\}_{n=0}^\infty$. We recall Picard’s criterion (see for example [1]): $y \in \overline{R(K)}$ is in the range of K if and only if

$$\sum_{n=0}^\infty \mu_n^2 |(y, u_n)|^2 < \infty.$$

Let $\hat{K} := KK^*$, then since $\{\mu_n^{-2}\}$ are the eigenvalues of \hat{K} and $\{u_n\}$ are the corresponding eigenvectors, we may define $\hat{K}^\nu: H_2 \rightarrow H_2$, for $\nu > 0$, by:

$$\hat{K}^\nu y = \sum_{n=0}^\infty \mu_n^{-2\nu} (Qy, u_n) u_n$$

and \tilde{K}^ν has a similar representation in terms of the $\{v_n\}$. It follows that \hat{K}^ν and \tilde{K}^ν are self-adjoint and satisfy $K^*\hat{K}^\nu = \tilde{K}^\nu K^*$. A straightforward argument shows that the nullspaces of K^* and $K^*\hat{K}^\nu$ are identical and hence

$$\overline{R(\hat{K}^\nu K)} = N(K^*\hat{K}^\nu)^\perp = N(K^*)^\perp = \overline{R(K)}.$$

Since $\{u_n, v_n; \mu_n^{2\nu+1}\}$ is the singular system for $\hat{K}^\nu K$, we conclude from Picard’s criterion that $y \in \overline{R(K)}$ is in the range of $\hat{K}^\nu K$ if and only if

$$(7) \quad \sum_{n=0}^\infty \mu_n^{4\nu+2} |(y, u_n)|^2 < \infty.$$

We are now in a position to give a partial converse of Theorem 1.

THEOREM 2. *Suppose K is compact and S_λ satisfies (6) for some $\nu_0 > 0$. If $\|e_\lambda\| = O(\omega(\lambda, \nu_0))$ then $Qb \in R(\hat{K}^{\nu_0}K)$. Moreover, if $\|e_\lambda\| = o(\omega(\lambda, \nu_0))$ then $Qb = 0$ and hence $x = x_\lambda = 0$.*

PROOF. Since $\{u_n\}$ is an orthonormal basis for $\overline{R(K)}$ we have

$$Qb = \sum_{n=0}^\infty (Qb, u_n) u_n.$$

Since $K^*b = K^*Qb$ and $K^*u_n = \mu_n^{-1}v_n$, it follows that

$$x_\lambda = S_\lambda(\hat{K})K^*b = \sum_{n=0}^\infty \mu_n^{-1}(Qb, u_n)S_\lambda(\mu_n^{-2})v_n.$$

The minimal norm least squares solution, $x = K^\dagger b$, has the representation (see for example [7])

$$x = \sum_{n=0}^\infty \mu_n(Qb, u_n)v_n$$

and hence

$$e_\lambda = (I - S_\lambda(\hat{K})\hat{K})x = \sum_{n=0}^\infty (1 - \mu_n^{-2}S_\lambda(\mu_n^{-2}))\mu_n(Qb, u_n)v_n.$$

For $\nu > 0$, we then have

$$\|e_\lambda\|^2 = \sum_{n=0}^\infty [\mu_n^{-2\nu}(1 - \mu_n^{-2}S_\lambda(\mu_n^{-2}))]^2 \mu_n^{4\nu+2} |(Qb, u_n)|^2.$$

Therefore if $\|e_\lambda\| = O(\omega(\lambda, \nu))$, there is a constant C_ν , which is independent of λ , such that

$$(8) \quad \sum_{n=0}^\infty \left[\frac{\mu_n^{-2\nu}(1 - \mu_n^{-2}S_\lambda(\mu_n^{-2}))}{\omega(\lambda, \nu)} \right]^2 \mu_n^{4\nu+2} |(Qb, u_n)|^2 \leq C_\nu.$$

But by (6c) we have

$$\liminf_{\lambda \rightarrow 0^+} \frac{\mu_n^{-2\nu_0} |1 - \mu_n^{-2}S_\lambda(\mu_n^{-2})|}{\omega(\lambda, \nu_0)} \geq \kappa_{\nu_0} > 0$$

so that

$$\sum_{n=0}^\infty \mu_n^{4\nu_0+2} |(Qb, u_n)|^2 \leq \kappa_{\nu_0}^{-1} C_{\nu_0} < \infty$$

and hence $Qb \in R(\hat{K}^{\nu_0}K)$ by (7).

If $\|e_\lambda\| = o(\omega(\lambda, \nu_0))$, then (8) is valid with C_{ν_0} replaced by $C_{\nu_0}(\lambda)$ where $C_{\nu_0}(\lambda) \rightarrow 0$ as $\lambda \rightarrow 0^+$. Using (6c) as before we then find that $(Qb, u_n) = 0$ for all n so that $Qb = 0$ and hence $0 = x = x_\lambda$.

In [4] error estimates are also obtained under somewhat different appearing assumptions on Qb . In Proposition 4 of [4] it is proved that if $Qb = \hat{K}^\nu w$ for some $\nu \geq 1$, then

$$(9) \quad \|e_\lambda\|^2 \leq \omega(\lambda, \nu)\omega(\lambda, \nu - 1)\|w\|^2.$$

In fact, for compact K it is not hard to show that $R(\hat{K}^\nu) = R(\hat{K}^{\nu-1/2}K)$ for $\nu \geq \frac{1}{2}$. It follows that (9) may be replaced for $\nu \geq \frac{1}{2}$ by

$$\|e_\lambda\| \leq \omega(\lambda, \nu - \frac{1}{2})\|z\|$$

for $\hat{K}^\nu w = \hat{K}^{\nu-1/2}Kz$.

Our next result gives, for the vector $\tilde{K}e_\lambda$, a relationship between the asymptotic order of accuracy and the smoothness of the data. The significance of this result resides in the fact that

$$\tilde{K}e_\lambda = K^*b - \tilde{K}x_\lambda$$

is a computable quantity, at least in a theoretical sense. The proof of the first part of the following theorem follows directly from the proof of Proposition 3 in [4]. The proof of the second part is similar to that of Theorem 2.

THEOREM 3. *Suppose S_λ satisfies (6a) and (6b) and $Qb = \tilde{K}^\nu K w$ for some $\nu \geq 0$, then*

$$\|\tilde{K}e_\lambda\| \leq \omega(\lambda, \nu + 1)\|w\|.$$

Conversely, if S_λ satisfies (6), K is compact and $\|\tilde{K}e_\lambda\| = O(\omega(\lambda, \nu_0 + 1))$, then $Qb \in R(\hat{K}^{\nu_0}K)$. Also, if $\|\tilde{K}e_\lambda\| = o(\omega(\lambda, \nu_0 + 1))$, then $x = x_\lambda = 0$.

3. Examples and applications

In this section we consider some specific examples of families of functions $\{S_\lambda\}$ to which the preceding results apply.

As a first example we consider the choice $S_\lambda(t) = (t + \lambda)^{-1}$ which results in ordinary Tikhonov regularization (see for example [8]). One can easily verify that $\omega(\lambda, \nu) = \lambda^\nu$ for $0 < \nu \leq 1$ satisfies (6a) and (6b) and for $\nu_0 = 1$ satisfies (6c) as well. For a compact operator Theorem 2 says that, except for the trivial case $Qb = 0$, ordinary Tikhonov regularization cannot converge at a rate faster than $O(\lambda)$. By Theorem 1 this rate is attained if $Qb \in R(\hat{K}K)$. Moreover the computable quantity $\|\tilde{K}e_\lambda\|$ is $O(\lambda^2)$ if and only if $Qb \in R(\hat{K}K)$.

For ordinary Tikhonov regularization it should be noted that the rate of convergence is $O(\lambda^{1/2})$ under the weaker hypothesis $Qb \in R(\hat{K})$ (see for example [7], [8]). We now show by example that Theorem 2 is not valid for this nonoptimal convergence rate. Suppose that K is a compact operator with singular system $\{u_n, v_n; n^{1/2}\}$ and let

$$b = Qb = \sum_{n=1}^{\infty} n^{-3/2}u_n.$$

Then $Qb \in R(\hat{K})$ if and only if the following sum is finite:

$$\sum_{n=1}^{\infty} (n^{1/2})^4 |(Qb, u_n)|^2 = \sum_{n=1}^{\infty} \frac{1}{n},$$

and hence $Qb \notin R(\hat{K})$. However,

$$e_{\lambda} = \sum_{n=1}^{\infty} \frac{\lambda}{1 + n\lambda} v_n$$

so that

$$\|e_{\lambda}\|^2 = \sum_{n=1}^{\infty} \left(\frac{\lambda}{1 + n\lambda}\right)^2 = \lambda \sum_{n=1}^{\infty} \frac{\lambda}{(1 + n\lambda)} < \lambda,$$

by the integral test and hence $\|e_{\lambda}\| = O(\lambda^{1/2})$.

In order to obtain approximations with the rate $\omega(\lambda, \nu) = \lambda^{\nu}$ for $\nu > 1$, one may use extrapolated regularization [6] or iterated regularization [9]. Saturation theorems for extrapolated regularization were given in [5].

Extrapolated Tikhonov regularization is defined as follows. Suppose $\lambda_i = \gamma_i \lambda$ where $\lambda > 0$ and γ_i are distinct numbers and let $a_i^{(k)}$, $i = 0, 1, \dots, k$, satisfy

$$\sum_{i=0}^k a_i^{(k)} = 1, \quad \sum_{i=0}^k a_i^{(k)} \lambda_i^j = 0, \quad j = 1, 2, \dots, k.$$

The k th extrapolated Tikhonov approximation is given by

$$x_{\lambda}^{(k)} = S_{\lambda}^{(k)}(\tilde{K})K^*b$$

where $S_{\lambda}^{(k)}(t) = \sum_{i=0}^k a_i^{(k)}(t + \lambda_i)^{-1}$. The function $S_{\lambda}^{(k)}$ satisfies (see [6])

$$1 - tS_{\lambda}^{(k)}(t) = \prod_{i=0}^k \lambda_i(t + \lambda_i)^{-1},$$

from which it follows immediately that

$$(t + \lambda_k)S_{\lambda}^{(k)}(t) = \lambda_k S_{\lambda}^{(k-1)}(t) + 1.$$

Therefore, for $k \geq 0$

$$(10) \quad (\tilde{K} + \lambda_k I)x_{\lambda}^{(k)} = \lambda_k x_{\lambda}^{(k-1)} + K^*b,$$

where, by convention, $x_{\lambda}^{(-1)} = 0$. The family $\{S_{\lambda}^{(k)}\}$ satisfies (6a) and (6b) with $\omega(\lambda, \nu) = \lambda^{\nu}$ if $k \geq \nu - 1$. It can be readily verified that (6c) is satisfied for $\nu_0 = k + 1$. Note that $S_{\lambda}^{(0)}$ corresponds to ordinary Tikhonov regularization.

Equation (10) shows that the extrapolants may be determined in an iterative manner and also suggests the iterative method (see [9])

$$(11) \quad (\tilde{K} + \lambda I)v_{\lambda}^{(k)} = \lambda v_{\lambda}^{(k-1)} + K^*b, \quad k = 0, 1, \dots,$$

where, again by convention, $v_\lambda^{(-1)} = 0$. This corresponds to the function

$$\mathfrak{S}_\lambda^{(k)}(t) = \sum_{i=1}^{k+1} \lambda^{i-1} (t + \lambda)^{-i},$$

that is, $v_\lambda^{(k)} = \mathfrak{S}_\lambda^{(k)}(\tilde{K})K^*b$. The function $\mathfrak{S}_\lambda^{(k)}$ satisfies

$$1 - t\mathfrak{S}_\lambda^{(k)}(t) = [\lambda / (t + \lambda)]^{k+1}$$

from which it follows that $\mathfrak{S}_\lambda^{(k)}$ satisfies (6a) and (6b) with $\omega(\lambda, \nu) = \lambda^\nu$ for $k \geq \nu - 1$. Also (6c) is satisfied for $\nu_0 = k + 1$. Again note that $\mathfrak{S}_\lambda^{(0)}$ corresponds to ordinary Tikhonov regularization.

We observe that $e_\lambda^{(k)} := K^\dagger b - v_\lambda^{(k)}$ satisfies

$$(\tilde{K} + \lambda I)e_\lambda^{(k)} = \lambda e_\lambda^{(k-1)}$$

and hence

$$\|\tilde{K}e_\lambda^{(k)}\| = \lambda \|v_\lambda^{(k)} - v_\lambda^{(k-1)}\|.$$

Therefore, by Theorem 3, $Qb \in R(\hat{K}^{k+1}K)$ if and only if the following asymptotic rate is valid for the computable quantity $\|v_\lambda^{(k)} - v_\lambda^{(k+1)}\|$:

$$\|v_\lambda^{(k)} - v_\lambda^{(k-1)}\| = O(\omega(\lambda, k + 2))/\lambda = O(\lambda^{k+1}).$$

In these examples of Tikhonov regularization and its variants our results say that the attainment of the optimal convergence rate is equivalent to a ‘‘smoothness’’ condition on the data b (or, equivalently, the vector $K^\dagger b$). Let us examine this more closely for the prototypical example wherein the compact operator is given by an integral operator with square integrable kernel.

To be precise we must introduce the appropriate Hilbert spaces. For simplicity we consider functions which are defined on the interval $[0, 1]$. For $r \geq 1$ let H^r denote the Sobolev space consisting of all functions ϕ whose j th derivatives, $\phi^{(j)}$, are absolutely continuous for $0 \leq j \leq r - 1$ and also satisfy $\phi^{(r)} \in L_2[0, 1]$. The inner product on H^r is given by

$$\langle \phi, \psi \rangle = \sum_{j=0}^r (\phi^{(j)}, \psi^{(j)}),$$

where (\cdot, \cdot) denotes the usual inner product on $L_2[0, 1]$.

Let $k(s, t)$ be a given function on $[0, 1] \times [0, 1]$ such that $k(s, \cdot) \in H^r$, almost everywhere in s . We consider the first kind integral equation

$$(12) \quad Kf(s) := \int_0^1 k(s, t)f(t) dt = g(s),$$

where $g \in L_2[0, 1]$ and $k \in L_2([0, 1] \times [0, 1])$ are given. Thus $K: L_2[0, 1] \rightarrow L_2[0, 1]$ is compact and has adjoint, K^* , given by

$$(13) \quad K^*w(t) = \int_0^1 k(s, t)w(s) ds.$$

Tikhonov regularization of order 0 applied to (12) consists in minimizing over $L_2[0, 1]$ the functional

$$(14) \quad F_\lambda(\phi) = \|K_\phi - g\|_{L_2[0, 1]}^2 + \lambda \|\phi\|_{L_2[0, 1]}^2.$$

If we denote the minimizer of (14) by x_λ , then by Theorem 2, $\|K^\dagger g - x_\lambda\| = O(\lambda)$ if and only if $Qg \in R(KK)$. Since $K^\dagger g \in N(K)^\perp$, it follows from (2) that $K^\dagger g \in R(K^*K)$. Hence by (13) and our assumption on the kernel we must have $K^\dagger g \in H^r$. In particular we see that the rate of convergence $O(\lambda)$ implies the existence of at least one smooth least squares solution.

We note that if $k(s, 1) = 0$ or $k(s, 0) = 0$ for almost all s , then the optimal rate of convergence imposes the boundary conditions $x(1) = 0$ or $x(0) = 0$, respectively, on $x = K^\dagger g$. A specific example of this type is given by the problem of numerical differentiation. Define the kernel $\gamma(\cdot, \cdot)$ by

$$\gamma(s, t) = \begin{cases} 1, & 0 \leq t \leq s, \\ 0, & s < t \leq 1, \end{cases}$$

and denote the corresponding integral operator by Γ . For a given $g \in L_2[0, 1]$ satisfying $g(0) = 0$, it follows that $\Gamma f = g$ if and only if $f = g' \in L_2[0, 1]$. Since the nullspace of Γ is trivial it follows that $\Gamma^\dagger g = g'$ for $g \in H^1$ satisfying $g(0) = 0$. Let

$${}_0H^r = \{g \in H^r: g(0) = 0\} \quad \text{and} \quad H_0^r = \{g \in H^r: g(1) = 0\}.$$

Then $R(\Gamma) = \mathcal{D}(\Gamma^\dagger) = {}_0H^1$ and $R(\Gamma^*) = H_0^1$.

Now suppose that $K: L_2[0, 1] \rightarrow L_2[0, 1]$ is a given integral operator whose kernel satisfies $k(s, \cdot) \in H^r$ and $k(s, 1) = 0$, for almost all s . Suppose also that zero order Tikhonov regularization applied to $Kf = g$ results in the optimal rate of convergence. Then $K^\dagger g \in H^r$ and use of higher order regularization is suggested (and justified). For simplicity we consider only regularization of order 1. Let \bar{K} denote the restriction of K to the Hilbert space H_0^1 with norm

$$\|w\|_1 = \|w'\|_{L_2[0, 1]}.$$

It is not difficult to see that \bar{K} is compact so that our results apply. Tikhonov regularization of order 1 applied to $\bar{K}f = g$ consists in minimizing

$$G_\lambda(\phi) := \|\bar{K}\phi - g\|_{L_2[0, 1]}^2 + \lambda \|\phi\|_1^2$$

over H_0^1 . If we denote the minimizer of G_λ by x_λ , then $\|\bar{K}^\dagger g - x_\lambda\|_1 = O(\lambda)$ implies that $\bar{K}^\dagger g \in R(\bar{K}^*\bar{K})$. In order to interpret this condition, note that for $\phi \in L_2[0, 1]$ and $\psi \in H_0^1$, we have

$$\begin{aligned} ((\Gamma^*\Gamma K^*\phi)', \psi') &= -(\Gamma K^*\phi, \psi') = -(K^*\phi, \Gamma^*\psi') \\ &= (K^*\phi, \psi) = (\phi, K\psi) = (\phi, \bar{K}\psi) \end{aligned}$$

and hence $\bar{K}^* = \Gamma^* \Gamma K^*$. It then follows that the condition $K^\dagger g = x \in R(\bar{K}^* \bar{K})$, requires the smoothness condition $x \in H^{r+2}$ as well as the boundary condition $x'(0) = x(1) = 0$.

Acknowledgement

The first author would like to thank Dr. C. T. H. Baker for making his visit to the University of Manchester possible.

References

- [1] C. T. H. Baker, *The numerical treatment of integral equations*, (Clarendon Press, Oxford, 1977).
- [2] R. A. DeVore, *The approximation of continuous functions by positive linear operators*, (Lecture Notes in Mathematics Vol. 293, Springer-Verlag, Berlin, 1972).
- [3] C. W. Groetsch, *Generalized inverses of linear operators: representation and approximation*, (Dekker, New York, 1977).
- [4] C. W. Groetsch, 'On a class of regularization methods', *Boll. Un. Mat. Ital.* (5) 17 (1980), 1411–1419.
- [5] C. W. Groetsch, *On the convergence of the method of regularization for equations of the first kind*, (Numerical Analysis Report No. 52, August, 1980, University of Manchester).
- [6] C. W. Groetsch and J. T. King, 'Extrapolation and the method of regularization for generalized inverses', *J. Approximation Theory* 25 (1979), 233–247.
- [7] J. W. Hilgers, *Noniterative methods for solving operator equations of the first kind*, (Tech. Summ. Rep. 1413, Mathematics Research Center, Univ. of Wisconsin-Madison, 1973).
- [8] V. V. Ivanov, *The theory of approximate methods and their application to the numerical solution of singular integral equations*, (Noordhoff, Leyden, 1976).
- [9] J. T. King and D. Chillingworth, 'Approximation of generalized inverses by iterated regularization', *Numer. Func. Anal. Optim.* 1 (1979), 499–513.

Department of Mathematical Sciences
 University of Cincinnati
 Cincinnati, Ohio 45221
 U.S.A.