Note di Matematica Vol. 16 - n. 1, 1-8 (1996)

A PARAMETER CHOICE FOR MINIMIZING THE ERROR BOUND OF TIKHONOV REGULARIZATION METHOD

HASAN M. YMERI

Abstract. *The paper considers posteriori strategies for choosing a parameter in a simplified in a simplified version of the Tikhonov regularization. An arror bound is minimized to select the parameter that regulates the* smoothness *of approximation in ill posed problems. The rate obtained in doing so is optimal for the first order and for the iterated case.*

1. INTRODUCTION

We are concerned with the inverse problems

$$Aw = g \tag{1}$$

A is a compact non-negative selfadjoint operator defined on a Hilbert space, g is the known data, and w is the minimal norm least squares solution that we want to find. This problem arises in physical situations as an integral equation of the first kind generated by a symmetric kernel on an L^2 space.

It is wellknown that equation (1) is in general an ill posed problem, except when the range of A is a closed set. By this we mean that even if the solution w exists, the mapping $g \to w$ is not continuous. Because ill-posedness leads to unstable numerical schemes, solving equation (1) requires a procedure called regularization, which is designed to make the solution continuously dependent upon the data g. The most widely known regularization is Tikhonov's method. It uses x_α, the minimizer of the functional

$$F_\alpha(x) = ||Ax - g||^2 + \alpha||x||^2, \alpha > 0,$$

or equivalently the solution to $(\alpha I + A^*A)x_\alpha = A^*g$, as an approximation to w. Here, $< \cdot, \cdot >$ is the inner product and $|| \cdot ||$ is the norm of the underlying Hilbert space. In this setting α controls the smoothness of approximations in the sense that it permits the trade off between fidelity (small α) and minimizing the norm of x_α (large α).

For a comprehensive study of inverse problems see e.g. [1, 6, 10].

Simplified regularization consists in minimizing the functional

$$G_\alpha(x) = ||A^{1/2}(x - w)||^2 + \alpha||x||^2, \tag{2}$$

instead of the Tikhonov functional; or equivalently in using w_α, the solution to $(\alpha I + A)w_\alpha = g$, as an approximation to w. This regularization is applicable only when the data is in the range of $A, g \in R(A)$. If this is not the case, there are examples of w_α not being convergent, [11]. Therefore, in this note we assume that the data is in $R(A)$.

As equation (1) is intended to model physical phenomena, we have to consider the inexact data case, because the data comes from experimental measurements that are frequently inexact. Let g^δ denote the inexact data and assume that the *noise level* δ, is much smaller than the data; i.e. $\|g^\delta - g\| \leq \delta << \|g^\delta\|$. The core of any regularization processes is the choice of the positive parameter α, depending on δ, in such way that approximations converge to the solution w. To meet this end there are essentially two ways: either using priori strategies that choose α beforehand according to some information of its asymptotic behaviour, or using posteriori strategies that decide the value of α as calculation proceed. Of course, the optimal parameter choice is the one that minimizes the difference $\|w - x_{\alpha\delta}\|$, but since w is not available we have to leave aside this optimal choice for some other tactic. Among the tactics we have:

Considering the residual $Ax_{\alpha\delta} - g^\delta$, or the residual $A^*Ax_{\alpha\delta} - A^*g^\delta$, to define discrepancy principles. Some discrepancy principles, that apply to Tikhonov's method, lead to divergent sequences in simplified regularization, for example Arcangeli's method [8, 11].

Minimizing an error bound rather than $\|w - x_{\alpha\delta}\|$, Engl [3] followed these tactics for approximations of the form $U_\alpha(A^*A)A^*g$ where $U_\alpha(\cdot)$ is a regularizing function.

Minimizing the differences between the approximation on hand and a faster convergent sequence (see Remark 2).

And using only some part of the data to select the parameter that best predicts the rest of data. This is known as the cross validation principle [10].

For parameter choices in the Tikhonov case see for example [4, 7, 9, 10]. In practice, a common procedure consists of performing the regularization for several values of α, e.g. $2^{-1}, 2^{-2}, \ldots, 2^{-2}$ and then selecting the value that satisfies (or almost satisfies) the criterion in mind. See Gorbonova and Morozov for some of these algorithms [5]. Numerical experiments show that for different examples different tactics perform better then the others, i.e. no one parameter choice is the best.

In this article we minimize a bound for the error $\|w - w_\alpha\|$. The value of α at this point is our a posteriori strategy. Comparisons of the numerical behaviour between the above strategies calles for a separate study.

2. THE MAIN RESULTS

The following two lemmas proved in [11] are the basis for our parameter choices.

Lemma 1. *If $g \in \mathbf{R}(A^2)$ then regularized solutions converge to w as $\alpha \to 0$, with the order* $\|w - w_\alpha\| = 0(\alpha)$.

Lemma 2. *For inexact data the error with respect to the idealized approximation w_α is bounded by* $\|w_\alpha - w_{\alpha\delta}\| \leq \delta / \alpha$.

By using the triangle inequality and by making the bounds in Lemma 1 and 2 of the same asymptotic order, i.e. $\alpha \approx \delta^{1/2}$, we obtain the order of convergence $0(\delta^{1/2})$, which cannot be improved as we will see next.

Theorem 1. *Suppose that A is of infinite rank and that $\|w - w_{\alpha\delta}\| = 0(\delta^{1/2})$ for every g^δ such that $\|g - g^\delta| \leq \delta$, then $w = 0$.*

Proof. Let $\{u_i, \lambda_1 > \lambda_2 > \dots\}$ be an orthonormal system for A. Suppose that $w \neq 0$. If we take $w = u_1$, $g = \lambda_1 u_1$ and $g^\delta = g + \delta u_n$, with $\delta = \lambda_n^2 \to 0$ as $n \to \infty$, then we will see a contradiction.

Since $(\alpha I + A)(w - w_{\alpha\delta}) = \alpha w + Aw - g^\delta = \alpha w + g - g^\delta$, $w \neq 0$ and A is a bounded operator the equalities

$$\|\alpha w\| = \|(\alpha I + A)(w - w_{\alpha\delta}) - (g - g^\delta)\| = 0(\|w - w_{\alpha\delta}\|) + 0(\delta) = 0(\delta^{1/2})$$

imply $\alpha = 0(\delta^{1/2})$.

From their definitions we have that $w - w_{\alpha\delta} = w - w_\alpha - (\alpha I + A)^{-1}\delta u_n$ and

$$\|w - w_{\alpha\delta}\|^2 = \|w - w_\alpha\|^2 - 2\delta(\alpha + \lambda_n)^{-1}\delta < w - w_\alpha, u_n > + \delta^2(\alpha + \lambda_n)^{-2}.$$

$$\|w - w_{\alpha\delta}\|^2 / \delta \geq -2(\alpha + \lambda_n)^{-1} < w - w_\alpha, u_n > + (\alpha\delta^{-1/2} + \lambda_n \delta^{-1/2})^{-2}.$$

Since $\|w - w_{\alpha\delta}\| = 0(\delta^{1/2})$, $\alpha = 0(\delta^{1/2})$ and $\delta = \lambda_n^2$ in the limit we have

$$0 \geq \limsup_{n \to 0} \frac{2}{\alpha + \delta^{1/2}} < w - w_\alpha, u_n > + 1.$$

By taking $g^\delta = g$ the assumption that $\|w - w_{\alpha\delta}\| = 0(\delta^{1/2})$ implies $\|w - w_\alpha\| = 0\,(\delta^{1/2})$. Therefore the right hand side on the above inequality is $0(\delta^{1/2})$ divided by $0(\delta^{1/2})$, and hence $0 \geq 1$, a contradiction that completes the proof.

Q.E.D.

In the following we assume that A is not of finite rank and that $w \neq 0$; therefore, the search is of strategies that provide rate of convergence $0(\delta^{1/2})$. We start by considering the error estimate

$$\frac{1}{2}\|w - w_{\alpha\delta}\|^2 \leq \|w - w_\alpha\|^2 + \frac{\delta^2}{\alpha^2} \tag{3}$$

that comes from Lemma 2 and the parallelogram law. By matching the two terms of the right hand side, we define our first criterion

$$h_1(\alpha) := \alpha^4 < (\alpha I + A)^{-2} w, w > = \delta^2. \tag{4}$$

Some properties of the continuous function h_1, that make (4) a valid strategy are: First

$$\lim_{\alpha \to 0} h_1(\alpha) = \lim_{\alpha \to 0} \alpha^2 < \alpha^2(\alpha I + A)^{-2} w, w > = 0$$

and $\lim_{\alpha \to \infty} h_1(\alpha) = \infty$, this limit because A is bounded. Second, the derivative $h_1'(\alpha) = 4\alpha^3 < (\alpha I + A)^{-3}\, w, w >$ which is positive for $\alpha > 0$; therefore, h_1 is strictly increasing and given δ there exists a unique $\alpha = \alpha(\delta)$ that satisfies (4).

And third, we have the following lemma

Lemma 3. *The parameter α selected by (4) converges to 0 as $\delta \to 0$.*

Proof. On the contrary assume that either $\alpha \to \infty$, then by (4) $\lim_{\alpha\to\infty} \alpha^2 < \alpha^2\ (\alpha I + A)^2 w, w> = 0$; or if there exists a subsequence of α's that converges to $c>0$, then in the limit we have $c^4 <(cI+A)^{-2}w, w> = 0$, implying $w = 0$, which contradicts our hypothesis.

Q.E.D.

In regard to convergence we have

Theorem 2. *If* $w \in \mathbf{R}(A)$ *and* $\alpha = \alpha(\delta)$ *is selected according to (4) then* $w_{\alpha\delta}$ *attains the order of convergence* $\|w_{\alpha\delta} - w\| = 0(\delta^{1/2})$.

Proof. As $w \in \mathbf{R}(A)$, say $w = Au$, the approximation $(\alpha I + A)^{-1}w$ converges to u as $\alpha \to 0$. Let $\delta \to 0$, then by Lemma 3 $\alpha \to 0$. Because of equation (4)

$$\lim_{\alpha,\delta\to 0} \frac{\delta^2}{\alpha^4} = \lim_{\alpha\to 0} \|(\alpha I + A)^{-1}w\|^2 = \|u\|^2.$$

This means that α and $\delta^{1/2}$ have the same asymptotic order. Finally, by Lemmas 1 and 2

$$\|w - w_{\alpha\delta}\| \leq 0(\alpha) + \frac{\delta}{\alpha} = 0(\delta^{1/2}).$$

Q.E.D.

The first order necessary condition for minimizing the error bound (3) is $2<w - w_\alpha, -\frac{d}{d\alpha}(w_\alpha) > -2\delta^2/\alpha^3 = 0$, which provides the criterion

$$h_2(\alpha) = \alpha^4 <(\alpha I + A)^{-2}w_\alpha, w> = \delta^2. \tag{5}$$

The above equation is like (4) with one of the w's replaced by w_α. This suggests that we replace both w's to obtain

$$h_3(\alpha) = \alpha^4 <(\alpha I + A)^{-2}w_\alpha, w_\alpha < = \delta^2. \tag{6}$$

Furthermore, since w_α is not available for computations we may use $w_{\alpha\delta}$ to obtain the compulable criterion

$$h_4(\alpha) = \alpha^4 <(\alpha I + A)^{-2}w_{\alpha\delta}, w_{\alpha\delta} > = p\delta^2, \tag{7}$$

where the positive constant p is arbitrary, but no larger than a noise ratio $\|g^\delta\|^2/\delta^2$, as we will see next.

The following properties for continuous functions $h_i, i = 2,3,4$, can be proved in the same fashion as for h_1: first, $\lim_{\alpha\to 0} h_i(\alpha) = 0$; second, the derivative $h_i'(\alpha)>0$ for $\alpha>0$, so we infer h_i is strictly increasing and each criterion (5), (6) and (7) determines a unique $\alpha = \alpha(\delta)$; third, $\alpha \to 0$ as $\delta \to 0$; and fourth, $\lim_{\alpha\to\infty} h_2(\alpha) = \infty$, $\lim_{\alpha\to\infty} h_3(\alpha) = \|g\|^2$ and $\lim_{\alpha\to\infty} h_4(\alpha) = \|g^\delta\|^2$. This limit implies that p in (7) should be smaller than the ratio $\|g^\delta\|^2/\delta^2$.

The proof of Theorem 2 applies to criteria (5) and (6) but is does not apply to (7). To show that (7) can also provide an α which produces the order $0(\delta^{1/2})$, it suffices to show that the quotient δ / α^2 is bounded from above and from below.

Lemma 4. *Assume that $w = Au$ for some vector u. If $\alpha = \alpha(\delta)$ is selected according to criterion (7) with $q < p$, then as $\alpha, \delta \to 0$ we have*

$$\frac{\|u\|^2}{p+1} \leq \frac{\delta^2}{\alpha^4} \leq \frac{\|u\|^2}{p-1}.$$

Proof. From the definition of h_3, h_4, w_α and $w_{\alpha\delta}$ it follows that

$$|h_3(\alpha) - h_4(\alpha)| \leq \alpha^4 \|(\alpha I + A)^{-2}(g - g^\delta)\|^2 \leq \|g - g^\delta\|^2 \leq \delta^2.$$

Since $\alpha \to 0, p > 1$, from (6) we have $(p-1)\delta^2 \leq h_3(\alpha) \leq (p+1)\delta^2$. This implies the existence of a function $q(\delta)$ with values in the interval $[p-1, p+1]$, such that $h_3(\alpha) = q(\delta)\delta^2$ and $h_4(\alpha) = p\delta^2 = ph_3(\alpha) / q(\delta)$. This last equality together with the limit $\lim_{\alpha\to 0} h_3(\alpha) / \alpha^4 = \|u\|^2$ implies

$$\frac{p\|u\|^2}{p+1} \leq \liminf_{\alpha\to 0} \frac{h_4(\alpha)}{\alpha^4} = \liminf_{\alpha,\delta\to 0} \frac{\delta^2}{\alpha^4}$$

$$\leq \limsup_{\alpha,\delta\to 0} \frac{\delta^2}{\alpha^4} \limsup_{\alpha\to 0} \frac{h_4(\alpha)}{\alpha^4} \leq \frac{p\|u\|^2}{p-1}$$

concluding the present proof.

Q.E.D.

As α and $\delta^{1/2}$ have the same asymptotic order we are ready for the statement on the best rate of convergence.

Theorem 3. *Under the hypothesis of Lemma 4 criterion (7) provides an α that gives $\|w - w_{\alpha\delta}\| = 0(\delta^{1/2})$.*

Proof. Since $\alpha = \alpha(\delta) \to 0$ by Lemmas 1,2 and 4 it follows that

$$\|w - w_{\alpha\delta}\| \leq 0(\alpha) + \frac{\delta}{\alpha} = 0(\delta^{1/2}).$$

Q.E.D.

Remark 1. *To compute α from (7) we use the equalities $h_4(\alpha) = \|\alpha^2(\alpha I + A)^{-1} w_\alpha\|^2 = \|\alpha^2 \frac{d}{d\alpha} w_{\alpha\beta}\|^2$. Then two possible ways are using finite differences of the already known $w_{\alpha\delta}$'s or storing the $L-U$ decomposition to iterate twice the recurrence formula $(\alpha I + A)x^{i+1} = x^i, x^0 = g^\delta$. In the second case finding the α is as expensive as computing a second order regularization. Our attention turns now to the parameter choice for a high order regularization.*

3. ITERATED SIMPLIFIED REGULARIZATION

The importance of the iterative Tikhonov methods is that they produce rates of convergence higher than $0(\alpha)$ [2,4], although at the expense of more computations and the need of selecting a new parameter n. In simplified regularization we consider the iterative scheme, for $i = 2, 3, \dots$,

$$w_\alpha^i = \alpha(\alpha I + A)^{-1} w_\alpha^{i-1} + w_\alpha^1 : \textit{with } w_\alpha^1 = (\alpha I + A)^{-1} g. \tag{8}$$

As in the previous case, we first study some properties of the approximation and then define criteria for choosing α.

Lemma 5. *If* $w \in R(A^n)$, *then* $\|w_\alpha^n - w\| = 0(\alpha^n)$.

Proof. By (8) we have

$$\begin{aligned} w_\alpha^n &= (\alpha^{n-1}(\alpha I + A)^{-(n-1)} + \dots + \alpha(\alpha I + A)^{-1} + I) w_\alpha^1 \\ &= (\alpha^n(\alpha I + A)^{-n} - I)(\alpha(\alpha I + A)^{-1} - I)^{-1} W_\alpha^1 \\ &= -(\alpha^n(\alpha I + A)^{-n} - I) w. \end{aligned}$$

Because $w \in R(A^n)$ then $(\alpha I + A)^{-n} w$ converges as $\alpha \to 0$, and

$$\|w - w_\alpha^n\| = \alpha^n \|(\alpha I + A)^{-n} w\| = 0(\alpha^n).$$

Let $w_{\alpha\delta}^n$ denote the iterated approximation when g^δ is used instead of g.

Lemma 6. *For inexact data the error with respect to the idealized value* w_α^n *is bounded by* $\|w_{\alpha\delta}^n - w_\alpha^n\| \leq n\delta / \alpha$.

Proof. Using the proof of Lemma 5, and Lemma 2 we see that

$$\begin{aligned} \|w_{\alpha\delta}^n - w_\alpha^n\| &= \|\alpha^{n-1}(\alpha I + A)^{-(n-1)}(w_{\alpha\delta}^1 - w_\alpha^1) + \\ &\dots + (w_{\alpha\delta}^1 - w_\alpha^1)\| \leq \frac{\delta}{\alpha} + \dots + \frac{\delta}{\alpha} = n\frac{\delta}{\alpha}. \end{aligned}$$

Q.E.D.

Our goal is now to select α so that the n-iterated approximations have rate of convergence $0(\delta^{n/(n+1)})$. Lemma 5 and 6 and the parallelogram law give the error bound

$$\frac{1}{2}\|w - w_{\alpha\delta}^n\|^2 \leq \|w - w_\alpha^n Vert^2 + \left(n\frac{\delta}{\alpha}\right)^2 .$$

By matching the two terms on the right hand side of the above inequality, we obtain the criterion

$$k_1(\alpha) = \alpha^{2n+2} <(\alpha I + A)^{-2n} w, w> = n^2\delta^2. \tag{9}$$

By differentiating the above error bound with respect to α, the first order necessary conditions for minimization provide the criterion

$$k_2(\alpha) = \alpha^{2n+2} <(\alpha I + A)^{-2n} w_\alpha, w> = n^2\delta^2. \tag{10}$$

On replacing w and w_α by $w_{\alpha\delta}$ in (10) we obtain

$$k_3(\alpha) = \alpha^{n+2} <(\alpha I + A)^{-2n} w_\alpha, w_\alpha> = n^2\delta^2. \tag{11}$$

and

$$k_4(\alpha) = \alpha^{n+2} <(\alpha I + A)^{-2n} w_{\alpha\delta}, w_{\alpha\delta}> = n^2 p\delta^2. \tag{12}$$

where p is a constant such that $1 < n^2 p < \|g^\delta\|^2 / \delta^2$. For $i = 1, 2, 3, 4$ the functions k_i and h_i have similar properties: the equivalent of Lemma 4 is the boundedness, from above and from below, of the equation δ^2 / α^{2n+2}. This fact is used in the next theorem which says that for large n the order of convergence becomes arbitrarily close to $0(\delta)$, the best possible in linear problems.

Theorem 3. *If $w \in R(A^n)$ and $\alpha(\delta)$ is selected in accordance with (12) then the iterated approximations $w^n_{\alpha\delta}$ converge with order $\|w - w^n_{\alpha\delta}\| = 0(\delta^{n/(n+1)})$.*

Proof. From Lemma 5 and the fact that δ and α^{n+1} are of the same asymptotic order, it follows that

$$\begin{aligned} \|w - w^n_{\alpha\delta}\| &= \|w - w^n_\alpha\| + \|e^n_\alpha - w^n_{\alpha\delta}\| \\ &\leq 0(\alpha^n) + n\frac{\delta}{\alpha} = 0(\delta^{n/(n+1)}). \end{aligned}$$

Q.E.D.

Remark 2. *The strategy of selecting the α that moves $w_{\alpha\delta}$ toward w^2_α, a more expensive computational approximation, can be realized by minimizing the error bound*

$$\frac{1}{2}\|w^2_\alpha - w_{\alpha\delta}\| \leq \|\alpha(\alpha I + A)^{-2} g^\delta\|^2 + (\alpha / \delta)^2.$$

The first order condition for minimizing this bound defines the criterion

$$h_5(\alpha) = 2\alpha^6 <(\alpha I + A)^{-2} w_{\alpha\delta}, w + \alpha\delta> = p\delta^2,$$

where w_α has been replaced by $w_{\alpha\delta}$. Note that h_5 has similar properties to those of h_1, and that the assumption $w \in R(A)$ implies boundedness from above and from below of the quotient δ^2 / α^6. This last statement implies

$$\|w - w_{\alpha\delta}\| \leq 0(\alpha) + \frac{\delta}{\alpha} = 0(\delta^{1/3}).$$

Consequently, the order obtained in this remark is not as good that one by criteria (7) and (12).

REFERENCES

[1] A.K. AMIROV, *On the question of solvability of inverse problems*, Soviet. Math. Dokl. 34 (1987), 258-259.

[2] I.F. DOROFEEV, *The accuracy of some regularizing algorithms*, USSR Comput. Math. and Math. Phys. 16 (1976), 224-229.

[3] H.W. ENGL, H. GFRERER, *A posteriori parameter choice for general regularization methods for solving linear ill-posed problems*, Appl. Numer. Math. 4 (1988), 395-417.

[4] H.W. ENGL, *On the choice of the regularization parameter for iterated Tikhonov regularization of ill-posed problems*, J. Approx. Theory 49 (1987), 55-63.

[5] V.I. GORDONOVA, V.A. MOROZOV, *Numerical algorithms for choosing a parameter in the regularization method*, USSR Comput. Math. and Math. Phys. 13 (1973), 1-9.

[6] C.W. GROETSCH, *Theory of Tikhonov Regularization for Fredholm Equations of the First Kind*, Boston, 1984.

[7] C.W. GROETSCH, *The parameter choice problem in linear regularization, a mathematical introduction*, preprint.

[8] C.W. GROETSCH, J. GUACANEME, *Aracangelis's method fpr Fredholm equations of the first kind*, Proc. Amer. Math. Soc. 99 (1987), 256-261.

[9] A.N. TIKHONOV, V.A. ARSENIN, *Solutions of Ill-posed Problems*, New York, 1977.

[10] G. WABA, *Constrained regularization for ill-posed linear operator equations with applications in meteorology and medicine*, In: S.S. Gupta and J.O. Berger (eds.), Statistical Decision Theory and Related Topics, III, Vol. 2, New York, 1982.

[11] H.M. YMERI, *On the finite dimensional simplified regularization*, Pun. J. Math. 27 (1993), 23-36.

Received March 1, 1996 and in revised form September 23, 1996
Higher Technical School
University of Prishtina
38220 Mitrovicë
Kosovë
EX YUGOSLAVIA