COMPUTATIONAL MATHEMATICS
BANACH CENTER PUBLICATIONS, VOLUME 13
PWN—POLISH SCIENTIFIC PUBLISHERS
WARSAW 1984

SOME REMARKS ON BEST REGULARIZATION

V. FRIEDRICH

Technische Hochschule, Sektion Mathematik,
Reichenhainerstr. 41, DDR-9022, Karl-Marx-Stadt, DDR

For the solution of ill-posed problems, arising in many fields of application, several various regularization techniques have been developed. Ill-posed operator equations

$$Ax = z$$

are characterized by unbounded inverse operators A^{-1}, if we use two norms $\| \|_x$ and $\| \|_z$, which may be interpreted in a natural physical sense.

If we want to determine an element $\hat{x}$ by measuring $A\hat{x}$, in problems which are faced in reality there will, in general, occur a measurement error y up to a level δ ($\|y\|_z \leqslant \delta$), so that $z = A\hat{x} + y$ is the result of the measuring process.

Consequently, we have to accept the set

$$X_\delta = \{x \mid \|Ax - z\| \leqslant \delta\}$$

as a set of formal solutions, which are compatible with result z relative to the given level δ. Different regularization techniques choose certain elements of this set, by means of some other a priori information about $\hat{x}$, in different ways. In some applications it is difficult to effect the requirement $\delta \to 0$ and we cannot use the well-known theorems about the convergence of the regularized elements to the unknown element $\hat{x}$ ([12], [13]). In these cases it is important to determine best procedures of regularization to a given level of measurement error and a given kind of a priori information. The knowledge of a best procedure allows one to characterize the usefulness of the mathematical treatment in a concrete situation of indirect measurement interpretation.

1. Criteria for regularizators

Inverse problems differ from one another by the amount of information about the unknown object and the measurement accuracy. In some cases we are interested in a complete reconstruction of the function $\hat{x}$,

in other situations it is enough to find the values of some functionals $l(\hat{x})$ on $\hat{x}$.

Thus we introduce different criteria for estimating the quality of a regularization method. We shall study these criteria in simplest model situations.

If the unknown element $\hat{x}$ belongs to a given set M and if $\|y\| \leqslant \delta$, a "pessimistic approach" consists in estimating regularization rules R by the upper bound for the reconstruction error in the most unfortunate situation, i.e. by the quantity

$$\max_{x \in M} \max_{\|y\|_Z \leqslant \delta} \|R(Ax+y)-x\|_x. \tag{1}$$

Another approach is based on the stochastic character of measurement error. If E denotes the mathematical expectation, for a random measurement error η we get an estimation by

$$\max_{x \in M} \mathrm{E}_\eta \|R(Ax+\eta)-x\|_x^2. \tag{2}$$

Some distribution moments of η are here assumed to be known or should be estimated parallelly with the unknown element.

In situations where the unknown function $\hat{x}$ may be interpreted as a realization of a random function (element) ξ, we accept the mean square error

$$\mathrm{E}_\xi \mathrm{E}_\eta \|R(A\xi+\eta)-\xi\|_x^2 \tag{3}$$

as an estimation of the regularization error.

If we are interested only in some functionals on $\hat{x}$, obviously we must substitute $|r(Ax+y)-l(x)|$ in place of the norm in (1)–(3). Here r denotes a special regularization rule for the determination of the functional l.

2. Results for the criterion (3)

Under the condition that the first and second moments of the random elements ξ and η are given (for simplicity they are assumed to be uncorrelated) optimal regularization rules are known. These results are extensions of well known results on optimal estimation rules for finite-dimensional spaces X and Z to Hilbert spaces [6]. Let be ξ and η random vectors in $\boldsymbol{R}^n$ resp. $\boldsymbol{R}^m$. Using the means $\bar{x} = \mathrm{E}\xi$, $\bar{y} = \mathrm{E}\eta$ ($= 0$ for simplicity) and the covariance matrices $B = \operatorname{cov}\xi$ and $C = \operatorname{cov}\eta$

$$B = \operatorname{cov}\xi = \mathrm{E}_\xi(\xi-\bar{x})(\xi-\bar{x})^T,$$

we have for any linear random functional $(l, \xi) = \sum_{i=1}^{n} l_i \xi_i$

$$\mathrm{E}(l, \xi) = (l, \bar{x}), \qquad \mathrm{E}[(l, \xi)-(l, \bar{x})]^2 = (l, Bl). \tag{4}$$

These properties may be also taken for the definition of the mean and the covariance operator of a random element ξ in a Hilbert space [10].

Using a linear rule

$$(r, z)+r_0 \tag{5}$$

to determine the value of the functional $(l, \hat{x})$, the estimation of type (3) may be written in the form

$$\begin{aligned} (6)\quad & \mathrm{E}_\xi \mathrm{E}_\eta \big((r, A\xi+\eta)+r_0-(l, \xi)\big)^2 \\ & = \mathrm{E}_\xi \mathrm{E}_\eta \big((A^*r-l, \xi-\bar{x})+(r, \eta)+(A^*r-l, \bar{x})+r_0\big)^2 \\ & = \big(A^*r-l, B(A^*r-l)\big)+(r, Cr)+\{(A^*r-l, \bar{x})+r_0\}^2. \end{aligned}$$

If C has a bounded inverse, expression (6) is minimal for

$$r = (ABA^*+C)^{-1}ABl \tag{7a}$$

$$= C^{-1}A(A^*C^{-1}A+B^{-1})^{-1}l \tag{7b}$$

and a corresponding value r_0. Obviously, formula (5) with the optimal r and r_0 can be transformed by (7a) so as to evaluate (l, x_{st}) for any linear functional l. Here

$$x_{\mathrm{st}} = \bar{x}+BA^*(ABA^*+C)^{-1}(z-A\bar{x}) \tag{8}$$

gives the best linear approximation of $(l, \hat{x})$ for the criteria of type (3).

The existence of a bounded inverse of C is a natural requirement for a random vector η. In general, the covariance operator of a random element in an infinite-dimensional Hilbert space is a nuclear operator ([10]). Thus we cannot use formula (7). To overcome the difficulties in this case, we can apply the concept of weakly random elements for the measurement error η ([1], [2]), because among weakly random elements there are elements with nonnuclear covariance operators. (The "white noise" with the covariance operator $\sigma^2 I$ may be treated as a weakly random element, too.)

Another way of approach is considered in [4]. In this paper an extension of the operator $BA^*(ABA^*+C)^{-1}$ has been constructed to an operator with a domain of probability measure 1 relative to the random element $A\xi+\eta$.

Although the Tikhonov regularization method was not founded on stochastic ideas, for the case $C = \sigma^2 I$ the element x_{st} minimizes the well known Tikhonov functional

$$\|Ax-z\|^2+\sigma^2\big(B^{-1}(x-\bar{x}), x-\bar{x}\big).$$

Let us mention the following important fact: For normally distributed random vectors ξ and η the vector x_{st} is the best estimator for $\hat{w}$

in the sense of the criteria of type (3) (not only among linear estimators of the type (6)). The same statement holds for Gaussian random elements in Hilbert spaces.

This approach has been used successfully in linearized models for the reconstruction of vertical temperature profiles by satellite measurement of radiation data. The comparison of the value (6) for optimal r, r_0 with the a priori dispersion (l, Bl) in an objective manner characterizes the usefulness of measured data.

In practice we usually have only statistical data as an approximation of $\bar{x}$ and B. The influence of such data on the accuracy of stochastic regularization was considered in [3]. For instance, we have not enough statistical data today for the reconstruction of vertical profiles of the earth structure.

3. Results on other criteria

The criterion of type (2) for linear functionals l has been considered in [11] in the special case of the set $M = \{x \mid (B^{-1}(x-\bar{x}), x-\bar{x}) \leqslant 1\}$ with a given element $\bar{x}$ and a given B. For this set (for simplicity let $\bar{x} = 0$) the criterion of type (2) leads to an expression similar to (6):

$$\max_{x \in M} \mathrm{E}_\eta [(r, Ax+\eta) - (l, x)]^2 = \max_{x \in M} (A^* r - l, x)^2 + (r, Cr)$$
$$= \|B^{1/2}(A^* r - l)\|^2 + (r, Cr).$$

Consequently the structure of an optimal linear rule for that criterion is similar to the structure of formula (8). Optimal linear rules R for the criterion (2) have to minimize the expression

$$\max_{x \in M} \mathrm{E}_\eta \|R(Ax+\eta) - x\|^2 = \max_{x \in M} \|(RA - I)x\|^2 + \operatorname{trace} RCR^*. \tag{9}$$

Under some special assumptions on M optimal linear rules R have been considered in [3] for an equation (1), where $A \in [\boldsymbol{R}^n \to \boldsymbol{R}^m]$. It seems to be of interest that the structure of optimal regularizators which minimize (6) or (9) differs from one another in the following sense: If we suppose $\operatorname{rank} B = \operatorname{rank} A = n \leqslant m$, then the approximations (8) (for all possible z) belong to the whole space $\boldsymbol{R}^n$, resp. to the whole set M. The optimal rule R which minimizes (9) in some sense gives an approximation $\tilde{x}$, which is similar to those constructed in [5], [7] and, in general, belongs to a subspace of $\boldsymbol{R}^n$ of dimension less than n.

4. Cross-validation and regularization

Without knowing the δ-level of the measurement accuracy a choice of regularization parameter can be based on the cross-validation-technique ([8]). To apply this approach, we decompose the system of n equations

$$Ax + y = z \tag{10}$$

into two disjoint subsystems, which we denote

$$A_1 x + y_1 = z_1,$$
$$A_2 x + y_2 = z_2.$$

From the first subsystem we calculate a regularized approximation

$$x_\alpha = (A_1^T A_1 + \alpha I)^{-1} A_1^T z_1$$

for $\hat{x}$ (for simplicity we use the regularization by means of $\alpha \|x\|^2$). The second subsystem is used to validate this approximation by means of the defect $A_2 x_\alpha - z_2$.

If we use different decompositions into subsystems $\{A_1^{(1)}, A_2^{(1)}\}, \ldots$ $\ldots, \{A_1^{(k)}, A_2^{(k)}\}$, we denote the corresponding regularized approximations by $x_\alpha^{(1)}, \ldots, x_\alpha^{(k)}$.

The cross-validation technique recommends the choice of a parameter α which minimizes the defect functional

$$\sum_{i=1}^{k} \|A_2^{(i)} x_\alpha^{(i)} - z_2^{(i)}\|^2. \tag{11}$$

The total calculation expense for minimizing (11) can be reduced by means of the factorization [9]

$$A = USV$$

with an upper bidiagonal matrix S and orthogonal matrices U and V, because we can evaluate the defect functional (11) without the knowledge of the regularized approximation $x_\alpha^{(i)}$.

We denote by U_1 and U_2 the portions of the orthogonal matrix U which correspond to A_1 and A_2 (we omit $^{(i)}$ for simplicity). For these matrices the equations $U_2 U_1^T = 0$, $U_2 U_2^T = I$ are satisfied.

One can easily prove the relation

$$x_\alpha = V^T S^T (SS^T + \alpha I)^{-1} (U_1^T z_1 + U_2^T \varrho), \tag{12}$$

where ϱ denotes the solution of the linear system

$$U_2 (SS^T + \alpha I)^{-1} \{U_2^T \varrho + U_1^T z_1\} = 0. \tag{13}$$

Using (12) and (13) the defect can be obviously transformed to

$$\begin{aligned} A_2 x_\alpha - z_2 &= U_2 [SVV^T S + \alpha I](SS^T + \alpha I)^{-1}\{U_2^T \varrho + U_1^T z_1\} - z_2 \\ &= U_2\{U_2^T \varrho + U_1^T z_1\} - z_2 = \varrho - z_2, \end{aligned}$$

and consequently, for finding the defect only the solution ϱ of the system (13) is needed.

Instead of a random decomposition of A into two submatrices A_1, A_2 we can use a complete family of systematic decompositions. For instance, we can choose the ith equation of the system (10) as $A_2^{(i)}$. In this case for all $i = 1, \ldots, n$ the systems (13) consist of one equation only.

References

[1] A. V. Balakrishnan, *Applied functional analysis*, Springer-Verlag, New York–Heidelberg–Berlin 1976.

[2] A. Bensoussan, *Filtrage optimal des systèmes linéaires*, Dunod, Paris 1971.

[3] V. Friedrich, B. Hofmann and U. Tautenhahn, *Möglichkeiten der Regularisierung bei der Auswertung von Meßdaten*, Wiss. Schriftenreihe der TH Karl-Marx-Stadt 10/1979 (ISSN 0323-6374).

[4] V. Friedrich und A. Uhlig, *Zur stochastischen Regularisierung linearer Gleichungen in Hilberträumen*, Beiträge zur Numerischen Mathematik 7 (1978), 33–48.

[5] R. J. Hanson, *A numerical method for solving Fredholm integral equations of the first kind using singular values*, SIAM J. Numer. Anal. 8 (1971), 616–622.

[6] O. N. Strand und E. R. Westwater, *Minimum RMS-estimation of the numerical solution of a Fredholm integral equation of the first kind*, ibid. 5 (1968), 287–295.

[7] J. M. Varah, *On the numerical solution of ill-conditioned linear systems with applications to ill-posed problems*, ibid. 10 (1973), 257–267.

[8] G. Wahba, *Practical approximate solution to linear operator equations when the data are noisy*, ibid. 14 (1977), 651–667.

[9] В. В. Воеводин, *О методе регуляризации*, Ж. вычисл. мат. и мат. физ. 9 (1969), 673–675.

[10] И. И. Гихман и А. В. Скороход, *Теория случайных процессов*, „Наука", Москва 1971.

[11] А. П. Петров, *Оценки линейных функционалов для решения некоторых обратных задач*, Ж. вычисл. мат. и мат. физ. 7 (1967), 648–654.

[12] А. Н. Тихонов, *О регуляризации некорректно поставленных задач*, Докл. АН СССР, 153 (1963), 49–52.

[13] А. Н. Тихонов и В. Я. Арсенин, *Методы решения некоторректных задач*, „Наука", Москва 1974.

Presented to the Semester
Computational Mathematics
February 20 – May 30, 1980