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

ON EMPIRICAL STOCHASTIC REGULARIZATION

BERND HOFMANN

Sektion Mathematik, Technische Hochschule Karl-Marx-Stadt, DDR

1. Introduction

A class of inverse problems from remote sounding, geophysics and material sciences consists in the interpretation of indirect measurements. An interesting but not directly measurable physical quantity x should be determined by measuring a dependent physical quantity b. Therefore in the case of linear dependence between x and b a Fredholm integral equation of the first kind

$$\int K(v, p)x(p)\,dp = b(v) \tag{1}$$

has to be solved. With continuous or square-integrable functions x and b over finite intervals and a smooth kernel k, the equation (1) is ill-posed in the sense of Hadamard (see [10]).

In satellite meteorology the solution of equation (1) provides an approximate determination of vertical temperature profiles of the atmosphere $x(p)$ as a function of the atmospheric pressure p. There the function $b(v)$ denotes the intensity of thermal radiation with wave-length v, which can be measured by a satellite (cf. [4]).

By using a quadrature formula, the integral equation (1) may be approximated by a linear algebraic system

$$Ax = b, \quad x \in \boldsymbol{R}^n,\ b \in \boldsymbol{R}^m, \text{ with } A \text{ an } (m, n)\text{-matrix}. \tag{2}$$

For sufficiently large n the vector x of the discretized problem (2) characterizes the solution function $x(p)$ sufficiently well.

To consider the measurement error in the sequel we examine instead of (2), the equation

$$Ax+y = z, \quad y \in \boldsymbol{R}^m,\ z \in \boldsymbol{R}^m \tag{3}$$

with the noise y and the real measured vector z. By means of suitable decision functions good approximations $\hat{x}(z)$ of the solution vector x can be found.

In consequence of the ill-posedness of the integral equation (1), the matrix A in systems (2) and (3) is ill-conditioned. Therefore and because of measurement economy the number m of rows will be small in many problems. In general, a full column-rank of A cannot be achieved for such problems.

In the followings we consider ways and properties of numerical solution of such ill-posed problems if the noise y and the solution have stochastic nature.

2. Stochastic regularization and ridge regression

We assume that the measured vector z has an error $y \in \boldsymbol{R}^m$ which is a realization of a centralized random vector η with the positive definite (m, m)-covariance-matrix C. The vector $x \in \boldsymbol{R}^n$ represents a realization of the random vector ξ with the expectation vector $\mathrm{E}\xi = \bar{x} \in \boldsymbol{R}^n$ and the positive definite (n, n)-covariance-matrix B. Under the condition that ξ and η are stochastic independent random vectors the equation (3) describes a random coefficient regression model (cf. [7], [9]). Approximations $\hat{x}(z)$ are estimations of the vector x. The quality of an estimator $\hat{x}$ should be evaluated by the Bayesian risk

$$r(\hat{x}) = \mathrm{E}\,\|\hat{x}(z) - x\|^2 \tag{4}$$

as an expected mean square error for the Euclidean norm.

For any inhomogeneous linear estimation $\hat{x}(z) = Qz + q$ with a constant vector $q \in \boldsymbol{R}^n$ and a constant (n, m)-matrix Q the Bayesian risk

$$r(\hat{x}) = \mathrm{Trace}\,\{(QA - I)B(QA - I)^T + QCQ^T\} + \|(QA - I)\bar{x} + q\|^2 \tag{5}$$

can be expressed explicitly. The *stochastic regularization* method (SR) (cf. [2], [3], [8], [11])

$$x_{\mathrm{SR}}(z) = \bar{x} + BA^T(ABA^T + C)^{-1}(z - A\bar{x}) \tag{6}$$

with

$$r(x_{\mathrm{SR}}) = \mathrm{Trace}\big(B - BA^T(ABA^T + C)^{-1}AB\big) \tag{7}$$

minimizes (5) relative to q and Q and provides the best such estimation for any kind of probability distribution when the first and second moments of ξ and η are given. The procedure (6) is defined for any dimension and rank of A. In practice, the matrix C has about diagonal structure and small condition number. Therefore (6) is numerically stable in most important cases.

It is difficult to apply SR in practice, because the moments $\bar{x}$ and B have to be known exactly. If such information is absent, the *ridge regression* method (RR) (cf. [5]) provides an approximation

$$x^L_{\text{ridge}}(z) = LA^T(ALA^T+C)^{-1}z \tag{8}$$

of SR. The quality of this procedure depends on the choice of a symmetric and positive semidefinite parameter matrix L. For $\bar{x} = 0$ and $L = B$ the estimations of SR and RR coincide. As the following lemma shows, there exists an upper bound for the Bayesian risk of ridge regression, which does not depend on L.

LEMMA. *Let $\bar{x} = 0$. Then for any symmetric and positive semidefinite matrix L the Bayesian risk* (4) *of a ridge regression estimation* (8) *is bounded by the inequality*

$$r(x^L_{\text{ridge}}) \leqslant \operatorname{Trace}\{B+(A^TC^{-1}A)^+\}. \tag{9}$$

Here the right-hand side is the matrix trace of the covariance matrix B and the pseudo-inverse of the matrix $A^TC^{-1}A$.

Proof (see [6]).

3. Empirical stochastic regularization

For the temperature profile determination the matrices A and C are available, whereas $\bar{x}$ and B have to be estimated from additional data. These additional data have empirical character. A sample $x^1, \ldots, x^N$ of a random vector ξ is obtained by direct temperature measurements from balloon ascents. The optimality of SR is essentially based on the knowledge of $\bar{x}$ and B. Therefore it is important to get good estimations for these two moments from the given sample.

The sample moments

$$\bar{x}^N = \frac{1}{N}\sum_{1}^{N} x^i \tag{10}$$

and

$$B^N = \frac{1}{N-1}\sum_{1}^{N}(x^i-\bar{x}^N)(x^i-\bar{x}^N)^T \tag{11}$$

are unbiased and consistent estimations for $\bar{x}$ and B (cf. [1], [9]). A useful modification of SR is found, if in formula (6) the moments $\bar{x}$ and B are replaced by the sample moments (10) and (11).

The procedure

$$x^N_{\text{ESR}}(z) = \bar{x}^N+B^NA^T(AB^NA^T+C)^{-1}(z-A\bar{x}^N) \tag{12}$$

will be called *empirical stochastic regularization* (ESR). For large sample size N the following theorem ensures the efficiency of ESR.

THEOREM. *The empirical stochastic regularization is asymptotically optimal, i.e. the limit condition*

$$\lim_{N\to\infty} r_{\mathrm{E}}(x_{\mathrm{ESR}}^N) = r(x_{\mathrm{SR}}) \tag{13}$$

holds. Here $r_{\mathrm{E}}(x_{\mathrm{ESR}}^N)$ *denotes the expectation value of* $r(x_{\mathrm{ESR}}^N)$ *relative to the probability distribution of the sample* $x^1, \ldots, x^N$.

Proof (see [6]).

The asymptotic optimality of ESR is based on the preceding lemma, i.e. on the existence of a common upper bound for ridge regression risk. A convergence theorem of Lebesgue (cf. [1]) allows to prove the theorem.

4. A Monte-Carlo simulation experiment

A computer simulation experiment helps to obtain information of the error behaviour $r_{\mathrm{E}}(x_{\mathrm{ESR}}^N)$ as a function of the sample size N. Under the condition that ξ is normally distributed the sample $x^1, \ldots, x^N$ may be generated by means of pseudo-random numbers. Thus the sample moments and, using formula (5), the risk $r(x_{\mathrm{ESR}}^N)$ are computable for the generated sample. Twenty examples of samples have been generated on the computer ES 1040 for any considered size N $(5 \leqslant N \leqslant 3200)$. The average values of $r(x_{\mathrm{ESR}}^N)$ serve as an approximation of $r_{\mathrm{E}}(x_{\mathrm{ESR}}^N)$.

In the experiment climatological data for temperature profile determination provide the matrices A, B and C with $m = 8$ and $n = 20$. Tables of numerical results of that simulation are published in the paper [6]. The experiment shows that, in general, for sufficiently large values of N, i.e. for $N > n$, the error of ESR may be expressed by the empirical formula

$$r_{\mathrm{E}}(x_{\mathrm{ESR}}^N) = r(x_{\mathrm{SR}}) + \frac{1}{N} \cdot \varkappa(A, B, C). \tag{14}$$

The value $\varkappa$ can be calculated by simulation for any special A, B and C. In most cases the weak dependence $\varkappa(\bar{x})$ will be neglected. Thus the formula (14) allows one to find minimal values of N, such that the risk $r_{\mathrm{E}}(x_{\mathrm{ESR}}^N)$ is bounded by a given value $r > r(x_{\mathrm{SR}})$.

References

[1] H. Bunke and J. Gladitz, *Empirical linear Bayes decision rules for a sequence of linear models with different regressor matrices*, Mathemat. Operationsforsch. Stat. 5 (1974), 235–244.

[2] V. Friedrich, *Zur stochastischen Regularisierung nichtkorrekter Gleichungen in Hilberträumen*; In: *Inverse and improperly posed problems*, Proc. of Conf. Halle 1979, ed. G. Anger, pp. 83–88, Akademie-Verlag, Berlin 1979.

[3] V. Friedrich, B. Hofmann und U. Tautenhahn, *Möglichkeiten der Regularisierung bei der Auswertung von Meßdaten*, Wiss. Schriftenreihe der TH Karl-Marx-Stadt, no. 10, 1979.

[4] В. А. Головко и др., *Одновременное определение температуры относительно геопотенциала и влагосодержания атмосферы по данным спектральных измерений*; In: *Дистационное зодирование атмосферы со спутника „Метеор"*, Гидр. изд., Ленинград 1979, 66–78.

[5] A. E. Hoerl and R. W. Kennard, *Ridge regression: Biased estimation for nonorthogonal problems*, Technometrics 12 (1970), 55–67.

[6] B. Hofmann, *Zur Methode der empirischen stochastischen Regularisierung*, Beiträge zur Numerischen Mathematik 11 (1983), 55–67.

[7] C. R. Rao, *Simultaneous estimation of parameters in different linear models and applications to biometric problems*, Biometrics 31 (1975), 545–554.

[8] O. N. Strand and E. R. Westwater, *Statistical estimation of the numerical solution of a Fredholm integral equation of the first kind*, JACM 15 (1968), 100–114.

[9] P. A. V. B. Swamy, *Statistical inference in random coefficient regression models*, Springer-Verlag, Berlin–Heidelberg–New York 1971.

[10] А. Н. Тихонов, В. Я. Арсенин, *Методы решения некорректных задач*, „Наука", Москва 1974.

[11] В. Ф. Турчип, В. П. Козиов, М. С. Маклевич, *Использование методов математический статистики для решения некоректных задач*, Успехи Физ. Наук 102 (1970), 345–386.

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