数理解析研究所講究録
第 215 巻 1974 年 190-209

STATISTICAL REGULARIZATION OF A NOISY ILL-CONDITIONED SYSTEM OF LINEAR EQUATIONS BY AKAIKE'S INFORMATION CRITERION

KUNIO TANABE

ABSTRACT

The problem of obtaining a reasonable solution of a noisy ill-conditioned system of linear equations whose coefficient matrix and right-hand-side vector are corrupted by random noise is considered.. It is well-known that direct application of usual numerical procedures to the equation will result in an oscilatory solution which is often not in good agreement with the nature of the problem. This paper introduces a statistical model which naturally leads to the approach associated with the singular value decomposition of the coefficient matrix and by using Akaike's information criterion an effective rank of the matrix is determined objectively to obtain a solution in which the gain in resolution is balanced against the amplification of the noise.

1. INTRODUCTION

The recovery of some definite structures from noisy data forms an important subject in experimental sciences. The "inverse" problems in physics are typical sources of this type of problems. These problems are inherently ill-conditioned in the sense that small disturbances in data have a disastrous effect on the direct estimate of the true structures, although sometimes the problems are considered improperly formulated. They are called "incorrectly posed problems" and extensive efforts have been devoted to them in the literature for reasonable estimate of the structure.

In this paper we consider a problem of obtaining a statistically reasonable solution of an ill-conditioned system of m linear equations in n unknowns

$$\bar{A}x = \bar{b} \tag{1}$$

whose $m \times n$ coefficient matrix $\bar{A}$ and the m-dimensional right-hand-side vector $\bar{b}$ are corrupted by random noise. The problem is closely associated with the numerical solution of the Fredholm integral equation of the first kind

$$\int K(s,t)\, x(t)dt = b(s) \tag{2}$$

with the singular or ill-conditioned kernel $K(s,t)$ and some disturbance in the right-hand-side $b(t)$, which arises in many branches of physical sciences, typically in the experimental sciences where physical data are measured by indirect sensing devices. To solve (2) numerically we discretize it either by numerical quadrature formula or by finite series expansion. These processes always introduce some errors into the original equation to produce an ill-conditioned equation (1).

As well demonstrated by Phillips [20], direct application of usual numerical procedures to the equation (1) will result in an oscillatory solution which is often not in good agreement with the nature of the problem. This subject was discussed extensively in the literature and various smoothing procedures were developed. [8 - 10, 12 - 20, 22 - 33]

The procedures are classified into two categories. The first one minimizes

$$||\bar{b} - \bar{A}x||^2 + \gamma||x - x_a||^2_W, \qquad (3)$$

where $||x|| = \sqrt{x^*x}$, $||x||_W = \sqrt{x^*Wx}$ and the positive difinite matrix W, an n-dimensional vector x_a and the smoothing parameter γ are determined from a prior information of the 'true' solution. The result thus obtained is given by

$$x = (A^*A + \gamma W)^{-1}(A^*b + \gamma W x_a). \qquad (4)$$

Phillips [20] and Tikhonov [28] originally devised this method. Foster [12] and Starnd & Westwater [26] gave statistical extensions and justifications of this approach to certain models in which disturbances in the coefficient matrix were not taken into account. Their method for determining the smoothing parameter is not entirely satisfactory since it depends on the knowledge of the covariance matrix of the statistical noise which is usually unknown. The other uses the singular value decomposition to invert the equation by the least squares method for the rank deficient case. Colub & Saunders [14], Hanson [16] and Varah [32] used an ad hoc method for determining an effective rank of $\bar{A}$. Both approaches will give satisfactory results so long as W, x_a and γ or an effective rank of $\bar{A}$ are suitably chosen. However, the choices depend more or less on subjective judgement of the investigator.

The purpose of this paper is to introduce a statistical model which naturally leads to the second approach and determine objectively an effective rank of $\bar{A}$ by using the Akaike's information criterion and obtain a reasonable solution in which the gain in resolution is balanced against the amplification of the noise. The criterion was first introduced successfully in the field of time series analysis and then extended to a general principle for fitting statistical models to data. [1-7]

2. SINGULAR VALUE DECOMPOSITION AND RANK CONSTRAINED INVERSES OF A MATRIX

In this section we define some matrix notations useful in the following sections. The singular value decomposition of a matrix is essential for our purpose.

Theorem 1. An $m \times n$ matrix A of rank r can be decomposed in the following way:

$$A = UDV^*,\ D = \begin{bmatrix} \sigma_1 & & & 0 \\ & \ddots & & \\ & & \sigma_r & \\ 0 & & & \end{bmatrix},\ \sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r \geq 0, \quad (5)$$

where U and V are unitary matrices of respective dimension m and n and V^* denotes the complex conjugate of V. If we denote the j-th column vectors of U and V by u_j and v_j respectively, the expression (5) is rewritten in

$$A = \sum_{j=1}^{r} \sigma_j u_j v_j^*, \quad (6)$$

in this case the Moore-Penrose inverse $A^\dagger$ of A, which is defined by

$$AA^\dagger A = A,\ A^\dagger AA^\dagger = A^\dagger,\ (A^\dagger A)^* = A^\dagger A \text{ and } (AA^\dagger)^* = AA^\dagger, \quad (7)$$

has the following expression:

$$A^\dagger = \sum_{i=1}^{r} \sigma_j^{-1} v_j u_j^*. \quad (8)$$

The Frobenius norm

$$\|A\|_F = \text{trace } (A^*A)^{1/2} = \text{trace } (AA^*)^{1/2}, \quad (9)$$

for $m \times n$ matrices is unitary invariant, i.e.,

$$||A||_F = ||U^*A||_F = ||AV||_F,$$

for any unitary matrices U and V.

The following theorem has been shown by Eckart and Young [11].

Theorem 2. Given an $m \times n$ matrix A of rank r and a non-negative integer $p(p \leq r)$, there exists uniquely an $m \times n$ matrix A_p of rank p that satisfies

$$||A - B||_F \geq ||A - A_p||_F = (\sum_{j=p+1}^{r} \sigma_j^2)^{1/2} \tag{10}$$

for any matrix B of rank p, and A_p can be expressed as

$$A_p = \sum_{i=1}^{p} \sigma_j u_j v_j^*.$$

This result is easily deduced from the singular decomposition of A, since $||\cdot||_F$ is unitary invariant. We will call A_p the rank p approximation to A.

Now we introduce rank constrained inverses of a matrix.

Definition Given a $m \times n$ matrix A of rank r and a non-negative integer p less than r, the rank p inverse $A_p^\dagger$ of A is the Moore-Penrose inverse of the rank p approximation A_p of A, i.e.,

$$A_p^\dagger = \sum_{j=1}^{p} \sigma_j^{-1} v_j u_j^*. \tag{11}$$

Note here that $A_p^\dagger$ satisfies

$$A_p^\dagger A A_p^\dagger = A_p^\dagger, \tag{12}$$

i.e. A is a generalized inverse of $A_p^\dagger$. See Rao & Mitra [21].

3. STOCHASTIC MODEL AND MAXIMUM LIKELIHOOD ESTIMATOR FOR THE COEFFICIENT MATRIX

We consider the following stochastic linear equation model:

$$(A + \tilde{E})x = b + \tilde{d} \tag{13}$$

where $A = (\alpha_{ij})$ is an m x n rank deficient constant matrix with rank r ($r < \min(m,n)$) and b is an m-dimensional constant vector such that

$$Ax = b \tag{14}$$

is solvable in x, i.e., b belongs to the image space $R(A)$ of A, $\tilde{E} = (\tilde{\varepsilon}_{ij})$ is an m x n random matrix whose elements $\tilde{\varepsilon}_{ij}$ are independent and identically normally distributed with mean zero and *unknown* variance η^2 and $\tilde{d} = (\tilde{\delta}_i)$ is an m-dimensional random vector whose elements $\tilde{\delta}_i$ are independent and identically normally distributed with mean zero and *unknown* variance θ^2. We assume $\tilde{E}$ and $\tilde{d}$ are distributed independently. We will denote $A + \tilde{E}$ and $b + \tilde{d}$ by $\tilde{A} = (\tilde{\alpha}_{ij})$ and $\tilde{b} = (\tilde{\beta}_j)$. $\tilde{E}$ and $\tilde{d}$ may be interpreted as the random disturbance of sensing devices in measurement and as the error of measurement respectively. In the case of numerical solution of the Fredoholm integral equation, they may be considered as discretization or trunction errors. We will denote sample values of $\tilde{A}$, $\tilde{E}$, $\tilde{b}$ and $\tilde{d}$ by $\bar{A} = (\bar{\alpha}_{ij})$, $\bar{E} = (\bar{\varepsilon}_{ij})$, $\bar{b} = (\bar{\beta}_j)$ and $\bar{d} = (\bar{\delta}_j)$ respectively.

Now the problem can be formulated as one of estimating the minimum norm solution $A^{\dagger}b$ of Eq. (14) from a sample equation

$$\bar{A}x = \bar{b}, \tag{15}$$

where $\bar{A}$ is almost surely of full rank and this equation is not always solvable. It should be noted $A^{\dagger}b$ is a constant vector. To solve this problem we first estimate the coefficient matrix A by the method of maximum likelihood.

Given a sample matrix $\bar{A}$, the likelihood function $L(\bar{A}: A,\eta)$ for the parameters A and η is given by

$$L(\bar{A}: A,\eta) = \prod_{i,j}^{m,n} (2\pi\eta^2)^{-1/2} exp(-(\bar{\alpha}_{ij} - \alpha_{ij})^2/2\eta^2)$$

$$= (2\pi\eta^2)^{-\frac{mn}{2}} exp(-\|\bar{A} - A\|_F^2/2\eta^2), \quad (16)$$

and the log-likelihood function is given by

$$logL(\bar{A}: A,\eta) = -\frac{mn}{2}log\ (2\pi\eta^2) - \frac{1}{2\eta^2}\|\bar{A} - A\|_F^2. \quad (17)$$

Here we consider the case where the rank r of the matrix A is known to be p, although we will treat in the next section a general case where the rank is unknown and estimated from the sample matrix $\bar{A}$. In this case, maximizing $logL(\bar{A}: A,\eta)$ with respect to A and η subject to rank$(A) = p$, we obtain the maximum likelihood estimates

$$\hat{A}_p(\bar{A}) = \bar{A}_p, \quad (18)$$

$$\hat{\eta}_p(\bar{A}) = \|\bar{A} - \bar{A}_p\|_F/\sqrt{mn} = (\sum_{i>p} \sigma_i^2(\bar{A})/mn)^{1/2} \quad (19)$$

of A and η, and we have

$$-2\ logL(\bar{A}: \hat{A}_p(\bar{A}),\hat{\eta}(\bar{A})) = mn(log\|\bar{A} - (\bar{A})_p\|_F^2 + log(2\pi e)/(mn))$$

$$= mn(log(\sum_{i>p} \sigma_i^2(\bar{A}))$$

$$+ log(2\pi e)/(mn)), \quad (20)$$

where $\bar{A}_p$ is the rank p approximation to $\bar{A}$ and $\sigma_i(\bar{A})$'s are the singular values of $\bar{A}$, numbered in decreasing order.

By the above discussion we have the following lemma.

Lemma 3. The maximum likelihood estimator $\hat{A}(\tilde{A})$ with rank p of A is the rank p approximation $\tilde{A}_p$ to $\tilde{A}$.

4. APPLICATION OF AKAIKE'S INFORMATION CRITERION FOR CHOOSING THE EFFECTIVE RANK

We are adopting $\tilde{A}_p^\dagger \tilde{b}$ as an estimator for $A^\dagger b$, but the problem of finding the rank p of the coefficient matrix remains to be settled. Before proceeding to this central problem, we give a brief sketch of the inherent difficulty of obtaining a reasonable solution of the equation (15).

We define a signal to noise (SN) ratio $\rho_{A/E}$ of the coefficient of the stochastic system (13) as

$$\rho_{A/E} = \frac{\sigma_r^2(A)}{m\eta^2}, \tag{21}$$

where $\sigma_r(A)$ is the smallest non-zero singular value of A. This definition is partly justified since we have

$$E(\tilde{A}^*\tilde{A}) = A^*A + m\eta^2 I, \tag{22}$$

whose singular values are

$$\sigma_i^2(A) + m\eta^2 \quad (i = 1, 2, \ldots n), \tag{23}$$

where $\sigma_i(A) = 0$ for $i > r$ and $E(\cdot)$ denotes the expectation. We also define an SN ratio $\rho_{A/d}$ of the stochastic system (13) as

$$\rho_{A/d} = \frac{\sigma_r^2}{\theta^2}. \tag{24}$$

When the rank of A is less than m and n, small errors $\tilde{\varepsilon}_{ij}$ in $\tilde{A}$ will yield small singular values of $\tilde{A}$, which together with the errors $\tilde{\delta}_j$ in $\tilde{b}$, can cause large errors in the least squares solution of the sample equations (15), although this phenomenon may be mitigated when the ratio θ/η is sufficiently small and the SN ratio $\rho_{A/E}$ is large. Also when SN ratio $\rho_{A/E}$ is of order less than one, the information corresponding to the singular value σ_r is covered by the random noise $\tilde{E}$, and it is difficult to recover A from $\bar{A}$, in which case however, the rank reduced approximation $\bar{A}_p (p < r)$ to $\bar{A}$ may retain some useful information about A. In any case the direct inversion of Eq. (15) by the usual least squares method is not an adequate approach and we must choose an effective rank p of $\bar{A}$, with which we may estimate $A^{\dagger}b$ by $\bar{A}_p^{\dagger}\bar{b}$. If we allow too large p, we pick up a great deal of noise. Contrarily if we choose p less than r, we lose some of the information about $A^{\dagger}b$. Hence the problem is how to balance the gain in resolution against the amplification of the noise in an estimate $\bar{A}_p^{\dagger}\bar{b}$ by choosing suitable p.

Akaike devised a new method for determining the degree of freedom of the statistical model which are to be fitted to finite number of data. Making use of the Kullback-Leibler's information quantity and the asymptotic theory of likelihood ratio test statistic, he proposed a new statistic which we will call Akaike's information criterion (*AIC*),

$$AIC = -2\log_e \text{(maximized likelihood of a model with respect to given data)} + 2\,\text{(degrees of freedom of the model)} \qquad (25)$$

and the model chosen is the one with the smallest *AIC* value among competetive models with varying degrees of freedom.[7] This procedure which we will call minimum *AIC* estimation (*MAICE*) differs from the existing methods in that it does not require such a subjective criterion as the significance level in statistical hypothesis testing. Its significance in the field of time series analysis was well demonstrated in Akaike's original papers. [1 - 7]

It is easily seen that the *MAICE* procedure can be incorporated to our problem of finding effective rank p of $\bar{A}$. Since $m \times n$ matrices with rank p have $p(m+n-p)$ degrees of freedom, our model (estimator for A) with rank p has $p(m+n-p)+1$ degree of freedom since η is taken into account. Hence given data $\bar{A}$, Akaike's information criterion $AIC(p:\bar{A})$ for our model with rank p is

$$\begin{aligned} AIC(p:\bar{A}) &= -2\log_e L(\bar{A}: \hat{A}_p(\bar{A}), \hat{\eta}(\bar{A})) + 2(p(m+n-p)+1) \\ &= mn \log_e \Big(\sum_{i>p} \sigma_i^2(\bar{A})\Big) + 2p(m+n-p)+k, \end{aligned} \tag{26}$$

where $k = mn \log_e (2\pi e/mn) + 2$.

Here it should be noted that the *MAICE* procedure is based on the assymptotic theory of likelihood ratio test statistics, and we must be careful to use it only in the situation where the number of parameters, $r(m+n-r)+1$ is relatively small compared with the number of data, mn, although it may be used as a heuristic procedure even when this condition is not satisfied. Hence it is sufficient to confine the alternative models to those whose rank p satisfy

$$\frac{p(m+n-p)}{mn} < \frac{1}{10} \text{ (say).} \tag{27}$$

We will denote the integer p which minimize $AIC(p:\bar{A})$ subjects to (27) by $p(\bar{A})$, and adopt it as an effective rank of $\bar{A}$.

Definition Given a matrix $\bar{A}$, the AIC inverse $\bar{A}^{\dagger}_{aic}$ of $\bar{A}$ is the rank $p(\bar{A})$ inverse $\bar{A}^{\dagger}_{p(\bar{A})}$ of $\bar{A}$. We adopt $\bar{A}^{\dagger}_{aic}\bar{b}$ as an estimator for $A^{\dagger}b$.

5. AIC RESOLVER

In this section we summarise the procedure, which we will call *AIC* resolver, for practical computation of $A^{\dagger}_{aic}b$ and give some modifications of it.

AIC RESOLVER

1. Compute the singular value decomposition

$$\overline{A} = \overline{U} \begin{bmatrix} \overline{\sigma}_1 & & & 0 \\ & \overline{\sigma}_2 & & \\ & & \overline{\sigma}_3 & \\ & & & \ddots \\ 0 & & & \end{bmatrix} \overline{V}^*, \quad \overline{\sigma}_1 \geq \overline{\sigma}_2 \geq \overline{\sigma}_3 \geq \ldots, \tag{28}$$

of $\overline{A}$, and compute $\overline{U}^*\overline{b} \equiv (\overline{\gamma}_1, \overline{\gamma}_2, \ldots \overline{\gamma}_m)^t$. An efficient algorithm for this was given by Golub and Reinsch. [13]

2. Find the positive integer p which minimizes

$$(mn)\,log_e(\sum_{i>p} \overline{\sigma}_i^2) + 2p(m+n-p), \tag{29}$$

subjects to the inequality (27). It should be noted that since only the differences of *AIC* are of interest an arbitrary common constant can be added to the definition of $AIC(p\colon \overline{A})$.

3. Compute

$$\overline{\eta}^2 \equiv \sum_{i>p} \overline{\sigma}_i^2/mn \tag{30}$$

$$\overline{\rho}_{A/E} \equiv \overline{\sigma}_p^2/(m\overline{\eta}^2), \tag{31}$$

$$\overline{\theta}^2 \equiv \sum_{i>p} \overline{\gamma}_i^2/(m-p) \text{ and} \tag{32}$$

$$\overline{\rho}_{A/d} \equiv \overline{\sigma}_p^2/\overline{\theta}^2, \tag{33}$$

which may serve as estimates for η^2, $\rho_{A/E}$, θ^2 and $\rho_{A/d}$ respectively.

4. Compute the vector

$$\sum_{i=1}^{p} (\overline{\gamma}_i/\overline{\sigma}_i)\overline{v}_i = \sum_{i=1}^{p} \sigma_i^{-1}(\overline{u}_i^t\overline{b})\overline{v}_i \qquad (34)$$

which is equal to $\overline{A}^{\dagger}_{aic}\overline{b}$, where $\overline{u}_i$ and $\overline{v}_i$ are the i-th column vectors of $\overline{U}$ and $\overline{V}$ respectively.

Now we will give some of heuristis modifications of the above procedure.

MODIFICATION 1: The choice of p in the above statement 2 depends only on the coefficient matrix $\overline{A}$ and not on the right-hand-side $\overline{b}$. This is justified when $\sigma_r(A)$ is comparatively larger than θ. But when $\rho_{A/d}$ is of smaller order than one, errors due to $\overline{d}$ will undesirably effect to the 'solution', in which case we should discard the information corresponding to σ_r. This suggests the following statement to be inserted after the statement 3.

3'. If $\overline{\rho}_{A/d}$ is less than one (say), reduce p by one and go to 3.

MODIFICATION 2: Since $\overline{\sigma}_i^2$'s are biased upwards by the amount about $m\eta^2$, we may use the reduced value $(\overline{\sigma}_i^2 - m\overline{\eta}^2)^{1/2}$ as an estimator for σ_i. In this case $\overline{\rho}_{A/E}$ should be larger than one or preferably more. This suggests the use of the following two statemetns in place of the statements 4.

3''. If $\overline{\rho}_{A/E}$ is less than ten (say), reduce p by one and go to 3.

4'. Compute the vector

$$\sum_{i=1}^{p} (\gamma_i/\sqrt{\bar{\sigma}_i^2 - m\bar{\eta}^2})\bar{v}_i \qquad (35)$$

which may serve as an estimator for $A^{\dagger}b$.

OTHER POSSIBLE MODIFICATIONS: Since given $\bar{A}$ we can calculate the estimates $\bar{\eta}^2$, $\bar{\rho}_{A/E}$, $\bar{\theta}^2$ and $\bar{\rho}_{A/d}$ by the formulae (30) - (33) for each value of p, if we have an a priori knowledge of the magnitude of either of the values η^2, $\rho_{A/E}$, θ^2 or $\rho_{A/d}$, we can easily incorporate this information into the estimation of r.

REFERENCES

[1] H. Akaike, "On a semi-automatic power spectrum estimation procedure", Proc. 3rd Hawaii Intl. Conf. on System Sciences pp. 974-977, 1970.

[2] H. Akaike. "Autoregressive model fitting for control", Ann. Inst. Statist. Math., Vol. 23, pp. 163-180, 1971.

[3] H. Akaike, "Determination of the number of factors by an extended maximum likelihood principle", Research Memorandum No. 44, The Institute of Statistical Mathematics pp. 1-11, 1971.

[4] H. Akaike, "Automatic data structure search by the maximum likelihood", Computers in Biomedicine Supplement to the Proceedings of the Fifth Hawaii International Conference on System Science, pp. 99-101, 1972.

[5] H. Akaike, "Use of an information theoretic quantity for statistical model identification", Proc. 5th Hawaii Intl. Conf. on System Sciences, pp. 249-250, 1972.

[6] H. Akaike, "A new look at the statistical model identification", submitted to IEEE Trans, Automatic Control.

[7] H. Akaike, "Information theory and an extension of the maximum likelihood principle", Problems of Control and Information Theory AKADÉMIAI KIADÓ.

[8] P. M. Anselone, "Uniform approximation the theory for integral equations with discontinuous kernels", SIAM J. Numer. Anal., Vol. 4, No. 2, pp. 245- , 1967.

[9] C. T. H. Baker, L. Fox, D. F. Mayers and K. Wright, "Numerical solution of Fredholm integral equations of first kind", The Computer Journal, Vol. 7, pp. 141-148, 1964-65.

[10] Lawrence Crone, "The singular value decomposition of matrices and cheap numerical filtering of systems of linear equations", J. Franklin Institute 294, pp. 133-136, 1972.

[11] C. Eckart and G. Young, "The approximation of one matrix by another of lower rank", Psychometrika 1, pp. 211-218, 1936.

[12] Manus Foster, "An application of the Wiener-Kolmogorov smoothing theory to matirx inversion", J. Soc. Indust. Appl. Math., Vol. 9, No. 3, pp. 387-392, 1961.

[13] G. H. Golub and C. Reinsch, "Contribution I/10 singular value decomposition and least squares solutions", Linear Algebra, Handbook for Automatic Computation, Vol. II. (Ed. J. H. Willkinson and C. Reinsch) pp. 134-151, 1970.

[14] G. H. Golub and M. A. Saunders, "Linear least squares and quadratic programming", Integer and Nonlinear Programming, pp. 229-256, 1970.

[15] Rudolf Gorenflo, "Computation of an integral trnasform of Abel's type in the presence of noise by quadratic programming", Proceedings of the IFIP Congress, pp. 575-576, 1965.

[16] Richard J. Hanson, "A numerical method for solving Fredholm integral equations of the first kind using singular values", SIAM. J. Numer. Anal., Vol. 8, No. 3, pp. 616-622, 1971.

[17] Richard J. Hanson, "Integral equations of immunology", Communications of the ACM, Vol. 15, No. 10, pp. 883-890, 1972.

[18] F. D. Kahn, "The correction of observational data for instrumental band width", P.C.P.S., Vol. 51, pp. 519-525, 1955.

[19] F. M. Larkin, "Optimal estimation of bounded linear functionals from noisy data", Information Processing 71, North-Holland, Publishing Company, pp. 1335-1345, 1972.

[20] David L. Phillips, "A technique for the numerical solution of certain integral equations of the first kind", J.A.C.M. Vol. 9, pp. 84-97, 1962.

[21] C. R. Rao and S. K. Mitra, "Generalized inverse of a matrix and its applications", John Wiley and Sons, New York, 1971.

[22] Philip W. Schaefer, "Improvable estimates in some non-well-posed problems for a system of elliptic equations", SIAM J. Math. Anal., Vol. 4, No. 3, pp. 447-455, 1973.

[23] C. B. Shaw, Jr., "Improvement of the resolution of an instrument by numerical solution of an integral equation", Journal of Mathematical Analysis and Applications 37, pp. 83-112, 1972.

[24] Harvey A. Smith, "Improvement of the resolution of a linear scanning device", J. SIAM Appl. Math., Vol. 14, No. 1, pp. 23-40, 1966.

[25] Otto Neall Strand and Ed R. Westwater, "Minimum-RMS estimation of the numerical solution of a Fredholm integral equation of the first kind", SIAM J. Numer. Anal., Vol. 5, No. 2, pp. 287-295, 1968.

[26] Otto Neall Strand and Ed R. Westwater, "Statistical estimation of the numerical solution of a Fredholm integral equation of the first kind", Journal of the Association for Computing Machinery, Vol. 15, No. 1, pp. 100-114, 1968.

[27] R. P. Tewarson, "Solution of linear equations in remote sensing and picture reconstruction", Computing 10, pp. 221-230, 1972.

[28] Tikhonov, A. N. "The stability of algorthms for the solution of degenerate systems of linear algebraic equations", USSR Computational Math. and Math. Physics 5, 181-188, 1965.

[29] V. F. Turchin, V. P. Kozlov, and M. S. Malkevich, "The use of mathematical-statistics methods in the solution of incor-

rectly posed problems", Soviet Physics USPEKHI, Vol. 13, No. 6, pp. 681-840, 1971.

[30] S. Twomey, "On the numerical solution of Fredholm integral equations of the first kind by the inversion of the linear system produced by quadrature", J.A.C.M., Vol. 10, pp. 97-101, 1963.

[31] S. Twomey, "The determination of aerosol size distributions from diffusional decay measurements", J. of Franklin Institute, Vol. 275, pp. 121-138, 1963.

[32] S. Twomey, "The application of numerical filtering to the solution of integral equations encountered in indirect sensing measurements", Journal of the Franklin Institute, Vol. 279, No. 2, pp. 95-109, 1965.

[33] J. M. Varah, "On the numerical solution of ill-conditioned linear systems with applications to ill-posed problems" SIAM. J. Numer. Anal., Vol. 2, pp. 257-267, 1973.

4'. Compute the vector

$$\sum_{i=1}^{p} (\gamma_i/\sqrt{\sigma_i^2 - m\bar{\eta}^2})\bar{v}_i \tag{35}$$

which may serve as an estimator for $A^{\dagger}b$.

OTHER POSSIBLE MODIFICATIONS: Since given $\bar{A}$ we can calculate the estimates $\bar{\eta}^2$, $\bar{\rho}_{A/E}$, $\bar{\theta}^2$ and $\bar{\rho}_{A/d}$ by the formulae (30) - (33) for each value of p, if we have an a priori knowledge of the magnitude of either of the values η^2, $\rho_{A/E}$, θ^2 or $\rho_{A/d}$, we can easily incorporate this information into the estimation of r.

Aknowledgement. The auther greatly aknowleges the valuable discussion with Dr. Akaike.