MATHEMATICAL FOUNDATIONS OF THE BURRUS TECHNIQUES FOR SPECTRAL UNFOLDING

Bert W. Rust
Computer Sciences Division
Union Carbide Corporation, Nuclear Division *
Oak Ridge, Tennessee

ABSTRACT

This paper treats the numerical solution of the first kind Fredholm integral equation in the case where the normal non-uniqueness and ill-conditioning problems are further aggravated by the presence of stochastic measuring errors. The basic ideas, due originally to Walter R. Burrus, of the techniques described are to seek statistical confidence interval estimates for the solution and to make the intervals as small as possible by using physically motivated *a priori* information to constrain the size of the set of permissible solutions. When these constraints are added to the classical linear regression model, the resulting interval estimation problems can be formulated in terms of parametric quadratic programming, but the solution of such problems is difficult and costly. Two more practicable methods have been developed which give suboptimal bounds, i.e., intervals which are somewhat wider than the optimally narrow ones that would be obtained from the quadratic programming procedure. One method uses various linear programming approximations while the other uses a constrained least squares procedure reminiscent of the Phillips-Twomey-Tikonov smoothing and regularizing techniques. Both methods are illustrated by mathematical and physically motivated examples. This work was supported by the U.S. Energy Research and Development Administration.

---

*Prime Contractor to the U.S. Energy Research and Development Administration

MATHEMATICAL FOUNDATIONS OF THE BURRUS TECHNIQUES FOR SPECTRAL UNFOLDING

B. W. Rust

Computer Sciences Division
at Oak Ridge National Laboratory
Union Carbide Corporation, Nuclear Division*

This paper is concerned with the numerical solution of Fredholm integral equations of the form

$$\int_a^b K(t,s)x(s)ds = \hat{y}(t) + \hat{\varepsilon}(t) \qquad (1)$$

where $\hat{\varepsilon}(t)$ is a stochastic measuring error. The kernel function $K(t,s)$ is assumed to be known quite accurately in comparison to the right-hand side function $\hat{y}(t)$. The methods herein described could be extended to handle errors in $K(t,s)$ also, but the implementation would be much more complicated, and at the present time it would not be possible to give completely rigorous statistical statements about the accuracy of the solutions obtained. Furthermore, Eq. (1) is a very good model for many physical situations, especially those involving measuring instruments whose response functions $K(t,s)$ can be accurately determined.

The first step in the numerical solution is to use a quadrature rule to replace the integral equation by a linear system

$$Kx = \hat{y} + \hat{\varepsilon} \qquad (2)$$

where K is an m x n matrix composed from the kernel function and the quadrature weights and x is an n-vector pointwise approximation to x(s).

---

*Prime contractor to the U.S. Energy Research and Development Administration.

In most physical situations $\hat{y}$ is an m vector of measurements, the number m being limited by experimental considerations. The number n, on the other hand, is limited only by the amount of work one is willing to do in order to get the solution. The linear system (2) is almost always very ill conditioned for any value of n large enough to give a reasonable approximation of (1) so the strategy adopted here is to try to make n so large that the discretization error is negligible in comparison to the statistical error. The usual result is a highly underdetermined linear system which will give stable solutions only if it is possible to impose a priori constraints. In physically motivated problems such constraints are almost always available, e.g., a simple non-negativity constraint on x is often remarkably effective in reducing the size of the solution set.

The stochastic error $\hat{\epsilon}$ is assumed to satisfy

$$E(\hat{\epsilon}) = 0, \quad E(\hat{\epsilon}\hat{\epsilon}^T) = S^{-2} \tag{3}$$

where S^{-2} is a known positive semidefinite matrix which, without loss of generality, can be assumed to be diagonal. Equations (2) and (3) constitute a statement of the classical linear regression model. Following the standard statistical practice we seek estimates for linear functions of the form

$$\phi(x) = w^T x \ , \tag{4}$$

which can be regarded as discrete approximations to integral quantities of the form

$$\phi[x(s)] = \int_a^b w(s)x(s)ds \ . \tag{5}$$

To find the solution vector x itself, one can simply successively take

$$w = e_i, \qquad i = 1,2,\ldots,n ,$$

where e_i is the $\underline{i}$th column of the identity matrix.

The basic idea of the methods described here is to seek statistical confidence interval estimates for the linear functions $\phi(x)$ and to make the intervals as small as possible by using the $\underline{a\ priori}$ information to constrain the size of the set of permissible solutions. This idea was first investigated by Walter R. Burrus[1] who used non-negativity constraints, reducing other types of constraints to non-negativity by a suitable transformation of variables. The first step is to pick a confidence level α and to consider the corresponding confidence ellipsoid in y-space, i.e.,

$$(y-\hat{y})^T S^{-2}(y-\hat{y}) \leq \mu^2 ,$$

where μ is determined by the value of α and the assumption that the error vector $\hat{\varepsilon}$ has a joint normal distribution. The x-space representation of this ellipsoid,

$$(Kx-\hat{y})^T S^{-2}(Kx-\hat{y}) \leq \mu^2 , \tag{6}$$

will in general be degenerate having axes of infinite length in the directions of the eigenvectors of $K^T S^{-2} K$ which correspond to zero eigenvalues. The intersection of this ellipsoid with the positive orthant will, hopefully, be bounded in most directions so one seeks the lower and upper bounds,

$$\begin{aligned} \phi^{lo} &= \min_x \{w^T x \mid x \geq 0, \quad (Kx-\hat{y})^T S^{-2}(Kx-\hat{y}) \leq \mu^2\} , \\ \phi^{up} &= \max_x \{w^T x \mid x \geq 0, \quad (Kx-\hat{y})^T S^{-2}(Kx-\hat{y}) \leq \mu^2\} , \end{aligned} \tag{7}$$

in the direction of the vector w.

The constraint equation (6) can also be written

$$(x-\hat{x})^T K^T S^{-2} K(x-\hat{x}) \leq \gamma^2 \tag{8}$$

where $\gamma^2 = \mu^2 - r_o$, r_o being the unconstrained least squares minimum value

$$r_o = \min_x \{(Kx-\hat{y})^T S^{-2}(Kx-\hat{y})\} .$$

Thus, the choice of μ^2 is equivalent to a choice of γ^2. It can be shown [cf. Rust and Burrus,[2] Chapter 6] that the normality assumption on $\hat{\varepsilon}$ implies that $(x-\hat{x})^T K^T S^{-2} K(x-\hat{x})$ follows a chi-square distribution with p degrees of freedom where p is the rank of $S^{-1}K$. Hence one can pick the value of γ^2 associated with a given confidence level α from a chi-square table using for p a conservative estimate of rank$(S^{-1}K)$, i.e., an estimate $\hat{p}$ such that rank$(S^{-1}K) \leq \hat{p}$.

Solving the constrained optimization problems (7) is an extremely difficult task. One approach [cf. Rust and Burrus[2], Chapter 5] is to calculate ϕ^{lo} and ϕ^{up} as the roots of the equation

$$L(\phi) = \mu^2$$

where $L(\phi)$ is the piecewise quadratic function

$$L(\phi) = \min_x \{(Kx-\hat{y})^T S^{-2}(Kx-\hat{y}) \mid x \geq 0, \quad w^T x = \phi\} . \tag{9}$$

This procedure amounts to parametric quadratic programming and would be extremely expensive. To avoid this difficulty two classes of suboptimal methods have been developed.[3] These methods seek to replace (7) with problems which can be solved more easily and which give interval solutions

that contain the solutions to (7) as subintervals. The price that must be paid in order to get the easier problem is a widening of the intervals obtained for any given confidence level α.

The first class of suboptimal methods requires that the user know from the outset lower and upper bounds (100% confidence bounds) for each component of the solution vector x. The lower bounds can always be taken to be zero, and the upper bounds do not have to be realistic. If the matrix K is nonnegative, as is often the case, upper bounds u_j can be computed from

$$u_j = \min_{1 \le i \le m} \left\{ \frac{\hat{y}_i + \mu s_i}{K_{ij}} \right\}, \quad j = 1,2,\ldots,n ,$$

where the s_i are the inverse square roots of the diagonal elements of S^{-2}. Denoting the lower bounds by v_i, one forms the diagonal matrix

$$Q = \operatorname{diag}\left(\frac{u_1 + v_1}{2} , \frac{u_2 + v_2}{2} , \ldots, \frac{u_n + v_n}{2} \right)$$

and considers convex combinations of the constraint ellipsoid (8) and the ellipsoid defined by

$$(x-Qe)^T Q^{-2} (x-Qe) \le n$$

where $e = (1,1,\ldots,1)^T$. These combinations, which are written

$$\eta \cdot \frac{1}{\mu^2} (Kx-\hat{y})^T S^{-2} (Kx-\hat{y}) + (1-\eta) \frac{1}{n} (x-Qe)^T Q^{-2} (x-Qe) \le 1 ,$$

with $0 < \eta < 1$, are themselves ellipsoids all of which contain as a subset the intersection of the constraint ellipsoid (8) and the positive orthant. Thus, for any value of η, the corresponding ellipsoid can be regarded as a confidence region for x with an associated confidence level greater than or equal to α. This means that each such ellipsoid can be used to generate confidence interval estimates for ϕ with ϕ^{lo} and ϕ^{up} being the values of ψ on the two support planes of the ellipsoid that are orthogonal to the vector w. These values can be computed by applying the standard, classical least squares interval estimation techniques to the augmented linear regression model

$$\begin{pmatrix} K \\ \frac{\tau}{\sqrt{n}} I \end{pmatrix} x = \begin{pmatrix} \hat{y} \\ \frac{\tau}{\sqrt{n}} Qe \end{pmatrix} + \begin{pmatrix} \hat{\varepsilon} \\ \varepsilon' \end{pmatrix} ,$$

with variance matrix

$$V^{-2} = \begin{pmatrix} S^{-2} & 0 \\ 0 & Q^{-2} \end{pmatrix} ,$$

where I is an identity matrix and τ is related to η by

$$\tau^2 = \frac{(1-\eta)}{\eta} \mu^2 \quad .$$

As η ranges from 1 to 0, τ ranges from 0 to ∞. At the present time there is no known strategy for choosing the best value of τ but clearly it is desirous to choose a value which makes the resulting convex combination ellipsoid in some sense a good approximation to the intersection of the

constraint ellipsoid with the positive orthant so that the resulting interval estimate is as close as possible to the optimal intervals that would be obtained from (7). If the vector w is successively taken to be the columns of the identity matrix, then the set of bounds that are obtained can be used start the procedure again and one can iterate the process as long as there is improvement in the bounds.

The second class of suboptimal methods is composed of linear programming approximations to (7). One way to produce such an approximation is to replace the confidence ellipsoid in y-space by a circumscribing polytope. If an infinity-norm, "square" circumscribing box is used, the linear programming problems are

$$\phi^{lo} = \min_x \left\{ w^T x \,\middle|\, x \geq 0, \begin{pmatrix} K \\ -K \end{pmatrix} x \leq \begin{pmatrix} \hat{y} + \mu Se \\ -\hat{y} + \mu Se \end{pmatrix} \right\} ,$$

$$\phi^{up} = \max_x \left\{ w^T x \,\middle|\, x \geq 0, \begin{pmatrix} K \\ -K \end{pmatrix} x \leq \begin{pmatrix} \hat{y} + \mu Se \\ -\hat{y} + \mu Se \end{pmatrix} \right\} .$$

If a one-norm, "diamond" box is used, the problems are

$$\phi^{\{lo\}_{up}} = \{\min_{max}\}_x \left\{ (w^T, 0^T) \begin{pmatrix} x \\ p \end{pmatrix} \middle| \begin{pmatrix} x \\ p \end{pmatrix} \geq 0, \begin{bmatrix} K & -I \\ -K & -I \\ 0^T & e^T S^{-1} \end{bmatrix} \begin{bmatrix} x \\ p \end{bmatrix} \leq \begin{bmatrix} \hat{y} \\ -\hat{y} \\ \mu\sqrt{m} \end{bmatrix} \right\}$$

where p is a dummy vector. Which box to use depends on the problem and at the present time it is not possible to tell a priori which box will give the narrowest interval estimates. For some problems narrower intervals can be obtained by first computing the singular value decomposition of $S^{-1}K$, i.e.,

$$S^{-1}K = L\Sigma^{\dagger}R^{T} ,$$

defining a new variable $x' = R^{T}x$, and circumscribing the box around the ellipsoid in x'-space. The resulting square box problems, for example, are

$$\phi^{\begin{Bmatrix}lo\\up\end{Bmatrix}} = \begin{Bmatrix}\min\\\max\end{Bmatrix}_{x} \left\{ w^{T}x \,\middle|\, x \geq 0, \begin{pmatrix} \Sigma^{\dagger}R^{T} \\ -\Sigma^{\dagger}R^{T} \end{pmatrix} x \leq \begin{pmatrix} L^{T}S^{-1}\hat{y} + \mu e \\ -L^{T}S^{-1}\hat{y} + \mu e \end{pmatrix} \right\} .$$

This procedure does not always give improved intervals, and, in fact, it sometimes makes them wider.

At the present time there is no way to tell which of the two classes of suboptimal methods will give the narrowest intervals for a given problem without trying them both. Furthermore, for both classes of methods there are a number of choices that can presently be made only by trial and error, e.g., the choice of τ or the choice of which circumscribing box to use. But in spite of these ambiguities good interval estimates can be obtained for many problems. Perhaps future developments will make the obtaining of such estimates less of an art and more of a science.

REFFRENCES

1. Burrus, Walter R., "Utilization of A Priori Information by Means of Mathematical Programming in the Statistical Interpretation of Measured Distributions," ORNL-3743 (June, 1965).

2. Rust, B. W. and Burrus, W. R., Mathematical Programming and the Numerical Solution of Linear Equations, American Elsevier (New York, 1972).

3. Rust, B. W. and Burrus, W. R., "Suboptimal Methods for Solving Constrained Estimation Problems," DASA 2604 (January, 1971).