CONFIDENCE INTERVALS FOR INEQUALITY-CONSTRAINED LEAST SQUARES PROBLEMS, WITH APPLICATIONS TO ILL-POSED PROBLEMS*

DIANNE P. O'LEARY† AND BERT W. RUST‡

Abstract. Computing confidence intervals for functions $\phi(x) = w^T x$, where $Kx = y + e$ and e is a normally distributed error vector, is a standard problem in multivariate statistics. In this work, we develop an algorithm for solving this problem if additional information, $x \geqq 0$, is given. Applications to estimating solutions to integral equations of the first kind are given.

Key words. confidence intervals, ill-posed problems, integral equations, quadratic programming

1. Introduction. Consider the linear model

$$(1.1) \qquad Kx = y + e$$

where K is a known $m \times n$ matrix, y is an $m \times 1$ vector of observations, x is the unknown solution vector, and e is an unknown $m \times 1$ error vector. The usual case is that $m \geqq n$. In this paper, we develop techniques for obtaining confidence interval estimates on functions $w^T x$ when extra information such as $x \geqq 0$ is known.

When the model above arises from discretization of ill-posed problems, such as integral equations of the first kind, the matrix K is highly ill-conditioned, and small changes in the right-hand side can make large changes in the solution x. Many approaches have been made to finding realistic solutions to this problem. Implicitly or explicitly, all of them make use of some extra information or side conditions, not included in (1.1), in order to eliminate unreasonable solutions. We summarize some of these methods below. Varah [20] gives a more extensive survey of some of the methods and the effects of ill-conditioning.

(a) Regularization techniques [8], [12], [17], [18] favor solutions that are smooth in the sense of having a small value of $\|Lx\|$, where L is a given matrix. Choosing L as a kth difference operator matrix forces the solution to have a small kth derivative. The solution is obtained by solving

$$\min_{x} \{\|Kx - y\|^2 + \eta \|Lx\|^2\}$$

where η is a given parameter. Large values of η force increasing smoothness; small values allow better fidelity to equation (1.1).

(b) Projection techniques [2], [16] restrict x to lie in some subspace:

$$x = Bu$$

where B is an $n \times p$ matrix of basis vectors, $p < n$. The objective is then to solve

$$\min_{u} \|KBu - y\|_2^2.$$

The columns of B are chosen to admit only those solutions which are smooth or have some other desirable property. The truncated singular value decomposition is one way

---

to choose B in a data-dependent way in order to force the solution vector to be of small norm.

(c) Side conditions [1], [7], [9], [13] may be used explicitly in order to eliminate undesirable solutions. For instance, in many physical applications the solution is known to be nonnegative, and may be estimated by

$$\min_{x \geqq 0} \|Kx - y\|_2^2.$$

Confidence interval methods use statistical information on the distribution of e in order to estimate the solution. In contrast to the methods above, these procedures give not only an estimate of the true solution x^*, but also a region which contains x^* with a given probability. That is, if a probability α is given and the experiment is repeated many times, then the region will contain the true result $100\alpha\%$ of the time. The vector e is assumed to be normally distributed with mean 0 and nonsingular variance matrix S^2 where S is symmetric. Then the best linear unbiased estimate of the true solution is that vector $\bar{x}$ which solves

$$\min_x \|Kx - y\|_S^2,$$

where $\|z\|_S^2 = z^T S^{-2} z$. For any linear function $\phi = w^T x$, the best linear unbiased estimate $w^T \tilde{x}$ is normally distributed about the true value $\phi^* = w^T x^*$ with variance $w^T (K^T S^{-2} K)^{-1} w$. For any given probability α, if κ is chosen so that

$$\alpha = \int_{-\kappa}^{\kappa} n(x; 0, 1)\, dx,$$

where $n(x; 0, 1)$ is the probability density function for the standard normal distribution, then the interval $[\phi_{\text{LO}}, \phi^{\text{UP}}]$ is a $100\alpha\%$ confidence interval for ϕ^* when ϕ_{LO} and ϕ^{UP} are given by

$$w^T \tilde{x} \pm \kappa \sqrt{w^T (K^T S^{-2} K)^{-1} w}.$$

We can write ϕ_{LO} and ϕ^{UP} in another way, similar to the forms we will be using in the next section. Using the technique of Lagrange multipliers, it can be shown that

$$\phi_{\text{LO}} = \min_x \{w^T x \colon \|K(x - \tilde{x})\|_S^2 = \kappa^2\},$$

$$\phi^{\text{UP}} = \max_x \{w^T x \colon \|K(x - \tilde{x})\|_S^2 = \kappa^2\}.$$

Note that a confidence interval can be given for any component x_i^* of the solution vector by choosing w equal to the ith unit vector.

In this paper we combine the techniques of confidence interval estimation and the use of inequality constraints as side conditions to develop a mathematical framework and some numerical techniques for computing ϕ_{LO} and ϕ^{UP} satisfying:

$$\text{Prob}\, \{\phi_{\text{LO}} < \phi^* < \phi^{\text{UP}}\} \geqq \alpha$$

when it is known a priori that

$$x^* \geqq 0.$$

This work can be considered an extension of the work of Cope and Rust [7] on confidence interval estimation with inequality constraints. The case of linear equality constraints has been rather well-studied; see, for example, Rao [13]. In this paper,

some statistical background is discussed in § 2, algorithms are presented in § 3, and computational results are given in § 4.

2. Statistical framework. Throughout this work we make the following assumptions:

(a) $Kx = y + e$.

(b) $e \sim N(0, S^2)$, where S is nonsingular symmetric.

(c) It is known a priori that x is nonnegative.

See Bard [3, p. 180] for a discussion of the implication of (c) on the form of the distribution.

In this section we present the results that define the computational task of computing a confidence interval for $\phi = w^T x$.

THEOREM 1. *Under the assumptions above, the probability that ϕ is contained in the interval $[\phi_{LO}, \phi^{UP}]$ is greater than or equal to α, where*

$$\phi_{LO}^{UP} = \frac{\max}{\min} \{w^T x : \|Kx - y\|_S \leq \mu, x \geq 0\},$$

$$\text{rank}(K) = q,$$

$$\int_0^{\gamma^2} \chi_q^2(\rho) \, d\rho = \alpha,$$

$$r_0 = \min_x \|Kx - y\|_S^2,$$

$$\mu^2 = r_0 + \gamma^2,$$

and χ_q^2 is the probability density function for the chi-squared distribution with q degrees of freedom.

Proof. If μ^2 is chosen as stated above, then $\{x : \|Kx - y\|_S \leq \mu\}$ is a $100\alpha\%$ confidence region (ellipsoidal) for the true x^* [15, § 6.4]. (Note that this makes no use of the information that x is nonnegative.) At the same time, the region $\{x : x \geq 0\}$ is a 100% confidence region. Intersecting the two, we find that $\{x : \|Kx - y\|_S \leq \mu, x \geq 0\}$ is a $100\alpha\%$ confidence region for x^*. Thus, for $\phi = w^T x$, a $100\alpha\%$ confidence interval is given by $[\phi_{LO}, \phi^{UP}]$, where ϕ_{LO} and ϕ^{UP} are, respectively, the maximum and minimum of ϕ over this set.

In some cases the procedure outlined by this theorem could fail to give useful information. The set over which the maximum and minimum are taken could be null, although a nonnull set always exists for γ large enough. Further, unless w is orthogonal to the null space of K (i.e., unless w satisfies an estimability condition), the interval could be infinite for all α; however, the interval could have one or two finite endpoints even if w does not satisfy this orthogonality condition, and this technique could give information on ϕ even in these cases where, without the side conditions, a minimum variance unbiased estimator fails to exist.

Note that if several vectors w_ν, $\nu = 1, \cdots, N$ are given, the procedure above gives *simultaneous* confidence intervals for all $\phi_\nu = w_\nu^T x$, since the region over which the maximum and minimum are taken is a $100\alpha\%$ confidence region for x^*; i.e., the probability that all $w_\nu^T x$ are in the computed intervals is $100\alpha\%$.

LEMMA 1. *Let*

$$L(\phi) = \min_x \{\|Kx - y\|_S^2 : x \geq 0, w^T x = \phi\}.$$

Then $L(\phi)$ is unimodal, piecewise quadratic, differentiable, and convex.

Proof. $L(\phi)$ is the objective function value for a convex parametric quadratic programming problem. This is the essence of the argument proving the lemma. A more detailed discussion follows, since this argument forms the basis for algorithms in the next section.

Let I be a set of indices, let $\bar{I}$ be its complement in the set $\{1, 2, \cdots, n\}$, and hold x_i at zero for $i \in \bar{I}$. We partition the x vector as $(x_I^T, x_{\bar{I}}^T)$, and partition K and w to conform to it. Then, using Lagrange multipliers, the solution to the problem

$$(2.1) \qquad L(\phi, I) = \min_x \{\|Kx - y\|_S^2 : x_{\bar{I}} = 0, w^T x = \phi\}$$

is determined by

$$(2.2) \qquad \begin{bmatrix} K_I^T S^{-2} K_I & w_I \\ w_I^T & 0 \end{bmatrix} \begin{bmatrix} x_I \\ \lambda \end{bmatrix} = \begin{bmatrix} K_I^T S^{-2} y \\ \phi \end{bmatrix}.$$

If K_I is not full column rank, then the solution may be nonunique, but we will take the solution $x_I = x_I^P + x_I^N$ with the smallest null space component x_I^N needed to keep $x_I \geqq 0$. For a given index set I, x_I^P is a linear function of ϕ, and $L(\phi, I)$ is convex and quadratic in ϕ. Furthermore, it is clear that

$$L(\phi) = \min_I \{L(\phi, I) : x_I \geqq 0\}.$$

The vector x for which this minimum is achieved is a continuous function of ϕ, and therefore $L(\phi)$ is a continuous and differentiable function of ϕ. Let $\min_x \{\|Kx - y\|_S^2 : x \geqq 0\}$ be attained at $\tilde{x}$. (If the minimum is achieved at a set of points, let $\tilde{x}$ be any one for which $w^T\tilde{x}$ is minimal.) Let $\tilde{\phi} = w^T\tilde{x}$. Then $L(\phi)$ has a global minimum at $\tilde{\phi}$. A sketch of $L(\phi)$ is given in Fig. 1. In the following discussion we assume that a designated point $\hat{\phi}$ is greater than $\tilde{\phi}$. The argument for $\hat{\phi} < \tilde{\phi}$ is analogous.
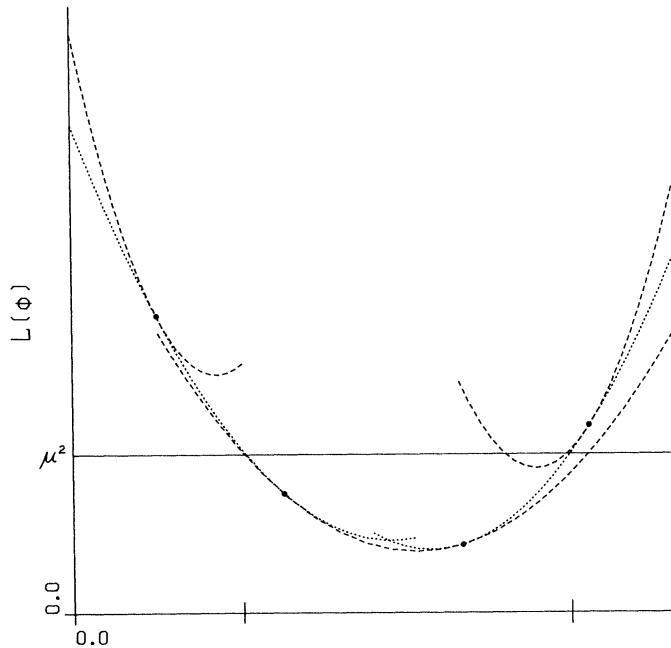

FIG. 1. *The $L(\phi)$ curve, showing various quadratic segments. The points ϕ_{LO} and ϕ^{UP} are indicated on the ϕ axis.*

Let $\hat{\phi} > \tilde{\phi}$ be the first point where an index j needs to be dropped from the index set I (because x_j has been driven to zero and would go negative if it were kept in the set), or a variable needs to be added to the index set (because using it produces a smaller function value without violating any constraint). Then the set of equations defining x as a function of ϕ has one fewer or one extra degree of freedom, but $L(\phi)$ over the next interval (i.e., until the index set changes again) is still convex and quadratic. Let I denote the previous index set, and $\hat{I}$ the index set for $\phi \geqq \hat{\phi}$. Since $dL(\hat{\phi}, \hat{I})/d\phi = dL(\hat{\phi}, I)/d\phi$ and $dL(\hat{\phi}, I)/d\phi \geqq 0$, then $L(\phi, \hat{I})$ must be monotone nondecreasing for $\phi \geqq \hat{\phi}$.

This argument can be repeated on all intervals beyond this one, thus establishing the result.

THEOREM 2. *The values* $\phi_{\mathrm{LO}}^{\mathrm{UP}}$ *are defined by the two extreme roots of* $L(\phi) - \mu^2 = 0$, *where* $L(\phi) = \min_x \{\|Kx - y\|_S^2 : x \geqq 0, w^T x = \phi\}$.

Proof. We sketch the proof for ϕ_{LO}; the discussion for ϕ^{UP} is analogous.

Let $\phi_{\mathrm{LO}} = \min_x \{w^T x : \|Kx - y\|_S^2 \leqq \mu^2, x \geqq 0\}$ and let x_{LO} be an x vector for which the minimum is attained. Similarly, let $\bar{\phi}_{\mathrm{LO}}$ and $\bar{x}_{\mathrm{LO}}$ be the corresponding minimal root and x values for $L(\phi) - \mu^2 = 0$. Then

$$L(\bar{\phi}_{\mathrm{LO}}) = \|K\bar{x}_{\mathrm{LO}} - y\|_S^2 = \mu^2, \quad \bar{x}_{\mathrm{LO}} \geqq 0, \quad \text{and} \quad w^T \bar{x}_{\mathrm{LO}} = \bar{\phi}_{\mathrm{LO}}.$$

Therefore,

$$\phi_{\mathrm{LO}} = \min_x \{w^T x : \|Kx - y\|_S^2 \leqq \mu^2, x \geqq 0\} \leqq \bar{\phi}_{\mathrm{LO}},$$

since $\bar{x}_{\mathrm{LO}}$ is one of the candidate points in the minimization.

On the other hand,

$$L(\phi_{\mathrm{LO}}) = \min_x \{\|Kx - y\|_S^2 : x \geqq 0, w^T x = \phi_{\mathrm{LO}}\} \leqq \|Kx_{\mathrm{LO}} - y\|_S^2 \leqq \mu^2,$$

since x_{LO} is one of the candidates in the minimization. Thus $L(\phi_{\mathrm{LO}}) \leqq \mu^2$, and, since by Lemma 1, $L(\phi)$ is unimodal and convex, this implies that $\phi_{\mathrm{LO}} \geqq \bar{\phi}_{\mathrm{LO}}$. Thus $\phi_{\mathrm{LO}} = \bar{\phi}_{\mathrm{LO}}$.

3. Algorithms for constrained interval estimation. Two approaches to solving $L(\phi) = \mu^2$ are discussed in this section. Methods for finding ϕ_{LO} are discussed; methods for computing ϕ^{UP} are analogous. The algorithms are written under the assumption that K has full column rank. Modifications for the rank deficient case are possible but tedious.

3.1. Tracing the $L(\phi)$ curve. Conceptually, the simplest approach for finding ϕ_{LO} is to start from a known lower bound ϕ_0 (or some other convenient point) and follow the $L(\phi)$ curve until it crosses μ^2. This is a parametric quadratic programming problem as ϕ varies:

$$(3.1) \qquad \min_x \{\|Kx - y\|_S^2 : x \geqq 0, w^T x = \phi\}.$$

The optimal values of the objective function form a piecewise quadratic function of ϕ; the nodes occur where variables x change from zero to positive or vice versa. The algorithm is as follows:

ALGORITHM TRACE
1. Solve (3.1) for $\phi = \phi_0$ and set $\phi_{\mathrm{NODE}} = \phi_{\mathrm{NEW}} = \phi_0$.
2. Until $L(\phi_{\mathrm{NEW}}) < \mu^2$
    Let $\phi_{\mathrm{NODE}} = \phi_{\mathrm{NEW}}$.

Find the next node, the minimum value $\phi_{\text{NEW}} > \phi_{\text{NODE}}$ such that a variable x_j changes from zero to positive or vice versa.

3. The value ϕ_{LO} is now known to satisfy $\phi_{\text{NODE}} \leqq \phi_{\text{LO}} < \phi_{\text{NEW}}$. It can be determined by solving a quadratic equation.

We now provide more details on each step.

Step 1. The problem (3.1) can be solved, for example, by the code WNNLS of Haskell and Hanson [10], but Step 2 is done more efficiently using an algorithm from which matrix factorizations are accessible. We will present such an algorithm in § 3.2.

Step 2. By equation (2.2), the value of x over a particular quadratic segment can be determined by

$$A \begin{bmatrix} x_I \\ \lambda \end{bmatrix} = \begin{bmatrix} K_I^T S^{-2} y \\ \phi \end{bmatrix},$$

$$x_{\bar{I}} = 0$$

where

$$A = \begin{bmatrix} K_I^T S^{-2} K_I & w_I \\ w_I^T & 0 \end{bmatrix}.$$

Thus

$$\begin{bmatrix} x_I(\phi_{\text{NODE}} + \Delta\phi) \\ \lambda(\phi_{\text{NODE}} + \Delta\phi) \end{bmatrix} = \begin{bmatrix} x_I(\phi_{\text{NODE}}) \\ \lambda(\phi_{\text{NODE}}) \end{bmatrix} + \Delta\phi A^{-1} \begin{bmatrix} 0 \\ 1 \end{bmatrix},$$

and it is easy to find the minimum positive $\Delta\phi$ which drives a variable in x_I to zero. Similarly, the minimum positive $\Delta\phi$ for which a variable in $x_{\bar{I}}$ becomes positive can also be determined. Note that the optimality conditions for (3.1) are

$$\begin{bmatrix} K^T S^{-2} K & w & I \\ w^T & 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ \lambda \\ \begin{bmatrix} 0 \\ \delta_{\bar{I}} \end{bmatrix} \end{bmatrix} = \begin{bmatrix} K^T S^{-2} y \\ \phi \end{bmatrix},$$

$$x_i \delta_i = 0, \qquad i \in I,$$

$$x_I \geqq 0, \qquad x_{\bar{I}} = 0,$$

$$\delta \geqq 0.$$

Thus a variable in $\bar{I}$ is released from zero when the corresponding δ value is driven to zero, and the first equation block above can be used to calculate $\delta_{\bar{I}}$ as a function of ϕ.

Step 3. To find ϕ_{LO}, note that

$$L(\phi) = (Kx(\phi) - y)^T S^{-2} (Kx(\phi) - y)$$

and, from Step 2,

$$x(\phi) = x_{\text{NODE}} + \Delta\phi d$$

where d is a computable column vector. Thus,

$$L(\phi) = L(\phi_{\text{NODE}}) + 2\Delta\phi d^T K^T S^{-2} (Kx_{\text{NODE}} - y) + (\Delta\phi)^2 d^T K^T S^{-2} Kd,$$

and solving $L(\phi) = \mu^2$ just involves finding the root of the quadratic equation in the interval $[\phi_{\text{NODE}}, \phi_{\text{NEW}}]$.

Algorithm Trace is appealing in its simplicity, and is quite efficient and adequate for well-conditioned problems. In practice, however, it does not work well on the ill-conditioned problems of interest to us. Degeneracies are often encountered in which the index set I makes several changes with no appreciable change in ϕ. Computationally, this can cause an infinite loop, with variables entering and leaving the set I without changing L or ϕ_{NODE}. A more robust although somewhat more expensive approach is given in the next section.

3.2. Rootfinding for $L(\phi) = \mu^2$. This approach amounts to applying a nonlinear equation solver to $L(\phi) = \mu^2$, solving a quadratic programming problem whenever a function evaluation is needed. The simplest example is:

ALGORITHM BRENT-WNNLS
    Brent's program ZEROIN [5] is used as the rootfinder, and Haskell and Hanson's WNNLS [10] is used to evaluate $L(\phi)$.

This algorithm has the advantage of being easy to program, but loses efficiency in two ways: ZEROIN does not take advantage of the piecewise quadratic nature of the function, and WNNLS solves each problem starting with the guess $x = 0$, taking no advantage of previous information. Efficiency can be gained by tailoring the rootfinder and the quadratic program solver to this application. We discuss these two alternate algorithms in turn.

Hanson and Haskell's WNNLS, based on Lawson and Hanson's NNLS [11], solves the problem

$$\min_x \| Kx - y \|_S^2, \qquad x \geqq 0, \quad w^T x = \phi$$

by converting it to

$$\min_x \left\| \begin{bmatrix} \alpha w^T \\ S^{-1} K \end{bmatrix} x - \begin{bmatrix} \alpha \phi \\ S^{-1} y \end{bmatrix} \right\|^2, \qquad x \geqq 0$$

where α is a large weight, related to the wordlength of the machine, so that the equation $w^T x = \phi$ is satisfied to within a very small tolerance. We modify the algorithm NNLS to

1) Start from the solution to the previous problem rather than from $x = 0$.
2) Compute and save the orthogonal factorization of the least squares matrix in a form usable for side calculations in the rootfinder.

A description of the algorithm follows. The numbering of the steps parallels Lawson and Hanson's description in [11, p. 161].

ALGORITHM LS.
Saved from previous solution: the index set I of nonzero x variables and a factorization

$$\begin{bmatrix} \alpha w_I^T \\ S^{-1} K_I \end{bmatrix} = QR, \quad \text{such that } Q^T Q = I, \quad \text{and } R = \text{upper triangular}$$

1. Compute x from

$$Rx_I = Q^T \begin{bmatrix} \alpha \phi \\ S^{-1} y \end{bmatrix}, \qquad x_{\bar{I}} = 0.$$

If any $x_j < 0$, drop such indices j from I and such columns j from the QR factors and repeat this step.

2. Compute the residual r and negative half gradient g:

$$r = \begin{bmatrix} \alpha\phi \\ S^{-1}y \end{bmatrix} - \begin{bmatrix} \alpha w^T \\ S^{-1}K \end{bmatrix} x, \qquad g = [\alpha w, K^T S^{-1}] r.$$

3. If $g_{\bar{I}} \leq 0$ or if $\bar{I}$ is null, then the optimal solution has been found.
4. Let $g_t = \max_{j \in \bar{I}} g_j$.
5. $I = I \cup \{t\}$. Add column t to the QR factors.
6. Compute z from

$$Rz_I = Q^T \begin{bmatrix} \alpha\phi \\ S^{-1}y \end{bmatrix}, \qquad z_{\bar{I}} = 0.$$

7. If $z_I \geq 0$ then let $x = z$ and go to Step 2.
8-9. Otherwise, a variable needs to be dropped. Find an index $q \in I$ such that

$$\hat{\beta} = \frac{x_q}{x_q - z_q} = \min_{\substack{z_j \leq 0 \\ j \in I}} \frac{x_j}{x_j - z_j}.$$

10. Let $x = x + \hat{\beta}(z - x)$.
11. Let $I = I \backslash \{j : x_j = 0\}$. Drop column q (and any others for which the x component is zero) from the QR factors. Go to 6.

The QR updating and downdating in LS can be performed using a modified Gram–Schmidt algorithm.

Tailoring the rootfinder to the piecewise quadratic nature of the $L(\phi)$ curve also enhances efficiency. We present an algorithm which keeps the root bracketed but bases its new prediction on a quadratic extrapolation of the function.

ALGORITHM BRACKET
Given: an initial interval $[\phi_a, \phi_b]$ containing a single root, and two convergence tolerances ε_1 and ε_2. Let $\Delta\phi = 0$ and let $\phi = \phi_b$ if looking for ϕ_{LO} and $\phi = \phi_a$ if looking for ϕ^{UP}. Until $|(\phi_a - \phi_b)/\phi| < \varepsilon_1$ or $|(L(\phi) - \mu^2)/\mu^2| < \varepsilon_2$, perform steps 1 through 5.
1. Let $\Delta\phi_{OLD} = \Delta\phi$.
2. Make a prediction of the root, $\phi + \Delta\phi$, based on the current quadratic segment: *Find $\Delta\phi_T$ such that $L(\phi + \Delta\phi_T) = \mu^2$ assuming that no basis changes occur.* This involves solving a quadratic equation for $\Delta\phi_T$. Set $\Delta\phi = \Delta\phi_T$.
3. If $\Delta\phi_T$ is too small, increase it:

$$\text{If } |\Delta\phi_T| < .9\varepsilon_1|\phi| \text{ then } \Delta\phi_T = .9\varepsilon_1|\phi| \text{ sgn } (\Delta\phi_T).$$

4. Reject the quadratic prediction if
   a) convergence is linear (i.e., $\Delta\phi_{OLD}/\Delta\phi$ is close to 1), or
   b) the new prediction $\phi + \Delta\phi_T$ does not fall within the current interval, or
   c) the change in ϕ is small and both the new and the old ϕ values are on the same side of the root (i.e., $\Delta\phi \neq \Delta\phi_T$ and $(L(\phi) - \mu^2) \times (L(\phi + \Delta\phi_T) - \mu^2) > 0$).
   In these cases, use the bisection algorithm (i.e., set $\phi = (\phi_a + \phi_b)/2$). Otherwise, ϕ is set equal to the quadratic prediction $\phi = \phi + \Delta\phi_T$.
5. Evaluate $L(\phi)$. Shorten the interval known to contain the root: if $(L(\phi_a) - \mu^2)(L(\phi) - \mu^2) > 0$ then $\phi_a = \phi$; otherwise, $\phi_b = \phi$.

Algorithm Bracket-LS has been successful on many ill-conditioned test problems. Examples are given in the next section.

4. Numerical results. We present the results of numerical experiments on two problems. The first is a standard test problem, and we use it to compare the performance of several algorithms. The second is a problem in radiation physics, presented to demonstrate the practical utility of the Bracket-LS algorithm.

Example 1. The Phillips equation [12] is as follows:

$$\int_{-6}^{6} K(t, s)x(s)\, ds = y(t), \qquad |t| \leqq 6,$$

where

$$K(t, s) = \begin{cases} 1 + \cos\left(\pi(s-t)/3\right), & |s-t| \leqq 3, \quad |t| \leqq 6, \\ 0, & |s-t| \geqq 3, \quad |t| \leqq 6, \end{cases}$$

and

$$y(t) = \begin{cases} (6-|t|)[1+.5\cos\left(\pi t/3\right)] + 9/(2\pi)\sin\left(\pi|t|/3\right), & |t| \leqq 6, \\ 0, & |t| > 6. \end{cases}$$

The solution is

$$x(s) = \begin{cases} 1 + \cos\left(\pi s/3\right), & |s| \leqq 3, \\ 0, & |s| > 3. \end{cases}$$

We approximate this problem by

$$Kx = y + e$$

where $y_i = y(t_i)$ and the t_i are midpoints of 78 intervals of equal length in the interval $[-6, 6]$, and the quadrature is performed by the trapezoidal rule with 48 intervals of equal length on $[-3, 3]$. Then K is a matrix of dimension 78×49 and $r_0 = 0$. The diagonal elements of the matrix S were taken to be $s_{ii} = 10^{-4} y_i$ and μ was taken to be 9.792, corresponding to $\alpha = .9999$. The a priori information is that the x vector is nonnegative, and the vectors w are chosen to give nonoverlapping three point averages of the x function using weights $(1/4, 1/2, 1/4)$ for each triplet of adjacent components of the x vector.

The initial interval was determined in two ways:

(1) The "naive interval" is

$$0 \leqq x_j \leqq \min_{1 \leqq i \leqq m} \frac{y_i + \mu s_i}{K_{ij}}, \qquad j = 1, \cdots, n.$$

For matrices with nonnegative components, this defines a box containing the intersection of the ellipsoid $\{x: \|Kx - y\|_S \leqq \mu\}$ with the positive orthant [7].

(2) The "FERDIT interval" is the interval determined by one iteration of the algorithm described in [7]. This is a method for determining suboptimal confidence intervals by minimizing or maximizing $w^T x$ for x· in an ellipsoidal region containing the intersection of $\{x: \|Kx - y\|_S \leqq \mu\}$ and the naive interval above.

We show the results of several experiments. Figure 2 shows the confidence intervals obtained by standard methods, without the inequality constraints. The lengths are of order 10^4, reflecting the 10^5 condition number of the rank 42 matrix K, although the true solution is of order 1. Figure 3 shows confidence intervals obtained by requiring that the solution be symmetric in its variable. This corresponds to adding equality constraints and reduces the condition number to 10^3. The confidence intervals now are of order 10^3. Even the naive bounds give more information than this: the solution
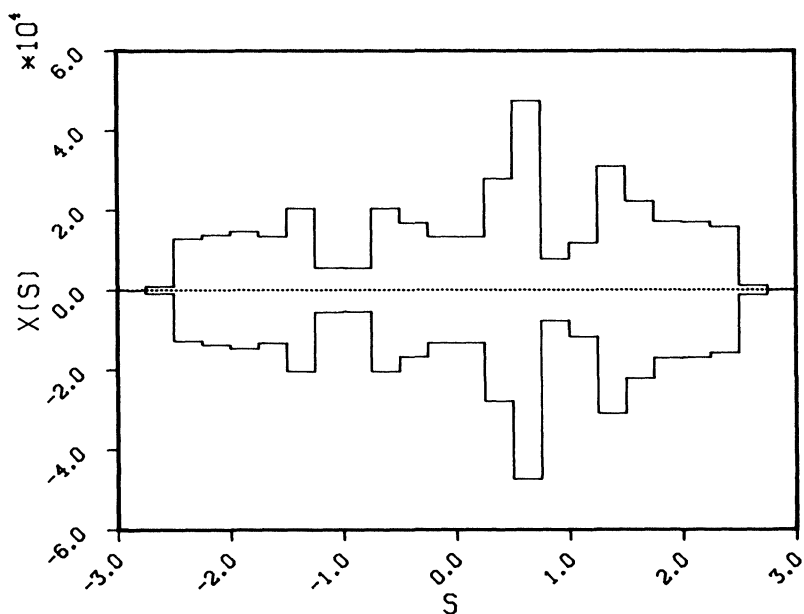
FIG. 2. *Standard confidence intervals for the* 78 × 49 *Phillips problem. The dotted line plots sample values of the true solution to the continuous problem.*
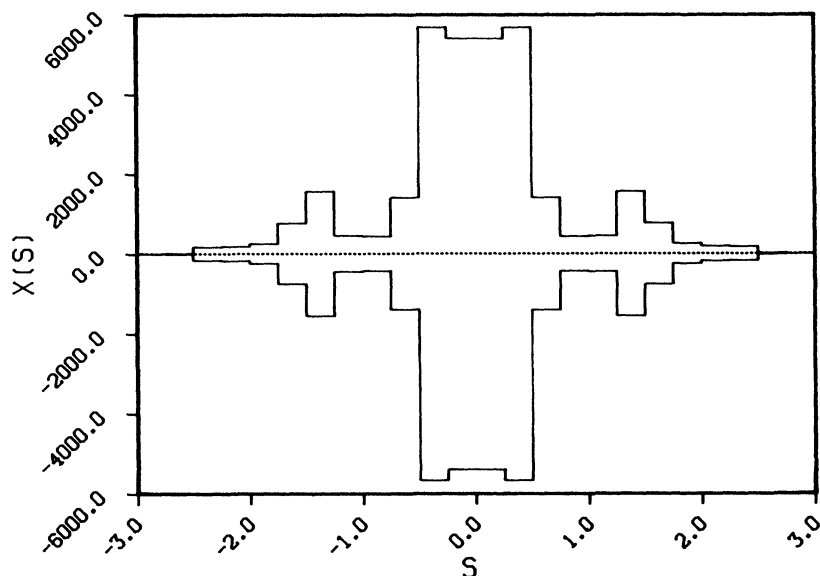

FIG. 3. *Standard confidence intervals for the* 78 × 49 *Phillips problem, using the constraint that x is symmetric around the point s* = 0. *The true solution is represented by the dotted line.*

is bounded by 0 and 35. Figure 4 gives the estimated 99.99% confidence bounds using the inequality constraints. The intervals now have a maximum length of less than 2. Figure 5 presents the bounds obtained using the inequality constraints and the symmetry condition.
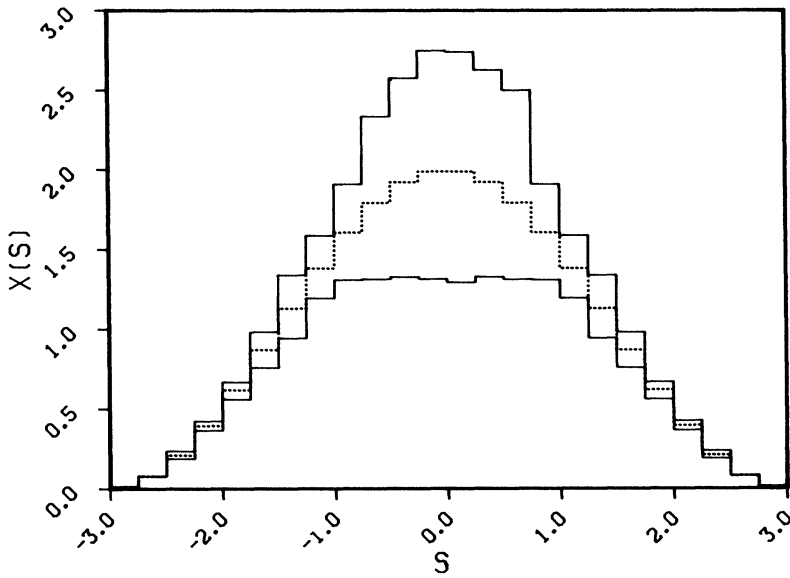
FIG. 4. *Confidence intervals for the* 78 × 49 *Phillips problem, using the constraint that x is nonnegative. The true solution is represented by the dotted line.*
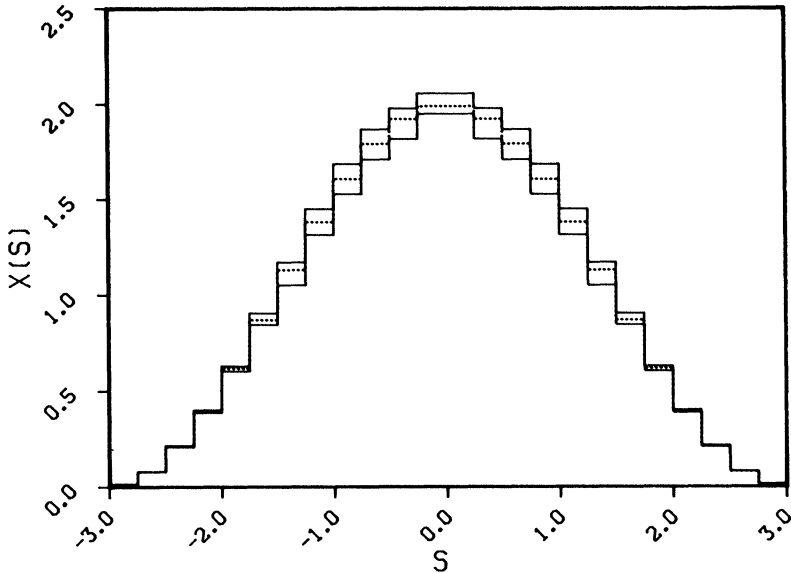

FIG. 5. *Confidence intervals for the* 78 × 49 *Phillips problem, using the constraints that x is nonnegative and symmetric around the point s* = 0. *The true solution is represented by the dotted line.*

Table 1 gives the performance statistics for various algorithms used to compute the results shown in Fig. 4. The runs were performed in FORTRAN single precision on a UNIVAC 1100/82. The table indicates the number of digits accuracy requested in the ϕ bounds, the run time (including FERDIT, if used), and the "basis changes", i.e., the number of changes of the index set I in Algorithm LS. The FERDIT iteration

TABLE 1
Algorithm performance on example 1.

Algorithm	Digits accuracy	Time (min.)	Basis changes
Brent–NNLS			
naive interval	4	25.0	53,314
Brent–LS			
naive interval	4	22.9	18,704
naive interval	3	13.9	10,993
Bracket–LS			
naive interval	4	8.9	6,220
naive interval	3	8.8	6,194
FERDIT interval	4	4.8	2,081
FERDIT interval	3	4.6	1,982

took approximately .6 minutes on this problem and gave upper bounds within a factor of 3.2 of the true three point averages, but most lower bounds were 0. The best algorithm, Bracket–LS with FERDIT, was 5 times faster than Brent–NNLS. Most of the improvement comes from the use of Bracket rather than Brent, and the use of initial interval information from FERDIT.

On other test problems, FERDIT continued to prove useful. If the initial bounds were close to optimal, FERDIT sometimes increased the cost of the algorithm, but on problems where the initial bounds were crude, it often led to large savings.

Example 2. We now consider the spectrum unfolding problem which arises in radiation physics. The integral equation is

$$\int_{E_{\text{LO}}}^{E_{\text{UP}}} K_i(E)x(E)\,dE = y_i + \varepsilon_i, \qquad i = 1, 2, \cdots, m,$$

where $x(E)$ is an unknown energy spectrum, y_i is the number of particles or photons counted in channel i of a multi-channel analyzer, and $K_i(E)$ is the spectrometer's energy response for channel i, (i.e. $K_i(E)\,dE$ is the probability that a particle or photon in the energy range $E \pm \frac{1}{2}\,dE$ will produce a pulse which gives a count in channel i). Figures 6a and 6b show two views of the response function for an NE-213 neutron spectrometer which has been described in detail by Verbinski et al. [21] and by Burrus and Verbinski [6]. The figures show a piecewise linear discretization K_{ij} to be used with the finite approximation equations

$$\sum_{j=1}^{n} K_{ij}x_j = y_i + \bar{\varepsilon}_i, \qquad i = 1, 2, \cdots, m,$$

where x_j is the total number of neutrons in the jth energy interval $(E_j \pm \frac{1}{2}\,dE_j)$. The values of m and n were taken to be 113 and 77 respectively, but in the interest of graphical clarity, the figure shows only every 3rd ordinate in each of the two abscissa directions. In order to show the structure for the higher energies, we have plotted the base 10 logarithm of $(1 + K)$ rather than the response function K itself. The mesh spacings for both energy and pulse height vary over their entire ranges, with narrower meshes being used for lower energies and pulse heights.

Ideally a response function should be a narrow, symmetric, sharply-peaked ridge centered along some linear relation between energy and pulse height. It is clear from the figures that the NE-213 spectrometer response function has none of these properties and that the measured pulse height spectrum will be a poor representation of the actual
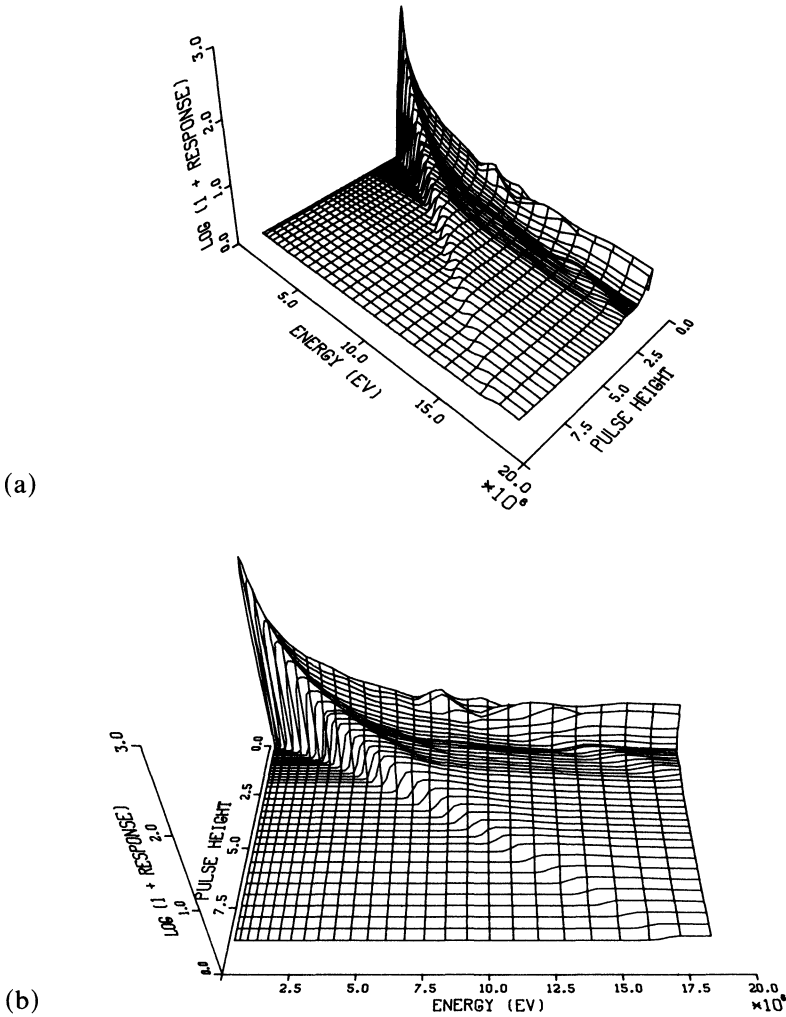
FIG. 6. *The instrument response function which provides data for the matrix K for the second example.*

energy spectrum. Figure 7 shows the plus and minus one standard deviation bounds for the measured spectrum of monoenergetic neutrons produced by the nuclear reaction $T(d, n)^4He$, (i.e. tritium nuclei bombarded with deuterons to produce helium nuclei and the neutrons whose spectrum was measured). It is assumed that there are no correlations between the numbers counted in separate channels and that the number in each channel is normally distributed. The variance matrix S is then an $m \times m$ diagonal matrix whose diagonal elements are one-half the widths shown in the figure. It is assumed that the standard deviations are known exactly so that chi-square statistics can be used rather than F-distribution statistics. In reality the standard deviations are not really known exactly but the estimates for most channels are very accurate because they are based on large numbers of counts in each channel. The standard deviation estimate for the number counted in each channel was the square root of the number counted (see Trumpler and Weaver [19, pp. 166–169] except that in channels containing only a few counts the estimates were chosen to be larger than this in order to assure that they were conservative. Figure 7 shows not the actual raw counts but rather a
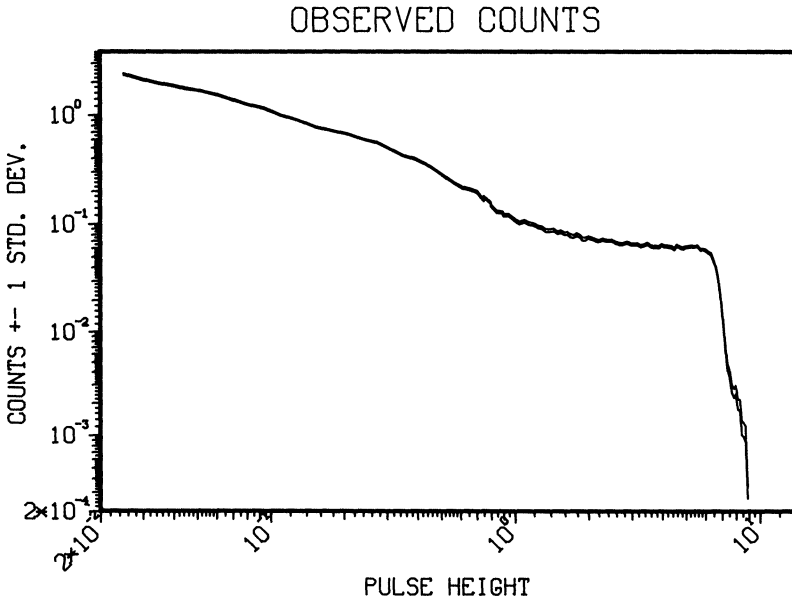
OBSERVED COUNTS

FIG. 7. *The instrument output, which provides data for the standard deviation matrix and the right-hand side function y for the second example.*

normalization which was required to correct for instrument gain effects that need not concern us here.

Since it is hopeless to try to estimate $x(E)$ at each energy E we seek instead to estimate averages of $x(E)$ over various energy ranges. A common practice is to estimate weighted averages with Gaussian weighting functions. Accordingly we seek estimates of the quantities

$$\phi_k(E) = \int_{-\infty}^{\infty} w_k(E)x(E) \, dE, \qquad k = 1, 2, \cdots, p,$$

where

$$w_k(E) = \frac{1}{\sigma_k\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma_k^2}(E - E_k)^2 \right], \qquad k = 1, 2, \cdots, p,$$

with the peak energies E_k chosen to cover the same energy range as the response functions and each peak width σ_k chosen to give the energy resolution desired in the neighborhood of E_k. The weighting functions $w_k(E)$ are called window functions and are tabulated at the same energy mesh points as the instrument response functions. The result is a set of window vectors

$$w_k = (w_{k1}, w_{k2}, \cdots, w_{kn})^T, \qquad k = 1, 2, \cdots, p,$$

and the "unfolded spectrum" is the set of p points (E_k, ϕ_k) where

$$\phi_k = \sum_{j=1}^{n} w_{kj}x_j = w_k^T x, \qquad k = 1, 2, \cdots, p.$$

Note that this procedure, in effect, replaces the real instrument response functions with the window functions so that the unfolded spectrum represents the measurements that would have been obtained using an instrument whose response functions were the latter rather than the former. The set of window vectors actually used is shown in

Fig. 8. The total number of window vectors was 105, with each vector being tabulated at $n = 77$ points. The windows widths σ_k were all chosen much smaller than the total energy range (0–20 Mev) so that most of the 77 elements in each vector were negligible and therefore set to zero. Note that the higher energy windows are wider than those at lower energies. The widths chosen reflect the experimentalist's estimate of the degree of energy resolution obtainable at each energy. All of the windows are normalized so that

$$\int_{-\infty}^{\infty} w_k(E)\, dE = 1.0, \qquad k = 1, 2, \cdots, p.$$

Note that we have again plotted $\log_{10}(1 + w_{kj})$ rather than the w_{kj} themselves.

To obtain an estimate of the "unfolded spectrum" we first computed the confidence intervals by the standard method. The results are shown in Fig. 9, which is a plot of
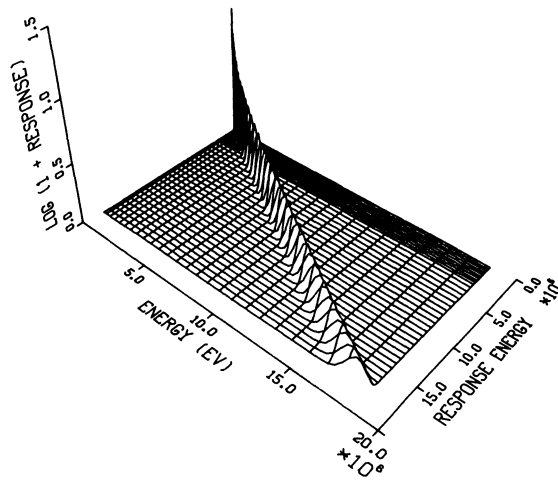

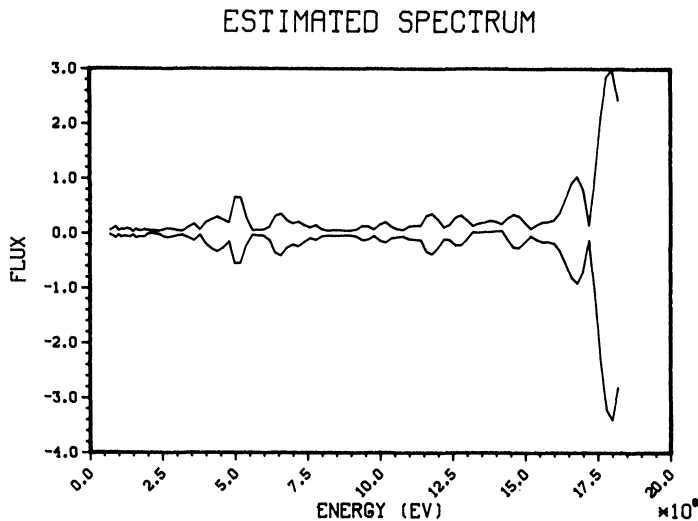FIG. 8. *The window vectors w for the second example.*

ESTIMATED SPECTRUM

FIG. 9. *Standard confidence intervals for the second example.*

the bounds ϕ_k^{LO} and ϕ_k^{UP} as a function of E_k, $k = 1, 2, \cdots, 105$. The value of μ^2 was chosen so that each of the intervals $[\phi_k^{LO}, \phi_k^{UP}]$ was an $\alpha = 99.99\%$ confidence interval. The value of $r_0 = \min_x \|Kx - y\|_S^2$ was 38.048, and we calculated γ^2 from the asymptotic formula

$$\gamma^2 = \tfrac{1}{2}[\kappa_\alpha + \sqrt{2N-1}]^2$$

where N is the number of degrees of freedom and κ_α is the percentage point of the cumulative normal distribution corresponding to confidence level α [4, p. 293]. For the present problem, the value of N should ideally be the rank of the response matrix but this is difficult to determine reliably, so we used the most conservative estimate $N = n = 77$. We used $\kappa_\alpha = 4.0$ which gives $\alpha = 0.9999$. The value of μ^2 was then computed by $\mu^2 = r_0 + \gamma^2$. In Fig. 9 the computed bounds ϕ_k^{LO} and ϕ_k^{UP} were joined by straight line segments to form an estimated uncertainty band.

    An experimentalist would expect a peak in the spectrum between 13 and 14 Mev, but there is no evidence of this in the standard confidence interval estimates. We applied Algorithm Bracket-LS to each of the window vectors, yielding the results shown in Fig. 10. The spectrum in Fig. 10 is dominated by a single peak centered at about 13.8 Mev and with a full width at half maximum of about 1.5 Mev. Part of this width (about 0.6 Mev) can be attributed to the inherent resolution limit of the instrument, but the remainder arises from the choice of the window function widths σ_k for the windows centered in the neighborhood of 14 Mev. The width of the peak can be reduced by choosing smaller window widths σ_k, but this procedure also produces wider confidence intervals $[\phi_k^{LO}, \phi_k^{UP}]$. Thus, in choosing window widths, it is necessary to balance statistical uncertainty against energy resolution. If the windows are too wide, details in the spectrum are smeared out and lost. If they are too small the widths of the confidence intervals become excessively large. Three or four digit accuracy can be achieved in 22.1 minutes using FERDIT and Bracket-LS, or in 26.7 minutes without FERDIT.
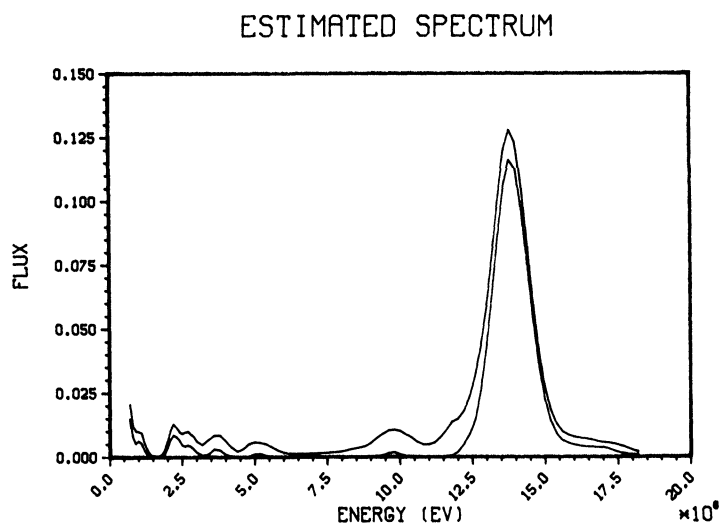
ESTIMATED SPECTRUM

FIG. 10. *Confidence intervals for the second example, using the constraint that x is nonnegative.*

REFERENCES

[1] A. M. AURELA AND J. TORSTI, *Stabilization of unstable resolution correction problems*, Ann. Acad. Sci. Fennical Series, A 255 (1967), pp. 1–17.

[2] C. T. H. BAKER, L. FOX, D. F. MAYERS AND K. WRIGHT, *Numerical solution of Fredholm integral equations of first kind*, Comput. J., 7 (1964), pp. 141–148.

[3] Y. BARD, *Nonlinear Parameter Estimation*, Academic Press, New York, 1974.

[4] W. H. BEYER, *Handbook of Tables for Probability and Statistics*, 2nd Edition, CRC Press, Boca Raton, FL, 1979.

[5] R. P. BRENT, *An algorithm with guaranteed convergence for finding a zero of a function*, Comput. J., 14 (1971), pp. 422–425.

[6] W. R. BURRUS AND V. V. VERBINSKI, *Fast-neutron spectroscopy with thick organic scintillators*, Nuclear Instruments and Methods, 67 (1969), pp. 181–196.

[7] J. E. COPE AND B. W. RUST, *Constrained least squares interval estimation*, this Journal, 6 (1985), pp. 670–683.

[8] M. FOSTER, *An application of the Wiener–Kolmogorov smoothing theory to matrix inversion*, J. Soc. Indust. Appl. Math, 9 (1961), pp. 387–392.

[9] R. J. HANSON, *Integral equations of immunology*, Comm. Assoc. Comput. Mach., 15 (1972), pp. 883–890.

[10] K. H. HASKELL AND R. J. HANSON, *An algorithm for linear least squares problems with equality and nonnegativity constraints*, Math. Programming, 21 (1981), pp. 98–118.

[11] C. L. LAWSON AND R. J. HANSON, *Solving Least Squares Problems*, Prentice-Hall, Englewood Cliffs, NJ, 1974.

[12] D. L. PHILLIPS, *A technique for the numerical solution of certain integral equations of the first kind*, J. Assoc. Comput. Mach., 9 (1962), pp. 84–97.

[13] C. RADHAKRISHNA RAO, *Linear Statistical Inference and Its Applications*, John Wiley, New York, 1973.

[14] J. REPLOGLE, B. D. HOLCOMB AND W. R. BURRUS, *The use of mathematical programming for solving singular and poorly conditioned systems of equations*, J. Mathematical Analysis and Applics., 20 (1967), pp. 310–324.

[15] B. W. RUST AND W. R. BURRUS, *Mathematical Programming and the Numerical Solution of Linear Equations*, American Elsevier, New York, 1972.

[16] O. STRAND AND E. R. WESTWATER, *Statistical estimation of the numerical solution of a Fredholm integral equation of the first kind*, J. Assoc. Comput. Mach., 15 (1968), pp. 100–114.

[17] A. N. TIKHONOV AND V. Y. ARSENIN, *Solutions of Ill-Posed Problems*, V. H. Winston, Washington DC, 1977.

[18] S. TWOMEY, *On the numerical solution of Fredholm integral equations of the first kind by the inversion of the linear systems produced by quadrature*, J. Assoc. Comput. Mach., 10 (1963), pp. 97–101.

[19] R. J. TRUMPLER AND H. F. WEAVER, *Statistical Astronomy*, Dover, New York 1962.

[20] J. M. VARAH, *A practical examination of some numerical methods for linear discrete ill-posed problems*, SIAM Rev., 21 (1979), pp. 100–111.

[21] V. V. VERBINSKI, W. R. BURRUS, T. A. LOVE, W. ZOBEL, N. W. HILL AND R. TEXTOR, *Calibration of an organic scintillator for neutron spectrometry*, Nuclear Instruments and Methods, 65 (1968), pp. 8–25.