

---

Computing and Applied Mathematics Laboratory

---



REFERENCE

NIST  
PUBLICATIONS

*CONFIDENCE INTERVALS FOR  
DISCRETE APPROXIMATIONS TO  
ILL-POSED PROBLEMS*

*Bert W. Rust and Dianne P. O'Leary*

*January 1992*

---

U.S. DEPARTMENT OF COMMERCE  
National Institute of Standards and Technology  
Gaithersburg, MD 20899

QC  
100  
.U56  
#4720  
1992



1992  
0-100  
186  
4720  
1992

# Confidence Intervals for Discrete Approximations to Ill-Posed Problems

**Bert W. Rust**  
**Dianne P. O'Leary**

U.S. DEPARTMENT OF COMMERCE  
National Institute of Standards  
and Technology  
Computing and Applied Mathematics Laboratory  
Applied and Computational  
Mathematics Division  
Gaithersburg, MD 20899

January 1992



**U.S. DEPARTMENT OF COMMERCE**  
**Rockwell A. Schnabel, Acting Secretary**  
**NATIONAL INSTITUTE OF STANDARDS**  
**AND TECHNOLOGY**  
**John W. Lyons, Director**



# CONFIDENCE INTERVALS FOR DISCRETE APPROXIMATIONS TO ILL-POSED PROBLEMS

Bert W. Rust and Dianne P. O’Leary  
Applied and Computational Mathematics Division  
National Institute of Standards and Technology  
Gaithersburg, MD 20899

## Abstract

Let the  $m \times n$  linear model

$$\hat{\mathbf{y}} = \mathbf{K} \bar{\mathbf{x}} + \hat{\mathbf{e}}$$

be obtained by discretizing a system of first-kind integral equations

$$\hat{y}_i = \int_a^b K_i(\xi) x(\xi) d\xi + \hat{e}_i, \quad i = 1, \dots, m,$$

with known functions  $K_i(\xi)$ , unknown function  $x(\xi)$ , and an  $m$ -vector  $\hat{\mathbf{y}}$  of measurements corrupted by random errors  $\hat{\mathbf{e}}$  drawn from a distribution with zero mean vector and known variance matrix. Consider the problem of estimating linear functions of the form

$$\phi = \mathbf{w}^T \bar{\mathbf{x}} \approx \int_a^b w(\xi) x(\xi) d\xi,$$

where  $w(\xi)$  is an averaging function designed to elicit some desired information about  $x(\xi)$ . For such problems, the least squares solution is a highly unstable function of the measurements, and the classical confidence intervals are too wide to be useful. The solution can often be stabilized by imposing physically motivated, *a priori* nonnegativity constraints on  $\bar{\mathbf{x}}$ . This paper will show how to extend the classical confidence interval estimation procedure to accommodate these nonnegativity constraints in order to obtain improved confidence intervals. The technique defines valid confidence intervals even for problems with  $m < n$ .

## Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Test Problems</b>	<b>4</b>
2.1	Problem I – An Overdetermined Problem . . . . .	5
2.2	Problem II – An Underdetermined Problem . . . . .	5
2.3	Window Functions and Window Vectors . . . . .	5
<b>3</b>	<b>Classical Point Estimation</b>	<b>7</b>
3.1	Full Rank Problems – The Best Linear Unbiased Estimator . . . . .	8
3.2	Underdetermined Problems . . . . .	8
<b>4</b>	<b>Classical One-at-a-Time Confidence Intervals</b>	<b>10</b>
4.1	Confidence Intervals for Normally Distributed Errors . . . . .	12
4.2	Confidence Intervals From Chebyshev’s Inequality . . . . .	12
<b>5</b>	<b>The Geometry of Confidence Interval Estimation</b>	<b>14</b>
5.1	Full Rank Problems . . . . .	14
5.2	Underdetermined Problems . . . . .	15

6	Classical Simultaneous Confidence Intervals	15
6.1	Simultaneous Intervals for Normally Distributed Errors . . . . .	16
6.2	Simultaneous Intervals from Chebyshev's Inequality . . . . .	17
7	Nonnegatively Constrained Estimation	17
7.1	Constrained Point Estimation . . . . .	19
7.2	Constrained Interval Estimation . . . . .	19
7.3	Simultaneous Confidence Intervals . . . . .	21
8	Constrained One-at-a-Time Confidence Intervals	23
8.1	The Burrus Conjecture . . . . .	23
8.2	The Duality Theorem . . . . .	23
8.3	Lower-biased and Upper-biased Estimators . . . . .	27
8.4	Proof of the Burrus Conjecture . . . . .	28
8.5	Examples . . . . .	30
9	A Real-World Problem	30

## 1. Introduction

Consider the linear model

$$\hat{y} = K\bar{x} + \hat{\epsilon} , \quad (1.1)$$

where  $\hat{y}$  is a measured  $m$ -vector,  $K$  is a known  $m \times n$  matrix,  $\bar{x}$  is an unknown  $n$ -vector, and  $\hat{\epsilon}$  is a random  $m$ -vector with zero mean and known positive definite variance matrix  $S^2$ , i.e.,

$$\mathcal{E}(\hat{\epsilon}) = \mathbf{0} , \quad \mathcal{E}(\hat{\epsilon}\hat{\epsilon}^T) = S^2 . \quad (1.2)$$

We are particularly interested in the case when this model arises in discretizing a system of first kind integral equations

$$\hat{y}_i = \int_a^b K_i(\xi)x(\xi) d\xi + \hat{\epsilon}_i , \quad i = 1, \dots, m . \quad (1.3)$$

We assume that  $n$  is chosen large enough to assure that the discretization errors are negligible in comparison with the  $\hat{\epsilon}_i$ .

Without severe *a priori* restrictions on  $x(\xi)$ , the  $m$  discrete relationships (1.3) cannot determine the value of  $x(\xi)$  at every point on the interval  $[a, b]$ . Therefore we seek instead to estimate a finite set of average values of  $x(\xi)$ , averaged on various subintervals of  $[a, b]$ . More precisely, we choose  $p$  representative points  $\xi_1 < \xi_2 < \dots < \xi_p$  in  $[a, b]$  and for each  $\xi_k$  choose a corresponding *window function*  $w_k(s)$  designed so that

$$\phi_k(\xi_k) = \int_a^b w_k(\xi)x(\xi) d\xi \quad (1.4)$$

is an average value of the unknown function in a subinterval around  $\xi_k$ . Discretizing these averaging integrals by the same method used to discretize (1.3) gives a set of  $p$  linear functions

$$\bar{\phi}_k = w_k^T \bar{x} , \quad k = 1, 2, \dots, p , \quad (1.5)$$

which approximate the desired averages. We are again assuming that the discretization errors are negligible in comparison to the  $\hat{\epsilon}_i$ . Note that the individual components of  $\bar{x}$  can be specified by choosing  $p = n$  and

$$w_k = e_k , \quad k = 1, 2, \dots, n , \quad (1.6)$$

where  $e_k$  is the  $k$ th column of the identity matrix.



Having accomplished the discretization, the task becomes one of estimating each of the  $p$  linear functions (1.5) subject to the statistical constraints (1.1) and (1.2). If  $\text{rank}(\mathbf{K}) = n$ , then these are classical linear regression problems with the well known best linear unbiased (least squares) estimates

$$\hat{\phi}_k = \mathbf{w}_k^T \hat{\mathbf{x}} \quad , \quad \hat{\mathbf{x}} = (\mathbf{K}^T \mathbf{S}^{-2} \mathbf{K})^{-1} \mathbf{K}^T \mathbf{S}^{-2} \hat{\mathbf{y}} \quad . \quad (1.7)$$

When each of these *point estimates* is associated with the corresponding subinterval about  $\xi_k$ , the resulting set of pairs comprises a discrete approximation to the unknown function  $x(\xi)$ . The reliability of such an approximation can be assessed by computing confidence interval estimates  $[\phi_k^{lo}, \phi_k^{up}]$  for each of the  $\hat{\phi}_k$ . There are two kinds of confidence intervals that can be considered [8, Chapt. 6]:

**one-at-a-time confidence intervals** : Each  $\hat{\phi}_k$  is treated individually, and for a pre-specified probability  $\alpha$  ( $0 < \alpha < 1$ ), the  $100\alpha\%$  confidence intervals are determined separately. Each individual interval contains the corresponding true value with probability  $\alpha$ , i.e.,

$$\Pr \{ \phi_k^{lo} \leq \mathbf{w}_k^T \bar{\mathbf{x}} \leq \phi_k^{up} \} = \alpha \quad , \quad k = 1, 2, \dots, p \quad . \quad (1.8)$$

Repeating the measurements a large number of times  $M$ , and calculating the full set of  $p$  intervals each time, would give roughly  $\alpha M p$  intervals containing the corresponding true values and roughly  $(1 - \alpha) M p$  failing to contain the true value.

**simultaneous confidence intervals** : All of the  $\hat{\phi}_k$  are treated together, and the intervals are constructed so that, with probability  $\alpha$ , they all contain the corresponding true values, i.e.,

$$\Pr \{ \phi_k^{lo} \leq \bar{\phi}_k \leq \phi_k^{up} \quad , \quad k = 1, 2, \dots, p \} \geq \alpha \quad . \quad (1.9)$$

Repeating the measurements  $M$  times, and calculating a full set of  $p$  confidence intervals each time, would give roughly  $\alpha M$  sets in which all of the intervals contain the corresponding true values, and roughly  $(1 - \alpha) M$  sets containing at least one interval which fails to contain the corresponding true value. It is clear that this is a more exacting requirement than that for one-at-a-time intervals, so it is not surprising that, for a given  $\alpha$ , each of the simultaneous intervals is wider than its one-at-a-time counterpart.

The inversion of first kind integral equations is an ill-posed problem, so the matrix  $\mathbf{K}$  is usually ill-conditioned or even rank deficient. As a result the confidence intervals are typically extremely wide or even unbounded. To get physically plausible solutions, it is necessary to impose *a priori* constraints on the solution. In many real-world problems the solution is known to be nonnegative, and good algorithms for computing nonnegatively constrained point estimates have been widely available for some time [10]. In previous work [12] we considered the problem of computing nonnegatively constrained, simultaneous confidence intervals when  $\hat{\epsilon}$  is normally distributed. We showed that the confidence intervals can be computed by solving certain constrained optimization problems and presented an algorithm for solving them. In this paper we extend those results to include nonnegatively constrained one-at-a-time intervals.

In §2 we present two example problems used throughout the paper to illustrate our results. In §3 we briefly review classical point estimation theory with an emphasis on the effects of ill-conditioning and rank-deficiency. In §4 we briefly review classical, one-at-a-time confidence intervals, and in §5 we consider the geometrical basis for interval estimation in order to characterize the bounds as solutions to certain constrained optimization problems which will later be augmented to include the nonnegativity constraints. In §6 we briefly review the estimation of classical, simultaneous confidence intervals, whose geometrical basis is identical to that for the one-at-a-time intervals. The purpose of the discussion in these initial sections is to note the deficiencies of the classical bounds and point estimates when used for discretizations of ill-posed problems. In §7 we formally append the constraints  $\bar{\mathbf{x}} \geq \mathbf{o}$  to the classical linear regression model and demonstrate their effectiveness in stabilizing the point estimates. We also show that these constraints almost always produce bounded confidence intervals, even for underdetermined problems, and show how to extend the theory for simultaneous intervals to accommodate them. The extension of the classical theory for one-at-a-time intervals is considerably more difficult, but in §8 we prove a conjecture, first made some 20 years ago by Walter R. Burrus [17], which allows the calculation of such intervals using the same algorithm

we originally developed for simultaneous intervals. Finally, a real-world example involving the measurement of a gamma-ray spectrum is presented in §9. Since we assume that the only significant errors are the random perturbations  $\hat{\epsilon}$  in the measured data, we implicitly assume that the model is correct. In particular, this implies that errors due to discretization of the integrals are negligible. *Thus we produce confidence intervals corresponding to the class of functions for which the quadrature is sufficiently accurate (or the instrument resolution sufficiently fine).*

## 2. Test Problems

To generate problems with known solutions, we use the integral equation studied by Phillips [14]

$$y(t) = \int_{-3}^3 K(t, \xi) x(\xi) d\xi, \quad -6 \leq t \leq 6 \quad (2.1)$$

where

$$K(t, \xi) = \begin{cases} 1 + \cos \left[ \frac{\pi(\xi-t)}{3} \right] & , \quad |\xi - t| \leq 3, |t| \leq 6, \\ 0 & , \quad \text{otherwise,} \end{cases} \quad (2.2)$$

and

$$y(t) = \begin{cases} (6 - |t|) \left[ 1 + \frac{1}{2} \cos \left( \frac{\pi t}{3} \right) \right] + \frac{9}{2\pi} \sin \left( \frac{\pi |t|}{3} \right) & , \quad |t| \leq 6, \\ 0 & , \quad \text{otherwise.} \end{cases} \quad (2.3)$$

The kernel  $K(t, \xi)$  is nonnegative, a property common in many real-world applications. The function  $y(t)$  is also nonnegative, symmetric about  $t = 0$ , and bell-shaped, with maximum value  $y(0) = 9$  and minimum value  $y(-6) = 0 = y(+6)$ . The exact solution is

$$x(\xi) = \begin{cases} 1 + \cos \left( \frac{\pi \xi}{3} \right) & , \quad |\xi| \leq 3, \\ 0 & , \quad \text{otherwise.} \end{cases} \quad (2.4)$$

This nonnegative function also defines a symmetric bell-shaped curve with a maximum  $x(0) = 2$  and minima  $x(-3) = 0 = x(+3)$ .

To get a system of integral equations, we chose  $m$  mesh points

$$-6 < t_1 < t_2 < t_3 < \dots < t_{m-1} < t_m < +6 \quad (2.5)$$

and define  $K_i(\xi) \equiv K(t_i, \xi)$ . This gave

$$y_i \equiv y(t_i) = \int_{-3}^3 K_i(\xi) x(\xi) d\xi, \quad i = 1, 2, \dots, m. \quad (2.6)$$

We used the trapezoidal rule on the mesh

$$\xi_1 = -3, \xi_2 = -2.95, \xi_3 = -2.90, \dots, \xi_{61} = 0, \dots, \xi_{121} = 3. \quad (2.7)$$

to reduce this system of integral equations to

$$\mathbf{y} \cong \mathbf{K} \bar{\mathbf{x}}, \quad (2.8)$$

where the  $y_i$  are calculated from (2.3) and the  $\bar{x}_j = x(\xi_j)$  are calculated from (2.4). In order to assure that every row of the matrix  $\mathbf{K}$  subtends at least one quadrature panel, we chose the  $t$ -mesh to be equally spaced on the interval  $[-5.925, 5.925]$ :

$$t_i = -5.925 + \frac{11.85}{m-1}(i-1), \quad i = 1, 2, \dots, m. \quad (2.9)$$

In order to obtain test problems in which the quadrature errors were completely negligible relative to the statistical errors, we did not use the  $\mathbf{y}$ -vector computed from (2.3), but rather used  $\bar{\mathbf{y}}$  defined by

$$\bar{\mathbf{y}} \equiv \mathbf{K} \bar{\mathbf{x}}. \quad (2.10)$$



The calculation of  $\bar{y}$  was double precision, with results rounded to single precision on a machine with  $\epsilon_{mach} \cong 7 \times 10^{-15}$ . The vector  $\bar{x}$  can be regarded as the exact quantity to be estimated, and the vector  $\bar{y}$  can be regarded as the measurements obtained without error.

To get the random “measuring” errors  $\hat{\epsilon}_i$ , we let  $s_i = (10^{-6})\bar{y}_i$ ,  $i = 1, 2, \dots, m$ , and picked sample vectors  $\hat{\epsilon}$  from a multivariate normal distribution with independently distributed elements, i.e.,

$$\hat{\epsilon} = (\hat{\epsilon}_1, \hat{\epsilon}_2, \dots, \hat{\epsilon}_m)^T \sim N(\mathbf{o}, \mathbf{S}^2) , \quad (2.11)$$

where  $\mathbf{S}^2 = \text{diag}(s_1^2, s_2^2, \dots, s_m^2)$ . Adding the random errors chosen in this manner to the system (2.10) gave

$$\hat{y} = \mathbf{K}\bar{x} + \hat{\epsilon} , \quad \hat{\epsilon} \sim N(\mathbf{o}, \mathbf{S}^2) \quad (2.12)$$

where  $\hat{y}$  is the vector of “measured” values. Since the calculation of the  $\bar{y}_i$  assures that (2.10) is satisfied to 14 digits, the above standard deviations are all from 6 to 8 orders of magnitude greater than the corresponding truncation errors that arose in forming  $\bar{y}$ . Thus the random errors were essentially the only errors in the problem.

### 2.1. Problem I – An Overdetermined Problem

For the first test problem we chose  $m = 150$  and  $n = 121$ , producing a matrix  $\mathbf{K}$  of rank  $n$ . For statistical studies, we generated several such problems by randomly choosing different error vectors  $\hat{\epsilon}$ , but picked a typical one for use in this paper. The least squares solution for this one is shown in the top of Figure 1, together with the exact solution  $x(s)$ . The maximum and minimum singular values of the scaled matrix  $\mathbf{S}^{-1}\mathbf{K}$  are  $\sigma_1 \cong 3.3950 \times 10^9$  and  $\sigma_{121} \cong 1.1610$  so the condition number is  $\text{cond}(\mathbf{S}^{-1}\mathbf{K}) \cong 2.924 \times 10^9$ . Since  $\epsilon_{mach} \cong 7 \times 10^{-15}$ , we could reasonably expect to compute  $\hat{x}$  accurate to 6 digits. This means that all of the variation of the  $\hat{x}(\xi_j)$  about the true solution  $x(\xi)$  can be attributed to the measurement errors  $\hat{\epsilon}_i$ . This variation is surprisingly small in view of the large condition number of the matrix, but is still distressingly large when viewed as relative error in the solution. Many of the  $\hat{x}_i$  are negative even though the exact solution  $x(\xi)$  is everywhere nonnegative.

### 2.2. Problem II – An Underdetermined Problem

Although the matrix in Problem I has a large condition number, the least squares problem is better than most that arise from first kind integral equations. Many problems have rank deficient or even underdetermined matrices. If the kernel is accessible, then it is always possible to choose  $n < m$ , but this may produce discretization errors larger than the measuring errors. Even if discretization error is acceptably small, choosing  $n < m$  may still generate a rank-deficient problem. Many real-world problems are quite naturally underdetermined, with  $m$  and  $n$  being fixed by hardware considerations or other physical constraints. Often the analyst is given not the kernel functions  $K_i(\xi)$ , but rather the matrix  $\mathbf{K}$  with  $m < n$ . To simulate such problems we discretized the Phillips equation with  $m = 108$  and  $n = 121$ . The matrix  $\mathbf{K}$  has rank 108, and the maximum and minimum non-zero singular values of  $\mathbf{S}^{-1}\mathbf{K}$  are  $\sigma_1 \cong 3.3728 \times 10^9$  and  $\sigma_{108} \cong 0.14635$ . The bottom frame of Figure 1 gives a plot of the generalized inverse (minimal length least squares) solution

$$\hat{x} = (\mathbf{S}^{-1}\mathbf{K})^\dagger \mathbf{S}^{-1}\hat{y} , \quad (2.13)$$

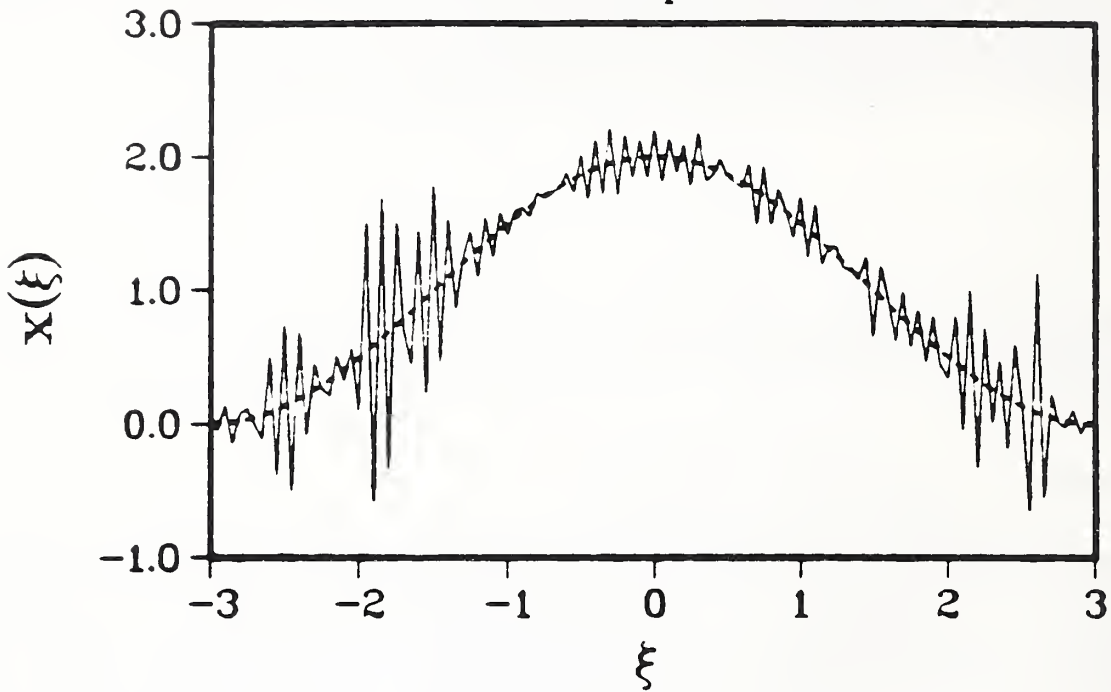
which oscillates about the true solution with wider variations than those obtained for Problem I. This higher noise level is due mostly to the loss of information in using fewer observations. Of course the estimate is not unbiased, but the oscillations tend to be centered on the true curve, so the bias is evidently small relative to the random scatter.

### 2.3. Window Functions and Window Vectors

The solutions presented in the preceding section were highly oscillatory because of the ill-conditioning of the regression model. These oscillations can be damped to some extent by seeking estimates for a set of average values of the function  $x(\xi)$  as in (1.4). The simplest set of window functions is

$$w_k(\xi) = \delta(\xi - \xi_k) , \quad k = 1, 2, \dots, n , \quad (2.14)$$

150 X 121 Phillips Problem  
Classical Least Squares Solution



108 X 121 Phillips Problem  
Generalized Inverse Solution

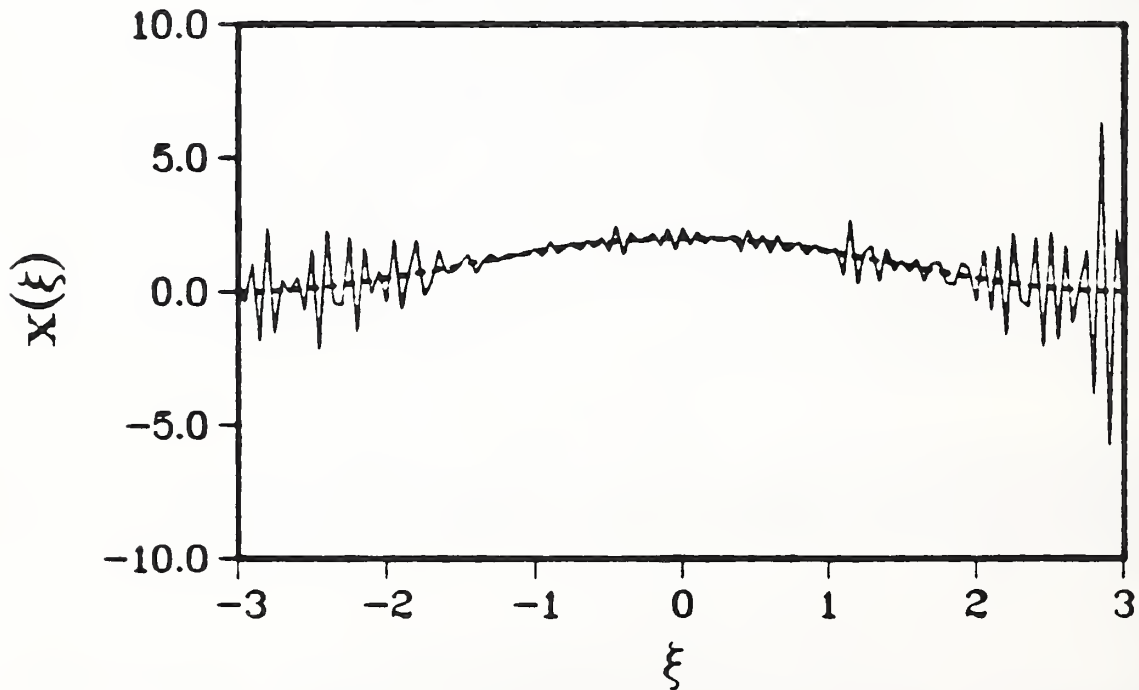


Figure 1: Classical solutions. Dashed lines are exact solutions  $x(\xi)$  and solid lines are estimates  $\hat{x}_k = e_k^T \hat{x}$ . Top: Problem I, least squares estimate. Bottom: Problem II, generalized inverse estimate.

where  $\delta(\xi - \xi_k)$  is the Dirac delta function centered at the point  $\xi_k$ . In this case the corresponding *window responses* are

$$\bar{\phi}_k = \int_a^b \delta(\xi - \xi_k) x(\xi) d\xi = \bar{x}_k , \quad (2.15)$$

and the corresponding window vectors (1.6) are just the columns of the  $n$ -th order identity matrix. No actual averaging is involved and the estimates  $(\xi_k, \hat{\phi}_k)$  correspond to those computed in the previous section.

The simplest set of real averages is defined by

$$\bar{\phi}_k = \frac{1}{\Delta\xi} \int_{\xi_k}^{\xi_{k+1}} x(\xi) d\xi , \quad k = 1, 2, \dots, n-1 , \quad (2.16)$$

where  $\Delta\xi = \xi_{k+1} - \xi_k$ . Each of these non-overlapping averages is combined with its corresponding  $\xi$ -interval to give

$$\mathcal{P}_2 = \{ ( [\xi_k, \xi_{k+1}] , \bar{\phi}_k ) : k = 1, 2, \dots, n-1 \} , \quad (2.17)$$

a histogram approximation to  $x(\xi)$ . Using the trapezoidal rule to discretize the integrals (2.16) gives the window vectors

$$\begin{aligned} \mathbf{w}_1 &= ( 0.5 \quad 0.5 \quad 0 \quad 0 \quad 0 \quad \dots \quad 0 )^T \\ \mathbf{w}_2 &= ( 0 \quad 0.5 \quad 0.5 \quad 0 \quad 0 \quad \dots \quad 0 )^T \\ \mathbf{w}_3 &= ( 0 \quad 0 \quad 0.5 \quad 0.5 \quad 0 \quad \dots \quad 0 )^T , \\ &\vdots \\ \mathbf{w}_{n-1} &= ( 0 \quad 0 \quad 0 \quad \dots \quad 0 \quad 0.5 \quad 0.5 )^T \end{aligned} \quad (2.18)$$

so we call this process *2-point averaging*.

Averaging can be extended to include more than one quadrature panel in each average, ending ultimately with *n-point averaging* which would seek to estimate only the average value of  $x(\xi)$  on the whole interval  $[a, b]$ . In general,  $\nu$ -point averaging uses window functions

$$w_k(\xi) \equiv \begin{cases} \frac{1}{(\nu-1)\Delta\xi} & , \quad \xi_{(k-1)(\nu-1)+1} \leq \xi \leq \xi_{k(\nu-1)+1} \\ 0 & , \quad \text{otherwise} \end{cases} , \quad k = 1, 2, \dots, p , \quad (2.19)$$

where  $p$  is the largest integer less than or equal to  $(n-1)/(\nu-1)$ . The typical window vector has the form

$$\mathbf{w}_k = \left( 0, \dots, 0, \frac{1}{2(\nu-1)}, \frac{1}{\nu-1}, \frac{1}{\nu-1}, \dots, \frac{1}{\nu-1}, \frac{1}{2(\nu-1)}, 0, \dots, 0 \right)^T , \quad (2.20)$$

with exactly  $\nu$  non-zero elements beginning with element number  $(k-1)(\nu-1)+1$ . The corresponding histogram approximation,

$$\mathcal{P}_\nu = \{ ( [\xi_{(k-1)(\nu-1)+1}, \xi_{k(\nu-1)+1}] , \bar{\phi}_k ) : k = 1, 2, \dots, p \} , \quad (2.21)$$

will cover the whole interval  $[a, b]$  only if  $(n-1)$  is an exact multiple of  $(\nu-1)$ . Otherwise, there will be one or more (but fewer than  $\nu$ ) quadrature panels that will not be included in the set of averages.

### 3. Classical Point Estimation

Consider the problem of estimating the linear function

$$\bar{\phi} = \mathbf{w}^T \bar{\mathbf{x}} \quad (3.1)$$

subject to the statistical constraints

$$\hat{\mathbf{y}} = \mathbf{K} \bar{\mathbf{x}} + \hat{\boldsymbol{\epsilon}} , \quad \mathcal{E}(\hat{\boldsymbol{\epsilon}}) = \mathbf{0} , \quad \mathcal{E}(\hat{\boldsymbol{\epsilon}} \hat{\boldsymbol{\epsilon}}^T) = \mathbf{S}^2 . \quad (3.2)$$

A *linear estimator*

$$\phi = \mathbf{u}^T \hat{\mathbf{y}} , \quad (3.3)$$

with mean and variance

$$\mathcal{E}(\phi) = \mathbf{u}^T \mathbf{K} \bar{\mathbf{x}} , \quad \text{Var}(\phi) = \mathbf{u}^T \mathbf{S}^2 \mathbf{u} . \quad (3.4)$$

is *unbiased* if  $\mathcal{E}(\phi) = \bar{\phi} = \mathbf{w}^T \bar{\mathbf{x}}$  identically in  $\bar{\mathbf{x}}$  [18, Chapt. 1]. Thus  $\phi$  is unbiased if and only if

$$\mathbf{u}^T \mathbf{K} = \mathbf{w}^T , \quad (3.5)$$

which means the vector  $\mathbf{w}$  is a linear combination of the rows of  $\mathbf{K}$ .

### 3.1. Full Rank Problems – The Best Linear Unbiased Estimator

If  $\text{rank}(\mathbf{K}) = n \leq m$ , an unbiased estimator exists for any linear function  $\mathbf{w}^T \bar{\mathbf{x}}$ . In fact, there will be many unbiased estimators

$$\mathbf{u}^T \hat{\mathbf{y}} = \left[ \mathbf{w}^T \mathbf{K}^\dagger + \mathbf{z}^T \left( \mathbf{I}_m - \mathbf{K} \mathbf{K}^\dagger \right) \right] \hat{\mathbf{y}} , \quad (3.6)$$

where  $\mathbf{K}^\dagger$  is the generalized inverse of  $\mathbf{K}$  and  $\mathbf{z}$  is any  $m$ -vector. The *best linear unbiased estimator* is the one that minimizes the variance, i.e., the one corresponding to the vector  $\hat{\mathbf{u}}$  that solves the constrained optimization problem

$$\text{Var}(\hat{\mathbf{u}}^T \hat{\mathbf{y}}) = \min_{\hat{\mathbf{u}}} \{ \hat{\mathbf{u}}^T \mathbf{S}^2 \hat{\mathbf{u}} \mid \hat{\mathbf{u}}^T \mathbf{K} = \mathbf{w}^T \} . \quad (3.7)$$

It is easy to see that the solution vector is

$$\hat{\mathbf{u}} = \mathbf{S}^{-2} \mathbf{K} (\mathbf{K}^T \mathbf{S}^{-2} \mathbf{K})^{-1} \mathbf{w} , \quad (3.8)$$

so the best linear unbiased estimator is

$$\hat{\phi} = \hat{\mathbf{u}}^T \hat{\mathbf{y}} = \mathbf{w}^T (\mathbf{K}^T \mathbf{S}^{-2} \mathbf{K})^{-1} \mathbf{K}^T \mathbf{S}^{-2} \hat{\mathbf{y}} . \quad (3.9)$$

This estimator can also be written

$$\hat{\phi} = \mathbf{w}^T \hat{\mathbf{x}} , \quad (3.10)$$

where

$$\hat{\mathbf{x}} = (\mathbf{K}^T \mathbf{S}^{-2} \mathbf{K})^{-1} \mathbf{K}^T \mathbf{S}^{-2} \hat{\mathbf{y}} \quad (3.11)$$

is the solution vector for the weighted least squares problem

$$\hat{\mathbf{r}}^2 = \min_{\hat{\mathbf{x}}} \{ (\hat{\mathbf{y}} - \mathbf{K} \hat{\mathbf{x}})^T \mathbf{S}^{-2} (\hat{\mathbf{y}} - \mathbf{K} \hat{\mathbf{x}}) \} . \quad (3.12)$$

It follows from (3.4) and (3.8) that the minimum variance is

$$\text{Var}(\hat{\phi}) = \hat{\mathbf{u}}^T \mathbf{S}^2 \hat{\mathbf{u}} = \mathbf{w}^T (\mathbf{K}^T \mathbf{S}^{-2} \mathbf{K})^{-1} \mathbf{w} . \quad (3.13)$$

As an example, consider Problem I. If the set of window vectors is taken to be the columns of  $\mathbf{I}_n$ , then the  $\hat{\phi}_k$  are just the elements of the least squares solution  $\hat{\mathbf{x}}$  as in the top frame of Figure 1. The estimates corresponding to the window vectors (2.20) for 2-point averaging are shown in the top frame of Figure 2, and the estimates for 3-point averaging are shown in the bottom frame. The noise level has been reduced by averaging, but this improvement is achieved only at the expense of a reduction in resolution for the independent variable  $\xi$ . Note also that the averaging does not prevent the occurrence of some negative estimates.

### 3.2. Underdetermined Problems

If  $\mathbf{K}$  has less than full column rank, then the set of all least squares solutions can be written

$$\mathbf{x}(z) = (\mathbf{S}^{-1} \mathbf{K})^\dagger \mathbf{S}^{-1} \hat{\mathbf{y}} + \left[ \mathbf{I}_n - (\mathbf{S}^{-1} \mathbf{K})^\dagger (\mathbf{S}^{-1} \mathbf{K}) \right] z , \quad (3.14)$$

where  $z$  is an arbitrary  $n$ -vector. An unbiased estimator for  $\mathbf{w}^T \bar{\mathbf{x}}$  exists only if  $\mathbf{w}^T$  is an exact linear combination of the rows of  $\mathbf{K}$ , in which case the best linear unbiased estimate is given by the  $\mathbf{u}$ -vector

$$\hat{\mathbf{u}} = \mathbf{S}^{-1} \left[ (\mathbf{S}^{-1} \mathbf{K})^\dagger \right]^T \mathbf{w} , \quad (3.15)$$

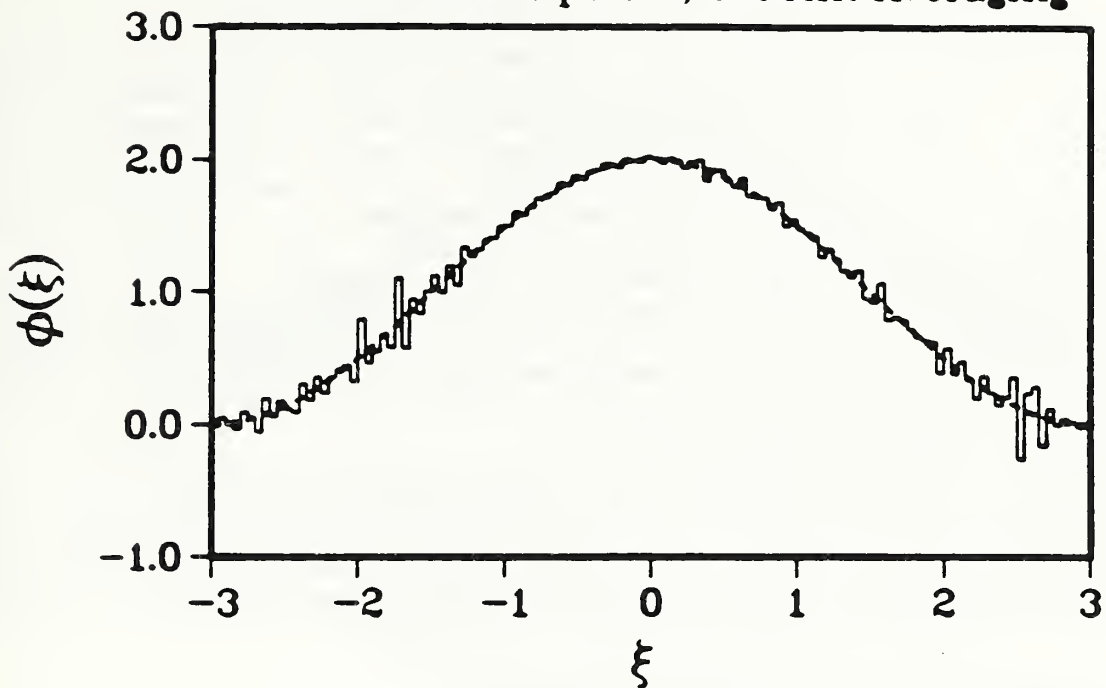
so

$$\hat{\phi} = \hat{\mathbf{u}}^T \hat{\mathbf{y}} = \mathbf{w}^T (\mathbf{S}^{-1} \mathbf{K})^\dagger \mathbf{S}^{-1} \hat{\mathbf{y}} = \mathbf{w}^T \hat{\mathbf{x}} , \quad (3.16)$$

where  $\hat{\mathbf{x}}$  is the minimal length solution (2.13). The variance of this estimator is

$$\text{Var}(\hat{\phi}) = \hat{\mathbf{u}}^T \mathbf{S}^2 \hat{\mathbf{u}} = \mathbf{w}^T (\mathbf{K}^T \mathbf{S}^{-2} \mathbf{K})^\dagger \mathbf{w} . \quad (3.17)$$

150 X 121 Phillips Problem  
 Classical Least Squares, 2-Point Averaging



150 X 121 Phillips Problem  
 Classical Least Squares, 3-Point Averaging

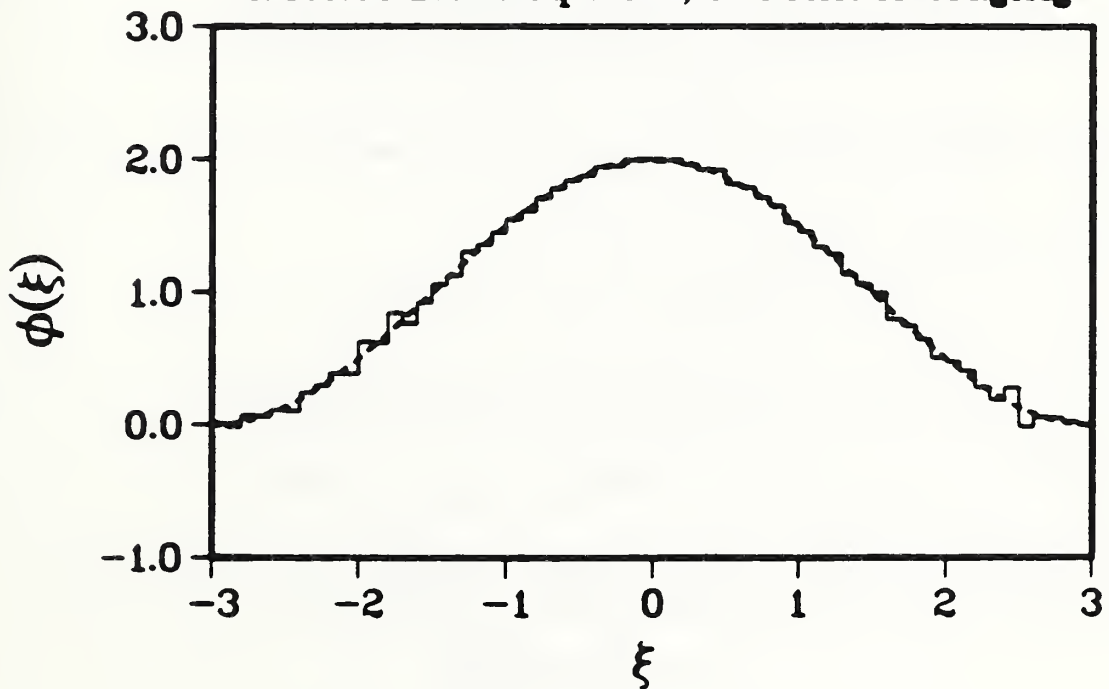


Figure 2: Best linear unbiased estimates for Problem I. Dashed lines are exact solutions  $x(\xi)$  and solid lines are histogram estimates  $\hat{\phi}_k$ .

Top: 2-point averages over intervals  $[\xi_k, \xi_{k+1}]$ . Bottom : 3-point averages over intervals  $[\xi_{2k-1}, \xi_{2k+1}]$ .



Being limited to window vectors expressible as linear combinations of the rows of  $\mathbf{K}$  places severe restrictions on the information that can be elicited about the unknown function  $\mathbf{x}(\xi)$ . This restriction is imposed in the method of Backus and Gilbert [19, Chapt. 7] that constructs just such a set of window vectors, called *averaging kernels*. Ideally, each averaging kernel should resemble a narrow Gaussian centered on the point  $\xi_k$ . In practice the rows of  $\mathbf{K}$  rarely admit good resemblances to Gaussians, so the estimates  $\hat{\phi}_k$  are usually non-symmetric weighting functions that change from one  $\xi_k$  to the next. The resulting set of point estimates  $(\xi_k, \hat{\phi}_k)$  comprise a *smoothed* discrete approximation to the unknown function  $\mathbf{x}(\xi)$ . This smoothing is difficult to characterize and non-uniform over the range of  $\xi$ .

If  $\mathbf{w}$  is not expressible as a linear combination of rows of  $\mathbf{K}$ , then the average  $\mathbf{w}^T \bar{\mathbf{x}}$  is said to be *inestimable*. This does not mean that no unbiased estimate exists, but rather that there exists no unbiased estimator of the form  $\mathbf{u}^T \hat{\mathbf{y}}$ . In theory, there exist unbiased estimates of the form  $\mathbf{w}^T \mathbf{x}(\bar{\mathbf{z}})$  where  $\mathbf{x}(\bar{\mathbf{z}})$  is some solution (3.14) of the underdetermined least squares problem. In reality, it is necessary to accept the fact that computable estimates will be biased and to try to assure that the errors due to bias are small in comparison to the random variations introduced by the  $\hat{\mathbf{e}}$ .

In Section 2 we presented the generalized inverse solution for Problem II. It is apparent from the bottom frame of Figure 1 that any bias in this approximation is smaller than the scatter due to the random errors. In Figure 3 we give the histogram approximations corresponding to the 2-point averaging windows and the 3-point averaging windows. Even though the averaging considerably reduces the scatter due to the random errors, it is still not possible to detect the bias. It would seem that the bias is not so large that it rules out the possibility of obtaining useful information about the function  $\mathbf{x}(\xi)$ .

#### 4. Classical One-at-a-Time Confidence Intervals

For any unbiased estimator  $\phi = \mathbf{u}^T \hat{\mathbf{y}}$ , one-at-a-time confidence intervals for  $\bar{\phi} = \mathbf{w}^T \bar{\mathbf{x}}$  can be constructed from the transformed random variable

$$\eta = \frac{\phi - \bar{\phi}}{\sqrt{\text{Var}(\phi)}} = \frac{\mathbf{u}^T \hat{\mathbf{y}} - \mathbf{w}^T \bar{\mathbf{x}}}{\sqrt{\mathbf{u}^T \mathbf{S}^2 \mathbf{u}}} , \quad (4.1)$$

where the latter equality follows from (3.4). It also follows from (3.4) and (3.5) that

$$\mathcal{E}(\eta) = 0 \quad , \quad \text{Var}(\eta) = 1 \quad . \quad (4.2)$$

For a given probability  $\alpha$ , a  $100\alpha\%$  confidence interval for  $\bar{\phi}$  can be constructed by determining a number  $\kappa$  such that

$$\Pr \{ -\kappa \leq \eta \leq +\kappa \} = \alpha \quad . \quad (4.3)$$

Substituting (4.1) and rearranging gives

$$\Pr \left\{ \left[ \phi - \kappa \sqrt{\text{Var}(\phi)} \right] \leq \bar{\phi} \leq \left[ \phi + \kappa \sqrt{\text{Var}(\phi)} \right] \right\} = \alpha \quad , \quad (4.4)$$

or

$$\Pr \left\{ \left( \mathbf{u}^T \hat{\mathbf{y}} - \kappa \sqrt{\mathbf{u}^T \mathbf{S}^2 \mathbf{u}} \right) \leq \mathbf{w}^T \bar{\mathbf{x}} \leq \left( \mathbf{u}^T \hat{\mathbf{y}} + \kappa \sqrt{\mathbf{u}^T \mathbf{S}^2 \mathbf{u}} \right) \right\} = \alpha \quad . \quad (4.5)$$

Thus, the interval

$$\mathcal{I}(\alpha; \kappa) = \left[ \left( \mathbf{u}^T \hat{\mathbf{y}} - \kappa \sqrt{\mathbf{u}^T \mathbf{S}^2 \mathbf{u}} \right) \quad , \quad \left( \mathbf{u}^T \hat{\mathbf{y}} + \kappa \sqrt{\mathbf{u}^T \mathbf{S}^2 \mathbf{u}} \right) \right] \quad (4.6)$$

is a  $100\alpha\%$  confidence interval for  $\bar{\phi}$ .

Given the value  $\kappa$ , the foregoing results are valid for any unbiased estimator  $\mathbf{u}^T \hat{\mathbf{y}}$ . The best linear unbiased estimator (3.9) gives the confidence interval  $\hat{\mathcal{I}}(\alpha; \kappa) = [\hat{\phi}^{lo}, \hat{\phi}^{up}]$ , where

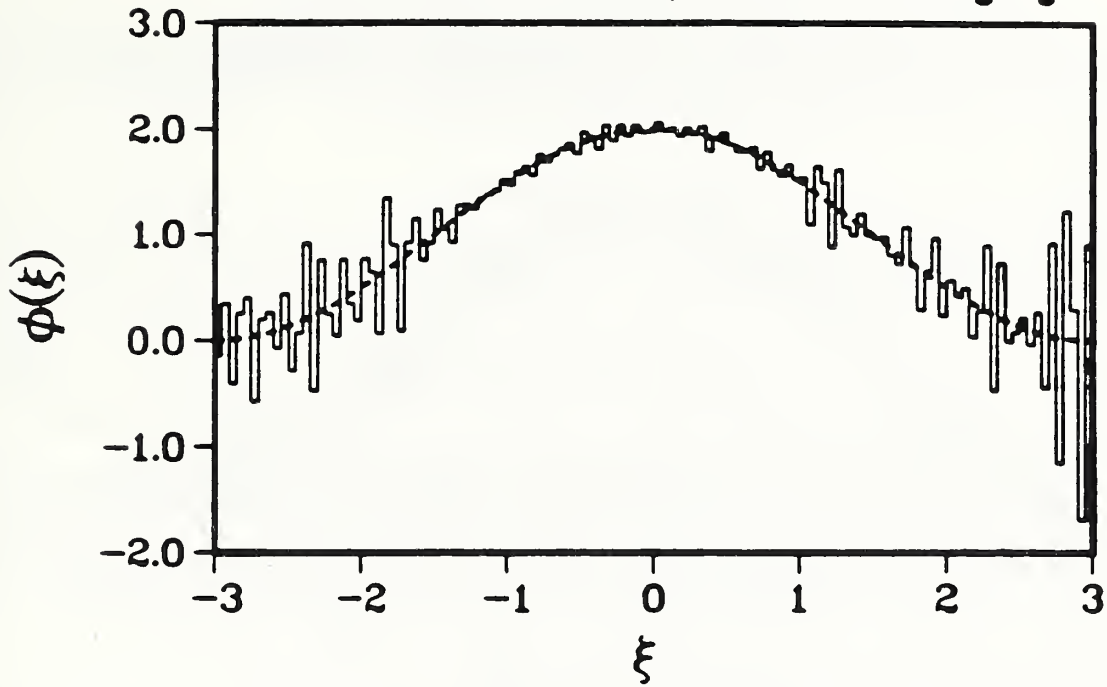
$$\begin{aligned} \hat{\phi}^{lo} &= \hat{\mathbf{u}}^T \hat{\mathbf{y}} - \kappa \sqrt{\hat{\mathbf{u}}^T \mathbf{S}^2 \hat{\mathbf{u}}} \quad , \\ \hat{\phi}^{up} &= \hat{\mathbf{u}}^T \hat{\mathbf{y}} + \kappa \sqrt{\hat{\mathbf{u}}^T \mathbf{S}^2 \hat{\mathbf{u}}} \quad . \end{aligned} \quad (4.7)$$

It is clear from Eqs. (3.8), (3.9) and (3.10) that these bounds can also be written

$$\begin{aligned} \hat{\phi}^{lo} &= \mathbf{w}^T \hat{\mathbf{x}} - \kappa \sqrt{\mathbf{w}^T (\mathbf{K}^T \mathbf{S}^{-2} \mathbf{K})^{-1} \mathbf{w}} \quad , \\ \hat{\phi}^{up} &= \mathbf{w}^T \hat{\mathbf{x}} + \kappa \sqrt{\mathbf{w}^T (\mathbf{K}^T \mathbf{S}^{-2} \mathbf{K})^{-1} \mathbf{w}} \quad , \end{aligned} \quad (4.8)$$



108 X 121 Phillips Problem  
Gen. Inverse Solution, 2-Point Averaging



108 X 121 Phillips Problem  
Gen. Inverse Solution, 3-Point Averaging

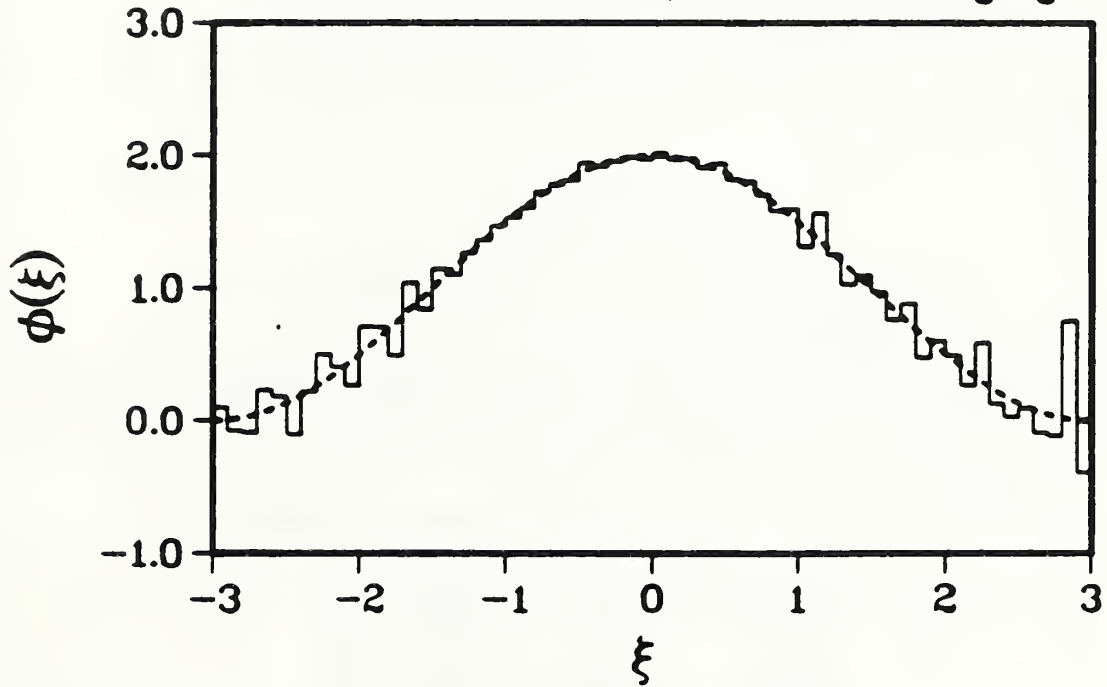


Figure 3: Generalized inverse estimates for Problem II. Dashed lines are the exact solution  $x(\xi)$  and solid lines are histogram estimates  $\hat{\phi}_k$ .  
Top: 2-point averages over intervals  $[\xi_k, \xi_{k+1}]$ . Bottom: 3-point averages over intervals  $[\xi_{2k-1}, \xi_{2k+1}]$ .

where  $\hat{\mathbf{x}}$  is the least squares solution vector (3.11).

The width of the confidence interval (4.6) is directly proportional to the value of  $\kappa$ . Optimally narrow confidence intervals can be computed only if the probability density function for  $\eta$  is known, and this requires a knowledge of the joint probability density function for the vector  $\hat{\mathbf{e}}$ .

#### 4.1. Confidence Intervals for Normally Distributed Errors

In many applications,  $\hat{\mathbf{e}}$  is known (or assumed) to have a multivariate normal distribution,  $\hat{\mathbf{e}} \sim N(\mathbf{0}, \mathbf{S}^2)$ , so  $\eta$  follows the standard normal distribution,  $\eta \sim n(0, 1)$ . This means that for any  $\alpha$  ( $0 < \alpha < 1$ ), it is possible to find a corresponding value  $\kappa$  ( $0 < \kappa < \infty$ ) satisfying

$$\frac{1}{\sqrt{2\pi}} \int_{-\kappa}^{+\kappa} \exp\left(-\frac{\eta^2}{2}\right) d\eta = \alpha, \quad (4.9)$$

which defines the relation between the  $\kappa$  and the  $\alpha$  required in (4.4).

As an example, consider Problem I and the best linear unbiased estimators for the 2-point averaging window vectors (2.18). Using  $\kappa = 1.960$ , for 95% intervals, Equations (4.8) give the confidence bounds plotted as solid lines in the top frame of Figure 4. There are 120 window vectors in the whole set, so we expect approximately 6 of the confidence intervals to fail to bracket the corresponding true averages. In this particular problem there were 9 bracket failures.

#### 4.2. Confidence Intervals From Chebyshev's Inequality

If the joint probability density function for  $\hat{\mathbf{e}}$  is not known, then valid, suboptimal confidence intervals can be constructed by using *Chebyshev's inequality* [9, Chapt. 1], which guarantees that, for any  $\kappa > 0$ ,

$$\Pr \{ |\eta - \mathcal{E}(\eta)| \geq \kappa \text{Var}(\eta) \} \leq \frac{1}{\kappa^2}. \quad (4.10)$$

It follows then that, for any unbiased estimator  $\mathbf{u}^T \hat{\mathbf{y}}$ ,

$$\Pr \left\{ \left| \mathbf{u}^T \hat{\mathbf{y}} - \mathbf{w}^T \bar{\mathbf{x}} \right| < \kappa \sqrt{\mathbf{u}^T \mathbf{S}^2 \mathbf{u}} \right\} \geq 1 - \frac{1}{\kappa^2}. \quad (4.11)$$

Therefore, for any  $\alpha$  ( $0 < \alpha < 1$ ), if

$$\kappa = +\sqrt{\frac{1}{1-\alpha}}, \quad (4.12)$$

then

$$\Pr \left\{ \left( \mathbf{u}^T \hat{\mathbf{y}} - \kappa \sqrt{\mathbf{u}^T \mathbf{S}^2 \mathbf{u}} \right) < \mathbf{w}^T \bar{\mathbf{x}} < \left( \mathbf{u}^T \hat{\mathbf{y}} + \kappa \sqrt{\mathbf{u}^T \mathbf{S}^2 \mathbf{u}} \right) \right\} \geq \alpha, \quad (4.13)$$

so the interval (4.6) is a  $100\alpha\%$  confidence interval for  $\bar{\phi}$ .

Confidence intervals calculated from Chebyshev's inequality are very conservative since they must be wide enough to accommodate any possible probability density for  $\eta$ . Table 1 compares the sizes of Chebyshev intervals and normal distribution intervals for some commonly used confidence levels. The ratio is the factor

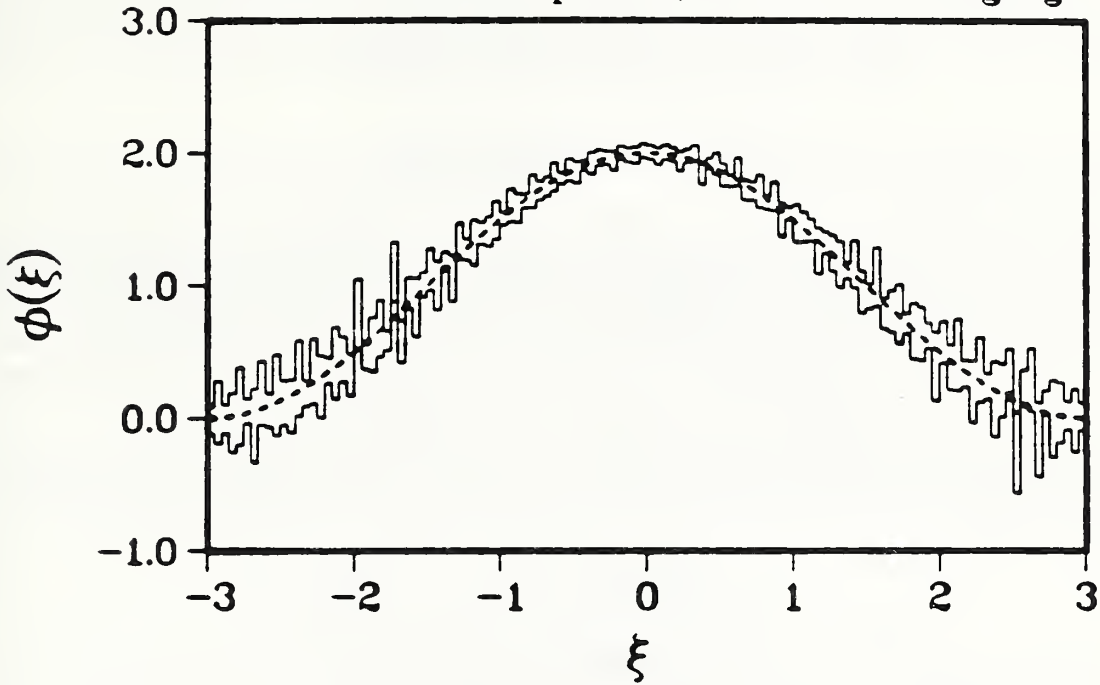
Table 1:  $\kappa$ -values for Chebyshev inequality and standard normal distribution

$\alpha$	$\kappa$ -Chebyshev	$\kappa$ -normal	ratio
0.6667	1.732	0.967	1.79
0.95	4.472	1.960	2.28
0.99	10.000	2.575	3.88
0.999	31.622	3.295	9.60

by which the interval width must be expanded if the form of the error distribution is not known.

As an example, consider again Problem I and the best linear unbiased estimators for the 2-point averaging window vectors. The bottom frame of Figure 4 is a plot of the 95% one-at-a-time confidence interval

150 X 121 Phillips Problem  
 One-at-a-time Confidence Intervals  
 Using Normal Distribution,  $\alpha = 0.95$   
 Classical Least Squares, 2-Point Averaging



150 X 121 Phillips Problem  
 One-at-a-time Confidence Intervals  
 Using Chebyshev Theorem,  $\alpha = 0.95$   
 Classical Least Squares, 2-Point Averaging

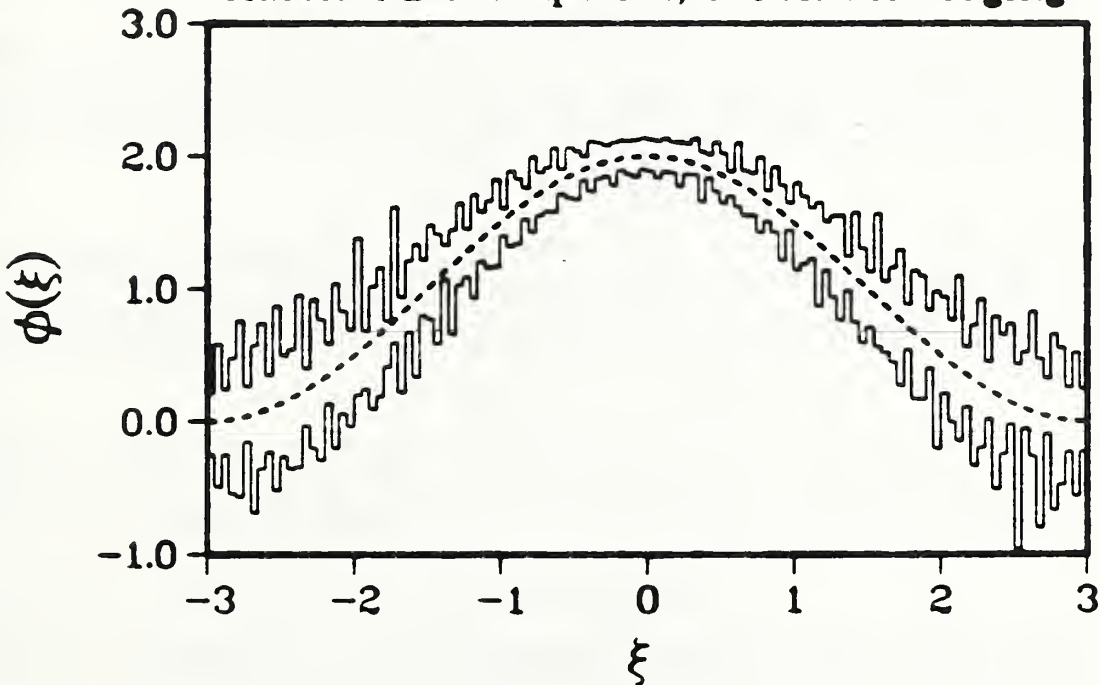


Figure 4: Classical 95% one-at-a-time confidence intervals for Problem I. Dashed lines are exact solution  $x(\xi)$ , and solid lines are histograms of estimated lower and upper bounds. Top: 2-point averages from normality assumption. Bottom: 2-point averages from Chebyshev's Inequality.

bounds obtained from Chebyshev's inequality for those averages. The solid lines are histograms of the upper and lower bounds calculated from (4.8) using  $\kappa = 4.472$ . These intervals are larger by a factor of  $4.472/1.960 \approx 2.28$  than the corresponding intervals (in the top frame) obtained by assuming normally distributed errors. The dashed line is the true solution. Note that none of the true values of the averages lies outside the corresponding confidence interval.

## 5. The Geometry of Confidence Interval Estimation

### 5.1. Full Rank Problems

No matter what the relation between  $\alpha$  and  $\kappa$ , the corresponding confidence interval bounds for the best linear unbiased estimator (3.9) can be calculated from either Eq. (4.7) or (4.8). It is not difficult to show that the end-points of this interval are also defined by the two constrained estimation problems:

$$\begin{aligned}\hat{\phi}^{lo} &= \min_{\mathbf{x}} \{ \mathbf{w}^T \mathbf{x} \mid (\hat{\mathbf{y}} - \mathbf{K} \mathbf{x})^T \mathbf{S}^{-2} (\hat{\mathbf{y}} - \mathbf{K} \mathbf{x}) = \hat{r}^2 + \kappa^2 \} , \\ \hat{\phi}^{up} &= \max_{\mathbf{x}} \{ \mathbf{w}^T \mathbf{x} \mid (\hat{\mathbf{y}} - \mathbf{K} \mathbf{x})^T \mathbf{S}^{-2} (\hat{\mathbf{y}} - \mathbf{K} \mathbf{x}) = \hat{r}^2 + \kappa^2 \} ,\end{aligned}\quad (5.1)$$

where  $\hat{r}^2$  is the minimum sum of squares (3.12). These problems can be solved by Lagrange multipliers to give solution vectors

$$\hat{\mathbf{x}}^{lo} = \hat{\mathbf{x}} - \frac{\kappa}{\sqrt{\mathbf{w}^T (\mathbf{K}^T \mathbf{S}^{-2} \mathbf{K})^{-1} \mathbf{w}}} (\mathbf{K}^T \mathbf{S}^{-2} \mathbf{K})^{-1} \mathbf{w} , \quad (5.2)$$

$$\hat{\mathbf{x}}^{up} = \hat{\mathbf{x}} + \frac{\kappa}{\sqrt{\mathbf{w}^T (\mathbf{K}^T \mathbf{S}^{-2} \mathbf{K})^{-1} \mathbf{w}}} (\mathbf{K}^T \mathbf{S}^{-2} \mathbf{K})^{-1} \mathbf{w} , \quad (5.3)$$

and values for  $\hat{\phi}^{lo}$  and  $\hat{\phi}^{up}$  that are the same as the values given by (4.8).

The common constraint region for problems (5.1), i.e.,

$$S(\kappa) = \{ \mathbf{x} \mid (\hat{\mathbf{y}} - \mathbf{K} \mathbf{x})^T \mathbf{S}^{-2} (\hat{\mathbf{y}} - \mathbf{K} \mathbf{x}) = \hat{r}^2 + \kappa^2 \} , \quad (5.4)$$

can also be written [17, Chapt. 2]

$$S(\kappa) = \left\{ \mathbf{x} \mid (\mathbf{x} - \hat{\mathbf{x}})^T \mathbf{K}^T \mathbf{S}^{-2} \mathbf{K} (\mathbf{x} - \hat{\mathbf{x}}) = \kappa^2 \right\} , \quad (5.5)$$

which defines the surface of an ellipsoid centered on the least squares solution  $\hat{\mathbf{x}}$ . The size of this ellipsoid varies with  $\kappa$ . For any value of the parameter  $\phi$ , the set of points

$$\mathcal{H}(\mathbf{w}, \phi) = \{ \mathbf{x} \mid \mathbf{w}^T \mathbf{x} = \phi \} \quad (5.6)$$

forms a hyperplane orthogonal to the vector  $\mathbf{w}$ . Therefore, the bounds  $\hat{\phi}^{lo}$  and  $\hat{\phi}^{up}$  are the values of  $\phi$  on the two tangential support planes of the ellipsoid  $S(\kappa)$  that are orthogonal to  $\mathbf{w}$ .

The location of the center of the ellipsoid  $S(\kappa)$  depends on  $\hat{\mathbf{y}}$ , and its size is scaled by the value of  $\kappa$ , but its orientation and shape are completely determined by the matrix  $\mathbf{K}^T \mathbf{S}^{-2} \mathbf{K}$ . Let the singular value decomposition of  $\mathbf{S}^{-1} \mathbf{K}$  be

$$\mathbf{S}^{-1} \mathbf{K} = \mathbf{U} \begin{pmatrix} \boldsymbol{\Sigma} \\ \mathbf{O} \end{pmatrix} \mathbf{V}^T , \quad (5.7)$$

with

$$\mathbf{U}^T \mathbf{U} = \mathbf{I}_m = \mathbf{U} \mathbf{U}^T , \quad \mathbf{V}^T \mathbf{V} = \mathbf{I}_n = \mathbf{V} \mathbf{V}^T , \quad \boldsymbol{\Sigma} = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_n) , \quad (5.8)$$

with  $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n$ . It follows then that

$$\mathbf{K}^T \mathbf{S}^{-2} \mathbf{K} = \mathbf{V} \boldsymbol{\Sigma}^2 \mathbf{V}^T . \quad (5.9)$$

It is easy to see [17, Chapt. 2] that the mutually orthogonal vectors  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$  define the directions of the major axes of  $S(\kappa)$ , and the lengths of those axes are given by

$$\ell_i = \frac{2\kappa}{\sigma_i} , \quad i = 1, 2, \dots, n . \quad (5.10)$$

From (4.8), (5.9) and the definition of  $\Sigma$ , it follows that for any window vector  $\mathbf{w}$ , the corresponding  $\alpha$ -level confidence interval has width

$$\hat{\phi}^{up} - \hat{\phi}^{lo} = 2\kappa \left[ \sum_{j=1}^n \frac{1}{\sigma_j^2} (\mathbf{v}_j^T \mathbf{w})^2 \right]^{\frac{1}{2}} = \left[ \sum_{j=1}^n \ell_j^2 (\mathbf{w}^T \mathbf{v}_j)^2 \right]^{\frac{1}{2}} . \quad (5.11)$$

Since  $\mathbf{v}_j$  is a unit vector (i.e.,  $\mathbf{v}_j^T \mathbf{v}_j = 1$ ), the scalar  $\mathbf{w}^T \mathbf{v}_j$  is just the projection of  $\mathbf{w}$  on the  $j$ th major axis of the ellipsoid, and  $\ell_j$  is the length of that major axis.

Linear regression models obtained from first kind integral equations are almost always poorly conditioned. The major axes of  $\mathcal{S}(\kappa)$  corresponding to the smaller singular values are, therefore, usually greatly elongated. From (5.11) it follows that if  $\mathbf{w}$  has a non-zero projection on any of the singular vectors corresponding to the smaller singular values, then the confidence interval for  $\mathbf{w}^T \bar{\mathbf{x}}$  will be very wide. Almost every window vector designed to elicit information about the function  $\mathbf{x}(\xi)$  will have non-zero components in the directions of these elongated axes. In the extreme case, one or more of the singular values are zero and the corresponding axes of  $\mathcal{S}(\kappa)$  are infinite in length, yielding infinite intervals. When such a problem is read into a computer, the machine truncation errors almost always produce a full rank matrix on the machine, but there is no danger of obtaining deceptively good confidence intervals as a result. If  $\epsilon_M$  is the *machine epsilon*, then typically for a  $\bar{\phi}_k$  whose true value is of order unity, the classical estimates (4.7) give intervals like  $[-\mathcal{O}(\epsilon_M^{-1}), +\mathcal{O}(\epsilon_M^{-1})]$  which, though shorter than  $[-\infty, +\infty]$ , are nevertheless useless.

## 5.2. Underdetermined Problems

For problems with  $m < n$ , the matrix is exactly rank deficient with  $\sigma_{m+1} = \sigma_{m+2} = \dots = \sigma_n = 0$ . The ellipsoid  $\mathcal{S}(\kappa)$  has at least  $n - m$  infinitely long major axes, and its center is not a single point  $\hat{\mathbf{x}}$  but rather a coset of the subspace spanned by the singular vectors  $\mathbf{v}_{m+1}, \mathbf{v}_{m+2}, \dots, \mathbf{v}_n$ , i.e., the set of points defined by (3.14). In that expression, the set of points

$$\mathcal{N}(\mathbf{S}^{-1} \mathbf{K}) = \left\{ \left[ \mathbf{I}_n - (\mathbf{S}^{-1} \mathbf{K})^\dagger (\mathbf{S}^{-1} \mathbf{K}) \right] \mathbf{z} \mid \mathbf{z} \text{ arbitrary} \right\} \quad (5.12)$$

is the subspace, and

$$\hat{\mathbf{x}} = (\mathbf{S}^{-1} \mathbf{K})^\dagger \mathbf{S}^{-1} \hat{\mathbf{y}} \quad (5.13)$$

is the displacement of the coset from the origin. If  $\mathbf{w}$  is orthogonal to the subspace  $\mathcal{N}(\mathbf{S}^{-1} \mathbf{K})$ , then it is an exact linear combination of the rows of  $\mathbf{K}$ , so the best linear unbiased estimator for  $\mathbf{w}^T \bar{\mathbf{x}}$  is given by (3.16), and its variance is given by (3.17). The confidence interval bounds are

$$\begin{aligned} \hat{\phi}^{lo} &= \mathbf{w}^T \hat{\mathbf{x}} - \kappa \sqrt{\mathbf{w}^T (\mathbf{K}^T \mathbf{S}^{-2} \mathbf{K})^\dagger \mathbf{w}} , \\ \hat{\phi}^{up} &= \mathbf{w}^T \hat{\mathbf{x}} + \kappa \sqrt{\mathbf{w}^T (\mathbf{K}^T \mathbf{S}^{-2} \mathbf{K})^\dagger \mathbf{w}} , \end{aligned} \quad (5.14)$$

where  $\hat{\mathbf{x}}$  is the minimal length solution (5.13).

If  $\mathbf{w}$  has a non-zero projection on any of the infinite axes of  $\mathcal{S}(\kappa)$ , then  $\mathbf{w}^T \bar{\mathbf{x}}$  is said to be inestimable because there exists no unbiased estimator of the form  $\phi = \mathbf{u}^T \hat{\mathbf{y}}$ . The center coset does contain at least one point

$$\mathbf{x}(\bar{\mathbf{z}}) = (\mathbf{S}^{-1} \mathbf{K})^\dagger \mathbf{S}^{-1} \hat{\mathbf{y}} + \left[ \mathbf{I}_n - (\mathbf{S}^{-1} \mathbf{K})^\dagger (\mathbf{S}^{-1} \mathbf{K}) \right] \bar{\mathbf{z}} , \quad (5.15)$$

for which  $\mathbf{w}^T \mathbf{x}(\bar{\mathbf{z}})$  gives an unbiased estimate. Taking  $\bar{\mathbf{z}} = \bar{\mathbf{x}}$  would give such an estimate, but of course  $\bar{\mathbf{x}}$  is unknown. Even if it were possible to find some  $\bar{\mathbf{z}}$  which gives an unbiased estimate, the confidence interval for  $\mathbf{w}^T \bar{\mathbf{z}}$  would still be unbounded.

## 6. Classical Simultaneous Confidence Intervals

The method for estimating a set of  $\alpha$ -level simultaneous confidence intervals for a whole set of window vectors  $\{\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_p\}$  is based on an  $\alpha$ -level *confidence ellipsoid* for the unknown vector  $\bar{\mathbf{x}}$ . For any positive value  $\gamma_n^2$ , the region in  $\mathbf{x}$ -space defined by

$$\mathcal{S}(\gamma_n) = \left\{ \mathbf{x} \mid (\mathbf{x} - \hat{\mathbf{x}})^T \mathbf{K}^T \mathbf{S}^{-2} \mathbf{K} (\mathbf{x} - \hat{\mathbf{x}}) = \gamma_n^2 \right\} . \quad (6.1)$$



is an ellipsoid centered on the least squares solution vector  $\hat{\mathbf{x}}$  (3.11). It becomes an  $\alpha$ -level confidence ellipsoid for  $\bar{\mathbf{x}}$  if  $\gamma_n^2$  is determined so that

$$\Pr \left\{ (\mathbf{x} - \hat{\mathbf{x}})^T \mathbf{K}^T \mathbf{S}^{-2} \mathbf{K} (\mathbf{x} - \hat{\mathbf{x}}) \leq \gamma_n^2 \right\} = \alpha . \quad (6.2)$$

The required set of simultaneous confidence intervals is constructed from the support planes for this ellipsoid. In particular, the set of intervals  $\hat{\mathcal{L}}_k(\alpha; \gamma_n^2) = [\hat{\phi}_k^{lo}, \hat{\phi}_k^{up}]$ , defined by

$$\left. \begin{aligned} \hat{\phi}_k^{lo} &= \min_{\mathbf{x}} \left\{ \mathbf{w}_k^T \mathbf{x} \mid (\mathbf{x} - \hat{\mathbf{x}})^T \mathbf{K}^T \mathbf{S}^{-2} \mathbf{K} (\mathbf{x} - \hat{\mathbf{x}}) = \gamma_n^2 \right\} , \\ \hat{\phi}_k^{up} &= \max_{\mathbf{x}} \left\{ \mathbf{w}_k^T \mathbf{x} \mid (\mathbf{x} - \hat{\mathbf{x}})^T \mathbf{K}^T \mathbf{S}^{-2} \mathbf{K} (\mathbf{x} - \hat{\mathbf{x}}) = \gamma_n^2 \right\} \end{aligned} \right\} , \quad k = 1, 2, \dots, p , \quad (6.3)$$

satisfy

$$\Pr \left\{ \hat{\phi}_k^{lo} \leq \bar{\phi}_k \leq \hat{\phi}_k^{up} , \quad k = 1, 2, \dots, p \right\} \geq \alpha , \quad (6.4)$$

and this result holds for any number of averaging vectors. These simultaneous confidence intervals are defined by the support planes of an ellipsoid in the same way as are the one-at-a-time intervals given by (5.1). The only difference is in the size of the defining ellipsoid. It is easy to see that if  $\text{rank}(\mathbf{K}) = n \leq m$ , then the solutions to the constrained optimization problems (6.3) are

$$\left. \begin{aligned} \hat{\phi}_k^{lo} &= \mathbf{w}_k^T \hat{\mathbf{x}} - \gamma_n \sqrt{\mathbf{w}_k^T (\mathbf{K}^T \mathbf{S}^{-2} \mathbf{K})^{-1} \mathbf{w}_k} , \\ \hat{\phi}_k^{up} &= \mathbf{w}_k^T \hat{\mathbf{x}} + \gamma_n \sqrt{\mathbf{w}_k^T (\mathbf{K}^T \mathbf{S}^{-2} \mathbf{K})^{-1} \mathbf{w}_k} \end{aligned} \right\} , \quad k = 1, 2, \dots, p , \quad (6.5)$$

where  $\hat{\mathbf{x}}$  is the least squares solution vector (3.11).

### 6.1. Simultaneous Intervals for Normally Distributed Errors

For the case where  $\hat{\mathbf{e}}$  is normally distributed and  $\mathbf{K}$  has full column rank,

$$(\hat{\mathbf{x}} - \bar{\mathbf{x}})^T \mathbf{K}^T \mathbf{S}^{-2} \mathbf{K} (\hat{\mathbf{x}} - \bar{\mathbf{x}}) \sim \chi^2(n) , \quad (6.6)$$

so

$$\Pr \left\{ (\hat{\mathbf{x}} - \bar{\mathbf{x}})^T \mathbf{K}^T \mathbf{S}^{-2} \mathbf{K} (\hat{\mathbf{x}} - \bar{\mathbf{x}}) \leq \gamma_n^2 \right\} = \int_0^{\gamma_n^2} \chi^2(n; q) dq , \quad (6.7)$$

where  $\chi^2(n; q)$  is the probability density function for the  $\chi^2(n)$  distribution. Thus, for any probability  $\alpha$ , if  $\gamma_n^2$  is defined by

$$\int_0^{\gamma_n^2} \chi^2(n; q) dq = \alpha , \quad (6.8)$$

then the corresponding ellipsoid  $\mathcal{S}(\gamma_n)$  will be a  $100\alpha\%$  confidence ellipsoid. Most statistics handbooks tabulate  $\gamma_n^2$  as a function of  $\alpha$  and  $n$  for values  $n \leq 30$  (cf. [1], Chapt. V.1). For larger values of  $n$ , the random variable

$$\eta = \sqrt{2\chi^2} - \sqrt{2n-1} \quad (6.9)$$

follows, to a good approximation, a standard normal distribution, so for a given  $\alpha$ , the corresponding  $\gamma_n^2$  can be computed by

$$\gamma_n^2 = \frac{1}{2} [\lambda_\alpha + \sqrt{2n-1}]^2 , \quad (6.10)$$

where  $\lambda_\alpha$  is the  $\alpha$ -point of the standard normal distribution, i.e.,

$$\int_{-\infty}^{\lambda_\alpha} n(0, 1; \eta) d\eta = \alpha . \quad (6.11)$$



Table 2: The ratio  $\gamma_n/\kappa$  for normally distributed errors

$\alpha$	$n = 20$	$n = 30$	$n = 50$	$n = 70$	$n = 100$	$n = 150$	$n = 200$
0.6667	4.9	5.9	7.6	8.9	10.6	13.0	14.9
0.95	2.9	3.4	4.2	4.9	5.7	6.8	7.8
0.99	2.4	2.8	3.4	3.9	4.5	5.4	6.1
0.999	2.0	2.3	2.8	3.2	3.7	4.4	4.9

It is interesting to compare the widths of the simultaneous intervals (6.5) with those of the one-at-a-time intervals (4.7). The ratio of the corresponding interval widths is

$$\frac{\text{simultaneous interval width}}{\text{single-dimensional interval width}} = \frac{\gamma_n}{\kappa} \quad (6.12)$$

This quantity is tabulated in Table 2 for several values of  $\alpha$  and  $n$ .

As an example, we again consider Problem I with 2-point averaging windows and seek 95% simultaneous intervals. The 95%-point for the standard normal distribution is  $\lambda_\alpha = 1.645$  which, by Eq. (6.10), gives  $\gamma_n^2 = 12.14$ . Using this value in (6.5) gives the intervals shown in the top frame of Figure 5. Comparing this plot with the one in the top frame of Figure 4 gives a rough idea of the price that must be paid for the additional certainty of simultaneous intervals.

## 6.2. Simultaneous Intervals from Chebyshev's Inequality

If the form of the joint probability density function for  $\hat{\epsilon}$  is unknown, then valid, though conservative, confidence ellipsoids can be constructed from Chebyshev's inequality. For a given  $\alpha$ , it is necessary to determine a corresponding  $\gamma_n$  such that

$$\Pr \{ \bar{\mathbf{x}} \in \mathcal{S}(\gamma_n) \} = \Pr \left\{ (\hat{\mathbf{x}} - \bar{\mathbf{x}})^T \mathbf{V} \Sigma^2 \mathbf{V}^T (\hat{\mathbf{x}} - \bar{\mathbf{x}}) \leq \gamma_n^2 \right\} \geq \alpha \quad (6.13)$$

A multivariate Chebyshev inequality given by Olkin and Pratt [13] ensures that for any  $\alpha$ , ( $0 < \alpha < 1$ ), this will be satisfied if

$$\gamma_n = \frac{n}{\sqrt{1-\alpha}} \quad (6.14)$$

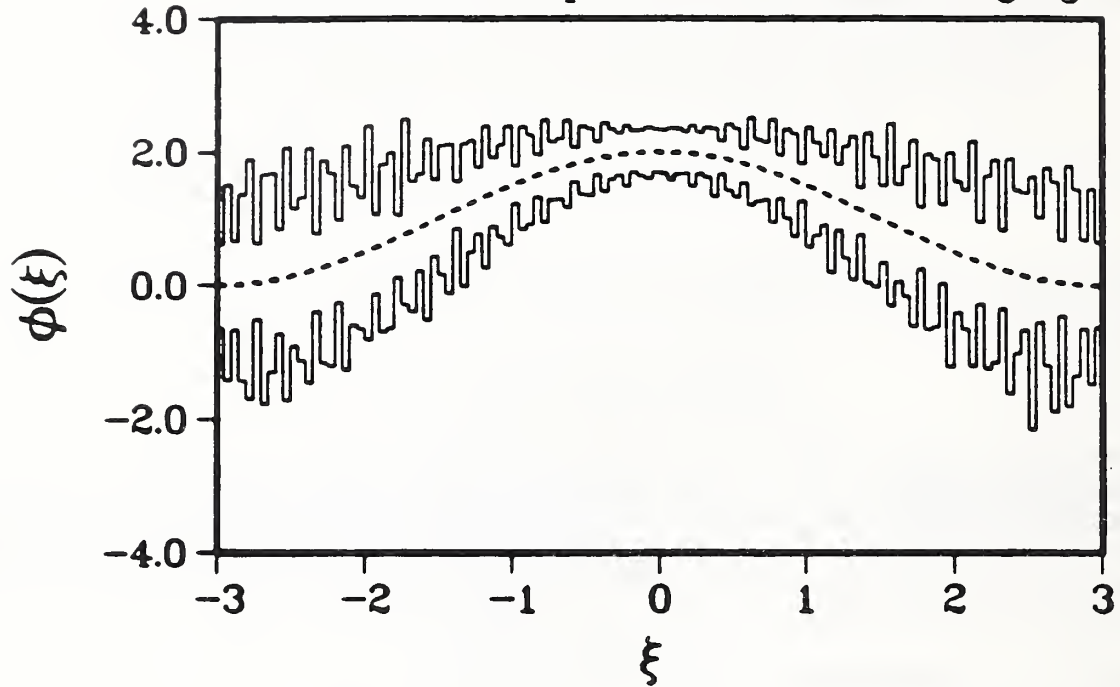
As an example we consider the simultaneous 95% confidence intervals corresponding to a set of 5-point averaging windows for Problem I. The window vectors are given by (2.20) with  $\nu = 5$ . For  $n = 121$  and  $\alpha = 0.95$ , the corresponding  $\gamma_n$  is  $\gamma_n = 541.1$ . The confidence bounds are plotted as histograms in the bottom frame of Figure 5. In spite of the high degree of averaging, the demand for the certainty of simultaneous intervals coupled with a lack of knowledge about the distribution of the errors has produced intervals that are almost useless.

## 7. Nonnegatively Constrained Estimation

The least squares problem arising from the discretization of a system of first kind integral equations is usually poorly conditioned with respect to small variations in the measured vector  $\hat{\mathbf{y}}$ . Quite often, the exact matrix  $\mathbf{K}$  is rank deficient which means, in the classical theory, that the linear functions  $\bar{\phi}_k = \mathbf{w}_k^T \bar{\mathbf{x}}$  may not even be estimable. Fortunately, in cases where the solution  $\mathbf{x}(\xi)$  of the integral equations is physically constrained to be nonnegative, the imposition of those constraints on the solution vector  $\hat{\mathbf{x}}$  almost always stabilizes the estimation problems and gives biased, but bounded, physically plausible solutions, no matter what the value of  $\text{rank}(\mathbf{K})$ . Therefore we append the constraint  $\bar{\mathbf{x}} \geq \mathbf{0}$  to the linear estimation model (1.1), (1.2), i.e.,

$$\hat{\mathbf{y}} = \mathbf{K} \bar{\mathbf{x}} + \hat{\epsilon} \quad , \quad \mathcal{E}(\hat{\epsilon}) = \mathbf{0} \quad , \quad \mathcal{E}(\hat{\epsilon} \hat{\epsilon}^T) = \mathbf{S}^2 \quad , \quad \Pr \{ \bar{\mathbf{x}} \geq \mathbf{0} \} = 1.0 \quad (7.1)$$

150 X121 Phillips Problem  
 Simultaneous Confidence Intervals  
 Using Normal Distribution,  $\alpha = 0.95$   
 Classical Least Squares, 2-Point Averaging



150 X 121 Phillips Problem  
 Simultaneous Confidence Intervals  
 Using Chebyshev Theorem,  $\alpha = 0.95$   
 Classical Least Squares, 5-Point Averaging

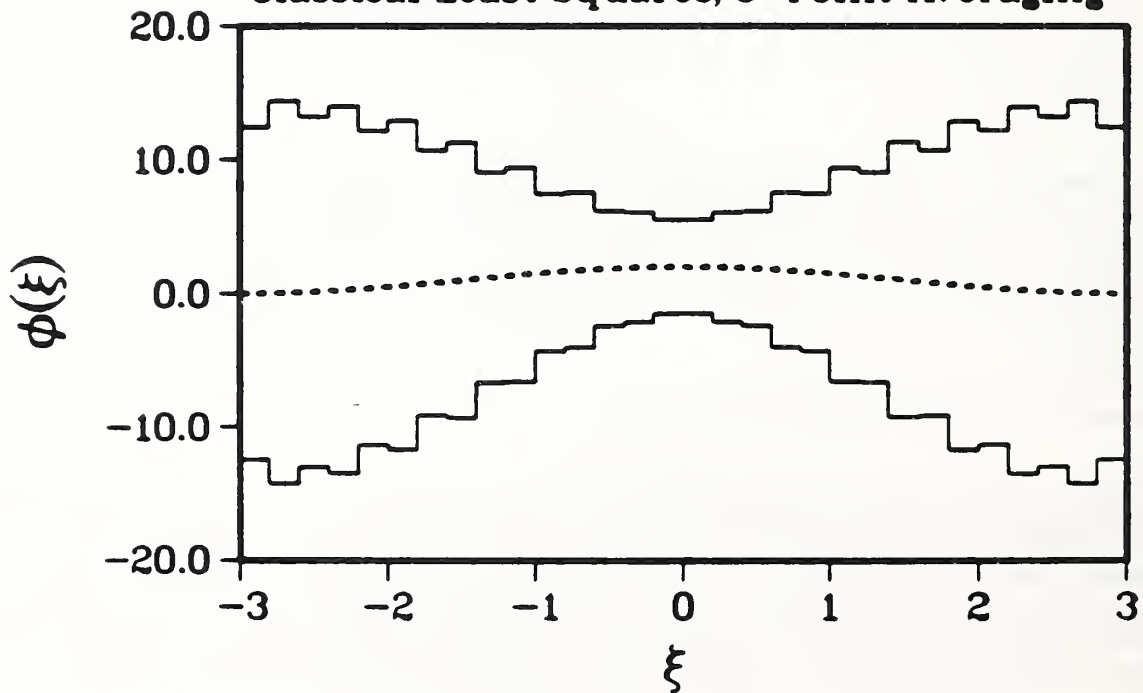


Figure 5: Classical 95% simultaneous confidence intervals for Problem I. Dashed lines are the exact solution  $z(\xi)$ , and solid lines are histograms of estimated lower and upper bounds. Top: 2-point averages from normality assumption. Bottom: 5-point averages from Chebyshev's Inequality.

## 7.1. Constrained Point Estimation

When the nonnegativity constraint is appended, the point estimation problem becomes

$$\rho_{\min} = \min_{\mathbf{x} \geq \mathbf{0}} \{ (\mathbf{K}\mathbf{x} - \hat{\mathbf{y}})^T \mathbf{S}^{-2} (\mathbf{K}\mathbf{x} - \hat{\mathbf{y}}) \} = \hat{r}^2 + \min_{\mathbf{x} \geq \mathbf{0}} \{ (\mathbf{x} - \hat{\mathbf{x}})^T \mathbf{K}^T \mathbf{S}^{-2} \mathbf{K} (\mathbf{x} - \hat{\mathbf{x}}) \} , \quad (7.2)$$

with  $\hat{r}^2$  defined by (3.12). If  $\check{\mathbf{x}}$  denotes the solution vector, i.e., if

$$\rho_{\min} = (\mathbf{K}\check{\mathbf{x}} - \hat{\mathbf{y}})^T \mathbf{S}^{-2} (\mathbf{K}\check{\mathbf{x}} - \hat{\mathbf{y}}) , \quad (7.3)$$

then estimates of the ensemble of averages can be computed by

$$\check{\phi}_k = \mathbf{w}_k^T \check{\mathbf{x}} , \quad k = 1, 2, \dots, p . \quad (7.4)$$

Equation (7.2) defines a nonnegatively constrained least squares problem for which there is no explicit closed form solution, but good algorithms for the numerical solution have been available since the mid-1970s [10]. Figure 6 gives plots of these solutions for Problems I and II. Comparison of these plots with those in Figure 1 shows that the nonnegativity constraints gave good, but not spectacular, improvement. For most real-world ill-posed problems, the improvement is much more dramatic.

## 7.2. Constrained Interval Estimation

The classical method for estimating confidence intervals can also be extended to take the nonnegativity constraints into account. For both one-at-a-time and simultaneous intervals, the classical, unconstrained interval estimation problems can be written in the form

$$\begin{aligned} \hat{\phi}^{lo} &= \min_{\mathbf{x}} \{ \mathbf{w}^T \mathbf{x} \mid (\hat{\mathbf{y}} - \mathbf{K}\mathbf{x})^T \mathbf{S}^{-2} (\hat{\mathbf{y}} - \mathbf{K}\mathbf{x}) = \beta^2 \} , \\ \hat{\phi}^{up} &= \max_{\mathbf{x}} \{ \mathbf{w}^T \mathbf{x} \mid (\hat{\mathbf{y}} - \mathbf{K}\mathbf{x})^T \mathbf{S}^{-2} (\hat{\mathbf{y}} - \mathbf{K}\mathbf{x}) = \beta^2 \} , \end{aligned} \quad (7.5)$$

where  $\beta^2$  is a constant chosen to give the desired confidence level  $\alpha$ . We will show in the following that valid nonnegatively constrained confidence intervals can be obtained by solving problems of the form

$$\begin{aligned} \check{\phi}^{lo} &= \min_{\mathbf{x}} \{ \mathbf{w}^T \mathbf{x} \mid (\hat{\mathbf{y}} - \mathbf{K}\mathbf{x})^T \mathbf{S}^{-2} (\hat{\mathbf{y}} - \mathbf{K}\mathbf{x}) = \mu^2 , \mathbf{x} \geq \mathbf{0} \} , \\ \check{\phi}^{up} &= \max_{\mathbf{x}} \{ \mathbf{w}^T \mathbf{x} \mid (\hat{\mathbf{y}} - \mathbf{K}\mathbf{x})^T \mathbf{S}^{-2} (\hat{\mathbf{y}} - \mathbf{K}\mathbf{x}) = \mu^2 , \mathbf{x} \geq \mathbf{0} \} , \end{aligned} \quad (7.6)$$

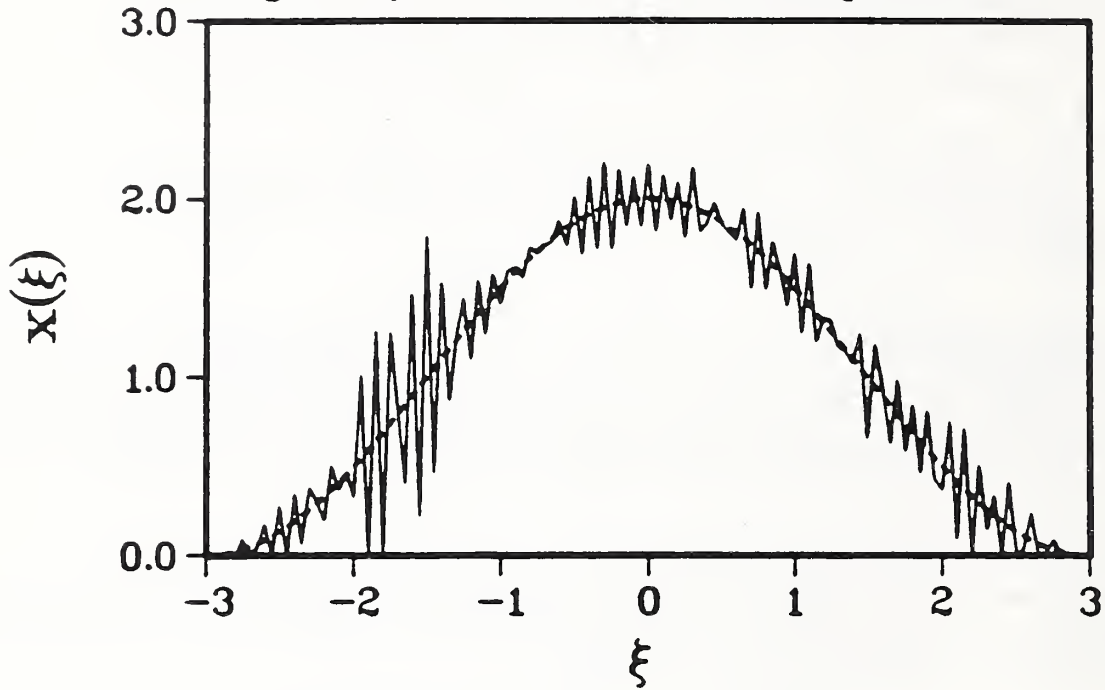
where  $\mu = \beta$  for simultaneous confidence intervals, but  $\mu$  differs from  $\beta$  for one-at-a-time intervals.

The common constraint region in problems (7.6) is the intersection of the  $n$ -dimensional  $\mu$ -ellipsoid with the positive orthant. For any reasonable discretization of the integral equations, the value of  $n$  will be large enough to assure that the positive orthant constitutes a very small fraction of  $n$ -space. The resulting intersection is generally much smaller in all directions than is the  $\mu$ -ellipsoid itself.

Even for rank deficient matrices  $\mathbf{K}$  this intersection is almost always bounded. If  $\mathbf{K}$  is rank deficient, then one or more of the smallest eigenvalues of  $\mathbf{K}^T \mathbf{S}^{-2} \mathbf{K}$  will be zero, and the ellipsoid will be unbounded in the directions of the corresponding eigenvectors. The intersection of the ellipsoid with the positive orthant will, however, be unbounded only if at least one of these degenerate eigenvectors lies in the positive orthant. Since  $\mathbf{x}$ -space has  $2^n$  orthants and  $\mathbf{K}^T \mathbf{S}^{-2} \mathbf{K}$  has fewer than  $n$  degenerate eigenvalues, it will be a very rare occurrence for one of the corresponding eigenvectors to extend into the positive orthant. In many physical applications the matrix  $\mathbf{K}$  is nonnegative, and in such cases,  $\mathbf{x} \geq \mathbf{0}, \mathbf{x} \neq \mathbf{0} \Rightarrow \mathbf{K}^T \mathbf{S}^{-2} \mathbf{K} \mathbf{x} \neq \mathbf{0} \cdot \mathbf{x}$ ; thus, for these applications, no degenerate eigenvector could lie in the positive orthant.

It is not possible to write closed form solutions for the problems (7.6), but numerical calculation of the solutions is possible. The first person to employ nonnegativity constraints in order to reduce the size of estimated confidence intervals was Walter R. Burrus [2] who used the technique in neutron and gamma-ray spectrum unfolding problems. Burrus did not give an algorithm for solving (7.6), but he and his colleagues developed several algorithms [2], [3], [4], [15], [16] for computing suboptimal interval approximations to  $[\check{\phi}^{lo}, \check{\phi}^{up}]$ . In spite of the suboptimality, the intervals thus obtained were uniformly much better than the classical intervals (4.7). O'Leary and Rust [12] have developed an algorithm called BRACKET-LS for computing the optimal intervals. This algorithm was used to compute the confidence intervals given in the remainder of this paper.

150 X 121 Phillips Problem  
 Nonnegatively Constrained Least Squares Solution



108 X 121 Phillips Problem  
 Nonnegatively Constrained Least Squares Solution

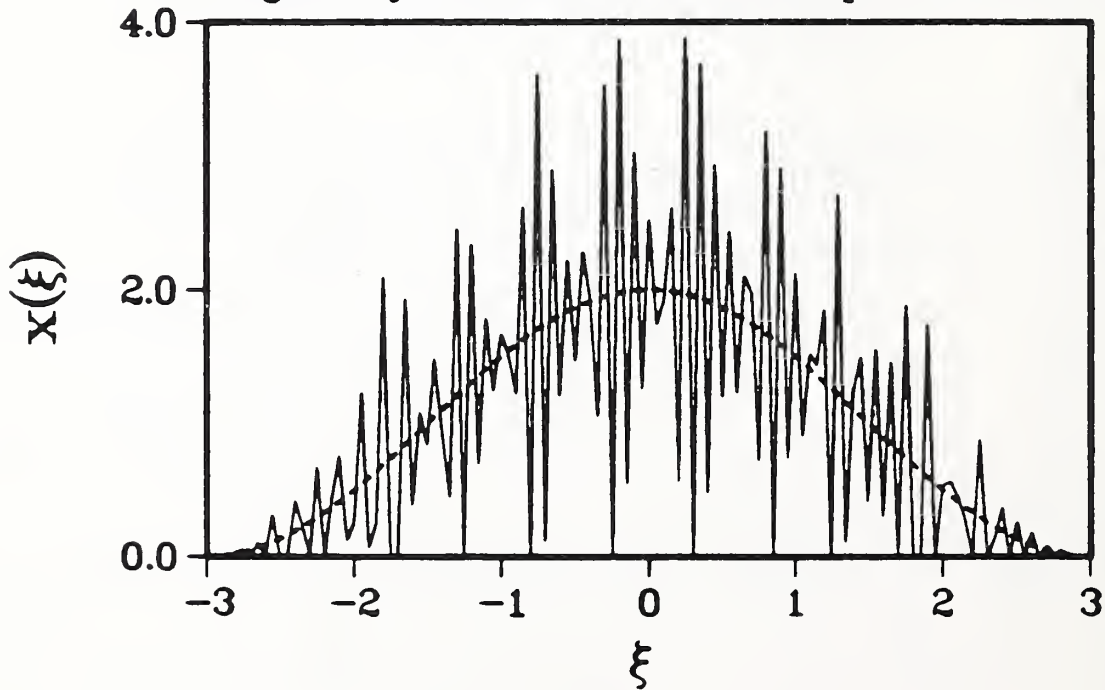


Figure 6: Nonnegatively constrained solutions to the test problems. Dashed lines are exact solutions  $x(\xi)$ , and solid lines are least squares estimates  $\check{x}_k = e_k^T \check{x}$ .  
 Top: Problem I. Bottom: Problem II.



### 7.3. Simultaneous Confidence Intervals

The extension of the classical theory to accommodate nonnegativity constraints in the estimation of one-at-a-time intervals is complicated, so we defer it to the next section. The extension to estimating simultaneous intervals is straightforward because the intersection of a confidence ellipsoid with the positive orthant is itself a confidence region with the same confidence level as the ellipsoid. Let  $\gamma$  be chosen so that

$$\begin{aligned} S(\gamma) &= \left\{ \mathbf{x} \mid (\hat{\mathbf{y}} - \mathbf{K}\mathbf{x})^T \mathbf{S}^{-2} (\hat{\mathbf{y}} - \mathbf{K}\mathbf{x}) \leq \hat{r}^2 + \gamma^2 \right\} \\ &= \left\{ \mathbf{x} \mid (\mathbf{x} - \hat{\mathbf{x}})^T \mathbf{K}^T \mathbf{S}^{-2} \mathbf{K} (\mathbf{x} - \hat{\mathbf{x}}) \leq \gamma^2 \right\} \end{aligned} \quad (7.7)$$

is a  $100\alpha\%$  confidence ellipsoid for  $\bar{\mathbf{x}}$ , and define events  $\mathcal{A}$  and  $\mathcal{B}$  by

$$\mathcal{A} = \{ (\hat{\mathbf{y}} - \mathbf{K}\bar{\mathbf{x}})^T \mathbf{S}^{-2} (\hat{\mathbf{y}} - \mathbf{K}\bar{\mathbf{x}}) \leq \hat{r}^2 + \gamma^2 \} \quad , \quad \mathcal{B} = \{ \bar{\mathbf{x}} \geq \mathbf{o} \} \quad . \quad (7.8)$$

These events have probabilities

$$P(\mathcal{A}) = \alpha \quad , \quad P(\mathcal{B}) = 1.0 \quad , \quad P(\mathcal{A} \cup \mathcal{B}) = 1.0 \quad , \quad (7.9)$$

so it follows from the elementary identity,

$$P(\mathcal{A} \cap \mathcal{B}) = P(\mathcal{A}) + P(\mathcal{B}) - P(\mathcal{A} \cup \mathcal{B}) \quad , \quad (7.10)$$

that

$$\Pr \{ (\hat{\mathbf{y}} - \mathbf{K}\bar{\mathbf{x}})^T \mathbf{S}^{-2} (\hat{\mathbf{y}} - \mathbf{K}\bar{\mathbf{x}}) \leq \hat{r}^2 + \gamma^2 \quad , \quad \bar{\mathbf{x}} \geq \mathbf{o} \} = \alpha \quad . \quad (7.11)$$

Therefore, if

$$\begin{aligned} \check{\phi}_k^{lo} &= \min_{\mathbf{x}} \{ \mathbf{w}_k^T \mathbf{x} \mid (\hat{\mathbf{y}} - \mathbf{K}\mathbf{x})^T \mathbf{S}^{-2} (\hat{\mathbf{y}} - \mathbf{K}\mathbf{x}) \leq \hat{r}^2 + \gamma^2 \quad , \quad \mathbf{x} \geq \mathbf{o} \} \quad , \\ \check{\phi}_k^{up} &= \max_{\mathbf{x}} \{ \mathbf{w}_k^T \mathbf{x} \mid (\hat{\mathbf{y}} - \mathbf{K}\mathbf{x})^T \mathbf{S}^{-2} (\hat{\mathbf{y}} - \mathbf{K}\mathbf{x}) \leq \hat{r}^2 + \gamma^2 \quad , \quad \mathbf{x} \geq \mathbf{o} \} \quad , \end{aligned} \quad (7.12)$$

for  $k = 1, 2, \dots, p$ , then it follows that

$$\Pr \left\{ \check{\phi}_k^{lo} \leq \mathbf{w}_k^T \bar{\mathbf{x}} \leq \check{\phi}_k^{up} \quad , \quad k = 1, 2, \dots, p \right\} = \alpha \quad , \quad (7.13)$$

and this result holds for any number  $p$  of window vectors.

In Section 6.1 we saw that for full-rank problems with normally distributed errors, a  $100\alpha\%$  confidence ellipsoid is obtained by choosing  $\gamma^2$  to be the  $\alpha$ -point of the  $\chi^2(n)$  distribution [cf. Eq. (6.8)]. With the addition of nonnegativity constraints, it becomes possible to estimate confidence intervals for underdetermined problems, but the choice of  $\gamma^2$  is different. Rust and Burrus [17, Chapt. 6] proved that if  $\hat{\mathbf{y}}$  is normally distributed, and

$$\hat{\mathbf{x}} = (\mathbf{S}^{-1} \mathbf{K})^\dagger \mathbf{S}^{-1} \hat{\mathbf{y}} = (\mathbf{K}^T \mathbf{S}^{-2} \mathbf{K})^\dagger \mathbf{K}^T \mathbf{S}^{-2} \hat{\mathbf{y}} \quad , \quad (7.14)$$

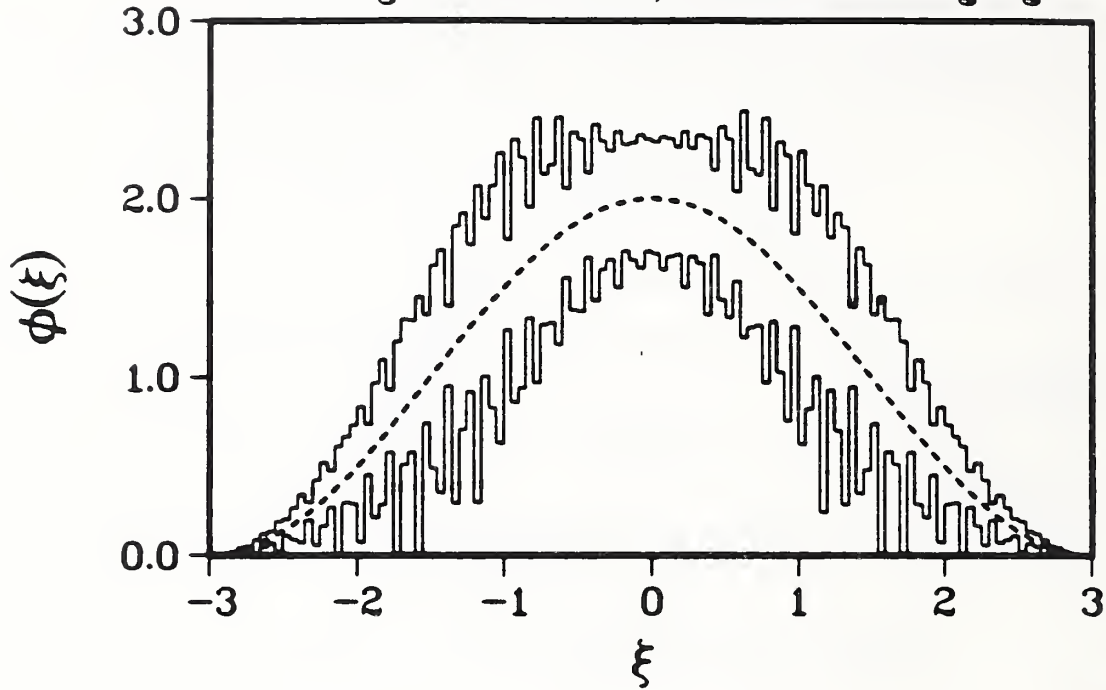
is the corresponding generalized inverse, least squares solution vector, then

$$(\bar{\mathbf{x}} - \hat{\mathbf{x}})^T \mathbf{K}^T \mathbf{S}^{-2} \mathbf{K} (\bar{\mathbf{x}} - \hat{\mathbf{x}}) \sim \chi^2(\nu) \quad , \quad (7.15)$$

where  $\nu = \text{rank}(\mathbf{K})$ . Therefore, if  $\gamma$  is taken to be the  $\alpha$ -point of the  $\chi^2(\nu)$  distribution, then Eqs. (7.12) define the confidence interval which can be computed with BRACKET-LS algorithm [12]. Determining the exact rank  $\nu$  is often a difficult and uncertain procedure, but it is better to overestimate the rank than to underestimate it. In the former case, the confidence intervals will be suboptimal (larger than necessary to assure the confidence level), but in the latter case they will be dishonest (too small to assure the claimed confidence level). A conservative procedure is to choose  $\nu = \min\{m, n\}$ .

Figure 7 gives plots of the simultaneous confidence intervals computed by BRACKET-LS for Problem I using 2-point averaging, and Problem II using 3-point averaging, when the measuring errors are normally distributed. The bounds for Problem I can be compared with those in the top frame of Figure 5 to assess the improvement obtained by using the nonnegativity constraints. The improvement is even more dramatic for Problem II because the intervals are all unbounded for the unconstrained problem. The nonnegativity

150 X121 Phillips Problem  
 Simultaneous Confidence Intervals  
 Using Normal Distribution,  $\alpha = 0.95$   
 Nonneg. Constraints, 2-Point Averaging



108 X 121 Phillips Problem  
 Simultaneous Confidence Intervals  
 Using Normal Distribution,  $\alpha = 0.95$   
 Nonneg. Constraints, 3-Point Averaging

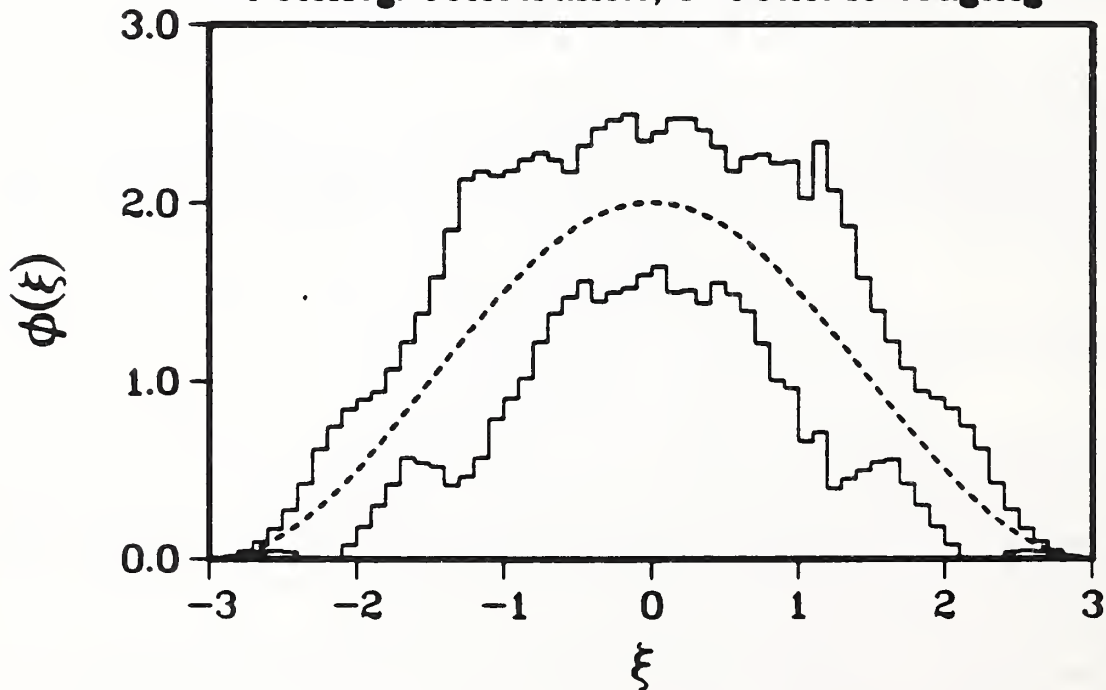


Figure 7: Nonnegatively constrained, simultaneous 95% confidence intervals from the normal distribution. Dashed lines are the exact solutions  $x(\xi)$ , and solid lines are histograms of estimated lower and upper bounds. Top: Problem I, 2-point averages. Bottom: Problem II, 3-point averages.



constraints make it possible to compute useful confidence intervals even though the underlying linear model is underdetermined.

When the form of the error distribution is unknown, the procedure is the same except  $\gamma$  is given by (6.14). Figure 8 gives plots of the bounds obtained for both Problems I and II using 5-point averaging. Even with this higher degree of averaging, the bounds are poor in comparison with those in Figure 7, but they are considerably stronger than those in the bottom frame of Figure 5 which were obtained without the nonnegativity constraints.

## 8. Constrained One-at-a-Time Confidence Intervals

### 8.1. The Burrus Conjecture

Nonnegatively constrained one-at-a-time confidence intervals cannot be defined in the same manner as constrained simultaneous intervals. More precisely, if  $\kappa$  is defined by (4.9) or by (4.12), then valid  $100\alpha\%$  one-at-a-time intervals cannot be calculated from Equations (7.6) with  $\mu^2 = \hat{r}^2 + \kappa^2$ . In fact, it almost always happens that

$$\hat{r}^2 + \kappa^2 < \rho_{\min} = \min_{\mathbf{x} \geq \mathbf{0}} \{(\hat{\mathbf{y}} - \mathbf{K}\mathbf{x})^T \mathbf{S}^{-2}(\hat{\mathbf{y}} - \mathbf{K}\mathbf{x})\} , \quad (8.1)$$

i.e., the ellipsoid (5.4) had no points in common with the positive orthant. This difficulty arises because the regression models almost never have  $m \gg n$ . Accordingly, the classical least squares procedure produces a solution vector  $\hat{\mathbf{x}}$  which models a large portion of the measurement errors  $\hat{\mathbf{e}}$  and hence gives an unrealistically low value for  $\hat{r}^2$ . This fact was pointed out by Walter R. Burrus [2] who also noted that the ellipsoid defined by

$$(\hat{\mathbf{y}} - \mathbf{K}\mathbf{x})^T \mathbf{S}^{-2}(\hat{\mathbf{y}} - \mathbf{K}\mathbf{x}) \leq \rho_{\min} + \kappa^2 \quad (8.2)$$

always has a non-empty intersection with the positive orthant. He conjectured [17, Chapt. 6] that valid confidence intervals could be obtained by using that intersection as the constraint region, i.e., by solving the problems

$$\begin{aligned} \check{\phi}^{lo} &= \min_{\mathbf{x}} \{ \mathbf{w}^T \mathbf{x} \mid (\hat{\mathbf{y}} - \mathbf{K}\mathbf{x})^T \mathbf{S}^{-2}(\hat{\mathbf{y}} - \mathbf{K}\mathbf{x}) = \rho_{\min} + \kappa^2, \mathbf{x} \geq \mathbf{0} \} , \\ \check{\phi}^{up} &= \max_{\mathbf{x}} \{ \mathbf{w}^T \mathbf{x} \mid (\hat{\mathbf{y}} - \mathbf{K}\mathbf{x})^T \mathbf{S}^{-2}(\hat{\mathbf{y}} - \mathbf{K}\mathbf{x}) = \rho_{\min} + \kappa^2, \mathbf{x} \geq \mathbf{0} \} . \end{aligned} \quad (8.3)$$

We shall show in the following that this conjecture is indeed correct.

### 8.2. The Duality Theorem

In order to prove the Burrus Conjecture, it will be necessary to restate problems (8.3) in equivalent forms defined by the *Duality Theorem for Nonlinear Programming* [20]:

**Wolfe's Duality Theorem.** Suppose  $\mathbf{z}$  is an  $N$ -vector of unknown variables,  $f(\mathbf{z})$  is a scalar function and  $\mathbf{g}(\mathbf{z})$  is an  $M$ -vector function of  $\mathbf{z}$ , and that  $f(\mathbf{z})$  and  $\mathbf{g}(\mathbf{z})$  are all convex and differentiable on an open set  $\mathcal{Z}$ . If

$$\phi_P = \min_{\mathbf{z} \in \mathcal{Z}} \{ f(\mathbf{z}) \mid \mathbf{g}(\mathbf{z}) \geq \mathbf{0} \} , \quad (8.4)$$

and the constraints  $\mathbf{g}(\mathbf{z}) \geq \mathbf{0}$  satisfy a constraint qualification, then

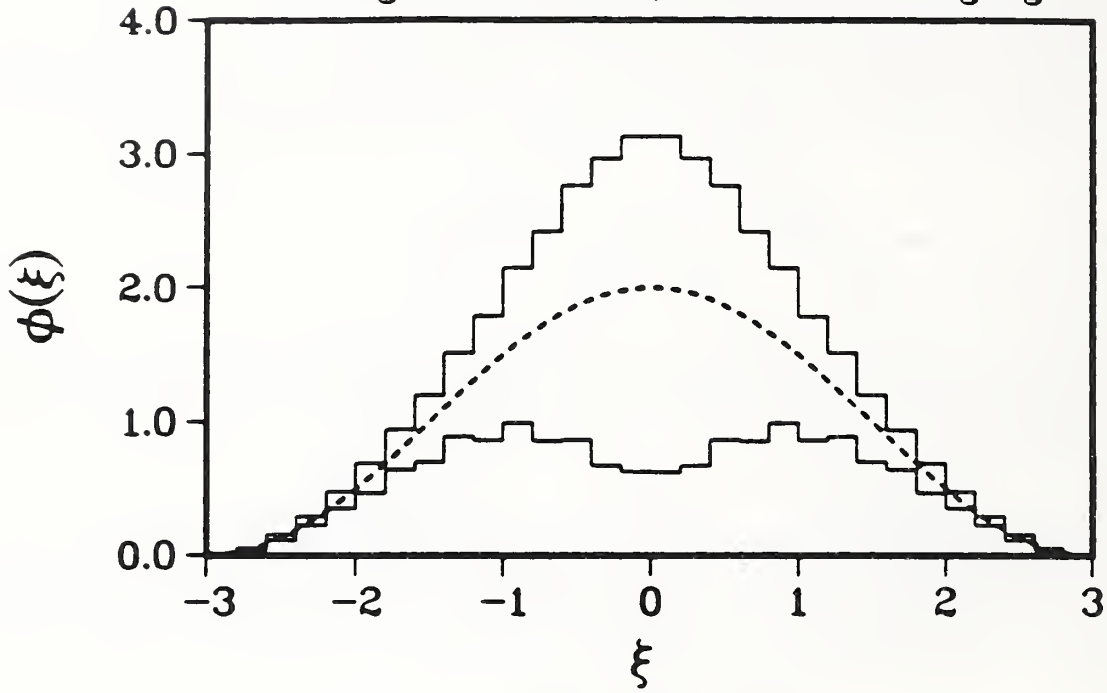
$$\phi_P = \phi_D , \quad (8.5)$$

where  $\phi_D$  is determined by the dual problem

$$\phi_D = \max_{\{\mathbf{z}, \mathbf{v}\}} \left\{ f(\mathbf{z}) - \mathbf{g}^T \mathbf{v} \mid \nabla f(\mathbf{z}) = \left( \frac{\partial \mathbf{g}^T}{\partial \mathbf{z}} \right) \mathbf{v}, \mathbf{v} \geq \mathbf{0} \right\} . \quad (8.6)$$

Equation (8.4) is called the primal problem. The constraint qualification can be any one of several regularity conditions on either the convexity or differentiability properties of the functions  $\mathbf{g}_i(\mathbf{z})$ . More details on such conditions can be found in [11, Chaps. 5, 7].

150 X 121 Phillips Problem  
 Simultaneous Confidence Intervals  
 Using Chebyshev Theorem,  $\alpha = 0.95$   
 Nonneg. Constraints, 5-Point Averaging



108 X 121 Phillips Problem  
 Simultaneous Confidence Intervals  
 Using Chebyshev Theorem,  $\alpha = 0.95$   
 Nonneg. Constraints, 5-Point Averaging

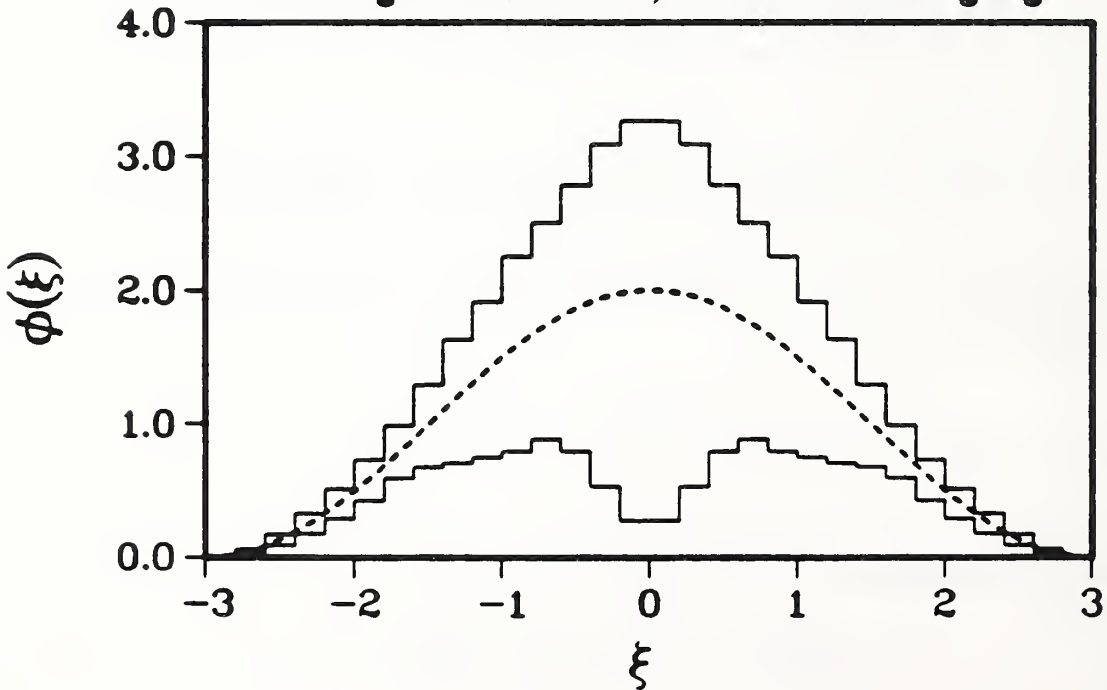


Figure 8: Nonnegatively constrained, simultaneous 95% confidence intervals from Chebyshev's Theorem. Dashed lines are exact solutions  $z(\xi)$ , and solid lines are histograms of estimated lower and upper bounds. Top: Problem I, 5-point averages. Bottom: Problem II, 5-point averages.

It is instructive to apply this theorem to the classical, full-rank interval estimation problems (5.1). To simplify the notation, let

$$\mu = +\sqrt{\hat{r}^2 + \kappa^2} , \quad (8.7)$$

and consider first the lower bound problem which can also be written with inequality constraints,

$$\hat{\phi}^{lo} = \min_{\mathbf{x}} \{ \mathbf{w}^T \mathbf{x} \mid (\hat{\mathbf{y}} - \mathbf{K} \mathbf{x})^T \mathbf{S}^{-2} (\hat{\mathbf{y}} - \mathbf{K} \mathbf{x}) \leq \mu^2 \} , \quad (8.8)$$

because the indicated minimum is attained on the boundary of the ellipsoid. Many equivalent forms of the dual problem can be derived, but it is easiest to get the form we need by using the artifice of writing the problem in terms of the original variables  $\mathbf{x}$  and the scaled residual vector

$$\mathbf{r} = \mathbf{S}^{-1} (\hat{\mathbf{y}} - \mathbf{K} \mathbf{x}) . \quad (8.9)$$

The vector  $\mathbf{r}$  can be specified by two vector inequalities,

$$\mathbf{r} - \mathbf{S}^{-1} (\hat{\mathbf{y}} - \mathbf{K} \mathbf{x}) \geq \mathbf{o} , \quad -\mathbf{r} + \mathbf{S}^{-1} (\hat{\mathbf{y}} - \mathbf{K} \mathbf{x}) \geq \mathbf{o} . \quad (8.10)$$

Reformulating (8.8) in terms of the ( $N = m + n$ ) variables

$$\mathbf{z} \equiv (\mathbf{x}^T , \mathbf{r}^T)^T , \quad (8.11)$$

gives

$$\mathbf{g}(\mathbf{z}) \equiv \begin{pmatrix} -\mathbf{S}^{-1}(\hat{\mathbf{y}} - \mathbf{K} \mathbf{x}) + \mathbf{r} \\ +\mathbf{S}^{-1}(\hat{\mathbf{y}} - \mathbf{K} \mathbf{x}) - \mathbf{r} \\ \mu - \sqrt{\mathbf{r}^T \mathbf{r}} \end{pmatrix} \geq \begin{pmatrix} \mathbf{o} \\ \mathbf{o} \\ 0 \end{pmatrix} , \quad (8.12)$$

for the constraints and

$$f(\mathbf{z}) \equiv (\mathbf{w}^T , \mathbf{o}^T) \mathbf{z} \quad (8.13)$$

for the objective function. The problem then becomes

$$\hat{\phi}^{lo} = \phi_P = \min_{\mathbf{z} \in \mathcal{Z}} \{ f(\mathbf{z}) \mid \mathbf{g}(\mathbf{z}) \geq \mathbf{o} \} , \quad (8.14)$$

where  $\mathcal{Z} = \mathcal{R}^N$ . The functions  $f(\mathbf{z})$  and  $\mathbf{g}(\mathbf{z})$  are convex and differentiable everywhere, and the constraints  $\mathbf{g}(\mathbf{z}) \geq \mathbf{o}$  satisfy Slater's constraint qualification (cf. [11, §5.4.3]), so the Duality Theorem is applicable. Therefore

$$\hat{\phi}^{lo} = \phi_D = \max_{\{\mathbf{z}, \mathbf{v}\}} \left\{ f(\mathbf{z}) - \mathbf{g}^T \mathbf{v} \mid \nabla f(\mathbf{z}) = \left( \frac{\partial \mathbf{g}^T}{\partial \mathbf{z}} \right) \mathbf{v}, \mathbf{v} \geq \mathbf{o} \right\} , \quad (8.15)$$

where  $\mathbf{v}$  is a ( $M = 2m + 1$ )-vector of dual variables that can also be written

$$\mathbf{v}^T = (\mathbf{v}_1^T , \mathbf{v}_2^T , v_3)^T , \quad (8.16)$$

with subvectors  $\mathbf{v}_1$  and  $\mathbf{v}_2$  of length  $m$ .

The derivatives required for the dual constraints are

$$\nabla f(\mathbf{z}) = \begin{pmatrix} \mathbf{w} \\ \mathbf{o} \end{pmatrix} , \quad \left( \frac{\partial \mathbf{g}}{\partial \mathbf{z}} \right) = \begin{pmatrix} \mathbf{S}^{-1} \mathbf{K} & \mathbf{I} \\ -\mathbf{S}^{-1} \mathbf{K} & -\mathbf{I} \\ \mathbf{o} & -\frac{\mathbf{r}}{\sqrt{\mathbf{r}^T \mathbf{r}}} \end{pmatrix} , \quad (8.17)$$

so the constraints themselves can be written

$$\mathbf{K}^T \mathbf{S}^{-1} (\mathbf{v}_1 - \mathbf{v}_2) = \mathbf{w} , \quad (\mathbf{v}_1 - \mathbf{v}_2) - v_3 \frac{\mathbf{r}}{\sqrt{\mathbf{r}^T \mathbf{r}}} = \mathbf{o} , \quad (8.18)$$

$$\mathbf{v}_1 \geq \mathbf{o} , \quad \mathbf{v}_2 \geq \mathbf{o} , \quad v_3 \geq 0 .$$

Note that the dual vectors  $\mathbf{v}_1$  and  $\mathbf{v}_2$  appear only as components of the difference  $\mathbf{v}_1 - \mathbf{v}_2$ . Even though  $\mathbf{v}_1$  and  $\mathbf{v}_2$  are constrained to be nonnegative, their difference is totally unconstrained. Defining new, unconstrained dual variables

$$\mathbf{v}_0 \equiv \mathbf{v}_1 - \mathbf{v}_2 , \quad (8.19)$$

allows the constraints to be written

$$\mathbf{K}^T \mathbf{S}^{-1} \mathbf{v}_0 = \mathbf{w} \quad , \quad \mathbf{v}_0 = v_3 \frac{\mathbf{r}}{\sqrt{\mathbf{r}^T \mathbf{r}}} \quad , \quad v_3 \geq 0 \quad , \quad (8.20)$$

and eliminating the variable  $\mathbf{v}_0$  gives

$$v_3 \frac{\mathbf{K}^T \mathbf{S}^{-1} \mathbf{r}}{\sqrt{\mathbf{r}^T \mathbf{r}}} = \mathbf{w} \quad , \quad v_3 \geq 0 \quad . \quad (8.21)$$

Note that  $\mathbf{r}/\sqrt{\mathbf{r}^T \mathbf{r}}$  is a unit vector, and  $v_3$  is simply a normalization parameter used to scale the length of  $\mathbf{K}^T \mathbf{S}^{-1}$  times this vector to equal the length of  $\mathbf{w}$ .

The objective function for the dual problem is

$$f(\mathbf{z}) - \mathbf{g}^T(\mathbf{z})\mathbf{v} = \mathbf{w}^T \mathbf{x} + \left[ \hat{\mathbf{y}}^T \mathbf{S}^{-1} - \mathbf{x}^T \mathbf{K}^T \mathbf{S}^{-1} - \mathbf{r}^T \right] (\mathbf{v}_1 - \mathbf{v}_2) + v_3 \left( \sqrt{\mathbf{r}^T \mathbf{r}} - \mu \right) \quad . \quad (8.22)$$

$$= \mathbf{w}^T \mathbf{x} + \left[ \hat{\mathbf{y}}^T \mathbf{S}^{-1} - \mathbf{x}^T \mathbf{K}^T \mathbf{S}^{-1} - \mathbf{r}^T \right] \mathbf{v}_0 + v_3 \left( \sqrt{\mathbf{r}^T \mathbf{r}} - \mu \right) \quad . \quad (8.23)$$

$$= v_3 \frac{\mathbf{r}^T \mathbf{S}^{-1} \mathbf{K} \mathbf{x}}{\sqrt{\mathbf{r}^T \mathbf{r}}} + \left( \hat{\mathbf{y}}^T \mathbf{S}^{-1} - \mathbf{x}^T \mathbf{K}^T \mathbf{S}^{-1} - \mathbf{r}^T \right) \frac{v_3 \mathbf{r}}{\sqrt{\mathbf{r}^T \mathbf{r}}} + v_3 \left( \sqrt{\mathbf{r}^T \mathbf{r}} - \mu \right) \quad (8.24)$$

$$= v_3 \left( \frac{\hat{\mathbf{y}}^T \mathbf{S}^{-1} \mathbf{r}}{\sqrt{\mathbf{r}^T \mathbf{r}}} - \mu \right) \quad . \quad (8.25)$$

To simplify, we introduce a change of variables

$$\mathbf{u} \equiv \frac{v_3 \mathbf{S}^{-1} \mathbf{r}}{\sqrt{\mathbf{r}^T \mathbf{r}}} \quad , \quad (8.26)$$

and note that the variables  $\mathbf{u}$  are constrained only by  $\mathbf{K}^T \mathbf{u} = \mathbf{w}$ . Then multiplying both sides of (8.26) by  $\mathbf{S}$  and “squaring” yields

$$\mathbf{u}^T \mathbf{S}^2 \mathbf{u} = v_3^2 \quad . \quad (8.27)$$

Now we can write the dual objective function as

$$f(\mathbf{z}) - \mathbf{g}^T(\mathbf{z})\mathbf{v} = \hat{\mathbf{y}}^T \mathbf{u} - \mu \sqrt{\mathbf{u}^T \mathbf{S}^2 \mathbf{u}} \quad , \quad (8.28)$$

yielding, after substitution of  $\mu = \sqrt{\hat{r}^2 + \kappa^2}$ , our final form of the dual problem

$$\hat{\phi}^{lo} = \max_{\mathbf{u}} \left\{ \mathbf{u}^T \hat{\mathbf{y}} - (\hat{r}^2 + \kappa^2)^{\frac{1}{2}} \sqrt{\mathbf{u}^T \mathbf{S}^2 \mathbf{u}} \quad | \quad \mathbf{u}^T \mathbf{K} = \mathbf{w}^T \right\} \quad . \quad (8.29)$$

In a similar manner, it can be shown that the dual problem for the upper bound problem in (5.1) can be written

$$\hat{\phi}^{up} = \min_{\mathbf{u}} \left\{ \mathbf{u}^T \hat{\mathbf{y}} + (\hat{r}^2 + \kappa^2)^{\frac{1}{2}} \sqrt{\mathbf{u}^T \mathbf{S}^2 \mathbf{u}} \quad | \quad \mathbf{u}^T \mathbf{K} = \mathbf{w}^T \right\} \quad . \quad (8.30)$$

It is easy to independently verify that these two dual problems are equivalent to the primals (5.1) because they too can be solved by Lagrange multipliers. The solution vectors are

$$\begin{aligned} \hat{\mathbf{u}}^{lo} &= \hat{\mathbf{u}} + \frac{\sqrt{\hat{\mathbf{u}}^T \mathbf{S}^2 \hat{\mathbf{u}}}}{\kappa} \mathbf{S}^{-2} (\hat{\mathbf{y}} - \mathbf{K} \hat{\mathbf{x}}) \quad , \\ \hat{\mathbf{u}}^{up} &= \hat{\mathbf{u}} - \frac{\sqrt{\hat{\mathbf{u}}^T \mathbf{S}^2 \hat{\mathbf{u}}}}{\kappa} \mathbf{S}^{-2} (\hat{\mathbf{y}} - \mathbf{K} \hat{\mathbf{x}}) \quad , \end{aligned} \quad (8.31)$$

where  $\hat{\mathbf{u}}$  is given by (3.8) and  $\hat{\mathbf{x}}$  by (3.11). The corresponding optimal values are

$$\begin{aligned} \hat{\phi}^{lo} &= (\hat{\mathbf{u}}^{lo})^T \hat{\mathbf{y}} - (\hat{r}^2 + \kappa^2)^{\frac{1}{2}} \sqrt{(\hat{\mathbf{u}}^{lo})^T \mathbf{S}^2 \hat{\mathbf{u}}^{lo}} = \hat{\mathbf{u}}^T \hat{\mathbf{y}} - \kappa \sqrt{\hat{\mathbf{u}}^T \mathbf{S}^2 \hat{\mathbf{u}}} \quad , \\ \hat{\phi}^{up} &= (\hat{\mathbf{u}}^{up})^T \hat{\mathbf{y}} + (\hat{r}^2 + \kappa^2)^{\frac{1}{2}} \sqrt{(\hat{\mathbf{u}}^{up})^T \mathbf{S}^2 \hat{\mathbf{u}}^{up}} = \hat{\mathbf{u}}^T \hat{\mathbf{y}} + \kappa \sqrt{\hat{\mathbf{u}}^T \mathbf{S}^2 \hat{\mathbf{u}}} \quad , \end{aligned} \quad (8.32)$$

which are the same as the values given by (4.7).

The primal problems (5.1) correspond to a Gaussian formulation of linear interval estimation while the corresponding dual problems (8.29) and (8.30) correspond to the Markov formulation. In the objective functions of the latter, the inner product  $\mathbf{u}^T \hat{\mathbf{y}}$  is a linear estimator of the unknown  $\bar{\phi}$  and the quantity  $\mathbf{u}^T \mathbf{S}^{-2} \mathbf{u}$  is the variance of that estimator. *The constraint functions common to both problems require the estimators to be unbiased.* When nonnegativity constraints are imposed on the problems, the forms of the objective functions are unaffected, but the constraint functions change in an interesting way.

### 8.3. Lower-biased and Upper-biased Estimators

Now consider the nonnegatively constrained interval estimation problems (8.3). The lower bound primal problem is

$$\check{\phi}^{lo} = \min_{\mathbf{x}} \{ \mathbf{w}^T \mathbf{x} \mid (\hat{\mathbf{y}} - \mathbf{K} \mathbf{x})^T \mathbf{S}^{-2} (\hat{\mathbf{y}} - \mathbf{K} \mathbf{x}) \leq \mu^2, \mathbf{x} \geq \mathbf{o} \}, \quad (8.33)$$

where

$$\mu = \sqrt{\rho_{\min} + \kappa^2}. \quad (8.34)$$

The transformation of this problem to its dual form is similar to that given in the preceding section for the unconstrained problem (8.8). The only difference is the additional set of  $n$  constraints  $\mathbf{x} \geq \mathbf{o}$ , which augment the primal constraint relations (8.12). The dual vector (8.16) is also augmented by  $n$  additional variables  $\mathbf{v}_4$ , i.e.,

$$\mathbf{v}^T = (\mathbf{v}_1^T, \mathbf{v}_2^T, \mathbf{v}_3, \mathbf{v}_4^T)^T, \quad (8.35)$$

and the constraint expressions (8.18) are changed to

$$\begin{aligned} \mathbf{K}^T \mathbf{S}^{-1} (\mathbf{v}_1 - \mathbf{v}_2) + \mathbf{v}_4 = \mathbf{w}, \quad (\mathbf{v}_1 - \mathbf{v}_2) - v_3 \frac{\mathbf{r}}{\sqrt{\mathbf{r}^T \mathbf{r}}} = \mathbf{o}, \\ \mathbf{v}_1 \geq \mathbf{o}, \quad \mathbf{v}_2 \geq \mathbf{o}, \quad v_3 \geq 0, \quad \mathbf{v}_4 \geq \mathbf{o}. \end{aligned} \quad (8.36)$$

The objective function remains the same as it was for the classical problem, and the definition of  $\mathbf{v}_0$  and the determination of  $\mathbf{v}_3$  are the same also. The net result is that

$$\check{\phi}^{lo} = \max_{\{\mathbf{v}_0, \mathbf{v}_4\}} \left\{ \hat{\mathbf{y}}^T \mathbf{S}^{-1} \mathbf{v}_0 - \mu \sqrt{\mathbf{v}_0^T \mathbf{v}_0} \mid \mathbf{K}^T \mathbf{S}^{-1} \mathbf{v}_0 + \mathbf{v}_4 = \mathbf{w}, \mathbf{v}_4 \geq \mathbf{o} \right\}. \quad (8.37)$$

The objective function does not depend on  $\mathbf{v}_4$ , so the constraints can be written more simply as

$$\mathbf{K}^T \mathbf{S}^{-1} \mathbf{v}_0 \leq \mathbf{w}. \quad (8.38)$$

Defining  $\mathbf{u}$  as in (8.26) and restoring the value of  $\mu$  gives

$$\check{\phi}^{lo} = \max_{\mathbf{u}} \left\{ \mathbf{u}^T \hat{\mathbf{y}} - (\rho_{\min} + \kappa^2)^{\frac{1}{2}} \sqrt{\mathbf{u}^T \mathbf{S}^2 \mathbf{u}} \mid \mathbf{u}^T \mathbf{K} \leq \mathbf{w}^T \right\}. \quad (8.39)$$

Similarly, the dual statement of the upper bound problem can be shown to be

$$\check{\phi}^{up} = \min_{\mathbf{u}} \left\{ \mathbf{u}^T \hat{\mathbf{y}} + (\rho_{\min} + \kappa^2)^{\frac{1}{2}} \sqrt{\mathbf{u}^T \mathbf{S}^2 \mathbf{u}} \mid \mathbf{u}^T \mathbf{K} \geq \mathbf{w}^T \right\}. \quad (8.40)$$

The objective functions for the current dual problems differ from those of the classical duals (8.29) and (8.30) only in the substitution of  $\rho_{\min}$  for  $\hat{r}^2$ . The constraints, however, are quite different. In the classical case, the estimators  $\mathbf{u}^T \hat{\mathbf{y}}$  are required to be unbiased in both cases. For the nonnegatively constrained problems, the estimator for the lower bound is required to be *lower biased*, and the estimator for the upper bound is required to be *upper biased*. This equivalence of bias constraints in the Markov formulation to nonnegativity constraints in the Gauss formulation is a surprising result whose significance we are at present unable to assess. It seems very logical to use lower biased estimators when seeking a lower bound and upper biased estimators when seeking an upper bound, but such constraints appear at first glance to be less restrictive than the classical unbiasedness constraints. In practical problems, nonnegatively constrained confidence intervals are smaller than the corresponding classical intervals, even though the substitution of  $\rho_{\min}$  for  $\hat{r}^2$  in the objective functions should tend to increase the size of the former. It would seem then that the upper and lower bias constraints are considerably stronger than the unbiasedness constraint.



#### 8.4. Proof of the Burrus Conjecture

Having established the equivalence of the dual problems (8.39) and (8.40) to the primals (8.3), it is not difficult to prove the Burrus Conjecture which we now state as a theorem.

**Theorem.** Let  $\hat{\mathbf{y}}$  be a given  $m$ -vector,  $\mathbf{K}$  a given  $m \times n$  matrix, and  $\bar{\mathbf{x}}$  an unknown  $n$ -vector satisfying

$$\hat{\mathbf{y}} = \mathbf{K}\bar{\mathbf{x}} + \hat{\boldsymbol{\epsilon}}, \quad \hat{\boldsymbol{\epsilon}} \sim N(\mathbf{0}, \mathbf{S}^2), \quad (8.41)$$

and

$$\Pr\{\bar{\mathbf{x}} \geq \mathbf{0}\} = 1. \quad (8.42)$$

Let  $\rho_{\min}$  be defined by

$$\rho_{\min} = \min_{\mathbf{x}} \{ (\hat{\mathbf{y}} - \mathbf{K}\mathbf{x})^T \mathbf{S}^{-2} (\hat{\mathbf{y}} - \mathbf{K}\mathbf{x}) \mid \mathbf{x} \geq \mathbf{0} \}. \quad (8.43)$$

Let  $\alpha$  be a given probability ( $0 < \alpha < 1$ ), and suppose that the value  $\kappa$  is chosen so that

$$\frac{1}{\sqrt{2\pi}} \int_{-\kappa}^{+\kappa} \exp\left(-\frac{\eta^2}{2}\right) d\eta = \alpha. \quad (8.44)$$

If  $\mathbf{w}$  is a given  $n$ -vector, and

$$\begin{aligned} \check{\phi}^{lo} &= \min_{\mathbf{x}} \{ \mathbf{w}^T \mathbf{x} \mid (\hat{\mathbf{y}} - \mathbf{K}\mathbf{x})^T \mathbf{S}^{-2} (\hat{\mathbf{y}} - \mathbf{K}\mathbf{x}) \leq \rho_{\min} + \kappa^2, \mathbf{x} \geq \mathbf{0} \}, \\ \check{\phi}^{up} &= \max_{\mathbf{x}} \{ \mathbf{w}^T \mathbf{x} \mid (\hat{\mathbf{y}} - \mathbf{K}\mathbf{x})^T \mathbf{S}^{-2} (\hat{\mathbf{y}} - \mathbf{K}\mathbf{x}) \leq \rho_{\min} + \kappa^2, \mathbf{x} \geq \mathbf{0} \}, \end{aligned} \quad (8.45)$$

then

$$\Pr\{\check{\phi}^{lo} \leq \mathbf{w}^T \bar{\mathbf{x}} \leq \check{\phi}^{up}\} \geq \alpha. \quad (8.46)$$

**Proof.** For any  $m$ -vector  $\mathbf{u}$ , consider the reduced random variable

$$\eta = \frac{\mathbf{u}^T \hat{\mathbf{y}} - \mathbf{u}^T \mathbf{K} \bar{\mathbf{x}}}{\sqrt{\mathbf{u}^T \mathbf{S}^2 \mathbf{u}}} \quad (8.47)$$

which follows the standard normal distribution  $n(0, 1)$ . It follows that

$$\Pr\left\{ \left( \mathbf{u}^T \hat{\mathbf{y}} - \kappa \sqrt{\mathbf{u}^T \mathbf{S}^2 \mathbf{u}} \right) \leq \mathbf{u}^T \mathbf{K} \bar{\mathbf{x}} \leq \left( \mathbf{u}^T \hat{\mathbf{y}} + \kappa \sqrt{\mathbf{u}^T \mathbf{S}^2 \mathbf{u}} \right) \right\} = \alpha. \quad (8.48)$$

Define constants  $\beta_1$  and  $\beta_2$  by

$$\begin{aligned} \Pr\left\{ \left( \mathbf{u}^T \hat{\mathbf{y}} - \kappa \sqrt{\mathbf{u}^T \mathbf{S}^2 \mathbf{u}} \right) \leq \mathbf{u}^T \mathbf{K} \bar{\mathbf{x}} \right\} &= \frac{1}{\sqrt{2\pi}} \int_{-\kappa}^{\infty} \exp\left(-\frac{\eta^2}{2}\right) d\eta = \alpha + \beta_2, \\ \Pr\left\{ \mathbf{u}^T \mathbf{K} \bar{\mathbf{x}} \leq \left( \mathbf{u}^T \hat{\mathbf{y}} + \kappa \sqrt{\mathbf{u}^T \mathbf{S}^2 \mathbf{u}} \right) \right\} &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\kappa} \exp\left(-\frac{\eta^2}{2}\right) d\eta = \beta_1 + \alpha, \end{aligned} \quad (8.49)$$

and note that

$$\beta_1 + \alpha + \beta_2 = 1. \quad (8.50)$$

We saw in the preceding section that  $\check{\phi}^{lo}$  and  $\check{\phi}^{up}$  can also be defined as

$$\begin{aligned} \check{\phi}^{lo} &= \max_{\mathbf{u}} \left\{ \mathbf{u}^T \hat{\mathbf{y}} - (\rho_{\min} + \kappa^2)^{\frac{1}{2}} \sqrt{\mathbf{u}^T \mathbf{S}^2 \mathbf{u}} \mid \mathbf{u}^T \mathbf{K} \leq \mathbf{w}^T \right\}, \\ \check{\phi}^{up} &= \min_{\mathbf{u}} \left\{ \mathbf{u}^T \hat{\mathbf{y}} + (\rho_{\min} + \kappa^2)^{\frac{1}{2}} \sqrt{\mathbf{u}^T \mathbf{S}^2 \mathbf{u}} \mid \mathbf{u}^T \mathbf{K} \geq \mathbf{w}^T \right\}. \end{aligned} \quad (8.51)$$

Consider first the upper bound problem and let  $\mathbf{u}_{up}$  be its solution vector, i.e.,

$$\check{\phi}^{up} = \mathbf{u}_{up}^T \hat{\mathbf{y}} + (\rho_{\min} + \kappa^2)^{\frac{1}{2}} \sqrt{\mathbf{u}_{up}^T \mathbf{S}^2 \mathbf{u}_{up}}, \quad (8.52)$$



with

$$\mathbf{u}_{up}^T \mathbf{K} \geq \mathbf{w}^T . \quad (8.53)$$

From the second relation in (8.49), since  $\kappa \leq \sqrt{\rho_{\min} + \kappa^2}$ , it follows that

$$\Pr \left\{ \mathbf{u}_{up}^T \mathbf{K} \bar{\mathbf{x}} \leq \left[ \mathbf{u}_{up}^T \hat{\mathbf{y}} + (\rho_{\min} + \kappa^2)^{\frac{1}{2}} \sqrt{\mathbf{u}_{up}^T \mathbf{S}^2 \mathbf{u}_{up}} \right] \right\} \geq \beta_1 + \alpha , \quad (8.54)$$

or, by (8.52), that

$$\Pr \left\{ \mathbf{u}_{up}^T \mathbf{K} \bar{\mathbf{x}} \leq \check{\phi}^{up} \right\} \geq \beta_1 + \alpha . \quad (8.55)$$

Since  $\bar{\mathbf{x}} \geq \mathbf{o}$ , it follows from (8.53) that

$$\mathbf{w}^T \bar{\mathbf{x}} \leq \mathbf{u}_{up}^T \mathbf{K} \bar{\mathbf{x}} . \quad (8.56)$$

Combining this inequality with (8.55) gives

$$\Pr \left\{ \mathbf{w}^T \bar{\mathbf{x}} \leq \check{\phi}^{up} \right\} \geq \beta_1 + \alpha . \quad (8.57)$$

In a similar manner, it is easy to show that

$$\Pr \left\{ \check{\phi}^{lo} \leq \mathbf{w}^T \bar{\mathbf{x}} \right\} \geq \alpha + \beta_2 . \quad (8.58)$$

Now,

$$\begin{aligned} \Pr \left\{ \check{\phi}^{lo} \leq \mathbf{w}^T \bar{\mathbf{x}} \leq \check{\phi}^{up} \right\} &= \Pr \left[ \left\{ \check{\phi}^{lo} \leq \mathbf{w}^T \bar{\mathbf{x}} \right\} \cap \left\{ \mathbf{w}^T \bar{\mathbf{x}} \leq \check{\phi}^{up} \right\} \right] \\ &= \Pr \left\{ \check{\phi}^{lo} \leq \mathbf{w}^T \bar{\mathbf{x}} \right\} + \Pr \left\{ \mathbf{w}^T \bar{\mathbf{x}} \leq \check{\phi}^{up} \right\} \\ &\quad - \Pr \left[ \left\{ \check{\phi}^{lo} \leq \mathbf{w}^T \bar{\mathbf{x}} \right\} \cup \left\{ \mathbf{w}^T \bar{\mathbf{x}} \leq \check{\phi}^{up} \right\} \right] , \end{aligned} \quad (8.59)$$

so by (8.57) and (8.58),

$$\Pr \left\{ \check{\phi}^{lo} \leq \mathbf{w}^T \bar{\mathbf{x}} \leq \check{\phi}^{up} \right\} \geq \beta_1 + 2\alpha + \beta_2 - \Pr \left[ \left\{ \check{\phi}^{lo} \leq \mathbf{w}^T \bar{\mathbf{x}} \right\} \cup \left\{ \mathbf{w}^T \bar{\mathbf{x}} \leq \check{\phi}^{up} \right\} \right] . \quad (8.60)$$

From definition (8.45), we have  $\check{\phi}^{lo} < \check{\phi}^{up}$ , so

$$\Pr \left[ \left\{ \check{\phi}^{lo} \leq \mathbf{w}^T \bar{\mathbf{x}} \right\} \cup \left\{ \mathbf{w}^T \bar{\mathbf{x}} \leq \check{\phi}^{up} \right\} \right] = 1 , \quad (8.61)$$

which, when substituted into the above expression, gives

$$\Pr \left\{ \check{\phi}^{lo} \leq \mathbf{w}^T \bar{\mathbf{x}} \leq \check{\phi}^{up} \right\} \geq \beta_1 + 2\alpha + \beta_2 - 1 . \quad (8.62)$$

It follows then from (8.50) that

$$\Pr \left\{ \check{\phi}^{lo} \leq \mathbf{w}^T \bar{\mathbf{x}} \leq \check{\phi}^{up} \right\} \geq \alpha , \quad (8.63)$$

which completes the proof of the theorem.  $\square$

Note that the theorem makes no restrictions on  $m$ ,  $n$ , or  $\text{rank}(\mathbf{K})$ . For an underdetermined system in which  $\mathbf{K}$  has some negative elements, it is possible, but not very probable, that one of the two problems (8.45) may be unbounded. In that case, the theorem is still true, but the confidence interval is semi-infinite.

In proving the theorem, we assumed that the errors  $\hat{\mathbf{e}}$  were normally distributed. This is generally not a serious limitation in practice, but it is worth noting that the theorem can be proved for a wider class of possible error distributions. The essential restriction required by the proof is that the reduced random variable  $\eta$  have the same probability distribution for all  $m$ -vectors  $\mathbf{u}$ .

**Corollary 1 .** Let  $\hat{\mathbf{y}}$  be a given  $m$ -vector,  $\mathbf{K}$  a given  $m \times n$  matrix, and  $\bar{\mathbf{x}}$  an unknown  $n$ -vector satisfying

$$\hat{\mathbf{y}} = \mathbf{K}\bar{\mathbf{x}} + \hat{\boldsymbol{\epsilon}} \ , \quad \mathcal{E}(\hat{\boldsymbol{\epsilon}}) = \mathbf{o} \ , \quad \mathcal{E}(\hat{\boldsymbol{\epsilon}}\hat{\boldsymbol{\epsilon}}^T) = \mathbf{S}^2 \quad (8.64)$$

and

$$\Pr \{ \bar{\mathbf{x}} \geq \mathbf{o} \} = 1 \ . \quad (8.65)$$

Let  $\rho_{\min}$  be defined by (8.43) and assume that the errors  $\hat{\boldsymbol{\epsilon}}$  are distributed in such a manner that the reduced random variable

$$\eta = \frac{\mathbf{u}^T \hat{\mathbf{y}} - \mathbf{u}^T \mathbf{K} \bar{\mathbf{x}}}{\sqrt{\mathbf{u}^T \mathbf{S}^2 \mathbf{u}}} \quad (8.66)$$

has the same probability density function  $f(\eta)$  for all  $m$ -vectors  $\mathbf{u}$ . Let  $\alpha$  be a given probability ( $0 < \alpha < 1$ ), and suppose that the value  $\kappa$  is chosen so that

$$\Pr \{ -\kappa \leq \eta \leq \kappa \} = \int_{-\kappa}^{\kappa} f(\eta) d\eta = \alpha \ . \quad (8.67)$$

If  $\mathbf{w}$  is a given  $n$ -vector, and if  $\check{\phi}^{lo}$  and  $\check{\phi}^{up}$  are defined by (8.45), then

$$\Pr \{ \check{\phi}^{lo} \leq \mathbf{w}^T \bar{\mathbf{x}} \leq \check{\phi}^{up} \} \geq \alpha \ . \quad (8.68)$$

**Proof.** The proof differs from that of the Theorem only in the definitions of  $\beta_1$  and  $\beta_2$ , i.e., Eqs. (8.49) are replaced by

$$\begin{aligned} \Pr \left\{ \left( \mathbf{u}^T \hat{\mathbf{y}} - \kappa \sqrt{\mathbf{u}^T \mathbf{S}^2 \mathbf{u}} \right) \leq \mathbf{u}^T \mathbf{K} \bar{\mathbf{x}} \right\} &= \int_{-\kappa}^{\infty} f(\eta) d\eta = \alpha + \beta_2 \ , \\ \Pr \left\{ \mathbf{u}^T \mathbf{K} \bar{\mathbf{x}} \leq \left( \mathbf{u}^T \hat{\mathbf{y}} + \kappa \sqrt{\mathbf{u}^T \mathbf{S}^2 \mathbf{u}} \right) \right\} &= \int_{-\infty}^{+\kappa} f(\eta) d\eta = \beta_1 + \alpha \ . \end{aligned} \quad (8.69)$$

The values  $\beta_1$  and  $\beta_2$  satisfy (8.50), and the remainder of the proof follows exactly as before.  $\square$

## 8.5. Examples

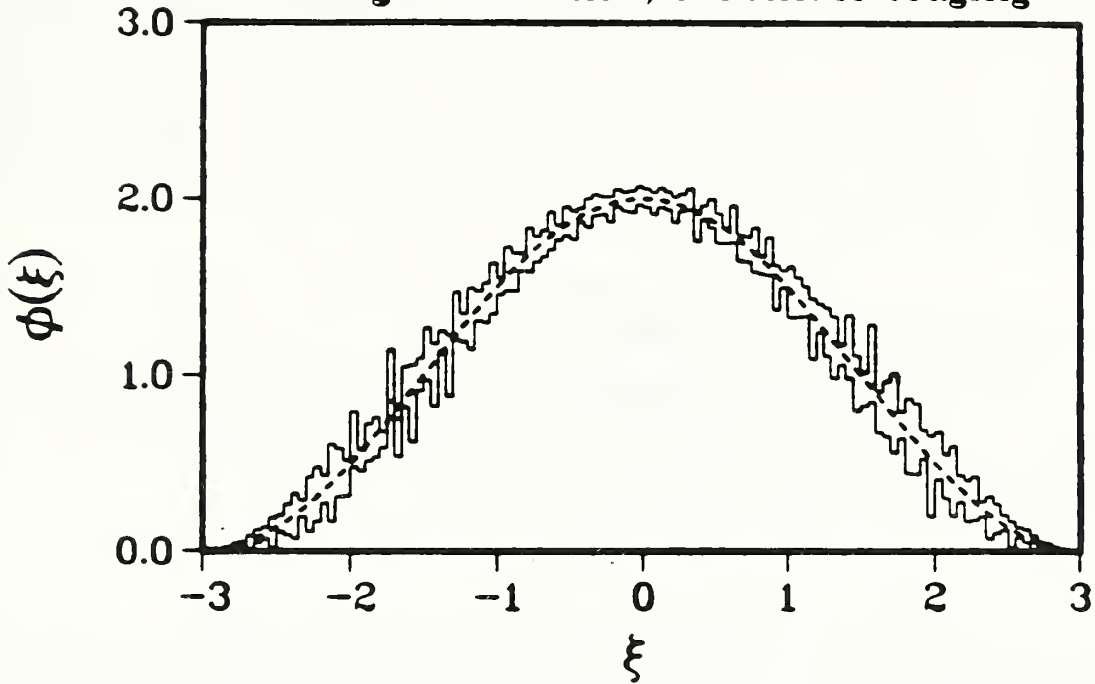
Consider first, confidence intervals based on the normality assumption. Figure 9 gives plots of 95% confidence intervals for both Problems I and II using 2-point averaging. The Problem I bounds can be compared with those in the top frame of Figure 4 in order to assess the improvement to be attributed to the nonnegativity constraints. Problem I is not a bad problem, so the improvement is less striking than that normally attained for real-world problems. The bounds for Problem II do represent a striking improvement since, in the absence of the nonnegativity constraints, all the confidence intervals are unbounded. The bounds for Problem II are unimpressive when compared with those for Problem I, but they could be significantly improved by using 3-point averaging.

## 9. A Real-World Problem

We conclude with a gamma-ray unfolding problem suggested by a series of measurements made by Faddegon, et al. [7]. The data from these experiments have been archived [5] and were made available to us by Dr. Faddegon. In one experiment, a beam of electrons with energies tightly clustered around 15 MeV was incident on the end of a cylinder of lead just thick enough to stop them. A sodium iodide scintillation detector coupled to a photomultiplier tube was used to measure the resulting bremsstrahlung photon spectrum at an angle of  $60^\circ$  from the axis of the cylinder. In this problem, the variable  $\xi$  becomes the energy  $\mathcal{E}$  of the photons incident on the detector, and the subscripts  $i$  become the bin numbers of the pulse height bins into which the photons are counted. The underlying integral equations are

$$N(E_i)dE_i = dE_i \int_{E_0}^{\infty} R(E_i, \mathcal{E}) \mathcal{N}(\mathcal{E}) d\mathcal{E} + \hat{\epsilon}_i \ , \quad i = 1, 2, \dots, m \ , \quad (9.1)$$

150 X 121 Phillips Problem  
 One-at-a-time Confidence Intervals  
 Using Normal Distribution,  $\alpha = 0.95$   
 Nonneg. Constraints, 2-Point Averaging



108 X 121 Phillips Problem  
 One-at-a-time Confidence Intervals  
 Using Normal Distribution,  $\alpha = 0.95$   
 Nonneg. constraints, 2-Point Averaging

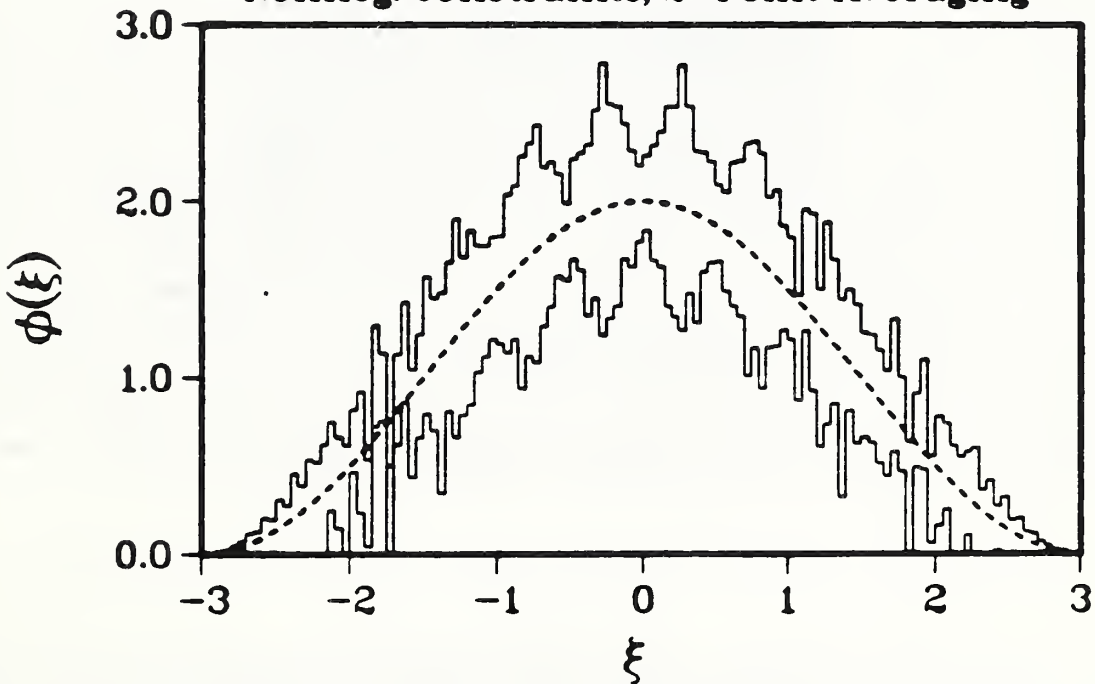


Figure 9: Nonnegatively constrained, one-at-a-time 95% confidence intervals from the normal distribution. Dashed lines are exact solutions  $z(\xi)$ , and solid lines are histograms of estimated lower and upper bounds. Top: Problem I, 2-point averages. Bottom: Problem II, 2-point averages.

where  $E_0$  is the threshold energy for detection,  $E_i$  is the central energy of the  $i$ th pulse-height bin,  $dE_i$  is the width of the pulse-height bin,  $N(E_i)dE_i$  is the total number of photons counted in the bin,  $\mathcal{N}(\mathcal{E})$  is the unknown photon spectrum (  $\mathcal{N}(\mathcal{E})d\mathcal{E}$  is the number of photons in an energy band of width  $d\mathcal{E}$  centered on energy  $\mathcal{E}$  ),  $R(E_i, \mathcal{E})dE_i$  is the instrument response function ( the probability that a photon of energy  $\mathcal{E}$  will produce a count in the  $i$ th bin ), and  $\hat{\epsilon}_i$  is the measurement error. The  $\hat{\epsilon}_i$  were due mostly to counting errors and thus were approximately Poisson distributed, but the counts were also corrected for background radiation and various hardware effects. The details on the calculation of the standard deviations  $s_i$  are given in [7]. We assume that the normal distribution  $n(0, s_i)$  gives a sufficiently accurate approximation to the  $\hat{\epsilon}_i$ -distribution and that the errors were not correlated with one another. Thus,

$$\hat{\epsilon} \sim N(\mathbf{0}, \mathbf{S}^2) , \quad \mathbf{S} = \text{diag}(s_1, s_2, \dots, s_m) , \quad (9.2)$$

Discrete approximations to the detector response functions, i.e.,

$$R_{i,j} = R(E_i, \mathcal{E}_j)dE_i , \quad j = 1, 2, \dots, n , \quad (9.3)$$

were determined by a combination of calculations and high-precision measurements described in [6]. Using those values to approximate the system (9.1) gives

$$N(E_i)dE_i = \sum_{j=1}^n R_{i,j}\mathcal{N}(\mathcal{E}_j)\Delta\mathcal{E}_j + \hat{\epsilon}_i , \quad i = 1, 2, \dots, m , \quad (9.4)$$

where  $\Delta\mathcal{E}_j = \mathcal{E}_j - \mathcal{E}_{j-1}$ . For the particular set of measurements that concerns us here, the values of  $m$  and  $n$  are 56 and 66 respectively. It was necessary to choose  $n > m$  because the response functions, which are shown in Figures 10 and 11, have nonzero values outside the range  $[\mathcal{E}_0, \mathcal{E}_n]$ . Thus, the numbers counted in the lower bins can be affected by photons with energies less than that of the lowest bin and counts in the upper bins can similarly be affected by photons with energies higher than that of the highest bin. These extra columns can be dropped from the response matrix if it is known with certainty that the spectrum of incident photons is zero at the corresponding energies.

The measured spectrum, normalized to unity, is plotted in Figure 12 as a histogram with two horizontal lines plotted for each bin. The top line gives the measured value plus one standard deviation and the bottom line gives the value minus one standard deviation. Since the actual value cannot be negative, it is possible to compute confidence interval estimates for the true spectrum even though the matrix of the system has more columns than rows. In order to compute 95% one-at-a-time confidence intervals for the estimates of the actual value, we took  $\kappa = 1.96$  and applied the BRACKET-LS algorithm, first with two-point averaging and then with no averaging. The top panel of Figure 13 gives the results for two-point averaging. Note that the averaging gave a bin structure similar to that of the measured spectrum. The maximum of the spectrum occurs in the bin containing the .511 *MeV* electron-positron annihilation energy. The pronounced shoulder to the right of this peak, together with the relatively small interval widths encouraged us to repeat the calculation with no averaging. The results are shown in the bottom panel of the Figure. The dark vertical lines represent the calculated confidence intervals, and the dashed line, which connects the non-negatively constrained point estimates, was added only as a visual aid. With the given data, it is not possible to estimate the spectrum at the intermediate points, but the results that were obtained suggest that higher energy resolution would be possible if the response functions could be computed on a denser  $\mathcal{E}$ -mesh. In particular, there is a hint that a smaller peak might be resolved in the vicinity 0.7 *MeV*.

# Sodium Iodide Photon Detector Instrument Response Functions

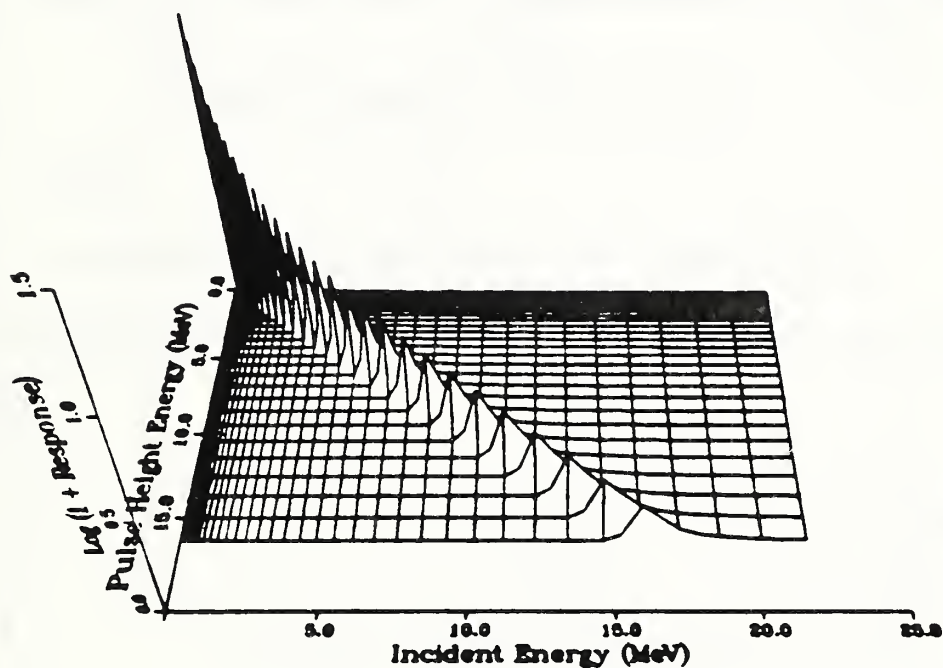
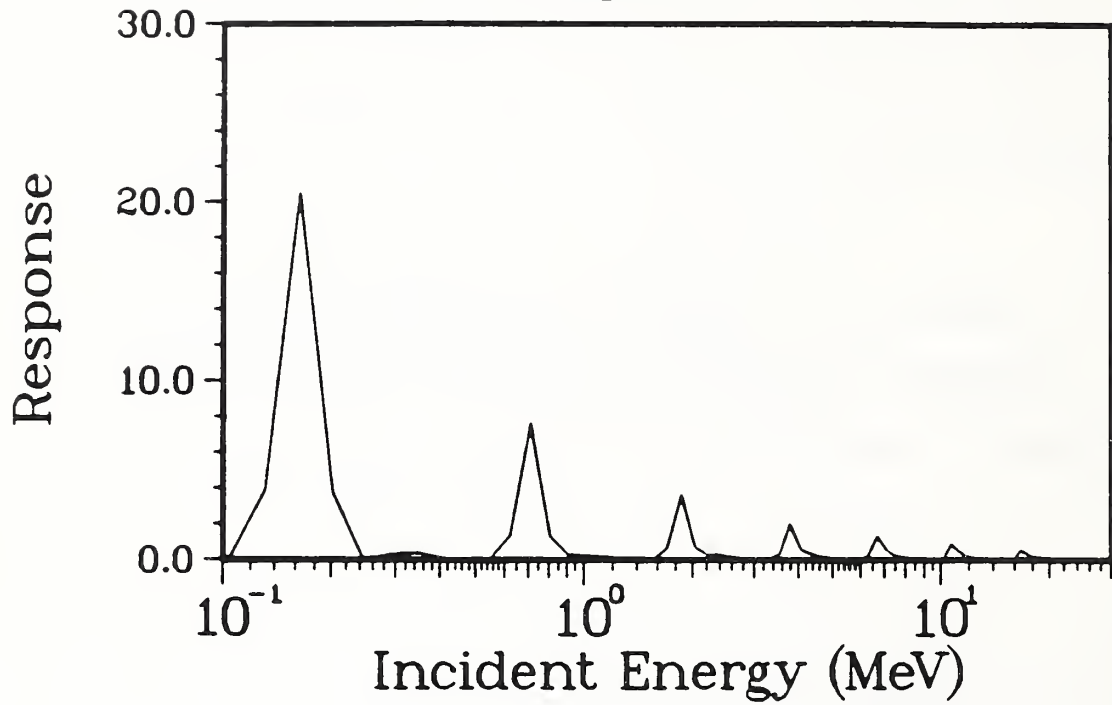


Figure 10: Response functions  $R_{i,j}$  for the NaI gamma-ray detector. For clarity, we have plotted only every 2nd ordinate in each of the abscissa directions. To show the structure at higher energies, we have plotted  $\log_{10}(1 + R_{i,j})$  rather than  $R_{i,j}$  itself. Note the finer mesh spacing and higher energy resolution at the lower energies.



# Sodium Iodide Photon Detector

## Selected Response Functions



## Selected Response Functions

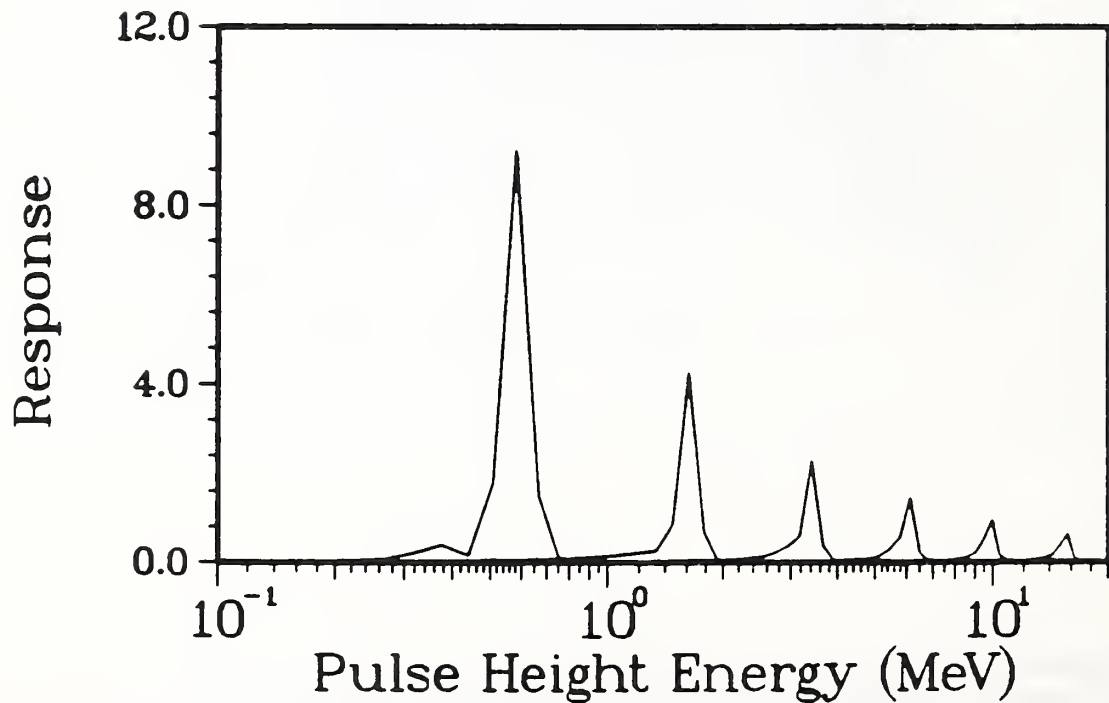


Figure 11: Cross-section views of the response functions  $R_{i,j}$ .

Top: Responses for pulse height bins centered on  $\mathcal{E}_j = 0.182, 0.758, 1.945, 3.93, 6.88, 10.95, 16.96$  MeV.

Bottom: Responses for incident energy bins centered on  $\mathcal{E}_i = 0.546, 1.545, 3.29, 5.95, 9.70, 15.30$  MeV.

Measured Bremsstrahlung Spectrum  
 15 MeV Electrons on Pb Cylinder  
 Photon Yield at 60-deg. Angle

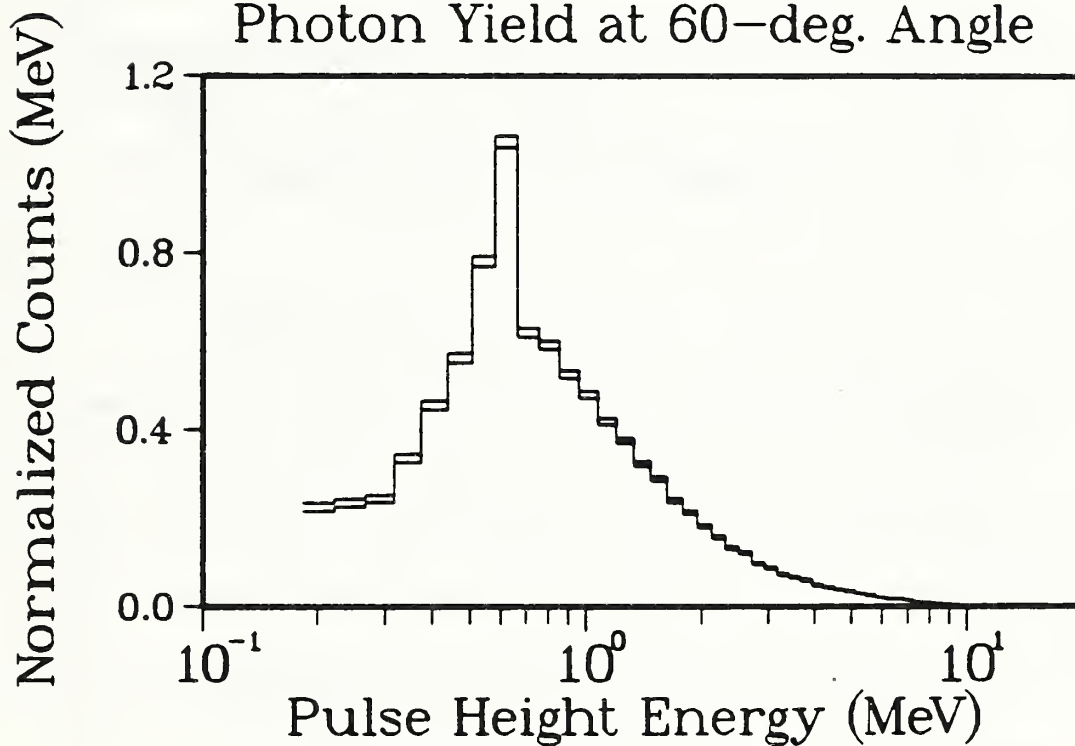


Figure 12: Measured pulse-height spectrum for bremsstrahlung yields produced at  $60^\circ$  from the axis by 15 MeV electrons incident on the end of the Pb cylinder. The two levels plotted in each bin are  $(dS/dE)_i \pm s_i$ .

# Unfolded Bremsstrahlung Spectrum

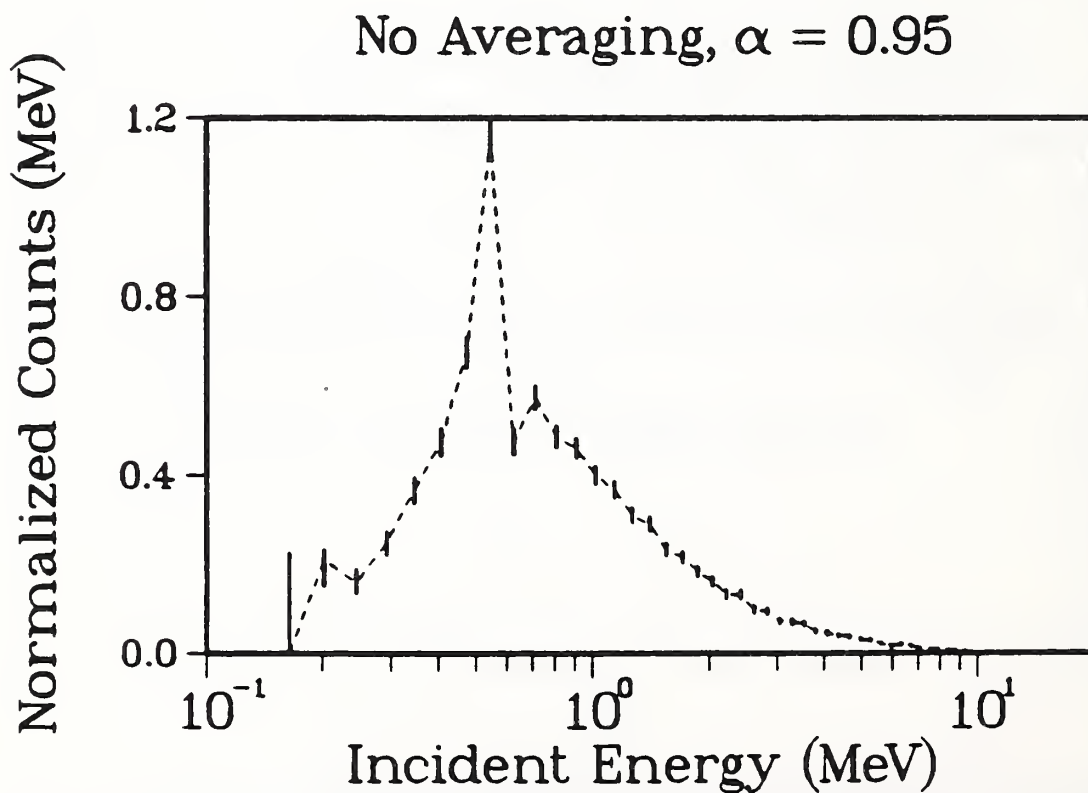
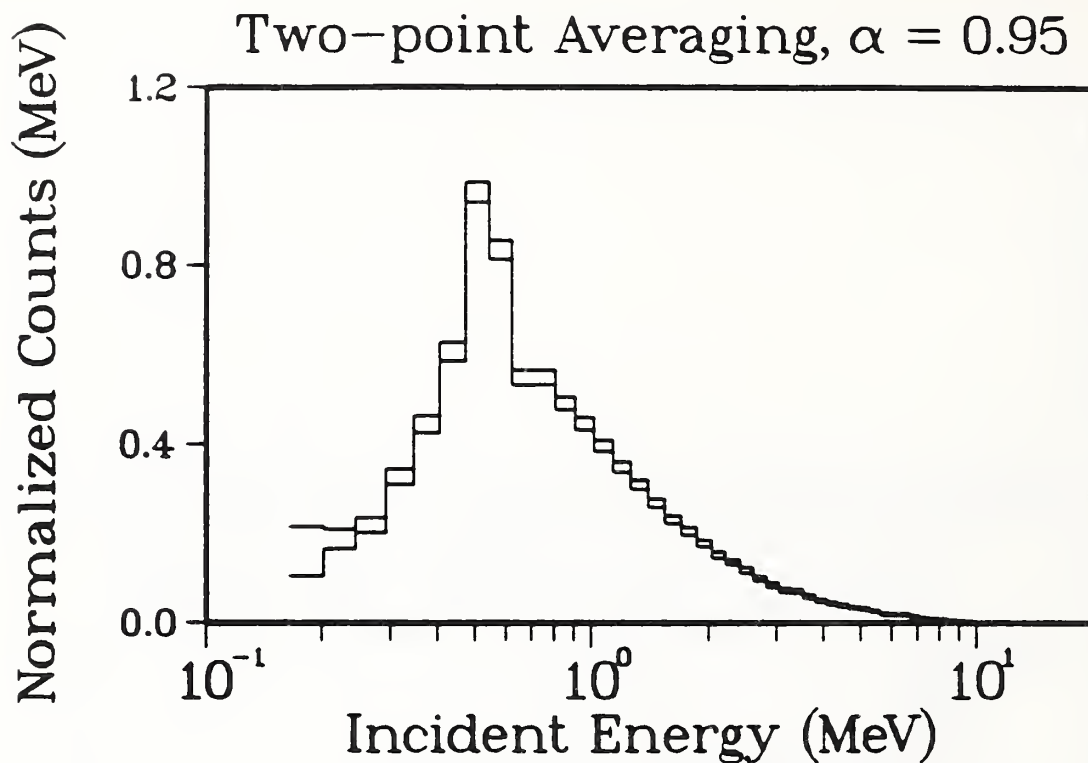


Figure 13: Estimated 95% one-at-a-time confidence intervals for the actual yields  $(dS/dE)_j$ .

Top: Average values over the intervals  $\mathcal{E}_j - \mathcal{E}_{j-1}$ .

Bottom: Confidence intervals for the individual  $(dS/dE)_j$ . The dashed line connects the point estimates.

## References

- [1] W.H. Beyer, editor. *Handbook of Tables for Probability and Statistics*. CRC Press, Inc., Boca Raton, Florida, 1979.
- [2] W. R. Burrus. Utilization of a priori information by means of mathematical programming in the statistical interpretation of measured distributions. Technical Report ORNL-3743, Oak Ridge National Laboratory, Ph.D. Thesis, Ohio State University, June 1965.
- [3] W. R. Burrus, B. W. Rust, and J. E. Cope. Constrained interval estimation for linear models with ill-conditioned equations. In A. L. Schoenstadt et al., editor, *Information Linkage between Applied Mathematics and Industry II*, pages 1–38. Academic Press, New York, 1980.
- [4] J. E. Cope and B. W. Rust. Bounds on solutions of linear systems with inaccurate data. *SIAM J. Numer. Anal.*, 16:950–963, 1979.
- [5] B. A. Faddegon. Megavoltage bremsstrahlung spectroscopy project: Archived software and data. Technical Report PIRS 0244, Institute for National Measurement Standards, National Research Council Canada, 1991.
- [6] B.A. Faddegon, L. Van der Zwan, D.W.O. Rogers, and C.K. Ross. Precision response estimation, energy calibration, and unfolding of spectra measured with a large NaI detector. *Nuclear Instruments and Methods in Physics Research*, A301:138–149, 1991.
- [7] B.A. Faddegon, C.K. Ross, and D.W.O. Rogers. Forward-directed bremsstrahlung of 10- to 30-mev electrons incident on thick targets of Al and Pb. *Medical Physics*, 17:773–785, 1990.
- [8] F.A. Graybill. *Theory and Application of the Linear Model*. Duxbury Press, North Scituate, Massachusetts, 1976.
- [9] R. V. Hogg and A. T. Craig. *Introduction to Mathematical Statistics*. Macmillan, New York, 1965.
- [10] C. L. Lawson and R. J. Hanson. *Solving Least Squares Problems*. Prentice-Hall, Englewood Cliffs, New Jersey, 1974.
- [11] Olvi L. Mangasarian. *Nonlinear Programming*. McGraw-Hill Book Co., New York, 1969.
- [12] D. P. O’Leary and B. W. Rust. Confidence intervals for inequality-constrained least squares problems, with applications to ill-posed problems. *SIAM Jour. Sci. Stat. Comput.*, 7:473–489, 1986.
- [13] Ingram Olkin and J. W. Pratt. A multivariate Tchebycheff inequality. *Annals of Math. Stat.*, 29:226–234, 1958.
- [14] D.L. Phillips. A technique for the numerical solution of certain integral equations of the first kind. *J. Assoc. Comput. Mach.*, 9:84–97, 1962.
- [15] J. E. Pierce and B. W. Rust. Constrained least squares interval estimation. *SIAM Jour. Sci. Stat. Comput.*, 6:670–683, 1985.
- [16] B. W. Rust and W. R. Burrus. Suboptimal methods for solving constrained estimation problems. Technical Report DASA 2604, Tennecomp Systems, Inc., Oak Ridge, Tennessee, 1971.
- [17] B. W. Rust and W. R. Burrus. *Mathematical Programming and the Numerical Solution of Linear Equations*. American Elsevier, New York, 1972.
- [18] H. Scheffé. *The Analysis of Variance*. John Wiley & Sons, New York, 1959.
- [19] A. Tarantola. *Inverse Problem Theory*. Elsevier, Amsterdam, 1987.
- [20] Phillip Wolfe. A duality theorem for non-linear programming. *Quarterly of Appl. Math.*, pages 239–244, 1961.





NIST-114A  
(REV. 3-90)

U.S. DEPARTMENT OF COMMERCE  
NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY

# BIBLIOGRAPHIC DATA SHEET

1. PUBLICATION OR REPORT NUMBER

NISTIR 4720

2. PERFORMING ORGANIZATION REPORT NUMBER

3. PUBLICATION DATE

JANUARY 1992

4. TITLE AND SUBTITLE

Confidence Intervals for Discrete Approximations to Ill-Posed Problems

5. AUTHOR(S)

Bert W. Rust and Dianne P. O'Leary

6. PERFORMING ORGANIZATION (IF JOINT OR OTHER THAN NIST, SEE INSTRUCTIONS)

U.S. DEPARTMENT OF COMMERCE  
NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY  
GAITHERSBURG, MD 20899

7. CONTRACT/GRANT NUMBER

8. TYPE OF REPORT AND PERIOD COVERED

9. SPONSORING ORGANIZATION NAME AND COMPLETE ADDRESS (STREET, CITY, STATE, ZIP)

10. SUPPLEMENTARY NOTES

11. ABSTRACT (A 200-WORD OR LESS FACTUAL SUMMARY OF MOST SIGNIFICANT INFORMATION. IF DOCUMENT INCLUDES A SIGNIFICANT BIBLIOGRAPHY OR LITERATURE SURVEY, MENTION IT HERE.)

We consider the linear regression model obtained by discretizing a system of first-kind integral equations with random measurement errors in the right hand side. The errors are assumed to have zero means and known variances. We consider the problem of estimating confidence intervals for linear functions of the solution vector. For such problems, the least squares solution is a highly unstable function of the measurements, and the classical confidence intervals are too wide to be useful. The solution can often be stabilized by imposing physically motivated, a priori nonnegativity constraints on the solution. This paper will show how to extend the classical confidence interval estimation procedure to accommodate these nonnegativity constraints in order to obtain improved confidence intervals. The technique defines valid confidence intervals even for problems with fewer measurements than unknowns.

12. KEY WORDS (6 TO 12 ENTRIES; ALPHABETICAL ORDER; CAPITALIZE ONLY PROPER NAMES; AND SEPARATE KEY WORDS BY SEMICOLONS)

Burrus conjecture; confidence intervals; first-kind integral equations; ill-posed problems; nonnegativity constraints; radiation spectrum unfolding

13. AVAILABILITY

UNLIMITED

FOR OFFICIAL DISTRIBUTION. DO NOT RELEASE TO NATIONAL TECHNICAL INFORMATION SERVICE (NTIS).

ORDER FROM SUPERINTENDENT OF DOCUMENTS, U.S. GOVERNMENT PRINTING OFFICE,  
WASHINGTON, DC 20402.

ORDER FROM NATIONAL TECHNICAL INFORMATION SERVICE (NTIS), SPRINGFIELD, VA 22161.

14. NUMBER OF PRINTED PAGES

40

15. PRICE

A03

ELECTRONIC FORM





