Solving Quadratically Constrained Least Squares Using Black Box Unconstrained Solvers

Tony F. Chan*
Dept. of Mathematics, University of California,
Los Angeles, California.

Julia A. Olkin and Donald W. Cooley
Remote Measurements Lab, SRI International,
Menlo Park, California.

September 1, 1989

Abstract We present algorithms for solving quadratically constrained linear least squares problems that only require calls to a black box unconstrained solver. One of our goals is to exploit special structures in the unconstrained problem which could not be used easily in the constrained problem. Our approach is to solve for the Lagrange multiplier as the root of a secular equation. We compare a linear and a rational (Hebden) local model and a Newton and secant method. We also derive a formula for estimating the Lagrange multiplier which depends on the amount the unconstrained solution violates the constraint and an estimate of the smallest singular value of the coefficient matrix A. The estimate can be used as a good initial guess for solving the secular equation. We also show conditions under which the estimate is guaranteed to be an acceptable solution without further refinement.

1 Introduction

The most general form of the problem we wish to consider is a least squares (LS) minimization with a quadratic inequality constraint — the LSQI problem — which arises whenever the solution to the ordinary least squares

*Research supported by SRI International and by the National Science Foundation under grant DMS-87-14612.

problem needs to be regularized. We define a simplified version of the LSQI problem as

$$\begin{aligned} &\text{minimize} \quad \|Ax - b\|_2^2 \\ &\text{subject to} \quad \|Bx\|_2^2 \leq c, \end{aligned} \tag{1}$$

where $A \in R^{m\mathrm{x}n}$, $b \in R^m$, $B \in R^{p\mathrm{x}n}$(full rank) and $c \geq 0$. For the remainder of this paper we will refer collectively to the objective function and constraint in equation (1) as Problem 1. Henceforth, all norms refer to the l_2 norms unless otherwise specified.

A subset of the LSQI problem occurs when A factors into the product of two rectangular matrices H and G, so that $A = HG$. A typical scenario from signal processing systems is shown in Figure 1.1.

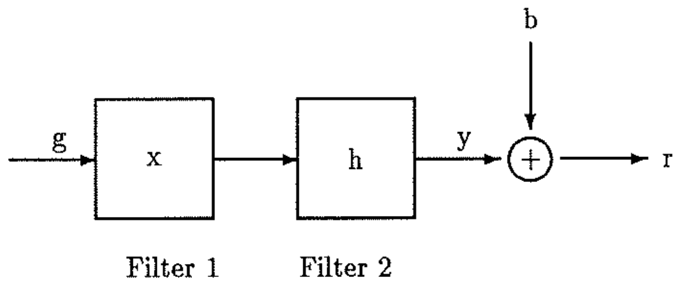

Figure 1.1: Single Channel Block Diagram

The output of the system is the convolution

$$y = h * x * g = h * g * x$$

or, in matrix form,

$$y = HGx.$$

For the single channel system shown in Figure 1.1, both matrices G and H are column circulant and rectangular, with their first columns respectively composed of the input (g) and filter 2 impulse response (h) suitably padded with zeros. The weights x of filter 1 are chosen to match the system output y to a known desired response b, that is, we find x by minimizing the residual r in

$$\min \|r\|^2 = \|HGx - b\|_2^2.$$

It is often necessary to restrict the norm of the input to filter 2, Gx, to ensure that the filter operates in its linear region. Hence the constraint takes the form $\|Gx\|_2^2 \le c$. In some events the matrix B in Problem (1) is actually one of the factors of the matrix A, thusly,

$$\begin{aligned} &\text{minimize} \quad \|HGx - b\|_2^2 \\ &\text{subject to} \quad \|Gx\|_2^2 \le c, \end{aligned} \tag{2}$$

referred to as Problem 2.

Our main motivation is to solve Problem 1 in $\mathcal{O}(n^2)$ time rather than the usual $\mathcal{O}(n^3)$ time. Our hope lies in the fact that it is easier to exploit the structure in A in the unconstrained problem than in the constrained problem (e.g., the matrix B may not have any structure). For instance, if A is Toeplitz, then the unconstrained problem could be solved in $\mathcal{O}(n^2)$ time using a fast Toeplitz solver. The idea, then, is to treat the solution to the unconstrained problem as a black box solver which will be called as needed to solve the constrained problem. Before we present shortcuts, though, we need to discuss how to solve Problem 1 accurately without regard to work involved.

If $\min \|Bx\|^2 \le c$ then the unconstrained solution to Problem 1 is feasible and we are finished. Otherwise, the solution satisfies the generalized normal equation,

$$(A^TA + \lambda B^TB)x = A^Tb,$$

where λ is chosen such that $\|Bx\|^2 = c$.

We can write the solution to Problem (1) with the aid of the generalized singular value decomposition (GSVD) of A and B [Gvl]. Let

$$\begin{aligned} &U^TAX = D_A = diag(\alpha_1, ..., \alpha_n), &&U^TU = I_m \\ &V^TBX = D_B = diag(\beta_1, ..., \beta_q), &&V^TV = I_p,\ q = \min\{p, n\} \end{aligned}$$

be the GSVD of A and B. Then Problem (1) transforms to

$$\begin{aligned} &\text{minimize} \quad \|D_Ay - \tilde{b}\|_2^2 \\ &\text{subject to} \quad \|D_By\|_2^2 \le c, \end{aligned}$$

where $\tilde{b} = U^Tb$ and $y = X^{-1}x$. The vector $y \in R^n$ defined by

$$y_i = \begin{cases} \tilde{b}_i/\alpha_i & \alpha_i \ne 0 \\ 0 & \alpha_i = 0 \end{cases} \quad i = 1, ..., n$$

minimizes $\|D_A y - \tilde{b}\|_2^2$. If y is also feasible then we have a solution to Problem (1). Otherwise we assume that

$$\sum_{\substack{i=1 \\ \alpha_i \neq 0}}^{q} \left(\beta_i \frac{\tilde{b}_i}{\alpha_i} \right)^2 > c, \tag{3}$$

thus implying that the solution to Problem (1) occurs on the boundary of the feasible set. We must now solve the LSQE problem,

$$\begin{aligned} &\text{minimize} && \|D_A y - \tilde{b}\|_2^2 \\ &\text{subject to} && \|D_B y\|_2^2 = c. \end{aligned} \tag{4}$$

Using Lagrange multiplier theory it is evident that the solution vector y as a function of the Lagrange multiplier λ has as its i^{th} component

$$y_i(\lambda) = \frac{\alpha_i \tilde{b}_i}{\alpha_i^2 + \lambda \beta_i^2}.$$

The parameter λ is chosen so that

$$\phi(\lambda) \equiv \|D_B y(\lambda)\|_2^2 = \sum_{i=1}^{r} \left(\frac{\alpha_i \beta_i \tilde{b}_i}{\alpha_i^2 + \lambda \beta_i^2} \right)^2 = c, \tag{5}$$

where $r = \text{rank}(B)$. Equation (5) is sometimes referred to as the *secular* equation.

Now, $\phi(0) > c$ from (3). Since $\phi(\lambda)$ is monotone decreasing for $\lambda > 0$, (3) implies the existence of a unique positive λ^* for which $\phi(\lambda^*) = c$. It can be shown that λ^* is the desired root [Gan]. Any root-finding technique can be applied to (5) to find λ^*, and then the solution to Problem (1) is $x = Xy(\lambda^*)$.

The derivation of the secular equation for Problem (2) closely resembles that for (1). However, because of the special form of B, only a SVD (and not a GSVD) is required. In particular, we first transform Problem (2) into a canonical form. Compute the QR factorization of G as

$$G = QR$$

and the SVD of HQ as

$$HQ = UJV^T,$$

where Q, U and V are orthogonal matrices, R is upper triangular and $J \equiv diag(\sigma_1, ...\sigma_n)$. Then, by defining $\tilde{b} = U^T b$ and $y = V^T R x$, Problem (2) is transformed to

$$\text{minimize} \quad \|Jy - \tilde{b}\|_2^2 \qquad\qquad (6)$$

$$\text{subject to} \quad \|y\|_2^2 \leq c.$$

The solution vector y is obtained by forming the Lagrangian of (6), taking its derivative with respect to y and λ, and simultaneously solving the two equations. Then y can be seen to solve the generalized normal equation,

$$\left(J^T J + \lambda I\right) y = J^T \tilde{b},$$

where again λ is chosen so that

$$\|y(\lambda)\|^2 \equiv \sum_{i=1}^{n} \frac{\sigma_i^2 \tilde{b}_i^2}{(\sigma_i^2 + \lambda)^2} = c. \qquad (7)$$

Note that the *unconstrained* solution y_u, at $\lambda = 0$, is

$$(y_u)_i = \frac{\tilde{b}_i}{\sigma_i}, \; i = 1, 2, ..., n$$

and therefore

$$c_u \equiv \|y_u\|^2 = \sum_{i=1}^{n} \frac{\tilde{b}_i^2}{\sigma_i^2}. \qquad (8)$$

One of the drawbacks of solving either form of the secular equation, (5) or (7), is that either a GSVD or SVD is required. Eldén considers using QR for solving the LSQI problem, but this still requires matrix factorization. Not only is this expensive, but any special structure in A, H or G will be lost in the decomposition. In some applications the matrices are large, Toeplitz or block-Toeplitz, and ill-conditioned. In such cases, the secular equation is not known explicitly. We must implicitly solve $\|y(\lambda)\|_2^2 = c$. Thus, we need a good estimate of λ without matrix factorization, and we need a good iterative method that improves on the estimate. For a more complete analysis of constrained least squares, including the existence and uniqueness of a solution, we refer the reader to [Gan] and [FG]. For a general discussion on the LSQI problem, refer to [Gvl] or [Bjo].

In section 2 we compare a linear versus a rational (Hebden) local model, and a Newton versus a secant method. In section 3 we derive an estimate for the optimal positive Lagrange multiplier which satisfies (7). This estimate is used as a first guess for various root-finders. Section 4 provides some theoretical results justifying this use of this estimate and gives conditions under which the estimate is satisfactory without refinements. Numerical results are included in section 5, followed by some concluding remarks in section 6.

2 Root-Finding Techniques

In this section we discuss various iterative methods used to solve for the unique positive root λ of

$$f(\lambda) \equiv \|Bx(\lambda)\|_2 - \sqrt{c} = 0, \tag{9}$$

where $x(\lambda)$ is the solution

$$x(\lambda) = \left(A^T A + \lambda B^T B\right)^{-1} A^T b. \tag{10}$$

To construct an iterative numerical algorithm for finding the root of (9) we choose two forms for the local model of the nonlinear function $f(\lambda)$, the usual linear model, and a rational model ([Heb] and [BNS]). For fitting the model to the current iterate, we use two forms of interpolation: (1) fitting the values of the function and its derivative at one point (à la Newton), and (2) secant interpolation. Then the next iterate is defined as the zero of the local model using a form of interpolation. Table 2.1 clarifies the four possibilities provided (two local models and two forms of interpolation).

Model	Interpolation	
	Point/Tangent	Two Points
Hebden	Hebnew	Hebsec
Linear	Newton	Secant

Table 2.1: Naming Convention for Iterative Methods

The four methods fall into the following general algorithmic framework:

Given initial guess λ_0
DO until convergence
 IF (method = *hebsec*)
  calculate λ_{n+1}
 ELSE
  calculate λ_{step}
  $\lambda_{n+1} = \lambda_n + \lambda_{step}$
 END
END DO

To determine λ_{step} for the Newton methods we need the first derivative of $\|Bx(\lambda)\|$ which, as it turns out, can be obtained by an extra matrix solution with the same coefficient matrix. We follow the derivation given in [Heb]. The trick is to write the derivative of $\|Bx(\lambda)\|^2$ in two different forms, and equate the two forms. Thus we have

$$\frac{\partial}{\partial\lambda}\|Bx(\lambda)\|^2 = 2\|Bx(\lambda)\|\frac{\partial}{\partial\lambda}\|Bx(\lambda)\| \tag{11}$$

and

$$\frac{\partial}{\partial\lambda}\|Bx(\lambda)\|^2 = \frac{\partial}{\partial\lambda}\left[(Bx(\lambda))^T Bx(\lambda)\right] = 2x(\lambda)^T B^T \frac{\partial}{\partial\lambda} Bx(\lambda). \tag{12}$$

Equating (11) and (12) gives

$$\frac{\partial}{\partial\lambda}\|Bx(\lambda)\| = \frac{x(\lambda)^T B^T B \frac{\partial}{\partial\lambda} x(\lambda)}{\|Bx(\lambda)\|}. \tag{13}$$

Now,

$$\begin{aligned} v \equiv \frac{\partial}{\partial\lambda}x(\lambda) &= -\left(A^TA + \lambda B^TB\right)^{-1} B^TB\left(A^TA + \lambda B^TB\right)^{-1} A^Tb \\ &= -\left(A^TA + \lambda B^TB\right)^{-1} B^TBx(\lambda), \end{aligned}$$

where we have used the relationship

$$\frac{\partial}{\partial\lambda}C^{-1}(\lambda) = -C^{-1}(\lambda)\left(\frac{\partial}{\partial\lambda}C(\lambda)\right)C^{-1}(\lambda).$$

The value of $v = \frac{\partial}{\partial\lambda}x(\lambda)$ in (13) is found by solving two systems of equations with the same coefficient matrix:

$$\left(A^TA + \lambda B^TB\right)x = A^Tb \tag{14}$$

and

$$\left(A^T A + \lambda B^T B\right) v = -B^T B x. \tag{15}$$

All the pieces to proceed are now in place to compute the function $\|Bx(\lambda)\|$ and its derivative, by using (14) in (15) to calculate $\frac{\partial}{\partial\lambda} x(\lambda)$.

The step in λ for Newton's method applied to $f(\lambda) = 0$ is given by

$$\lambda^n_{step} = \frac{(\|Bx(\lambda_c)\| - \sqrt{c})\|Bx(\lambda_c)\|}{-x(\lambda_c)^T B^T B v}.$$

The form of $\|Bx(\lambda)\|$ is $\sum \left(\sigma_i^2 \tilde{b}_i^2 / \sigma_i^2 + \lambda\right)$, so it may appear that the linear model may not be the best. Several people have suggested using the rational approximation ([Heb] and [Rei])

$$\|Bx(\lambda)\| = \frac{a}{b+\lambda}.$$

Then

$$\begin{aligned} \frac{\partial}{\partial\lambda}\|Bx(\lambda)\| &= \frac{-a}{(b+\lambda)^2} \\ &= \frac{-a}{b+\lambda}\,\frac{1}{b+\lambda} \\ &= -\|Bx(\lambda)\|\frac{1}{b+\lambda}. \end{aligned}$$

Thus,

$$\frac{1}{b+\lambda} = -\frac{\frac{\partial}{\partial\lambda}\|Bx(\lambda)\|}{\|Bx(\lambda)\|}.$$

Solving the model for $\delta\lambda$ we get

$$\|Bx(\lambda + \delta\lambda)\| = \frac{a}{b+\lambda+\delta\lambda} = \sqrt{c}$$

This is achieved by

$$\begin{aligned} \delta\lambda &= \tfrac{a-\sqrt{c}(b+\lambda)}{\sqrt{c}} \\ &= -\left(\tfrac{\|Bx(\lambda)\|-\sqrt{c}}{\sqrt{c}}\right)\left(\|Bx(\lambda)\| / \tfrac{\partial}{\partial\lambda}\|Bx(\lambda)\|\right) \qquad (16) \\ &= -\left(\tfrac{\|Bx(\lambda)\|}{\sqrt{c}} - 1\right)\left(\|Bx(\lambda)\|^2 / x(\lambda)^T B^T B v\right). \end{aligned}$$

The step in λ for *Hebnew* is given by the last equality in (16),

$$\lambda^{hn}_{step} = \frac{(\|Bx(\lambda_c)\| - \sqrt{c})\|Bx(\lambda_c)\|^2}{-x(\lambda_c)^T B^T B v \sqrt{c}}.$$

The Newton methods, which require derivative information, need two black box solves per step. Thus, an advantage to secant methods, which use a finite difference approximation to $f'(\lambda_c)$, is that fewer matrix solutions may be required. The iterations are started with two points, for example $\lambda = 0$ and $\lambda = \lambda^*$ (a good estimate λ^* for the Lagrange multiplier will be derived in the next section). The step for *Secant* is

$$\lambda^{s}_{step} = -\frac{(\|Bx(\lambda_{new})\| - \sqrt{c})\,(\lambda_{old} - \lambda_{new})}{\|Bx(\lambda_{new})\| - \|Bx(\lambda_{old})\|}.$$

For the rational model *Hebsec*, the term $\|Bx(\lambda)\|$ is approximated by $a/(b+\lambda)$. Matching this at two points gives

$$a = \frac{\|Bx(\lambda_{old})\|\|Bx(\lambda_{new})\|(\lambda_{old} - \lambda_{new})}{\|Bx(\lambda_{new})\| - \|Bx(\lambda_{old})\|},$$
$$b = \frac{\lambda_{old}\|Bx(\lambda_{old})\| - \lambda_{new}\|Bx(\lambda_{new})\|}{\|Bx(\lambda_{new})\| - \|Bx(\lambda_{old})\|}.$$

Equating $a/(b+\lambda) = \sqrt{c}$ gives

$$\lambda^{hs}_{new} = \frac{a}{\sqrt{c}} - b$$

All four methods run until

$$\Big|\|Bx\| - c\Big| < tol * c,$$

for a user-specified tolerance *tol*.

3 A Lagrange Multiplier Estimate

The secular equation is a scalar nonlinear equation in λ and in general has no explicit solution. Its solution therefore must require an iterative procedure, starting from an appropriate initial guess. If $\|y_u\| > c$, then it is easy to show that there must exist a unique positive solution λ. Without any further knowledge of the problem, the value $\lambda = 0$ is often used as an initial guess for the iterative procedure. In this section, we derive an estimate of the solution of the secular equation and prove (in the next section) that it is better than $\lambda = 0$ for many problems.

Let $\sigma_i = \alpha_i/\beta_i$ be the generalized singular values of A and B in Problem 1, or the singular values of HQ in Problem 2. Assume that these singular values are ordered in the usual convention,

$$\sigma_n \leq \sigma_{n-1} \ldots \leq \sigma_1.$$

Consider the regularization of an ill-conditioned problem for which the constraint is violated due to the occurrence of a small singular value. Suppose that $\sigma_n \ll \sigma_{n-1}$ and $\tilde{b}_n \neq 0$. Then

$$c_u \approx \left(\frac{\tilde{b}_n^2}{\sigma_n^2}\right) \gg \frac{\tilde{b}_i^2}{\sigma_i^2} \qquad \forall i < n.$$

Therefore, for small enough λ,

$$\frac{\sigma_i^2 \tilde{b}_i^2}{(\sigma_i^2 + \lambda)^2} \ll \frac{\sigma_n^2 \tilde{b}_n^2}{(\sigma_n^2 + \lambda)^2} \qquad \forall i < n.$$

Then

$$c \approx \frac{\sigma_n^2 \tilde{b}_n^2}{(\sigma_n^2 + \lambda)^2} \approx \frac{\sigma_n^4 c_u}{(\sigma_n^2 + \lambda)^2}.$$

Solving for λ yields our estimate $\hat{\lambda}$

$$\lambda_{true} \approx \sigma_n^2 \left(\sqrt{\frac{c_u}{c}} - 1\right) \equiv \hat{\lambda}. \tag{17}$$

Note that, in practice, σ_n must be estimated by some procedure, such as inverse iteration. This leads to the following procedure:

1. Solve unconstrained problem to obtain y_u and compute $c_u = \|y_u\|^2$.

2. If $c_u \leq c$ we are finished. Otherwise,

3. Estimate σ_n and set $\hat{\lambda} = \sigma_n^2 \left(\sqrt{\frac{c_u}{c}} - 1\right)$

4. Solve the regularized least squares problem:

$$\min_y \left\| \begin{pmatrix} J \\ \sqrt{\lambda} I \end{pmatrix} y - \begin{pmatrix} \tilde{b} \\ 0 \end{pmatrix} \right\|_2^2,$$

or the equivalent problem:

$$\min_x \left\| \begin{pmatrix} A \\ \sqrt{\lambda} B \end{pmatrix} x - \begin{pmatrix} b \\ 0 \end{pmatrix} \right\|_2^2. \tag{18}$$

We note that this procedure can be applied even if J is bidiagonal or triangular instead of diagonal and hence it is not necessary to compute the SVD of HQ or GSVD of A and B. In fact, the procedure can be applied to any "black box" solver for the original unconstrained problem involving H and G. We shall pursue this further in a subsequent section when we discuss Hebden's method.

This section closes with a simple example to illustrate the usefulness of the Lagrange multiplier estimate. Consider a problem with:

$$J = \begin{pmatrix} 1 & 0 \\ 0 & \epsilon \\ 0 & 0 \end{pmatrix}, \qquad \tilde{b} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \epsilon \ll 1.$$

$$y_u = \begin{pmatrix} 1 \\ \epsilon^{-1} \end{pmatrix} \qquad c_u = \|y_u\|^2 = 1 + \epsilon^{-2}$$

Suppose $0 < c < c_u$ and c is given. Then, for small ϵ,

$$\hat{\lambda} = \epsilon^2 \left(\sqrt{\frac{1+\epsilon^{-2}}{c}} - 1 \right) \approx \frac{\epsilon}{\sqrt{c}}.$$

Therefore,

$$\begin{aligned} y(\hat{\lambda}) &= (J^T J + \hat{\lambda} I)^{-1} J^T \tilde{b} \\ &\cong \begin{pmatrix} 1 + \frac{\epsilon}{\sqrt{c}} & 0 \\ 0 & \epsilon^2 + \frac{\epsilon}{\sqrt{c}} \end{pmatrix}^{-1} \begin{pmatrix} 1 \\ \epsilon \end{pmatrix} = \begin{pmatrix} \frac{1}{1+\frac{\epsilon}{\sqrt{c}}} \\ \frac{1}{\epsilon + \frac{1}{\sqrt{c}}} \end{pmatrix} \approx \begin{pmatrix} 1 \\ \sqrt{c} \end{pmatrix}, \end{aligned}$$

so

$$\frac{\|y(\hat{\lambda})\|_2^2}{c} = 1 + \frac{1}{c}.$$

For $c \gg 1$, this says that the solution with the Lagrange multiplier estimate comes very close to satisfying the constraint exactly.

An interesting observation is that our Lagrange multiplier estimate can be viewed as applying one iteration of *Hebnew* with a starting point of $\lambda_0 = 0$

to (1) with $B \equiv I$. Recall that at $\lambda = 0$ we have the unconstrained solution c_u, so that

$$\begin{aligned} \lambda_1^{hn} &= \frac{(\|x(\lambda_0)\| - \sqrt{c})\,\|x(\lambda_0)\|^2}{x(\lambda_0)^T\left(A^TA + \lambda_0 I\right)^{-1} x(\lambda_0)\sqrt{c}} \\ &= \left(\sqrt{\frac{c_u}{c}} - 1\right)\frac{x(0)^T x(0)}{x(0)^T(A^TA)^{-1}x(0)} \\ &= \left(\sqrt{\frac{c_u}{c}} - 1\right)\sigma^2. \end{aligned}$$

Generally we have the inequality

$$\sigma_n^2 \leq \frac{x^T x}{x^T(A^TA)^{-1}x} \leq \sigma_1^2$$

and σ^2 can be interpreted as an inverse Rayleigh-Ritz approximation to σ_n^2. With this approximation the expression for λ_1^{hn} becomes the same as λ^* in (17).

4 Theoretical Results For Our Estimate

In this section, we derive some theoretical results justifying the use of the Lagrange multiplier estimate derived in the previous section.

First we prove that $\hat{\lambda}$ is a lower bound for λ^* or, equivalently, that $\|y(\hat{\lambda})\|^2$ always overestimates c.

Theorem 4.1

$$\hat{\lambda} \equiv \sigma_n^2\left(\sqrt{\frac{c_u}{c}} - 1\right) \leq \lambda^* \leq \sigma_1^2\left(\sqrt{\frac{c_u}{c}} - 1\right) \quad \textit{and} \quad \|y(\hat{\lambda})\|^2 \geq c$$

Proof. Rewrite (7) as follows:

$$\sum_{i=1}^{n} \frac{\tilde{b}_i^2}{\sigma_i^2\left(1 + \dfrac{\lambda}{\sigma_i^2}\right)^2} = c.$$

Now

$$\frac{1}{\max\limits_{1\leq i\leq n}\left(1 + \frac{\lambda}{\sigma_i^2}\right)^2}\sum_{i=1}^{n}\frac{\tilde{b}_i^2}{\sigma_i^2} \leq \sum_{i=1}^{n}\frac{\tilde{b}_i^2}{\sigma_i^2\left(1 + \frac{\lambda}{\sigma_i^2}\right)^2} \leq \frac{1}{\min\limits_{1\leq i\leq n}\left(1 + \frac{\lambda}{\sigma_i^2}\right)^2}\sum_{i=1}^{n}\frac{\tilde{b}_i^2}{\sigma_i^2}.$$

Therefore,

$$\frac{c_u}{\left(1+\frac{\lambda}{\sigma_n^2}\right)^2} \le c \le \frac{c_u}{\left(1+\frac{\lambda}{\sigma_1^2}\right)^2}.$$

Solving for λ gives the desired result [Gol]. Also,

$$\|y(\hat{\lambda})\|^2 = \sum_{i=1}^{n} \frac{\tilde{b}_i^2}{\sigma_i^2} \frac{1}{(1+\frac{\hat{\lambda}}{\sigma_i^2})^2} \ge \frac{1}{(1+\frac{\hat{\lambda}}{\sigma_n^2})^2} \sum_{i=1}^{n} \frac{\tilde{b}_i^2}{\sigma_i^2} = c. \qquad \square$$

One implication of this result is that if we start Newton's method with $\hat{\lambda}$ or the secant method with $\lambda = 0$ and $\hat{\lambda}$, then we are guaranteed to have monotonic convergence towards λ^*. This follows from the convexity of $\|y(\lambda)\|^2$.

Next we shall show that $\|y(\hat{\lambda})\|^2$ cannot be too much larger than c. Define

$$g(\sigma) \equiv \left(\sigma + \frac{\hat{\lambda}}{\sigma}\right)^2.$$

Note that

$$\|y(\hat{\lambda})\|^2 = \sum_{i=1}^{n} \frac{\tilde{b}_i^2}{g(\sigma_i)}.$$

It can be easily verified that $g(\sigma)$ has a unique minimum (for $\sigma > 0$) at $\sigma = \sigma^*$ given by

$$0 = g'(\sigma^*) = 2\left(\sigma^* + \frac{\hat{\lambda}}{\sigma^*}\right)\left(1 - \frac{\hat{\lambda}}{(\sigma^*)^2}\right)$$

or

$$\sigma^* = \sqrt{\hat{\lambda}}.$$

At the minimum

$$g(\sigma^*) = \left(\sqrt{\hat{\lambda}} + \frac{\hat{\lambda}}{\sqrt{\hat{\lambda}}}\right)^2 = 4\hat{\lambda}.$$

At $\sigma = \sigma_n$, we have

$$g(\sigma_n) = \left(\sigma_n + \sigma_n(\sqrt{\frac{c_u}{c}} - 1)\right)^2 = \sigma_n^2 \, \frac{c_u}{c}.$$

Define $\bar{\sigma}$ by $g(\bar{\sigma}) = g(\sigma_n)$, $\bar{\sigma} \ne \sigma_n$. Then we find

$$\bar{\sigma} + \frac{\hat{\lambda}}{\bar{\sigma}} = \sigma_n + \frac{\hat{\lambda}}{\sigma_n}$$

or

$$\bar{\sigma} - \sigma_n = \hat{\lambda}\left(\frac{\bar{\sigma} - \sigma_n}{\sigma_n\,\bar{\sigma}}\right)$$

or

$$\bar{\sigma} = \frac{\hat{\lambda}}{\sigma_n} = \sigma_n(\sqrt{\frac{c_u}{c}} - 1).$$

Now define $\gamma \equiv \sqrt{\frac{c_u}{c}} - 1$. Note that $\gamma > 0$, and $\sigma^* \equiv \sigma_n\sqrt{\gamma}$ and $\bar{\sigma} \equiv \sigma_n\,\gamma$.

The situation is summarized in Figure 4.2 for both the cases of $\gamma > 1$ and $\gamma \leq 1$.

Generally speaking, if any singular values lie near the minimum σ^*, (especially in the interval $(\sigma_n, \sigma_n\gamma)$), then $\|y(\hat{\lambda})\|^2$ could be much larger than c. Since $\sigma_1 \geq \sigma_2 \geq \sigma_{n-1} \geq \sigma_n$, we see that this situation is only possible if $\gamma > 1$. We shall prove this in the next theorem. First define

$$\rho \equiv \frac{\|\tilde{b}\|^2}{\tilde{b}_n^2} \geq 1.$$

Theorem 4.2

IF	*Case 1.*	$\gamma \leq 1$	*then* $\frac{\|y(\hat{\lambda})\|^2}{c} \leq \rho$
ELSEIF	*Case 2.*	$\gamma > 1$ $\sigma_{n-1} \geq \bar{\sigma} \equiv \sigma_n\gamma$	*then* $\frac{\|y(\hat{\lambda})\|^2}{c} \leq \rho$
ELSE	*Case 3.*	$\gamma > 1$ $\exists\sigma_i \in (\sigma_n, \sigma_n, \gamma)$	*then* $\frac{\|y(\hat{\lambda})\|^2}{c} \leq \frac{\rho}{2}\sqrt{\frac{c_u}{c}}$

Proof.

1. Since $\gamma \leq 1$, it follows that

$$g(\sigma_i) \geq g(\sigma_n), \quad \text{for } i < n.$$

Therefore

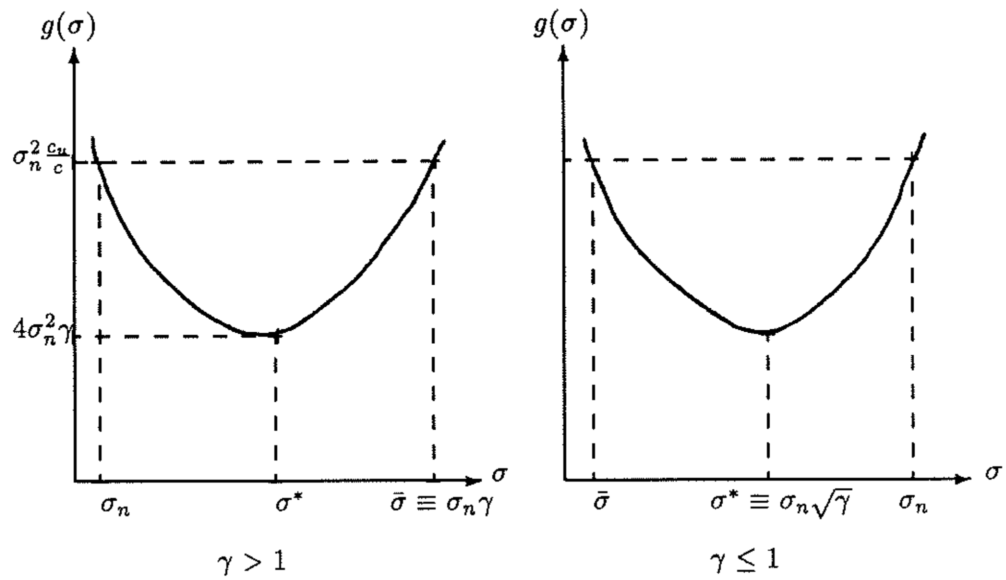

Figure 4.2: The function $g(\sigma)$ versus σ

$$
\begin{aligned}
\|y(\hat\lambda)\|^2 &= \sum_{i=1}^{n} \frac{\tilde b_i^2}{g(\sigma_i)} \le \frac{1}{g(\sigma_n)}\|\tilde b\|^2 \\
&= \frac{\|\tilde b\|^2}{\sigma_n^2}\cdot\frac{c}{c_u} = \frac{\|\tilde b\|^2}{\tilde b_n^2}\cdot\frac{\tilde b_n^2}{\sigma_n^2}\,\frac{c}{c_u} \le \rho c,
\end{aligned}
$$

where we have used the fact that

$$c_u = \sum_{i=1}^{n} \frac{\tilde b_i^2}{\sigma_i^2} \ge \frac{\tilde b_n^2}{\sigma_n^2}.$$

2. The proof is exactly the same as in case 1, because again we have $g(\sigma_i) \ge g(\sigma_n)$ for $i < n$.

3. In this case we have

$$g(\sigma_i) \ge g(\sigma^*) = 4\sigma_n^2\gamma \quad \text{for} \quad i \le n$$

Therefore,

$$
\begin{aligned}
\frac{\|y(\hat\lambda)\|^2}{c} \le \frac{(\sum_{i=1}^{n} \tilde b_i^2)}{4\sigma_n^2\,\gamma c} &= \frac{\|\tilde b\|^2}{\tilde b_n^2}\,\frac{\tilde p_n^2}{4\sigma_n^2}\cdot\frac{1}{\gamma c} \\
&\le \frac{\rho}{4}\frac{c_u}{c}\frac{1}{\gamma} = \frac{2^{-1}}{4}\sqrt{\frac{c_u}{c}}\frac{1}{(1-\sqrt{\frac{c}{c_u}})} \\
&\le \frac{\rho}{2}\sqrt{\frac{c_u}{c}},
\end{aligned}
$$

where we have used the fact that

$$\gamma > 1 \Rightarrow \sqrt{\frac{c_u}{c}} > 2 \Rightarrow \sqrt{\frac{c}{c_u}} < \frac{1}{2}. \qquad \square$$

Case 3 in Theorem 4.2 gives the general worst case for the $\hat\lambda$ estimation formula. Essentially, it guarantees that using $\hat\lambda$ as the regularization parameter will reduce the "constraint violation factor" $\frac{c_u}{c}$ to a value close to its square root.

Theorem 4.2 assumes a general distribution of singular values, other than exclusion/inclusion in the interval $(\sigma_n, \sigma_n\gamma)$. For special distributions of singular values, we can have tighter bounds.

Theorem 4.3 *If* $\sigma_1 = \sigma_2 = \cdots = \sigma_n \equiv \sigma$ *then* $\|y(\hat{\lambda})\|^2 = c$.

Proof. In this case $c_n = \|\tilde{b}\|^2/\sigma^2$ and

$$\|y(\hat{\lambda})\|^2 = \frac{\|\tilde{b}\|^2}{(\sigma + \sigma(\sqrt{\frac{c_u}{c}} - 1))^2} = c. \qquad \Box$$

Results for other special distributions are possible but we shall not pursue that here.

To summarize the situation, the effectiveness of the $\hat{\lambda}$ formula depends on the value of $\frac{c_u}{c}$. It helps to view c_u as $\|y(0)\|^2$. If $\frac{c_u}{c} \leq 4$ (the case $\gamma \leq 1$), then $\|y(\hat{\lambda})\|^2$ is guaranteed to approximate c to within a factor of ρ. If $\frac{c_u}{c} > 4$ (the case $\gamma > 1$), then it depends on whether there are singular values in the interval $(\sigma_n, \sigma_n\gamma)$. If there are none, then the $\hat{\lambda}$ formula is again guaranteed to work; otherwise it at least reduces the factor $\|y(\hat{\lambda})\|^2/c$ to approximately $\frac{\rho}{2}\sqrt{\frac{c_u}{c}}$.

5 Numerical Results

In this section we present numerical results. The first portion deals with several artificial examples so that we may demonstrate the three cases of $\gamma = \sqrt{\frac{c_u}{c}} - 1$ discussed in Theorem 4.2.

Assume that the original problem has already been transformed to (6), and our dimension size is 10. Thus, the matrix J is a diagonal matrix with ten singular values along the diagonal. The singular values have been chosen to lie between 1 and 10. For case 1 the singular values are bunched up by 1; for case 2 they are bunched up by 10, and for case 3 they are evenly spaced. Table 5.2 shows σ_i for $i = 1, ..., 10$ for the three case examples. For all three examples discussed below, the transformed right hand side vector $\tilde{b}$ has been randomly hand-picked to be

$$\tilde{b} = (2.1,\ 1,\ 1,\ 5,\ 4.4,\ 3.7,\ 0,\ 9,\ 2.8,\ 3)^T$$

We would like to investigate the behavior of applying one iteration of $\hat{\lambda}$.

The pertinent variables are shown in Table 5.3 for each case.

We run the same example again, with a different right hand side vector,

$$\tilde{b} = (.1,\ .1,\ .1,\ .1,\ .1,\ .1,\ .1,\ .1,\ .1,\ 1.0)^T.$$

i	Three Cases of γ		
	1	2	3
1	10	10	10
2	9	9.9	9
3	8	9.8	8
4	7	9.7	7
5	1.5	9.6	6
6	1.4	9.5	5
7	1.3	9.4	4
8	1.2	9.3	3
9	1.1	9.2	2
10	1	1	1

Table 5.2: Three Sets Of Singular Values

Case	Variables			
	c_u/c	$\|y(\hat{\lambda})\|^2/c$	ρ	$\rho/2\sqrt{\frac{c_u}{c}}$
1	2.75	1.32	18.0	-
2	5.36	1.68	18.0	-
3	100.0	16.6	-	90.2

Table 5.3: Values of the Pertinent Variables

Case	Variables			
	c_u/c	$\|y(\hat{\lambda})\|^2/c$	ρ	$\rho/2\sqrt{\frac{c_u}{c}}$
1	2.75	1.01	1.09	-
2	5.36	1.004	1.09	-
3	100.0	1.16	-	5.45

Table 5.4: Values of the Pertinent Variables; Tighter Bounds

The constant c is chosen so that the ratio of c_u/c is the same as in Table 5.3. The new values of all the variables are displayed in Table 5.4. Notice that in five of the six test cases the ratio $\|\lambda^*\|^2/c$ is close enough to 1.

We now turn from test data to measured data for Problem 2. The data arise from a signal processing application at SRI International. The matrix dimensions are $H \in R^{382\text{x}255}$ and $G \in R^{255\text{x}128}$. A semilog plot of the spectrum of HQ is shown in Figure 5.3. Notice that there is no clean break between any two small singular values, and that they taper off towards zero.

We would like to compare numerical results of the four techniques, with varying starting points. For the more realistic experiments we estimate σ_n. We estimate $\sigma_n(A)$ using one step of inverse iteration by choosing a random vector b, solving $x = A^{-1}b$, and setting

$$\sigma_n \approx \frac{\|b\|}{\|x\|}.$$

The inverse iterations could be continued to approximate σ_n more accurately, but the added accuracy in λ^*_{est} ends up costing more computations. Note, however, that λ^*_{est} could be greater than λ^{true} and we could lose the monotonicity of Newton and secant mentioned in Section 2. Therefore, we add some safeguards to the code to ensure convergence to the positive root.

We decided *a priori* to choose four values of a reduction factor, representing the amount of reduction required in the unconstrained solution c_u. The four values are 10, 100, 1000, and 10^6. Thus, the constraint c is determined by dividing c_u by the given reduction factor.

For the purpose of this paper, all unconstrained black box solvers are MATLAB calls. However, in production runs with large problems we use the Argonne Toeplitz solvers [Arg].

Table 5.5 gives the number of iterations needed for the four techniques for the four reduction factors, all with a starting guess of λ^*_{est} using an estimate of σ_n. The number of iterations for *Hebnew* are the same or lower for all four reduction factors.

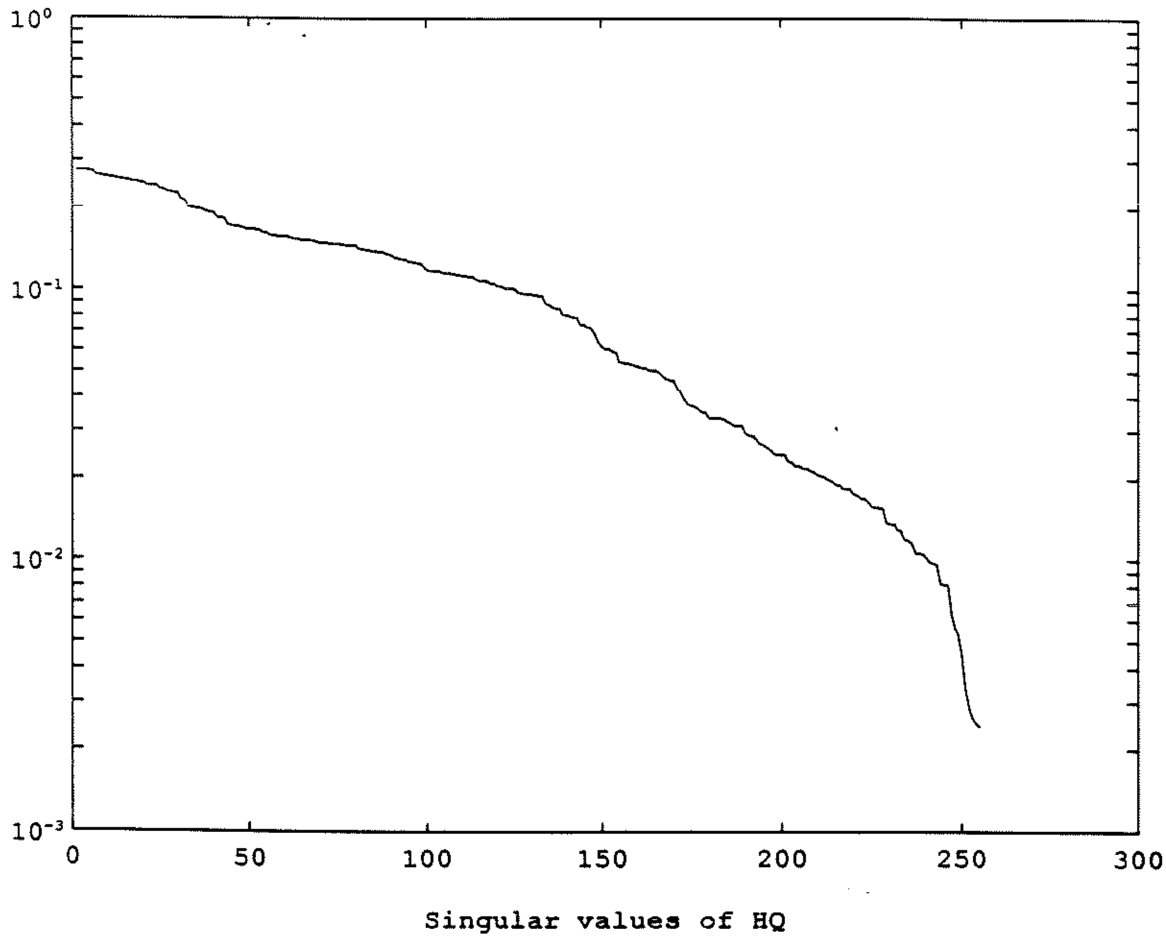

Figure 5.3: Spectrum of HQ

Method	Reduction in c_u/c			
	10	100	1,000	10^6
Hebnew	4	4	3	3
Newton	6	7	9	10
Hebsec	5	4	4	3
Secant	8	10	12	14

Table 5.5: Iteration Count Using λ^*_{est} as the Starting Guess(Tol = 10^{-4})

Method	Starting Values For Four Reduction Factors											
	10			100			1,000			1,000,000		
	0	λ^*_{est}	λ^*	0	λ^*_{est}	λ^*	0	λ^*_{est}	λ^*	0	λ^*_{est}	λ^*
Hebnew	5	4	5	5	4	4	4	3	4	4	3	3
Newton	8	6	7	10	7	8	11	9	10	16	10	13

Table 5.6: A Comparison of Three Starting Points

A feature of the matrix equation solver must be noted when keeping count of the iterations. Notice that the derivative information needed in the Newton methods requires solving (14) and (15), two matrix equations sharing the same matrix and two different right-hand side vectors. Some black box solvers retain the factored information to be used again for multiple right-hand side vectors (e.g. LINPACK), whereas in other packages no explicit matrix factoring is computed and hence cannot be saved (e.g. Toeplitz solvers). Thus, a Newton iteration could include two separate calls to the matrix solver. Depending on the particular black box solver, a Newton iteration could be twice as much work as a secant iteration. For example, when using Toeplitz solvers a secant iteration is cheaper than a Newton iteration. In fact, *Hebsec* is the cheapest method for our purposes.

Table 5.6 compares three different starting guesses for the two Newton methods, $\lambda = 0$, λ^*_{est}, which uses one inverse iteration to estimate σ_n, and λ^*, which uses the exact value of σ_n. The secant methods are excluded from this table as they require two starting points.

These tables demonstrate the advantage of using the estimates λ^* and λ^*_{est} over the starting guess $\lambda = 0$. In the case of *Hebnew*, the number of iterations using λ^* is either equal to or one less than that for $\lambda = 0$. This agrees with the observation made at the end of Section 3. The advantage is more pronounced for the Newton case. This may be due to the higher sensitivity to the initial guess for Newton's method. For our particular data, λ^*_{est} is at least as good as λ^* and better than $\lambda = 0$. The reason for this behavior is that the estimated smallest singular value is larger than the true smallest singular value. Hence $\lambda^*_{est} > \lambda^*$ but still lies to the left of the root, so that convergence is faster for λ^*_{est}. Figure 5.4 clarifies the situation. Ideally we would like an estimate, λ^*_{est}, of the root to lie in the interval $[0, root]$ so that Newton's method will converge monotonically to the unique positive root due to the convexity of the curve. Once the estimate falls to the right of the root, then some safety measures must be taken to move

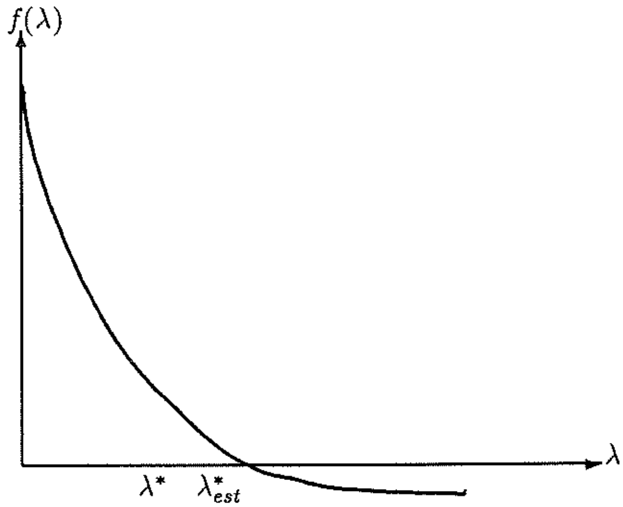

Figure 5.4: Graph of the secular equation

back to the correct interval. Also, if the estimate falls to the left of the origin, then again safety measures must be invoked to avoid converging to a negative root. In this particular case, using an estimate of the smallest singular value moved λ^*_{est} closer to the root than did λ^*, which used the exact minimum singular value.

6 Conclusions

We have shown how to solve quadratically constrained linear least squares problems that only require calls to a black box unconstrained solver. Our approach enables the user to exploit the matrix structure inherent in the unconstrained problem. We derive a formula for estimating the Lagrange multiplier which depends on the amount the unconstrained solution violates the constraint and an estimate of the smallest generalized singular value for Problem 1 or the smallest singular value for Problem 2. The method of choice depends on the nature of the black box solver, on whether it can retain factored information about the matrix equation. If information can be stored, then *Hebnew* starting with λ^* is the method of choice, where λ^*

is our estimate. This is a Hebden (rational) model with a Newton iteration. If the black box treats the right-hand side vector as an intrinsic part of the matrix equation, then $Hebsec$ starting with 0 and λ^* is the method of choice. Again, this is a Hebden model, with a secant iteration, so that derivative information is not required.

We have presented some theorems discussing under what conditions our estimate gives a satisfactory answer. Our numerical results verify the theory.

Another feature we have demonstrated is that, because our coefficient matrices are Toeplitz, the approach of using an iterative method with a starting guess of λ^* is an $\mathcal{O}(n^2)$ method due to the use of Toeplitz solvers as our black box unconstrained solvers.

References

[Arg] O.B. Arushanian, M.K. Samarin, V.V. Voevodin, E.E. Tyrtyshnikov, B.S. Garbow, J.M. Boyle, W.R. Cowell, and K.W. Dritz(1983), *The Toeplitz Package User's Guide*, Argonne National Laboratory, Argonne, Illinois.

[Bjo] A. Bjorck(1988), *Least Squares Methods.*, To appear in Handbook Of Numerical Analysis, Vol. II. Finite Difference Methods-Solution of Equations in R^n. Editors. P.G. Ciarlet and J.L. Lions, Elsevier/North Holland.

[BNS] J.R. Bunch and C.P.Nielsen and D.C. Sorensen (1978), *Rank-One Modification of the Symmetric Eigenproblem.*, Numer. Math. 31, 31-48.

[Eld] L. Eldén(1977), *Algorithms for the regularization of ill-conditioned least squares problems*, BIT 17, 134-145.

[FG] G.E. Forsythe and G.H. Golub(1965), *On the Stationary Values of a Second-Degree Polynomial on the Unit Sphere.*, J. Soc. Indust. Appl. Math, 13, No. 4, 1050-1068.

[Heb] M.D. Hebden(1973), *An algorithm for minimization using exact second derivatives*, Atomic Energy Research Establishment, Report T.P. 515, Harwell, England.

[Gan] W. Gander(1981), *Least Squares with a quadratic constraint.*, Numer. Math. 36, 291-307.

[Gol] G.H. GOLUB(1988), Private communication, SRI Interational, Menlo Park, California.

[Gvl] G.H. GOLUB AND C.F. VAN LOAN(1983), *Matrix Computations,* The Johns Hopkins University Press, Baltimore, Maryland.

[Rei] C.H. REINSCH(1971), *Smoothing by spline functions II*, Numer. Math. 16, 451-454.