Multiple Column Partitioned Min Max

E. Gomez[1], Y. Karant[1], K.E. Schubert[1]
{egomez, ykarant, schubert}@csci.csusb.edu
Department of Computer Science
California State University
San Bernardino, CA 92407

ABSTRACT: *This paper considers the problem of estimation and identification with bounded uncertainties in the data and system matrices. In particular, we will examine how to solve the $Ax \approx b$ when we have bounds on the errors of column blocks of A in the min max sense. Error bounds on column blocks of the system matrix A are caused by numerous situations, such as different sensor or parameter uncertainties, filter order updating, and image separation. Performance of the method is compared with existing techniques.*

KEYWORDS: *Regression, Least Squares, Robust Estimation*

1. Introduction

Estimation and identification are important areas of almost every problem in science and engineering. A typical way of stating an estimation or identification problem is that there is a system, described by a matrix, A, with inputs, x, and outputs, b. The inputs and outputs could either be matrices or vectors. The equation which describes this is thus $Ax = b$. The outputs of the system are considered measurable, and from them and the matrix A, it is desired to find the unknown inputs, x. In real systems the equality rarely holds because b is never measured perfectly, modeling and identification do not produce an exact A, and the basic equation $Ax = b$ is a linear approximation. The fundamental problem considered is thus $Ax \approx b$, where both A and b are assumed to have errors associated with them. In particular, let the "true" system, A_{true}, be related to the nominal model, A, by an error matrix E_A. Similarly let the true outputs, b_{true} be related to the measured outputs, b, by E_b. Since the true system is not a mathematical model, the resulting equation is still approximate, $(A + E_A)x \approx (b + E_b)$, but is the best approximation possible. The goal is to find the best x, in the resulting minimization problem, $\min_x \|(A + E_A)x - (b + E_b)\|$. The min max problem was proposed and solved in [1] by secular equation techniques and in [5] by Linear Matrix Inequality techniques. This paper will concentrate on the secular equation formulation.

This paper considers the case where bounds are known on groups of columns of E_A, which are referred to as block columns. This can arise if a new block column is added to A corresponding to an increase in the order of the filter. The new block column will not necessarily have the same uncertainty as the original block, thus partitioning is needed so different errors may be assigned to each block. Alternately, the column partitioning case could be used to model a series of geophones in a seismology problem that have different uncertainties due to geometry or surface geology conditions. Column partitioning also describes signal separation with different uncertainties associated with each signal. The column partitioning case also could be dealing with various polynomials in a polynomial fitting problem. In short, many problems satisfy the basic conditions of the multi-column problem. A simplified case of this, where one of the columns was unperturbed is considered in [1].

[1]The authors gratefully acknowledge the support of the NSF under award CISE 98-10708

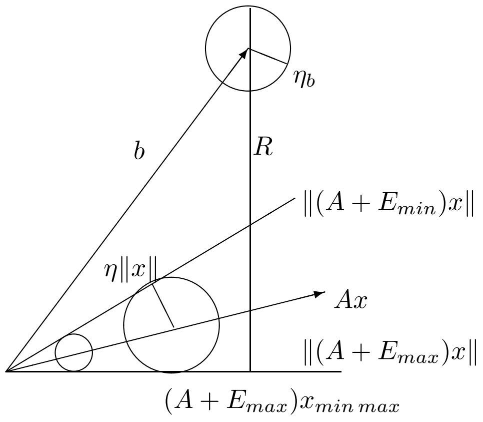

Figure 1: Geometric Interpretation of Min Max Solution

2. Min Max

The min max problem is to find the worst model of a system in a bounded region, and then solve the system based on this worst case scenario. Mathematically it is written as

$$\min_{x} \max_{\substack{\|E_A\| \leqslant \eta \\ \|E_b\| \leqslant \eta_b}} \|(A+E_A)x - (b+E_b)\|. \quad (1)$$

This problem can be shown to be equivalent to solving a problem with similar form to the Tikhonov problem, see [1]. Equation 1 can be interpreted geometrically by Figure 1.

The maximization forms the hyperspheres around Ax and b. The cone around A is formed by varying the size of x. The solution, x, and the residual, R, are found by connecting the furthest points on the hyperspheres. The maximization restricts the problem to the lower line of the cone. The minimization selects the point on the lower cone such that the line segment from the furthest point on the hypersphere around b to the lower cone is perpendicular to the lower cone. The norm used in [1] is the 2-norm, though [13] extends it to other norms. The min max problem becomes

$$\min_{x}(\|Ax-b\| + \eta\|x\| + \eta_b), \quad (2)$$

which differs from the typical Tikhonov problem in that the norms are not squared. As opposed to the Tikhonov problem, the term η now has a physical intuition also, that being the amount of uncertainty in the matrix. Computing the min max solution takes longer than computing the solution to a Tikhonov problem if a simple choice of regression parameter is chosen for the Tikhonov problem, so it is logical to ask why one would want to spend the extra operations to do so. The simple answer is that the two problems can give arbitrary differences, which we will examine in Section 3.

In the form of Equation 2 it is easy to see that the problem is continuous and convex but non-smooth, since it is non-differentiable whenever $x = 0$ or when $Ax = b$. The solution to Equation 2 and thus Equation 1, is summarized in Table 1. For Table 1, let the SVD of A be given by

$$A = \begin{bmatrix} U_1 U_2 \end{bmatrix} \begin{bmatrix} \Sigma \\ 0 \end{bmatrix} V^T.$$

Partition the vector $U^T b$ into

$$\begin{bmatrix} U_1 U_2 \end{bmatrix}^T b = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix},$$

and introduce the secular equation

$$g(\psi) = b_1^T(\Sigma^2 - \eta^2 I)(\Sigma^2 + \psi I)^{-2} b_1 - \frac{\eta^2}{\psi^2}\|b_2\|^2,$$

which has a unique positive root, denoted $\bar{\psi}$ under the conditions noted in Table 1. Define the weighted pseudo-inverse by

$$A_w^\dagger = (A^T A + \bar{\psi} I)^{-1} A^T.$$

Finally define

$$\tau_1 = \frac{\|\Sigma^{-1} b_1\|}{\|\Sigma^{-2} b_1\|} \qquad \text{and} \qquad \tau_2 = \frac{\|A^T b\|}{\|b\|}.$$

The solution is thus given below. Notice that the least squares solution, $A^\dagger b$, is the min max solution under special conditions. In one case a scaled family of the least squares solution solves the problem. In general though the solution is given by finding the unique root of the secular equation, $g(\psi)$ in the positive quadrant. When η is large the solution is zero.

3. Comparison to Tikhonov

The min max solution has a similar form to the Tikhonov solution and computing the min

	$b \in \mathcal{R}(A)$	$b \notin \mathcal{R}(A)$
$\eta \geqslant \tau_2$	0	0
$\tau_1 < \eta < \tau_2$	$x = A_w^\dagger b$	$x = A_w^\dagger b$
$\eta \leqslant \tau_1$	$x = A^\dagger b$	$x = A_w^\dagger b$
$\eta = \tau_1 = \tau_2$	$x = \beta A^\dagger b$ with $0 \leqslant \beta \leqslant 1$	$x = A_w^\dagger b$

Table 1: Min Max Solution

max solution takes longer than computing the solution to a Tikhonov problem if a simple choice of regression parameter is chosen for the Tikhonov problem. At this point, it is reasonable to ask if there is a similar, but simpler way to solve the problem, which exhibits the desired behavior of min max that can be solved instead of the min max methodology of [1, 13, 5]. Tikhonov regulation has a large body of literature, such as [6, 9], and a closed form solution.

A reasonable choice for the parameter λ in the Tikhonov problem is to chose it to be equal to the square of the uncertainty, since all the other terms are squared and this will account for the size of the uncertainty. In this case the model has a closed form solution which is given by

$$\hat{x} = \left(A^T A + \eta^2 I\right)^{-1} A^T b. \tag{3}$$

Note that for the min max problem that if $Ax \neq b$ and $x \neq 0$ then the min max problem also has a solution with a similar form given by

$$\hat{x} = \left(A^T A + \alpha I\right)^{-1} A^T b \tag{4}$$

$$\alpha = \eta \frac{\|Ax - b\|}{\|x\|}. \tag{5}$$

If the Tikhonov problem's parameter can be arbitrarily larger, then the solution can be over regularized and thus valuable information can be lost. If the Tikhonov parameter can be arbitrarily smaller, then the solution can be under regularized and thus the solution might not be robust. To compare the two, examine the ratio of the regularization parameters

$$\frac{\alpha}{\eta^2} = \frac{\|Ax_{mm} - b\|}{\eta \|x_{mm}\|}. \tag{6}$$

3.1. Over-Regularization

First, see if the Tikhonov problem can be over regularized, which is the more dangerous problem. This corresponds to the ratio being arbitrarily small. Note that $\|Ax_{mm} - b\| \leqslant \|b\|$ at the solution, by noting the cost at the solution must be less than the cost at the point $x = 0$. Thus,

$$\frac{\alpha}{\eta^2} \leqslant \frac{\|b\|}{\eta \|x_{mm}\|}. \tag{7}$$

It is clearly possible to pick A and b such that $\eta\|x_{mm}\| \gg \|b\|$. For example consider the following simple system,

$$A = \begin{bmatrix} \delta \\ 0 \end{bmatrix} \qquad b = \begin{bmatrix} \frac{1}{\delta} \\ \delta \end{bmatrix} \qquad \eta = \frac{\delta}{2} \tag{8}$$

with $\delta \ll 1$. For this system note that the least squares (LS) solution is given by $x_{LS} = \frac{1}{\delta^2}$, and the min max system is $x_{mm} = \frac{1}{\delta^2} - \frac{1}{\delta\sqrt{3}}$. Note that since $\delta \ll 1$, the min max estimate is extremely close to the LS solution. The Tikhonov problem solution is given by $x_T = \frac{4}{5\delta^2}$, which is easily seen to be arbitrarily far from the desired solution, since for $\delta \ll 1$ the two candidate solutions differ by almost 20% of an arbitrarily large number. Moreover, the ratio of regularization parameters is approximately given by the arbitrarily small number,

$$\frac{\alpha}{\eta^2} \approx \frac{4}{\sqrt{3}}\delta^2. \tag{9}$$

3.2. Under-Regularization

The second area to be considered is if the Tikhonov problem can be under-regularized. This corresponds to the ratio of α over η^2 being arbitrarily large. Note that $\|Ax_{mm} - b\| \geqslant \|P_{A^\perp} b\|$, thus

$$\frac{\alpha}{\eta^2} \geqslant \frac{\|P_{A^\perp} b\|}{\eta \|x_{mm}\|}. \tag{10}$$

It is clearly possible to pick A and b such that $\|x_{mm}\| \ll \|P_{A^\perp} b\|$. For example consider the following simple system,

$$A = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \qquad b = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \qquad \eta = 1. \tag{11}$$

Note that since the perturbation is as large as the norm of the A matrix, $x_{mm} = 0$, which corresponds to $\alpha \to \infty$. This is intuitively pleasing, as it confirms the belief that no valid information exists for a system with uncertainty as large as the system. Note also that $x_{LS} = 1$. Now the Tikhonov problem has the solution $x_T = \frac{1}{2}$. Not only is this clearly too optimistic an answer, the ratio is infinite and thus arbitrarily large, as was desired to be shown. Thus while the Tikhonov problem has nice properties for calculation, its estimator can be arbitrarily different than the min max problem. Additionally, the Tikhonov problem does not correspond to physical intuition as can be seen in the last example above. The min max problem can thus not be altered to an apparently similar problem and solved for that system.

4. The Partitioned Problem

After having shown the uniqueness of the general form of the min max problem, we will examine the column partitioned case. Consider the partitioned min max problem,

$$\min_x \max_{\substack{\|E_i\|_2 \leqslant \eta_i, \\ \|E_b\|_2 \leqslant \eta_b}} \left\| \sum_{i=1}^{p} (A_i + E_i) x_i - (b + E_b) \right\|,$$

where $A_i, E_i \in \mathbb{R}^{m \times n_i}$, $x_i \in \mathbb{R}^{n_i}$, and $n = \sum_{i=1}^{p} (n_i)$. Note that given the bounds on the maximization

$$\left\| \sum_{i=1}^{p} (A_i + E_i) x_i - (b + E_b) \right\| \leqslant \|Ax - b\| + \sum_{i=1}^{p} \eta_i \|x_i\| + \eta_b.$$

By considering the following perturbations,

$$E_i^o = \frac{\eta_i (Ax - b) x_i^T}{\|Ax - b\| \, \|x_i\|}$$
$$E_b^o = \frac{-\eta_b (Ax - b)}{\|Ax - b\|}$$

it can be shown that these perturbations achieve the upper bound and thus the cost function can be simplified to

$$J = \min_x \left(\|Ax - b\| + \sum_{i=1}^{p} \eta_i \|x_i\| + \eta_b \right). \tag{12}$$

The cost function is clearly convex, as it is the sum of convex functions, so a solution must exist but does not have to be unique.

5. Quadratically Convergent Method

The cost function is not only convex, but it is also a sum of Euclidean norms. A large body of literature exists for solving the sum of Euclidean norms problem. The problem dates back to Fermat, who posed a special case of it. Various methods have been proposed which range from a sequence of linear least squares problems [7, 3, 12], successive over-relaxation [11], hyperbolic approximation procedure [4], subgradients [8, 2]. All of these have, at best, linear convergence, however there is a quadratically convergent method proposed by Michael Overton in [10].

A quadratically convergent method with good properties exists, so why look further? The major reason is that method operates on the size of the original problem (m), while a secular equation solution will operate on a smaller problem (p, with $p \ll m$ usually).

6. Column Dependence

Given the similarity of the problem structure to the non-partitioned case, some have concluded that the solution conditions should be the same. In particular, the non-partitioned problem has two simple conditions on x that do not carry into the partitioned case,

1. the solution, x, is non-zero if and only if $\|A^T b\| > \eta \|b\|$,
2. the solution, x, has a smaller norm than the least squares solution.

6.1. When x Is Zero

First consider the simple relation that the solution x is non-zero if and only if $\|A^T b\| > \eta\|b\|$. This is not true for the partitioned case, which can be seen by considering the following

$$A_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \quad A_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \quad b = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}.$$

It is readily apparent that $A_2^T b = 0$ and thus from the original problem, $x_2 = 0$ for all η_2. Now consider $\eta_1 = \eta_2 = \frac{1}{4}$, and consider the cost, J, for $x_2 = 0$ and $x_1 \neq 0$.

$$J(x_1, x_2 = 0) \quad = \quad \sqrt{2(x_1^2 - x_1 + 1)} + \frac{|x_1|}{4}$$

The minimum can be found by taking the derivative of $J(x_1, x_2 = 0)$ and setting it equal to zero, thus the minimum cost for $x_2 = 0$ is

$$\begin{aligned} J\left(\frac{1}{2} - \sqrt{\frac{3}{124}}, 0\right) &= (3.875)\sqrt{\frac{3}{31}} + \frac{1}{8} \\ &\approx 1.330. \end{aligned}$$

Now consider the case when $x_2 \neq 0$. Note that when $x_2 \neq 0$, it must be that $x_1 \neq 0$ because if not, it is easily verified that $J(x_1 = 0, x_2 \neq 0) > J(x_1 = 0, x_2 = 0)$. To start, the expression for the cost is given by

$$\begin{aligned} J(x_1, x_2 \neq 0) = &\frac{|x_1| + |x_2|}{4} \\ &+ \sqrt{(1 - x_1)^2 + (x_1 + x_2)^2 + 1}. \end{aligned}$$

By examining the cost it can be concluded that the optimal $0 < x_1 < 1$ and optimal $-x_1 < x_2 < 0$. These relations can be used to simplify the derivatives about to be taken. Since both x_1 and x_2 are not zero, we take the derivative of the cost with respect to each variable in turn and set the result equal to zero and do some algebra to obtain

$$\begin{aligned} J\left(1 - \frac{2}{\sqrt{11}}, \frac{3}{\sqrt{11}} - 1\right) &= \frac{\sqrt{11}}{4} + \frac{1}{2} \\ &\approx 1.329. \end{aligned}$$

Thus, the cost for $x_2 \neq 0$ is less than the cost for $x_2 = 0$, so while in the original problem it would have been predicted that $x_2 = 0$ this is not the case.

6.2. Size of $\|x\|$

The second relation to consider is that the size of the multi-column partitioned min max solution, x_Ψ, should be smaller than the least squares solution, x_{LS}, since both have the same numerator and the denominator of x_Ψ is larger. This is not always the case. To demonstrate this, consider a simple problem.

Let A and b be the matrices defined below with each column of A a separate partition.

$$A = \begin{bmatrix} 1 & 0 & 0.1 \\ 1 & -1 & 1 \\ 0 & 0 & 0.1 \\ 0 & 0 & 0 \end{bmatrix} \quad b = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 10 \end{bmatrix}$$

The least squares solution is given by

$$x_{LS} = \begin{bmatrix} 1 & 1 & 0 \end{bmatrix}^T.$$

Now consider the case when $\eta_1 = 2$,$\eta_2 = 0$, and $\eta_3 = 0$. The solution, x_Ψ is given by

$$x_\Psi = \begin{bmatrix} 0 & 5 & 5 \end{bmatrix}^T.$$

It is trivial to see that $\|x_{LS}\| < \|x_\Psi\|$, and thus the idea is disproved. The question remains then as to what can be said about the size of x_Ψ and thus where it lies. The following lemma is not tight in its bound but it does provide a good starting point for the analysis.

Lemma 1 *For a matrix A, a vector b, and scalars η_i, the solution to the multiple column partitioned min max problem, x_Ψ satisfies*

$$\|x_\Psi\| \leqslant \kappa^2 \|x_{LS}\|.$$

where κ is the condition number of A.

Proof:

$$\begin{aligned} \|x_\Psi\| &= \|(A^T A + \Psi)^{-1} A^T b\| \\ &= \|(A^T A + \Psi)^{-1}\| \|A^T A\| \|(A^T A)^{-1} A^T b\| \\ &\leqslant \frac{1}{\sigma_{min}(A^T A + \Psi)} \sigma_1^2 \|x_{LS}\| \\ &\leqslant \frac{\sigma_1^2}{\sigma_n^2} \|x_{LS}\| \\ &\leqslant \kappa^2 \|x_{LS}\| \end{aligned}$$

Other such bounds exist and can be used to tighten the starting condition. A key point of developing this lemma is that bounds exist on the size of the estimate, and can be calculated a priori. Such bounds could be used to start methods like the ellipsoidal algorithm.

7. Form of Solution for Multiple Columns

The solution could be at either a differentiable point or a non-differentiable point. The non-differentiable points are located at $\|x_i\| \neq 0 \quad \forall i \in \{1, \cdots, p\}$ and $\|Ax - b\| \neq 0$. This section will consider the case when the solution is at a differentiable point. A necessary condition for a minimum at a differentiable point is obtained by taking the gradient and setting it equal to zero, which yields

$$A^T b = \left(A^T A + \Psi\right) \hat{x}$$

with

$$\Psi = \begin{bmatrix} \psi_1 I & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \psi_p I \end{bmatrix}$$

$$\psi_i = \frac{\eta_i \|A\hat{x} - b\|}{\|\hat{x}_i\|} > 0.$$

Note that $A^T A$ is positive semi-definite and Ψ is positive definite, so that the matrix multiplying $\hat{x}$ is invertible. Since, $\left(A^T A + \Psi\right)$ is invertible, $\hat{x}$ can be solved for,

$$\hat{x} = \left(A^T A + \Psi\right)^{-1} A^T b$$

8. General Column Form Secular Equation

The secular equations for this problem are now developed. First, square the definition of ψ_i.

$$\psi_i^2 \|\hat{x}_i\|^2 = \eta_i^2 \|A\hat{x} - b\|^2$$

Then using the expressions derived for $\hat{x}_i$ and $A\hat{x} - b$, define the secular equations, G_i ($\forall i \in 1, \ldots, p$), to be

$$G_i(\psi) = b^T F N_i F b \qquad (13)$$

with

$$\begin{aligned} F &= \left(I + A\Psi^{-1} A^T\right)^{-1} \\ N_i &= \left(A_i A_i^T - \eta_i^2 I\right). \end{aligned}$$

Note that the definition of F is positive definite for all positive values of ψ_i. Note also that the secular equations ($G_i(\psi)$,$i = 1, 2, \cdots, p$) have no singularities in the first quadrant and since the equations are rational expressions of ψ_i the functions are C^1 in the first quadrant. All that remains is to show the existence and uniqueness of the solution.

8.1. Uniqueness

To prove uniqueness, it will be shown that the cost function is strictly convex and thus any solution to the original problem is unique. Since the secular equations only have a root when the original problem has a solution, this will show that any solution to the secular equation is unique. To show the original problem is strictly convex in the region of interest for the problem, consider the Hessian of the cost, H,

$$H = \frac{A^T P_{Ax-b}^{\perp} A + \Psi \mathrm{diag}(P_{x_i}^{\perp})}{\|Ax - b\|},$$

where P is a projection matrix and its subscript specifies the space it projects onto. In order for the Hessian to be positive semi-definite there must be a column of A, say A_{i_k}, that is in the i^{th} partition and a corresponding element of x called x_{i_k} for which both

1. $P_{Ax-b}^{\perp} A_{i_k} = 0$,
2. $\psi_i e_{i_k}^T P_{x_i}^{\perp} e_{i_k} = 0$,

where e_{i_k} is a vector that is zero everywhere except the component in the i_k^{th} position, which is 1. In order for item 1 to hold, A_{i_k} must be in the direction of the residual, which means that $b \in \mathcal{R}(A)$. By assuming the standard condition that $b \notin \mathcal{R}(A)$, the first term is positive and thus the Hessian is positive definite. Even if $b \in \mathcal{R}(A)$ this only corresponds to the least squares case.

The only thing left is to observe that the problem is strictly convex and does not have any place it is undefined, thus there is always a solution. It can be shown that if $\eta_i \leqslant \frac{\|A_i^T b\|}{\|b\|}$ the solution was at the extremum. The solution is thus characterized. Any multidimensional root finder can be used to calculate the actual location of the roots of G_i.

9. A Numerical Example

The following problem is based on an example of Dr. Ali Sayed in an unpublished paper entitled "Estimation in the Presence of Multiple Sources of Uncertainties with Applications". Assume that there are two different signals that need to be estimated from a series of three simultaneous observations. The relation between the signals and the observations are known approximately and are the A matrix. Additionally, assume the first signal is stronger and that the errors associated with the first signal are smaller.

First consider the case of singular A. In Figure 2(a), least squares can only estimate the stronger signal, but does a reasonable job at it. The multi-column solution does quite well for the first signal, and gets basic features and is a reasonable scale for the second. Note that as is typical for a pessimistic problem, the multi-column min max tends to underestimate the size of the signal, but this underestimation is better than the alternatives. Total least squares is shown in Figure 2(b) because it is not even close, notice the order of magnitude is off by around 14.

Now consider the case of a near singular A. This is shown in Figure 2(c). Least squares and total least squares are almost identical for this problem, and off by a factor of two to seven. The multi-column solution is very good for first signal and reasonable for the second. Note the multi-column min max does not change significantly between the two cases. This is a result of the robustness of the solution. A solution for the min max problem works for nearby problems, so it tends not to change for small alterations in the problem, even when the change tends to cause a major change in other methods.

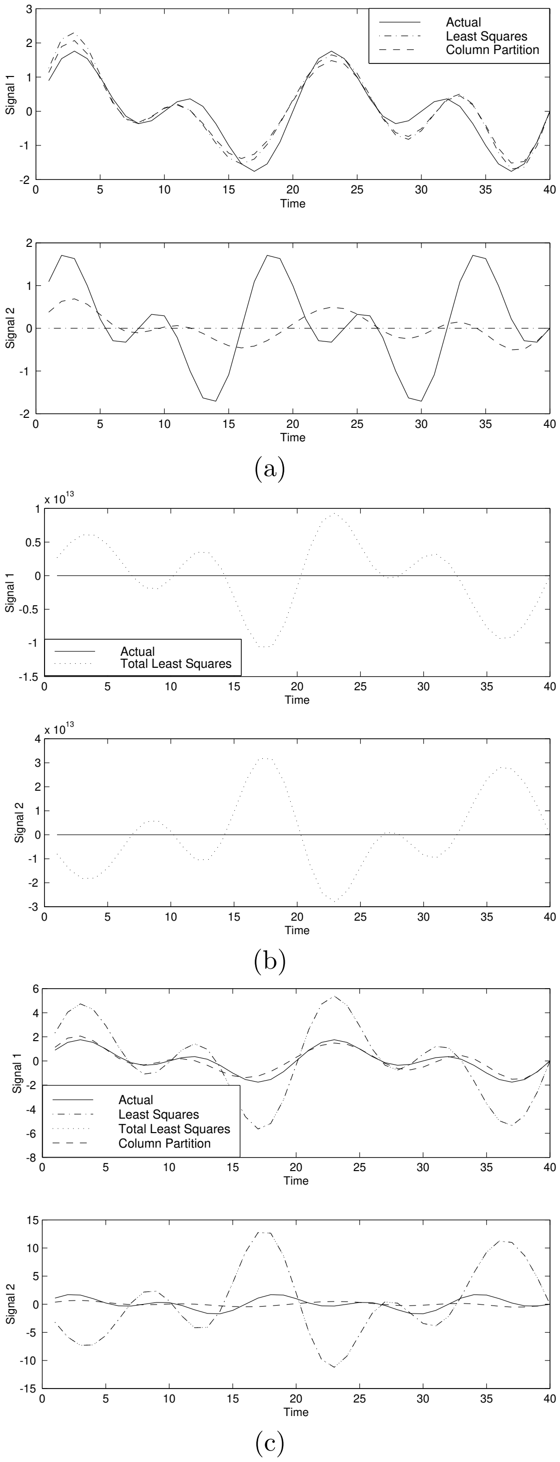

Figure 2: Matrix Signal Separation Problem

10. Summary

The multiple column min max problem has been posed and solved. Several techniques for solution are presented but the best technique is to use the secular equation because it is usually a much smaller problem. Overton's quadratically convergent method for the sum of Euclidean norms could be used, and can be faster if $m \approx p$ and the problem is ill-conditioned. Overton's method can converge faster when the problem is ill-conditioned, but note that for $m \approx p$ the problem must be at least nearly square with the partitions being individual columns of A. The conditions for the secular equation to work better are much more likely and thus are the advised solution technique.

The multiple column min max problem should be used instead of the regular min max problem if there is a significant difference in the bounds on some block columns. If the bounds are similar there is not a significant difference, but there is a processing cost difference. The usual case when the min max formulation has significant advantages over the least squares and total least squares is when the model has conditioning problems. Without conditioning problems the standard techniques give answers, which are reasonably close to the min max, and sometimes give better answers if the error bounds are over-estimated. When the conditioning problems exist, however the min max solution can maintain a reasonably good solution into areas where the other techniques are not capable. In cases where matrix structure is the key goal, and robustness is desired, the LMI techniques should be used.

References

[1] S. Chandrasekaran, G. H. Golub, M. Gu, and A. H. Sayed. Parameter Estimation in the Presence of Bounded Data Uncertainties. *SIMAX*, 19(1):235–252, 1998.

[2] F. Cordellier and J.Ch. Giorot. On the Fermat-Weber Problem with Convex Cost Functionals. *Mathematical Programming*, 14:295–311, 1978.

[3] U. Eckhardt. Weber's Problem and Weiszfeld's Algorithm in General Spaces. *Mathematical Programming*, 18:186–196, 1980.

[4] J.W. Eyster, J.A. White, and W.W. Wierwille. On Solving Multifacility Location Problems Using a Hyperboloid Approximation Procedure. *AIIE Transactions*, 5:1–6, 1973.

[5] L. El Ghaoui. and H. Lebret. Robust Solutions to Least-Squares Problems with Uncertain Data. *SIMAX*, 18(4):1035–1064, 1997.

[6] G. H. Golub, P. C. Hansen, and D. P. O'Leary. Tikhonov Regularization and Total Least Squares. *SIMAX*, 30(1):185–194, 1999.

[7] H.W. Kuhn. A Note on Fermat's Problem. *Mathematical Programming*, 4:98–107, 1973.

[8] R.F. Love. Locating Facilities in Three-dimensional Space by Convex Programming. *Naval Research Logistics Quarterly*, 16:503–516, 1969.

[9] D. P. O'Leary. Near-Optimal Parameters for Tikhonov and Other Regularization Methods. Technical report, CS Dept., University of Maryland, 1999. CS-TR-4004.

[10] M. L. Overton. A Quadratically Convergent Method for Minimizing a Sum of Euclidean Norms. *Mathematical Programming*, 27:34–63, 1983.

[11] S. Schechter. Minimization of a Convex Function by Relaxation. In J. Abadie, editor, *Integer and Nonlinear Programming*, pages 177–190. North-Holland, 1970.

[12] H. Voss and U. Eckhardt. Linear Convergence of Generalized Weiszfeld's Method. *Computing*, 25:243–251, 1980.

[13] G. A. Watson. Data Fitting Problems with Bounded Uncertainties in the Data. *SIMAX*, 22(4):1274–1293, 2001.