Estimation in the Presence of Multiple Sources of Uncertainties with Applications*

ALI H. SAYED
Electrical Engineering Department
University of California
Los Angeles, CA 90024

SHIVKUMAR CHANDRASEKARAN
Dept. Electrical and Computer Engineering
University of California
Santa Barbara, CA 93106

Abstract

We develop an estimation technique for problems that involve multiple sources of uncertainties or errors in the data. The method allows the designer to explicitly incorporate into the problem formulation bounds on the sizes of the uncertainties; thus leading to solutions that will not over-emphasize the effects of the uncertainties beyond what is assumed by the prior information. Applications in array signal processing and image processing are considered.

1. Introduction and problem formulation

This paper deals with the development of an estimation technique for models with bounded data uncertainties. The method will be referred to as a BDU estimation method for brevity. It is based on a new constrained game-type formulation that allows the designer to explicitly incorporate into the problem statement a-priori information about bounds on the sizes of the uncertainties in the model. In this way, the effect of uncertainties will not be unnecessarily over-emphasized beyond what is implied by the a-priori bounds; consequently, overly conservative designs, as well as overly sensitive designs, will be avoided.

A key feature of the BDU formulation is that geometric insights (such as orthogonality conditions and projections) and recursive (adaptive or online) techniques, which are widely known and appreciated for classical quadratic-cost designs, can be pursued in this new framework. More details on these aspects can be found in [1]. Also, algorithms for computing optimal solutions with the same computational effort as standard least-squares solutions exist, thus making the new formulations attractive for practical use. An SVD-based solution is developed rather fully in [2].

In this paper, we introduce the following optimization problem

$$\min_x \max_{\substack{\|\delta A_i\| \leq \eta_i \\ 1 \leq i \leq K}} \left\| \begin{bmatrix} A_1 + \delta A_1 & \ldots & A_K + \delta A_K \end{bmatrix} x - b \right\| \tag{1}$$

where x is an $n-$dimensional column vector, the $\{A_j\}$ denote submatrices (column-wise) of an $N \times n$ known or nominal matrix A, and the $\{\delta A_j\}$ denote submatrices of an $N \times n$ perturbation matrix δA. The notation $\|\cdot\|$ denotes the Euclidean norm of its vector argument or the maximum singular value of its matrix argument. We also partition the vector x accordingly with A, say $x = \text{col}\{x_1, \ldots, x_K\}$, and further assume that

$$\text{rank}(A) = n \quad \text{and} \quad b \notin \mathcal{R}(A) . \tag{2}$$

The analysis can be extended to cases where (2) is violated, but we shall focus here on (2) in order to highlight the main ideas. Due to space limitations, our proofs are brief and details will be provided elsewhere.

Problem (1) seeks a solution $\hat{x}$ that performs "best" in the worst-possible scenario. It can be regarded as a constrained two-player game problem, with the designer trying to pick an x that minimizes the residual norm while the opponents $\{\delta A_i\}$ try to maximize the residual norm. The game problem is constrained since it imposes a limit on how large (or how damaging) the opponents can be.

2. The case of a unique zero solution

We shall establish that when the uncertainty set $\{\|\delta A_i\| \leq \eta_i\}$ is large enough to include a perturbed

*This material was based on work supported in part by the National Science Foundation under Awards MIP-9796147, CCR-9732376, and CCR-9734290.

0-7803-5148-7/98/$10.00©1998 IEEE

matrix $(A+\delta A)$ that is orthogonal to b, then the unique solution of (1) is $\hat{x} = 0$. Otherwise, the solution is nonzero and has an interesting regularized form.

Lemma 1 *The uncertainty set $\{\|\delta A_i\| \leq \eta_i\}$ contains a perturbation δA such that $(A+\delta A)^T b = 0$ iff*

$$\eta_i \geq \|A_i^T b\|/\|b\| \qquad \text{for all } \ 1 \leq i \leq K \tag{3}$$

Proof: Assume there exists a perturbation δA, say $\overline{\delta A}$ with $\{\|\overline{\delta A}_i\| \leq \eta_i\}$, such that $(A+\overline{\delta A})^T b = 0$. Then $(A_i + \overline{\delta A}_i)^T b = 0$ and, consequently,

$$\|A_i^T b\| = \|\overline{\delta A}_i^T b\| \leq \|\overline{\delta A}_i^T\| \cdot \|b\| = \|\overline{\delta A}_i\| \cdot \|b\| ,$$

which implies that $\|\overline{\delta A}_i\| \geq \|A_i^T b\|/\|b\|$ and, hence, condition (3) must hold for each i. Conversely, assume (3) holds and choose $\overline{\delta A}_i = -bb^T A_i/\|b\|^2$. Then $\|\overline{\delta A}_i\| \leq \eta_i$ and $A + \overline{\delta A} = P_b^{\perp} A$, where $P_b^{\perp} \triangleq (I - bb^T/\|b\|^2)$ is the projector onto the orthogonal complement space of b. Then $(A+\overline{\delta A})^T b = 0$, as desired.

◊

Note that if we set x equal to zero in the BDU cost function (1), we obtain that the cost is equal to $\|b\|$ regardless of δA. We now show that when (3) holds, the cost for any nonzero x will be strictly larger than $\|b\|$ so that $\hat{x} = 0$ has to be the unique solution.

Lemma 2 *The* BDU *estimation problem (1) has a unique solution at $\hat{x} = 0$ if, and only if, (3) holds.*

Proof: Assume first that (3) holds and choose the matrix $\overline{\delta A}$ from the proof of Lemma 1. It can be shown that $A + \overline{\delta A}$ is full rank. Now since b is orthogonal to $(A+\overline{\delta A})$, it follows that $\|(A+\overline{\delta A})x - b\| > \|b\|$, for any nonzero vector x. Therefore,

$$\max_{\{\|\delta A_i\| \leq \eta_i\}} \|(A+\delta A)x - b\| \geq \|(A+\overline{\delta A})x - b\| > \|b\| ,$$

which shows that $\hat{x} = 0$ is the unique solution of (1).

Conversely, assume $\hat{x} = 0$ is the unique solution of (1) then, for every x, we can show that

$$\max_{\|\delta A_i\| \leq \eta_i} \left[-2b^T A_i A_i^T b - 2b^T \delta A_i A_i^T b\right] \geq 0 ,$$

from which we can conclude that $\eta_i \geq \|A_i^T b\|/\|b\|$.

◊

3. Worst-case perturbations

Now assume that (3) is violated at least for some i, say

$$\eta_{i_o} < \|A_{i_o}^T b\|/\|b\| , \qquad \text{for some } i_o . \tag{4}$$

Hence, if the problem has a solution $\hat{x}$ then it has to be nonzero. We shall in fact show that a unique *nonzero* solution exists in this case.

Returning to (1), we shall first identify the perturbations $\{\delta A_i\}$ that maximize the residual norm. Indeed define, for every x, the column vector

$$q(x) \triangleq \text{col}\{\eta_1 \|x_1\|^{\dagger} x_1, \ldots, \eta_K \|x_K\|^{\dagger} x_K\} , \tag{5}$$

where $a^{\dagger}$ denotes the pseudo-inverse of the scalar a (equal to its inverse if $a \neq 0$ and equal to 0 if $a = 0$). Then it can be verified that the following rank-one modification of A,

$$A + \delta A^o(x) \triangleq A + \frac{(Ax-b)\, q^T(x)}{\|Ax-b\|} , \tag{6}$$

achieves the maximum residual, viz.,

$$\max_{\{\|\delta A_i\| \leq \eta_i\}} \|(A+\delta A)x - b\| = \|[A + \delta A^o(x)]x - b\|$$

$$= \|Ax - b\| + \sum_{i=1}^{K} \eta_i \|x_i\| \triangleq J(x) .$$

Moreover, the following facts hold.

Lemma 3 *For any x, the matrix $\delta A^o(x)$ is such that $A + \delta A^o(x)$ is full rank and the residual vectors $[A + \delta A^o(x)]x - b$ and $Ax - b$ are collinear; they also point in the same direction (i.e., one is a positive multiple of the other).*

4. The Orthogonality Condition

We are thus reduced to studying the equivalent problem

$$\min_x \left(\|Ax-b\| + \eta_1\|x_1\| + \ldots + \eta_K\|x_K\|\right) . \tag{7}$$

Lemma 4 *Assume condition (4) holds. Then a unique nonzero solution $\hat{x}$ of (1) (or (7)) exist.*

Proof: The cost function $J(x)$ can be shown to be strictly convex in x since $b \notin \mathcal{R}(A)$. This implies that $J(x)$ has a unique global minimum. When (4) holds we already know from Lemma 2 that $\hat{x} = 0$ can not be the global minimum. Therefore, the global minimum is necessarily unique and nonzero.

◊

We shall denote the worst-case perturbed matrix $A + \delta A^o(x)$ by $A(x)$ so that (7) is equivalent to

$$\min_x \|A(x)x - b\| . \tag{8}$$

This statement looks similar to a least-squares problem except that the coefficient matrix is dependent on the unknown x. Hence, what we have is a nonlinear least-squares problem with a special form for the coefficient matrix $A(x)$. If $A(x)$ were a constant matrix, and therefore not dependent on x, say $\bar{A}$, then we know from the geometry of least-squares estimation that the solution $\hat{x}$ is obtained by imposing the orthogonality condition $\bar{A}^T(\bar{A}\hat{x} - b) = 0$. In the BDU case (8), the coefficient matrix is a nonlinear function of x. Interestingly enough, however, it turns out that the solution $\hat{x}$ can still be characterized by similar orthogonality conditions.

In the following we distinguish between two classes of vectors x. We define

$$\mathcal{X} = \{\text{set of vectors } x \text{ such that all } x_i \neq 0\}.$$

Regular solutions.

The following result shows that the unique solution $\hat{x}$ belongs to $\mathcal{X}$ if, and only if, an element $\hat{x}$ of $\mathcal{X}$ satisfies the following orthogonality condition (see Fig. 1)

$$A^T(\hat{x})\,[A(\hat{x})\hat{x} - b] = 0. \qquad (9)$$

Since, from Lemma 3, the residual vector $A(\hat{x})\hat{x} - b$ is collinear with $A\hat{x} - b$, we obtain the equivalent condition

$$\left[A + \frac{(A\hat{x} - b)\,q^T(\hat{x})}{\|A\hat{x} - b\|}\right]^T [A\hat{x} - b] = 0. \qquad (10)$$

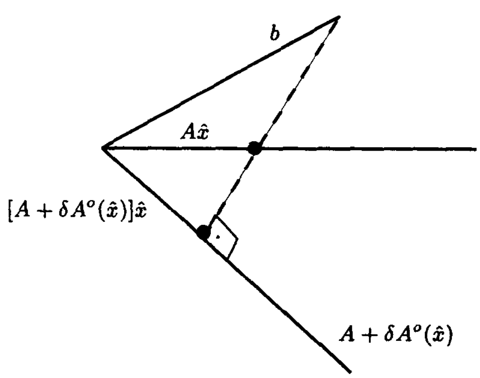

Figure 1. Orthogonality condition for BDU estimation.

Compared with least-squares theory, we can interpret the result (10) as requiring the residual vector to be orthogonal to a rank-one modification of A.

Theorem 1 *Assume (4) holds. Then an $\hat{x} \in \mathcal{X}$ is the unique solution of (7) if, and only if, (10) holds.*

Proof: Let $\hat{x} \in \mathcal{X}$ be a vector that satisfies the orthogonality condition $A^T(\hat{x})[A\hat{x} - b] = 0$ and pick any other vector x (in $\mathcal{X}$ or otherwise). Then we necessarily have

$$\|A(x)x - b\| \;\geq\; \|A(\hat{x})x - b\|.$$

This is because we already know that for a given x, $A(x)$ is a matrix that maximizes $\|(A + \delta A)x - b\|$ over δA. We can further verify that $\|A(\hat{x})\hat{x} - b\| \leq \|A(\hat{x})x - b\|$ so that $\|A(\hat{x})\hat{x} - b\| \leq \|A(x)x - b\|$, which means that $\hat{x}$ is the unique minimizer.

Conversely, suppose that $\hat{x} \in \mathcal{X}$ is a nonzero minimizer of $J(x)$. The gradient of $J(x)$ is defined at all $x \in \mathcal{X}$. Thus, setting the gradient of $J(x)$ at $x = \hat{x}$ equal to zero we obtain (10).

◇

Boundary solutions.

If a vector $\hat{x} \in \mathcal{X}$ satisfying the orthogonality condition (10) does not exist, i.e., one with all its entries $\{\hat{x}_i\}$ nonzero, then the unique minimizer belongs to the set $\mathcal{R}^n - \mathcal{X} - \{0\}$. That is, we need to examine the possibility of a solution $\hat{x}$ with one or more zero entries $\{\hat{x}_i\}$.

We illustrate this point by considering the simple case of $K = 2$. If a solution $\hat{x} \in \mathcal{X}$ does not exist, then we need to check for solutions of the form $\{0, \hat{x}_2\}$ or $\{\hat{x}_1, 0\}$. We shall refer to these as boundary solutions. In the first case, with $x_1 = 0$, the cost function collapses to

$$J_1(x) = \|A_2x_2 - b\| + \eta_2\|x_2\|.$$

A unique nonzero minimum of this cost exists if, and only if, $\eta_2 < \|A_2^Tb\|/\|b\|$, in which case it is given by the solution of the orthogonality condition

$$\left[\frac{(A_2\hat{x}_2 - b)\,\frac{\eta_2\hat{x}_2^T}{\|\hat{x}_2\|}}{\|A_2\hat{x}_2 - b\|}\right]^T (A_2\hat{x}_2 - b) = 0. \qquad (11)$$

Likewise, in the second case with $x_2 = 0$, the cost function collapses to

$$J_2(x) = \|A_1x_1 - b\| + \eta_1\|x_1\|.$$

A unique nonzero minimum of this cost exists if, and only if, $\eta_1 < \|A_1^Tb\|/\|b\|$, in which case it is given by the solution of the orthogonality condition

$$\left[\frac{(A_1\hat{x}_1 - b)\,\frac{\eta_1\hat{x}_1^T}{\|\hat{x}_1\|}}{\|A_1\hat{x}_1 - b\|}\right]^T (A_1\hat{x}_1 - b) = 0.$$

Once the unique nonzero minimizers of $J_1(x)$ and $J_2(x)$ have been determined (when they exist), we pick that solution $\{0, \hat{x}_2\}$ or $\{\hat{x}_1, 0\}$ that has the smallest cost as the unique minimizer of the original problem (1).

5. Statement of the solution

Returning to the orthogonality condition (10), we introduce the auxiliary nonnegative numbers

$$\hat{\alpha}_i \triangleq \eta_i \|A\hat{x} - b\| \cdot \|\hat{x}_i\|^{\dagger} . \qquad (12)$$

Then we can rewrite (10) in the form

$$\left(A^T A + \text{diag}\{\hat{\alpha}_1 I, \ldots, \hat{\alpha}_K I\}\hat{x}\right) = A^T b . \qquad (13)$$

Expressions (12)-(13) define a system of equations in the unknowns $\{\hat{x}, \hat{\alpha}_i\}$.

Theorem 2 *Under (2), the solution of the* BDU *problem (1) is always unique. The following facts hold:*

I. *The solution is zero ($\hat{x} = 0$) if, and only if, each η_i satisfies $\eta_i \geq \|A_i^T b\|/\|b\|$.*

II. *The solution is nonzero if, and only if, at least one η_i satisfies $\eta_i < \|A_i^T b\|/\|b\|$.*

   II.1 *The unique solution $\hat{x}$ is in $\mathcal{X}$ (i.e., with all $\hat{x}_i \neq 0$) if, and only if, an $\hat{x}$ exists that solves (10). Alternatively, this unique $\hat{x}$ can be found by solving the nonlinear system of equations (12)-(13) in $\hat{x}$ and $\hat{\alpha}_i$.*

   II.2 *If a solution $\hat{x} \in \mathcal{X}$ does not exist, then the unique minimizer is a boundary solution.*

◇

Note that the solution of the BDU problem performs automatic regularization by determining regularization parameters $\{\hat{\alpha}_i\}$. If we replace the expression for $\hat{x}$ from (13) into (12) we obtain a nonlinear system of equations in the $\{\hat{\alpha}_i\}$. Such equations can be solved by any appropriate zero-finding technique (*e.g.*, the command fsolve of Matlab[1] has been used in the simulation results below). The SVD of A can also be used to obtain better conditioned numerical solutions (cf. [2]).

6. Two Applications

Our first application is in the context of co-channel interference cancelation in array signal processing, as depicted in a simplified form in Fig. 2 for the case of two sources and four antenna elements.

[1]Matlab©is a trademark of the MathWords Inc.

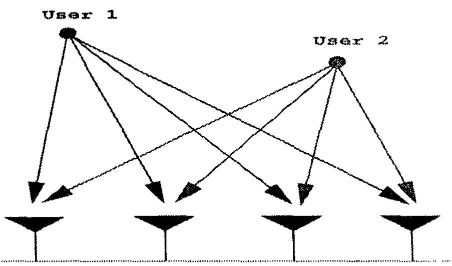

Figure 2. Spatial-processing with multiple users.

The figure shows two emitters sending at time i the signals $\{x_{1,i}, x_{2,i}\}$ from different angles to an antenna array. The antenna array has 4 elements that are equally spaced. The signal received by the elements of the antenna array can be presented in vector form as

$$b_i = A_1 x_{1,i} + A_2 x_{2,i} + v_i , \qquad (14)$$

where v_i denotes a measurement noise vector. Moreover, A_1 and A_2 are 4×1 column vectors. The $j-$th entry of A_1 is the gain from source x_1 to the $j-$th antenna. Likewise, the $j-$th entry in A_2 is the gain from source x_2 to the $j-$th antenna. In practice, these gains are estimated by a variety of methods (*e.g.*, MUSIC, ESPRIT, and many others – see [3, 4] and the many references therein).

Once the $\{A_1, A_2\}$ are known (or estimated), the common techniques in the literature proceed to recover the transmitted signals $\{x_{1,i}, x_{2,i}\}$, at each time instant i, by solving any of the following problems: least-squares, regularized least-squares, TLS [5], or generalized cross-validation (GCV) [6].

Now the data matrix A is subject to errors since it is the result of an identification procedure. There can also be different levels of errors in the different columns of A. The BDU formulation of this paper allows us to handle such situations with multiple sources of uncertainties rather naturally and allows us to incorporate into the problem formulation a-priori bounds on the sizes of the uncertainties in the estimated A_1 and A_2 (these bounds can be obtained from the identification procedure for A_1 and A_2), say $\|\delta A_1\| \leq \eta_1$ and $\|\delta A_2\| \leq \eta_2$. We can then recover the $\{x_{1,i}, x_{2,i}\}$ by solving

$$\min_{x_{1,i}, x_{2,i}} \max_{\|\delta A_1\| \leq \eta_1, \|\delta A_2\| \leq \eta_2} \left\| \begin{bmatrix} A_1 + \delta A_1 & A_2 + \delta A_2 \end{bmatrix} \begin{bmatrix} x_{1,i} \\ x_{2,i} \end{bmatrix} - b_i \right\| ,$$

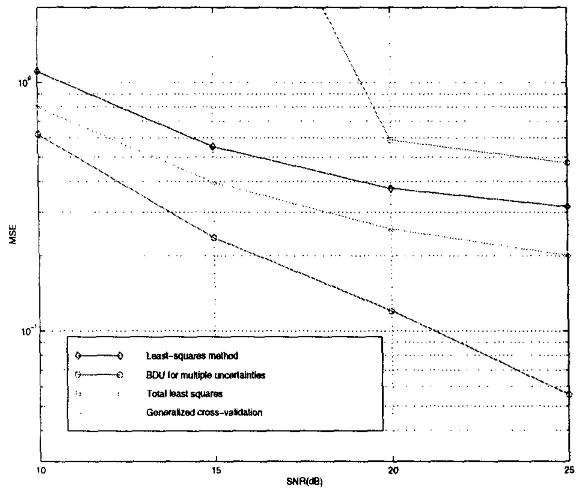

Figure 3. 4PAM modulation, $N = 4000$ **runs,** $\eta_1 \approx 7\%$, $\eta_2 \approx 22\%$.

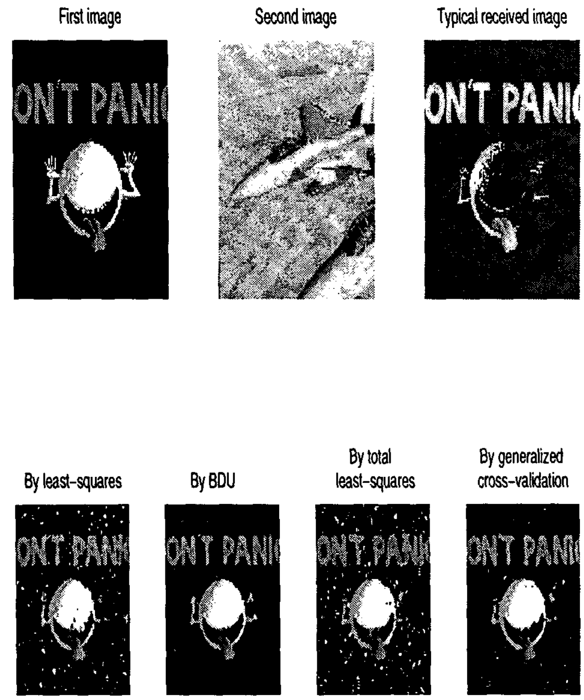

Figure 4. Image separation after median filtering.

which is a special case of (1). Fig. 3 compares the performance (in terms of mean-square error) of the BDU solution with the above alternative methods for 4PAM modulation with 7% and 22% relative uncertainties in the path gains (by relative uncertainty we mean $\eta_1/\|A_1\|$ and $\eta_2/\|A_2\|$). The top curve corresponds to total-least-squares while the bottom curve corresponds to BDU. The second curve from top is least-squares and the third curve is generalized cross-validation.

Our second application is in the context of image processing. Figure 4 repeats the same experiment as above where now the signals $\{x_{1,i}, x_{2,i}\}$ represent the pixels of two 128 × 128 images that are being transmitted over different paths. Hence, the purpose is to identify and separate the superimposed images. In this particular simulation, we took $\eta_1 = \eta_2 = 7\%$. We see that the result from the BDU solution is the clearest.

Acknowledgment. The authors are grateful to Mr. F. Xu of ULCA for his help with the simulations of Figs. 3 and 4.

References

[1] A. H. Sayed, V. H. Nascimento, and S. Chandrasekaran, Estimation and control with bounded data uncertainties, *Linear Algebra and Its Applications*, vol. 284, pp. 259–306, Nov. 1998.

[2] S. Chandrasekaran, G. H. Golub, M. Gu, and A. H. Sayed, Parameter estimation in the presence of bounded data uncertainties, *SIAM J. Matrix Analysis and Applications*, vol. 19, no. 1, pp. 235–252, Jan. 1998.

[3] H. Krim and M. Viberg, Two decades of array signal processing research, *IEEE Signal Processing Magazine*, vol. 13, no. 4, pp. 67–94, July 1996.

[4] A. J. Paulraj and C. B. Papadias, Space-time processing for wireless communications, *IEEE Signal Processing Magazine*, vol. 14, no. 6, pp. 49–83, Nov. 1997.

[5] S. Van Huffel and J. Vandewalle, *The Total Least Squares Problem: Computational Aspects and Analysis*, SIAM, Philadelphia, 1991.

[6] G. H. Golub and M. Heath, Generalized cross-validation as a method for choosing a good ridge parameter, *Technometrics*, vol. 21, no. 2, pp. 195–7, May 1979.