ROBUST ESTIMATION FOR UNCERTAIN MODELS IN A DATA FUSION SCENARIO

A. H. Sayed * **T. Y. Al-Naffouri** ** **T. Kailath** **

* *Department of Electrical Engineering, UCLA*
** *Department of Electrical Engineering, Stanford University*

Abstract: This paper formulates and solves a parameter estimation problem that shows how to combine, in a certain optimal and robust manner, measurements that arise from a finite collection of uncertain models. This scenario occurs, for example, in data fusion applications and in cases that involve systems that can operate under different failure conditions. An example in the context of macroscopic diversity in wireless cellular systems is considered. ©2000 IFAC.

Keywords: Robust estimation, data fusion, uncertain models, least-squares, regularization, CDMA, macroscopic diversity, cellular system.

1. INTRODUCTION

Modeling errors in the data are common in practice and they can be due to several factors including the approximation of complex models by simpler ones, the introduction of experimental errors while collecting data, or even the presence of unmodeled or unknown effects. Regardless of their source, modeling errors can adversely affect the performance of otherwise optimal estimators. This fact has motivated recent works on robust least-squares methods, especially in Chandrasekaran et al. (1997,1998), Ghaoui and Lebret (1997), and Sayed et al. (1998,1999,2000).

[1] **Published in *System Identification*, R. Smith, editor, vol. 3, pp. 899–904, Pergamon, UK, May 2001. [Proceedings of the 12th IFAC Symposium on System Identification, Santa Barbara, CA, June 2000.]** This work was partially support by NSF ECS-9820765 and by ARO DAAH04-96-1-0176-P00005. The work of Al-Naffouri was also supported by King Fahd University of Petroleum and Minerals, Saudi Arabia.

In the works by Sayed et al. (1998,1999,2000), general cost functions that allow for different levels and sources of bounded parametric uncertainties in the data have been proposed. These cost functions are based on constrained game-type formulations and they have been shown to lead to regularized least-squares solutions; albeit ones where the regularization parameters are constructed optimally from the nominal data and from the available information about the uncertainties. In particular, the work by Sayed and Nascimento (1999) developed a framework that can handle different classes of uncertainties in the data, as well as allow for data weighting. In the current paper, it is shown how to extend this framework to situations that involve a multitude of uncertain models; both cases of a data fusion scenario and a probabilistic (operation under failure) scenario are considered.

1.1 *Data Fusion Formulation*

Consider vector measurements $\{b_1, b_2, \ldots, b_L\}$ that arise from L uncertain models of the form

$$b_i = (A_i + \delta A_i)x + v_i, \quad i = 1, 2, \ldots, L,$$

where, for each i, v_i accounts for measurement noise and δA_i accounts for discrepancies between the given nominal matrix A_i and its actual value. Constraints on δA_i are described in the sequel.

The n-dimensional unknown parameter vector x is the same for all measurements $\{b_i\}$. This description corresponds to a situation where several distorted measurements of a single unknown vector x arise from different sources with different uncertainty models, as depicted in Fig. 1 for $L = 4$.

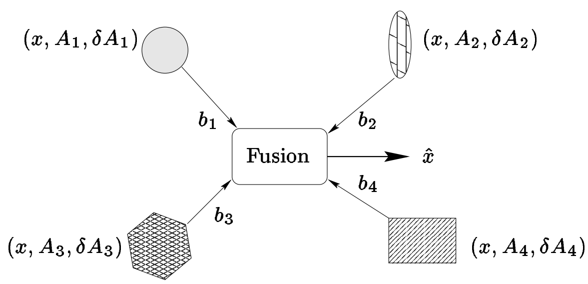

Fig. 1. *Measurements $\{b_i\}$ arising from different uncertain sources.*

Given the $\{b_i\}$, one can then pose the problem of estimating x optimally by minimizing the following worst-case criterion over all uncertainties:

$$\min_x \max_{\substack{\{\delta A_i\} \\ \{\delta b_i\}}} \left[\|x\|_Q^2 + \sum_{i=1}^{L} \|(A_i + \delta A_i)\, x - (b_i + \delta b_i)\|_{W_i}^2 \right] \quad (1)$$

where $\{Q > 0, W_i \geq 0\}$ are weighting matrices and the notation $\|z\|_\Sigma^2$ stands for the weighted norm $z^T \Sigma z$. Also, for generality, uncertainties $\{\delta b_i\}$ are included in the problem statement in order to account for possible additional distortion sources in the measurements $\{b_i\}$. In this way, each triple $\{A_i, \delta A_i, \delta b_i\}$, with the $\{\delta A_i, \delta b_i\}$ belonging to a certain set, describes an uncertain model for the data.

The above formulation amounts to a constrained game-type problem where the designer attempts to minimize the cost by picking x while the opponent attempts to maximize the cost through the selection of $\{\delta A_i, \delta b_i\}$. The game problem is constrained since bounds will be imposed on the sizes of the $\{\delta A_i, \delta b_i\}$, as explained further ahead. It is also worth remarking that problem (1) can be rewritten as

$$\min_x \max_{\substack{\{\delta A_i\} \\ \{\delta b_i\}}} \|x\|_Q^2 + \left\| \begin{bmatrix} A_1 + \delta A_1 \\ A_2 + \delta A_2 \\ \vdots \\ A_L + \delta A_2 \end{bmatrix} x - \begin{bmatrix} b_1 + \delta b_1 \\ b_2 + \delta b_2 \\ \vdots \\ b_L + \delta b_L \end{bmatrix} \right\|_{\mathcal{W}}^2$$

where $\mathcal{W} = (W_1 \oplus \ldots \oplus W_L)$. This indicates that problem (1) can be further interpreted as corresponding to solving an optimization problem with uncertainties in the rows of the data matrix $\mathrm{col}\{A_1, \ldots, A_L\}$.

1.2 *Probabilistic Formulation*

A related problem is one in which a single measurement b is available that could have risen from a selection of L models, say

$$b = (A_i + \delta A_i)x + v_i, \quad i = 1, 2, \ldots, L,$$

with a probability p_i for each possible nominal model $\{A_1, A_2, \ldots, A_L\}$, as depicted in Fig. 2 for $L = 4$. This formulation corresponds, for example, to a situation where failures can occur and the uncertain model giving rise to the measurement is therefore subject to changes according to a probability distribution.

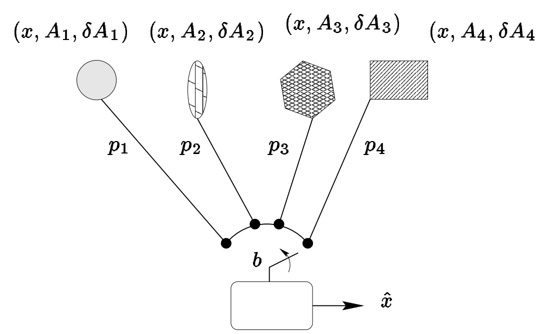

Fig. 2. *A single measurement b that could have risen from any of the models according to a certain probability distribution.*

In this case, one can seek to estimate x by minimizing the worst-case expected residual energy,

$$\min_x \max_{\substack{\{\delta A_i\} \\ \{\delta b_i\}}} \mathrm{E}\left[\|x\|_Q^2 + \|(A + \delta A)x - (b + \delta b)\|^2 \right] \quad (2)$$

where the expectation is over the nominal data $\{A_1, \ldots, A_L\}$; each having probability p_i. In other words, the quantities $\{A, \delta A, \delta b\}$ appearing in the above cost refer to each uncertainty model

$\{A_i, \delta A_i, \delta b_i\}$ with probability p_i. The maximization is performed over any specified (deterministic) characterization of the uncertainties $\{\delta A_i, \delta b_i\}$, so that problem (2) reduces to

$$\min_x \max_{\substack{\{\delta A_i\} \\ \{\delta b_i\}}} \left[\|x\|_Q^2 + \sum_{i=1}^{L} \|(A_i + \delta A_i)x - (b + \delta b_i)\|_{p_i I}^2 \right] \quad (3)$$

1.3 *Conditions on the Uncertainties*

Of course, there are several ways to describe the uncertainty sets $\{\delta A_i, \delta b_i\}$. Two convenient ones are the following (more general ones are described below):

(1) Bounded uncertainties. In this case one assumes that the uncertainties lie within balls of known radii, say

$$\|\delta A_i\| \leq \eta_{a,i}, \quad \|\delta b_i\| \leq \eta_{b,i}, \quad (4)$$

for some known scalars $\{\eta_{a,i}, \eta_{b,i}\}$. Here, the notation $\|\cdot\|$ denotes the Euclidean norm of its vector argument or the maximum singular value of its matrix argument.

(2) Factored form. In this case one assumes that the uncertainties satisfy a model of the form

$$\left[\, \delta A_i \;\; \delta b_i \,\right] \;=\; H_i S \left[\, E_{a,i} \;\; E_{b,i} \,\right], \quad (5)$$

where the $\{H_i, E_{a,i}, E_{b,i}\}$ are known matrices and S is an arbitrary contraction. The above form allows the designer to impose some structure on the uncertainty set. For example, (5) forces $[\delta A_i \;\; \delta b_i]$ to lie in the column span of H_i.

1.4 *A General Formulation*

The optimization problems (1) and (3), with uncertainties of either forms (4) or (5), can be regarded as special cases of the following more general formulation (see Sayed and Nascimento (1999)). Given data $\{A_i, b_i, H_i, W_i, Q\}$, and non-negative functions $\phi_i(x)$, solve

$$\hat{x} = \arg\min_x \max_{\{\|y_i\| \leq \phi_i(x)\}} \left[\|x\|_Q^2 + \sum_{i=1}^{L} R_i(x, y_i) \right] \quad (6)$$

where the residual energy terms $R_i(x, y_i)$ are defined by

$$R_i(x, y_i) = \|A_i x - b_i + H_i y_i\|_{W_i}^2 \quad (7)$$

In this formulation, the vectors $\{y_i\}$ play the role of uncertainties whose norms are bounded by the $\{\phi_i(x)\}$. For example, problem (3) with constraints (4) corresponds to the choices $W_i = p_i I$, $H_i = I$, and $\phi_i(x) = \eta_{a,i}\|x\| + \eta_{b,i}$. Using the constraints (5) instead, problem (3) would correspond to the choice $\phi_i(x) = \|E_{a,i}x - E_{b,i}\|$. Other choices for $\phi_i(x)$ are of course possible. It is assumed in the sequel that none of the $\phi_i(x)$ is identically zero so that none of the maximizations in (6) is trivialized.

2. SOLVING THE OPTIMIZATION PROBLEM

Since each $R_i(x, y_i)$ is dependent only on a single y_i, the maximization in (6) over the various $y_i's$ can be done independently and (6) is therefore equivalent to

$$\hat{x} = \arg\min_x \left\{ \|x\|_Q^2 + \sum_{i=1}^{L} C_i(x) \right\} \quad (8)$$

where the functions $C_i(x)$ are defined by

$$C_i(x) \triangleq \max_{\|y_i\| \leq \phi_i(x)} R_i(x, y_i) \quad (9)$$

Now since for any y_i, each residual cost $R_i(x, y_i)$ is convex in x, one concludes that each $C_i(x)$ is convex in x. Moreover, since $x^T Q x$ is strictly convex in x when $Q > 0$, it follows that the cost in (8) is strictly convex in x. This shows that problem (6) has a unique global minimum $\hat{x}$.

2.1 *The Maximization Step*

The first step towards computing $\hat{x}$ involves determining the $C_i(x)$. Now since $R_i(x, y_i)$ is convex in y_i, the maximum over y_i is achieved at the boundary, i.e., when $\|y_i\| = \phi_i(x)$. This means that the constrained problem (9) can be replaced by the unconstrained problem

$$C_i(x) = \quad (10)$$
$$\max_{(y_i, \lambda_i)} \left[\|A_i x - b_i + H_i y_i\|_{W_i}^2 - \lambda_i(\|y_i\|^2 - \phi_i^2(x)) \right]$$

where λ_i denotes a nonnegative Lagrange multiplier. This expression was used in Sayed and Nascimento (1999) to show that the dimensionality of the optimization problem (8) can be reduced by determining $C_i(x)$ through the following alternative construction.

For any nonnegative number $\lambda_i \geq \|H_i^T W_i H_i\|$, define the modified weighting matrix

$$W_i(\lambda_i) \triangleq W_i + W_i H_i \left(\lambda_i I - H_i^T W_i H_i\right)^\dagger H_i^T W_i \quad (11)$$

where the notation $X^\dagger$ denotes the pseudo-inverse of its argument. Introduce also the two-variable function

$$C_i(x, \lambda_i) \triangleq \|A_i x - b_i\|^2_{W_i(\lambda_i)} + \lambda_i \phi_i^2(x) \quad (12)$$

where x and λ_i are now treated as independent variables. Then it can be shown that the desired $C_i(x)$ in (10) can also be obtained via

$$C_i(x) = \min_{\lambda_i \geq \|H_i^T W_i H_i\|} C_i(x, \lambda_i). \quad (13)$$

In other words, $C_i(x)$ can be determined by minimizing $C_i(x, \lambda_i)$ over the single scalar parameter λ_i in the interval $[\|H_i^T W_i H_i\|, \infty)$. With this in hand, problem (8) then reduces to the following $(n+L)$-dimensional minimization problem:

$$\min_{\lambda_i \geq \|H_i^T W_i H_i\|} \min_x \left[\|x\|^2_Q + \sum_{i=1}^L C_i(x, \lambda_i)\right]. \quad (14)$$

A key property of this equivalent characterization of (8) is that the variables $\{\lambda_i, x\}$ are independent of each other. In addition, by performing the inner minimization with respect to x first, the above procedure can be further reduced to an equivalent L-dimensional minimization problem. Indeed, setting the gradient of the objective function with respect to x equal to zero leads to the equality

$$\left(\sum_{i=1}^L M_i(\lambda_i)\right) x + \frac{1}{2}\sum_{i=1}^L \lambda_i \bigtriangledown \phi_i^2(x) = \sum_{i=1}^L d_i(\lambda_i) \quad (15)$$

where $\{M_i(\lambda_i), d_i(\lambda_i)\}$ are defined by

$$M_i(\lambda_i) \triangleq \frac{1}{L} Q + A_i^T W_i(\lambda_i) A_i, \quad d_i(\lambda_i) \triangleq A_i^T W_i(\lambda_i) b_i.$$

Also, the notation $\bigtriangledown \phi_i^2(x)$ denotes the gradient of $\phi_i^2(x)$ with respect to x. Observe that the $M_i(\lambda_i)$ are positive-definite matrices. When (15) has a unique solution, say x^o (which is dependent on the $\{\lambda_i\}$), then substituting its expression into the cost function in (14) leads to a minimization problem over the $\{\lambda_i\}$ alone.

3. A SPECIAL CASE

In order to illustrate the above solution, consider the case of uncertainties $\{\delta A_i, \delta b_i\}$ that satisfy the constraints (5). Then, in this case,

$$\phi_i(x) = \|E_{a,i} x - E_{b,i}\| \quad (16)$$

so that $\nabla \phi_i^2(x) = 2E_{a,i}^T (E_{a,i} x - E_{b,i})$. Substituting into (15) shows that there is a unique solution x^o, dependent on the $\{\lambda_i\}$, and is given by

$$x^o = \left[\sum_{i=1}^L [M_i(\lambda_i) + \lambda_i E_{a,i}^T E_{a,i}]\right]^{-1} \cdot \sum_{i=1}^L \left[A_i^T W_i(\lambda_i) b_i + \lambda_i E_{a,i}^T E_{b,i}\right]. \quad (17)$$

By substituting this expression for x^o into the cost function in (14), one determines the minimum value of the innermost minimization over x. Let G denote the resulting minimum value, i.e.,

$$G(\lambda_1, \ldots, \lambda_L) = \left[\|x^o\|^2_Q + \sum_{i=1}^L C_i(x^o, \lambda_i)\right]. \quad (18)$$

Then problem (14) is reduced to the equivalent problem of minimizing this cost function over the $\{\lambda_1, \ldots, \lambda_L\}$ in the respective intervals $\lambda_i \geq \|H_i^T W_i H_i\|$. In summary, one arrives at the following conclusion.

Theorem 1. *Consider problem (6) and assume the uncertainties satisfy the constraints (5) (or, equivalently, that $\phi_i(x)$ is given by (16)). The unique global minimum $\hat{x}$ can be determined as follows. Determine first the scalars $\{\hat{\lambda}_1, \ldots, \hat{\lambda}_L\}$ that minimize the function G in (18) over the intervals $\lambda_i \geq \|H_i^T W_i H_i\|$, where x^o is defined by (17). Then set*

$$\hat{x} = \left[\widehat{Q} + \sum_{i=1}^L A_i^T \widehat{W}_i A_i\right]^{-1} \sum_{i=1}^L \left[A_i^T \widehat{W}_i b_i + \hat{\lambda}_i E_{a,i}^T E_{b,i}\right]$$

where the modified weighting matrices $\{\widehat{Q}, \widehat{W}_i\}$ are computed from the given $\{Q, W_i\}$ in terms of the $\{\hat{\lambda}_i\}$ as follows:

$$\widehat{Q} = Q + \sum_{i=1}^L \hat{\lambda}_i E_{a,i}^T E_{a,i}$$

$$\widehat{W}_i = W_i + W_i H_i (\hat{\lambda}_i I - H_i^T W_i H_i)^\dagger H_i^T W_i$$

4. REGULARIZED LEAST-SQUARES

It is worth comparing the above solution to the one that would be obtained had the presence of the data uncertainties been ignored, which corresponds to assuming all the y_i or the H_i to be zero in (6). In this case, problem (6) reduces to the standard regularized least-squares problem

$$\hat{x} = \arg\min_x \left[\|x\|_Q^2 + \sum_{i=1}^{L} \|A_i x - b_i\|_{W_i}^2 \right] \quad (19)$$

whose unique solution is easily seen to be

$$\hat{x} = \left[Q + \sum_{i=1}^{L} A_i^T W_i A_i \right]^{-1} \left[\sum_{i=1}^{L} A_i^T W_i b_i \right]$$

Comparing this expression with the one for $\hat{x}$ in the statement of Theorem 1, it is seen that the solution to the constrained game problem essentially amounts to correcting the given weighting matrices $\{Q, W_i\}$ to new matrices $\{\widehat{Q}, \widehat{W}_i\}$ that account for the uncertainty model. The additional terms $\{\hat{\lambda}_i E_{a,i}^T E_{b,i}\}$ that appear in the expression for $\hat{x}$ in the theorem are typical of regularized least-squares problems with coupling.

5. AN APPLICATION: DIVERSITY IN A WIRELESS CELLULAR SYSTEM

An interesting application for the framework developed in this article arises in the context of cellular wireless systems. Fig. 3 shows a typical (hexagonal) cell in this system with multiple mobile units and antennas. One of the main difficulties in this application is that of fading in which the strength of the received signal changes as the mobile unit moves in the cell (see Stuber (1996)). Two mechanisms contribute to this variation and they are called slow and fast fading.

Microdiversity is used to combat fast fading. Here, multiple antennas are employed at both the mobile unit and the base station to receive *uncorrelated* copies of the signal.[2] The signals are then combined constructively enhancing the overall signal strength.

While we can use the same mechanism to combat slow fading, the various antennas have to be placed in *geographically different* locations in this case to ensure that the received signals are uncorrelated; hence the name macrodiversity (see Stuber (1996)). While such distributed processing is possible for the mobile-to-network link (uplink), it is obviously impossible to implement on the reverse link (downlink). In what follows we shall discuss in more detail how macrodiversity manifests itself in a CDMA cellular system (see Weiss (1999)).

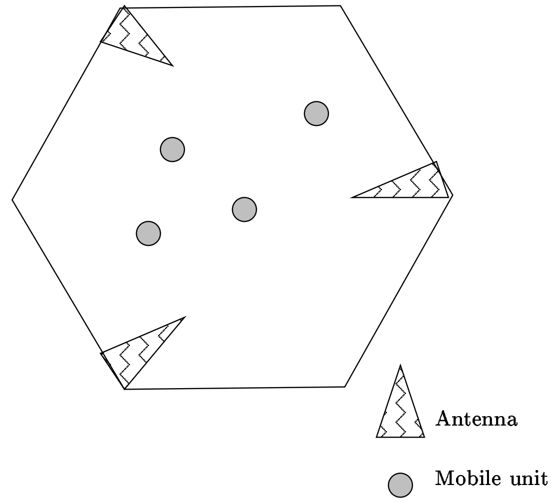

Fig. 3. *A hexagonal cell with multiple mobile units and antennas for macrodiversity processing.*

5.1 *Macrodiversity on the Uplink*

To enhance reception on the uplink (mobile to base station), several antennas are deployed in geographically different locations in the cell. In the work of Turkmani (1992), it is argued that three antennas are enough. In a CDMA system, the signal b_i that is received at the ith antenna has the form[3] $b_i = A_i x + v_i$, where A_i is the channel matrix (channel gains) from the various mobiles in the cell to the ith antenna, and v_i is additive noise. Here, x is a vector of transmitted signals from the mobile units. The matrix A_i is usually known up to an uncertainty δA_i due to synchronization errors and multipath propagation. The measurement vectors $\{b_i\}$ at the antennas are relayed to a central base station whose task is to *fuse* this information and to come up with an estimate for x, thus giving rise to problem (1).

[2] The copies are said to be uncorrelated in the sense that they experience independent channel gains.

[3] The receiver actually receives a *scalar* signal. It is only after passing this signal through a bank of matched filters (one for each user) that we get the *vector* b_i.

5.2 *Macrodiversity on the Downlink*

The mobile unit cannot have antennas at geographically different locations. Thus, at a first glance, one would suggest that all antennas of the cell transmit the same signal to the mobile unit. While this might enhance operation at a particular mobile unit, it will overwhelm other units in a CDMA system due to multiuser interference.

Instead, as suggested in Weiss (1999), the central base station decides on one antenna for downlink transmission; the one that guarantees strongest reception at the mobile unit. The most that the mobile station knows is that the signal b_i that it has received arrives from antenna i with probability p_i, i.e. $b_i = (A_i + \delta A_i)x + v_i$ where δA_i is again some uncertainty due to synchronization errors and multipath propagation. Thus, the problem that we have to solve now has a probabilistic formulation similar to (3). The probabilities p_i can be calculated using the direction of arrival information for example.

5.3 *A Numerical Example*

In this example we assume that x is 2-dimensional, which corresponds to the case of 2 mobile units in the cell. Each unit transmits data that is chosen uniformly from the PAM distribution $\{-3, -1, 1, 3\}$ and, hence, the data from each unit has variance $\sigma_x^2 = 5$. The noise vector v_i is assumed to have a correlation matrix of the form $\sigma_v^2 I$, for some noise variance σ_v^2 that is chosen to enforce desired SNR levels.

Two antennas are assumed to exist in the cell and the nominal channels from the users to each antenna are given by

$$A_1 = \begin{bmatrix} 1.0 & -0.5 \\ 0.2 & 0.1 \end{bmatrix}, \quad A_2 = \begin{bmatrix} -0.32 & 0.76 \\ 0.08 & 0.024 \end{bmatrix}.$$

The uncertainties $\phi_i(x)$ in (16) are chosen such that $E_{a,1} = 0.171 I_2$, $E_{a,2} = 0.2072 I_2$ and $E_{b,i} = 0$, which correspond to the maximal relative error bounds of approximately 15% and 25% for A_1 and A_2, respectively (i.e., $\|\delta A_1\| / \|A_1\|$ is at most 15%).

The simulation was run for several values of the SNR between -30 and 35 dBs. For each value of the SNR, 100 data points x were generated, in addition to random $\{\delta A_1, \delta A_2\}$ within the specified bounds. The estimate $\hat{x}$ was obtained by the method outlined in Theorem 1 and also by the regularized least-squares method of Sec. 4.

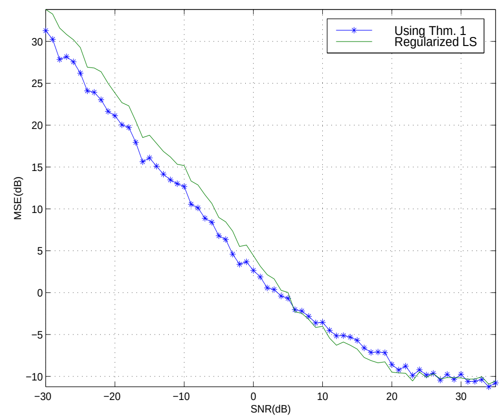

Fig. 4. *MSE curves for 4PAM modulation in a cellular wireless system scenario.*

The comparison between both methods is done in terms of the relative mean-square error:

$$\text{Relative MSE} = 10 \log_{10} \left(\frac{\frac{1}{100} \sum_{j=1}^{100} \|x_j - \hat{x}_j\|^2}{\sigma_x^2} \right)$$

where x_j stands for the data at iteration j, and $\hat{x}_j$ denotes the corresponding estimate. The resulting MSE curves are shown in Fig. 4. A point to stress here is that the uncertainties $\{\delta A_1, \delta A_2\}$ were generated randomly during the experiments; by design, a more robust performance would result when these uncertainties are close to their worst-case values.

6. CONCLUDING REMARKS

This paper formulates a robust estimation problem for data fusion applications. The solution turns out to be in regularized form, albeit one that is applied to modified weighting matrices rather than the original weighting matrices (cf. the statement of Theorem 1). An example in the context of macrodiversity design in a wireless cellular network was considered. Further studies are required to better understand the statistical properties of the proposed estimator; as well as to suggest alternative computational methods and approximations.

REFERENCES

Chandrasekaran, S., G. Golub, M. Gu, and A. H. Sayed (1997). Parameter estimation in the presence of bounded modeling errors, *IEEE Signal Processing Letters*, **4**, no. 7, pp. 195–197, July 1997. See also *SIAM J. Matrix Analysis and Applications*, **19**(1), pp. 235–252, 1998.

Ghaoui, L. E. and H. Hebret (1997). Robust solutions to least-squares problems with uncertain data. *SIAM J. Matrix Anal. Appl.*, **18**, pp. 1035–1064.

Sayed, A. H., V. H. Nascimento, and S. Chandrasekaran (1998). Estimation and control with bounded data uncertainties. *Linear Algebra and Its Applications*, **284**, pp. 259–306.

Sayed, A. H. and V. H. Nascimento (1999). Design criteria for uncertain models with structured and unstructured uncertainties. In *Robustness in Identification and Control*, A. Garulli, A. Tesi, and A. Vicino, editors, pp. 159–173, Springer Verlag, London.

Sayed, A. H. and S. Chandrasekaran (2000). Parameter estimation with multiple sources and levels of uncertainties. *IEEE Transactions on Signal Processing*, **48**.

Stuber, G. L. (1996). *Principles of Mobile Communication,* Kluwer Academic Publishing.

Turkmani, A. M. D. (1992). Probability of error for M-branch macroscopic diversity. *IEE Proceedings-I*, **139**, no. 1, pp. 71–78.

Weiss, U. (1999). Designing macroscopic diversity cellular systems. *Proc. IEEE Vehicular Technology Conference* **3**, pp. 2054 -2058.