Iterative Solutions of Min-Max Parameter Estimation with Bounded Data Uncertainties

Ali H. Sayed — Dept. of Electrical Engineering, University of California, Los Angeles, CA 90095, USA

Andrea Garulli — Facolta' di Ingegneria, Universita' di Siena, Via Roma 56, 53100 Siena, Italy

and

S. Chandrasekaran — Dept. of Elec. and Comp. Eng., University of California, Santa Barbara, CA 93106, USA

ABSTRACT

This paper deals with the important problem of parameter estimation in the presence of bounded data uncertainties. Its recent closed-form solution in [1] leads to more meaningful results than alternative methods (e.g., total least-squares and robust estimation), when a priori bounds about the uncertainties are available. The derivation in [1] requires the computation of the SVD of the data matrix and the determination of the unique positive root of a nonlinear equation. This paper establishes the existence of a fundamental contraction mapping and uses this observation to propose an approximate recursive algorithm that avoids the need for explicit SVDs and for the solution of the nonlinear equation. Simulation results are included to demonstrate the good performance of the recursive scheme.

I. INTRODUCTION

The central problem in estimation is to recover, to good accuracy, a set of unobservable parameters from corrupted data. Several optimization criteria have been used for estimation purposes, but the most important, at least in the sense of having had the most applications, are criteria that are based on quadratic cost functions. The most striking among these is the linear least-squares criterion, which enjoys widespread popularity in many diverse areas as a result of its attractive computational and statistical properties. But many alternative optimization criteria have been proposed over the years in order to improve the performance of standard least-squares estimators in the presence of data uncertainties (e.g., [2-4]). Among these criteria we mention regularized least-squares, Ridge regression, total least-squares, and robust (or H^{∞}) estimation. They all allow, in one way or another, to incorporate some a priori information about the unknown parameter into the problem statement. Nevertheless, they still may unnecessarily over-emphasize the effect of noise and of the uncertainties and can, therefore, lead to very conservative results.

In [1], a new formulation for parameter estimation in the presence of bounded data uncertainties has been posed and solved. The new method is especially useful when the measured data and the used model are uncertain *and* when a priori bounds on the uncertainties are available. In this way, the new problem formulation leads to solutions that are more meaningful especially when compared with other methods such as total least-squares and robust estimation. The reason for the more meaningful results is that the new formulation guarantees a robust performance with respect to uncertainties that are known to lie within certain bounds. This is in contrast to earlier robust designs that try to achieve a robust performance for any possible uncertainty and can therefore lead to overly conservative solutions.

The solution in [1] requires the computation of the SVD of a data matrix and the determination of the unique positive root of a nonlinear equation. In this paper, we show that some fundamental equations in [1] induce a contractive mapping. By invoking the Contraction Mapping Theorem [5], we further show that the unique fixed point of the mapping can be approximated to good accuracy via an iterative scheme. In so doing, we derive an approximate recursive scheme, similar in nature to RLS (recursive least-squares), that allows us to update the solution of the new estimation problem without the need for explicit SVDs and for the solution of the nonlinear equation.

II. PROBLEM FORMULATION

In [1], the following new estimation problem has been formulated and solved; it allows a priori bounds on the uncertain data to be explicitly incorporated into the problem formulation.

Let $A \in \mathbf{R}^{m\times n}$ be a given full rank matrix with $m \geq n$ and let $b \in \mathbf{R}^m$ be a given vector. The quantities (A, b) are assumed to be linearly related via an unknown vector of parameters $x \in \mathbf{R}^n$, $b = A\cdot x + v$, where $v \in \mathbf{R}^m$ explains the mismatch between $A\cdot x$ and b. We assume that the "true" coefficient matrix is $A + \delta A$, and that we only know an upper bound on the perturbation δA, say $\|\delta A\|_2 \leq \eta$. Likewise, we assume that the "true" observation vector is $b + \delta b$, and that we know an upper bound η_b on the

This material is based on work supported by the National Science Foundation under Award No. MIP-9409319.

perturbation δb, say $\|\delta b\|_2 \leq \eta_b$. The notation $\|\cdot\|_2$ denotes either the 2−induced norm of its matrix argument or the Euclidean norm of its vector argument.

We pose the problem of finding an estimate $\hat{x}$ that performs "well" for any possible perturbation $(\delta A, \delta b)$. That is, we would like to determine, if possible, an $\hat{x}$ that solves

$$\min_{\hat{x}} \left(\max_{\|\delta A\|_2 \leq \eta,\ \|\delta b\|_2 \leq \eta_b} \|\,(A + \delta A) \cdot \hat{x} - (b + \delta b)\|_2 \right) . \quad (1)$$

Any value that we pick for $\hat{x}$ would lead to many residuals norms, $\|\,(A + \delta A)\cdot\hat{x} - (b+\delta b)\|_2$, one for each possible choice of A in the disc $(A+\delta A)$ and b in the disc $(b+\delta b)$. We want to determine the particular value(s) for $\hat{x}$ whose maximum residual is the least possible. It turns out that this problem always has a unique solution except in a special degenerate case in which the solution is nonunique.

The problem also admits an interesting geometric formulation. For this purpose, and for the sake of illustration, assume we have a unit-norm vector b, $\|b\|_2 = 1$, with no uncertainties in it ($\eta_b = 0$; it turns out that the solution does not depend on η_b). Assume further that A is simply a column vector, say a, with $\eta \neq 0$, and consider (1) in this setting:

$$\min_{\hat{x}} \left(\max_{\|\delta a\|_2 \leq \eta} \|\,(a + \delta a) \cdot \hat{x} - b\|_2 \right) .$$

The situation is depicted in Fig. 1. The vectors a and b are indicated in thick black lines. The vector a is shown in the horizontal direction and a circle of radius η around its vertex indicates the set of all possible vertices for $a + \delta a$. It can be verified that the solution can be obtained by drawing a perpendicular from b to the lower tangential line θ_1. The segment r_1 denotes the optimum residual. More details can be found in [1].

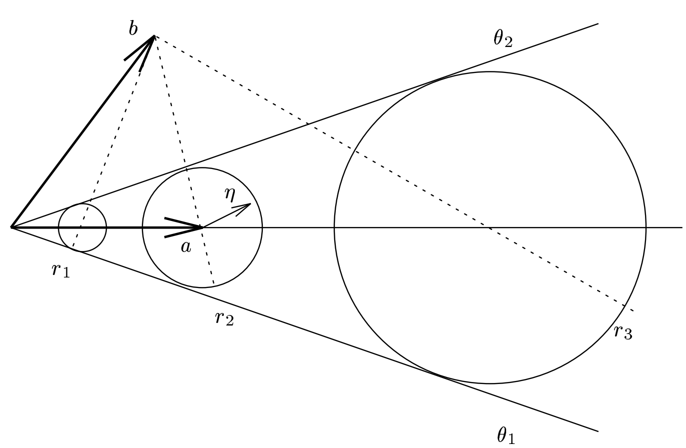

Fig. 1. *Geometric construction of the solution for a simple example.*

III. AN ALGEBRAIC SOLUTION

It can be verified that problem (1) reduces to the equivalent minimization problem:

$$\min_{\hat{x}} \left(\|A \cdot \hat{x} - b\|_2 + \eta \cdot \|\hat{x}\|_2 + \eta_b \right) , \quad (2)$$

where the cost function $\mathcal{L}(\hat{x}) = \|A\cdot\hat{x} - b\|_2 + \eta\cdot\|\hat{x}\|_2 + \eta_b$ is convex in $\hat{x}$. Note that it involves the Euclidean norms of certain vectors rather than their *squared* Euclidean norms (as in regularized least-squares problems). The following theorem summarizes the main result in [1].

Theorem 1. *Let $A \in \mathbb{R}^{m\times n}$, with $m \geq n$, be full rank, and $b \in \mathbb{R}^m$. Assume that b does not belong to the column span of A. Then the solution of the min-max estimation problem can be constructed as follows. Introduce the SVD of A: $A = U \cdot \begin{bmatrix} \Sigma^T & 0 \end{bmatrix}^T \cdot V^T$, partition the vector $U^T b$ into $U^T \cdot b = \begin{bmatrix} c^T & d^T \end{bmatrix}^T$, where $c \in \mathbf{R}^n$ and $d \in \mathbf{R}^{m-n}$, and introduce the secular equation*

$$\alpha = f(\alpha) \quad (3)$$

where

$$f(\alpha) = \eta\, \frac{\left\{\|\, d\,\|_2^2 + \alpha^2 \cdot \|\, \left(\Sigma^2 + \alpha I\right)^{-1} c\,\|_2^2\right\}^{1/2}}{\|\,\Sigma\,(\Sigma^2 + \alpha I)^{-1} c\,\|_2} . \quad (4)$$

Define $\tau = \frac{\|A^T b\|_2}{\|b\|_2}$. Then

1. *If $\eta \geq \tau$, the unique solution of (1) is $\hat{x} = 0$.*
2. *If $\eta < \tau$, the secular equation (3) has a unique positive solution $\hat{\alpha}$ and the unique solution of (1) is given by*

$$\hat{x} = \left(A^T A + \hat{\alpha} I\right)^{-1} A^T b . \quad (5)$$

It also follows that $\hat{\alpha}$ is equal to

$$\hat{\alpha} = \eta\, \frac{\|\, A\hat{x} - b\,\|_2}{\|\, \hat{x}\,\|_2} . \quad (6)$$

Remark. If b belongs to the column space of A, the solution of problem (1) is only slightly more involved (see [1] for details). The basic task, however, is still to find the unique positive solution of the secular equation (3).

According to Theorem 1, the solution of the min-max estimation problem (1) requires the determination of the unique positive solution $\hat{\alpha}$ of the secular equation (3). This task can be performed within any desired precision by using, for example, a bisection search method. This procedure may, however, require a large number of evaluations of the function $f(\cdot)$ since an a priori upper bound on $\hat{\alpha}$ is not available.

We now show that a good approximation for $\hat{\alpha}$ can be obtained by alternatively iterating the map defined by $f(\cdot)$. This will lead us to propose a recursive scheme for updating the parameter estimates as well.

IV. A FUNDAMENTAL CONTRACTION MAPPING

Define the recursive equation

$$\alpha^{(i+1)} = f(\alpha^{(i)}) , \quad \alpha^{(0)} = \text{initial condition.} \quad (7)$$

The following central result can be established by invoking the Contraction Mapping Theorem [5].

Theorem 2. *Assume $\eta < \tau$. For any positive initial value $\alpha^{(0)}$, it holds that $\lim_{i\to\infty} \alpha^{(i)} = \hat{\alpha}$, where $\hat{\alpha}$ is the unique positive solution of the secular equation (3).*

Proof: In view of Thm. 1, the condition $\eta < \tau$ guarantees the existence of a unique $\hat{\alpha} > 0$ satisfying $\hat{\alpha} = f(\hat{\alpha})$. Moreover, it can be verified that $f(0) > 0$, $f(\alpha)$ is continuous in α, and $f'(\alpha) \geq 0$ for any $\alpha \geq 0$ (the proof of this last property involves some tedious calculations that we omit here).

It then follows that $f(\alpha) \geq \alpha$ for every $\alpha \leq \hat{\alpha}$. Indeed, if for some $\tilde{\alpha} < \hat{\alpha}$ we have $f(\tilde{\alpha}) < \tilde{\alpha}$, and since $f(0) > 0$, we conclude by the continuity of f that there must exist an $0 < \bar{\alpha} < \tilde{\alpha} < \hat{\alpha}$ such that $f(\bar{\alpha}) - \bar{\alpha} = 0$. This contradicts the fact that $\hat{\alpha}$ is the only positive root of $f(\alpha) - \alpha = 0$.

Consequently, for any initial condition $\alpha^{(0)} < \hat{\alpha}$ we obtain that the resulting $\alpha^{(i)}$ is a nondecreasing sequence. Let I be an index such that $\alpha^{(I)} \leq \hat{\alpha}$. The fact that $f(\alpha)$ is a nondecreasing function shows that $\alpha^{(I+1)} = f(\alpha^{(I)}) \leq f(\hat{\alpha}) = \hat{\alpha}$ and, hence, $\alpha^{(I+1)} \leq \hat{\alpha}$. This establishes that $\alpha^{(i)} \leq \hat{\alpha}$ for all i, which means that the sequence $\{\alpha^{(i)}\}$ is bounded from above and therefore converges to some point $\alpha^{(\infty)} \leq \hat{\alpha}$. By continuity of f, we must have $\alpha^{(\infty)} = f(\alpha^{(\infty)})$ and, by uniqueness of the positive root $\hat{\alpha}$ we conclude that $\alpha^{(\infty)} = \hat{\alpha}$.

Similar arguments can be used to establish the convergence of the sequence $\{\alpha^{(i)}\}$ to $\hat{\alpha}$ for any initial condition $\alpha^{(0)} > \hat{\alpha}$.

■

We should note that the secular equation (3) is obtained by substituting (5) into (6). The iterative scheme (7) then corresponds to a successive approximation procedure with repeated applications of the function f. Alternative iterative schemes can be developed by combining expressions (5) and (6) differently. We forgo the details here.

Theorem 2 suggests that recursion (7) can be used to approximate the exact solution of the min-max estimation problem. Starting from any $\alpha^{(0)} > 0$ and computing p iterations of the map (7), we can approximate $\hat{x}$ in (5) with $x^{(p)} = (A^T A + \alpha^{(p)} I)^{-1} A^T b$. Several simulations on randomly generated data (see further ahead) have shown that in general good approximations can be obtained with very few iterations. This is particularly useful in recursive estimation contexts, as we explain in the next section.

V. RECURSIVE MIN-MAX ESTIMATION WITH BOUNDED DATA UNCERTAINTIES

Consider the linear regression model

$$y_t = (a_t + \delta a_t)^T x + v_t \quad , \quad t = 1, 2, \ldots \tag{8}$$

where $y_t \in \mathbb{R}$ is the output, $(a_t + \delta a_t) \in \mathbb{R}^n$ the regression vector, $x \in \mathbb{R}^n$ the unknown parameter vector, and $v_t \in \mathbb{R}$ a measurement noise affecting the output. Assume that the regression vector is not known exactly, while a_t and y_t are observed and a bound on the perturbation δa_t is available. In particular, a time-variant upper bound on the 2-induced norm of the matrix

$$\delta A_t = \begin{bmatrix} \delta a_1^T \\ \vdots \\ \delta a_t^T \end{bmatrix} \tag{9}$$

is known, i.e.$\| \delta A_t \|_2 \leq \eta_t$, where $\{\eta_t\}$ is a sequence of positive real numbers. Also, $b_t = \text{col}\{y_1, \ldots, y_t\}$. The recursive min-max estimation problem that we are interested in is to recursively time-update the solutions $\hat{x}_t$ of:

$$\min_{x_t} \max_{\|\delta A_t\|_2 \leq \eta_t} \| (A_t + \delta A_t) x_t - b_t \|_2 \tag{10}$$

Define $\tau_t = \frac{\|A_t^T b_t\|_2}{\|b_t\|_2}$. Let $\{\hat{x}_t\}_{t=t_0}^N$ denote the successive solutions for $t = t_0, \ldots, N$ of problem (10), where we are assuming that each b_t does not belong to the column space of the corresponding A_t. Define also $h_{t+1} = \left(A_{t+1}^T A_{t+1} + \hat{\alpha}_t I\right)^{-1} A_{t+1}^T b_{t+1}$. Comparing with the expression for $\hat{x}_{t+1} = \left(A_{t+1}^T A_{t+1} + \hat{\alpha}_{t+1} I\right)^{-1} A_{t+1}^T b_{t+1}$, we see that h_{t+1} approximates $\hat{x}_{t+1}$ by using $\hat{\alpha}_t$ instead of $\hat{\alpha}_{t+1}$.

Theorem 3. *At any particular time instant t, given $\hat{x}_t$, we can update it to $\hat{x}_{t+1}$ as follows: $\hat{x}_{t+1} = 0$ if $\eta_{t+1} \geq \tau_{t+1}$. Otherwise,*

$$\begin{aligned} h_{t+1} &= \hat{x}_t + \frac{P_t a_{t+1}}{1 + a_{t+1}^T P_t a_{t+1}} (y_{t+1} - a_{t+1}^T \hat{x}_t) \ , \\ \hat{x}_{t+1} &= [I - (\hat{\alpha}_{t+1} - \hat{\alpha}_t) P_{t+1}] \, h_{t+1} \ , \\ P_{t+1}^{-1} &= P_t^{-1} + a_{t+1} a_{t+1}^T + (\hat{\alpha}_{t+1} - \hat{\alpha}_t) I \ , \end{aligned} \tag{11}$$

where $\{\hat{\alpha}_t, \hat{\alpha}_{t+1}\}$ are the unique positive solutions of $\alpha_t = f_t(\alpha_t)$ and $\alpha_{t+1} = f_{t+1}(\alpha_{t+1})$.

Proof: Define $P_t^{-1} = A_t^T A_t + \hat{\alpha}_t I$. Then, since $A_{t+1}^T A_{t+1} = A_t^T A_t + a_{t+1} a_{t+1}^T$, we obtain (11). Moreover, by Thm. 1, $\hat{x}_{t+1} = P_{t+1} A_{t+1}^T b_{t+1}$. But since $A_{t+1}^T b_{t+1} = A_t^T b_t + a_{t+1} y_{t+1}$, we obtain by applying the matrix inversion formula to (11) the desired time-update expression for $\hat{x}_{t+1}$.

■

The recursive algorithm of Thm. 3 still requires the computation of the unique positive solution $\hat{\alpha}_t$ of the secular equation equation (3) at each time instant t. This task can be avoided if we replace the exact solution $\hat{\alpha}_t$ by an approximate solution, say $\overline{\alpha}_t$, that we obtain via an iterative scheme.

Suppose that at time t an approximation $\overline{\alpha}_t$ of $\hat{\alpha}_t$ is available. Then, one can consider computing a fixed number, say p, of iterations of the map

$$\alpha^{(i+1)} = f_{t+1}(\alpha^{(i)}) \tag{12}$$

with initial condition $\alpha^{(0)} = \overline{\alpha}_t$, and then choose $\overline{\alpha}_{t+1} = \alpha_{t+1}^{(p)}$ as an approximation for the exact value $\hat{\alpha}_{t+1}$. In particular, if we choose $p = 1$, we obtain a recursive relation for updating the approximations in time:

$$\overline{\alpha}_{t+1} = f_{t+1}(\overline{\alpha}_t). \tag{13}$$

This expression can be further reworked as follows. Let $\overline{x}_t = (A_t^T A_t + \overline{\alpha}_t I)^{-1} A_t^T b_t$ be the approximation of $\hat{x}_t$ at time t. Define $\overline{h}_{t+1} = \left(A_{t+1}^T A_{t+1} + \overline{\alpha}_t I\right)^{-1} A_{t+1}^T b_{t+1}$. Then, since f in Thm. 1 is obtained by substituting (5) into (6), the map (13) can be written as

$$\overline{\alpha}_{t+1} = f_{t+1}(\overline{\alpha}_t) = \eta_{t+1} \frac{\| A_{t+1}\overline{h}_{t+1} - b_{t+1} \|_2}{\| \overline{h}_{t+1} \|_2} \tag{14}$$

where the dependence on $\overline{\alpha}_t$ is implicit in $\overline{h}_{t+1}$. Defining $\overline{P}_t = (A_t^T A_t + \overline{\alpha}_t I)^{-1}$ and following the proof of Thm. 3, we get

$$\overline{h}_{t+1} = \overline{x}_t + \frac{\overline{P}_t a_{t+1}}{1 + a_{t+1}^T \overline{P}_t a_{t+1}} (y_{t+1} - a_{t+1}^T \overline{x}_t) \; . \tag{15}$$

A fully recursive expression of (13) can be obtained by substituting (15) into (14), and taking into account that $\overline{x}_t = \overline{P}_t A_t^T b_t$.

Iterative Min-Max Algorithm.*Set* $\overline{x}_{t_0} = \hat{x}_{t_0}$ *and* $\overline{\alpha}_{t_0} = \hat{\alpha}_{t_0}$, *and let* $\overline{P}_{t_0} = (A_{t_0}^T A_{t_0} + \overline{\alpha}_{t_0} I)^{-1}$, $z_{t_0}^2 = \|b_{t_0}\|_2^2$. *For* $t = t_0, \ldots, N$, *do*

$$\begin{aligned}
\overline{h}_{t+1} &= \left\{ \overline{x}_t + \frac{\overline{P}_t a_{t+1}}{1 + a_{t+1}^T \overline{P}_t a_{t+1}} (y_{t+1} - a_{t+1}^T \overline{x}_t) \right\} \\
z_{t+1}^2 &= z_t^2 + y_{t+1}^2 \\
p_t &= \overline{h}_{t+1}^T (\overline{P}_t^{-1} + a_{t+1} a_{t+1}^T - \overline{\alpha}_t I) \overline{h}_{t+1} \\
q_t &= 2\overline{h}_{t+1}^T (\overline{P}_t^{-1} \overline{x}_t + a_{t+1} y_{t+1}) \\
\overline{\alpha}_{t+1} &= \frac{\eta_{t+1}}{\| \overline{h}_{t+1} \|_2} \left[z_{t+1}^2 + p_t - q_t \right]^{1/2} \\
\overline{P}_{t+1}^{-1} &= \overline{P}_t^{-1} + a_{t+1} a_{t+1}^T + [\overline{\alpha}_{t+1} - \overline{\alpha}_t] I \qquad (16) \\
\overline{x}_{t+1} &= \left[I - \Delta\overline{\alpha}_{t+1} \overline{P}_{t+1} \right] \overline{h}_{t+1}
\end{aligned}$$

The above recursive algorithm requires the inversion of the $(n \times n)$-matrix $\overline{P}_t^{-1}$ at every step. Therefore, the computational complexity of a single iteration is $O(n^3)$ as it stands. This cost can be reduced to $O(n^2)$ by using (16) in order to efficiently update the eigendecomposition of $\overline{P}_t^{-1}$ (or of $\overline{P}_t$). This is because $\overline{P}_{t+1}^{-1}$ is obtained as a rank-one update of $\overline{P}_t^{-1}$ in addition to a scalar multiple of the identity. In this case, the numerically stable $O(n^2)$ algorithm developed in [6] for updating the SVD of rank-one matrix updates can be used to reduce the cost of the algorithm to $O(n^2)$. The details will be pursued elsewhere.

VI. SIMULATIONS

Consider the recursive min-max estimation problem in the simple case when the $\{a_t, \delta a_t\}$, $t = 1, 2, \ldots$, are scalars ($n = 1$) and the aim is to estimate the real parameter $x = 1$. The data a_t, the perturbation δa_t, and the noise v_t are generated randomly.

In Fig. 2(a), the exact solution $\hat{x}_t$ provided by Theorem 3 is compared to the approximation $\overline{x}_t$ computed according to the above algorithm, which has been initialized at time $t_0 = 1$ with a random positive value $\overline{\alpha}_1$. It can be seen that in few steps $\overline{x}_t$ gets very close to the exact solution and then tracks it almost perfectly. Fig. 2(b) shows that the same happens to $\overline{\alpha}_t$ with respect to $\hat{\alpha}_t$. As one might expect, simulations show that the approximation error can be further reduced by iterating the map (12) more than once every time instant, i.e. by choosing $p > 1$.

Several simulations have shown that the convergence rate of the map (12) becomes slower when η is close to τ. Therefore, the same experiment described above has been repeated with the choice $\eta_t = 0.9\tau_t$, and the results are reported in Fig. 3. Once again, the approximate solution is able to track the exact one very well.

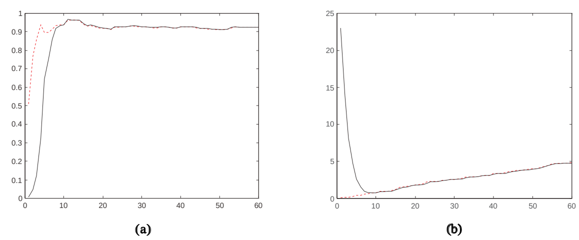

Fig. 2. (a) Exact solution $\hat{x}_t$ (dashed line) and its approximation $\overline{x}_t$ (continuous line) for the recursive min-max estimation problem. (b) $\hat{\alpha}_t$ (dashed line) and $\overline{\alpha}_t$ (continuous line).

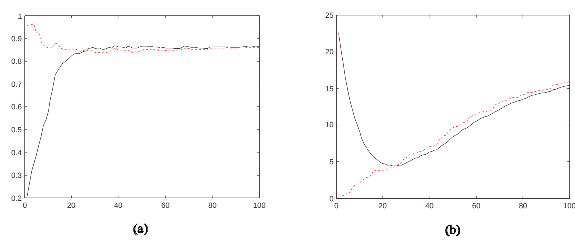

Fig. 3. (a) $\hat{x}_t$ (dashed line) and $\overline{x}_t$ (continuous line) for the recursive min-max estimation problem with $\eta_t = 0.9\tau_t$. (b) $\hat{\alpha}_t$ (dashed line) and $\overline{\alpha}_t$ (continuous line).

References

[1] S. Chandrasekaran, G. Golub, M. Gu, and A. H. Sayed, Parameter estimation in the presence of bounded data uncertainties, *submitted for publication.*

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

[3] M. Green and D. J. N. Limebeer, *Linear Robust Control*, Prentice Hall, NJ, 1995.

[4] G. Golub and C. Van Loan, *Matrix Computations*, The Johns Hopkins University Press, 1989.

[5] D. G. Luenberger. *Optimization by Vector Space Methods*, Wiley, NY, 1969.

[6] M. Gu *Studies in Numerical Linear Algebra*, Ph.D. Dissertation, Computer Science Dept., Yale University, 1993.