

EXPONENTIALLY-WEIGHTED ITERATIVE SOLUTIONS FOR WORST-CASE PARAMETER ESTIMATION *

A. H. SAYED, A. GARULLI, V. H. NASCIMENTO, AND S. CHANDRASEKARAN

Abstract

This paper proposes an iterative scheme for worst-case parameter estimation in the presence of bounded model uncertainties. The algorithm distinguishes itself from other estimation schemes, such as errors-in-variables and H_∞ methods, in that it leads to less conservative designs since it explicitly incorporates an a-priori bound on the size of the uncertainties. It also employs an exponential weighting scheme where data in the remote past are given less weight than the most recent measurements. This feature is especially useful in tracking problems where recent observations carry more information about the current value of the unknown parameter. Simulation results are included to demonstrate the performance of the recursive scheme.

1. INTRODUCTION

In recent work [1]-[3], we have introduced a new formulation for parameter estimation in the presence of bounded uncertainties in the data matrix (or regression vectors). Compared to earlier approaches in robust estimation, such as H_∞ and errors-in-variables methods (or Total Least Squares) [4]-[9], the new formulation incorporates an additional bound on the size of the mismatch between the nominal model and the accurate model. In so doing, the resulting solution is guaranteed to be robust with respect to uncertainties that lie within a specified domain, thus avoiding an overly conservative design.

A complete closed-form solution for the new estimation problem, in terms of the unique positive root of a nonlinear equation and the SVD of the nominal data matrix, has been obtained in [1]-[2]. A recursive procedure that is useful for on-line estimation contexts has been derived in [3] by establishing that some of the fundamental equations in [1] induce a contractive mapping. By invoking the Contraction Mapping Theorem [10], the reference [3] further showed that the unique fixed point of the mapping can be approximated to good accuracy via an iterative

scheme. In this way, an approximate recursive solution, similar in nature to RLS (recursive least-squares) [11,12], was derived that allows to update the solution of the new estimation problem without the need for explicit SVDs and for the solution of the nonlinear equation.

In this paper, we provide an overview of some of these results and then address some new issues by modifying the original formulation in order to incorporate exponential weighting of the data. This modification is better suited for tracking applications, where the unknown weight vector may vary with time and, correspondingly, the effect of earlier regression vectors on the current solution should be minimized.

2. PROBLEM FORMULATION

Let $x \in \mathbf{R}^n$ be a column vector of unknown parameters, $b \in \mathbf{R}^m$ a vector of measurements, and $A \in \mathbf{R}^{m \times n}$, $m \geq n$, a known full rank matrix. The matrix A represents nominal data in the sense that the true matrix that relates b to x is not A itself but rather a perturbed version of A , say $b = (A + \delta A)x + v$. The perturbation δA is not known. What is known is a bound on how far the true matrix $(A + \delta A)$ can be from the assumed nominal value A , say $\|\delta A\|_2 \leq \eta$ (in terms of the 2-induced norm of δA , or equivalently, its maximum singular value).

For example, x can denote the coefficients of an FIR filter, the rows of $A + \delta A$ can denote the successive state vectors, and the entries of b will then denote the resulting noisy measurements of the output of the filter. Due to several reasons (e.g., measurement errors, modeling errors), we do not have access to the true rows of $(A + \delta A)$ but only to the nominal data A . The perturbation in b is therefore due to both v and the mismatch between the nominal and true data.

We pose the problem of finding an estimate \hat{x} that performs "well" for any possible perturbation δA . That is, we would like to determine, if possible, an \hat{x} that meets the worst-case criterion

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

Any value that we pick for \hat{x} would lead to many residual norms, $\|(A + \delta A) \cdot \hat{x} - b\|_2$, one for each possible choice

* A. H. Sayed and V. Nascimento are with Dept. of EE, UCLA, CA 90095. A. Garulli is with Facolta' di Ingegneria, Universita' di Siena, Via Roma 56, 53100 Siena, Italy. S. Chandrasekaran is with Dept. of ECE, UCSB, CA 93106. The work of A. H. Sayed was in part supported by the National Science Foundation under Award No. MIP-9796147.

of A in the disc $(A + \delta A)$. 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 that is described in [1,2].

3. AN ALGEBRAIC SOLUTION

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

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

where the cost function $\mathcal{L}(\hat{x}) = \|A \cdot \hat{x} - b\|_2 + \eta \cdot \|\hat{x}\|_2$ 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). For this reason, the solution is more involved. The following theorem is a special case of the main result in [1].

Theorem 1. *Assume that b does not belong to the column space of A . Then the solution of the min-max estimation problem can be constructed as follows. Introduce the SVD of A : $A = U \cdot [\Sigma^T \ 0]^T \cdot V^T$, partition the vector $U^T b$ into $U^T \cdot b = [c^T \ d^T]^T$, where $c \in \mathbf{R}^n$ and $d \in \mathbf{R}^{m-n}$, and introduce the secular equation*

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

where

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

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

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

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

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

Remarks. The assumption that b does not belong to the column space of A is equivalent to $d \neq 0$. It also requires $m > n$ since A is assumed full rank. If b belongs to the column space of A , then 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).

Also, expressions (5) and (6) correspond to two equations in the two unknowns \hat{x} and $\hat{\alpha}$. By substituting (5) into (6), and using the SVD of A , we obtain the nonlinear equation (3) and (4).

Observe further from (5) that the expression for \hat{x} has the form of a regularized solution, except that the regularization parameter $\hat{\alpha}$ is not given a priori but has to be determined as the unique positive root of a secular equation. In this sense, we say that the solution involves automatic regularization.

4. EXPONENTIAL WEIGHTING

The rows of the data matrix A can be further weighted in order to give more or less significance to individual rows. Consider, for example, a tracking application where it is required to keep track of a weight vector x that might be changing with time. In this case, additional data measurements, in the form of additional rows to the data matrix A and additional entries to the data vector b , are continuously collected and used in updating the estimates for x . It is then natural to require that the top rows in A and the top entries in b , which correspond to data in the remote past, be given less weight than the most recent measurements since the latter measurements carry more information about the current value of the unknown x .

We use here one particular form of data weighting that has been used extensively in the adaptive filtering literature, viz., exponential weighting [11,12].

Let $0 \ll \lambda < 1$ be a number that is close to one and define $\Lambda_m = \text{diag}\{\lambda^{m-1}, \lambda^{m-2}, \dots, \lambda, 1\}$. Introduce also the weighted quantities

$$\bar{A} = \Lambda_m^{1/2} A , \quad \bar{b} = \Lambda_m^{1/2} b , \quad \delta \bar{A} = \Lambda_m^{1/2} \delta A .$$

Since $\|\Lambda_m^{1/2}\|_2 = 1$, it still holds that $\|\delta \bar{A}\|_2 \leq \eta$. We can then pose the weighted min-max problem

$$\min_{\hat{x}} \left(\max_{\|\delta \bar{A}\|_2 \leq \eta} \|(\bar{A} + \delta \bar{A}) \cdot \hat{x} - \bar{b}\|_2 \right) . \quad (7)$$

The solution is still given by Thm. 1 with A replaced by \bar{A} , the SVD of A replaced by the SVD of \bar{A} , and b replaced by \bar{b} .

5. EXACT TIME-UPDATES

We now apply the above results in the context of recursive estimation. So consider the linear regression model

$$y_t = (a_t + \delta a_t)^T x + v_t , \quad t = 1, 2, \dots \quad (8)$$

where $y_t \in \mathbf{R}$ is the output, $(a_t + \delta a_t) \in \mathbf{R}^n$ the regression vector, $x \in \mathbf{R}^n$ the unknown parameter vector, and $v_t \in \mathbf{R}$ 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, say $\|\delta a_t\|_2 \leq \epsilon$ for all t .

At each time instant t , we construct the quantities

$$A_t = \begin{bmatrix} a_1^T \\ \vdots \\ a_t^T \end{bmatrix}, \quad b_t = \begin{bmatrix} y_1 \\ \vdots \\ y_t \end{bmatrix}, \quad \delta A_t = \begin{bmatrix} \delta a_1^T \\ \vdots \\ \delta a_t^T \end{bmatrix}. \quad (9)$$

They contain the data up to time t . We then pose a min-max estimation problem of the form:

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

for some bound η_t on the size of δA_t . The corresponding solution is denoted by \hat{x}_t , and it is the best estimate that can be obtained in the above sense by using the data up to time t . Now, in view of the properties of matrix norms [5], we have $\|\delta A_t\|_2 \leq \epsilon \cdot \sqrt{t}$, which provides a possible choice for η_t in terms of the given value of ϵ . This choice might be inconvenient since it grows with time.

We instead pose an exponentially-weighted min-max problem, with

$$\bar{A}_t = \Lambda_t^{1/2} A_t, \quad \bar{b}_t = \Lambda_t^{1/2} b_t, \quad \delta \bar{A}_t = \Lambda_t^{1/2} \delta A_t.$$

In this case, we easily verify that

$$\|\delta \bar{A}_t\|_2 \leq \sqrt{\sum_{j=1}^t \lambda^{(t-j)} \epsilon^2} = \epsilon \cdot \sqrt{\frac{1-\lambda^t}{1-\lambda}} \leq \frac{\epsilon}{\sqrt{1-\lambda}}.$$

Hence, we can use $\eta_t = \frac{\epsilon}{\sqrt{1-\lambda}} \equiv \eta$, for all t , and the problem of interest becomes

$$\min_{x_t} \max_{\|\delta \bar{A}_t\|_2 \leq \frac{\epsilon}{\sqrt{1-\lambda}}} \|(\bar{A}_t + \delta \bar{A}_t)x_t - \bar{b}_t\|_2 \quad (11)$$

We shall also denote its solution by \hat{x}_t . We further assume that, for all t , b_t does not lie in the range space of A_t in order to avoid degenerate solutions.

We solve (11) for time instants $t > n$ so that the successive matrices \bar{A}_t have more rows than columns (as required by the statement of Thm. 1 and the remark after it). So let t_o denote our initial time instant (say, $t_o = n + 1$). Let $\{\hat{x}_t\}_{t=t_o}^N$ denote the successive solutions for $t = t_o, \dots, N$ of problem (11), viz.,

$$\hat{x}_t = (\bar{A}_t^T \bar{A}_t + \hat{\alpha}_t I)^{-1} \bar{A}_t^T \bar{b}_t, \quad (12)$$

where, in view of (6), $\hat{\alpha}_t$ is related to \hat{x}_t via

$$\hat{\alpha}_t = \eta \frac{\|\bar{A}_t \hat{x}_t - \bar{b}_t\|_2}{\|\hat{x}_t\|_2}. \quad (13)$$

Likewise, at time $(t + 1)$, the optimal solution is given by

$$\hat{x}_{t+1} = (\bar{A}_{t+1}^T \bar{A}_{t+1} + \hat{\alpha}_{t+1} I)^{-1} \bar{A}_{t+1}^T \bar{b}_{t+1}, \quad (14)$$

where $\hat{\alpha}_{t+1}$ satisfies a similar relation to (13) with time t replaced by $(t + 1)$. Using (12) and (14), we can find an expression that relates the optimal solution \hat{x}_{t+1} to the optimal solution \hat{x}_t as follows. Define

$$h_{t+1} = (\bar{A}_{t+1}^T \bar{A}_{t+1} + \hat{\alpha}_t I)^{-1} \bar{A}_{t+1}^T \bar{b}_{t+1}.$$

Comparing with the expression (14) for \hat{x}_{t+1} we see that h_{t+1} approximates \hat{x}_{t+1} by using $\hat{\alpha}_t$ instead of $\hat{\alpha}_{t+1}$. By further noting that

$$\begin{aligned} \bar{A}_{t+1}^T \bar{A}_{t+1} &= \lambda \bar{A}_t^T \bar{A}_t + a_{t+1} a_{t+1}^T, \\ \bar{A}_{t+1}^T \bar{b}_{t+1} &= \lambda \bar{A}_t^T \bar{b}_t + a_{t+1} y_{t+1}, \end{aligned}$$

we obtain the following result for $t \geq t_o$.

Theorem 2. For $t \geq t_o$. Let $\eta = \epsilon/\sqrt{1-\lambda}$. If $\eta \geq \frac{\|\bar{A}_t^T \bar{b}_t\|_2}{\|\bar{b}_t\|_2}$ then $\hat{x}_{t+1} = 0$. Otherwise,

$$\begin{aligned} h_{t+1} &= \hat{x}_t + \frac{\lambda^{-1} P_t a_{t+1}}{1 + \lambda^{-1} a_{t+1}^T P_t a_{t+1}} (y_{t+1} - a_{t+1}^T \hat{x}_t), \\ \hat{x}_{t+1} &= [I - (\hat{\alpha}_{t+1} - \lambda \hat{\alpha}_t) P_{t+1}] h_{t+1}, \\ P_{t+1}^{-1} &= \lambda P_t^{-1} + a_{t+1} a_{t+1}^T + (\hat{\alpha}_{t+1} - \lambda \hat{\alpha}_t) I, \end{aligned} \quad (15)$$

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

The recursive algorithm of Thm. 2 still requires the computation of the unique positive solution $\hat{\alpha}_t$ of the secular equation (3) at each time instant t (using the exponentially weighted data). The complexity of this task can be reduced if we replace the exact solution $\hat{\alpha}_t$ by an approximate solution, say $\bar{\alpha}_t$. In the next section, we first derive such an approximate scheme and then justify our reasoning by noting that it relies on a useful contractive property of the secular function f_t .

6. APPROXIMATE TIME UPDATES

Recall that the optimal value of $\hat{\alpha}_{t+1}$ is given by (cf. (13))

$$\hat{\alpha}_{t+1} = \eta \frac{\|\bar{A}_{t+1} \hat{x}_{t+1} - \bar{b}_{t+1}\|_2}{\|\hat{x}_{t+1}\|_2}, \quad (16)$$

where \hat{x}_{t+1} is given by (14). If we replace (14) into (16) we obtain the secular equation whose unique positive root is $\hat{\alpha}_{t+1}$. We would like instead to come up with an approximate procedure that avoids the need for explicitly working with the secular equation.

For this purpose, assume that we have solved the estimation problem exactly up to time t , and hence have available $\{\hat{x}_t, h_{t+1}, \hat{\alpha}_t\}$. We then replace \hat{x}_{t+1} in the expression (16) for $\hat{\alpha}_{t+1}$ by its approximation h_{t+1} and, in this way, define an approximate value for $\hat{\alpha}_{t+1}$. We shall denote this approximate value by $\bar{\alpha}_{t+1}$,

$$\bar{\alpha}_{t+1} = \eta \frac{\|\bar{A}_{t+1} h_{t+1} - \bar{b}_{t+1}\|_2}{\|h_{t+1}\|_2} \quad (17)$$

This step amounts to applying one iteration of the secular function f_{t+1} using as initial condition $\hat{\alpha}_t$,

$$\bar{\alpha}_{t+1} = f_{t+1}(\hat{\alpha}_t). \quad (18)$$

The resulting value is taken as an approximant for $\hat{\alpha}_{t+1}$. We shall show in Sec. 8. that, for any positive initial condition $\hat{\alpha}_t$, if we repeatedly apply f_{t+1} we can get arbitrarily close to the desired optimal value $\hat{\alpha}_{t+1}$. This fact provides an analytical justification for the approximation (18), except that here we are limiting ourselves to a single iteration for ease of implementation. More iterations would certainly lead to different procedures, with higher complexity and better accuracy. Simulations in [3] show the good accuracy of the single iteration procedure.

Once $\bar{\alpha}_{t+1}$ is obtained, we need to proceed to the next step and evaluate $\hat{\alpha}_{t+2}$. Now, however, we do not have the exact value for $\hat{\alpha}_{t+1}$, but only its approximation. For this reason, we can not compute the exact quantity h_{t+2} but only an approximant for it, say

$$\bar{h}_{t+2} = (\bar{A}_{t+2}^T \bar{A}_{t+2} + \bar{\alpha}_{t+1} I)^{-1} \bar{A}_{t+2}^T \bar{b}_{t+2}.$$

This in turn is used in an expression similar to (17) to produce $\bar{\alpha}_{t+2}$, and so on. In summary, our iterative procedure takes the form

$$\bar{\alpha}_{t+2} = \eta \frac{\|A_{t+2} \bar{h}_{t+2} - \bar{b}_{t+2}\|_2}{\|\bar{h}_{t+2}\|_2}, \quad (19)$$

which corresponds to $\bar{\alpha}_{t+2} = f_{t+2}(\bar{\alpha}_{t+1})$. Let $\bar{x}_t = (\bar{A}_t^T \bar{A}_t + \bar{\alpha}_t I)^{-1} \bar{A}_t^T \bar{b}_t$ denote the resulting approximation for \hat{x}_t at time t . [Note that we are using long overbars for approximate quantities, and short overbars for exponentially weighted quantities]. The following recursions for the approximate quantities hold.

Iterative Algorithm. Set $\bar{x}_{t_0} = \hat{x}_{t_0}$ and $\bar{\alpha}_{t_0} = \hat{\alpha}_{t_0}$, and let $\bar{P}_{t_0} = (\bar{A}_{t_0}^T \bar{A}_{t_0} + \bar{\alpha}_{t_0} I)^{-1}$, $z_{t_0}^2 = \|\bar{b}_{t_0}\|_2^2$, and $\eta = \epsilon/\sqrt{1-\lambda}$. That is, at time t_0 we have an exact weighted min-max solution. Now, for $t = t_0, \dots, N$ we repeat:

$$\begin{aligned} \bar{h}_{t+1} &= \left\{ \bar{x}_t + \frac{\lambda^{-1} \bar{P}_t a_{t+1}}{1 + \lambda^{-1} a_{t+1}^T \bar{P}_t a_{t+1}} (y_{t+1} - a_{t+1}^T \bar{x}_t) \right\} \\ z_{t+1}^2 &= \lambda z_t^2 + y_{t+1}^2 \\ p_t &= \bar{h}_{t+1}^T (\lambda \bar{P}_t^{-1} + a_{t+1} a_{t+1}^T - \lambda \bar{\alpha}_t I) \bar{h}_{t+1} \\ q_t &= 2 \bar{h}_{t+1}^T (\lambda \bar{P}_t^{-1} \bar{x}_t + a_{t+1} y_{t+1}) \\ \bar{\alpha}_{t+1} &= \frac{\eta}{\|\bar{h}_{t+1}\|_2} [z_{t+1}^2 + p_t - q_t]^{1/2} \\ \bar{P}_{t+1}^{-1} &= \lambda \bar{P}_t^{-1} + a_{t+1} a_{t+1}^T + [\bar{\alpha}_{t+1} - \lambda \bar{\alpha}_t] I \\ \bar{x}_{t+1} &= [I - (\bar{\alpha}_{t+1} - \lambda \bar{\alpha}_t) \bar{P}_{t+1}] \bar{h}_{t+1} \end{aligned} \quad (20)$$

7. A FAST ITERATIVE ALGORITHM

The recursions can be interpreted as follows. The algorithm computes an \bar{h}_{t+1} first. Its expression is very similar to the update expression of an exponentially-weighted recursive least-squares (RLS) algorithm that updates an estimate \bar{x}_t to a new estimate \bar{x}_{t+1} according to the rule:

$$\bar{x}_{t+1} = \left\{ \bar{x}_t + \frac{\lambda^{-1} \bar{P}_t a_{t+1}}{1 + \lambda^{-1} a_{t+1}^T \bar{P}_t a_{t+1}} (y_{t+1} - a_{t+1}^T \bar{x}_t) \right\}.$$

The corresponding \bar{P}_{t+1}^{-1} in RLS would be obtained via a rank-one update of the form $\bar{P}_{t+1}^{-1} = \lambda \bar{P}_t^{-1} + a_{t+1} a_{t+1}^T$. The iterative min-max algorithm, on the other hand, updates \bar{x}_t into \bar{h}_{t+1} first and then uses \bar{h}_{t+1} to compute the new estimate \bar{x}_{t+1} . The value of \bar{h}_{t+1} is further used to update $\bar{\alpha}_t$ to $\bar{\alpha}_{t+1}$. Moreover, the new matrix \bar{P}_{t+1}^{-1} is obtained from the older matrix not just through a rank-one update but also through an additional scalar multiple of the identity matrix.

Since the min-max iterative scheme requires the inversion of the $(n \times n)$ -matrix \bar{P}_t^{-1} at every step, we see that the computational complexity of a single iteration is $O(n^3)$ as it stands. This is an order of magnitude higher than the traditional recursive least-squares (RLS) algorithm. The lower cost in RLS is obtained by propagating \bar{P}_t rather than its inverse through a (simplified) Riccati recursion. This is possible in RLS since \bar{P}_{t+1}^{-1} is obtained only through a rank-one update of $\lambda \bar{P}_t^{-1}$.

For the iterative min-max algorithm, however, the fact that the computation of \bar{P}_{t+1}^{-1} also involves a multiple of the identity matrix does not allow for an immediate fast recursion for the explicit update of \bar{P}_{t+1} .

A way out of this difficulty is to employ a numerically stable $O(n^2)$ algorithm developed in [13] for updating the SVD of rank-one matrix updates. This would allow us to reduce the computational cost to $O(n^2)$. It can be achieved as follows.

Let $U_t \Lambda_t U_t^T$ denote the eigendecomposition of $\lambda \bar{P}_t^{-1}$. Let also $V_t \Gamma_t V_t^T$ denote the eigendecomposition of the rank-one update $\lambda \bar{P}_t^{-1} + a_{t+1} a_{t+1}^T$. The algorithm developed in [13] allows us to update U_t to V_t and Λ_t to Γ_t in $O(n^2)$:

$$U_t \longrightarrow V_t, \quad \Lambda_t \longrightarrow \Gamma_t \text{ in } O(n^2) \text{ operations.}$$

Now recognizing that

$$\lambda \bar{P}_t^{-1} + a_{t+1} a_{t+1}^T + (\bar{\alpha}_{t+1} - \lambda \bar{\alpha}_t) I = V_t [\Gamma_t + (\bar{\alpha}_{t+1} - \lambda \bar{\alpha}_t) I] V_t^T$$

we see that we can make the identifications:

$$U_{t+1} = V_t, \quad \Lambda_{t+1} = \lambda \Gamma_t + \lambda (\bar{\alpha}_{t+1} - \lambda \bar{\alpha}_t) I.$$

This allows us to update $\{U_t, \Lambda_t\}$ to $\{U_{t+1}, \Lambda_{t+1}\}$ in $O(n^2)$. The $\{U_t, \Lambda_t\}$ completely specify $\{\bar{P}_t, \bar{P}_t^{-1}\}$ and the algorithm can be completed in this way.

8. CONTRACTION MAPPING

We now return to our earlier remark after (18) that if the function f_{t+1} is applied repeatedly to an arbitrary positive initial condition, then the result gets arbitrarily close to its unique positive root. This fact was used to justify the approximation (18).

Indeed, returning to Thm. 1, we now argue that a good approximation for $\hat{\alpha}$ can be obtained by alternatively iterating the map defined by f [3].

Define the recursive equation

$$\alpha^{(i+1)} = f(\alpha^{(i)}), \quad \alpha^{(0)} = \text{positive initial condition.} \quad (21)$$

The following result can be established by invoking the Contraction Mapping Theorem [10].

Theorem 3. *Consider the setting of Thm. 1. Assume $\eta < \tau$. For any positive initial value $\alpha^{(0)}$, it holds that $\lim_{i \rightarrow \infty} \alpha^{(i)} = \hat{\alpha}$, where $\hat{\alpha}$ is the unique positive solution of the secular equation (3).*

A proof is given in [3]. Here we demonstrate the validity of the theorem in the scalar-case, i.e., when x is one-dimensional (and therefore a scalar). So let

$$A = a \in \mathbf{R}^{m \times 1}, \quad \Sigma = \sqrt{a^T a} \in \mathbf{R},$$

$$A^T b = a^T b = \sqrt{a^T a} \cdot c, \quad c \in \mathbf{R}, \quad d \in \mathbf{R}^{m-1}.$$

Using the identity $\|b\|_2^2 = d^T d + c^2$, we can verify that expression (4) for $f(\alpha)$ reduces to

$$f(\alpha) = \beta \cdot \sqrt{\alpha^2 + 2\alpha\kappa + (a^T a)\kappa},$$

where we have defined

$$\beta = \frac{\eta \sqrt{b^T b}}{|a^T b|}, \quad \kappa = \frac{(a^T a)(d^T d)}{b^T b}.$$

It is clear that $\beta < 1$ since $\eta < \tau$, and $0 \leq \kappa \leq (a^T a)$ since $(d^T d)/(b^T b) \leq 1$. The unique positive root $\hat{\alpha}$ satisfies $\hat{\alpha} = f(\hat{\alpha})$. It then follows from (21), and from the mean-value theorem, that

$$\alpha^{(i+1)} - \hat{\alpha} = f(\alpha^{(i)}) - f(\hat{\alpha}) = f'(\bar{\alpha}) \cdot (\alpha^{(i)} - \hat{\alpha}),$$

for some point $\bar{\alpha}$ between $\alpha^{(i)}$ and $\hat{\alpha}$. Here, f' denotes the derivative function. We claim that $0 < f'(\alpha) \leq \beta < 1$ for all $\alpha > 0$. Indeed, the expressions for the first- and second-order derivatives of f are given by

$$f'(\alpha) = \frac{\beta(\alpha + \kappa)}{\sqrt{\alpha^2 + 2\alpha\kappa + (a^T a)\kappa}},$$

and

$$f''(\alpha) = \frac{\beta[(a^T a)\kappa - \kappa^2]}{[\alpha^2 + 2\alpha\kappa + (a^T a)\kappa]^{3/2}} \geq 0.$$

This shows that $f'(\alpha)$ is a non-decreasing function of α . However,

$$f'(0) = \frac{\beta\sqrt{\kappa}}{\sqrt{a^T a}} \leq \beta \quad \text{and} \quad \lim_{\alpha \rightarrow \infty} f'(\alpha) = \beta.$$

We therefore conclude that $0 < f'(\alpha) \leq \beta < 1$ for all $\alpha > 0$ and

$$|\alpha^{(i+1)} - \hat{\alpha}| \leq \beta \cdot |\alpha^{(i)} - \hat{\alpha}|, \quad \beta < 1.$$

This relation establishes the convergence of $\alpha^{(i)}$ to $\hat{\alpha}$ from any positive initial condition $\alpha^{(0)}$. In the vector case, the argument is a bit more involved (e.g., [3]).

The result of Thm. 3 suggests that recursion (21) 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 (21), we can approximate \hat{x} in (5) with $x^{(p)} = (A^T A + \alpha^{(p)} I)^{-1} A^T b$. We have used an approximation that is based on a single iteration in (18). Several simulations on randomly generated data have shown that in general good approximations can be obtained with very few iterations [3]. This is particularly useful in recursive estimation contexts, as we explained in an earlier section.

9. SOME SIMULATIONS

In [3] we compared the performance of the iterative algorithm with the exact solution of Thm. 2. Here we provide alternative simulations that compare the performance of the algorithm with TLS and RLS.

To begin with, consider an FIR filter that is perturbed by an additive noise term v_t ,

$$y_t = x_1 u_t + x_2 u_{t-1} + v_t,$$

where $x = [x_1 \ x_2]$ is the vector of unknown parameters that we want to estimate. Assume that at time t , noisy measurements are available for the $\{u_t\}$, say $\{u_t + \delta u_t\}$, where v_t and δu_t are taken as i.i.d. noise sequences satisfying $|v_t| < v_{\max}$, $|\delta u_t| < w_{\max}$, respectively.

The recursive min-max estimation algorithm can be applied by setting

$$A_t = \begin{bmatrix} u_1 + \delta u_1 & u_0 + \delta u_0 \\ u_2 + \delta u_2 & u_1 + \delta u_1 \\ \vdots & \vdots \\ u_t + \delta u_t & u_{t-1} + \delta u_{t-1} \end{bmatrix}, \quad b_t = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_t \end{bmatrix},$$

and by taking $\epsilon = \sqrt{2} w_{\max}$ and $\eta = \epsilon \sqrt{t}$.

In Fig. 1(a) and (b), the estimates provided by the recursive min-max algorithm (solid lines) have been compared to those obtained using the well-known Total Least Squares algorithm (dashed lines). In the simulations, a square wave of amplitude 1 has been chosen as the input

u_t , and the noise sequences have been randomly generated with $w_{\max} = 0.1$ and $v_{\max} = 1$. The reported results are averaged over 10 different noise realizations. It can be noticed that the min-max estimates outperform the corresponding TLS estimates in this case. One reason for this behaviour can be found in Fig. 1(c), where the worst-case min-max residual (solid line) has been compared to the corresponding TLS square error (dashed line) $\|(A_t + \delta \hat{A}_t) \hat{x}_{t|S} - b_t\|_2$, where $\delta \hat{A}_t$ is the TLS estimated correction for the matrix A_t . This correction results overly conservative, especially when the measurement noise v_t is much greater than the input noise w_t .

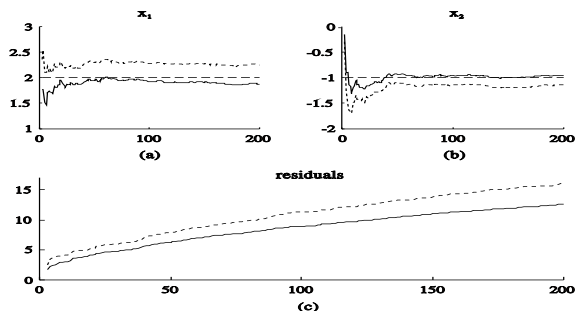


Figure 1: (a),(b) Min-max (solid lines) and total least squares (dashed lines) estimates of the FIR parameters. (c) Comparison between the residuals.

In another set of simulations, we have compared the performance of the min-max recursive solution with the exponentially weighted RLS solution. Figs. 2-4 show some typical results. They plot the norm of the error vector $x - \bar{x}_t$ as a function of time, averaged over 40 trials. The numerical values used were $n = 5$ and $\lambda = 0.95$.

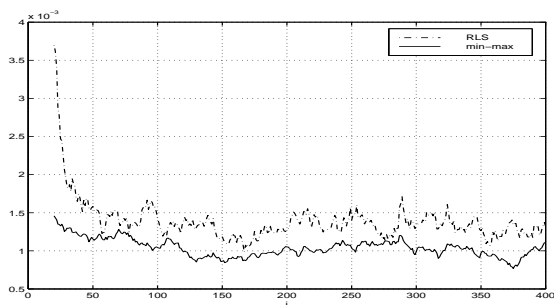


Figure 2: Dashed line is the RLS weight error norm, while the solid line is the min-max weight error norm. This simulation used $\epsilon = 0.25$, $\|A\|_2 = 20.7$, $\eta = 1.34$, and $x = [-0.0134, -0.0316, 0.0265, -0.0114, 0.0132]^T$.

References

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

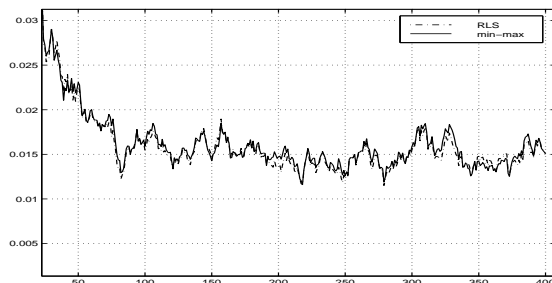


Figure 3: Dashed line is the RLS weight error norm, while the solid line is the min-max weight error norm. This simulation used $\epsilon = 0.25$, $\|A\|_2 = 64$, $\eta = 1.1$, and $x = [0.5924, -0.6835, -0.2023, -0.0924, 0.3010]^T$.

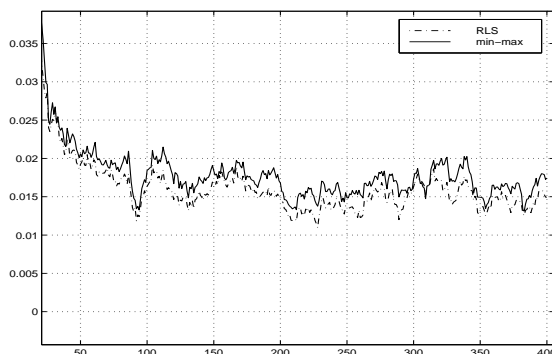


Figure 4: Dashed line is the RLS weight error norm, while the solid line is the min-max weight error norm. This simulation used $\epsilon = 0.4$, $\|A\|_2 = 68$, $\eta = 1.79$, and $x = [0.4202, 0.0954, 0.5985, -0.1797, 0.6546]^T$.

SIAM J. Matrix Anal. Appl., to appear, 1998. Also in Proc. American Control Conference, June 1997.

[2] S. Chandrasekaran, G. H. Golub, M. Gu, and A. H. Sayed, *Best-fit parameter estimation for a bounded errors-in-variables model*, in Proc. American Control Conference, June 1997.

[3] A. H. Sayed, A. Garulli, and S. Chandrasekaran, *A fast iterative solution for worst-case parameter estimation with bounded model uncertainties*, in Proc. American Control Conference, June 1997.

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

[5] G. H. Golub and C. F. Van Loan, *Matrix Computations*, The Johns Hopkins University Press, Baltimore, third ed., 1997.

[6] B. Hassibi, A. H. Sayed, and T. Kailath, *Linear estimation in Krein spaces - Part I: Theory*, IEEE Trans. Automatic Control, 41 (1996), pp. 18-33.

[7] P. Khargonekar and K. M. Nagpal, *Filtering and smoothing in an H^∞ -setting*, IEEE Trans. on Automatic Control, AC-36 (1991), pp. 151-166.

[8] U. Shaked and Y. Theodor, *H^∞ -optimal estimation: A tutorial*, in Proc. IEEE Conference on Decision and Control, Tucson, AZ, Dec. 1992, pp. 2278-2286.

- [9] M. Green and D. J. N. Limebeer, *Linear Robust Control*, Prentice Hall, NJ, 1995.
- [10] D. G. Luenberger, *Optimization by Vector Space Methods*, Wiley, NY, 1969.
- [11] S. Haykin, *Adaptive Filter Theory*, Prentice Hall, 3rd edition, NJ 1996.
- [12] A. H. Sayed and T. Kailath, *A state-space approach to RLS adaptive filtering*, IEEE Signal Processing Magazine, vol. 11, no. 3, pp. 18–60, July 1994.
- [13] M. Gu, *Studies in Numerical Linear Algebra*, Ph.D. Dissertation, Computer Science Dept., Yale University, 1993.