Proceedings of the American Control Conference
Albuquerque, New Mexico June 1997
0-7803-3832-4/97/$10.00 © 1997 AACC

# Unfalsified Weighted Least Squares Estimates in Set-Membership Identification

Er-Wei Bai[+*], Li Qiu[*] and Roberto Tempo[!]
[+]Dept. of ECE, Univ. of Iowa, Iowa City, IA, 52242, U.S.A.
erwei@icaen.uiowa.edu
[*]Dept of EEE, HKUST, Clear Water Bay, Kowloon, Hong Kong
eeqiu@ee.ust.hk
[!]CENS-CNR, Politecnico di Torino, Italy
tempo@polito.it

## Abstract

It is well-known that the Weighted Least Squares (WLS) identification algorithm provides estimates that are in general not in the membership set and in this sense are falsified estimates. This paper shows that: (1) If the noise bound is known, the WLS estimates can be made to lie in or converge to the membership set by choosing the weights properly. (2) If the noise bound is unknown, the same results can still be achieved by using white input signals for Finite Impulse Response systems.

## 1. Problem Statement

In this paper, we consider a discrete time scalar system

$$y_i = \phi_i^T \theta + v_i, \quad i = 1, 2, \ldots, N \tag{1.1}$$

where $y_i \in \mathbf{R}$ is the system output, $\phi_i \in \mathbf{R}^n$ the measurable regressor consisting of current and past input signals and (possibly) past output signals, $\theta \in \mathbf{R}^n$ the unknown parameter vector to be identified and $v_i \in R$ the measurement noise. It is assumed that $\phi_i$'s are bounded so that there exists a constant $M > 0$, independent of $N$ and

$$\|\phi_i\|^2 \le M \tag{1.2}$$

for all $i$. The equation (1.1) can be re-written in a compact vector form as

$$Y_N = \Phi_N \theta + V_N \tag{1.3}$$

where

$$Y_N = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{pmatrix}, \ \Phi_N = \begin{pmatrix} \phi_1^T \\ \phi_2^T \\ \vdots \\ \phi_N^T \end{pmatrix}, \ V_N = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_N \end{pmatrix}.$$

The purpose of system identification is to design an algorithm $\mathcal{A}$ which maps the input-output measurements $y_i$ and $\phi_i$ into the estimate $\widehat{\theta}$ of the unknown system parameter vector $\theta$. Depending on the specific assumptions on the noise, many identification algorithms can be constructed. For instance, in the stochastic setting, the noise $v_i$ is assumed to be a random sequence with some known probabilistic properties and Maximum Likelihood Estimators can be derived [17]. In set-membership identification, see e.g. the special issues [1], [2], [3] and the survey papers [20], [21], [22], [26], the noise is assumed to be unknown but bounded by $\epsilon$, i.e.,

$$|v_i| \le \epsilon \tag{1.4}$$

for all $i$. In this case, for the presence of noise, it is in general not possible to determine whether the obtained estimate $\widehat{\theta}$ coincides with the true but unknown $\theta$ but we can only detect whether $\widehat{\theta}$ is compatible with the observed input-output data. To this end, the membership set is defined as follows:

$$S_{i_0,i_1}(\epsilon) = \bigcap_{i=i_0}^{i_1} \{\widehat{\theta} \in R^n \ : \ |y_i - \phi_i^T \widehat{\theta}| \le \epsilon\}. \tag{1.5}$$

An estimate $\widehat{\theta}$ is compatible with the input-output data from the $i_0$-th observation to the $i_1$-th observation if and only if $\widehat{\theta} \in S_{i_0,i_1}(\epsilon)$.

Besides systems and control, set-membership identification proves to be a valuable tool in other areas, including digital signal processing, when a noise-bounded description of the errors is suitable; see e.g., the survey paper [10]. In this case, one example of paramount importance is when the measurements are affected by roundoff errors given by A/D converters [24]. More classical identification algorithms than set-membership identification include the celebrated Least Squares (LS), more generally, the Weighted Least Squares (WLS) algorithm. For given data $Y_N$ and $\Phi_N$, the WLS estimate $\widehat{\theta}_N$ is the solution of the minimization problem

$$\widehat{\theta}_N = \arg\min_{\widehat{\theta}} \sum_{i=1}^{N} q_i (y_i - \phi_i^T \widehat{\theta})^2$$

where $q_i$'s are non-negative weights. Letting

$$Q_N = \begin{pmatrix} q_1 & 0 & \dots & 0 \\ 0 & q_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & q_N \end{pmatrix},$$

a closed form solution of the WLS estimate $\widehat{\theta}_N$ can be easily derived [17] if $\Phi_N^T Q_N \Phi_N$ is non-singular,

$$\widehat{\theta}_N = (\Phi_N^T Q_N \Phi_N)^{-1} \Phi_N^T Q_N Y_N. \tag{1.6}$$

One of the powerful features of the WLS algorithm is that it can be implemented recursively [17] using only the current input-output measurement $y_i$, $\phi_i$ and the previous estimate $\widehat{\theta}_{i-1}$

$$\widehat{\theta}_i = \widehat{\theta}_{i-1} + \frac{q_i P_{i-1}\phi_i}{1 + q_i \phi_i^T P_{i-1}\phi_i}(y_i - \phi_i^T \widehat{\theta}_{i-1}) \tag{1.7}$$

where the matrix $P_i \in \mathbf{R}^{n\times n}$ is also computed recursively

$$P_i = P_{i-1} - \frac{q_i P_{i-1}\phi_i \phi_i^T P_{i-1}}{1 + q_i \phi_i^T P_{i-1}\phi_i}. \tag{1.8}$$

with some $P_0 > 0$.

The WLS algorithm does not need any *a priori* assumption on the noise $v_i$ and enjoys several worst-case optimality identification properties; see, [4] and [18]. It is also well-known that the WLS estimates are in general *not* in the membership set (1.5). In other words, under the assumption that the noise $|v_i| \leq \epsilon$ for some known $\epsilon$, the WLS estimates may be incompatible with the observed input-output data. In model validation terminology, we can say that the WLS estimates may be falsified by the input-output data. This observation leads us to the following question: For the noise bound (1.4), can we choose the weights $q_i \geq 0$ properly so that the WLS estimate $\widehat{\theta}_i$ either lies within the membership set for all $i$ or converges to the membership set asymptotically? The motivation of studying this problem is obvious: If such choice of $q_i$'s is possible, then the resulting WLS estimate enjoys the stochastic identification properties of the original WLS estimate and it is also an unfalsified estimate compatible with the observed input-output measurements.

The idea of finding a compatible estimate is not new; for example, in Information-Based Complexity [25], such estimates are called interpolatory algorithms. In the context of system identification and model validation, several interpolatory algorithms have been proposed, see, e.g., [7] and [8]. However, due to the complex nature of the problem, all these algorithms are off-line type. Continuing our previous work [5], the main contribution of this work is to find *recursive* interpolatory WLS algorithms. That is, the algorithms presented in this paper choose weights $q_i$'s on line so that the resulting recursive WLS estimates either lie within the membership set or converge to it asymptotically. Clearly, the proposed algorithms are different than the ellipsoid-outer-bounding ones since they are Least Squares type, but the weights are chosen to minimize the "volume" of the outer-bounding ellipsoid. Therefore, not every point inside the outer-bounding set belongs to the actual membership set. Moreover, it is well-known that the recursive implementation of outer bounding algorithms may introduce some conservatism [26]. The implication is that there is no guarantee that a point inside the outer-bounding set is also in the membership set.

The results of this paper can be summarized as follows: If the system is Finite Impulse Response (FIR) and the input is at designer's disposal, in Section 2 we show: (a) If the input is chosen to be periodic, the WLS estimates can be made to lie within the membership set by a proper choice of $q_i$'s, provided that the bound on the noise is known; (b) If the input is chosen to be an independent identically distributed (i.i.d.) random sequence with zero mean, the WLS estimate converges to the true but unknow parameter $\theta$ asymptotically almost surely (a.s.) for any bounded noise sequence $v_i$ with *unknown noise bound* $\epsilon$.

Consequently, the WLS estimate $\widehat{\theta}_i$ converges to the membership set almost surely. If the system is Infinite Impulse Response (IIR) (see, e.g., [15] for definitions of FIR and IIR systems) and the noise bound is known, in Section 3 we show that the WLS estimates converge to the membership set for arbitrary input if $q_i$'s are suitably chosen. The proofs are provided in Section 4 and some concluding remarks are outlined in Section 5.

## 2. Finite Impulse Response Systems

In this section, we consider the FIR system

$$y_i = \phi_i^T \theta + v_i = (u_{i-1}, \dots, u_{i-n})\theta + v_i, \quad i = 1, \dots, N. \tag{2.9}$$

Before presenting the results, we need to define persistent excitation (PE), see [5] and [9].

**Definition 2.1:** The regressor $\phi_i$ is said to be persistently exciting (PE) if there exist some $\alpha > 0$ and some positive integer $p$ such that

$$\alpha I \leq \sum_{i=i_0}^{i_0+p-1} \phi_i \phi_i^T$$

for all $i_0 \geq 0$.

**Theorem 2.1** *Consider the FIR system (2.9) with the noise $v_i$ bounded as in (1.4) by some known $\epsilon > 0$. Assume that the input $u_i$ is periodic with period $n$ and is persistently exciting. Consider the recursive WLS algorithm (1.7) and (1.8) with the weights*

$$q_i = \begin{cases} \frac{|y_i - \phi_i^T \widehat{\theta}_{i-1}| - \epsilon}{\epsilon \phi_i^T P_{i-1}\phi_i} & \text{if } |y_i - \phi_i^T \widehat{\theta}_{i-1}| > \epsilon; \\ 0 & \text{if } |y_i - \phi_i^T \widehat{\theta}_{i-1}| \leq \epsilon \end{cases} \tag{2.10}$$

*for $i \geq 1$, the initial conditions $P_0 = q_0^{-1}\Phi_n\Phi_n^T$ (note $\Phi_n$ is defined in (1.3) with N replaced by n), any arbitrary $\widehat{\theta}_0$ and any positive constant $q_0 > 0$. Then, the WLS estimate $\widehat{\theta}_i$ lies in the membership set for all i, i.e.,*

$$\widehat{\theta}_i \in S_{1,i}(\epsilon) = \bigcap_{m=1}^{i} \{\widehat{\theta} \in R^n \ : \ |y_m - \phi_m^T \widehat{\theta}| \leq \epsilon\}.$$

Next, we observe the following two facts:

1. $P_0 > 0$. To show this, notice that $\phi_i$ is periodic with period $n$. Let $kn > p$ for some $k$,

$$\alpha I \leq \sum_{i=i_0}^{i_0+p-1} \phi_i\phi_i^T \leq k \sum_{i=i_0}^{i_0+n-1} \phi_i\phi_i^T = k\Phi_n^T\Phi_n.$$

   The matrix $\Phi_n$ is non-singular and this implies $P_0 > 0$.

2. The weights $q_i$'s in (2.10) are well defined because $\epsilon\phi_i^T P_{i-1}\phi_i \neq 0$ if $|y_i - \phi_i^T\widehat{\theta}_{i-1}| > \epsilon$. This can be easily seen as follows: Since $P_i^{-1} = P_{i-1}^{-1} + q_i\phi_i\phi_i^T$ (see [11], page 58), $P_0^{-1} > 0$ and $q_i \geq 0$, we have $P_i > 0$ and that $\phi_i^T P_{i-1}\phi_i = 0$ implies $\phi_i = 0$. However, $\phi_i = 0$ implies $|y_i - \phi_i^T\widehat{\theta}_{i-1}| = |v_i| < \epsilon$ which is a contradiction.

The theorem above shows that if the bound $\epsilon$ on the unknown noise $v_i$ is available, then the WLS estimate $\widehat{\theta}_i$ can be made to lie within the membership set by choosing a periodic input and a proper weighting sequence $q_i \geq 0$. Here, the availability of the noise bound $\epsilon$ is the key. The following result shows that even when the noise bound $\epsilon$ is unknown, the WLS estimates can still be made to converge to the membership set asymptotically.

**Theorem 2.2** *Consider the FIR systems (2.9) with the noise $v_i$ bounded as in (1.4) by some unknown bound $\epsilon > 0$. Let the input sequence $\{u_i\}$ be an i.i.d. random sequence with zero mean and finite variance. Consider the WLS algorithm (1.6) with the weights $q_i$'s lower and upper bounded*

$$0 < \underline{q} \leq q_i \leq \overline{q} < \infty$$

*for all i. Then, the WLS estimate $\widehat{\theta}_i$ satisfies*

$$||\widehat{\theta}_i - \theta|| \rightarrow 0 \quad a.s.$$

*as $i \rightarrow \infty$ and, consequently, the WLS estimate $\widehat{\theta}_i$ converges to the membership set a.s. as $i \rightarrow \infty$.*

Theorem 2.2 shows that the effect of any bounded noise sequence with *known or unknown* bound can be averaged out asymptotically by an i.i.d. input sequence with zero mean. The important thing of this result is that the bound $\epsilon$ on $v_i$ may be unknown and it is indeed *not required* in the WLS algorithm. A similar result is reported [16] if the input sequence is deterministic and the noise is i.i.d. with zero mean.

## 3. Infinite Impulse Response Systems with Arbitrary Input

In the previous section, we studied FIR systems assuming that the inputs were at designer's disposal. In this section, we relax this assumption and study general IIR systems with arbitrary inputs. In this case, the result shown in Theorem 2.1 that the estimate $\widehat{\theta}_i$ always lies in the membership set $S_{1,i}(\epsilon)$ does not hold in general. Therefore, we present an algorithm that, for any given small positive number $\delta > 0$, provides an estimate that lies in or converges to the set $S_{N_0,\infty}(\epsilon + \delta)$ for some $N_0$. The hope is that for small $\delta$, the set $S_{N_0,\infty}(\epsilon+\delta)$ is "very close" to $S_{N_0,\infty}(\epsilon)$. This is certainly true for any fixed $N_0$ and $\delta \rightarrow 0$. To see this observe that the membership set has the inclusive property $S_{N_0,\infty}(\epsilon) \subseteq S_{N_0,\infty}(\epsilon + \delta)$ for all $\delta \geq 0$. Now, for the sake of contradiction, suppose that the membership set $S_{N_0,\infty}(\epsilon+\delta)$ is not a continuous function of $\delta$ as $\delta \rightarrow 0$. Then, by the inclusive property of the membership set, there exists some $\bar{\theta} \notin S_{N_0,\infty}(\epsilon)$ but $\bar{\theta} \in S_{N_0,\infty}(\epsilon + \delta)$ for any $\delta > 0$, i.e., for all $i \geq N_0$, $|\phi_i^T\bar{\theta} - y_i| \leq \epsilon + \delta$, for all $\delta > 0$. Then, it follows that $|\phi_i^T\bar{\theta} - y_i| \leq \epsilon$ and this would imply $\bar{\theta} \in S_{N_0,\infty}(\epsilon)$ which is a contradiction. Therefore, the two sets are "almost identical" for small $\delta$.

**Theorem 3.1** *Consider the system (1.1) with the noise $v_i$ bounded as in (1.4) by some known $\epsilon > 0$. Consider the recursive WLS algorithm (1.7) and (1.8) with $P_0 = P_0^T > 0$ and arbitrary $\widehat{\theta}_0$. For any $\delta > 0$, let $q_i$ be*

$$q_i = \begin{cases} \frac{|y_i - \phi_i^T\widehat{\theta}_{i-1}| - \epsilon}{\epsilon\phi_i^T P_{i-1}\phi_i} & \text{if } |y_i - \phi_i^T\widehat{\theta}_{i-1}| > \epsilon + \delta; \\ 0 & \text{if } |y_i - \phi_i^T\widehat{\theta}_{i-1}| \leq \epsilon + \delta \end{cases} \tag{3.1}$$

*for $i \geq 1$. Then, the WLS estimate $\widehat{\theta}_i$ converges to the membership set asymptotically in the following sense: For any $\delta > 0$, there exists a finite number $N_0 = N_0(\delta)$ such that for all $i \geq N_0$*

$$\widehat{\theta}_i \in S_{N_0,\infty}(\epsilon+\delta) = \bigcap_{m=N_0}^{\infty} \{\widehat{\theta} \in R^n \ : \ |y_m - \phi_m^T\widehat{\theta}| \leq \epsilon + \delta\}.$$

The above result is a continuation of our previous work on gradient type identification algorithms [5] which was motivated by the papers [6] and [13]. Even though in this work we have restricted our attention to WLS algorithms, remarks similar to those made in [5] apply as well.

**Remark 1:** The asymptotic estimate of $\widehat{\theta}_i$ given by the WLS algorithm is not necessarily in the membership set $S_{1,\infty}(\epsilon + \delta)$. Instead, it is only guaranteed to be in the membership set $S_{N_0,\infty}(\epsilon + \delta)$ where $N_0$, the learning period, is the instance of final update of $\theta$. The "learning period" $N_0$ of the above algorithm, after which no parameter update takes place, depends on the data $(\{\phi_i\}, \{y_i\})$ and the slack variable $\delta > 0$. For the above algorithm, without additional information, it is not possible to know on-line whether the estimate has converged. The slack variable $\delta > 0$ represents the tradeoff between the learning period $N_0$ and the estimation accuracy. Since the parameter estimate converges to the set $S_{N_0,\infty}(\epsilon+\delta)$, the final estimate would be more accurate if a smaller $\delta$ is chosen. However, in this case, the learning period $N_0$ would be larger.

The algorithm presented in Theorem 3.1 has a clear geometric interpretation. From equation (1.7), we notice that the parameter update at $i$-th iteration of the (weighted) least squares method is always along the direction of the vector $P_i\phi_i$ with $|\phi_i^T\hat{\theta}-y_i| = \epsilon$ representing two planes in the parameter space that bound all $\hat{\theta}$ compatible with the new data $\phi_i$ and $y_i$. The true parameter vector $\theta$ is unknown but always lies between these two planes. The algorithm presented in Theorem 3.1 updates the parameter $\hat{\theta}_{i-1}$ in the direction of $P_i\phi_i$ indicated in the figure. However, if $|\phi_i^T\hat{\theta}_{i-1}-y_i| \le \epsilon+\delta$, then $\hat{\theta}_{i-1}$ is compatible with the new data $\phi_i$ and $y_i$ and there is no need to update $\hat{\theta}_{i-1}$. In this case, $q_i = 0$ and $\hat{\theta}_i = \hat{\theta}_{i-1}$. Now, if $|\phi_i^T\hat{\theta}_{i-1}-y_i| > \epsilon+\delta$, then $\hat{\theta}_{i-1}$ is not compatible with the new data $\phi_i$ and $y_i$. In this case, we choose $q_i = \frac{|y_i-\phi_i^T\widehat{\theta}_{i-1}|-\epsilon}{\epsilon\phi_i^TP_{i-1}\phi_i}$ such that

$$\begin{aligned}|\phi_i^T\hat{\theta}_i - y_i| &= |\phi_i^T\hat{\theta}_{i-1}-y_i|\cdot\left|1-\frac{q_i\phi_i^TP_{i-1}\phi_i}{1+q_i\phi_i^TP_{i-1}\phi_i}\right| \\ &= |\phi_i^T\hat{\theta}_{i-1}-y_i|\frac{\epsilon}{|\phi_i^T\hat{\theta}_{i-1}-y_i|} = \epsilon.\end{aligned}$$

Pictorially, the new estimate $\hat{\theta}_i$ is on the plane $|\phi_i^T\hat{\theta}-y_i| = \epsilon$ that is closer to $\hat{\theta}_{i-1}$.

A unique feature of the algorithm given in Theorem 3.1 is that no parameter update takes place if the new data does not contradict the hypothesis of the noise model. For a large data set, the consequence of the cessation of updating at time $N_0$ is that a very small percentage of data is used. This fact can be interpreted in two ways: (a) The algorithm is very efficient in terms of computational burden since there is no or little computation at most iterations. (b) In many applications, the noise may have small "averaging effect". Sufficient use of all the data likely helps reducing estimation errors, especially if the bound $\epsilon$ on $v_i$ is over-estimated. Notice that the standard LS or WLS algorithm does not assume any *a priori* knowledge on $v_i$ and gives the true estimate $\widehat{\theta}_i = \theta$ if $N \ge n$ and $V_N = 0$. In other words, the standard LS and WLS algorithms belong to the set of Correct Identification Algorithms [19] as defined by

$$\mathcal{C} = \{\mathcal{A} \;:\; \mathcal{A}(Y_N) = \widehat{\theta} = \theta \;\text{ if } V_N = 0\} \tag{3.2}$$

where $\mathcal{A}$ is any identification algorithm which maps the data $Y_N$ into the estimate $\widehat{\theta}$. It is also well-known that the standard LS estimate is worst-case optimal [5] in terms of output prediction error, i.e., let $\hat{\theta}_{LS} = (\Phi_N^T\Phi_N)^{-1}\Phi_N^TY_N$ denote the least squares estimate, then

$$\hat{\theta}_{LS} = \arg\min_{\widehat{\theta}=\mathcal{A}(Y_i)}\sup_{v_i}\frac{\sum_{k=1}^{i}(\phi_k^T\theta-\phi_k^T\widehat{\theta})^2}{\sum_{k=1}^{i}v_k^2}.$$

Now, the algorithm proposed in Theorem 3.1 is guaranteed to provide an estimate in the membership set $S_{N_0,\infty}(\epsilon+\delta)$ but it is not necessarily a correct algorithm as defined in (3.2) and neither worst-case optimal because many of $q_i$'s may be zero. Motivated by this observation, in the following, we modify the algorithm (3.1) replacing weights that are zero by some positive constant $q$ so that this modified algorithm achieves the following multiple objectives: (a) It is correct and worst-case optimal in terms of output error, similarly to the LS estimate; (b) It produces estimates satisfying pointwise noise constraint.

**Theorem 3.2** *Consider the system (1.1) with the noise $v_i$ bounded as in (1.4) by some known $\epsilon > 0$. Consider the recursive WLS algorithm (1.7) and (1.8) with $P_0 = P_0^T > 0$ and arbitrary $\widehat{\theta}_0$. Let $q_i$ be*

$$q_i = \begin{cases} \frac{|y_i-\phi_i^T\widehat{\theta}_{i-1}|-\epsilon}{\epsilon\phi_i^TP_{i-1}\phi_i} & \text{if } |y_i-\phi_i^T\widehat{\theta}_{i-1}| > \epsilon; \\ q & \text{if } |y_i-\phi_i^T\widehat{\theta}_{i-1}| \le \epsilon \end{cases} \tag{3.3}$$

*for $i \ge 1$, where $q > 0$ is any positive constant. Then, the proposed WLS estimate $\widehat{\theta}_i$ satisfies the following conditions:*
*(a) The algorithm is worst-case optimal, i.e., for each $i \ge n$*

$$\widehat{\theta}_i = \arg\min_{\widehat{\theta}=\mathcal{A}(Y_i)}\sup_{v_i}\frac{\sum_{k=1}^{i}(\phi_k^T\theta-\phi_k^T\widehat{\theta})^2}{\sum_{k=1}^{i}q_kv_k^2} \tag{3.4}$$

*where $\mathcal{A}$ is any identification algorithm and $Y_i$ is defined in (1.3) for $N = i$.*
*(b) For all $i$,*

$$|y_i-\phi_i^T\widehat{\theta}_i| \le \epsilon.$$

**Remark 2:** Note that the denominator in the cost function (3.4) can be interpreted as the weighted $\ell_2$ norm of the uncertainty corrupting the measurements; similarly the numerator can be interpreted as the $\ell_2$ norm of the output estimation error. Thus, the algorithm (3.3) minimizes the worst-case amplification from the weighted noise to the output estimation error. We remark, however, that another WLS algorithm with different weights than those given in (3.3) might provide smaller mean-square residuals even though not worst-case optimal according to the definition given in Theorem 3.2.

## 4. Concluding Remarks

In this paper, we have addressed the issue of weights selections for Weighted Least Squares algorithms in a set-membership context. The objective was to obtain an estimate that is compatible with the data and the *a priori* information available about the system model and the measurement noise.

In the first part of the paper, we focused on FIR systems. In particular, we have shown that the WLS estimate can be made to lie within the membership set for all iterations if the bound on the noise is known. If the bound is unknown, the estimate can be still made to converge to the membership set asymptotically by utilizing i.i.d. random inputs with zero mean. In the second part of the paper, we presented algorithms for IIR systems with arbitrary inputs. These algorithms converge to the membership set after a finite "learning period" and therefore are useful for both identification applications and adaptive control. Finally, we developed algorithms that have good worst-case output estimation performance (like the standard WLS) and, at the same time, satisfy pointwise noise constraint. In this regard, we feel that more research needs to be performed for obtaining algorithms that enjoy "good performance" properties and are also compatible with the *a priori* information.

## References


[1] Special issue on bounded-error estimation, *Int. J. Adap. Contr. and Signal Proc.*, Vol. 8, No. 1, pp.1-103, 1994.

[2] Special issue on bounded-error estimation (Part II), *Int. J. Adap. Contr. and Signal Proc.*, Vol. 9, No.1, pp.1-132, 1995.

[3] Special issue on trends in system identification, *Automatica*, Vol. 31, pp.1689-1883, 1995.

[4] H. Akcay and P.P. Khargonekar, "The least squares algorithm, parametric system identification and bounded noise," *Automatica*, Vol. 29, pp.1535-1540, 1993.

[5] E.W. Bai, K.M. Nagpal and R. Tempo, "Bounded error parameter estimation: noise models and recursive algorithms," *Automatica*, Vol. 32, pp. 985-999, 1996.

[6] V.A. Bondarko and V.A. Yakubovich, "The method of recursive aim inequalities in adaptive control theory," *Int. J. Adap. Contr. and Signal Proc.* , Vol. 6, pp.141-160, 1992.

[7] J. Chen, C.N. Nett, "The Carathéodory-Fejér Problem and $H_\infty/\ell_1$ Identification: a time domain approach," *IEEE Transactions on Automatic Control*, Vol. 40, pp. 729-735, 1995.

[8] J. Chen, C.N. Nett and M.K.H. Fan, "Worst case system identification in $H_\infty$: Validation of a priori information, essentially optimal algorithms and error bounds," *IEEE Trans. on Auto. Control*, Vol. 40, pp.1260-1265, 1995.

[9] S. Dasgupta and Y. Huang, "Asymptotically convergent modified recursive least squares with data dependent updating and forgetting for systems with bounded noise," *IEEE Trans. on Info. Theory*, Vol. 33, pp. 383-392, 1987.

[10] J. Deller, "Set membership identification in digital signal processing," *Acoust. Speech and Signal Process. Mag.*, Vol. 6, pp. 4-20, 1990.

[11] G. Goodwin and K. Sin, ADAPTIVE FILTERING, PREDICTION AND CONTROL, Prentice-Hall, 1984.

[12] L. Guo, D.W. Huang and E.J. Hannan, "On arx($\infty$) approximation", *J. of Multivariate Anal.*, Vol. 32, pp. 17-47, 1990.

[13] S.V. Gusev, "A finite convergent algorithm for estimating the regression function and its application in adaptive control", *Automatic and Remote Control*, pp. 99-108, 1989.

[14] H. Hjalmarsson and L. Ljung, "A unifying view of disturbances in identification", *10th IFAC Symposium on System Identification*, Copenhagen, Denmark, 1994

[15] H. Kwakernaak and R. Sivan, MODERN SIGNALS AND SYSTEMS, *Prentice Hall*, Englewood Cliffs, NJ, 1991.

[16] T.L. Lai, H. Robbins and C.Z. Wei, "Strong consistency of least squares estimates in multiple regression II," *J of Multivariate Anal*, Vol. 19, pp. 343-361, 1979.

[17] L. Ljung, SYSTEM IDENTIFICATION: THEORY FOR THE USER, Prentice-Hall, Englewood Cliffs, NJ, 1987.

[18] M. Milanese, "Properties of least squares estimates in set membership identification," *Automatica*, Vol.31, pp.327-332, 1994.

[19] M. Milanese and R. Tempo, "Optimal algorithms theory for robust estimation and prediction," *IEEE Transactions on Automatic Control*, Vol. AC-30, pp. 730-738, 1985.

[20] M. Milanese and A. Vicino, "Optimal estimation theory for dynamic systems with set membership uncertainty:an overview," *Automatica*, Vol. 27, pp. 997-1009, 1991.

[21] B. Ninness and G.C. Goodwin, "Estimation of model quality," *Automatica*, Vol. 31, pp. 1771-1797, 1995.

[22] J.P. Norton, "Identification and application of bounded parameter models," *Automatica*, Vol. 23, pp. 497-507, 1987.

[23] W.F. Stout, ALMOST SURE CONVERGENCE, Academic Press, New York, 1974.

[24] R. Tempo, "Worst-case optimality of smoothing algorithms for parametric system identification," *Automatica*, Vol. 31, pp. 759-764, 1995.

[25] J.F. Traub, G.W. Wasilkowski and H. Woźniakowski, INFORMATION-BASED COMPLEXITY, Academic Press, New York, 1988.

[26] E. Walter and H. Piet-Lahanier, "Estimation of parameter bounds from bounded-error data: A survey," *Mathematics and Computers in Simulation*, Vol. 32, pp. 449-468, 1990.