Unfalsified Weighted Least Squares Estimates in Set-Membership Identification

Er-Wei Bai, Li Qiu, *Member, IEEE,* and Roberto Tempo, *Member, IEEE*

***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 and 2) if the noise bound is unknown, the same results can still be achieved by using white input signals for finite impulse response systems (FIR).**

***Index Terms*—Identification, least square, set membership.**

I. Problem Statement

In this paper, we consider a discrete time scalar system

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

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 $v_i \in R\mathbf{R}$ the measurement noise. It is assumed that ϕ_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. Equation (1.1) can be rewritten in a compact vector form as

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

where

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

The purpose of system identification is to design an algorithm $\mathcal{A}$ which maps the input–output measurements y_i and ϕ_i into the estimate $\hat{\theta}$ of the unknown system parameter vector θ. 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 (MLE) can be derived [16]. In set-membership identification (see e.g., the special issues [1]–[3] and the survey papers [19]–[21] and [25]) the noise is assumed to be unknown but bounded by ϵ, 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 $\hat{\theta}$ coincides with the true but unknown θ but we can only detect whether $\hat{\theta}$ is compatible with the observed input–output data. To this end, the membership set is defined as follows:

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

An estimate $\hat{\theta}$ is compatible with the input–output data from the i_0th observation to the i_1th observation if and only if $\hat{\theta} \in S_{i_0, i_1}(\epsilon)$.

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., [10]. In this case, one example of paramount importance is when the measurements are affected by roundoff errors given by A/D converters [23]. 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 Φ_N, the WLS estimate $\hat{\theta}_N$ is the solution of the minimization problem

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

where q_i's are nonnegative weights. Letting

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

a closed form solution of the WLS estimate $\hat{\theta}_N$ can be easily derived [16] if $\Phi_N^T Q_N \Phi_N$ is nonsingular,

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

One of the powerful features of the WLS algorithm is that it can be implemented recursively [16] using only the current

Manuscript received March 5, 1996; revised September 10, 1996. The work of E.-W. Bai and L. Qiu was supported by the National Science Foundation under Grant ECS9710297 and HKUST. The work of R. Tempo was supported by CENS-CNR of Italy. This paper was recommended by Associate Editor U. Helmke.

E.-W. Bai is with the University of Iowa, Iowa City, IA 52242 USA (e-mail: erwei@icaen.uiowa.edu).

L. Qiu is with the Department of Electrical Engineering, Hong Kong Univeristy of Science and Technology, Clear Water Bay, Kowloon, Hong Kong (e-mail: eeqiu@ee.ust.hk).

R. Tempo is with National Research Council of Italy, Politecnico di Torino, Torino, Italy (e-mail: tempo@polito.it).

Publisher Item Identifier S 1057-7122(98)00237-2.

input–output measurement y_i, ϕ_i and the previous estimate $\hat{\theta}_{i-1}$

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

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 [17]. 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 $|v_i| \le \epsilon$ for some known ϵ, 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 $q_i \ge 0$ properly so that the WLS estimate $\hat{\theta}_i$ 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 (IBC) [24], 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 (LS) 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 [25]. 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 II we show 1) 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, 2) 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 unknown parameter θ asymptotically almost surely (a.s.)[1] for any bounded noise sequence v_i with *unknown noise bound* ϵ.

Consequently, the WLS estimate $\hat{\theta}_i$ converges to the membership set almost surely. If the system is infinite impulse response (IIR) (see, e.g., [14] for definitions of FIR and IIR systems) and the noise bound is known, in Section III 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 IV and some concluding remarks are outlined in Section V.

II. Finite Impulse Response Systems

In this section, we consider the FIR system

$$y_i = \phi_i^T\theta + v_i = (u_{i-1}, u_{i-2}, \cdots, u_{i-n})\theta + v_i, \quad i = 1, 2, \cdots, N. \tag{2.1}$$

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

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

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

for all $i_0 \ge 0$.

Theorem 2.1: Consider the FIR system (2.1) 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

$$q_i = \begin{cases} \dfrac{|y_i - \phi_i^T\hat{\theta}_{i-1}| - \epsilon}{\epsilon\phi_i^T P_{i-1}\phi_i} & \text{if } |y_i - \phi_i^T\hat{\theta}_{i-1}| > \epsilon; \\ 0 & \text{if } |y_i - \phi_i^T\hat{\theta}_{i-1}| \le \epsilon \end{cases} \tag{2.2}$$

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

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

Proof: See Section IV.

Next, we observe the following two facts.

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

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

The matrix Φ_n is nonsingular and this implies $P_0 > 0$.

[1] Here the definition of the almost sure convergence of the WLS estimate $\hat{\theta}_i$ to the true but unknown system vector θ is standard [22]

$$\text{Prob}\left\{\lim_{i\to\infty} \hat{\theta}_i = \theta\right\} = 1.$$

2) The weights q_i's in (2.2) are well defined because $\epsilon\phi_i^T P_{i-1}\phi_i \neq 0$ if $|y_i - \phi_i^T\hat{\theta}_{i-1}| > \epsilon$. This can be easily seen as follows: Since $P_i^{-1} = P_{i-1}^{-1} + q_i\phi_i\phi_i^T$ [11, p. 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 $|y_i - \phi_i^T\hat{\theta}_{i-1}| = |v_i| < \epsilon$ which is a contradiction.

The theorem above shows that if the bound ϵ on the unknown noise v_i is available, then the WLS estimate $\hat{\theta}_i$ 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 ϵ is the key. The following result shows that even when the noise bound ϵ is unknown, the WLS estimates can still be made to converge to the membership set asymptotically.

Theorem 2.2: Consider the FIR systems (2.1) 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 $\hat{\theta}_i$ satisfies

$$\|\hat{\theta}_i - \theta\| \to 0 \qquad \text{a.s.}$$

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

Proof: See Section IV.

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 ϵ on v_i may be unknown and it is indeed *not required* in the WLS algorithm. A similar result is reported [15] if the input sequence is deterministic and the noise is i.i.d. with zero mean.

III. Infinite Impulse Response Systems with Arbitrary Input

In Section III, 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 $\hat{\theta}_i$ 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 δ, 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 $\delta \to 0$. 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 δ as $\delta \to 0$. Then, by the inclusive property of the membership set, there exists some $\overline{\theta} \notin S_{N_0,\infty}(\epsilon)$ but $\overline{\theta} \in S_{N_0,\infty}(\epsilon+\delta)$ for any $\delta > 0$, i.e., for all $i \geq N_0$, $|\phi_i^T\overline{\theta} - y_i| \leq \epsilon + \delta$, for all $\delta > 0$. Then, it follows that $|\phi_i^T\overline{\theta} - y_i| \leq \epsilon$ and this would imply $\overline{\theta} \in S_{N_0,\infty}(\epsilon)$ which is a contradiction. Therefore, the two sets are "almost identical" for small δ.

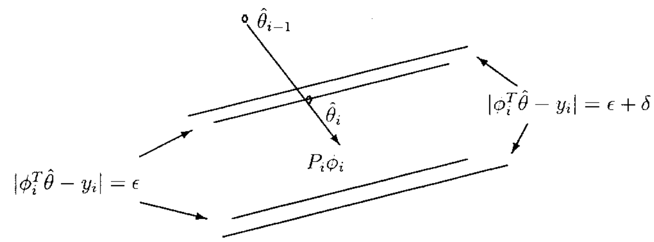

Fig. 1. Geometric interpretation of algorithm (3.1).

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 $\hat{\theta}_0$. For any $\delta > 0$, let q_i be

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

for $i \geq 1$. Then, the WLS estimate $\hat{\theta}_i$ 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 $i \geq N_0$

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

Proof: See Section IV.

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 $\hat{\theta}_i$ 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 θ. 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 online 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 δ 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 (1.7), we notice that the parameter update at ith iteration of the (weighted) least squares method is always along the direction of the vector $P_i\phi_i$; see Fig. 1, where $|\phi_i^T\hat{\theta} - y_i| = \epsilon$ represents two planes in the parameter space that bound all $\hat{\theta}$ compatible with the new data ϕ_i and y_i. The true parameter vector θ 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 $|\phi_i^T\hat{\theta}_{i-1} - y_i| \leq \epsilon+\delta$, then $\hat{\theta}_{i-1}$ is compatible with the new data ϕ_i and y_i and

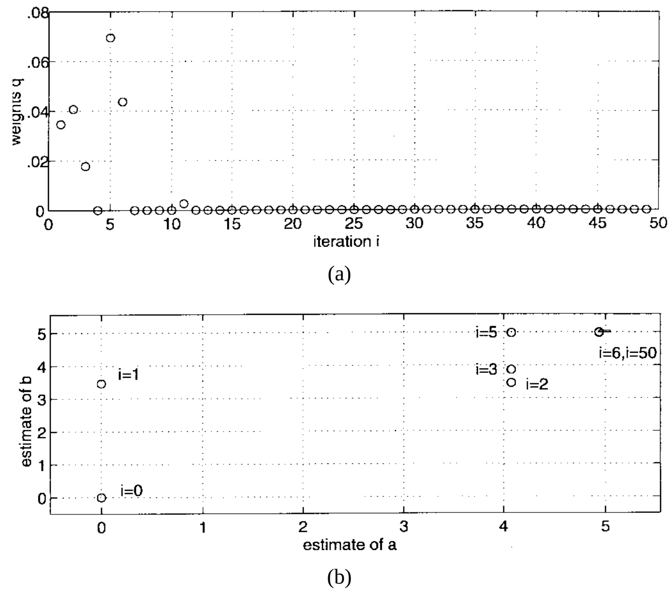

Fig. 2. Recursive estimates $\hat{a}_i$, $\hat{b}_i$, and weights q_i.

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 ϕ_i and y_i. In this case, we choose $q_i = (|y_i - \phi_i^T\hat{\theta}_{i-1}| - \epsilon)/(\epsilon\phi_i^T P_{i-1}\phi_i)$ such that

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

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

We now illustrate the performance of the algorithm presented in the Theorem 3.1 using a simulation example. We consider a second order system of the form

$$y_i = au_i + bu_{i-1} + v_i,$$

with (unknown) $\theta = \binom{a}{b} = \binom{5}{5}$. The initial estimate was chosen to be $\hat{\theta}_0 = (0, 0)^T$, $\delta = 0.01$ and $P_0 = \begin{pmatrix}100 & 0\\ 0 & 100\end{pmatrix}$. The input sequence applied was $\{u_i\} = \{1, 0, 1, 0, 1, 0, \cdots\}$ and v_i were chosen as independent random variables uniformly distributed in the interval $[-\epsilon, \epsilon]$ with $\epsilon = 1$. Fig. 2 shows the estimate $\hat{\theta}_i = \binom{\hat{a}_i}{\hat{b}_i}$ derived using the algorithm of Theorem 3.1 for 50 iterations from $i = 1$ to $i = 50$. The bottom diagram shows the estimate $\hat{b}_i$ versus the estimate $\hat{a}_i$ for $i = 0, 1, 2, 3, 5, 6$, and 50. The actual membership set $S_{1,50}(1)$ is depicted by a solid line box around the true parameters $a = 5$ and $b = 5$. The top figure shows the weights q_i versus time i. From Fig. 2 we see that after a few iterations, the weights q_i become 0 and update ceases.

As illustrated in the above example, 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. 1) The algorithm is very efficient in terms of computational burden since there is no or little computation at most iterations. 2) 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 ϵ on v_i is overestimated. Notice that the standard LS or WLS algorithm does not assume any *a priori* knowledge on v_i and gives the true estimate $\hat{\theta}_i = \theta$ if $N \geq n$ and $V_N = 0$. In other words, the standard LS and WLS algorithms belong to the set of correct identification algorithms (CIA) [18] as defined by

$$\mathcal{C} = \{\mathcal{A}\colon \mathcal{A}(Y_N) = \hat{\theta} = \theta \text{ if } V_N = 0\} \tag{3.2}$$

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

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

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 $\hat{\theta}_0$. Let q_i be

$$q_i = \begin{cases} \dfrac{|y_i - \phi_i^T\hat{\theta}_{i-1}| - \epsilon}{\epsilon\phi_i^T P_{i-1}\phi_i} & \text{if } |y_i - \phi_i^T\hat{\theta}_{i-1}| > \epsilon; \\ q & \text{if } |y_i - \phi_i^T\hat{\theta}_{i-1}| \leq \epsilon \end{cases} \tag{3.3}$$

for $i \geq 1$, where $q > 0$ is any positive constant. Then, the proposed WLS estimate $\hat{\theta}_i$ satisfies the following conditions.

1) The algorithm is worst case optimal, i.e., for each $i \geq n$

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

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

2) For all i,

$$|y_i - \phi_i^T\hat{\theta}_i| \leq \epsilon.$$

Proof: See Section IV.

Remark 2: Note that the denominator in the cost function (3.4) can be interpreted as the weighted ℓ_2 norm of the uncertainty corrupting the measurements; similarly the numerator can be interpreted as the ℓ_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.

IV. Proofs

Proof of Theorem 2.1: The proof relies on the following lemma that is a special case of Theorem 2.1 when the input $u(kn) = 1$ and $u(kn+1) = u(kn+2) = \cdots = u(kn+n-1) = 0$ for all $k = 1, 2, \cdots$. We then extend the result to general periodic input signal in the proof of Theorem 2.1.

Lemma 4.1: Consider the recursive WLS algorithm (1.7) and (1.8) with q_i and P_0^{-1} given in Theorem 2.1. Assume that u_i is periodic with period n so that

$$\phi_{kn+1} = \begin{bmatrix}1\\0\\\vdots\\0\end{bmatrix}, \quad \phi_{kn+2} = \begin{bmatrix}0\\1\\\vdots\\0\end{bmatrix}, \cdots, \phi_{kn+n} = \begin{bmatrix}0\\\vdots\\0\\1\end{bmatrix},$$

for all $k \geq 0$. Then, the WLS estimate $\hat{\theta}_i$ satisfies

$$\hat{\theta}_i \in S_{1,i}(\epsilon) = \bigcap_{m=1}^{i} \{\hat{\theta}; |y_m - \phi_m^T \hat{\theta}| \leq \epsilon\}$$

for all $i > 0$.

Proof: $P_0^{-1} = q_0 I$ and by the matrix inversion lemma (see [11, p. 58]), we have

$$P_{i-1}^{-1} = P_{i-2}^{-1} + q_{i-1}\phi_i\phi_i^T = q_0 I + \sum_{j=1}^{i-1} q_j \phi_j \phi_j^T.$$

For any $i \geq 1$, there exists some $k \geq 0$ and $1 \leq l \leq n$ so that $i = kn + l$. Thus,

$$P_{i-1}^{-1} = \operatorname{diag}\Bigg(q_0 + \sum_{j=0}^{k} q_{jn+1}, \cdots, q_0 + \sum_{j=0}^{k} q_{jn+l-1}, q_0 + \sum_{j=0}^{k-1} q_{jn+l}, \cdots, q_0 + \sum_{j=0}^{k-1} q_{jn+n}\Bigg),$$

$$P_{i-1}\phi_i = \left(0, \cdots, 0, \frac{1}{q_0 + \sum_{j=0}^{k-1} q_{jn+l}}, 0, \cdots, 0\right)^T$$

and

$$\phi_i^T P_{i-1}\phi_i = \frac{1}{q_0 + \sum_{j=0}^{k-1} q_{jn+l}}.$$

Letting $\theta = [\theta(1), \theta(2), \cdots, \theta(n)]^T$ and $\hat{\theta}_i = [\hat{\theta}_i(1), \hat{\theta}_i(2), \cdots, \hat{\theta}_i(n)]^T$, we obtain

$$y_i - \phi_i^T\hat{\theta}_{i-1} = \phi_i^T\theta + v_i - \phi_i^T\hat{\theta}_{i-1} = \theta(l) + v_i + \hat{\theta}_{i-1}(l)$$

and

$$y_i - \phi_i^T\hat{\theta}_i = \theta(l) + v_i - \hat{\theta}_i(l).$$

Therefore, $|y_i - \phi_i^T\hat{\theta}_i| \leq \epsilon$ holds if and only if

$$-\epsilon + \theta(l) + v_i \leq \hat{\theta}_i(l) \leq \theta(l) + \epsilon + v_i.$$

Now, decompose the noise sequence $\{v_i\}$ into n subsequences $\{v_{kn+1}\}, \{v_{kn+2}\}, \cdots, \{v_{kn+n}\}$, for $k \geq 0$. For any $i = kn + l$, $\hat{\theta}_i \in \bigcap_{m=1}^{i} \{\hat{\theta}; |y_m - \phi_m^T\hat{\theta}| \leq \epsilon\}$ if and only if

$$-\epsilon + \max_{0\leq j\leq k} v_{jn+l} + \theta(l) \leq \hat{\theta}_i(l) \leq \theta(l) + \min_{0\leq j\leq k} v_{jn+l} + \epsilon \tag{4.1}$$

for all $l = 1, 2, \cdots, n$ and $k \geq 1$ so that $kn + l \leq i$. We now prove the lemma by induction. We first check the case when $i = 1$. For any initial condition $\hat{\theta}_0$, if $|y_1 - \phi_1^T\hat{\theta}_0| \leq \epsilon$, then $q_1 = 0$. This implies $\hat{\theta}_1 = \hat{\theta}_0$ and $|y_1 - \phi_1^T\hat{\theta}_1| \leq \epsilon$. On the other hand, if $|y_1 - \phi_1^T\hat{\theta}_0| \geq \epsilon$, then $q_1 = (|y_1 - \phi_1^T\hat{\theta}_0| - \epsilon)/\epsilon\phi_1^T P_0 \phi_1$ and this implies

$$\begin{aligned}|y_1 - \phi_1^T\hat{\theta}_1| &= \left|y_1 - \phi^T\hat{\theta}_0 - \frac{q_1\phi_1^T P_0\phi_i}{1 + q_1\phi_1^T P_0\phi_i}(y_1 - \phi_1^T\hat{\theta}_0)\right| \\ &= |y_1 - \phi_1^T\hat{\theta}_0| \cdot \left|\frac{1}{1 + q_1\phi_1^T P_0\phi_i}\right| \\ &= |y_1 - \phi_1^T\hat{\theta}_0| \frac{\epsilon}{|y_1 - \phi_1^T\hat{\theta}_0|} = \epsilon.\end{aligned}$$

Therefore, for any y_1 and θ_0, $|y_1 - \phi_1^T\theta_1| \leq \epsilon$. We now show that if the (4.1) holds at $i-1$, then it is also true at i. In other words, let $i = kn + l$ and

$$\hat{\theta}_{i-1} \in \bigcap_{m=0}^{i-1} \{\hat{\theta}; |\phi_m^T\hat{\theta} - y_m| \leq \epsilon\} \tag{4.2}$$

then

$$\hat{\theta}_i \in \bigcap_{m=0}^{i} \{\hat{\theta}; |\phi_m^T\hat{\theta} - y_m| \leq \epsilon\}.$$

This is obvious if $|y_i - \phi_i^T\hat{\theta}_{i-1}| \leq \epsilon$ simply because $q_i = 0$ and $\hat{\theta}_i = \hat{\theta}_{i-1}$. For $|y_i - \phi_i^T\hat{\theta}_{i-1}| > \epsilon$, we consider two different cases:

Case 1: If $y_i - \phi_i^T\hat{\theta}_{i-1} > \epsilon$,

$$\hat{\theta}_{i-1}(l) < \theta(l) + v_{kn+l} - \epsilon. \tag{4.3}$$

However, from (4.1) and (4.2), we have

$$\hat{\theta}_{i-1}(l) \geq -\epsilon + \theta(l) + \max_{0\leq j\leq (k-1)} v_{jn+l}. \tag{4.4}$$

By comparing (4.3) and (4.4), it easily follows that

$$v_{kn+l} > \max_{0\leq j\leq (k-1)} v_{jn+l}$$

and

$$\phi_i^T\hat{\theta}_i - y_i = \phi_i^T\hat{\theta}_{i-1} + \frac{q_i\phi_i^T P_{i-1}\phi_i}{1+q_i\phi_i^T P_{i-1}\phi_i}(y_i - \phi_i^T\hat{\theta}_{i-1}) - y_i = \epsilon\frac{\phi_i^T\hat{\theta}_{i-1} - y_i}{|y_i - \phi_i^T\hat{\theta}_{i-1}|} = -\epsilon.$$

From this, we have

$$\hat{\theta}_i(l) = \theta(l) + v_{kn+l} - \epsilon \geq \theta(l) - \epsilon + \max_{0\leq j\leq k} v_{jn+l}.$$

Also, since $|v_i| \leq \epsilon$,

$$\hat{\theta}_i(l) = \theta(l) + v_{kn+l} - \epsilon \leq \theta(l) + \epsilon + \min_{0\leq j\leq k} v_{jn+l}$$

and this implies

$$-\epsilon + \max_{0\leq j\leq k} v_{jn+l} + \theta(l) \leq \theta(l) \leq \theta(l) + \min_{0\leq j\leq k} v_{jn+l} + \epsilon.$$

Notice that at the ith ($i = kn+l$) update only the lth component of $\hat{\theta}_i$ is affected, not other components $\hat{\theta}_i(m)$, $m \neq l$ that are compatible with all the previous and current data. Therefore, $\hat{\theta}_i \in \bigcap_{m=0}^i \{\hat{\theta};\, |\phi_m^T\hat{\theta} - y_m| \leq \epsilon\}$.

Case 2: $y_i - \phi_i^T\hat{\theta}_{i-1} < -\epsilon$. The proof for this case is similar to Case 1.

Combining Cases 1 and 2, we have completed the induction and consequently the proof of Lemma 4.1.

We now prove Theorem 2.1. Recall that $i = kn + l$ and $\Phi_n\Phi_n^{-1} = \Phi_n^{-1}\Phi_n = I$. Hence,

$$\phi_i^T\begin{pmatrix}\phi_1^T\\ \vdots \\ \phi_n^T\end{pmatrix}^{-1} = \Big(0, \cdots, 0, \underbrace{1}_{i\text{th}}, 0, \cdots, 0\Big)$$

and

$$(\phi_1, \cdots, \phi_n)^{-1}\phi_i = \Big(0, \cdots, 0, \underbrace{1}_{i\text{th}}, 0, \cdots, 0\Big)^T.$$

Defining $O_i = \Phi_n^{-1}P_i\Phi_n^{T^{-1}}$, $\sigma_i = \Phi_n^T\phi_i$, $\xi_i = \Phi_n^{-1}\hat{\theta}_i$, $\eta = \Phi_n^{-1}\theta$, we have

$$O_i^{-1} = \Phi_n^T\left(q_0\Phi_n^{T^{-1}}\Phi_n^{-1} + \sum_{j=1}^i q_j\phi_j\phi_j^T\right)\Phi_n = q_0 I + \sum_{j=1}^i q_j\sigma_j\sigma_j^T,$$

$$\xi_i = \Phi_n^{-1}\hat{\theta}_i = \xi_{i-1} + \frac{q_i O_{i-1}\sigma_i}{1+q_i\sigma_i^T O_{i-1}\sigma_i}(y_i - \sigma_i^T\xi_{i-1}).$$

Note that $y_i - \phi_i^T\hat{\theta}_{i-1} = y_i - \sigma_i^T\xi_{i-1}$ and

$$\frac{|y_i - \phi_i^T\hat{\theta}_{i-1}| - \epsilon}{\epsilon\phi_i^T P_{i-1}\phi_i} = \frac{|y_i - \sigma_i^T\xi_{i-1}| - \epsilon}{\epsilon\sigma_i^T O_{i-1}\sigma_i}.$$

The conclusion follows immediately from Lemma 4.1.

Proof of Theorem 2.2: The proof is reminiscent of that which appeared in [15] for a symmetric problem where the input sequence is deterministic and the noise is i.i.d. with zero mean. Thus, we will only provide a sketch of proof here as much of details are similar and can be found in [15]. Write

$$\hat{\theta}_i - \theta = (\Phi_i^T Q_i\Phi_i)^{-1}\Phi_i^T Q_i V_i.$$

Since $0 < \underline{q} \leq q_i \leq \overline{q} < \infty$ and u_i is i.i.d., by [12], we have

$$\sigma_{\min}(\Phi_i^T Q_i\Phi_i) \geq \underline{q}\sigma_{\min}(\Phi_i^T\Phi_i) \propto \mathcal{O}(i), \qquad \text{a.s.}$$

where $\sigma_{\min}$ indicates the minimum eigenvalue. On the other hand, each row of the matrix $(\Phi_i^T Q_i V_i)$ is given by $\sum_{j=1}^i u_{j+k}q_jv_j$ for some k. The u_{j+k}'s are i.i.d. with zero mean and this implies $u_{j+k}q_jv_j$'s are independent with zero mean and $\text{Var}(u_{j+k}q_jv_j) = q_j^2v_j^2\sigma^2 \leq \overline{q}^2\epsilon^2\sigma^2 < \infty$, where σ^2 is the variance of v_i. From [22, Corollary 3.4.2], it follows

$$\frac{1}{i}\sum_{j=1}^i u_{j+k}q_jv_j \to 0, \qquad \text{a.s.}$$

as $i \to \infty$. Therefore

$$\begin{aligned}\|\hat{\theta}_i - \theta\| &= \|(\Phi_i^T Q_i\Phi_i)^{-1}\Phi_i^T Q_i V_i\| \\ &\leq \|(\Phi_i^T Q_i\Phi_i)^{-1}\|\cdot\|\Phi_i^T Q_i V_i\| \\ &\propto \mathcal{O}\left(\frac{1}{i}\sum_{j=1}^i u_{j+k}q_jv_j\right) \to 0, \qquad \text{a.s.}\end{aligned}$$

This completes the proof.

Proof of Theorem 3.1: Let $\tilde{\theta}_i = \hat{\theta}_i - \theta$ and $L(i) = \tilde{\theta}_i^T P_i^{-1}\tilde{\theta}_i$. The following equations can be checked easily:

a) $$\tilde{\theta}_i = \tilde{\theta}_{i-1} + \frac{q_iP_{i-1}\phi_i}{1+q_i\phi_i^T P_{i-1}\phi_i}(y_i - \phi_i^T\hat{\theta}_{i-1}) = P_iP_{i-1}^{-1}\tilde{\theta}_{i-1} + \frac{q_iv_iP_{i-1}\phi_i}{1+q_i\phi_i^T P_{i-1}\phi_i}.$$

b) If $|y_i - \phi_i^T\hat{\theta}_{i-1}| > \epsilon + \delta$, then

$$1 + q_i\phi_i^T P_{i-1}\phi_i = \frac{|y_i - \phi_i^T\hat{\theta}_{i-1}|}{\epsilon}$$

and

$$\frac{q_i}{1+q_i\phi_i^T P_{i-1}\phi_i} = \frac{|y_i - \phi_i^T\hat{\theta}_{i-1}| - \epsilon}{|y_i - \phi_i^T\hat{\theta}_{i-1}|\phi_i^T P_{i-1}\phi_i}.$$

c) $$\tilde{\theta}_{i-1}^T P_{i-1}^{-1}P_iP_{i-1}^{-1}\tilde{\theta}_{i-1} = \tilde{\theta}_{i-1}^T P_{i-1}^{-1}\tilde{\theta}_{i-1} - \frac{q_i\tilde{\theta}_{i-1}\phi_i\phi_i^T\tilde{\theta}_{i-1}}{1+q_i\phi_i^T P_{i-1}\phi_i}.$$

$$\frac{\phi_i^T P_{i-1}P_i^{-1}P_{i-1}\phi_i}{(1+q_i\phi_i^T P_{i-1}\phi_i)^2} = \frac{\phi_i^T P_{i-1}\phi_i}{1+q_i\phi_i^T P_{i-1}\phi_i};$$

d) $$L(i) = \tilde{\theta}_{i-1}^T P_{i-1}^{-1}\tilde{\theta}_{i-1} - \frac{q_i|\phi_i^T\tilde{\theta}_{i-1}|^2}{1+q_i\phi_i^T P_{i-1}\phi_i} + \frac{2q_iv_i\phi_i^T\tilde{\theta}_{i-1}}{1+q_i\phi_i^T P_{i-1}\phi_i} + \frac{q_i^2v_i^2\phi_i^T P_{i-1}\phi_i}{1+q_i\phi_i^T P_{i-1}\phi_i}.$$

Therefore, if $|y_i - \phi_i^T\hat{\theta}_{i-1}| \le \epsilon + \delta$, $\Delta L(i) = L(i) - L(i-1) = 0$ and if $|y_i - \phi_i^T\hat{\theta}_{i-1}| > \epsilon + \delta$, from the above equations a)–d), we have

$$\begin{aligned} L(i) - L(i-1) &= \frac{q_i}{1+q_i\phi_i^T P_{i-1}\phi_i} \\ &\quad \cdot \left(-|\phi_i^T\tilde{\theta}_{i-1}|^2 + 2v_i\phi_i^T\tilde{\theta}_{i-1} + q_i v_i^2 \phi_i^T P_{i-1}\phi_i\right) \\ &= \frac{q_i}{1+q_i\phi_i^T P_{i-1}\phi_i} \\ &\quad \cdot \left\{-(\phi_i^T\tilde{\theta}_{i-1} - v_i)^2 + v_i^2 + q_i v_i^2 \phi_i^T P_{i-1}\phi_i\right\} \\ &= \frac{q_i}{1+q_i\phi_i^T P_{i-1}\phi_i} \\ &\quad \cdot \left\{-(\phi_i^T\hat{\theta}_{i-1} - y_i)^2 + v_i^2(1 + q_i\phi_i^T P_{i-1}\phi_i)\right\} \\ &= \frac{q_i}{1+q_i\phi_i^T P_{i-1}\phi_i} \\ &\quad \cdot \left\{-(\phi_i^T\hat{\theta}_{i-1} - y_i)^2 + v_i^2 \frac{|y_i - \phi_i^T\hat{\theta}_{i-1}|}{\epsilon}\right\} \\ &\le \frac{q_i}{1+q_i\phi_i^T P_{i-1}\phi_i} \\ &\quad \cdot \left\{-(\phi_i^T\hat{\theta}_{i-1} - y_i)^2 + \epsilon|y_i - \phi_i^T\hat{\theta}_{i-1}|\right\} \\ &= -\frac{|y_i - \phi_i^T\hat{\theta}_{i-1}| - \epsilon}{|y_i - \phi_i^T\hat{\theta}_{i-1}|\phi_i^T P_{i-1}\phi_i}|y_i - \phi_i^T\hat{\theta}_{i-1}| \\ &\quad \cdot \left\{|y_i - \phi_i^T\hat{\theta}_{i-1}| - \epsilon\right\} \\ &= -\frac{\{|y_i - \phi_i^T\hat{\theta}_{i-1}| - \epsilon\}^2}{\phi_i^T P_{i-1}\phi_i}. \end{aligned}$$

Note $||\phi_i||^2 \le M$ for all i. Also, $P_i \le P_{i-1} \le \cdots \le P_0$. Thus

$$\frac{1}{\phi_i^T P_{i-1}\phi_i} \ge \frac{1}{M\sigma_{\max}(P_0)}$$

and

$$\begin{aligned} \Delta L(i) &= L(i) - L(i-1) \\ &\le \frac{-1}{M\sigma_{\max}(P_0)}\left\{|y_i - \phi_i^T\hat{\theta}_{i-1}| - \epsilon\right\}^2. \end{aligned}$$

Since $|y_i - \phi_i^T\hat{\theta}_{i-1}| > \epsilon + \delta$ implies $|y_i - \phi_i^T\hat{\theta}_{i-1}| - \epsilon > \delta$ and

$$\Delta L(i) \le \frac{-\delta^2}{M\sigma_{\max}(P_0)}$$

if $|y_i - \phi_i^T\hat{\theta}_{i-1}| > \epsilon + \delta$. Note again $L(0)$ is bounded. Therefore, the number of changes of the estimate $\hat{\theta}_i$ for the case $|y_i - \phi_i^T\hat{\theta}_{i-1}| > \epsilon + \delta$ is finite. This completes the proof.

Proof of Theorem 3.2: Let $\Phi_i = U\binom{D}{0}W^T$, where U and W are unitary matrices and $Q_i = U\overline{Q}U^T = U\begin{pmatrix}\overline{Q}_{11} & \overline{Q}_{12} \\ \overline{Q}_{21} & \overline{Q}_{22}\end{pmatrix}U^T$. Notice that the cost function (3.4) is finite only if $\mathcal{A}$ a correct identification algorithm. Therefore, we only have to show that Algorithm 3.3 achieves the minimum in the set $\mathcal{C}$. We demonstrate by the following.

1) For each $\mathcal{A} \in \mathcal{C}$ and $\hat{\theta} = \mathcal{A}(Y_i)$

$$\sup_{v_i} \frac{\sum_{k=1}^{i}(\phi_k^T\theta - \phi_k^T\hat{\theta})^2}{\sum_{k=1}^{i} q_k v_k^2} \ge \frac{1}{\sigma_{\min}(\overline{Q}_{11})}. \tag{4.5}$$

2) The WLS estimate $\hat{\theta}_i$ provided by (3.3) satisfies

$$\sup_{v_i} \frac{\sum_{k=1}^{i}(\phi_k^T\theta - \phi_k^T\hat{\theta}_i)^2}{\sum_{k=1}^{i} q_k v_k^2} \le \frac{1}{\sigma_{\min}(\overline{Q}_{11})}. \tag{4.6}$$

Now, DW^T is nonsingular. Let $\xi \in R^n$ such that $DW^T(\xi - \theta)$ is an eigenvector of $\overline{Q}_{11}$ associated with the smallest eigenvalue $\sigma_{\min}(\overline{Q}_{11})$ of $\overline{Q}_{11}$. Let the noise be $V_i = \Phi_i(\xi - \theta)$. Then,

$$Y_i = \Phi_i\theta + V_i = \Phi_i\xi.$$

If $\mathcal{A} \in \mathcal{C}$, we have $\mathcal{A}(Y_i) = \hat{\theta} = \xi$ and

$$\Phi_i\hat{\theta} - \Phi_i\theta = U\begin{pmatrix}D\\0\end{pmatrix}W^T(\xi - \theta).$$

Consequently, (4.7) shown at the bottom of the page. Thus,

$$\sup_{v_i} \frac{\sum_{k=1}^{i}(\phi_k^T\theta - \phi_k^T\hat{\theta})^2}{\sum_{k=1}^{i} q_k v_k^2} \ge \frac{1}{\sigma_{\min}(\overline{Q}_{11})}$$

for any $\mathcal{A} \in \mathcal{C}$.

$$\begin{aligned} \frac{||\Phi_i\hat{\theta} - \Phi_i\theta||^2}{\sum_{k=1}^{i} q_k v_k^2} &= \frac{(\xi-\theta)^T W(D^T 0)U^T U\begin{pmatrix}D\\0\end{pmatrix}W^T(\xi-\theta)}{(\xi-\theta)^T\Phi^T(i)Q_i\Phi_i(\xi-\theta)} \\ &= \frac{(\xi-\theta)^T W D^T D W^T(\xi-\theta)}{(\xi-\theta)^T W(D^T 0)U^T U\begin{pmatrix}\overline{Q}_{11} & \overline{Q}_{12}\\ \overline{Q}_{21} & \overline{Q}_{22}\end{pmatrix}U^T U\begin{pmatrix}D\\0\end{pmatrix}W^T(\xi-\theta)} \\ &= \frac{(\xi-\theta)^T W D^T D W^T(\xi-\theta)}{(\xi-\theta)^T W D^T\overline{Q}_{11} D W^T(\xi-\theta)} = \frac{1}{\sigma_{\min}(\overline{Q}_{11})} \end{aligned} \tag{4.7}$$

On the other hand, $\Phi_i\hat{\theta}_i - \Phi_i\theta = \Phi_i(\Phi_i^T Q_i \Phi_i)^{-1}\Phi_i^T Q_i V_i$ and

$$\begin{aligned}\Phi_i(\Phi_i^T Q_i^T \Phi_i)^{-1}\Phi_i^T &= U\begin{pmatrix} D \\ 0\end{pmatrix} W^T \left\{ W(D^T 0)U^T U\begin{pmatrix}\overline{Q}_{11} & \overline{Q}_{12} \\ \overline{Q}_{21} & \overline{Q}_{22}\end{pmatrix} \cdot U^T U \begin{pmatrix} D \\ 0\end{pmatrix} W^T\right\}^{-1} W(D^T 0)U^T \\ &= U\begin{pmatrix} D \\ 0\end{pmatrix} W^T W D^{-1} \overline{Q}_{11}{}^{-1} D^{-1} W^T W (D^T 0) U^T \\ &= U\begin{pmatrix}\overline{Q}_{11}^{-1} & 0 \\ 0 & 0\end{pmatrix} U^T.\end{aligned}$$

Hence

$$\begin{aligned}\Phi_i\hat{\theta}_i - \Phi_i\theta &= U\begin{pmatrix}\overline{Q}_{11}^{-1} & 0 \\ 0 & 0\end{pmatrix} U^T Q_i Q_i^{-1/2} Q_i^{1/2} V_i \\ &= O Q_i^{1/2} V_i\end{aligned}$$

where $O = U\begin{pmatrix}\overline{Q}_{11}^{-1} & 0 \\ 0 & 0\end{pmatrix} U^T Q_i Q_i^{-1/2}$. Now for any V_i,

$$\begin{aligned}\frac{\|\Phi_i\hat{\theta}_i - \Phi_i\theta\|^2}{\sum_{k=1}^{i} q_k v_k^2} &= \frac{V^T Q_i^{1/2} O^T O Q_i^{1/2} V}{V^T Q_i^{1/2} Q_i^{1/2} V} \le \sigma_{\max}(O^T O) \\ &= \sigma_{\max}\left[U\begin{pmatrix}\overline{Q}_{11}^{-1} & 0 \\ 0 & 0\end{pmatrix} U^T Q_i Q_i^{-1/2} \cdot Q_i^{-1/2} Q_i U \begin{pmatrix}\overline{Q}_{11}^{-1} & 0 \\ 0 & 0\end{pmatrix} U^T\right] \\ &= \sigma_{\max}\left[U\begin{pmatrix}\overline{Q}_{11}^{-1} & 0 \\ 0 & 0\end{pmatrix} U^T U \begin{pmatrix}\overline{Q}_{11} & \overline{Q}_{12} \\ \overline{Q}_{21} & \overline{Q}_{22}\end{pmatrix} \cdot U^T U \begin{pmatrix}\overline{Q}_{11}^{-1} & 0 \\ 0 & 0\end{pmatrix} U^T\right] \\ &= \sigma_{\max}\left[U\begin{pmatrix}\overline{Q}_{11}^{-1} & 0 \\ 0 & 0\end{pmatrix} U^T\right] \\ &= \sigma_{\max}(\overline{Q}_{11}^{-1}) = \frac{1}{\sigma_{\min}(\overline{Q}_{11})}.\end{aligned}$$

This completes the proof.

V. Concluding Remarks

In this paper, we have addressed the issue of weights selections for WLS 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. Adapt. Control Signal Process.,* vol. 8, no. 1, pp. 1–103, 1994.

[2] Special issue on bounded-error estimation (Part II), *Int. J. Adapt. Control Signal Process.,* 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. Adapt. Control Signal Process.,* vol. 6, pp. 141–160, 1992.

[7] J. Chen and C. N. Nett, "The Carathéodory–Fejér problem and H_∞/ℓ_1 identification: A time domain approach," *IEEE Trans. Automat. Contr.,* vol. 40, pp. 729–735, 1995.

[8] J. Chen, C. N. Nett, and M. K. H. Fan, "Worst case system identification in H_∞: Validation of *a priori* information, essentially optimal algorithms and error bounds," *IEEE Trans. Automat. Contr.,* vol. 40, pp. 1260–1265, 1995.

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

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

[11] G. Goodwin and K. Sin, *Adaptive Filtering, Prediction and Control.* Englewood Cliffs, NJ: Prentice-Hall, 1984.

[12] L. Guo, D. W. Huang, and E. J. Hannan, "On arx approximation," *J. 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," *Autom. Remote Contr.,* pp. 99–108, 1989.

[14] H. Kwakernaak and R. Sivan, *Modern Signals and Systems.* Englewood Cliffs, NJ: Prentice-Hall, 1991.

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

[16] L. Ljung, *System Identification: Theory for the User.* Englewood Cliffs, NJ: Prentice-Hall, 1987.

[17] M. Milanese, "Properties of least squares estimates in set membership identification," *Automatica,* vol. 31, pp. 327–332, 1994.

[18] M. Milanese and R. Tempo, "Optimal algorithms theory for robust estimation and prediction," *IEEE Trans. Automat. Contr.,* vol. AC-30, pp. 730–738, 1985.

[19] M. Milanese and A. Vicino, "Optimal estimation theory for dynamic systems with set membership uncertainty: An overview," *Automatica,* vol. 27, pp. 997–1009, 1991.

[20] B. Ninness and G. C. Goodwin, "Estimation of model quality," *Automatica,* vol. 31, pp. 1771–1797, 1995.

[21] J. P. Norton, "Identification and application of bounded parameter models," *Automatica,* vol. 23, pp. 497–507, 1987.

[22] W. F. Stout, *Almost Sure Convergence.* New York: Academic, 1974.

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

[24] J. F. Traub, G. W. Wasilkowski, and H. Woźniakowski, *Information-Based Complexity.* New York: Academic, 1988.

[25] E. Walter and H. Piet-Lahanier, "Estimation of parameter bounds from bounded-error data: A survey," *Math. Comput. Simul.* vol. 32, pp. 449–468, 1990.

Er-Wei Bai received his education from Fudan University, Shanghai Jiaotong University, China, and the University of California, Berkeley.

He is currently Associate Professor of Electrical Engineering at the University of Iowa, Iowa City, where he teaches and conducts research in the area of systems, controls, and identifications.

Li Qiu (S'85–M'90) received the B.Eng. degree in electrical engineering from Hunan University, China, and the M.A.Sc. and Ph.D. degrees in electrical engineering from the University of Toronto, Toronto, Ont., Canada, in 1987 and 1990, respectively.

Since 1990, he has had research and teaching positions at the University of Toronto, Canadian Space Agency, University of Waterloo, Waterloo, Ont., and University of Minnesota, Minneapolis. At present, he is an Assistant Professor at the Department of Electrical and Electronic Engineering, Hong Kong University of Science and Technology. His current research interests include robust control, digital control, signal processing, and motor control.

Dr. Qiu is an Associate Editor of the IEEE TRANSACTIONS ON AUTOMATIC CONTROL.

Roberto Tempo (M'90) was born in Cuorgné, Italy, in 1956. In 1980, he graduated in electrical engineering at Politecnico di Torino, Italy.

From 1981 to 1983, he was with the Dipartimento di Automatica e Informatica, Politecnico di Torino. In 1984, he joined the National Research Council of Italy (CNR) at the research center CENS, Torino, where he is currently a Director of Research of Systems and Computer Engineering. He has held visiting and research positions at Columbia University, New York, NY, German Aerospace Research Organization, Oberpfaffenhofen, and the University of Wisconsin, Madison. His research interests lie in the area of modeling, identification, and control of uncertain systems.

Dr. Tempo is currently an Editor of *Automatica* and an Associate Editor of *Systems and Control Letters* and *Journal of Complexity*. In 1993, he received the outstanding Paper Prize from the International Federation of Automatic Control for a paper published in *Automatica*.