Parameter Convergence in Nonlinearly Parameterized Systems

Chengyu Cao, Anuradha M. Annaswamy, *Fellow, IEEE*, and Aleksandar Kojic, *Associate Member, IEEE*

***Abstract*—A large class of problems in parameter estimation concerns nonlinearly parametrized systems. Over the past few years, a stability framework for estimation and control of such systems has been established. We address the issue of parameter convergence in such systems in this paper. Systems with both convex/concave and general parameterizations are considered. In the former case, sufficient conditions are derived under which parameter estimates converge to their true values using a min–max algorithm as in a previous work by Annaswamy, *et al.* In the latter case, to achieve parameter convergence a hierarchical min–max algorithm is proposed where the lower level consists of a min–max algorithm and the higher level component updates the bounds on the parameter region within which the unknown parameter is known to lie. Using this hierarchical algorithm, a necessary and sufficient condition is established for global parameter convergence in systems with a general nonlinear parameterization. In both cases, the conditions needed are shown to be stronger than linear persistent excitation conditions that guarantee parameter convergence in linearly parametrized systems. Explanations and examples of these conditions and simulation results are included to illustrate the nature of these conditions. A general definition of nonlinear persistent excitation that leads to parameter convergence is proposed at the end of this paper.**

***Index Terms*—Author, please supply your own keywords or send a blank e-mail to keywords@ieee.org to receive a list of suggested keywords..**

I. Introduction

RECENTLY, a stability framework has been established for studying estimation and control of nonlinearly parameterized (NLP) systems in [1]–[8]. In [1]–[7], various NLP systems were considered and the conditions for global stability, regulation and tracking were derived using a min–max algorithm, while in [8], stability and parameter convergence in a class of discrete-time systems was considered. In this paper, we consider parameter convergence in a class of continuous-time dynamic systems. We begin with systems that have convex/concave parameterization and derive sufficient conditions under which parameter convergence can occur in such systems. These conditions are related to linear persistent excitation (LPE) conditions relevant for convergence in linearly parameterized systems [9], and are shown to be stronger, with the additional complexity being a function of the underlying nonlinearity.

We also propose a new hierarchical min–max algorithm in this paper in order to relax the sufficient conditions for parameter convergence. The lower level of this algorithm consists of the same min–max algorithm as in [1] and [6]. An additional higher level component is included in the hierarchical algorithm that consists of updating the bounds on the parameter region that the unknown parameter is assumed to belong to. We then show, using the hierarchical algorithm, that parameter convergence can be accomplished globally under a necessary and sufficient condition on the system variables and the underlying nonlinearity f. Examples of functions that satisfy such a condition, which we denote as a condition of nonlinear persistent excitation (NLPE), and relations to LPE are also presented in this paper.

The paper is organized as follows. Section II gives the statement of the problem, the estimator based on the min–max algorithm and the properties. In Section III, parameter estimation in functions that are concave/convex is considered, and a sufficient condition for parameter convergence is derived. In Section IV, a hierarchical min–max algorithm is proposed and necessary and sufficient conditions for parameter convergence are proposed. Examples and relation to LPE are also presented in this section. Simulation results are included in Section V. Summary and concluding remarks are stated in Section VI. Proofs of all properties, lemmas, and theorems can be found in Appendix A.

II. Statement of the Problem

The problem considered is the estimation of unknown parameters in a class of nonlinear systems of the form

$$\begin{aligned} \dot{y} &= -\alpha(y,u)y + f(\theta_0, \omega(y,u)) \\ 0 < &\alpha_{\min} \le \alpha(y,u) \le \alpha_{\max} \end{aligned} \tag{1}$$

where $\theta_0 \in \Omega^0 \subset \mathbb{R}^n$ are bounded unknown parameters, $u, y \in \mathbb{R}$ are input and output respectively, and the functions ω and f are given by $\omega : \mathbb{R} \times \mathbb{R} \to \mathbb{R}^m$ and $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$.

We make the following assumptions regarding ω and f.

Assumption 1: The function $\omega(y(t), u(t))$ is Lipschitz in t so that

$$\|\omega(t_1) - \omega(t_2)\| \le U_b \|t_1 - t_2\| \qquad \forall\, t_1, t_2 \in \mathbb{R}^+.$$

Assumption 2: f is Lipschitz with respect to its arguments, i.e.,

$$|f(\theta + \Delta\theta, \omega + \Delta\omega) - f(\theta, \omega)| \le B_\theta \|(\Delta\omega, \Delta\theta)\|$$

Manuscript received November 27, 2001; revised June 27, 2002 and Sptember 5, 2002. Recommended by Associate Editor A. Datta. This work was supported in part by the U.S. Army Research Office under Grant DAAG55-98-1-0235 and in part by the National Science Foundation under Grant ECS-0070039.

C. Cao and A. M. Annaswamy are with the Active-Adaptive Control Laboratory, Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: aanna@mit.edu).

A. Kojic is with the Robert Bosch Corporation, Palo Alto, CA 94130 USA.

Digital Object Identifier 10.1109/TAC.2003.809146

0018-9286/03$17.00 © 2003 IEEE

IEEE Proof

$$\leq B_\theta(\|\Delta\omega\| + \|\Delta\theta\|).$$

Assumption 3: $\omega(y,u)$ is a bounded, continuous function of its arguments, and u is bounded and continuous.

Assumption 4: The system in (1) has bounded solutions if u is bounded.

Assumption 5: $\theta_0 \in \Omega^0 \subset \mathbb{R}^n$, and Ω^0 is a known compact set.

Let a set U_I be defined as follows:

$$U_I = \{\omega_i, i = 1, \ldots, I, \omega_i \neq \omega_j \text{ if } i \neq j, \omega_i \in \mathbb{R}^m\}. \quad (2)$$

We introduce the definition of an identifiable function which is necessary for parameter convergence.

Definition 1: A function $f(\theta,\omega), \theta \in \Omega^0 \subset \mathbb{R}^n$ is identifiable over parameter region Ω^0 with respect to U_I if there does not exist $\theta_1, \theta_2 \in \Omega^0$ and $\theta_1 \neq \theta_2$ such that

$$\lim_{\theta \to \theta_1} f(\theta,\omega_i) = \lim_{\theta \to \theta_2} f(\theta,\omega_i) \; \forall\, \omega_i \in U_I, i = 1, \ldots, I.$$

Definition 1 implies that identifiability follows if the system of equations:

$$f(\hat{\theta},\omega_i) - f(\theta_0,\omega_i) = 0 \quad \forall\, \omega_i \in U_I \quad (3)$$

has a *unique* solution $\hat{\theta} = \theta_0$ for any $\theta_0 \in \Omega^0$. Equation (3) suggests a procedure for constructing U_I such that for a given Ω^0, f can become identifiable over Ω^0. That is, the number I and the value ω_i, for $i = 1, \ldots, I$ must be chosen such that (3) has a unique solution.

We also note that for a given Ω^0, identifiability of f is dependent on the choice of U_I. For example, if f is linear, then f is identifiable over any $\theta \in \mathbb{R}^n$ if elements of U_I span the entire space of $\mathbb{R}^m$; for a nonlinear f, identifiability may be possible even if these elements span only a subspace. We notice that if f is not identifiable with respect to U_I, it implies that we have no way of identifying θ_0 using any input ω_i in U_I.

In Sections II-A–C, we propose a min–max parameter estimation algorithm, and its properties. For simplicity, we omit the arguments of ω, and note that it is a measurable continuous function of time that satisfies Assumption 1.

A. Min–Max Parameter Estimation Algorithm

The dynamics of parameter estimation algorithm that we propose is the same as the min–max algorithm in [1] and is as follows:

$$\begin{aligned}\dot{\hat{y}} =& -\alpha(y,u)\left(\hat{y} - \epsilon \operatorname{sat}\left(\frac{\tilde{y}}{\epsilon}\right)\right) + f(\hat{\theta},\omega) - a^* \operatorname{sat}\left(\frac{\tilde{y}}{\epsilon}\right)\\ \dot{\hat{\theta}} =& -\tilde{y}_\epsilon \phi^*\end{aligned} \quad (4)$$

where

$$\tilde{y} = \hat{y} - y \quad \tilde{y}_\epsilon = \tilde{y} - \epsilon \operatorname{sat}\left(\frac{\tilde{y}}{\epsilon}\right) \quad (5)$$

ϵ is an arbitrary positive number, $\operatorname{sat}(\cdot)$ denotes the saturation function and is given by $\operatorname{sat}(x) = \operatorname{sign}(x)$ if $|x| \geq 1$ and $\operatorname{sat}(x) = x$ if $|x| < 1$, and a^* and ϕ^* come from the solution of an optimization problem

$$\begin{aligned}a^* =& \min_{\phi \in \mathbb{R}^m} \max_{\theta \in \Omega^0} g(\theta,\omega,\phi)\\ \phi^* =& \arg \min_{\phi \in \mathbb{R}^m} \max_{\theta \in \Omega^0} g(\theta,\omega,\phi)\\ g(\theta,\omega,\phi) =& \operatorname{sat}\left(\frac{\tilde{y}}{\epsilon}\right)\\ & \times \left(f(\hat{\theta},\omega) - f(\theta,\omega) - \phi^T(\widehat{\theta} - \theta)\right).\end{aligned} \quad (6)$$

The choices of ϕ^* and a^* imply the following inequality:

$$\operatorname{sat}\left(\frac{\tilde{y}}{\epsilon}\right)\left(f(\hat{\theta},\omega) - f(\theta,\omega) - \phi^{*T}(\widehat{\theta} - \theta)\right) - a^* \leq 0. \quad (7)$$

We define

$$\tilde{\theta} = \hat{\theta} - \theta_0$$

and rewrite the dynamics of the whole parameter estimation algorithm as

$$\begin{aligned}\dot{\tilde{y}} =& -\alpha(y,u)\tilde{y}_\epsilon + f(\hat{\theta},\omega) - f(\theta_0,\omega) - a^* \operatorname{sat}\left(\frac{\tilde{y}}{\epsilon}\right)\\ \dot{\tilde{\theta}} =& -\tilde{y}_\epsilon \phi^*.\end{aligned} \quad (8)$$

Let $x = [\tilde{y}_\epsilon, \tilde{\theta}^T]^T$. The problem is therefore to determine the conditions on ω under which the system (8) has uniform asymptotic stability in the large (u.a.s.l.) at $x = 0$.

B. Solutions of a^* and ϕ^*

In [1] and [6], closed form solutions to (6) when f is a concave/convex function of θ_0 and when f is a general function of θ_0 were derived, respectively. In both [1] and [6], these solutions were derived under the assumption that $\hat{\theta} \in \Omega^0$. In this paper, results are extended to the case when this assumption is omitted. For ease of exposition, we present the results for the cases when a) θ is a scalar, and f is a general function of θ and b) θ is a vector, and f is a convex/concave function of θ. We define a convex set $C(\Omega_0)$ which is constructed as follows: If $H_f(\Omega_0)$ is the convex hull, which is the smallest convex set in $\mathbb{R}^{n+1}$ that contains $\{(\theta, f(\theta,\omega)) | \theta \in \Omega^0\}$, then $C(\Omega_0)$ is the projection of $H_f(\Omega_0)$ on $\mathbb{R}^n$ which contains Ω_0. Such a convex set is needed since i) the hierarchical algorithm discussed in Section IV-C can allow the parameter estimate to wander outside Ω_0, and ii) the solutions to the min–max algorithm differ depending whether $\hat{\theta}$ lies within this convex set $C(\Omega_0)$ or outside.

a) $\theta \in \Omega^0 \subset \mathbb{R}$, and f is a general function of θ: In this case, $C(\Omega^0) = [\theta_{\min}, \theta_{\max}]$. Same as in [6], the following two definitions are useful.

Definition 2: A point $\theta^0 \in \theta_c$ if $\theta^0 \in C(\Omega^0)$ and

$$\nabla f_{\theta^0}(\theta - \theta^0) \leq f(\theta,\omega) - f(\theta^0,\omega) \qquad \forall\, \theta \in C(\Omega^0) \quad (9)$$

where $\nabla f_{\theta^0} = (\partial f/\partial\theta)|_{\theta^0}$.

Definition 3: $\tilde{\theta}_c = \bar{\theta}_c \bigcap C(\omega^0)$, where $\bar{\theta}_c$ denotes the complement of θ_c.

We now state the solutions to (6) in case a), when $\tilde{y} > 0$. The solutions when $\tilde{y} < 0$ can be derived in a similar manner using the concave cover.

IEEE Proof

Denoting $\breve{\theta}_c = \left\{\theta^{12}, \theta^{34}, \ldots, \theta^{mn}\right\}$, $\theta^{ij} = \left[\theta^i, \theta^j\right]$ as in [6], we obtain

$$\left.\begin{array}{ll} \phi^* = \nabla f_{\hat{\theta}} & \\ a^* = 0, & \text{if } \hat{\theta} \in \theta_c \\ \phi^* = \theta^{ij} & \\ a^* = f(\hat{\theta}, \omega) - f(\theta^i, \omega) - \phi^*(\hat{\theta} - \theta^i), & \text{if } \hat{\theta} \in \theta^{ij} \end{array}\right\} \text{if } \hat{\theta} \in C(\Omega^0) \tag{10}$$

and if $\hat{\theta} > \theta_{\max}$, (11), shown at the bottom of the page, holds, and if $\hat{\theta} < \theta_{\min}$,, (12), shown at the bottom of the page, holds.

b) $\theta \in \Omega^0 \subset \mathbb{R}^n$, f is a concave function of θ: The solutions to (6) are easier to find when Ω^0 is a simplex, and are presented first.

Case i): Ω^0 is a simplex: Very similar to [1], we have the following solutions:

$a^* = 0$ $\phi^* = \nabla f_{\hat{\theta}}$	$\tilde{y} < 0,\ \hat{\theta} \in C\left(P\left(\Omega^0\right)\right)$
$a^* = 0$ $\phi^* = \nabla f_{\hat{\theta}}$	$\tilde{y} < 0,\ \hat{\theta} \notin C\left(P\left(\Omega^0\right)\right)$
$a^* = A_1$ $\phi^* = A_2$	$\tilde{y} < 0,\ \hat{\theta} \in C\left(P\left(\Omega^0\right)\right)$
$a^* = 0$ $\phi^* = A_2$	$\tilde{y} < 0,\ \hat{\theta} \notin C\left(P\left(\Omega^0\right)\right)$

where $[A_1, A_2]^T = G^{-1}b$, $A_1 \in \mathbb{R}$, $A_2 \in \mathbb{R}^n$

$$G = \begin{bmatrix} -1 & -(\hat{\theta} - \theta_{S1})^T \\ -1 & -(\hat{\theta} - \theta_{S2})^T \\ \vdots & \vdots \\ -1 & -(\hat{\theta} - \theta_{Sn+1})^T \end{bmatrix}$$

$$b = \begin{bmatrix} -(f(\hat{\theta}, \omega) - f_{S1}) \\ -(f(\hat{\theta}, \omega) - f_{S2}) \\ \vdots \\ -(f(\hat{\theta}, \omega) - f_{Sn+1}) \end{bmatrix}.$$

$\theta_{Si}, i = 1, \ldots, n+1$ are the vertices of Ω^0, and $f_{Si} = f(\theta_{Si}, \omega)$.

Case ii) Ω^0 is a compact set in $\mathbb{R}^n$: We define a polygon $P(\Omega^0)$ which contains Ω^0, whose vertices are given by $P_1, P_2, \ldots, P_K$. Denoting $L = \begin{pmatrix} K \\ n+1 \end{pmatrix}$, we note that L hyperplanes can be constructing using a combination of $n+1$ points from the K vertices of the polygon. Denoting the vertices of the ith hyperplane as $P_{i_1}, \ldots, P_{i_{n+1}}$, and ϕ_i as the slope of this hyperplane, we choose J as a set of the L hyperplanes such that

$$J = \Big\{ i | 1 \le i \le L, f(\hat{\theta}, \omega) - f(P_{i_1}, \omega) - \phi_i^T(\hat{\theta} - P_{i_1}) \ge 0, \ \forall\, \hat{\theta} \in P(\Omega^0) \Big\}.$$

We can derive the solutions to (6) as

$a^* = 0$ $\phi^* = \nabla f_{\hat{\theta}}$	$\tilde{y} < 0,\ \hat{\theta} \in C\left(P\left(\Omega^0\right)\right)$
$a^* = 0$ $\phi^* = \nabla f_{\hat{\theta}}$	$\tilde{y} < 0,\ \hat{\theta} \notin C\left(P\left(\Omega^0\right)\right)$
$a^* = A_1$ $\phi^* = A_2$	$\tilde{y} < 0,\ \hat{\theta} \in C\left(P\left(\Omega^0\right)\right)$
$a^* = 0$ $\phi^* = A_2$	$\tilde{y} < 0,\ \hat{\theta} \notin C\left(P\left(\Omega^0\right)\right)$

where

$$\begin{aligned} A_2 &= \phi_j \\ A_1 &= f(\hat{\theta}, \omega) - f(P_{j_1}, \omega) - \phi_j^T(\hat{\theta} - P_{j_1}) \\ j &= \arg\max_{i \in J} f(P_{i_1}, \omega) + \phi_i^T(\hat{\theta} - P_{i_1}). \end{aligned}$$

The solutions for the case when f is a convex function of θ can be derived in a similar manner.

c) $\theta \in \Omega^0 \subset \mathbb{R}^n$, f is a general function of θ: Using the above two cases, and in particular, a combination of concave and convex covers, convex hull, and polygons, the solutions to (6) can be found.

C. Properties of the Min–Max Estimator

In [1], the min–max estimator and therefore the resulting error model in (8) was shown to be stable. The stability properties of this error model are summarized in Properties 1 and 2 below. In what follows, the quadratic function V is useful:

$$V = \frac{1}{2}\left(\tilde{y}_\epsilon^2 + \tilde{\theta}^T\tilde{\theta}\right) = \frac{1}{2}\|x(t)\|^2. \tag{13}$$

Property 1 summarizes the stability properties of (8).

$$\begin{aligned} a* &= 0 \\ \phi^* &= \begin{cases} \phi^*(\theta_{\max}), & \text{if } f(\theta_{\max}, \omega) + \phi^*(\theta_{\max})(\hat{\theta} - \theta_{\max}) \ge f(\hat{\theta}, \omega) \\ \frac{f(\theta_{\max}, \omega) - f(\hat{\theta}, \omega)}{\theta_{\max} - \hat{\theta}}, & \text{otherwise} \end{cases} \end{aligned} \tag{11}$$

$$\begin{aligned} a* &= 0 \\ \phi^* &= \begin{cases} \phi^*(\theta_{\min}), & \text{if } f(\theta_{\min}, \omega) + \phi^*(\theta_{\min})(\hat{\theta} - \theta_{\min}) \ge f(\hat{\theta}, \omega) \\ \frac{f(\theta_{\min}, \omega) - f(\hat{\theta}, \omega)}{\theta_{\min} - \hat{\theta}}, & \text{otherwise} \end{cases} \end{aligned} \tag{12}$$

IEEE Proof

Property 1:

$$\dot{v} \leq -\alpha \tilde{y}_\epsilon^2. \tag{14}$$

Property 1 implies that the min–max estimator is stable. However, whether the parameter estimates will converge to their true values, that is, whether x will converge to the origin is yet to be established. To facilitate parameter convergence discussions, an additional property of the min–max estimator is stated in Property 2.

Property 2: If in (8)

$$|\tilde{y}_\epsilon(t_1)| \geq \gamma \tag{15}$$

then

$$V(t_1 + T') \leq V(t_1) - \frac{\alpha_{\min}\gamma^3}{3(M + \alpha_{\max}\gamma)} \tag{16}$$

where $T' = (\gamma/M + \alpha_{\max}\gamma)$, $0 < \alpha_{\min} \leq \alpha(y, u) \leq \alpha_{\max}$, and

$$\begin{aligned} M &= \max\{|m(t)|\} \\ m(t) &= f(\hat{\theta}, \omega) - f(\theta_0, \omega) - a^* \mathrm{sat}\left(\frac{\tilde{y}_\epsilon}{\epsilon}\right). \end{aligned} \tag{17}$$

Property 2 implies that for parameter convergence to occur, $\tilde{y}_\epsilon$ must become periodically large. For this in turn to occur, examining the dynamics in (8) and defining $\tilde{f}(\hat{\theta}, \theta_0, \omega) = f(\hat{\theta}, \omega) - f(\theta_0, \omega)$, i) $\tilde{f}(\hat{\theta}, \theta_0, \omega)$ must be large when $\|\tilde{\theta}\|$ is large and ii) a^* must be small compared to $\tilde{f}(\hat{\theta}, \theta_0, \omega)$. Condition i) is related to persistent excitation, and is similar to parameter convergence conditions in linearly parameterized systems. Condition ii) is specific to the min–max algorithm. In order to facilitate the latter, a few properties of a^* are worth deriving, and are enumerated below in Properties 3 and 4.

Noting that a^* is defined as in (6), we denote

$$\begin{aligned} a_-^*(\hat{\theta}, \omega) &= a^* \text{ if } \tilde{y}_\epsilon < 0 \\ a_+^*(\hat{\theta}, \omega) &= a^* \text{ if } \tilde{y}_\epsilon > 0. \end{aligned}$$

It follows that a_-^* and a_+^* are well defined functions of $\hat{\theta}$ and ω. We establish the following properties about a^*, $a_-^*(\hat{\theta}, \omega)$ and $a_+^*(\hat{\theta}, \omega)$.

Property 3:

$$1 a_-^*(\hat{\theta}, \omega) \leq a^* \mathrm{sat}\left(\frac{\tilde{y}}{\epsilon}\right) \leq a_+^*(\hat{\theta}, \omega).$$

Suppose for a given ω, $f(\theta, \omega)$ retains its curvature as θ varies. We define

$$\beta(\omega) = \begin{cases} 1, & \text{if } f(\theta, \omega) \text{ is convex} \\ -1, & \text{if } f(\theta, \omega) \text{ is concave} \end{cases}. \tag{18}$$

Property 4: For a^* and β defined as in (6) and (18), respectively, the following holds:

$$\begin{aligned} &\text{i) } a_-^* = 0 \text{ if } \beta = -1; \quad \text{ii) } a_+^* = 0 \text{ if } \beta = 1; \\ &\text{(iii) } \beta a^* \tilde{y} \leq 0, \text{ for any } \beta. \end{aligned} \tag{19}$$

Both Properties 3 and 4 are used in Section III for the proof of Theorem 1.

III. Parameter Convergence in Systems With Convex/Concave Parameterization

We first focus on parameter convergence of the system (8) when f is convex/concave for any $\theta \in \Omega^0$. For the sake of completeness, we include the definition of a concave/convex function.

Definition 4: A function $f(\theta)$ is said to be i) convex on Θ if it satisfies the inequality

$$f(\lambda\theta_1 + (1-\lambda)\theta_2) \leq \lambda f(\theta_1) + (1-\lambda) f(\theta_2) \quad \forall\, \theta_1, \theta_2 \in \Theta$$

and ii) concave if it satisfies the inequality

$$f(\lambda\theta_1 + (1-\lambda)\theta_2) \geq \lambda f(\theta_1) + (1-\lambda) f(\theta_2) \quad \forall\, \theta_1, \theta_2 \in \Theta$$

where $0 \leq \lambda \leq 1$.

We make a few qualitative comments regarding the solutions of (8) and their convergence properties before establishing the main result. The main difficulty in establishing parameter convergence is due to the presence of the time-varying function a^* in (8). As shown in Properties 3–4 in Section II-C, the magnitude of a^* changes with the curvature of f. As mentioned in Section II-C, in order to establish parameter convergence, in addition to $\tilde{f}(\hat{\theta}, \theta_0, \omega)$ being large when $\tilde{\theta}$ is large a^* has to remain small. Property 3 shows that for any nonzero value of $\tilde{y}$, a^* can periodically take the value zero if f switches periodically between concavity and convexity. This in turn implies that a^* can periodically become small if f continues to change its curvature, that is, β changes from $+1$ to -1. As will be shown in Section III-A, the conditions for parameter convergence not only require that $\tilde{f}$ become large for a large $\tilde{\theta}$ but also require f to switch between convexity and concavity over any given interval.

Yet another feature of the min–max algorithm is the use of the error $\tilde{y}_\epsilon$ for adjusting the parameter $\hat{\theta}$ instead of the traditional estimation error $\tilde{y}$. This was introduced in the estimation algorithm to ensure a continuous estimator in the presence of a discontinuous solution that can be obtained from the min–max optimization problem. The introduction of a nonzero ϵ can cause the parameter estimation to stop if $|\tilde{y}|$ becomes smaller than ϵ. As a result, the trajectories are shown to converge to a neighborhood D_ϵ of the origin rather the origin itself.

In Section III-A, we state and prove the convergence result. In Section III-B, we discuss the sufficient condition that results in parameter convergence, specific examples of f and counterexamples, and the relation to persistent excitation conditions that guarantee parameter convergence in the case of linear parameterization.

A. Proof of Convergence

The first convergence result in this paper is stated in Theorem 1.

Theorem 1: If i) $f(\theta, \omega(t))$ is convex (or concave) on θ for any $\omega(t) \in \mathbb{R}^m$, and ii) for every $t_1 > t_0$, there exist positive constants T_0, ϵ_u and a time instant $t_2 \in [t_1, t_1 + T_0]$ such that for any θ

$$\beta(\omega(t_2))\left(f(\theta, \omega(t_2)) - f(\theta_0, \omega(t_2))\right) \geq \epsilon_u \|\theta - \theta_0\| \tag{20}$$

IEEE Proof

where $\beta(\omega(t_2))$ is defined as in (18), then all trajectories of (8) will converge uniformly to

$$D_\epsilon = \{x | V(x) \le \gamma_1\} \tag{21}$$

where

$$x = [\tilde{y}_\epsilon \quad \tilde{\theta}^T]^T, \gamma_1 = \frac{2\epsilon}{\epsilon_u^2}\left(16B_\theta U_b + 8B_\theta B_\phi + 4B_\phi^2\right) \tag{22}$$

ϵ is defined as in (8), ϵ_u is given by (20), U_b and B_θ are defined as in Assumptions 1 and 2, and B_ϕ is the bound on ϕ^* in (4) so that

$$\|\phi^*(t)\| \le B_\phi \qquad \forall\, t \ge t_0. \tag{23}$$

The proof of Theorem 1 follows by showing that if ω and f are such that condition (20) is satisfied, then $\tilde{y}_\epsilon(t)$ becomes large at some time t over the interval $[t_1, t_1 + T_0]$. Once $\tilde{y}_\epsilon(t)$ becomes large, it follows from Property 2 that $V(t)$ decreases over the interval $[t_1, t_1 + T_0]$ by a finite amount.

Remark 1: If f is concave (or convex) and if f satisifes the inequality in (20), we shall define that f satisfies the convex persistent excitation (CPE) condition with respect to ω. Theorem 1 implies that if f satisfies the CPE condition with respect to ω, then parameter convergence to a desired precision follows.

Remark 2: From the definition of D_ϵ, it automatically follows that as $\epsilon \to 0$, all trajectories converge to the region $x = 0$ and hence u.a.s.l. follows.

B. Sufficient Condition for Parameter Convergence

The CPE condition specifies certain requirements on f in order to achieve parameter convergence. For a given f, Theorem 1 does not state how ω should behave over time in order to satisfy (20). In this section, we state some observations and examples of ω that satisfies (20) for a general f.

Equation (20) consists of two separate requirements. Denoting $\tilde{f} = f(\theta, \omega) - f(\theta_0, \omega)$, the first requirement is that the magnitude of $\tilde{f}$ must be large. The second requirement is that $\tilde{f}$ must have the same sign as β. The first component states that for a large parameter error, there must be a large error in $\tilde{f}$. It is straightforward to demonstrate that this condition is equivalent to linear persistent excitation condition in [10], and is shown in Section III-B.2. The second requirement states what the sign of $\tilde{f}$ should be in relation to the convexity/concavity of f. If f is convex, $\widetilde{f}$ should be positive, and conversely, if f is concave, $\widetilde{f}$ should be negative.

The coupling of convexity/concavity and the sign of the integral of $\widetilde{f}$ has the following practical implications. To ensure parameter convergence, ω must be such that one of the following occurs: At least at one instant $t_2 \in [t_1, t_1 + T]$.

a) For the given $\widetilde{\theta}$, ω must change in such a way that the sign of $\widetilde{f}$ is reversed, while keeping the convexity/concavity of f the same.
b) Or, for the given $\widetilde{\theta}$, ω must reverse the convexity/concavity of f, while preserving the sign of $\widetilde{f}$

1) Examples: We illustrate the aforementioned comments using specific examples of f. Suppose

$$f = e^{-\omega^T\theta} \tag{24}$$

where $\omega(t) : \mathbb{R} \to \mathbb{R}^n$, $\theta \in \Omega \subset \mathbb{R}^n$. It can be checked that f given in (24) is always convex with respect to θ for all ω. Therefore, option *b)* is not possible. Hence, ω must be such that $\widetilde{f}$ can switch sign for any $\widetilde{\theta}$ as required by option *(a)*. One example of such an ω is if for any t_1, there exists $t_2 \in [t_1, t_1 + T]$ such that

$$\omega^T(t_2)v \ge \epsilon_u \tag{25}$$

where v is any unit vector in $\mathbb{R}^n$. Another example which satisfies condition (20) is given by

$$f = \theta^\omega, \qquad \omega \in \mathbb{R}.$$

It is easy to show that for such an f, condition b) is satisfied if ω switches between ω_1 and ω_2 where $0 < \omega_1 < 1$ and $\omega_2 > 1$.

The previous examples show that the condition on ω that satisfies (20) varies with f.

2) Relation to Conditions of Linear Persistent Excitation : The relation between CPE and LPE is worth exploring. For this purpose, we consider a linearly parameterized system, which is given by (1) with

$$f(\theta_0, \omega) = \theta_0^T\phi(\omega)$$

where $\phi(\omega) \in \mathbb{R}^n$. In this case, it is well known that the corresponding estimator is given by (4) with $a^* = 0$ and $\phi^* = \phi$ [9]. The resulting error equations are summarized by

$$\begin{aligned} \dot{\tilde{y}} &= -\alpha(y, u)\tilde{y}_\epsilon + \tilde{\theta}^T\phi(\omega) \\ \dot{\tilde{\theta}} &= -\tilde{y}_\epsilon\phi(\omega). \end{aligned} \tag{26}$$

In [10], it is shown that u.a.s.l. of (26) follows under an LPE condition. For the sake of completeness, we state this condition now.

Definition (LPE): ϕ is said to be linearly persistently exciting (l.p.e.) if for every $t_1 > t_0$, there exists positive constants T_0, δ_0, ϵ_0 and a subinterval $[t_2, t_2 + \delta_0] \in [t_1, t_1 + T_0]$ such that

$$\left| \int_{t_2}^{t_2+\delta_0} \theta^T\phi(\omega(\tau))d\tau \right| \ge \epsilon_0\|\theta\|. \tag{27}$$

We now show the relation between the LPE condition and the CPE condition in (20). When $f(\theta, \omega) = \theta^T\phi(\omega)$, if Assumption 1 holds, it can be shown that the LPE condition is equivalent to the following inequality: For every $t_1 > t_0$ and θ, there exists positive constants T_0, ϵ_0 and a time instant $t_2 \in [t_1, t_1 + T_0]$ such that

$$|f(\theta, \omega(t_2)) - f(\theta_0, \omega(t_2))| \ge \epsilon_u\|\theta - \theta_0\|. \tag{28}$$

Since a linear function can be considered to be either convex or concave, the inequality in (28) is equivalent to the CPE condition in (20). This equivalence is summarized in the following lemma:

Lemma 1: When $f(\theta, \omega) = \theta^T\phi(\omega)$, if Assumption 1 holds, the CPE condition in (20) is equivalent to the LPE condition in (27).

It should be noted that for a general nonlinear f, the CPE condition becomes more restrictive than the LPE condition. For example, for f as in (24), the CPE condition implies that ω must

IEEE Proof

satisfy (25). On the other hand, if $f=\omega^T\theta$, even if ω is such that $|\omega^T(t_2)v|$ is periodically large, the LPE condition is satisfied.

3) Counterexample: For a general function f, it may not be possible to find a ω that satisfies either condition a) or b) previously mentioned. A simple example is

$$f=\cos(\theta\omega)$$

where $|\omega|\le\omega_{\max}$ and $\theta\in[0,\pi/(2\omega_{\max})]$. We note that f is concave and monotonically decreasing for any ω with $|\omega|\le\omega_{\max}$. Hence, neither a) nor b) is satisfied. That is, it is possible for the parameter estimate $\hat{\theta}$ of the min–max algorithm to get "stalled" in a region in Ω^0. This motivates the need for an improved min–max algorithm, and is outlined in Section IV.

IV. Parameter Convergence in Systems With a General Parameterization

In Section III, we showed that if a function f is convex (or concave), and if f and ω satisfy the CPE condition, then parameter convergence follows. However, as we saw in Section III-B-III, not all convex/concave functions can satisfy the CPE condition. In this section, we present a new algorithm which not only allows the persistent excitation condition to be relaxed but also enables parameter convergence for nonconvex and nonconcave functions.

The algorithm we present in this section is hierarchical in nature, and consists of a lower level and a higher level. In the lower level, for a given unknown parameter region Ω^0, the parameter estimate $\hat{\theta}$ is updated using the min–max algorithm as in (4). In the higher level, using information regarding the parameter estimate $\hat{\theta}$ obtained from the lower level, the unknown parameter region is updated as Ω^1. Iterating between the lower and higher levels, the overall hierarchical algorithm guarantees a sequence of parameter region Ω^k. The properties of these two levels are discussed in Sections IV-A and B, respectively. In Section IV-D, we discuss conditions under which $\hat{\theta}$ converges to θ_0. Using these conditions, the definition of persistent excitation for nonlinearly parameterized systems is introduced. In Section IV-E, we present examples of such an NLPE. The relation between NLPE and CPE is discussed in Section IV-F.

A. Lower Level Algorithm

The lower level algorithm consists of the min–max parameter estimation as in (4) with the unknown parameter $\theta_0\in\Omega^k$. We show in this section that when this algorithm is used, $\tilde{y}_\epsilon(t)$ becomes small in a finite time, which is denoted as lower level convergence. Once this occurs, the parameter estimate $\hat{\theta}$, derived from the lower level convergence, remains nearly steady. This estimate, in turn, is used in the higher level part of hierarchical algorithm to update the unknown parameter region from Ω^k to Ω^{k+1}. The convergence of $\tilde{y}_\epsilon$ is stated in Lemma 2, and the characterization of the unknown parameter is stated in Lemma 3.

Lemma 2: For the system in (1) and the estimator in (4), given any positive T and δ, there exists a finite time t_1 such that

$$|\tilde{y}_\epsilon(t)|\le\delta \text{ for } t_1\le t\le t_1+T. \tag{29}$$

We note that for every specific ω, a time t_1 that satisfies (29) exists. However, the value of t_1 will depend on the choice of ω. Since our goal is parameter convergence, we require ω to assume distinct values, i.e., persistently span a set of interest. This is stated in the definition later.

Let U_I be defined as in (2).

Definition 5: ω is said to persistently span U_I if for any $\omega_i\in U_I$ and any t_1, there exist a finite T_i and τ_i such that

$$\omega(\tau_i)=\omega_i\quad \tau_i\in[t_1,t_1+T_i]\quad i=1,\dots I. \tag{30}$$

Definition 5 implies that ω periodically visits all points in U_I.

Let

$$B_t=2B_\theta(\delta B_\phi+2U_b)+\delta B_\phi^2 \tag{31}$$

where U_b, B_θ and B_ϕ are defined in Assumption 1, Assumption 2 and (23) and δ is any positive number. If we choose T as

$$T=\max_{1\le i\le I}T_i+\frac{2\sqrt{B_t(\delta+\epsilon)}}{B_t}$$

where T_i is given by (30), then Lemma 2 implies that there exists a finite time t_1 such that

$$|\tilde{y}_\epsilon(t)|\le\delta\quad t_1\le t\le t_1+T. \tag{32}$$

When $\tilde{y}_\epsilon$ satisfies (32), we refer to it as lower level convergence. If ω persistently spans U_I, then Definition 5 and the choice of τ_i implies that at $\tau_i\in[t_1,t_1+T]$, $\omega(\tau_i)=\omega_i$, $i=1,\dots,I$. The parameter estimate $\hat{\theta}(\tau_i)$ at time instances are defined as

$$\hat{\theta}_i^c=\hat{\theta}(\tau_i)\quad i=1,\dots,I$$

and are denoted as low-level convergent estimates. We characterize the region where the unknown parameters lie in lLemma 3 using these lower level convergent estimates.

Lemma 3: For the system in (1) and the estimator in (4), let Ω be the unknown parameter region and $\hat{\theta}_i^c, i=1,\dots,I$, be the lower level convergent estimates. If the input ω persistently spans U_I, then

$$\theta_0\in\bigcap_{i=1}^{I}\Phi_\epsilon(\Omega,\omega_i,\epsilon,\delta,\hat{\theta}_i^c).$$

where

$$\begin{aligned}\Phi_\epsilon(\Omega,\omega_i,\epsilon,\delta,\hat{\theta}_i^c)&=\left\{\theta\in\Omega|\underline{f}_i\le f(\theta,\omega_i)\le\bar{f}_i\right\}\\ \underline{f}_i&=f(\hat{\theta}_i^c,\omega_i)r-a_+^*(\hat{\theta}_i^c,\omega)\\ &\quad-\alpha_{\max}\delta-2\sqrt{B_t(\delta+\epsilon)}\\ \bar{f}_i&=f(\hat{\theta}_i^c,\omega_i)+a_-^*(\hat{\theta}_i^c,\omega)\\ &\quad+\alpha_{\max}\delta+2\sqrt{B_t(\delta+\epsilon)}\end{aligned} \tag{33}$$

and B_t as in (31).

Lemma 3 implies that the unknown parameter θ_0 lies in Φ_ϵ for a given ω_i. It should be noted that in general, Φ_ϵ need not be smaller than Ω. However other properties of Φ_ϵ are useful for characterizing the convergence behavior of the min–max algorithm. These are enumerated as follows.

IEEE Proof

P1) For $\delta = \epsilon = 0$, if $a_+^* = a_-^* = 0$, then Φ_ϵ reduces to the manifold

$$f(\theta_0, \omega_i) = f(\hat{\theta}_i^c, \omega_i).$$

P2) Property P3) implies that if i) ω is p.e. in U_I, ii) f is identifable w.r.t. U_I, iii) $\delta = \epsilon = 0$, and iv) $a_+^* = a_-^* = 0$, then

$$\bigcap_{i=1}^{I} \Phi_\epsilon(\Omega^k, \omega_i, \epsilon, \delta, \hat{\theta}_i^c) = \{\theta_0\}.$$

These properties are made judicious use of in designing the higher level algorithm in Section IV-B.

B. Higher Level Algorithm

We now present the higher level component of the hierarchical algorithm. Here, our goal is to start from a known parameter region Ω^k that the unknown parameter θ_0 lies in, and update it as Ω^{k+1} using all available information from the lower level component. In particular, we use Φ_ϵ defined in (33) to update Ω^k. In order to reduce the parameter uncertainty, different Φ_ϵ's are computed by varying ω_i, $i = 1, \ldots, I$. The resulting Ω^{k+1} is, therefore, chosen as

$$\Omega^{k+1} = \bigcap_{i=1}^{I} \Phi_\epsilon(\Omega^k, \omega_i, \epsilon, \delta, \hat{\theta}_i^c). \quad (34)$$

C. Hierarchical Algorithm

The complete hierarchical algorithm is stated in Table I.

It should be noted that Steps 2) and 3) correspond to the lower level and the higher level parts of the hierarchical algorithm, respectively. Also, we note that Step 2) requires the closed-form solutions of a^* and ϕ^* which can be found as outlined in Section II-A.

In order to obtain parameter convergence using the hierarchical algorithm, what remains to be shown is whether Ω^{k+1} is a strict subset of Ω^k.

D. Parameter Convergence With the Hierarchical Algorithm

We now address the parameter convergence of the hierarchical algorithm. We introduce a definition for a "stalled" parameter region Δ_i:

For any $\Omega \subseteq \Omega^0$, define $\underline{f}_i^*$ and $\bar{f}_i^*$ as

$$\underline{f}_i^* = \min_{\theta \in \Omega} f(\theta, \omega_i) \quad \bar{f}_i^* = \max_{\theta \in \Omega} f(\theta, \omega_i). \quad (35)$$

Then, we define $\Delta_i(\Omega)$ to be a "stalled" estimate-region of Ω as

$$\Delta_i(\Omega) = \Big\{\theta | f(\theta, \omega_i) - a_+^*(\theta, \omega_i) - D(\epsilon, \delta) \le \underline{f}_i^* \text{ and } f(\theta, \omega_i) + a_-^*(\theta, \omega_i) + D(\epsilon, \delta) \ge \bar{f}_i^* \Big\} \quad (36)$$

where

$$D(\epsilon, \delta) = \alpha_{\max}\delta + 2\sqrt{B_t(\delta + \epsilon)}. \quad (37)$$

TABLE I
COMPLETE HIERARCHICAL ALGORITHM

Step 1:	Set $k = 0$ and $\Omega^k = \Omega^0$, and $T = \max_i T_i + 2\sqrt{B_t(\delta + \epsilon)}/B_t$ where B_t is defined as in (31).
Step 2:	Run the estimator in (4) where $a^* = \min_{\phi \in \mathbb{R}^m} \max_{\theta \in \Omega^k} g(\theta, \omega, \phi)$ $\phi^* = arg \min_{\phi \in \mathbb{R}^m} \max_{\theta \in \Omega^k} g(\theta, \omega, \phi)$ $g(\theta, \omega, \phi) =$ $sat\left(\frac{\tilde{y}}{\epsilon}\right)\left(f(\hat{\theta}, \omega) - f(\theta, \omega) - \phi^T(\hat{\theta} - \theta)\right)$. Wait until time t_k^* where $\lvert\tilde{y}_\epsilon(t)\rvert \le \delta \quad$ for $t \in [t_k^*,\ t_k^* + T]$, and record the low level convergent estimate $\hat{\theta}_i^c$ as $\hat{\theta}_i^c = \hat{\theta}(\tau_i)$ where $\omega(\tau_i) = \omega_i, \quad \forall\ \tau_i \in [t_k^*,\ t_k^* + T]$.
Step 3:	Calculate Ω^{k+1} from Ω^k and $\hat{\theta}_i^c, i = 1, \ldots I$, as follows: $\Omega^{k+1} = \bigcap_{i=1}^{I} \Phi_\epsilon(\Omega^k, \omega_i, \epsilon, \delta, \hat{\theta}_i^c)$ $\Phi_\epsilon(\Omega, \omega_i, \epsilon, \delta, \hat{\theta}_i^c) = \left\{\theta \in \Omega \lvert \underline{f}_i \le f(\theta, \omega_i) \le \bar{f}_i\right\}$ $\underline{f}_i = f(\hat{\theta}_i^c, \omega_i) - a_+^*(\hat{\theta}_i^c, \omega) - Q(\delta, \epsilon)$ $\bar{f}_i = f(\hat{\theta}_i^c, \omega_i) + a_-^*(\hat{\theta}_i^c, \omega) + Q(\delta, \epsilon)$ where $Q(\delta, \epsilon) = \alpha_{max}\delta + 2\sqrt{B_t(\delta + \epsilon)}$.
Step 4:	If $\Omega^{k+1} = \Omega^k$, stop. Otherwise, set $k = k + 1$ and return to step 2.

We prove a property of $\Delta_i(\Omega)$ which explains why it corresponds to a "stalled" region in Ω.

Lemma 4: For some k, if $\hat{\theta}_i^c \in \Delta_i(\Omega^k)$, $\forall\, i = 1, \ldots, I$, then

$$\Omega^{k+1} = \Omega^k.$$

In order to establish parameter convergence, we first characterize the region L that the parameter estimate converges to in Lemma 5, and then establish the conditions under which L simply coincides with the true parameter θ_0 in Theorem 2. The set L is defined as follows:

$$L(\Omega^0, \epsilon, \delta) = \bigcup_{\forall\ \Omega \in \Omega^0, \theta_0 \in \Omega} \left(\bigcap_{i=1,\ldots,I} B(\Delta_i(\Omega)) \right) \quad (38)$$

where $B(X)$ is a box that contains any set X and is defined as

$$B(X) = \{\theta | \|\theta - \bar{\theta}\| \le \delta T B_\phi,\ \forall\ \bar{\theta} \in X\}. \quad (39)$$

Lemma 5: For the system in (1) and estimator in (4), under Assumptions 1–4, the hierarchical algorithm outlined in Table I guarantees that

$$\lim_{t \to \infty} \hat{\theta}(t) \in L(\Omega^0, \epsilon, \delta). \quad (40)$$

Since ϵ and δ are arbitrary positive numbers, they can be chosen to be as small as possible. When $\epsilon \to 0$, $\delta \to 0$, it follows directly from (36), (39), and (38) that

$$\lim_{\epsilon \to 0, \delta \to 0} L(\Omega^0, \epsilon, \delta) = \bigcup_{\forall\ \Omega \in \Omega^0, \text{ with } \theta_0 \in \Omega} \ \bigcap_{i=1,\ldots,I} \bar{\Delta}_i(\Omega) \quad (41)$$

IEEE Proof

where

$$\bar{\Delta}_i(\Omega) = \Big\{\theta | f(\theta,\omega_i) - a_+^*(\theta,\omega_i) \le \underline{f}_i^* \\ \text{and } f(\theta,\omega_i) + a_-^*(\theta,\omega_i) \ge \bar{f}_i^* \Big\}. \quad (42)$$

with $\underline{f}_i^*$ and $\bar{f}_i^*$ defined as in (35). From (41), we have the following theorem.

Theorem 2: For the system in (1) and the estimator in (4), under Assumptions 1–4

$$\lim_{t\to\infty,\epsilon\to 0,\delta\to 0} \hat{\theta} = \theta_0 \quad (43)$$

if and only if for any $\Omega \subseteq \Omega_0$ where $\theta_0 \in \Omega$

$$\bigcap_{i=1,\ldots,I} \bar{\Delta}_i(\Omega) = \phi \text{ or } \bigcap_{i=1,\ldots,I} \bar{\Delta}_i(\Omega) = \{\theta_0\}. \quad (44)$$

where ϕ denotes the null set and $\bar{\Delta}$ is defined as in (42).

Theorem 2 gives us a method to check if the hierarchical algorithm can estimate the true parameters to any desired precision when we set ϵ and δ small enough for a specific problem. We note that $L(\theta_0,\epsilon,\delta)$ is a continuous function of ϵ and δ, and that as ϵ and δ becomes small, L becomes arbitrarily close to the set $\{\theta_0\}$. Hence, the parameter estimate converges to the true value with a desired precision.

Remark 3: If $f(\theta,\omega)$ is identifiable over Ω with respect to U_I, ω persistently spans U_I, and f satisfies the inequality (44), we shall define that f satisfies the NLPE condition with respect to ω. Theorem 2 implies that NLPE of f with respect to ω is necessary and sufficient for parameter convergence to take place.

Remark 4: The requirement on ω for f to satisfy the NLPE can sometimes be less stringent than that on ω for LPE. An example of this fact is for the parameter $\theta = [\theta_1,\theta_2]^T$, and the cases (i) $l(\theta) = \theta^T\omega$, and (ii) $f(\theta) = \theta_1\omega_1\cos(\theta_2\omega_2)$ where ω_1 and ω_2 are the elements of ω. Clearly, for a ω such that $\omega_1 = k\omega_2$, where k is a constant, ω does not satisfy LPE, but f does satsify NLPE with respect to ω. As shown in Section III-B-I, NLPE can impose more stringent conditions on ω as well.

Remark 5: It should be noted that the NLPE condition guarantees parameter convergence for any general nonlinear function f that is identifiable. This implies that the min–max algorithm outlined in [6], which is applicable for even a nonconvex (or a nonconcave) function, can be used to establish parameter convergence. We include simulation results of such an example in Section V.

Remark 6: It should be noted that a fairly extensive treatment of conditions of persistent excitation has been carried out in [11], [12]for a class of nonlinear systems. The systems under consideration in this paper do not belong to this class. The most distinct features of the system (1) is the presence of the quantity a^* and the quantity $f(\hat{\theta},\omega) - f(\theta_0,\omega)$, where the former can introduce equilibrium points other than zero and the latter is not Lipschitz with respect to $\hat{\theta} - \theta$. As a result, an entirely different set of conditions and properties have had to be derived to establish parameter convergence.

Remark 7: The closed-form solutions of a^* and ϕ^* can be calculated as shown in Section II-A. It should be noted that these solutions have been derived without requiring that $\hat{\theta} \in \Omega^k$, thereby expanding the results of [1]. Since $\hat{\theta}$ can lie anywhere, subsequent iterations of the hierarchical algorithm can be carried out during which time the corresponding min–max solutions can be derived.

As is evident from (44), (35), and (42), to check if indeed the NLPE condition is satisfied for every $\Omega \subseteq \Omega_0$ for a given f and ω is a difficult task. In Section IV-E, we show that when $\theta \in \mathbb{R}^2$, if f is monotonic function of θ, identifiable with respect to U_I, and f is convex/concave, then the NLPE condition is satisfied.

E. Parameter Convergence When $\theta \in \mathbb{R}^2$: An Example

When $\theta = [\theta_1,\theta_2] \in \mathbb{R}^2$, the following lemma provides sufficient conditions for (44) to hold and, hence, for the hierarchical algorithm to guarantee convergence.

Lemma 6: For system in (1), the estimator in (4) where $\theta \in \mathbb{R}^2$, let $\theta_0 \in \Omega^0$ and f be identifiable over Ω^0 with respect to U_2. If

i) $f(\theta,\omega_i)$ is convex (or concave) over all θ in Ω^0

$$\omega_1,\omega_2 \in U_I; \quad (45)$$

ii) $f(\theta,\omega_i)$ is monotonic with respect to θ in Ω^0

$$\omega_1,\omega_2 \in U_I; \quad (46)$$

then (44) holds for any $\Omega \subseteq \Omega^0$ where $\theta_0 \in \Omega$.

The reader is refered to [13] for the proof.

F. Relation Between NLPE and CPE

In what follows, we compare the NLPE and the CPE conditions. In order to facilitate this comparison, we restate the CPE condition in a simpler form.

Definition 6: f is said to satisfy the CPE' condition with respect to ω if i) $f(\theta,\omega(t))$ is convex (or concave) for any $\omega(t) \in \mathbb{R}^m$, and ii) ω is persistently spanning with respect to U_I, and (iii) for any θ, there exists $\omega_i \in U_I$ such that

$$\beta(\omega_i)\left(f(\theta,\omega_i) - f(\theta_0,\omega_i)\right) \ge \epsilon_u\|\theta - \theta_0\|. \quad (47)$$

We note that the only distinction between the inequalities in (20) and (47) is in the value taken by $\omega(t_2)$ for some t_2 in the interval $[t, t+T]$. In (47) it implies that $\omega(t_2)$ assumes one of the finite values ω_i in U_I while in (20), the corresponding U_I can consist of infinite values. If ω is "ergodic" in nature so that it visits all typical values that it will assume for all t over one interval, then it implies that the two conditions (20) and (47) are equivalent. We shall assume in the following that the input is "ergodic."

Lemma 7: Let $f(\theta,\omega_i)$ be convex (or concave) for all $\theta \in \Omega^0$, then the CPE' condition implies the NLPE condition.

Remark 8: Lemma 7 shows that the CPE' condition is sufficient for the NLPE to hold if f is convex (or concave). Clearly, the CPE' condition is not necessary, as shown by the counterexample in Section III-B-III. The NLPE condition therefore represents the most general definition of persistent excitation in nonlinearly parameterized systems.

IEEE Proof

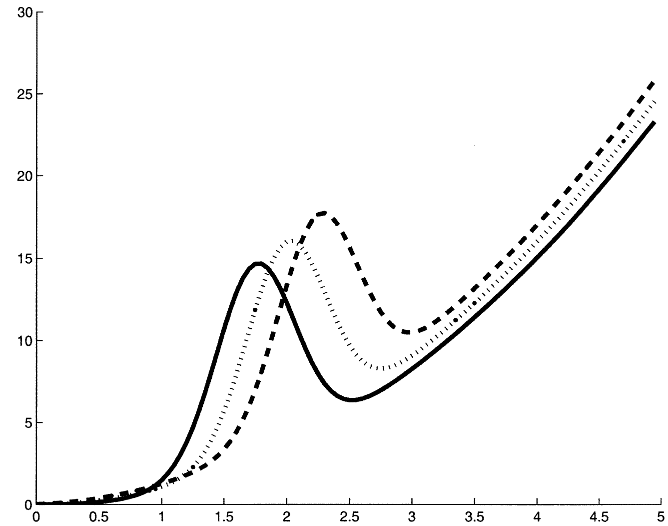

Fig. 1. Nonconcave (and nonconvex) function $f(\theta, u)$ versus θ, for $u = 1, 0, -1$. $f(\theta, 1)$:—, $f(\theta, -1)$: - - -, $f(\theta, 0)$:.

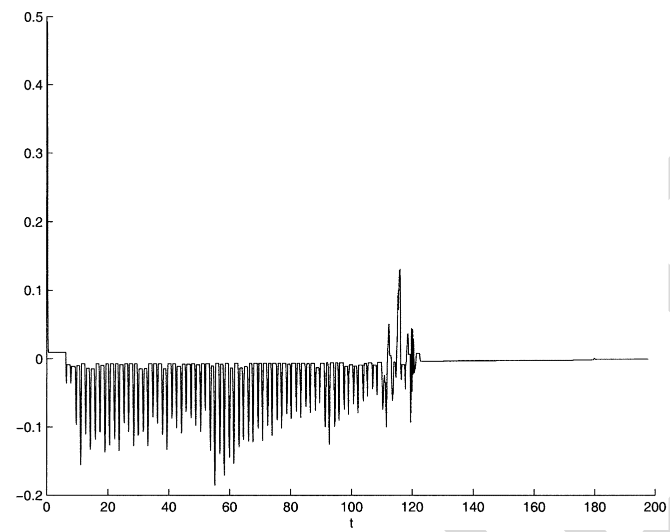

Fig. 2. Output error $\tilde{y}_\epsilon(t)$ with t using the hierarchical algorithm. $\epsilon = 0.001$ and $\delta = 0.02$.

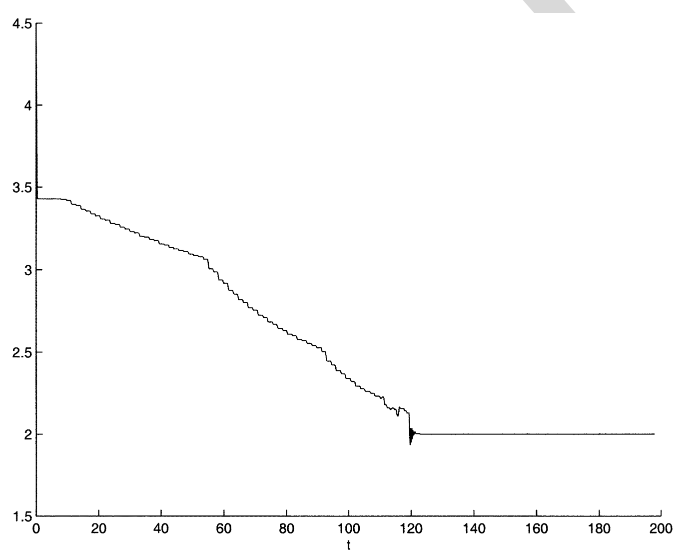

Fig. 3. Parameter estimate $\hat{\theta}(t)$ with t using the hierarchical algorithm. True parameter value $\theta_0 = 2$.

V. Simulation Results

We consider the system in (1) and the estimator in (4) to evaluate the performance of the hierarchical algorithm. The system parameters are chosen as follows:

$$f = \left(\theta_0 - \frac{\omega}{8}\right)^2 + 12 \exp\left\{-5\left(\theta_0 - 2 + \frac{\omega}{4}\right)^2\right\}$$

where θ_0 is an unknown parameter that belongs to a known interval $\Omega^0 = [0, 5]$. System variable ω is chosen as a sinusoidal function $\omega = 1.1\sin(2t)$ and the true unknown parameter θ_0 equals 2. We note that the function f is nonconvex (and nonconcave), whose values are shown in Fig. 1 for $\omega = 1, -1, 0$. It can be shown that f is identifable with respect to Ω^0 and that ω is persistently spanning with respect to $U_I = \{1, -1, 0\}$. The hierarchical algorithm in Table I was implemented to estimate θ_0. The parameters $\epsilon = 0.001$ and $\delta = 0.02$. Since ω is a sinusoid, the parameter T was set to the corresponding period π. The resulting output error $\tilde{y}_\epsilon$, parameter estimate $\hat{\theta}$, and the update of the parameter region Ω^k are shown in Figs. 2–4, respectively. The evolutions of the lower and upper bounds $\underline{f}_i^k$ and $\bar{f}_i^k$, $i = 1, 2, 3$ with respect to t are also shown in Fig. 5. A similar convergence was observed to occur for any θ_0 in Ω^0. These figures show that the update of Ω^k is not necessarily periodic. Once $\tilde{y}_\epsilon$ becomes smaller than δ over an interval T, the corresponding parameter estimates and the upper and lower bounds on f_i and therefore the unknown parameter region are computed. It was also observed that just the min–max algorithm without the higher level component did not result in parameter convergence.

VI. Summary

In this paper, the problem of parameter estimation in systems with general nonlinear parameterization is considered. In systems with convex/concave parameterization, sufficient conditions are derived under which parameter estimates converge to their true values using a min–max algorithm as in [1]. In systems with a general nonlinear parameterization, a hierarchical min–max algorithm is proposed where the lower level consists of a min–max algorithm and the higher level component updates the bounds on the parameter region within which the unknown parameter is known to lie. Using this algorithm, a necessary and sufficient condition is established for parameter convergence in systems with a general nonlinear parameterization. In both cases, the conditions needed are shown to be stronger than linear persistent excitation conditions that guarantee parameter convergence in linearly parametrized systems, thereby leading to a general definition of NLPE.

The results in this paper establish parameter estimation in a system of the form (1). Even though the output is a scalar, as is shown in [6], a wide variety of adaptive control and estimation problems can be reduced to an error model of the form of (1). As a result, the persistent excitation conditions presented in this paper are applicable to all of these problems.

IEEE Proof

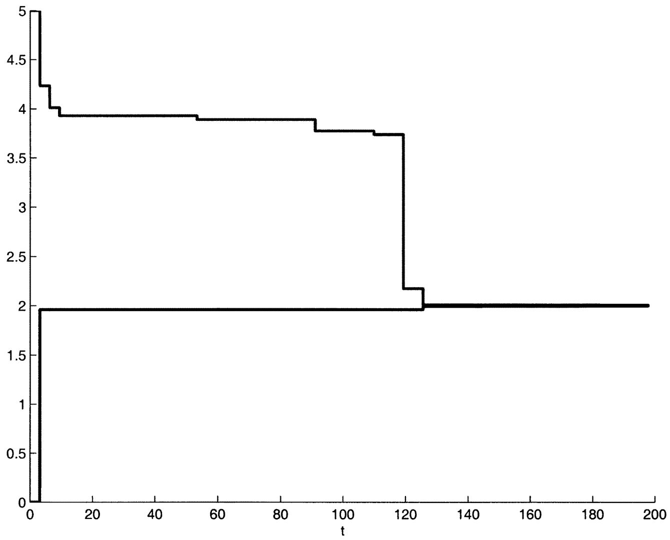

Fig. 4. Evolution of the parameter region Ω^k with t, using the hierarchical algorithm. Note that Ω^k is updated at instants t_k^* such that $|\tilde{y}_\epsilon(t)| \leq \delta$ for $t \in [t_k^* - T, t_k^*]$.

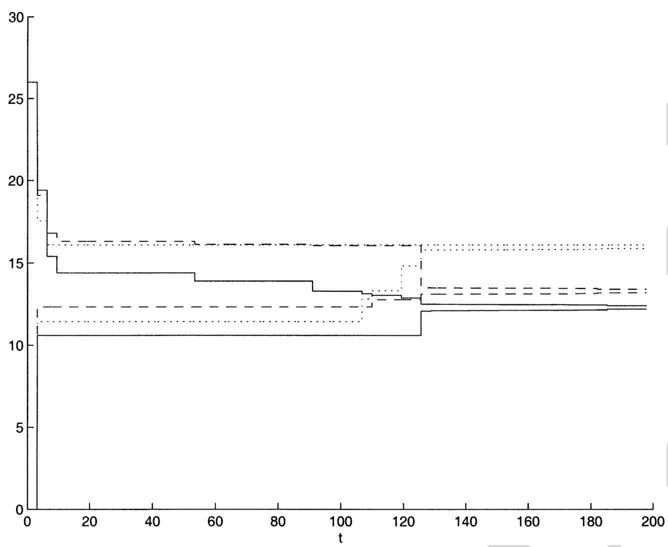

Fig. 5. Upper-bounds $\overline{f}_i^k$ and lower bounds $\underline{f}_i^k$ of $f(\theta, u_i)$ with t using the hierarchical algorithm, for $u_i = 1, -1, 0$. $\overline{f}_1, \underline{f}_1$:—, $\overline{f}_2, \underline{f}_2$: - - -, $\overline{f}_3, \underline{f}_3$:….

APPENDIX

A. Proof of Property 1

From (8) and (13), it follows that

$$\dot{V} = -\alpha(y,u)\tilde{y}_\epsilon^2 + \tilde{y}_\epsilon \times \left(f(\hat{\theta}, \omega) - f(\theta_0, \omega) - \phi^{*T}(\hat{\theta} - \theta_0) - a^* \mathrm{sat}\left(\frac{y}{\epsilon}\right)\right). \quad (48)$$

When $|\tilde{y}| \leq \epsilon$, it follows that $\tilde{y}_\epsilon = 0$ and, hence, $\dot{V} = 0$. When $|\tilde{y}| > \epsilon$, it follows that $\mathrm{sat}\,(\tilde{y}/\epsilon) = \mathrm{sign}(\tilde{y})$. Then, (48) is transformed into

$$\begin{aligned}\dot{V} = &- \alpha(y,u)\tilde{y}_\epsilon^2 \tilde{y}_\epsilon \mathrm{sat}\left(\frac{\tilde{y}}{\epsilon}\right) \\ &\times \left(\mathrm{sat}\left(\frac{\tilde{y}}{\epsilon}\right) + (f(\hat{\theta}, \omega) - f(\theta_0, \omega) \right. \\ &\left. - \phi^{*T}(\hat{\theta} - \theta_0)) - a^* \right). \end{aligned} \quad (49)$$

Combining (7) and (49), Property 1 is established. •

B. Proof of Property 2

To prove Property 2, the following sublemma is needed.

Sublemma 2:1: For given systems

$$\begin{aligned}\dot{x} &= - k(t)x + z(t) \\ \dot{x}_m &= - k_m x_m + z_m\end{aligned}$$

where $k(t) > 0, k_m > 0$ and

$$\begin{aligned} &|z(t)| \leq z_m \quad \forall\, t \geq t_0 \\ &\text{if } x(t_0) \leq x_m(t_0) < 0, k(t) \leq k_m \\ &\text{then } x(t) \leq x_m(t) \qquad \forall\, t \geq t_0 \text{ where } x_m(t) \leq 0.\end{aligned}$$

The proof of the sublemma is straight forward and is omitted. Now, let us prove Property 2.

Without loss of generality, we assume that

$$\tilde{y}_\epsilon(t_1) \leq -\gamma. \quad (50)$$

From (8), it follows that

$$\dot{\tilde{y}}_\epsilon = -\alpha(y,u)\tilde{y}_\epsilon + m(t) \quad (51)$$

where $m(t)$ is defined as in (17). From Assumption 2, because Ω^0 is bounded, $|f(\hat{\theta}, u) - f(\theta_0, u)|$, a^* and therefore $m(t)$ are also bounded, with $|m(t)|$ bounded by M. Let $y_m(t)$ be specified as the solution of the following differential equation for $t \geq t_1$:

$$\dot{y}_m = -\alpha_{\max} y_m + M \quad y_m(t_1) = -\gamma. \quad (52)$$

From (50), (51), and (52), Sublemma 2.1 implies that

$$\tilde{y}_\epsilon(t_1 + \tau) \leq y_m(t_1 + \tau) \qquad \forall\, \tau \geq 0 \text{ and } y_m(t_1 + \tau) \leq 0. \quad (53)$$

From (52), it follows that

$$y_m(t_1 + \tau) = \left(-\frac{M}{\alpha_{\max}} - \gamma \right) e^{-\alpha_{\max}\tau} + \frac{M}{\alpha_{\max}}.$$

We note that $y_m(t_1 + \tau)$ is a concave function of τ for $\tau \geq 0$. From properties of concave functions, it can be shown that $y_m(t_1 + \tau)$ satisfies the inequality

$$y_m(t_1 + \tau) \leq y_m(t_1) + \nabla_\tau y_m(t_1 + \tau)\,|_{\tau=0}. \quad (54)$$

From (53) and (54), we obtain that

$$\tilde{y}_\epsilon(t_1 + \tau) \leq -\gamma + (M + \alpha_{\max}\gamma)\tau, \qquad \text{for } \tau \geq 0. \quad (55)$$

For $T' = (\gamma / M + \alpha_{\max}\gamma)$, we can verify easily from (55) that

$$\tilde{y}_\epsilon(t) \leq 0 \quad \forall\, t \in [t_1, t_1 + T'].$$

From (55), we have that

$$\int_{t_1}^{t_1+T'} |\tilde{y}_\epsilon(\tau)|^2 d\tau \geq \frac{\gamma^3}{3(M + \alpha_{\max}\gamma)}.$$

Integrating (14) over $[t_1, t_1+T']$, we have that

$$V(t_1+T') \leq V(t_1) - \frac{\alpha_{\min}\gamma^3}{3(M+\alpha_{\max}\gamma)}.$$

For

$$\tilde{y}_\epsilon(t_1) \geq \gamma$$

we can obtain a similar result. This proves Property 2. •

C. Proof of Property 3

Let us first prove that

$$-a_-^*(\hat{\theta},\omega) \leq a^* \mathrm{sat}\left(\frac{\tilde{y}}{\epsilon}\right). \tag{56}$$

Since $g(\hat{\theta},\omega,\phi)=0$, it follows that for at least one value of θ in Ω^0, $g(\theta,\omega,\phi)=0$. This proves

$$a^* \geq 0. \tag{57}$$

If $\tilde{y} \geq 0$, from (57), (56) holds. If $\tilde{y}<0$, it follows that

$$a^* \mathrm{sat}\left(\frac{\tilde{y}}{\epsilon}\right) = a_-^*(\hat{\theta},\omega)\mathrm{sat}\left(\frac{\tilde{y}}{\epsilon}\right) \geq -a_-^*(\hat{\theta},\omega).$$

Similarly, we can prove $a^* \leq a_+^*(\hat{\theta},\omega)$ and Property 3 is established. •

D. Proof of Property 4

Since $\beta=-1$, $f(\theta,\omega)$ is concave. It follows from the solutions of the min–max algorithm in Section II-A that

$$a^*=0, \text{ if } \tilde{y}<0 \tag{58}$$

which proves Property 4-i). When $\tilde{y}>0$, it follows from the solutions of the min–max algorithm that a^* is nonnegative, hence

$$\beta a^* \tilde{y} \leq 0, \text{ if } \tilde{y}>0. \tag{59}$$

When $\beta=1$, proceeding in the same manner as before, it follows that

$$a^*=0, \text{ if } \beta=1, \text{and } \tilde{y}>0 \tag{60}$$

and

$$\beta a^* \tilde{y} \leq 0, \text{ if } \beta=1, \text{and } \tilde{y}<0. \tag{61}$$

prove Property 4-ii) and 4-iii). •

E. Proof of Theorem 1

For any t_1 and $\hat{\theta}(t_1)$, it follows from (20) that there exists $t_2 < t_1+T_0$ such that

$$\beta(\omega(t_2))\left(f(\hat{\theta}(t_1),\omega(t_2)) - f(\theta_0,\omega(t_2))\right) \geq \epsilon_u\|\hat{\theta}(t_1)-\theta_0\|. \tag{62}$$

Without loss of generality, we assume that $\beta(\omega(t_2))=1$ which means that $f(\theta,\omega(t_2))$ is convex (or linear) over θ. The proof can be given in a similar manner if $\beta(t_2)=-1$. When $\beta(t_2)=1$, (62) can be rewritten as

$$f(\hat{\theta}(t_1),\omega(t_2)) - f(\theta_0,\omega(t_2)) \geq \bar{\epsilon} \tag{63}$$

where

$$\bar{\epsilon} = \epsilon_u\|\hat{\theta}(t_1)-\theta_0\|. \tag{64}$$

If $x(t_1)\in D_\epsilon$, we note that $x(t)\in D_\epsilon$ for all $t\geq t_1$, since V is a Lyapunov function. Hence, we assume that $x(t_1)\notin D_\epsilon$. It follows from the definition of V in (13) that either

$$\text{i) } \|\tilde{\theta}(t_1)\| > \sqrt{\gamma_1} \text{ or ii) } |\tilde{y}_\epsilon(t_1)| > \sqrt{\gamma_1}. \tag{65}$$

If (65)-ii) holds, it is easy to show that V decreases. If (65)-i) holds, we show below that $\tilde{y}_\epsilon(t)$ will become large for some $t>t_1$. Using the definitions of γ_1 in (22), it follows from (64) and (65)-i) that

$$\bar{\epsilon}^2 > 2\epsilon(16B_\theta U_b + 8B_\theta B_\phi + 4B_\phi^2). \tag{66}$$

We shall show that if (66) holds, there exists $t_3\in[t_2,t_2+T_1]$ such that

$$|\tilde{y}_\epsilon(t_3)| \geq \min\{1,\bar{\delta}\} \tag{67}$$

where (68) and (69), shown at the bottom of the page, hold. From (66), we can verify easily that both $\bar{\delta}$ and T_1 are positive numbers. We prove (67) by contradiction.

Suppose (66) holds and (67) is not true. Then, it follows that

$$\text{(a) } |\tilde{y}_\epsilon(t_2+\tau)|<1 \text{ and (b) } |\tilde{y}_\epsilon(t_2+\tau)|<\bar{\delta} \tag{70}$$

for any $\tau\in[0,T_1]$. From (8) and (70) b), it follows that

$$\dot{\tilde{y}} \geq -\alpha_{\max}\bar{\delta} + \left[f(\hat{\theta},\omega)-f(\theta_0,\omega)\right] - a^*\mathrm{sat}\left(\frac{\tilde{y}}{\epsilon}\right). \tag{71}$$

We prove that $\tilde{y}_\epsilon(t)$ must become large over $[t_2,t_2+T_1]$ by establishing lower bounds on the bracketed term and the last term on the right-hand side of (71).

$$\bar{\delta} = \min\left\{\frac{\bar{\epsilon}}{2(B_\theta B_\phi T_0+\alpha_{\max})}, \frac{\bar{\epsilon}^2 - \epsilon(16B_\theta U_b+8B_\theta B_\phi+4B_\phi^2)}{2\bar{\epsilon}B_\theta B_\phi T_0 + 2\bar{\epsilon}\alpha_{\max} + 16B_\theta U_b + 8B_\theta B_\phi + 4B_\phi^2}\right\} \tag{68}$$

$$T_1 = \frac{\bar{\epsilon} - (B_\theta B_\phi T_0+\alpha_{\max})\bar{\delta}}{4B_\theta U_b + 2B_\theta B_\phi + B_\phi^2}. \tag{69}$$

IEEE Proof

It follows from (4), (70) b), Assumption 2, and the fact that $t_2 - t_1 \leq T_0$, that

$$|f(\hat{\theta}(t_2), \omega(t_2)) - f(\hat{\theta}(t_1), \omega(t_2))| \leq B_\theta B_\phi T_0 \bar{\delta}. \tag{72}$$

Combining (63) and (72), we have that

$$f(\hat{\theta}(t_2), \omega(t_2)) - f(\theta_0, \omega(t_2)) \geq \bar{\epsilon} - B_\theta B_\phi T_0 \bar{\delta}. \tag{73}$$

From Assumption 1, it follows that

$$\|\omega(t_2 + \tau) - \omega(t_2)\| \leq U_b \tau. \tag{74}$$

For $\tau \in [0, T_1]$, since $|\tilde{y}_\epsilon(t_2 + \tau)| \leq 1$ from (70) (a), by integrating (4) over $[t_2, t_2 + \tau]$, we obtain that

$$\|\hat{\theta}(t_2 + \tau) - \hat{\theta}(t_2)\| \leq B_\phi \tau. \tag{75}$$

By combining (74), (75), and Assumption 2, it follows that

$$\begin{aligned} \left| \hat{f}_{2\tau} - \hat{f}_2 - (f(\theta_0, \omega(t_2 + \tau)) - f(\theta_0, \omega(t_2))) \right| \\ \leq B_\theta (2U_b + B_\phi) \tau \end{aligned} \tag{76}$$

which can be rewritten as

$$\begin{aligned} &\hat{f}_{2\tau} - f(\theta_0, \omega(t_2 + \tau)) \\ &\quad \geq \hat{f}_2 - f(\theta_0, \omega(t_2)) - B_\theta (2U_b + B_\phi) \tau \end{aligned} \tag{77}$$

where

$$\begin{aligned} \hat{f}_{2\tau} &= f(\hat{\theta}(t_2 + \tau), \omega(t_2 + \tau)) \\ \hat{f}_2 &= f(\hat{\theta}(t_2), \omega(t_2)). \end{aligned} \tag{78}$$

Combining (73) and (77), it follows that for any $\tau \in [0, T_1]$

$$\begin{aligned} &\hat{f}_{2\tau} - f(\theta_0, \omega(t_2 + \tau)) \\ &\quad \geq \bar{\epsilon} - B_\theta B_\phi T_0 \bar{\delta} - B_\theta (2U_b + B_\phi) \tau \end{aligned} \tag{79}$$

which establishes a lower bound for the bracketed term in (71).

We now derive a lower bound for the third term in (71). For any θ, using the same procedure as for (76), it can be shown that

$$\begin{aligned} |\hat{f}_{2\tau} - \hat{f}_2 - (f(\theta, \omega(t_2 + \tau)) - f(\theta, \omega(t_2)))| \\ \leq B_\theta (2U_b + B_\phi) \tau \end{aligned} \tag{80}$$

where $\hat{f}_{2\tau}, \hat{f}_2$ are defined in (78). It follows from (75) that

$$|\phi^*(\tau_i)(\hat{\theta}(t_2 + \tau) - \hat{\theta}(t_2))| \leq B_\phi^2 \tau. \tag{81}$$

We know that at t_2, because $\beta(t_2) = 1$, it follows from Property 4-ii) that

$$a_+^*(\hat{\theta}(t_2), \omega(t_2)) = 0. \tag{82}$$

From the definition of a_+^*, and the optimization problem in (6), we obtain that

$$\begin{aligned} a_+^*(\hat{\theta}(t_2), \omega(t_2)) &= \max_{\theta \in \Omega^0} \Big(\hat{f}_2 - f(\theta, \omega(t_2)) \\ &\qquad - \phi^*(t_2)(\hat{\theta}(t_2) - \theta) \Big) \\ a_+^*(\hat{\theta}(t_2 + \tau), \omega(t_2 + \tau)) &= \max_{\theta \in \Omega^0} \Big(\hat{f}_{2\tau} - f(\theta, \omega(t_2 + \tau)) \\ &\qquad - \phi^*(t_2 + \tau) \\ &\qquad \times (\hat{\theta}(t_2 + \tau) - \theta) \Big) \end{aligned} \tag{83}$$

where $\hat{f}_{2\tau}, \hat{f}_2$ are defined as in (78). Because $\phi^*(\tau_i + t)$ is the value that result in the minimum value of $a_+^*(\hat{\theta}(\tau_i + t), \omega(\tau_i + t))$, it follows that

$$\begin{aligned} a_+^*(\hat{\theta}(t_2 + \tau), \omega(t_2 + \tau)) \leq \max_{\theta \in \Omega^0} \Big(\hat{f}_{2\tau} - f(\theta, \omega(t_2 + \tau)) \\ - \phi^*(t_2)(\hat{\theta}(t_2 + \tau) - \theta) \Big). \end{aligned} \tag{84}$$

Combining (83) and (84), it follows that

$$\begin{aligned} &a_+^*(\hat{\theta}(t_2 + \tau), \omega(t_2 + \tau)) \\ &\quad \leq a_+^*(\hat{\theta}(t_2), \omega(t_2)) \\ &\qquad + \max_{\theta \in \Omega^0} \Big(\hat{f}_{2\tau} - f(\theta, \omega(t_2 + \tau)) \\ &\qquad\qquad - (\hat{f}_2 - f(\theta, \omega(t_2))) - \phi^*(\tau_i) \\ &\qquad\qquad \times (\hat{\theta}(t_2 + \tau) - \hat{\theta}(t_2)) \Big) \end{aligned} \tag{85}$$

where $\hat{f}_{2\tau}, \hat{f}_2$ are defined in (78). From (81), (85), and (80), it follows that

$$\begin{aligned} &a_+^*(\hat{\theta}(t_2 + \tau), \omega(t_2 + \tau)) \\ &\quad \leq a_+^*(\hat{\theta}(t_2), \omega(t_2)) + B_\theta (2U_b + B_\phi) \tau + B_\phi^2 \tau. \end{aligned} \tag{86}$$

Combining (82) and (86), it follows that

$$\begin{aligned} &-a_+^*(\hat{\theta}(t_2 + \tau), \omega(t_2 + \tau)) \mathrm{sat}\left(\frac{\tilde{y}}{\epsilon}\right) \\ &\qquad \geq -B_\theta (2U_b + B_\phi) \tau - B_\phi^2 \tau. \end{aligned} \tag{87}$$

It follows from (87) and Property 3 that for all $\tau \in [0, T_1]$

$$\begin{aligned} &-a^*(\hat{\theta}(t_2 + \tau), \omega(t_2 + \tau)) \mathrm{sat}\left(\frac{\tilde{y}}{\epsilon}\right) \\ &\qquad \geq -B_\theta (2U_b + B_\phi) \tau - B_\phi^2 \tau \end{aligned} \tag{88}$$

which establishes a lower bound on the last term on the right-hand side of (71).

Using (79) and (88), (71) leads to the inequality

$$\begin{aligned} \dot{\tilde{y}}(t_2 + \tau) \geq &\bar{\epsilon} - (B_\theta B_\phi T_0 + \alpha_{\max}) \bar{\delta} \\ &- (4B_\theta U_b + 2B_\theta B_\phi + B_\phi^2) \tau \\ \forall\, \tau \in &[0, T_1]. \end{aligned} \tag{89}$$

Integrating both sides of (89) over $[t_2, t_2 + T_1]$ where T_1 is defined in (69), we have

$$\begin{aligned} &\tilde{y}(t_2 + T_1) - \tilde{y}(t_2) \\ &\quad \geq \int_{t_2}^{t_2 + T_1} \big(\bar{\epsilon} - (B_\theta B_\phi T_0 + \alpha_{\max}) \bar{\delta} \\ &\qquad\qquad - (4B_\theta U_b + 2B_\theta B_\phi + B_\phi^2) \tau \big) \, d\tau \end{aligned}$$

which can be simplified as

$$\tilde{y}(t_2 + T_1) - \tilde{y}(t_2) \geq \frac{(\bar{\epsilon} - (B_\theta B_\phi T_0 + \alpha_{\max}) \bar{\delta})^2}{2(4B_\theta U_b + 2B_\theta B_\phi + B_\phi^2)}. \tag{90}$$

IEEE Proof

Since (70) holds for all $\tau \in [0, T_1]$, we have that

$$\tilde{y}(t_2) > -\epsilon - \bar{\delta}. \tag{91}$$

It follows from the definition of $\bar{\delta}$ in (68) that

$$\bar{\delta} \leq \frac{\bar{\epsilon}^2 - \epsilon(16B_\theta U_b + 8B_\theta B_\phi + 4B_\phi^2)}{2\bar{\epsilon}B_\theta B_\phi T_0 + 2\bar{\epsilon}\alpha_{\max} + 16B_\theta U_b + 8B_\theta B_\phi + 4B_\phi^2}. \tag{92}$$

Equation (92) can be rewritten as

$$\frac{\bar{\epsilon}^2 - 2\bar{\epsilon}(B_\theta B_\phi T_0 + \alpha_{\max})\bar{\delta}}{2(4B_\theta U_b + 2B_\theta B_\phi + B_\phi^2)} \geq 2\epsilon + 2\bar{\delta}. \tag{93}$$

It follows from (90) and (91) that

$$\tilde{y}(t_2 + T_1) \geq \frac{\bar{\epsilon}^2 - 2\bar{\epsilon}(B_\theta B_\phi T_0 + \alpha_{\max})\bar{\delta}}{2(4B_\theta U_b + 2B_\theta B_\phi + B_\phi^2)} - \epsilon - \bar{\delta}. \tag{94}$$

From (93), (94) can be simplified as

$$\tilde{y}(t_2 + T_1) \geq \epsilon + \bar{\delta}. \tag{95}$$

Equation (95) implies $\tilde{y}_\epsilon(t_2 + T_1) \geq \bar{\delta}$ which contradicts (70). Thus, we have shown that (67) must hold.

In summary, we have shown that if $V(t_1) > \gamma_1$, then either

$$\begin{aligned} &\text{i) } |\tilde{y}_\epsilon(t_3)| \geq \min\{1, \bar{\delta}\} t_3 \in [t_1, t_1 + T_0 + T_1], \text{ or} \\ &\text{ii) } |\tilde{y}_\epsilon(t_1)| > \sqrt{\gamma_1}. \end{aligned} \tag{96}$$

where $t_3 = t_2 + T_1$. From Property 2 , it follows that if (96)-i) holds, then there exists $T_1' = (\bar{\delta}/M + \alpha_{\max}\bar{\delta})$ such that

$$V(t_3 + T_1') \leq V(t_3) - \frac{\alpha_{\min}\bar{\delta}^3}{3(M + \alpha_{\max}\bar{\delta})}. \tag{97}$$

Similarly, if (96)-ii) holds, then

$$V(t_1 + T_2') \leq V(t_1) - \frac{\alpha_{\min}\sqrt{\gamma_1}^3}{3(M + \alpha_{\max}\sqrt{\gamma_1})} \tag{98}$$

where $T_2' = (\sqrt{\gamma_1}/M + \alpha_{\max}\sqrt{\gamma_1})$. Because $V(t)$ is nonincreasing, it follows from (97) and (98) that for any $V(t_1) > \gamma_1$

$$V(t_1 + T_3') \leq V(t_1) - \Delta V \tag{99}$$

where

$$\begin{aligned} T_3' &= \max\{T_0 + T_1 + T_1', T_2'\} \\ \Delta V &= \min\left\{\frac{\alpha_{\min}\bar{\delta}^3}{3(M + \alpha_{\max}\bar{\delta})}, \frac{\alpha_{\min}\gamma_1^3}{3(M + \alpha_{\max}\gamma_1)}\right\}. \end{aligned}$$

This implies that $V(t)$ decreases by a finite amount over erery interval T_3' until trajectories reach D_ϵ. This proves Theorem 1. •

F. Proof of Lemma 2

For any m, if

$$|\tilde{y}_\epsilon(t)| \leq \delta \quad \forall\, t \in [t_0 + mT, t_0 + (m+1)T] \tag{100}$$

we are done. Otherwise, it means there exists $t_1 \in [t_0 + mT, t_0 + (m+1)T]$ such that

$$|\tilde{y}_\epsilon(t_1)| \geq \delta.$$

It follows from Property 2 that there exists $T' = (\delta/M + \alpha_{\max}\delta)$ such that

$$V(t_1 + T') \leq V(t_1) - \frac{\alpha_{\min}\delta^3}{3(M + \alpha_{\max}\delta)}.$$

This implies that everytime when $|\tilde{y}_\epsilon(t)| \geq \delta$, Lyapunov function will decrease a small amount. Now, that $V(t_0)$ is finite, these kind of situation can only happen finite times. It means that we can find a finite m^* such that

$$|\tilde{y}_\epsilon(t)| \leq \delta \quad \forall\, t \in [t_0 + m^*T, t_0 + (m^*+1)T].$$

This establishes Lemma 2. •

G. Proof of Lemma 3

We shall prove by contradiction that Lemma 3 holds. Assume that $\theta_0 \notin \Phi_\epsilon(\Omega, \omega_i, \epsilon, \delta, \hat{\theta}_i^c)$ for some $1 \leq i \leq I$. That is

$$\text{i) } f(\theta_0, \omega_i) < \underline{f}_i, \text{ or ii) } f(\theta_0, \omega_i) > \bar{f}_i$$

for some $1 \leq i \leq I$. Suppose (i) is true. Since $\hat{\theta}_i^c = \hat{\theta}(\tau_i)$, case (i) implies that

$$f(\hat{\theta}(\tau_i), \omega) - f(\theta_0, \omega) - a_+^*(\hat{\theta}(\tau_i), \omega) - \alpha_{\max}\delta > 2\sqrt{B_t(\delta + \epsilon)}. \tag{101}$$

From (8), Property 3 and the fact $|\tilde{y}_\epsilon| \leq \delta$, it follows that

$$\dot{\tilde{y}} \geq y_l \tag{102}$$

where

$$y_l = -\alpha_{\max}\delta + f(\hat{\theta}, \omega) - f(\theta_0, \omega) - a_+^*(\hat{\theta}, \omega) \tag{103}$$

represents the lower bound of $\dot{\tilde{y}}$. Combining (101) and (103), it follows that

$$y_l(\tau_i) > 2\sqrt{B_t(\delta + \epsilon)}. \tag{104}$$

From the definition of a_+^*, and the optimization problem in (6), we obtain that

$$\begin{aligned} a_+^*(\hat{\theta}(\tau_i), \omega(\tau_i)) &= \max_{\theta \in \Omega^0} \Big(\hat{f}_\tau - f(\theta, \omega(\tau_i)) \\ &\qquad - \phi^*(\tau_i)(\hat{\theta}(\tau_i) - \theta)\Big) \\ a_+^*(\hat{\theta}(\tau_i + t), \omega(\tau_i + t)) &= \max_{\theta \in \Omega^0} \Big(\hat{f}_{\tau t} - f(\theta, \omega(\tau_i + t)) \\ &\qquad - \phi^*(\tau_i + t) \\ &\qquad \times (\hat{\theta}(\tau_i + t) - \theta)\Big) \end{aligned} \tag{105}$$

where

$$\begin{aligned} \hat{f}_\tau &= f(\hat{\theta}(\tau_i), \omega(\tau_i)) \\ \hat{f}_{\tau t} &= f(\hat{\theta}(\tau_i + t), \omega(\tau_i + t)). \end{aligned} \tag{106}$$

Because $\phi^*(\tau_i + t)$ is the value that result in the minimum value of $a_+^*(\hat{\theta}(\tau_i + t), \omega(\tau_i + t))$, it follows that

$$\begin{aligned} a_+^*(\hat{\theta}(\tau_i + t), \omega(\tau_i + t)) \leq \max_{\theta \in \Omega^0} \Big(&\hat{f}_{\tau t} - f(\theta, \omega(\tau_i + t)) \\ &- \phi^*(\tau_i)(\hat{\theta}(\tau_i + t) - \theta)\Big). \end{aligned} \tag{107}$$

IEEE Proof

Combining (105) and (107), it follows that

$$\begin{aligned} a_+^*(\hat{\theta}(\tau_i+t),\omega(\tau_i+t)) &\le a_+^*(\hat{\theta}(\tau_i),\omega(\tau_i)) \\ &+ \max_{\theta\in\Omega^0}\Big(\hat{f}_{\tau t} - f(\theta,\omega(\tau_i+t)) \\ &\quad - (\hat{f}_\tau - f(\theta,\omega(\tau_i))) \\ &\quad - \phi^*(\tau_i)(\hat{\theta}(\tau_i+t)-\hat{\theta}(\tau_i))\Big) \end{aligned} \tag{108}$$

where $\hat{f}_\tau$ and $\hat{f}_{\tau t}$ are defined as in (1). From Assumption 1, it follows that

$$\|\omega(\tau_i+t)-\omega(\tau_i)\| \le U_b t. \tag{109}$$

Since $|\tilde{y}_\epsilon(\tau_i+t)| \le \delta$, by integrating (4) over $[\tau_i, \tau_i+t]$, we obtain that

$$\|\hat{\theta}(\tau_i+t)-\hat{\theta}(\tau_i)\| \le \delta B_\phi t. \tag{110}$$

By combining (109), (110), and Assumption 2, it follows that

$$\begin{aligned} |\hat{f}_{\tau t} - \hat{f}_\tau - (f(\theta,\omega(\tau_i+t)) - f(\theta,\omega(\tau_i)))| \\ \le B_\theta(2U_b+\delta B_\phi)t \end{aligned} \tag{111}$$

where $\hat{f}_\tau$ and $\hat{f}_{\tau t}$ are defined as in (106). From (110), (111), and (108), we get

$$\begin{aligned} a_+^*(\hat{\theta}(\tau_i+t),\omega(\tau_i+t)) \le a_+^*(\hat{\theta}(\tau_i),\omega(\tau_i)) \\ +B_\theta(2U_b+\delta B_\phi+\delta B_\phi^2)t. \end{aligned} \tag{112}$$

Incorporating (111) and (112) into (103), we have that

$$y_l(\tau_i+t)-y_l(\tau_i) > -(4B_\theta U_b + 2\delta B_\theta B_\phi + \delta B_\phi^2)t$$

which can be simplified using (31) and (104) as

$$y_l(\tau_i+t) > 2\sqrt{B_t(\delta+\epsilon)} - B_t t. \tag{113}$$

It follows from (102) and (113) that

$$\dot{\tilde{y}}(\tau_i+t) > 2\sqrt{B_t(\delta+\epsilon)} - B_t t. \tag{114}$$

Integrating both sides of (114) over $[\tau_i, \tau_i+\tau_B]$ where $\tau_B = 2\sqrt{B_t(\delta+\epsilon)}/B_t$,

$$\tilde{y}(\tau_i+\tau_B)-\tilde{y}(\tau_i) = \int_{\tau_i}^{\tau_i+\tau_B} 2\sqrt{B_t(\delta+\epsilon)} - B_t t. \tag{115}$$

Since $\tilde{y}(\tau_i) \ge -(\delta+\epsilon)$, we can rewrite (115) as

$$\tilde{y}(\tau_i+\tau_B) > (\delta+\epsilon). \tag{116}$$

Equation (116) implies

$$\tilde{y}_\epsilon(\tau_i+\tau_B) > \delta$$

and this contradicts the fact that $|\tilde{y}_\epsilon(t)| \le \delta$ over $[t_1, t_1+T]$. Thus, we conclude that the assumption (i) that $f < \underline{f}_i$ for some $1 \le i \le I$ is not true and, hence

$$\begin{aligned} f(\theta_0,\omega) \ge \underline{f}_i &= f(\hat{\theta}_i^c,\omega) - a_+^*(\hat{\theta}_i^c,\omega) \\ &- \alpha_{\max}\delta - 2\sqrt{B_t(\delta+\epsilon)} \\ &\forall\, i=1,\ldots,I. \end{aligned} \tag{117}$$

In the same manner, we can prove that

$$\begin{aligned} f(\theta,\omega) \le \bar{f}_i &= a_-^*(\hat{\theta}_i^c,\omega) + f(\hat{\theta}_i^c,\omega) \\ &+ \alpha_{\max}\delta + 2\sqrt{B_t(\delta+\epsilon)} \\ &\forall\, i=1,\ldots,I. \end{aligned} \tag{118}$$

Equations (117) and (118) conclude the proof of Lemma 3. •

H. Proof of Lemma 4

This proof follows directly from the definition of Δ_i in (36) and the construction of Ω^{k+1} in (34).

I. Proof of Lemma 5

We start with the hierarchical algorithm shown in Table I. Because $\Omega^{k+1} \subseteq \Omega^k$, there exists Ω^l such that

$$\Omega^l = \lim_{k\to\infty} \Omega^k. \tag{119}$$

Corresponding to Ω^l, if the lower level convergent estimate of θ_0 is given by $\hat{\theta}^l$, it follows that

$$\lim_{t\to\infty} \hat{\theta}(t) = \hat{\theta}^l. \tag{120}$$

Suppose (40) does not hold, it implies that

$$\hat{\theta}^l \notin \bigcap_{i=1}^{I} B(\Delta_i(\Omega^l)). \tag{121}$$

Then, there exists an $i, 1 \le i \le I$ such that

$$\hat{\theta}_i^l \notin \Delta_i(\Omega^l) \tag{122}$$

where $\hat{\theta}_i^l = \hat{\theta}(\tau_i)$ with $\omega(\tau_i) = \omega_i$. We can prove (122) by contradiction. We assume that

$$\hat{\theta}_i^l \in \Delta_i(\Omega^l) \quad \forall\, i = 1,\ldots,I.$$

Because

$$\|\hat{\theta}^l - \hat{\theta}_i^l\| \le \delta T B_\phi$$

combining the defintion of $L(\Omega^0,\epsilon,\delta)$ in (38), it follows that

$$\hat{\theta}^l \in L(\Omega^0,\epsilon,\delta)$$

which is a contradiction to (121). Thus, (122) must be true if (40) does not hold.

Let $\underline{f}_i^l$ and $\bar{f}_i^l$ be lower and upper bounds in Ω^l specified as

$$\underline{f}_i^l = \min_{\theta\in\Omega^l} f(\theta,\omega_i) \quad \bar{f}_i^l = \max_{\theta\in\Omega^l} f(\theta,\omega_i).$$

If we define

$$\begin{aligned} \widehat{\underline{f}}_i &= f(\hat{\theta}_i^l,\omega_i) - a_+^*(\hat{\theta}_i^l,\omega_i) - D(\epsilon,\delta) \\ \widehat{\bar{f}}_i &= f(\hat{\theta}_i^l,\omega_i) + a_-^*(\hat{\theta}_i^l,\omega_i) + D(\epsilon,\delta) \end{aligned}$$

where $D(\epsilon,\delta)$ is defined as in (37), (122) together with the definition of Δ_i, imply that

$$\widehat{\underline{f}}_i > \underline{f}_i^l \text{ or } \widehat{\bar{f}}_i < \bar{f}_i^l. \tag{123}$$

IEEE Proof

Equation (123) implies that tighter bounds $\underline{\hat{f}}_i$ or $\hat{\bar{f}}_i$ can be found for $f(\theta, \omega_i)$ for $\theta \in \Omega^l$ which implies that a smaller set Ω^{l+1} can be found using $\underline{\hat{f}}_i$ or $\hat{\bar{f}}_i$. This contradicts the assumption in (119) and Lemma 5 is proved. •

J. Proof of Theorem 5

Sufficiency follows directly from Lemma 5 and (41).

To prove necessity, we assume that (44) does not hold. That is, there exists $\Omega \subseteq \Omega^0$ where $\theta_0 \in \Omega$ such that

$$\text{i)} \bigcap_{i=1,\ldots,I} \bar{\Delta}_i(\Omega) \neq \phi, \text{ and}$$
$$\text{ii)} \bigcap_{i=1,\ldots,I} \bar{\Delta}_i(\Omega) \neq \{\theta_0\}. \tag{124}$$

It implies that there exists some $\bar{\theta} \in \bigcap_{i=1,\ldots,I} \bar{\Delta}_i(\Omega)$ and $\bar{\theta} \neq \theta_0$. Assume that at iteration k, the unknown parameter region $\Omega^k = \Omega$ and the lower level convergent parameter estimate at this iteration is given by $\hat{\theta}^{ck} = \bar{\theta}$. Then, condition ii) in (124) implies that

$$\hat{\theta}^{ck} \in \bigcap_{i=1,\ldots,I} \bar{\Delta}_i(\Omega)$$

since $\bigcap_{i=1,\ldots,I} \bar{\Delta}^i(\Omega)$ is not empty. From Lemma 4 , it follows that

$$\Omega^j = \Omega^k = \Omega \quad j = k+1, k+2, \ldots$$

and $\hat{\theta}$ will remains at $\hat{\theta}^{ck}$ always. Since $\hat{\theta}^{ck} \neq \theta_0$, the parameter estimate will not converge to θ_0 even ϵ and δ approaches 0. This implies that (44) is a necessary condition for (43). •

K. Proof of Lemma 7

For any $\Omega \subseteq \Omega^0$ where $\theta_0 \in \Omega$, if (44) does not hold, it follows that

$$\bigcap_{i=1}^{I} \bar{\Delta}_i(\Omega) \neq \phi \text{ and } \bigcap_{i=1}^{I} \bar{\Delta}_i(\Omega) \neq \{\theta_0\}. \tag{125}$$

From (125), it follows that there exists $\theta \neq \theta_0$ such that

$$\theta \in \bigcap_{i=1}^{I} \bar{\Delta}_i(\Omega). \tag{126}$$

For this choice of θ, from (47) we have that there exists a $\omega_i(\theta)$ such that

$$\beta(f(\theta, \omega_i) - f(\theta_0, \omega_i)) > 0. \tag{127}$$

Without loss of generality, we assume that $\beta = -1$. It follows from Property 4-i) that

$$a_{-}^{*}(\theta, \omega_i) = 0. \tag{128}$$

It follows from (128) and the definition of $\bar{\Delta}_i(\Omega)$ in (42) that

$$f(\theta, \omega_i) \geq \bar{f}_i^* \qquad \forall\, \theta \in \bigcap_{i=1}^{I} \bar{\Delta}_i(\Omega). \tag{129}$$

From the definition of $\bar{f}_i^*$ in (35) and the fact that $\theta_0 \in \Omega$, it follows that

$$\bar{f}_i^* \geq f(\theta_0, \omega_i). \tag{130}$$

Combining (129) and (130), it follows that

$$f(\theta, \omega_i) \geq f(\theta_0, \omega_i) \tag{131}$$

which is a contradiction to (127) since $\beta = -1$. This proves Lemma 7. •

REFERENCES

[1] A. M. Annaswamy, A. P. Loh, and F. P. Skantze, "Adaptive control of continuous time systems with convex/concave parametrization," *Automatica*, vol. 34, pp. 33–49, Jan. 1998.

[2] R. Ortega, "Some remarks on adaptive neuro-fuzzy systems," *Int. J. Adapt. Control Signal Processing*, vol. 10, pp. 79–83, 1996.

[3] A. M. Annaswamy, C. Thanomsat, N. Mehta, and A. P. Loh, "Applications of adaptive controllers based on nonlinear parametrization," *ASME J. Dyna. Syst., Measure., Control*, vol. 120, pp. 477–487, December 1998.

[4] J. D. Bosković, "Adaptive control of a class of nonlinearly parametrized plants," *IEEE Trans. Automat. Contr.*, vol. 43, pp. 930–934, July 1998.

[5] A. Kojić, A. M. Annaswamy, A.-P. Loh, and R. Lozano, "Adaptive control of a class of nonlinear systems with convex/concave parameterization," *Syst. Control Lett.*, vol. 37, pp. 267–274, 1999.

[6] A. P. Loh, A. M. Annaswamy, and F. P. Skantze, "Adaptation in the presence of a general nonlinear parametrization: An error model approach," *IEEE Trans. Automat. Contr.*, vol. 44, pp. 1634–1652, Sept. 1999.

[7] M. S. Netto, A. M. Annaswamy, R. Ortega, and P. Moya, "Adaptive control of a class of nonlinearly parametrized systems using convexification," *Int. J. Control*, vol. 73, pp. 1312–1321, 2000.

[8] F. P. Skantze, A. Kojic, A. P. Loh, and A. M. Annaswamy, "Adaptive estimation of discrete-time systems with nonlinear parametrization," *Automatica*, vol. 36, pp. 1879–1887, 2000.

[9] K. S. Narendra and A. M. Annaswamy, *Stable Adaptive Systems*. Upper Saddle River, NJ: Prentice-Hall, 1989.

[10] A. P. Morgan and K. S. Narendra, "On the stability of nonautonomous differential equations $\dot{x} = [A + B(t)]x$ with skew-symmetric matrix $B(t)$," *SIAM J. Control Optim.*, vol. 15, pp. 163–176, Jan. 1977.

[11] E. Panteley, A. Loria, and A. Teel, "Relaxed persistency of excitation for uniform asymptotic stability," *IEEE Trans. Automat. Contr.*, vol. 46, pp. 1874–1888, Dec. 2001.

[12] Y. Zhang, P. Ioannou, and C. Chien, "Parameter convergence of a new class of adaptive controllers," *IEEE Trans. Automat. Contr.*, vol. 41, pp. 1489–1493, Oct. 1996.

[13] C. Cao, A. M. Annaswamy, and A. Kojic, "Active and adaptive control lab," Mass. Inst. Technol., Cambridge, MA, 2002.

Chengyu Cao received the B.E. degree in information and control engineering from Xi'an Jiaotong University, P.R. China, and the M.S. degree in manufacturing engineering from Boston University, Boston, MA, in 1995 and 1999, respectively. He is currently working towards the Ph.D. degree in the Mechanical Engineering Department, the Massachusetts Institute of Technology, Cambridge, MA.

Between 1995 and 1997, he was a Graduate Student in the System Engineering Institute of Xi'an Jiaotong University, P.R. China. His main research interests are in the fields of adaptive control, control and identification of nonlinear systems, and neural networks.

IEEE Proof

Anuradha M. Annaswamy (S'82–M'85–SM'01–F'02) received the Ph.D. degree in electrical engineering from Yale University, New Haven, CT, in 1985.

She has been a member of the faculty at Yale University, Boston University, Boston, MA, and the Massachusetts Institute of Technology, Cambridge, MA, where she is currently the Director of the Active-Adaptive Control Laboratory and a Principal Research Scientist in the Department of Mechanical Engineering. Her research interests pertain to adaptive control, active control of resonant thermo-fluid systems including combustion processes and supersonic flows, and neural networks. She has authored numerous journal and conference papers and coauthored a graduate textbook on adaptive control.

Dr. Annaswamy has received several awards including the Alfred Hay Medal from the Indian Institute of Science in 1977, the Stennard Fellowship from Yale University in 1980, the IBM Postdoctoral Fellowship in 1985, the George Axelby Outstanding Paper Award from the IEEE Control Systems Society in 1988, and the Presidential Young Investigator award from the National Science Foundation in 1991. Dr. Annaswamy is a Member of AIAA.

Aleksandar Kojic (S'01–A'01) was born in 1974 and is a native of Kragujevac, Yugoslavia. He received the B.Sc. degree from the Mechanical Engineering Department, the University of Kragujevac, Kragujevac, Yugoslavia, and the M.S.M.E. and Ph.D. degrees from the in Mechanical Engineering Department, the Massachusetts Institute of Tecnology, Cambridge, MA, in 1995, 1998, and 2001, respectively.

He is currently a Senior Research Engineer at the Robert Bosch Corporation, Palo Alto, CA. His research interests are in the area of identification and control of nonlinearly parameterized systems.

IEEE Proof