Deterministic Approach to Robust Adaptive Learning of Fuzzy Models

Mohit Kumar, *Associate Member, IEEE*, Regina Stoll, and Norbert Stoll

***Abstract*—This study is concerned with the adaptive learning of an interpretable Sugeno-type fuzzy inference system, in a deterministic framework, in the presence of data uncertainties and modeling errors. The authors explore the use of H^∞ estimation theory and least squares estimation for online learning of membership functions and consequent parameters without making any assumption and requiring *a priori* knowledge of upper bounds, statistics, and distribution of data uncertainties and modeling errors. The issues of data uncertainties, modeling errors, and time variations have been considered mathematically in a sensible way. The proposed robust approach to the adaptive learning of fuzzy models has been illustrated through the examples of adaptive system identification, time-series prediction, and estimation of an uncertain process.**

***Index Terms*—Fuzzy identification, gradient descent, H^∞-optimality, interpretability, least squares estimation.**

I. Introduction

FUZZY identification is considered a powerful tool for the approximation of nonlinear systems from the input–output measured data. This has stimulated many studies in the data-driven construction of fuzzy models using *ad hoc* approaches [1], [2], neural networks [3]–[5], genetic algorithms [6]–[9], clustering techniques [10]–[12], and Kalman filtering [13], [14]. Typically, gradient-descent-based algorithms (e.g., backpropagation) are used for the adaptive learning of fuzzy-model parameters [15]. Slow convergence, poor generalization, noise sensitivity, and loss of interpretability are main drawbacks of these algorithms. A robust backpropagation was introduced by Chen and Jain [16] to resist the noise effects, and Wang *et al.* [17] introduced a robust objective function. The issue of stability has been addressed by Yu and Li [18] to suggest a stable learning algorithm.

The problem of adaptive learning of fuzzy models is ill-posed [19], and regularization is suggested a general method to obtain stable and well-behaved solutions [19]–[21]. However, the crucial choice of regularization parameter is usually not obvious and application dependent. Wang and coworkers [22]–[23] use projection operator to keep the parameter estimates bounded. Also, different modifications (e.g., switching σ modification, e_1 modifications, parameter projections, and dead zones [24]) could be used in order to render robustness in the learning of fuzzy models [25]. Despite the above cited work on the subject of robust adaptive fuzzy learning, we feel the following needs.

1) There is a need to take into account that fuzzy learning is an ill-posed problem.
2) The issues, such as data uncertainties and modeling errors, should be considered mathematically in a sensible way without making any assumption and requiring *a priori* knowledge of upper bounds, statistics, and distribution of data uncertainties and modeling errors.

The above two concerns are relatively easy to address if the membership functions are kept constant and only the linear parameters (consequents) are tuned, since a large body of literature is available that deals with the robust estimation of linear parameters. There are mainly two approaches to the estimation of linear parameters: least squares estimation and H^∞-optimal estimation. The main contribution of this paper is to address the above listed concerns by extending the well-developed linear estimation theory (least squares estimation and H^∞-optimal estimation) to the particular structure of nonlinear fuzzy models in a deterministic framework. Sugeno-type fuzzy inference systems combine simplicity with good analytical properties [26], and allow qualitative insight into the relationships [27]–[30]. Therefore, we have considered the problem of robust adaptive learning of a Sugeno-type fuzzy inference system.

II. Sugeno Fuzzy Inference System

Let us consider an explicit mathematical formulation of a Sugeno-type fuzzy inference system that assigns to each crisp value (vector) in input space a crisp value in output space. Consider a Sugeno fuzzy inference system $(F_\mathrm{s} : X \to Y)$, mapping n-dimensional input space $(X = X_1 \times X_2 \times \cdots \times X_n)$ to one-dimensional real line, consisting of K different rules. The ith rule is in the form

$$\text{If } x_1 \text{ is } A_{i1} \text{ and } x_2 \text{ is } A_{i2} \cdots \text{ and } x_n \text{ is } A_{in}, \text{ then } y = c_i$$

for all $i = 1, 2, \ldots, K$, where $A_{i1}, A_{i2}, \ldots, A_{in}$ are nonempty fuzzy subsets of $X_1, X_2, \ldots, X_n$, respectively, such that the membership function $\mu_{A_{ij}} : X_j \to [0, 1]$ fulfills $\sum_{i=1}^{K} \prod_{j=1}^{n} \mu_{A_{ij}}(x_j) > 0$ for all $x_j \epsilon X_j$, and values $c_1, \ldots, c_K$ are real numbers. The different rules, by using "product" as conjunction operator, can be aggregated as

$$F_\mathrm{s}(x_1, x_2, \ldots, x_n) = \frac{\sum_{i=1}^{K} c_i \prod_{j=1}^{n} \mu_{A_{ij}}(x_j)}{\sum_{i=1}^{K} \prod_{j=1}^{n} \mu_{A_{ij}}(x_j)}. \qquad (1)$$

We assume that x_j belongs to a nonempty real intervals, i.e., $x_j \in [a_j, b_j]$ for all $j = 1, \ldots, n$. Let us define a real vector θ such that the membership functions can be constructed from

Manuscript received March 28, 2005; revised August 22, 2005 and December 16, 2005. This paper was recommended by Associate Editor D. Nauck.

M. Kumar and R. Stoll are with the Institute of Occupational and Social Medicine, Faculty of Medicine, University of Rostock, 18055 Rostock, Germany (e-mail: mohit.kumar@uni-rostock.de; regina.stoll@uni-rostock.de).

N. Stoll is with the Institute of Automation, College of Computer Science and Electrical Engineering, University of Rostock, 18119 Rostock–Warnemünde, Germany (e-mail: norbert.stoll@uni-rostock.de).

Digital Object Identifier 10.1109/TSMCB.2006.870625

1083-4419/$20.00 © 2006 IEEE

the elements of vector θ. To illustrate the construction of the membership functions based on knot vector (θ), consider the following examples.

1) Trapezoidal Membership Function: Let $\theta = (a_1, t_1^1, \ldots, t_1^{2P_1-2}, b_1, \ldots, a_n, t_n^1, \ldots, t_n^{2P_n-2}, b_n)$, such that for ith input, $a_i \le t_i^1 \le \cdots \le t_i^{2P_i-2} \le b_i$ holds $\forall \; i = 1, \ldots, n$. Now, P_i trapezoidal membership functions for ith input ($\mu_{A_{1i}}$, $\mu_{A_{2i}}, \ldots, \mu_{A_{P_i i}}$) can be defined as

$$\mu_{A_{1i}}(x_i, \theta) = \begin{cases} 1, & \text{if } x_i \in \left[a_i, t_i^1\right] \\ \frac{-x_i + t_i^2}{t_i^2 - t_i^1}, & \text{if } x_i \in \left[t_i^1, t_i^2\right] \\ 0, & \text{otherwise} \end{cases}$$

$$\mu_{A_{ji}}(x_i, \theta) = \begin{cases} \frac{x_i - t_i^{2j-3}}{t_i^{2j-2} - t_i^{2j-3}}, & \text{if } x_i \in \left[t_i^{2j-3}, t_i^{2j-2}\right] \\ 1, & \text{if } x_i \in \left[t_i^{2j-2}, t_i^{2j-1}\right] \\ \frac{-x_i + t_i^{2j}}{t_i^{2j} - t_i^{2j-1}}, & \text{if } x_i \in \left[t_i^{2j-1}, t_i^{2j}\right] \\ 0, & \text{otherwise} \end{cases}$$

$$\mu_{A_{P_i i}}(x_i, \theta) = \begin{cases} \frac{x_i - t_i^{2P_i-3}}{t_i^{2P_i-2} - t_i^{2P_i-3}}, & \text{if } x_i \in \left[t_i^{2P_i-3}, t_i^{2P_i-2}\right] \\ 1, & \text{if } x_i \in \left[t_i^{2P_i-2}, b_i\right] \\ 0, & \text{otherwise.} \end{cases}$$

2) Gaussian Membership Functions With Unit Dispersion: Let $\theta = (a_1, t_1^1, \ldots, t_1^{P_1-2}, b_1, \ldots, a_n, t_n^1, \ldots, t_n^{P_n-2}, b_n)$, such that for ith input, $a_i \le t_i^1 \le \cdots \le t_i^{P_i-2} \le b_i$ holds for all $i = 1, \ldots, n$. For ith input, P_i Gaussian membership functions assuming unit dispersion ($\mu_{A_{1i}}, \mu_{A_{2i}}, \ldots, \mu_{A_{P_i i}}$) can be defined as

$$\begin{aligned} \mu_{A_{1i}}(x_i) &= e^{-(x_i - a_i)^2} \\ \mu_{A_{ji}}(x_i, \theta) &= e^{-\left(x_i - t_i^j\right)^2} \\ \mu_{A_{P_i} i}(x_i) &= e^{-(x_i - b_i)^2}. \end{aligned}$$

The total number of possible K rules depends on the number of membership functions for each input, i.e., $K = \Pi_{i=1}^n P_i$, where P_i is the number of membership functions defined over ith input. For any choice of membership functions (which can be constructed from a vector θ), (1) can be rewritten as a function of θ

$$\begin{aligned} F_{\mathrm{s}}(x_1, x_2, \ldots, x_n) &= \sum_{i=1}^{K} c_i G_i(x_1, x_2, \ldots, x_n, \theta) \\ G_i(x_1, x_2, \ldots, x_n, \theta) &= \frac{\prod_{j=1}^n \mu_{A_{ij}}(x_j)}{\sum_{i=1}^K \prod_{j=1}^n \mu_{A_{ij}}(x_j)}. \end{aligned}$$

Let us introduce the following notation: $\alpha = [c_i]_{i=1,\ldots,K} \in R^K$, $x = [x_i]_{i=1,\ldots,n} \in R^n$, $G = [G_i(x, \theta)]_{i=1,\ldots,K} \in R^K$. Now, the above expression becomes $F_{\mathrm{s}}(x) = G^T(x, \theta)\alpha$. In this expression, θ is not allowed to be any arbitrary vector, since the elements of θ must ensure the following:

1) in case of trapezoidal membership functions

$$a_i \le t_i^1 \le \cdots \le t_i^{2P_i-2} \le b_i \qquad \forall i = 1, \ldots, n \qquad (2)$$

2) in case of Gaussian membership functions

$$a_i \le t_i^1 \le \cdots \le t_i^{P_i-2} \le b_i \qquad \forall i = 1, \ldots, n \qquad (3)$$

to preserve the linguistic interpretation of fuzzy rule base [31]. In other words, there must exist some $\epsilon_i > 0$ for all $i = 1, \ldots, n$, such that for trapezoidal membership functions

$$\begin{aligned} t_i^1 - a_i &\ge \epsilon_i \\ t_i^{j+1} - t_i^j &\ge \epsilon_i \qquad \text{for all } j = 1, 2, \ldots, (2P_i - 3) \\ b_i - t_i^{2P_i-2} &\ge \epsilon_i. \end{aligned}$$

The above inequalities can be written in terms of a matrix inequality $c\theta \ge h$ (as those in [19], [21], and [32]–[34]). Hence, the output of a Sugeno-type fuzzy model

$$F_{\mathrm{s}}(x) = G^T(x, \theta)\alpha, \qquad c\theta \ge h$$

is linear in consequents (i.e., α) but nonlinear in antecedents (i.e., θ).

III. Motivation for Deterministic Learning of Fuzzy

Let us consider the fuzzy identification of a time-varying process $y(k) = f_k(x(k) + \delta x_k) + \delta y_k$, using a Sugeno-type fuzzy model. The identification data consist of sequence $\{x(j), y(j)\}_{j=0}^k$, and data uncertainties $\{\delta x_j, \delta y_j\}_{j=0}^k$ are unknown but bounded signals. We assume that there exist some optimum parameters of a Sugeno fuzzy model, say ($\alpha_k^*, \theta_k^*,\ c\theta_k^* \ge h$), to approximate the time-varying process: $y(k) = G^T(x(k) + \delta x_k, \theta_k^*)\alpha_k^* + \delta y_k, c\theta_k^* \ge h$. Given the identification data $\{x(j), y(j)\}_{j=0}^k$, we want to estimate the parameter sequence $\{\alpha_j^*, \theta_j^*, c\theta_k^* \ge h\}_{j=0}^k$, in the presence of uncertainties $\{\delta x_j, \delta y_j\}_{j=0}^k$, without making any assumption and requiring *a priori* knowledge of upper bounds, statistics, and distribution of signals. Let us denote the estimation strategy by E and estimated parameters by $\{\alpha_j, \theta_j\}_{j=0}^k$, i.e., $\{\alpha_j, \theta_j\}_{j=0}^k = E(\{x(j), y(j)\}_{j=0}^k)$.

In the literature of neural networks and fuzzy modeling, instantaneous-gradient-based algorithms (such as backpropagation) are probably the most commonly used techniques for the adaptive learning of nonlinear model parameters. It is obvious that any choice of estimation strategy (e.g., gradient descent) will induce a transfer operator T_E from data uncertainties $\{\delta x_j, \delta y_j\}_{j=0}^k$ to the estimation errors $\{G^T(x(j), \theta_k^*)\alpha_k^* - G^T(x(j), \theta_j)\alpha_j\}_{j=0}^k$

$$\left\{G^T(x(j), \theta_k^*)\alpha_k^* - G^T(x(j), \theta_j)\alpha_j\right\}_{j=0}^k = T_E\left(\{\delta x_j, \delta y_j\}_{j=0}^k\right).$$

An ideal estimation strategy E^* would be obviously the one that results in the zero value of estimation errors, no matter what the uncertainties are. However, in practical problems, it is almost impossible to design an ideal estimation strategy since uncertainty signals are unknown. Therefore, we are concerned to choose an estimation strategy that is least sensitive to the uncertainties, i.e.,

$$\min_E \max_{\{\delta x_j, \delta y_j\}_{j=0}^k} \frac{\text{Estimation Error Energy}}{\text{Energy of Uncertainties}}.$$

In other words, the estimation strategy should minimize the maximum possible value of energy gain (from uncertainties to the estimation errors). Such an estimation strategy (being called as robust) will safeguard against the worst case effect of uncertainties on estimation performance. Also, the robust estimation strategy must be designed in a deterministic framework, since we want to accommodate all possible uncertainties and we do not have any *a priori* knowledge of the upper bounds, statistics, and distribution of uncertainty signals. One important requirement on the estimation strategy is that the interpretability of fuzzy model during the identification of membership functions should remain preserved. There would be a total loss of interpretability if (2) and (3) do not hold good, since in this case, membership functions constructed from the knot sequence would not have any linguistic interpretation. Thus, we must constrain our estimation strategy to ensure that $c\theta_j \geq h, j = 0, \ldots, k$. Therefore, we are motivated to develop robust algorithms for the adaptive learning of nonlinear (in terms of membership function parameters) interpretable fuzzy models in a deterministic framework. The next part of the paper presents two different robust fuzzy estimation strategies and their deterministic robustness analysis. The first estimation strategy is based on H^∞-optimal criterion, where H^∞ norm of a transfer operator, which maps the uncertainties to estimation errors, is minimized. The second one is based on the classical least square estimation criterion. Our aim is to achieve robustness not only against data uncertainties but also against other unknown disturbance, e.g., time variation of process parameters, and deviation of initial guess from true values.

IV. Mathematical Formulation

Let us consider the fuzzy approximation of any time-varying physical process described at kth time instant by

$$y(k) = f_k\left(x(k) + \delta x_k\right) + \delta y_k \tag{4}$$

where δx_k is the uncertainty present in input vector $x(k)$, δy_k is the uncertainty in output measurement $y(k)$, and f_k is an unknown function, to be identified by an interpretable fuzzy model. At this end, consider that the data uncertainties $(\delta x_k, \delta y_k)$ and unknown function $f_k(\cdot)$ are bounded, therefore it is always possible to define a time-varying unknown signal n_k based on (4), such that $y(k) = f_k(x(k)) + n_k$. Assume that there exists an interpretable Sugeno-type fuzzy model, say (α_k^*, θ_k^*) with $c\theta_k^* \geq h$, for approximating the process. Therefore

$$y(k) = G^T\left(x(k), \theta_k^*\right)\alpha_k^* + n_k. \tag{5}$$

Given the uncertain input–output identification data set $\{x(j), y(j)\}_{j=0}^k$, we are concerned to estimate the fuzzy parameter sequence $\{(\alpha_j^*, \theta_j^*), c\theta_j^* \geq h\}_{j=0}^k$, say $\{(\alpha_j, \theta_j), c\theta_j \geq h\}_{j=0}^k$.

A. H^∞-Optimal Estimation

For the H^∞-optimal estimation of fuzzy parameter causal sequence $\{(\alpha_j^*, \theta_j^*), c\theta_j^* \geq h\}_{j=0}^k$, we try to minimize the H^∞ norm of a transfer operator mapping the unknown disturbances $\{\mu^{-1/2}(\alpha_0^* - \alpha_{-1}), q^{-1/2}\{\alpha_{j+1}^* - \alpha_j^*\}_{j=0}^k, \mu_\theta^{-1/2}\{\theta_j^* - \theta_{j-1}\}_{j=0}^k, \{n_j\}_{j=0}^k\}$ to the estimation errors $\{G^T(x(j), \theta_j^*)\alpha_j^* - G^T(x(j), \theta_j)\alpha_j\}_{j=0}^k$. Here, the parameter $\mu > 0$ reflects *a priori* knowledge as to how close α_0^* is to initial guess α_{-1}, the parameter $q > 0$ reflects *a priori* knowledge as to how fast α_k^* varies with time, and the parameter $\mu_\theta > 0$ reflects *a priori* knowledge as to how close the parameters $\{\theta_j^*\}_{j=0}^k$ are to the initial guess $\{\theta_{j-1}\}_{j=0}^k$. Mathematically, we want to solve the following $\min - \max$ estimation problem (shown at the bottom of the page), where the estimation strategy $\{(\alpha_j, \theta_j)\}_{j=0}^k$ and unknown sequences $(\{\alpha_{j+1}^* - \alpha_j^*\}_{j=0}^k, \{\theta_j^*\}_{j=0}^k, \{n_j\}_{j=0}^k)$ are causal.

B. Least Squares Estimation

Consider the fuzzy modeling of the time-varying unknown processes as

$$\begin{aligned} \alpha_{k+1}^* &= \alpha_k^* + \delta\alpha_k^* \\ y(k) &= G^T\left(x(k), \theta_k^*\right)\alpha_k^* + n_k \end{aligned}$$

where the unknown quantity $\delta\alpha_k^*$ can be regarded as the process noise. Now, our concern is to estimate the unknown vectors (α_k^*, θ_k^*) and to track its variation with time k. This will be done via solving a regularized least squares estimation problem. We formulate the problem of constrained estimation of parameters $\alpha_0^*, \{\delta\alpha_j^*\}_{j=0}^k, \{\theta_j^*\}_{j=0}^k$ as the solution to the following quadratic problem:

$$\begin{aligned} [\alpha_0, \{\delta\alpha_j\}, \{\theta_j\}] &= \arg \min_{\left[\alpha_0^*, \{\delta\alpha_j^*\}, \{\theta_j^*\}\right]} \left[J_k, \left\{c\theta_j^* \geq h\right\}\right] \\ J_k &= q^{-1}\sum_{j=0}^k \|\delta\alpha_j^*\|^2 \\ &\quad + \mu_\theta^{-1}\sum_{j=0}^k \|\theta_j^* - \theta_j^0\|^2 + \mu^{-1}\|\alpha_0^* - \alpha_{-1}\|^2 \\ &\quad + \sum_{j=0}^k \left|y(j) - G^T\left(x(j), \theta_j^*\right)\alpha_j^*\right|^2 \end{aligned}$$

$$\min_{\{(\alpha_j, \theta_j), c\theta_j \geq h\}} \max_{\left\{\alpha_0^*, \left\{\alpha_{j+1}^* - \alpha_j^*\right\}, \left\{\theta_j^*, c\theta_j^* \geq h\right\}, \{n_j\}\right\}} \mathcal{T}_k$$

$$\mathcal{T}_k = \frac{\sum_{j=0}^k \left|G^T\left(x(j), \theta_j^*\right)\alpha_j^* - G^T\left(x(j), \theta_j\right)\alpha_j\right|^2}{\mu^{-1}\left\|\alpha_0^* - \alpha_{-1}\right\|^2 + \sum_{j=0}^k \left(q^{-1}\left\|\alpha_{j+1}^* - \alpha_j^*\right\|^2 + \mu_\theta^{-1}\left\|\theta_j^* - \theta_{j-1}\right\|^2 + |n_j|^2\right)}$$

subject to the constraint $\alpha_{j+1}^* = \alpha_j^* + \delta\alpha_j^*$, where the parameters $\{\theta_j^0\}_{j=0}^k$ are initial guess about $\{\theta_j^*\}_{j=0}^k$, α_{-1} is an initial guess about α_0^*, the unknown sequences $(\{\delta\alpha_j^*\}_{j=0}^k, \{\theta_j^*\}_{j=0}^k, \{n_j\}_{j=0}^k)$ and the estimation strategy $\{\alpha_0, \delta\alpha_j, \theta_j\}_{j=0}^k$ are causal. The positive parameters (μ, μ_θ, q), as stated above, penalize the initial guess about some true fuzzy-model parameters and speed of their variation. The above least square estimation criterion will be shown robust to unknown disturbances in the latter part of the paper. The least square approach to the adaptive estimation of nonlinear parameters has been already explored in the literature of state space filtering (see, e.g., [35] and [36]).

V. H^∞-Suboptimal Estimation

Let us solve a suboptimal H^∞ estimation problem. That is, given a scalar $\gamma \geq 0$, find a causal estimation strategy $\{\alpha_j, \theta_j, c\theta_j \geq h\}_{j=0}^k$ that achieves (6), shown at the bottom of the page, for any vector α_0^* and for all causal sequences $(\{n_j\}_{j=0}^k, \{\alpha_{j+1}^* - \alpha_j^*\}_{j=0}^k, \{\theta_j^*, c\theta_j^* \geq h\}_{j=0}^k)$. If we are only concerned to estimate the linear parameters (i.e. $\{\alpha_j\}_{j=0}^k$), then the standard results of H^∞ linear estimation theory can be used (see, e.g., [37] and [38]). Now, for the nonlinear constrained fuzzy-model parameter estimation, we extend the H^∞ linear estimation approach by defining an indefinite quadratic form

$$\begin{aligned}
&J_k\left(\left\{\alpha_j^*, \theta_j^*\right\}_{j=0}^k, \{x(j), y(j)\}_{j=0}^k, \{\alpha_j, \theta_j\}_{j=0}^k\right) \\
&= \sum_{j=0}^k \left|y(j) - G^T\left(x(j), \theta_j^*\right)\alpha_j^*\right|^2 + \mu^{-1}\left\|\alpha_0^* - \alpha_{-1}\right\|^2 \\
&\quad + q^{-1}\sum_{j=0}^k \left\|\alpha_{j+1}^* - \alpha_j^*\right\|^2 + \mu_\theta^{-1}\sum_{j=0}^k \left\|\theta_j^* - \theta_{j-1}\right\|^2 \\
&\quad - \gamma^{-2}\sum_{j=0}^k \left|G^T\left(x(j), \theta_j^*\right)\alpha_j^* - G^T\left(x(j), \theta_j\right)\alpha_j\right|^2.
\end{aligned}$$

Noting $n_j = y(j) - G^T(x(j), \theta_j^*)\alpha_j^*$, it can be seen that the suboptimality condition (6) is satisfied, if and only if

$$J_k\left(\left\{\alpha_j^*, \theta_j^*\right\}_{j=0}^k, \{x(j), y(j)\}_{j=0}^k, \{\alpha_j, \theta_j\}_{j=0}^k\right) > 0$$

for any vector α_0^* and all causal sequences $(\{\alpha_{j+1}^* - \alpha_j^*\}_{j=0}^k, \{\theta_j^*, c\theta_j^* \geq h\}_{j=0}^k, \{x(j), y(j)\}_{j=0}^k)$. Therefore, any H^∞-suboptimal estimation strategy $\{\alpha_j, \theta_j, c\theta_j \geq h\}_{j=0}^k$ that achieves a robustness level of γ, for a given fixed data sequence $\{x(j), y(j)\}_{j=0}^k$, must ensure that

$$\min_{\alpha_0^*, \left\{\alpha_{j+1}^* - \alpha_j^*\right\}_{j=0}^k, \left\{\theta_j^*, c\theta_j^* \geq h\right\}_{j=0}^k} J_k > 0. \tag{7}$$

For a given parameter sequence $\{\theta_j^*, c\theta_j^* \geq h\}_{j=0}^k$, we define

$$\begin{aligned}
&J_k^{\min}\left(\left\{\theta_j^*\right\}_{j=0}^k, \{x(j), y(j)\}_{j=0}^k, \{\alpha_j, \theta_j\}_{j=0}^k\right) \\
&= \min_{\alpha_0^*, \left\{\alpha_{j+1}^* - \alpha_j^*\right\}_{j=0}^k} J_k\Big(\left\{\alpha_j^*, \theta_j^*\right\}_{j=0}^k, \{x(j), y(j)\}_{j=0}^k, \\
&\qquad \{\alpha_j, \theta_j\}_{j=0}^k\Big).
\end{aligned}$$

That is,

$$\begin{aligned}
J_k^{\min} &= \mu_\theta^{-1}\sum_{j=0}^k \left\|\theta_j^* - \theta_{j-1}\right\|^2 + \min_{\alpha_0^*, \left\{\alpha_{j+1}^* - \alpha_j^*\right\}_{j=0}^k} T_1 \\
T_1 &= \sum_{j=0}^k \left|y(j) - G^T\left(x(j), \theta_j^*\right)\alpha_j^*\right|^2 + \mu^{-1}\left\|\alpha_0^* - \alpha_{-1}\right\|^2 \\
&\quad - \gamma^{-2}\sum_{j=0}^k \left|G^T\left(x(j), \theta_j^*\right)\alpha_j^* - G^T\left(x(j), \theta_j\right)\alpha_j\right|^2 \\
&\quad + q^{-1}\sum_{j=0}^k \left\|\alpha_{j+1}^* - \alpha_j^*\right\|^2.
\end{aligned}$$

Now, any H^∞-suboptimal estimation strategy $\{\alpha_j, \theta_j, c\theta_j \geq h\}_{j=0}^k$, must ensure that

$$\min_{\left\{\theta_j^*, c\theta_j^* \geq h\right\}_{j=0}^k} J_k^{\min} > 0. \tag{8}$$

For solving the above formulated H^∞-suboptimal fuzzy estimation problem, we first need to find the functional value of $J_k^{\min}$ by making use of the results of state-space H^∞ estimation theory developed in the literature. Therefore, we first give a brief review of the results from [37].

Theorem 1: Consider a quadratic form

$$\begin{aligned}
&\mathcal{J}_k\left(x_0, \{u_j\}_{j=0}^k, \{y_j\}_{j=0}^k\right) = x_0^T \Pi_0^{-1} x_0 \\
&\quad + \sum_{j=0}^k \begin{bmatrix} u_j \\ y_j - H_j x_j \end{bmatrix}^* \begin{bmatrix} Q_j & S_j \\ S_j^T & R_j \end{bmatrix}^{-1} \begin{bmatrix} u_j \\ y_j - H_j x_j \end{bmatrix}
\end{aligned} \tag{9}$$

over x_0 and $\{u_j\}_{j=0}^k$, subject to the state-space constraints $x_{j+1} = F_j x_j + G_j u_j$, $j = 0, 1, \ldots, k$. If $\Pi_0 > 0$, $Q_j > 0$, R_j is invertible, $Q_j - S_j R_j^{-1} S_j^T > 0$ and $[F_j \; G_j]$ has a full rank for all j, then the quadratic forms (9) will have a unique minimum, if and only if

$$P_j^{-1} + H_j^T R_j^{-1} H_j > 0, \qquad 0 \leq j \leq k \tag{10}$$

where

$$\begin{aligned}
P_{j+1} &= F_j P_j F_j^T + G_j Q_j G_j^T - K_{p,j} R_{e,j} K_{p,j}^T \\
P_0 &= \Pi_0
\end{aligned} \tag{11}$$

$$\frac{\sum_{j=0}^k \left|G^T\left(x(j), \theta_j^*\right)\alpha_j^* - G^T\left(x(j), \theta_j\right)\alpha_j\right|^2}{\mu^{-1}\left\|\alpha_0^* - \alpha_{-1}\right\|^2 + \sum_{j=0}^k \left(q^{-1}\left\|\alpha_{j+1}^* - \alpha_j^*\right\|^2 + \mu_\theta^{-1}\left\|\theta_j^* - \theta_{j-1}\right\|^2 + |n_j|^2\right)} < \gamma^2 \tag{6}$$

with

$$R_{e,j} = R_j + H_j P_j H_j^T$$

and

$$K_{p,j} = \left(F_j P_j H_j^T + G_j S_j\right) R_{e,j}^{-1}.$$

It also follows in the minimum case that $P_{j+1} > 0$ for all $0 \leq j \leq k$. Also, the minimum value of $\mathcal{J}_k(x_0, \{u_j\}_{j=0}^k, \{y_j\}_{j=0}^k)$ over $(x_0, \{u_j\}_{j=0}^k)$ is given by

$$\sum_{j=0}^{k}(y_j - H_j\widehat{x}_j)^T R_{e,j}^{-1}(y_j - H_j\widehat{x}_j)$$

$$\widehat{x}_{j+1} = F_j\widehat{x}_j + K_{p,j}(y_j - H_j\widehat{x}_j), \qquad \widehat{x}_0 = 0. \quad (12)$$

Proof: This is the result of Theorem 6 and Lemma 13 in [37]. ■

The minimization problem

$$\min_{\alpha_0^*, \{\alpha_{j+1}^* - \alpha_j^*\}_{j=0}^k} T_1 \quad (13)$$

can be identified as a special case of (9) by considering for all $0 \leq j \leq k$, $\alpha_{-1} = 0$, $F_j = I$, $G_j = I$, $x_j = \alpha_j^*$, $\Pi_0 = \mu I$, $y_j = \begin{bmatrix} y(j) \\ G^T(x(j),\theta_j)\alpha_j \end{bmatrix}$, $H_j = \begin{bmatrix} G^T(x(j),\theta_j^*) \\ G^T(x(j),\theta_j^*) \end{bmatrix}$, $Q_j = qI, S_j = 0$, $R_j = \begin{bmatrix} 1 & 0 \\ 0 & -\gamma^2 \end{bmatrix}$. Let $\widehat{\alpha}_j$ denotes a vector that corresponds to a variable $\widehat{x}_j$ in (12). Here, $\widehat{\alpha}_j$ should not be confused with the estimation strategy α_j. Note that, it is known from the results of state-space estimation that $\widehat{\alpha}_j$ is an H^∞-optimal estimation of α_j^*. However, our concern is to find only the minimum value of the quadratic form and then to design an H^∞-optimal estimation strategy not only for α_j^* but also for θ_j^*. We can now apply (10), to check whether a minimum exists for (13). That is, $J_k(\{\alpha_j^*, \theta_j^*\}_{j=0}^k, \{x(j), y(j)\}_{j=0}^k, \{\alpha_j, \theta_j\}_{j=0}^k)$ will have a minimum over $\alpha_0^*, \{\alpha_{j+1}^* - \alpha_j^*\}_{j=0}^k$ for all $0 \leq j \leq k$, iff

$$\left(P_j^{-1} + \left[G\left(x(j),\theta_j^*\right) G\left(x(j),\theta_j^*\right)\right] \times \begin{bmatrix} 1 & 0 \\ 0 & -\gamma^2 \end{bmatrix}^{-1} \begin{bmatrix} G^T\left(x(j),\theta_j^*\right) \\ G^T\left(x(j),\theta_j^*\right) \end{bmatrix}\right) > 0 \quad (14)$$

where $P_0 = \mu I$ and

$$P_{j+1} = P_j + qI - P_j\left[G\left(x(j),\theta_j^*\right) G\left(x(j),\theta_j^*\right)\right] T_2^{-1} \times \begin{bmatrix} G^T\left(x(j),\theta^*\right) \\ G^T\left(x(j),\theta_j^*\right) \end{bmatrix} P_j$$

$$T_2 = \begin{bmatrix} 1 & 0 \\ 0 & -\gamma^2 \end{bmatrix} + \begin{bmatrix} G^T\left(x(j),\theta_j^*\right) \\ G^T\left(x(j),\theta_j^*\right) \end{bmatrix} \times P_j\left[G\left(x(j),\theta_j^*\right) G\left(x(j),\theta_j^*\right)\right].$$

Applying matrix inversion lemma, it implies that

$$P_{j+1} = \left[P_j^{-1} + (1-\gamma^{-2})G\left(x(j),\theta_j^*\right) G^T\left(x(j),\theta_j^*\right)\right]^{-1} + qI.$$

The existence condition according to (14) is given by

$$P_j^{-1} + (1-\gamma^{-2})G\left(x(j),\theta_j^*\right) G^T\left(x(j),\theta_j^*\right) > 0, \quad j = 0,\ldots,k. \quad (15)$$

A sufficient condition for the existence of minimum is given by following theorem.

Theorem 2: If $\gamma > 1$, then for the recursions

$$P_{j+1} = \left[P_j^{-1} + (1-\gamma^{-2})G\left(x(j),\theta_j^*\right) G^T\left(x(j),\theta_j^*\right)\right]^{-1} + qI$$

$P_0 = \mu I$, the inequality $P_j^{-1} + (1-\gamma^{-2})G(x(j),\theta_j^*) G^T(x(j),\theta_j^*) > 0$ holds good $\forall j > 0$.

Proof: This is a standard result. One possible proof is stated in Appendix. ■

We see that for $\gamma > 1$, the existence condition is satisfied and so the minimum value of function $J_k(\{\alpha_j^*, \theta_j^*\}_{j=0}^k, \{x(j), y(j)\}_{j=0}^k, \{\alpha_j, \theta_j\}_{j=0}^k)$ over $\alpha_0^*, \{\alpha_{j+1}^* - \alpha_j^*\}_{j=0}^k$ can be calculated using (12). For this, consider

$$R_{e,j} = \begin{bmatrix} 1 + G^T P_j G & G^T P_j G \\ G^T P_j G & -\gamma^2 + G^T P_j G \end{bmatrix}$$

where $G = G(x(j),\theta_j^*)$ has been written because of the space limitations. Using the block triangular factorization of $R_{e,j}$ and then finding its inverse, we have

$$R_{e,j}^{-1} = \begin{bmatrix} 1 & \frac{-G^T P_j G}{1+G^T P_j G} \\ 0 & 1 \end{bmatrix} T_3^{-1} \begin{bmatrix} 1 & 0 \\ \frac{-G^T P_j G}{1+G^T P_j G} & 1 \end{bmatrix}$$

$$T_3 = \begin{bmatrix} 1 + G^T P_j G & 0 \\ 0 & -\gamma^2 + G^T\left(P_j^{-1} + GG^T\right)^{-1} G \end{bmatrix}.$$

Now, the minimum value $J_k^{\min}(\{\theta_j^*\}_{j=0}^k, \{x(j), y(j)\}_{j=0}^k, \{\alpha_j, \theta_j\}_{j=0}^k)$ is

$$J_k^{\min} = \mu_\theta^{-1}\sum_{j=0}^{k}\left\|\theta_j^* - \theta_{j-1}\right\|^2 + \sum_{j=0}^{k} T_4^T R_{e,j}^{-1} T_4$$

$$T_4 = \begin{bmatrix} y(j) - G^T\left(x(j),\theta_j^*\right)\widehat{\alpha}_j \\ G^T\left(x(j),\theta_j\right)\alpha_j - G^T\left(x(j),\theta_j^*\right)\widehat{\alpha}_j \end{bmatrix}.$$

From Theorem 1, $\widehat{\alpha}_0 = 0$ and

$$\widehat{\alpha}_{j+1} = \widehat{\alpha}_j + P_j\left[G\left(x(j),\theta_j^*\right) \quad G\left(x(j),\theta_j^*\right)\right] R_{e,j}^{-1} T_4.$$

By substituting the value of $R_{e,j}^{-1}$, we have (16), shown at the bottom of the next page, and

$$\widehat{\alpha}_{j+1} = \widehat{\alpha}_j + \frac{P_j G\left(x(j),\theta_j^*\right)\left[y(j) - G^T\left(x(j),\theta_j^*\right)\widehat{\alpha}_j\right]}{1 + G^T\left(x(j),\theta_j^*\right) P_j G\left(x(j),\theta_j^*\right)} - \frac{P_j G\left(x(j),\theta_j^*\right)\left[G^T\left(x(j),\theta_j\right)\alpha_j - G^T\left(x(j),\theta_j^*\right)\overline{\alpha}_j\right]}{\left(1+G^T\left(x(j),\theta_j^*\right)P_j G\left(x(j),\theta_j^*\right)\right)\left(\gamma^2 - G^T\left(x(j),\theta_j^*\right)T_5\right)}$$

$$T_5 = \left[P_j^{-1} + G\left(x(j),\theta_j^*\right) G^T\left(x(j),\theta_j^*\right)\right]^{-1} G\left(x(j),\theta_j^*\right)$$

$$\overline{\alpha}_j = \widehat{\alpha}_j + \frac{P_j G\left(x(j),\theta_j^*\right)\left[y(j) - G^T\left(x(j),\theta_j^*\right)\widehat{\alpha}_j\right]}{1 + G^T\left(x(j),\theta_j^*\right) P_j G\left(x(j),\theta_j^*\right)}.$$

After calculating $J_k^{\min}$, we return to the original H^∞ fuzzy estimation problem (8). Therefore, all we have to do is to choose any causal estimation strategy $\{\alpha_j, \theta_j, c\theta_j \geq h\}_{j=0}^k$ that ensures $\min_{\{\theta_j^*, c\theta_j^* \geq h\}_{j=0}^k} J_k^{\min} > 0$, where $J_k^{\min}$ is given by (16) and the minimization is over causal sequence $\{\theta_j^*, c\theta_j^* \geq h\}_{j=0}^k$. Before we further minimize $J_k^{\min}$ with respect to the causal sequence of parameters $\{\theta_j^*, c\theta_j^* \geq h\}_{j=0}^k$, let us put a constraint on the estimation strategy that

$$\alpha_j = A_\alpha \overline{\alpha}_j \quad \theta_j = A_\theta \theta_j^*, \qquad j = 0, \dots, k$$

where A_α and A_θ are some matrices of suitable dimensions which operate on $\overline{\alpha}_j$ and θ_j^*, respectively, to define an estimation strategy (α_j, θ_j). Let the minimization of $J_k^{\min}$ with respect to $\{\theta_j^*, c\theta_j^* \geq h\}_{j=0}^k$, for a given choice of A_α and A_θ, be denoted by $\{\theta_{A_\theta}^{A_\alpha}(j)\}_{j=0}^k$. Let us consider an example of computing the causal sequence of parameters $\{\theta_{A_\theta}^{A_\alpha}(j)\}_{j=0}^k$ when $A_\alpha = I$ and $A_\theta = I$.

Example 1: When $A_\alpha = I$ and $A_\theta = I$, then

$$\left\{\theta_I^I(j)\right\}_{j=0}^k = \arg \min_{\left\{\theta_j^*, c\theta_j^* \geq h\right\}_{j=0}^k} \sum_{j=0}^k \Psi_j$$

$$\Psi_j = \frac{\left[y(j) - G^T\left(x(j), \theta_j^*\right)\widehat{\alpha}_j\right]^2}{1 + G^T\left(x(j), \theta_j^*\right) P_j G\left(x(j), \theta_j^*\right)} + \mu_\theta^{-1}\left\|\theta_j^* - \theta_{j-1}^*\right\|^2 \quad (17)$$

where θ_{-1}^* denotes the initial guess θ_{-1}

$$\widehat{\alpha}_{j+1} = \widehat{\alpha}_j + \frac{P_j G\left(x(j), \theta_j^*\right)\left[y(j) - G^T\left(x(j), \theta_j^*\right)\widehat{\alpha}_j\right]}{1 + G^T\left(x(j), \theta_j^*\right) P_j G\left(x(j), \theta_j^*\right)}$$

$$P_{j+1} = \left[P_j^{-1} + (1-\gamma^{-2}) G\left(x(j), \theta_j^*\right) G^T\left(x(j), \theta_j^*\right)\right]^{-1} + qI$$

$\widehat{\alpha}_0 = 0$, $P_0 = \mu I$. Since $\{\theta_I^I(j)\}_{j=0}^k$ is a causal sequence, therefore

$$\theta_I^I(0) = \arg\min_\theta \left[\Psi_0(\theta), c\theta \geq h\right]$$

$$\Psi_0(\theta) = \frac{\left[y(0) - G^T\left(x(0), \theta\right)\widehat{\alpha}_0\right]^2}{1 + G^T\left(x(0), \theta\right) P_0 G\left(x(0), \theta\right)} + \mu_\theta^{-1}\|\theta - \theta_{-1}\|^2$$

$\widehat{\alpha}_0 = 0$, $P_0 = \mu I$. Now, the value (and so the values $\widehat{\alpha}_1, P_1$) are fixed. Therefore, the estimation of $\theta_I^I 1$ follows as

$$\theta_I^I(1) = \arg\min_\theta \left[\Psi_1(\theta), c\theta \geq h\right]$$

$$\Psi_1(\theta) = \frac{\left[y(1) - G^T\left(x(1), \theta\right)\widehat{\alpha}_1\right]^2}{1 + G^T\left(x(1), \theta\right) P_1 G\left(x(1), \theta\right)} + \mu_\theta^{-1}\left\|\theta - \theta_I^I(0)\right\|^2$$

and so on follows the estimation of other parameters. Hence, the parameter sequence $\{\theta_I^I(j)\}_{j=0}^k$ can be recursively computed by solving $(k+1)$ minimization problems, i.e., for $j = 0, \dots, k$

$$\theta_I^I(j) = \arg\min_\theta \left[\Psi_j(\theta), c\theta \geq h\right] \quad (18)$$

$$\Psi_j(\theta) = \frac{\left[y(j) - G^T\left(x(j), \theta\right)\widehat{\alpha}_j\right]^2}{1 + G^T\left(x(j), \theta\right) P_j G\left(x(j), \theta\right)} + \mu_\theta^{-1}\left\|\theta - \theta_I^I(j-1)\right\|^2$$

$$\widehat{\alpha}_{j+1} = \widehat{\alpha}_j + \frac{P_j G\left(x(j), \theta_I^I(j)\right)\left[y(j) - G^T\left(x(j), \theta_I^I(j)\right)\widehat{\alpha}_j\right]}{1 + G^T\left(x(j), \theta_I^I(j)\right) P_j G\left(x(j), \theta_I^I(j)\right)} \quad (19)$$

$$P_{j+1} = qI + \left[P_j^{-1} + (1-\gamma^{-2}) G\left(x(j), \theta_I^I(j)\right) \times G^T\left(x(j), \theta_I^I(j)\right)\right]^{-1} \quad (20)$$

starting with $\theta_I^I(-1) = \theta_{-1}$, $\widehat{\alpha}_0 = 0$, and $P_0 = \mu I$.

Now, any causal estimation strategy $\{\alpha_j = A_\alpha \overline{\alpha}_j, \theta_j = A_\theta \theta_j^*\}_{j=0}^k$ is H^∞ suboptimal (i.e., achieves a robustness level of $\gamma > 1$) if (21), shown at the bottom of the next page, is satisfied. $\widehat{\alpha}_0 = 0$, $P_0 = \mu I$. There may exist different estimation strategies that satisfy the said H^∞-suboptimal condition [i.e., (21)]. One of such estimation strategies is to choose $A_\alpha = I$ (i.e., $\alpha_j = \overline{\alpha}_j$) and to define the operator A_θ in such a way that $\theta_j = A_\theta \theta_j^* = \theta_{A_\theta}^I(j)$. This results in

$$J_k^{\min}\left(\left\{\theta_{A_\theta}^I(j)\right\}_{j=0}^k, \{x(j), y(j)\}_{j=0}^k, \left\{\overline{\alpha}_j, \theta_{A_\theta}^I(j)\right\}_{j=0}^k\right) = \left(\sum_{j=0}^k \frac{\left[y(j) - G^T\left(x(j), \theta_{A_\theta}^I(j)\right)\widehat{\alpha}_j\right]^2}{1 + G^T\left(x(j), \theta_{A_\theta}^I(j)\right) P_j G\left(x(j), \theta_{A_\theta}^I(j)\right)} + \mu_\theta^{-1}\sum_{j=0}^k \left\|\theta_{A_\theta}^I(j) - \theta_{j-1}\right\|^2\right) > 0, \text{ since } P_j > 0$$

$$\widehat{\alpha}_{j+1} = \widehat{\alpha}_j + \frac{P_j G\left(x(j), \theta_{A_\theta}^I(j)\right)\left[y(j) - G^T\left(x(j), \theta_{A_\theta}^I(j)\right)\widehat{\alpha}_j\right]}{1 + G^T\left(x(j), \theta_{A_\theta}^I(j)\right) P_j G\left(x(j), \theta_{A_\theta}^I(j)\right)}$$

$$J_k^{\min}\left(\left\{\theta_j^*\right\}_{j=0}^k, \{x(j), y(j)\}_{j=0}^k, \{\alpha_j, \theta_j\}_{j=0}^k\right) = \sum_{j=0}^k \left(\frac{\left[y(j) - G^T\left(x(j), \theta_j^*\right)\widehat{\alpha}_j\right]^2}{1 + G^T\left(x(j), \theta_j^*\right) P_j G\left(x(j), \theta_j^*\right)} + \mu_\theta^{-1}\left\|\theta_j^* - \theta_{j-1}\right\|^2 - \frac{\left[G^T\left(x(j), \theta_j\right)\alpha_j - G^T\left(x(j), \theta_j^*\right)\overline{\alpha}_j\right]^2}{\gamma^2 - G^T\left(x(j), \theta_j^*\right)\left[P_j^{-1} + G\left(x(j), \theta_j^*\right) G^T\left(x(j), \theta_j^*\right)\right]^{-1} G\left(x(j), \theta_j^*\right)}\right) \quad (16)$$

$\widehat{\alpha}_0 = 0$, and $\overline{\alpha}_j = \widehat{\alpha}_{j+1}$. The choice of operator A_θ that satisfies

$$A_\theta \theta_j^* = \theta_{A_\theta}^I(j), \qquad \text{for } j = 0, \ldots, k$$

is still not clear. We have seen in example 1 that the causal parameter sequence $\{\theta_I^I(j)\}_{j=0}^k$ can be recursively computed using (18)–(20). Therefore, we motivate the choice $A_\theta = I$ by defining $\theta_j^* = \theta_I^I(j)$, i.e., we model the unknown process [see (5)] as

$$y(j) = G^T\left(x(j), \theta_I^I(j)\right)\alpha_j^* + n_j. \tag{22}$$

Please note that n_j here is also accommodating any mismatch between θ_j^* and $\theta_I^I(j)$. This modeling of the unknown process is justified by the definition of $\theta_I^I(j)$. To see this, note that when $\theta_j = \theta_I^I(j)$ and $\alpha_j = \overline{\alpha}_j$, then $\alpha_j = \widehat{\alpha}_{j+1}$. Then, it follows from (18)–(20) that

$$\theta_j = \arg\min_\theta\left[\Psi_j(\theta), c\theta \geq h\right] \tag{23}$$

$$\begin{aligned}
\Psi_j(\theta) &= \frac{\left[y(j) - G^T(x(j),\theta)\,\alpha_{j-1}\right]^2}{1 + G^T(x(j),\theta)\,P_j G(x(j),\theta)} + \mu_\theta^{-1}\|\theta - \theta_{j-1}\|^2\\
\alpha_j &= \alpha_{j-1} + \frac{P_j G(x(j),\theta_j)\left[y(j) - G^T(x(j),\theta_j)\,\alpha_{j-1}\right]}{1 + G^T(x(j),\theta_j)\,P_j G(x(j),\theta_j)}\\
P_{j+1} &= \left[P_j^{-1} + (1-\gamma^{-2})\,G(x(j),\theta_j)\,G^T(x(j),\theta_j)\right]^{-1} + qI
\end{aligned} \tag{24}$$

$\alpha_{-1} = 0$, $P_0 = \mu I$. Using the matrix inversion lemma, it can be seen that

$$P_{j+1} = P_j - \frac{P_j G(x(j),\theta_j)\,G^T(x(j),\theta_j)\,P_j}{(1-\gamma^{-2})^{-1} + G^T(x(j),\theta_j)\,P_j G(x(j),\theta_j)} + qI. \tag{25}$$

The definition of $\theta_I^I(j)$ given by (23) can be interpreted, as follows, as the solution of a least square-based adaptive-filtering problem. Given an initial guess $(\alpha_{j-1}, \theta_{j-1})$, we seek to improve upon θ_{j-1} by incorporating the additional information that is provided by the new data $(x(j), y(j))$, via solving a regularized nonlinear least square estimation problem (23). Thus, the modeling of the unknown process based on (22) makes intuitive engineering make sense. Hence, a H^∞-suboptimal estimation strategy that achieves a robustness level of $\gamma > 1$ consists of recursive computation of parameters $\{\alpha_j, \theta_j\}_{j=0}^k$ using (23)–(25).

VI. Least Square Estimation

Consider a least square optimization problem

$$\begin{aligned}
&\left[\alpha_0, \{\delta\alpha_j\}_{j=0}^k, \{\theta_j\}_{j=0}^k\right]\\
&\qquad = \arg\min_{\left[\alpha_0^*, \{\delta\alpha_j^*\}_{j=0}^k, \{\theta_j^*\}_{j=0}^k\right]}\left[J_k, \{c\theta_j^* \geq h\}_{j=0}^k\right]
\end{aligned}$$

$$\begin{aligned}
J_k = &\sum_{j=0}^k \left|y(j) - G^T\left(x(j),\theta_j^*\right)\alpha_j^*\right|^2 + \mu^{-1}\|\alpha_0^*\|^2\\
&+ q^{-1}\sum_{j=0}^k \|\delta\alpha_j^*\|^2 + \mu_\theta^{-1}\sum_{j=0}^k \|\theta_j^* - \theta_j^0\|^2
\end{aligned} \tag{26}$$

$$\begin{aligned}
&J_k^{\min}\left(\left\{\theta_{A_\theta}^{A_\alpha}(j)\right\}_{j=0}^k, \{x(j),y(j)\}_{j=0}^k, \{\alpha_j,\theta_j\}_{j=0}^k\right) > 0\\
&J_k^{\min}\left(\left\{\theta_{A_\theta}^{A_\alpha}(j)\right\}_{j=0}^k, \{x(j),y(j)\}_{j=0}^k, \{\alpha_j,\theta_j\}_{j=0}^k\right) = \mu_\theta^{-1}\sum_{j=0}^k\left\|\theta_{A_\theta}^{A_\alpha}(j) - \theta_{j-1}\right\|^2 + \sum_{j=0}^k \frac{\left[y(j) - G^T\left(x(j),\theta_{A_\theta}^{A_\alpha}(j)\right)\widehat{\alpha}_j\right]^2}{1 + G^T\left(x(j),\theta_{A_\theta}^{A_\alpha}(j)\right)P_j G\left(x(j),\theta_{A_\theta}^{A_\alpha}(j)\right)}\\
&\qquad - \sum_{j=0}^k \frac{\left[G^T(x(j),\theta_j)\,\alpha_j - G^T\left(x(j),\theta_{A_\theta}^{A_\alpha}(j)\right)\overline{\alpha}_j\right]^2}{\gamma^2 - G^T\left(x(j),\theta_{A_\theta}^{A_\alpha}(j)\right)T_6}\\
&T_6 = \left[P_j^{-1} + G\left(x(j),\theta_{A_\theta}^{A_\alpha}(j)\right)G^T\left(x(j),\theta_{A_\theta}^{A_\alpha}(j)\right)\right]^{-1} G\left(x(j),\theta_{A_\theta}^{A_\alpha}(j)\right)\\
&\widehat{\alpha}_{j+1} = \widehat{\alpha}_j + \frac{P_j G\left(x(j),\theta_{A_\theta}^{A_\alpha}(j)\right)\left[y(j) - G^T\left(x(j),\theta_{A_\theta}^{A_\alpha}(j)\right)\widehat{\alpha}_j\right]}{1 + G^T\left(x(j),\theta_{A_\theta}^{A_\alpha}(j)\right)P_j G\left(x(j),\theta_{A_\theta}^{A_\alpha}(j)\right)}\\
&\qquad - \frac{P_j G\left(x(j),\theta_{A_\theta}^{A_\alpha}(j)\right)\left[G^T(x(j),\theta_j)\,\alpha_j - G^T\left(x(j),\theta_{A_\theta}^{A_\alpha}(j)\right)\overline{\alpha}_j\right]}{\left(1 + G^T\left(x(j),\theta_{A_\theta}^{A_\alpha}(j)\right)P_j G\left(x(j),\theta_{A_\theta}^{A_\alpha}(j)\right)\right)\left(\gamma^2 - G^T\left(x(j),\theta_{A_\theta}^{A_\alpha}(j)\right)T_6\right)}\\
&\overline{\alpha}_j = \widehat{\alpha}_j + \frac{P_j G\left(x(j),\theta_{A_\theta}^{A_\alpha}(j)\right)\left[y(j) - G^T\left(x(j),\theta_{A_\theta}^{A_\alpha}(j)\right)\widehat{\alpha}_j\right]}{1 + G^T\left(x(j),\theta_{A_\theta}^{A_\alpha}(j)\right)P_j G\left(x(j),\theta_{A_\theta}^{A_\alpha}(j)\right)}\\
&P_{j+1} = qI + \left[P_j^{-1} + (1-\gamma^{-2})G\left(x(j),\theta_{A_\theta}^{A_\alpha}(j)\right)G^T\left(x(j),\theta_{A_\theta}^{A_\alpha}(j)\right)\right]^{-1}
\end{aligned} \tag{21}$$

subject to the constraints

$$\alpha^*_{j+1} = \alpha^*_j + \delta\alpha^*_j, \qquad j = 0, \dots, k. \tag{27}$$

Theorem 3: The estimation of parameters $\{\alpha^*_j, \theta^*_j\}_{j=0}^k$, say $\{\alpha_j, \theta_j\}_{j=0}^k$, based on the solution of constrained least square minimization problem (26) subject to the constraints (27), by taking $\theta^0_j = \theta_{j-1}$ (i.e., initial guess of θ^*_j is equal to the estimate of θ^*_{j-1}), follows for all $j = 0, \dots, k$ as

$$\theta_j = \arg\min_\theta \left[\Psi_j(\theta), c\theta \geq h\right] \tag{28}$$

$$\Psi_j(\theta) = \frac{\left[y(j) - G^T\left(x(j), \theta\right)\alpha_j\right]^2}{1 + G^T\left(x(j), \theta\right) P_j G\left(x(j), \theta\right)} + \mu_\theta^{-1}\|\theta - \theta_{j-1}\|^2$$

$$\alpha_{j+1} = \alpha_j + \frac{P_j G(x(j), \theta_j)\left[y(j) - G^T(x(j), \theta_j)\alpha_j\right]}{1 + G^T(x(j), \theta_j) P_j G(x(j), \theta_j)}, \quad \alpha_0 = 0 \tag{29}$$

$$P_{j+1} = P_j - \frac{P_j G(x(j), \theta_j)\, G^T(x(j), \theta_j) P_j}{1 + G^T(x(j), \theta_j) P_j G(x(j), \theta_j)} + qI, \quad P_0 = \mu I. \tag{30}$$

Proof: Define

$$J_k^{\min}\left(\{\theta^*_j\}_{j=0}^k, \{\theta^0_j\}_{j=0}^k\right) = \mu_\theta^{-1}\sum_{j=0}^k \|\theta^*_j - \theta^0_j\|^2 + \min_{\alpha^*_0, \{\delta\alpha^*_j\}_{j=0}^k} T_7\left(\alpha^*_0, \{\delta\alpha^*_j\}_{j=0}^k, \{\theta^*_j\}_{j=0}^k\right)$$

$$T_7\left(\alpha^*_0, \{\delta\alpha^*_j\}_{j=0}^k, \{\theta^*_j\}_{j=0}^k\right) = \mu^{-1}\|\alpha^*_0\|^2 + \sum_{j=0}^k \left|y(j) - G^T\left(x(j), \theta^*_j\right)\alpha^*_j\right|^2 + q^{-1}\sum_{j=0}^k \|\delta\alpha^*_j\|^2.$$

Now, the causal estimation strategy $\{\alpha_j, \theta_j\}_{j=0}^k$ is given as

$$\{\theta_j\}_{j=0}^k = \arg\min_{\{\theta^*_j\}_{j=0}^k} \left[J_k^{\min}, \{c\theta^*_j \geq h\}_{j=0}^k\right] \tag{31}$$

$$\left[\alpha_0, \{\delta\alpha_j\}_{j=0}^k\right] = \arg\min_{\alpha^*_0, \{\delta\alpha^*_j\}_{j=0}^k} T_7\left(\alpha^*_0, \{\delta\alpha^*_j\}_{j=0}^k, \{\theta_j\}_{j=0}^k\right),$$

$$\text{subject to } \alpha^*_{j+1} = \alpha^*_j + \delta\alpha^*_j, \qquad j = 0, \dots, k. \tag{32}$$

The functional value $J_k^{\min}(\{\theta^*_j\}_{j=0}^k, \{\theta^0_j\}_{j=0}^k)$ can be computed using Theorem 1. The minimization problem $\min_{\alpha^*_0, \{\delta\alpha^*_j\}_{j=0}^k} T_7(\alpha^*_0, \{\delta\alpha^*_j\}_{j=0}^k, \{\theta^*_j\}_{j=0}^k)$ can be identified a special case of quadratic form (9) by considering for all $0 \leq j \leq k$, $F_j = I, G_j = I, x_j = \alpha^*_j, \Pi_0 = \mu I, Q_j = qI, S_j = 0, R_j = 1, u_j = \delta\alpha^*_j, y_j = y(j), H_j = G^T(x(j), \theta^*_j)$. The existence condition (10) follows as

$$P_j^{-1} + G\left(x(j), \theta^*_j\right) G^T\left(x(j), \theta^*_j\right) > 0, \qquad 0 \leq j \leq k$$

$$P_{j+1} = P_j - \frac{P_j G\left(x(j), \theta^*_j\right) G^T\left(x(j), \theta^*_j\right) P_j}{1 + G^T\left(x(j), \theta^*_j\right) P_j G\left(x(j), \theta^*_j\right)} + qI, \qquad P_0 = \mu I$$

which, using matrix inversion lemma, can be written as

$$P_{j+1} = \left[P_j^{-1} + G\left(x(j), \theta^*_j\right) G^T\left(x(j), \theta^*_j\right)\right]^{-1} + qI, \quad P_0 = \mu I.$$

It can be seen using Theorem 2 for $\gamma = \infty$ that existence condition $P_j^{-1} + G(x(j), \theta^*_j) G^T(x(j), \theta^*_j) > 0$ holds good for $0 \leq j \leq k$. Let $\widehat{\alpha}_j$ denote a vector that corresponds to a variable $\widehat{x}_j$ in (12). Now, the minimum value according to (12) is given as

$$\sum_{j=0}^k \frac{\left[y(j) - G^T\left(x(j), \theta^*_j\right)\widehat{\alpha}_j\right]^2}{1 + G^T\left(x(j), \theta^*_j\right) P_j G\left(x(j), \theta^*_j\right)}$$

$$\widehat{\alpha}_{j+1} = \widehat{\alpha}_j + \frac{P_j G\left(x(j), \theta^*_j\right)\left[y(j) - G^T\left(x(j), \theta^*_j\right)\widehat{\alpha}_j\right]}{1 + G^T\left(x(j), \theta^*_j\right) P_j G\left(x(j), \theta^*_j\right)}, \quad \widehat{\alpha}_0 = 0 \tag{33}$$

and thus

$$J_k^{\min}\left(\{\theta^*_j\}_{j=0}^k, \{\theta^0_j\}_{j=0}^k\right) = \sum_{j=0}^k \frac{\left[y(j) - G^T\left(x(j), \theta^*_j\right)\widehat{\alpha}_j\right]^2}{1 + G^T\left(x(j), \theta^*_j\right) P_j G\left(x(j), \theta^*_j\right)} + \mu_\theta^{-1}\sum_{j=0}^k \|\theta^*_j - \theta^0_j\|^2.$$

Let us choose for all $j = 0, \dots, k$, $\theta^0_j = \theta_{j-1}$, where θ_{-1} is an initial guess about θ^*_0. Since the minimization in (31) is over a causal sequence, therefore, the parameters $\{\theta_j\}_{j=0}^k$ can be recursively computed as

$$\theta_j = \arg\min_\theta \left[\Psi_j(\theta), c\theta \geq h\right] \tag{34}$$

$$\Psi_j(\theta) = \frac{\left[y(j) - G^T\left(x(j), \theta\right)\widehat{\alpha}_j\right]^2}{1 + G^T(x(j), \theta) P_j G\left(x(j), \theta\right)} + \mu_\theta^{-1}\|\theta - \theta_{j-1}\|^2$$

$$\widehat{\alpha}_{j+1} = \widehat{\alpha}_j + \frac{P_j G(x(j), \theta_j)\left[y(j) - G^T(x(j), \theta_j)\widehat{\alpha}_j\right]}{1 + G^T(x(j), \theta_j) P_j G(x(j), \theta_j)}, \quad \widehat{\alpha}_0 = 0 \tag{35}$$

$$P_{j+1} = P_j - \frac{P_j G(x(j), \theta_j) G^T(x(j), \theta_j)\, P_j}{1 + G^T(x(j), \theta_j) P_j G(x(j), \theta_j)} + qI, \quad P_0 = \mu I. \tag{36}$$

Once the parameters $\{\theta_j\}_{j=0}^k$ have been computed using (34)–(36), the computation of consequent parameters follows from (32). The solution of (32) is given by the well-known extended recursive least squares (RLS) algorithm (see [39] and [40])

$$\alpha_{j+1} = \alpha_j + \frac{P_j G\left(x(j), \theta_j\right)\left[y(j) - G^T(x(j), \theta_j)\alpha_j\right]}{1 + G^T\left(x(j), \theta_j\right) P_j G(x(j), \theta_j)}, \quad \alpha_0 = 0$$

$$P_{j+1} = P_j - \frac{P_j G(x(j), \theta_j) G^T(x(j), \theta_j)\, P_j}{1 + G^T(x(j), \theta_j) P_j G(x(j), \theta_j)} + qI, \quad P_0 = \mu I.$$

We see that $\widehat{\alpha}_j = \alpha_j$ and thus the parameters $\{\alpha_j, \theta_j\}_{j=0}^k$ could be recursively computed as (28)–(30). ■

The use of least square estimation criterion is most common and has a long history in engineering fields [41]. However, we still would like to explore the robustness properties of least square estimation in a deterministic framework. For this, assume that the time-varying process could be modeled as

$$\alpha_{j+1}^* = \alpha_j^* + \delta\alpha_j^* \quad (37)$$

$$y(j) = G^T(x(j), \theta_j)\,\alpha_j^* + n_j \quad (38)$$

where θ_j is given by (28) and n_j accommodates any mismatch between θ_j^* and θ_j. If we take α_j as an estimate of α_j^* in (37) and (38), then an instantaneous estimation error is given as $e_s(j) = G^T(x(j), \theta_j)\alpha_j^* - G^T(x(j), \theta_j)\alpha_j$. Also, we have

$$y(j) - G^T(x(j), \theta_j)\,\alpha_j = e_s(j) + n_j. \quad (39)$$

Theorem 4: For a model of (37) and (38), the estimation of α_j^* using (29) and (30) ensures an upper bound on the value of energy gain from the unknown disturbances to the estimation errors, such that for all $\{\alpha_0^*, \{\delta\alpha_j^*\}_{j=0}^k, \{n_j\}_{j=0}^k\}$

$$\frac{\sum_{j=0}^k \left|G^T(x(j), \theta_j)\,\alpha_j^* - G^T(x(j), \theta_j)\,\alpha_j\right|^2}{\mu^{-1}\|\alpha_0^*\|^2 + q^{-1}\sum_{j=0}^k \|\delta\alpha_j^*\|^2 + \sum_{j=0}^k |n_j|^2} \leq (1+\sqrt{u})^2$$

$$u = 1 + \max_j G^T(x(j), \theta_j)\, P_j G(x(j), \theta_j).$$

Proof: To derive this result, we follow the approach of [42], where the H^∞ bounds for least square estimators have been derived. Consider a minimization problem

$$\min_{\{\alpha_0^*, \{\delta\alpha_j^*\}_{j=0}^k\}} D\left(\alpha_0^*, \{\delta\alpha_j^*\}_{j=0}^k, \{x(j), y(j), \theta_j\}_{j=0}^k\right)$$

$$D\left(\alpha_0^*, \{\delta\alpha_j^*\}_{j=0}^k, \{x(j), y(j), \theta_j\}_{j=0}^k\right) = \mu^{-1}\|\alpha_0^*\|^2 + q^{-1}\sum_{j=0}^k \|\delta\alpha_j^*\|^2 + \sum_{j=0}^k \left|y(j) - G^T(x(j), \theta_j)\alpha_j^*\right|^2$$

subject to $\alpha_{j+1}^* = \alpha_j^* + \delta\alpha_j^*$. The minimum value of this cost function, as calculated in Theorem 3, follows from (33) (by replacing θ_j^* with θ_j and noting that $\widehat{\alpha}_j = \alpha_j$) as $\sum_{j=0}^k([y(j) - G^T(x(j), \theta_j)\alpha_j]^2)/(1 + G^T(x(j), \theta_j)P_j G(x(j), \theta_j))$. Now, for any $(\alpha_0^*, \{\delta\alpha_j^*\}_{j=0}^k, \{x(j), y(j), \theta_j\}_{j=0}^k)$, we have

$$D \geq \sum_{j=0}^k \frac{\left[y(j) - G^T(x(j), \theta_j)\,\alpha_j\right]^2}{1 + G^T(x(j), \theta_j)\, P_j G(x(j), \theta_j)} \geq \frac{1}{u}\sum_{j=0}^k \left[y(j) - G^T(x(j), \theta_j)\,\alpha_j\right]^2.$$

Using (38) and (39), we have

$$\mu^{-1}\|\alpha_0^*\|^2 + q^{-1}\sum_{j=0}^k \|\delta\alpha_j^*\|^2 + \sum_{j=0}^k |n_j|^2 \geq \frac{1}{u}\sum_{j=0}^k [e_s(j) + n_j]^2. \quad (40)$$

Since $u > 1$ (due to $P_j > 0$), therefore

$$\left[\frac{e_s(j)}{\sqrt{1+\sqrt{u}}} + \left(\sqrt{1+\sqrt{u}}\right) n_j\right]^2 \geq 0$$

$$\frac{1}{1+\sqrt{u}}|e_s(j)|^2 + (1+\sqrt{u})|n_j|^2 + 2e_s(j)n_j \geq 0.$$

Therefore, $[e_s(j) + n_j]^2 \geq (\sqrt{u}/(1+\sqrt{u}))|e_s(j)|^2 - \sqrt{u}|n_j|^2$ and, thus, (40) results in

$$\mu^{-1}\|\alpha_0^*\|^2 + q^{-1}\sum_{j=0}^k \|\delta\alpha_j^*\|^2 + \sum_{j=0}^k |n_j|^2 \geq \frac{1}{\sqrt{u}}\left(\frac{1}{1+\sqrt{u}}\right)\sum_{j=0}^k |e_s(j)|^2 - \frac{1}{\sqrt{u}}\sum_{j=0}^k |n_j|^2.$$

Now, it can be seen using the above inequality that

$$\sum_{j=0}^k |e_s(j)|^2 \leq (1+\sqrt{u})^2 \times \left[\mu^{-1}\|\alpha_0^*\|^2 + q^{-1}\sum_{j=0}^k \|\delta\alpha_j^*\|^2 + \sum_{j=0}^k |n_j|^2\right].$$

Hence

$$\frac{\sum_{j=0}^k \left|G^T(x(j), \theta_j)\,\alpha_j^* - G^T(x(j), \theta_j)\,\alpha_j\right|^2}{\mu^{-1}\|\alpha_0^*\|^2 + q^{-1}\sum_{j=0}^k \|\delta\alpha_j^*\|^2 + \sum_{j=0}^k |n_j|^2} \leq (1+\sqrt{u})^2.$$

■

Remark 1: The RLS algorithms used in literature for the estimation of linear parameters can be seen as a particular case of recursions (29) and (30) with $q = 0$. For $q > 0$, it is a special case of extended RLS algorithm [40].

VII. Algorithm and Further Remarks

We present a Gauss–Newton method-based algorithm for the H^∞-suboptimal estimation of fuzzy-model parameters [i.e., expressions (23)–(25)] that could be modified for least square estimation [i.e., expressions (28)–(30)] without loss of any generality. The algorithm consists of the following steps.

1) Choose $\alpha_{-1} = 0, \theta_{-1}, c, h, \mu > 0, \mu_\theta > 0, q > 0, \gamma > 1, P_0 = \mu I$, and $j = 0$.
2) Define

$$r_H(\theta) = \begin{bmatrix} \left[\frac{[y(j) - G^T(x(j),\theta)\alpha_{j-1}]^2}{1 + G^T(x(j),\theta)P_j G(x(j),\theta)}\right]^{1/2} \\ \left[\mu_\theta^{-1}\right]^{1/2}(\theta - \theta_{j-1}) \end{bmatrix}$$

and let $s^*(\theta)$ be the unique solution of the following constrained optimization problem solved by the algorithm suggested in [43]:

$$s^*(\theta) = \arg\min_s \left[\|r_H(\theta) + r_H'(\theta)s\|^2 ; cs \geq h - c\theta\right]$$

where $r'_H(\theta)$ is the Jacobian matrix of vector r_H with respect to θ determined by the method of finite differences. The Jacobian $r'_H(\theta)$ is a full-rank matrix, as a result of using regularization.

3) Compute

$$\theta_j = \theta_{j-1} + s^*(\theta_{j-1})$$

$$\alpha_j = \alpha_{j-1} + \frac{P_j G(x(j),\theta_j)\left[y(j) - G^T(x(j),\theta_j)\,\alpha_{j-1}\right]}{1 + G^T(x(j),\theta_j)\,P_j G(x(j),\theta_j)}$$

$$P_{j+1} = P_j - \frac{P_j G(x(j),\theta_j)\,G^T(x(j),\theta_j)\,P_j}{(1-\gamma^{-2})^{-1} + G^T(x(j),\theta_j) P_j G(x(j),\theta_j)} + qI.$$

4) $j := j+1$ and go to step 2).

A matlab code was developed to implement the above algorithm. The constrained linear least square problem has been solved by transforming it first to a least distance programming (see [43] for details). To solve the least distance programming problem, a simple algorithm based on matlab routine "lsqnonneg," was developed. The average time in seconds required for the one pass of algorithm on a Pentium 2.53-GHz computer was approximately equal to (number of rules in fuzzy model/4000).

We have considered a deterministic framework to study the robust adaptive fuzzy-identification problem. The framework provides a possibility of developing different new robust adaptive model identification algorithms. To illustrate this, consider the following examples.

A. Robust Adaptive Identification of Semilinear Models

We have considered for simplicity a zero-order Takagi–Sugeno fuzzy model. However, it is straightforward to apply the developed techniques to the robust identification of any semilinear model with linear inequality constraints, e.g., first-order Takagi–Sugeno fuzzy models, radial basis function (RBF) neural networks, B-spline models, etc. Our approach is valid for any model characterized by parameter set Θ such that $\Theta = \Theta_l \oplus \Theta_n$, $y = G^T(x,\Theta_n)\Theta_l$, $c\Theta_n \geq h$.

B. Exponentially Windowed Adaptive Fuzzy Identification

method to account for time variations in the linear identification literature is to exponentially weight the data with a forgetting factor $0 < \lambda < 1$. A larger weight is given to more recent data than the earlier ones. To use an exponential window in least square fuzzy-model optimization, we solve

$$[\alpha_0, \{\delta\alpha_j\}, \{\theta_j\}] = \arg \min_{\left[\alpha_0^*, \{\delta\alpha_j^*\}, \{\theta_j^*\}\right]} \left[J_k, \{c\theta_j^* \geq h\}\right]$$

$$J_k = \sum_{j=0}^{k} \lambda^{k-j} \left|y(j) - G^T\left(x(j),\theta_j^*\right)\alpha_j^*\right|^2 + \mu^{-1}\lambda^{k+1}\left\|\alpha_0^*\right\|^2 + q^{-1}\sum_{j=0}^{k}\lambda^{k-j}\left\|\delta\alpha_j^*\right\|^2 + \mu_\theta^{-1}\sum_{j=0}^{k}\lambda^{k-j}\left\|\theta_j^* - \theta_j^0\right\|^2$$

subject to the constraints $\alpha_{j+1}^* = \alpha_j^* + \delta\alpha_j^*$. The solution in this case will be given as

$$\theta_j = \arg\min_\theta \left[\Psi_j(\theta), c\theta \geq h\right] \tag{41}$$

$$\Psi_j(\theta) = \frac{\left[y(j) - G^T(x(j),\theta)\,\alpha_j\right]^2}{1 + \lambda^{-1} G^T(x(j),\theta)\,\overline{P}_j G(x(j),\theta)} + \mu_\theta^{-1}\|\theta - \theta_{j-1}\|^2$$

$$\alpha_{j+1} = \alpha_j + \frac{\lambda^{-1}\overline{P}_j G(x(j),\theta_j)\left[y(j) - G^T(x(j),\theta_j)\alpha_j\right]}{1 + \lambda^{-1} G^T(x(j),\theta_j)\,\overline{P}_j G(x(j),\theta_j)} \tag{42}$$

$$\overline{P}_{j+1} = \frac{1}{\lambda}\left[\overline{P}_j - \frac{\lambda^{-1}\overline{P}_j G(x(j),\theta_j)\,G^T(x(j),\theta_j)\,\overline{P}_j}{1 + \lambda^{-1}G^T(x(j),\theta_j)\,\overline{P}_j G(x(j),\theta_j)}\right] + qI \tag{43}$$

$\alpha_0 = 0$, $\overline{P}_0 = \mu I$. Similarly, an exponential window could be used for H^∞-suboptimal fuzzy identification. Further, a variable-forgetting-factor strategy could be adopted for an automatic tradeoff between fast parameter tracking (small λ) and good noise suppression (large λ) [44].

VIII. Simulation Studies

The proposed approach to the adaptive learning of fuzzy models is applied to the examples of adaptive system identification, time-series prediction, and estimation of an uncertain process. The simulation studies consider: 1) H^∞-suboptimal estimation for $\gamma = 1.1$; 2) H^∞-suboptimal estimation for $\gamma = 2$; and 3) least squares estimation. Our methods are compared with the most commonly used gradient-descent technique for the adaptive learning of nonlinear parameters. A gradient-descent-learning law seeks to decrease the value of the objective function based on the instantaneous error $\mathrm{Er}(\alpha,\theta,k) = (1/2)[y(k) - G^T(x(k),\theta)\alpha]^2$. A gradient-descent-learning law, for estimating the parameter set $\Theta_k = [\alpha_k^T \theta_k^T]^T$, has the form $\Theta_k = \Theta_{k-1} - \mu(\partial\mathrm{Er}(\Theta,k)/\partial\Theta)_{\Theta_{k-1}}$, where μ is a step size. During the gradient-descent learning of membership functions, in the presence of disturbances, the knots may attempt to come close to (or even cross) one another, thus leading to a loss of interpretability and learning performance. Thus, for a better performance of the gradient-descent learning, the knots must be prevented from crossing one another by modifying the learning law as

$$\Theta_k = \begin{cases} \Theta_{k-1} - \mu\left(\frac{\partial \mathrm{Er}(\Theta,k)}{\partial\Theta}\right)_{\Theta_{k-1}}, & \text{if } c\theta_k \geq h \\ \Theta_{k-1}, & \text{otherwise.}\end{cases}$$

Let us first consider the fuzzy identification of a process described as

$$y = f(x,p) = \frac{-10x}{2p + x^2} + p^2\tanh(x), \qquad x \in [-0.5, 2.5] \tag{44}$$

where the parameter p is time varying. The process is simulated for $t = 0$ to $t = 40$ as $x(t) = -0.5 + |3\sin 10t|$, and parameter

p varies with time as

$$p(t) = \begin{cases} 1.5, & 0 \le t < 10 \\ 2, & 10 \le t < 20 \\ 1, & 20 \le t < 30 \\ 1.5, & 30 \le t < 40. \end{cases} \quad (45)$$

The identification data are generated by sampling the process with a sampling period of $T = 0.01$. The uncertain input–output identification data are obtained by generating the sequence $\{(1+\delta x_k)x(kT), (1+\delta y_k)y(kT)\}_{k=0,1,\ldots}$, where δx_k and δy_k are random entries, chosen from a normal distribution with zero mean and 0.01 variance. The instantaneous absolute estimation error (AE) at kth sampled time instant is defined as $\mathrm{AE}(k) = (1/300)\sum_{j=1}^{300} |f(x^j, p(0.01k)) - G^T(x^j,\theta_k)\alpha_k|$, where the points $\{x^j\}_{j=1}^{300}$ are uniformly distributed in $[-0.5,2.5]$. $\mathrm{AE}(k)$ indicates the instantaneous estimation performance at kth sampled time instant. The estimation performance, however, over a considered time span ($t = 0$ to $t = 40$) can be assessed by defining, e.g., the energy of estimation-error signal AE as $\sum_{k=0}^{4000} |\mathrm{AE}(k)|^2$. A smaller value of estimation-error energy indicates the smaller values of squared instantaneous AE during the considered time span and vice versa. A good estimation strategy corresponds to small values of $\{\mathrm{AE}(k)\}_{k=0}^{N}$ and, hence, a small value of estimation-error energy $\sum_{k=0}^{N} |\mathrm{AE}(k)|^2$ during the time span of $k = 0$ to $k = N$. Thus, the performances of different estimation strategies, over the considered time span, can be compared by computing their estimation-error energies.

Consider a fuzzy model with four Gaussian-shaped membership functions with an initial guess of $\theta_{-1} = [-0.5\ 0.5\ 1.5\ 2.5]$. We choose c and h in a way that two knots must be separated at least by a distance of 0.1. For a fair comparison of different techniques, same step-size $\mu = \mu_\theta = 0.05$ is taken for all. For H^∞-suboptimal and least square estimation, let us take $q = 0.001$. Fig. 1 shows the simulation results where $\mathrm{AE}(k)$ is plotted with time index k. The adaptation of knot sequence vector θ_k with time index k has been shown in Fig. 2. It was indicated in Fig. 2 that knot curves come close to one another but remain separated at least by a distance of 0.1 (as a result of interpretability constraints). Thus, in case the interpretability constraints are not put, the knot curves may cross one another resulting in a total loss of interpretability and performance. A comparison of different estimation strategies has been shown in Table I.

Consider a four-dimensional example to predict the future values of a chaotic time series. The time series is generated by simulating the chaotic Mackey-Glass differential delay equation, i.e., $dx/dt = (0.2x(t-17)/(1+x^{10}(t-17))) - 0.1x(t),\ x(0) = 1.2,\ x(t) = 0$, for $t < 0$. A fourth-order Runge–Kutta method was used for the simulation of the above equation. The aim of the problem is to predict the value of $x(t+6)$ by using a set of past values i.e., $x(t-18)$, $x(t-12)$, $x(t-6)$, $x(t)$. The uncertain input–output identification data are a sequence of 500 elements from $t = 118$ to $t = 617$, i.e.,

$$\left\{ \begin{bmatrix} (1+\delta x_{1k})x(t-18) \\ (1+\delta x_{2k})x(t-12) \\ (1+\delta x_{3k})x(t-6) \\ (1+\delta x_{4k})x(t) \end{bmatrix}, (1+\delta y_k)x(t+6) \right\}_{t=118,\ldots,617}$$

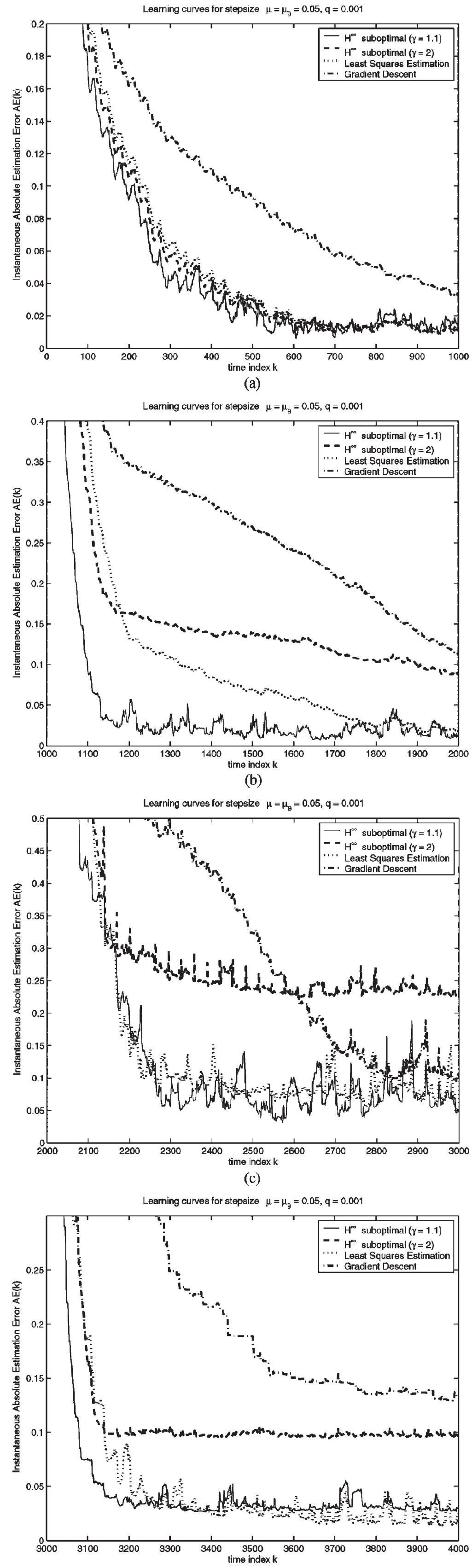

Fig. 1. Plot of instantaneous AE for a time-varying process. (a) Learning curves from $t = 0$ to $t = 10$. (b) Learning curves from $t = 10$ to $t = 20$. (c) Learning curves from $t = 20$ to $t = 30$. (d) Learning curves from $t = 30$ to $t = 40$.

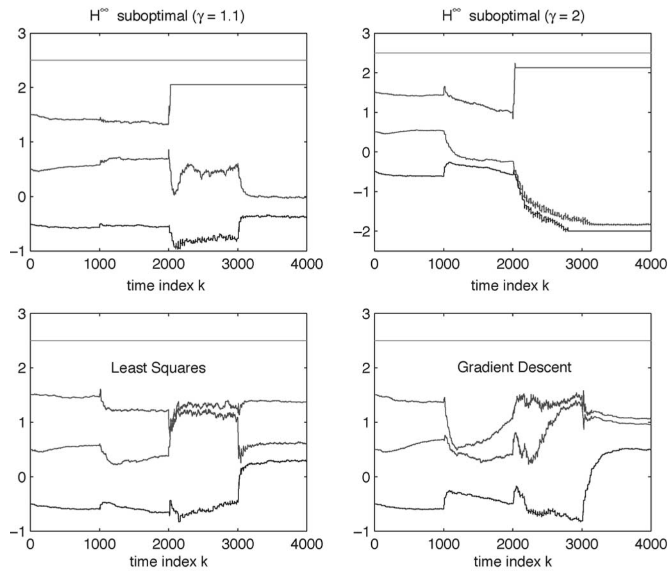

Fig. 2. Time variation of parameter θ_k for different estimation strategies.

TABLE I
COMPARISON OF DIFFERENT ESTIMATION STRATEGIES FOR THE FUZZY IDENTIFICATION OF A TIME-VARYING PROCESS

Estimation strategy	$\sum_{k=0}^{4000} \lvert AE(k)\rvert^2$
H^∞-suboptimal estimation for $\gamma = 1.1$	225.49
H^∞-suboptimal estimation for $\gamma = 2$	423.42
Least squares estimation	406.19
Gradient descent	783.36

TABLE II
COMPARISON OF DIFFERENT ESTIMATION STRATEGIES FOR A CHAOTIC TIME-SERIES PREDICTION

Estimation strategy	$\sum_{k=0}^{500} \lvert AE(k)\rvert^2$
H^∞-suboptimal estimation for $\gamma = 1.1$	69.50
H^∞-suboptimal estimation for $\gamma = 2$	74.55
Least squares estimation	76.94
Gradient descent	191.06

where δx_{1k}, δx_{2k}, δx_{3k}, δx_{4k}, and δy_k are random entries chosen from a normal distribution with zero mean and 0.01 variance. Let us choose the trapezoidal type of membership functions such that the number of membership functions assigned to each of the four inputs [i.e., $x(t-18)$, $x(t-12)$, $x(t-6)$, $x(t)$] is equal to three. Again, we choose c and h in such a way that two knots must be separated at least by a distance of 0.01. Define instantaneous AE at index k as $\mathrm{AE}(k) = (1/500)\sum_{t=618}^{1117} |x(t+6) - F_{\mathrm{s}}(k)|$, $F_{\mathrm{s}}(k) = G^T\left([x(t-18)\; x(t-12)\; x(t-6)\; x(t)]^T, \theta_k\right)\alpha_k$. We run the simulations from $k = 1$ to $k = 500$, taking $\mu = \mu_\theta = 0.01$, and $q = 0.001$. The simulation results have been shown in Fig. 3 by plotting $\mathrm{AE}(k)$ with k. Table II shows the comparison of different estimation strategies.

We have considered the fuzzy identification of a time-varying process (44), where the time variation of the process parameter p is according to (45). Now, we consider the case that the process parameter p is uncertain. That is, parameter p is different at every sampled time instant and takes random values. Let the uncertain process be described by the following system of unknown equations:

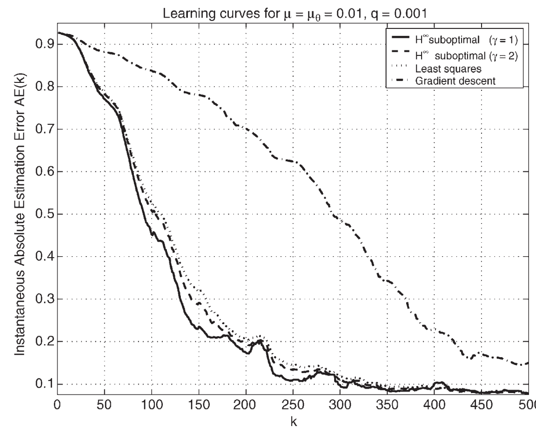

Fig. 3. Plot of instantaneous AE for a chaotic time series.

$$\xi_{k+1} = \begin{bmatrix} 0 & -0.5 \\ 1 & 1+0.3\delta_k \end{bmatrix}\xi_k + 0.02\begin{bmatrix} -6 \\ 1 \end{bmatrix} w_k, \quad \xi_0 = 0$$

$$p(k) = 1.5 + [\,0.2 \quad 0\,]\xi_k$$

$$y(k) = \frac{-10x(k)}{2p(k) + x^2(k)} + p^2(k)\tanh(x(k)) + 0.02v(k)$$

where δ_k, w_k, $v(k)$ are random entries chosen from a uniform distribution on the interval [−1,1], and $x(k)$ is chosen from a uniform distribution on the interval [−0.5,2.5]. The aim is to filter the noise $v(k)$ from measurement $y(k)$ to estimate the signal $s(k) = (-10x(k)/(2p(k) + x^2(k))) + p^2(k)\tanh(x(k))$, using a fuzzy model. Thus, the estimation error is given as $\mathrm{AE}(k) = |s(k) - G^T(x(k), \theta_k)\alpha_k|$. The problem of estimating the state vector ξ_k in the above system using a measurement signal $y_k = C\xi_k + Dv_k$, given the knowledge of system equations and (C, D), has been studied in the literature of robust filtering (see, e.g., [45]). Our concern, however, is to estimate a nonlinear function of ξ_k using $\{x(j), y(j)\}_{j=0}^{k}$, but without any knowledge of system equations. Again, we consider a fuzzy model with four Gaussian-shaped membership functions with an initial guess of $\theta_{-1} = [-0.5\ 0.5\ 1.5\ 2.5]$. The interpretability constraints are so chosen that the two knots must be separated at least by a distance of 0.1. Let us take $\mu = \mu_\theta = 0.9$. The process is simulated from $k = 0$ to $k = 1000$ at different values of q ranging from $q = 0.01$ to $q = 10$. For comparing the different estimation strategies, the estimation-error energy $\sum_{k=0}^{1000} |\mathrm{AE}(k)|^2$ has been computed and plotted in Fig. 4 for different values of q. Fig. 4 shows that, at a given value of step-s ize μ, robust techniques perform better at higher values of q. This, as expected, is due to the fact that q reflects *a priori* knowledge as to how fast the process parameters vary with time. Therefore, for fast varying processes, a higher value of q should be taken and vice versa.

The above simulation results (Tables I and II and Fig. 4) have clearly shown that the proposed learning methods of fuzzy models in a deterministic framework have a better performance than the gradient descent in the presence of data uncertainties and modeling errors. The better performance of H^∞-suboptimal (corresponding to $\gamma = 1.1$) estimation strategy

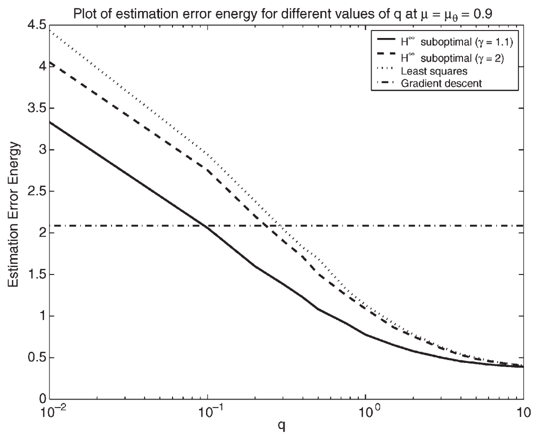

Fig. 4. Plot of estimation-error energy of different estimation strategies for an uncertain process.

is obviously due to the robust nature of H^∞ optimization theory, since H^∞ fuzzy estimation strategy has been derived via bounding the energy gain (from disturbances to estimation errors) by a value γ. The performance of an estimation strategies in a time-varying environment can be further improved by using an exponential window. As an illustration, we study the least square estimation strategy (41)–(43) with $\lambda = 0.99$ for the above defined adaptive fuzzy-identification problem. The estimation-error energy in this case, as expected, is lowered from 406.19 ($\lambda = 1$) to 353.13 ($\lambda = 0.99$).

IX. Conclusion

This study has outlined a deterministic approach to the robust adaptive fuzzy identification of time-varying processes based on H^∞-optimization and least square estimation criterion. The main contributions are as follows.

1) Unlike many *ad hoc* approaches to the learning of fuzzy models, this study rests upon a mathematical basis for addressing the issues of the data uncertainties, modeling errors, and time variations.
2) The study proposes new fuzzy learning algorithms that are robust against data uncertainties, modeling errors, and time variations.
3) The study offers a new opening to know how to develop the robust adaptive fuzzy-identification algorithms in a "rigorous" manner by extending the robust linear estimation theory to the nonlinear interpretable fuzzy models.

Appendix

Define $\hat{P}_j = [P_j^{-1} + (1-\gamma^{-2})G(x(j),\theta_j^*)G^T(x(j),\theta_j^*)]^{-1}$ and consider the minimum eigenvalue of matrix $(\hat{P}_{j-1})^{-1}$. To do this, use the fact that $\lambda_{\min}(\hat{P}_{j-1}^{-1}) = \min_{u^Tu=1}\ u^T\hat{P}_{j-1}^{-1}u$. Therefore, for every j, we can determine a vector $u_0 \ \epsilon\ R^K$ with $u_0^Tu_0 = 1$, such that $\lambda_{\min}(\hat{P}_{j-1}^{-1}) = u_0^T\hat{P}_{j-1}^{-1}u_0$. Substituting $\hat{P}_{j-1}^{-1}$, we have $\lambda_{\min}(\hat{P}_{j-1}^{-1}) = u_0^TP_{j-1}^{-1}u_0 + u_0^T(1-\gamma^{-2})$ $G(x(j-1),\theta_{j-1}^*)G^T(x(j-1),\theta_{j-1}^*)u_0$. If we assume that $\gamma > 1$ and $P_{j-1}^{-1} > 0$, then $u_0^TP_{j-1}^{-1}u_0 > 0$ and hence

$$\lambda_{\min}\left(\hat{P}_{j-1}^{-1}\right) > (1-\gamma^{-2})\left|G^T\left(x(j-1),\theta_{j-1}^*\right)u_0\right|^2 > 0.$$

The above inequality implies that

$$\lambda_{\max}(\hat{P}_{j-1}) < \frac{1}{(1-\gamma^{-2})\left|G^T\left(x(j-1),\theta_{j-1}^*\right)u_0\right|^2}$$

and therefore

$$\frac{1}{q+\lambda_{\max}(\hat{P}_{j-1})} > \frac{(1-\gamma^{-2})\left|G^T\left(x(j-1),\theta_{j-1}^*\right)u_0\right|^2}{q(1-\gamma^{-2})\left|G^T\left(x(j-1),\theta_{j-1}^*\right)u_0\right|^2+1}.$$

Since $P_j = \hat{P}_{j-1} + qI$, therefore $\lambda_{\min}(P_j^{-1}) = 1/\lambda_{\max}(\hat{P}_{j-1} + qI) = 1/(q+\lambda_{\max}(\hat{P}_{j-1}))$. Thus

$$\lambda_{\min}\left(P_j^{-1}\right) > \frac{(1-\gamma^{-2})\left|G^T\left(x(j-1),\theta_{j-1}^*\right)u_0\right|^2}{q(1-\gamma^{-2})\left|G^T\left(x(j-1),\theta_{j-1}^*\right)u_0\right|^2+1}. \tag{46}$$

For any nonzero vector $\alpha \in R^K$, consider

$$\begin{aligned}\alpha^T&\left[P_j^{-1} + (1-\gamma^{-2})G\left(x(j),\theta_j^*\right)G^T\left(x(j),\theta_j^*\right)\right]\alpha\\ &= \alpha^TP_j^{-1}\alpha + (1-\gamma^{-2})\left|G^T\left(x(j),\theta_j^*\right)\alpha\right|^2\\ &\geq \lambda_{\min}\left(P_j^{-1}\right)\|\alpha\|^2 + (1-\gamma^{-2})\left|G^T\left(x(j),\theta_j^*\right)\alpha\right|^2.\end{aligned}$$

Using (46), it follows that $\alpha^T[P_j^{-1} + (1-\gamma^{-2})G(x(j),\theta_j^*)$ $G^T(x(j),\theta_j^*)]\alpha > ((1-\gamma^{-2}|G^T(x(j-1),\theta_{j-1}^*)u_0|^2\|\alpha\|^2)/$ $(q(1-\gamma^{-2})|G^T(x(j-1),\theta_{j-1}^*)u_0|^2+1)) + (1-\gamma^{-2})|G^T$ $(x(j),\theta_j^*)\alpha|^2$ and, therefore

$$\alpha^T\left[P_j^{-1} + (1-\gamma^{-2})G\left(x(j),\theta_j^*\right)G^T\left(x(j),\theta_j^*\right)\right]\alpha > 0$$

for all nonzero $\alpha \in R^K$. This implies that

$$P_j^{-1} + (1-\gamma^{-2})G\left(x(j),\theta_j^*\right)G^T\left(x(j),\theta_j^*\right) > 0.$$

Thus, we see that if $P_{j-1}^{-1} > 0$ for $\gamma > 1$, then $P_j^{-1} + (1-\gamma^{-2})G(x(j),\theta_j^*)G^T(x(j),\theta_j^*) > 0$ and $P_j^{-1} > 0$ [from (46)]. Since $P_0^{-1} > 0$, therefore, by induction, the theorem is proven.

References

[1] L. X. Wang and J. M. Mendel, "Generating fuzzy rules by learning from examples," *IEEE Trans. Syst., Man, Cybern.*, vol. 22, no. 6, pp. 1414–1427, Nov./Dec. 1992.

[2] K. Nozaki, H. Ishibuchi, and H. Tanaka, "A simple but powerful heuristic method for generating fuzzy rules from numerical data," *Fuzzy Sets Syst.*, vol. 86, no. 3, pp. 251–270, Mar. 1997.

[3] J. J. Shan and H. C. Fu, "A fuzzy neural network for rule acquiring on fuzzy control systems," *Fuzzy Sets Syst.*, vol. 71, no. 3, pp. 345–357, May 1995.

[4] D. Nauck and R. Kruse, "A neuro-fuzzy approach to obtain interpretable fuzzy systems for function approximation," in *Proc. IEEE Int. Conf. Fuzzy Systems (FUZZ-IEEE)*, Anchorage, AK, May 1998, pp. 1106–1111.

[5] J.-S. R. Jang, "ANFIS: Adaptive-network-based fuzzy inference systems," *IEEE Trans. Syst., Man, Cybern.*, vol. 23, no. 3, pp. 665–685, May 1993.

[6] P. Thrift, "Fuzzy logic synthesis with genetic algorithms," in *Proc. 4th Int. Conf. Genetic Algorithms*, 1991, pp. 509–513.

[7] J. Liska and S. S. Melsheimer, "Complete design of fuzzy logic systems using genetic algorithms," in *Proc. 3rd IEEE Int. Conf. Fuzzy Systems*, 1994, pp. 1377–1382.

[8] F. Herrera, M. Lozano, and J. Verdegay, "Generating fuzzy rules from examples using genetic algorithms," in *Proc. 5th Int. Conf. Information Processing and Management Uncertainty Knowledge-Based Systems (IPMU)*, Paris, France, Jul. 1994, pp. 675–680.

[9] A. González and R. Peréz, "Completeness and consistency conditions for learning fuzzy rules," *Fuzzy Sets Syst.*, vol. 96, no. 1, pp. 37–51, May 1998.

[10] R. Babuška and H. Verbruggen, "Constructing fuzzy models by product space clustering," in *Fuzzy Model Identification: Selected Approaches*, H. Hellendoorn and D. Driankov, Eds. Berlin, Germany: Springer-Verlag, 1997, pp. 53–90.

[11] R. Babuška, *Fuzzy Modeling for Control*. Boston, MA: Kluwer, 1998.

[12] J. Abonyi, R. Babuška, and F. Szeifert, "Modified Gath–Geva fuzzy clustering for identification of Takagi–Sugeno fuzzy models," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 32, no. 5, pp. 612–621, Oct. 2002.

[13] D. Simon, "Design and rule base reduction of a fuzzy filter for the estimation of motor currents," *Int. J. Approx. Reason.*, vol. 25, no. 2, pp. 145–167, Oct. 2000.

[14] ——, "Training fuzzy systems with the extended Kalman filter," *Fuzzy Sets Syst.*, vol. 132, no. 2, pp. 189–199, Dec. 2002.

[15] J.-S. R. Jang, C. T. Sun, and E. Mizutani, *Neuro-Fuzzy and Soft Computing; A Computational Approach to Learning and Machine Intelligence*. Upper Saddle River, NJ: Prentice-Hall, 1997.

[16] D. S. Chen and R. C. Jain, "A robust back propagation learning algorithm for function approximation," *IEEE Trans. Neural Netw.*, vol. 5, no. 3, pp. 467–479, May 1994.

[17] W. Y. Wang, T. T. Lee, C. L. Liu, and C. H. Wang, "Function approximation using fuzzy neural networks with robust learning algorithm," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 27, no. 4, pp. 740–747, Sep. 1997.

[18] W. Yu and X. Li, "Fuzzy identification using fuzzy neural networks with stable learning algorithms," *IEEE Trans. Fuzzy Syst.*, vol. 12, no. 3, pp. 411–420, Jun. 2004.

[19] M. Burger, H. Engl, J. Haslinger, and U. Bodenhofer, "Regularized data-driven construction of fuzzy controllers," *J. Inverse Ill-Posed Probl.*, vol. 10, no. 4, pp. 319–344, 2002.

[20] T. Johansen, "Robust identification of Takagi–Sugeno–Kang fuzzy models using regularization," in *Proc. IEEE Conf. Fuzzy Systems*, New Orleans, LA, 1996, pp. 180–186.

[21] M. Kumar, R. Stoll, and N. Stoll, "Regularized adaptation of fuzzy inference systems. Modelling the opinion of a medical expert about physical fitness: An application," *Fuzzy Optim. Decis. Mak.*, vol. 2, no. 4, pp. 317–336, Dec. 2003.

[22] Y. G. Leu, T. T. Lee, and W. Y. Wang, "Observer-based adaptive fuzzy neural control for unknown nonlinear dynamical systems," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 29, no. 5, pp. 583–591, Oct. 1999.

[23] L. X. Wang, *Adaptive Fuzzy Systems and Control*. Englewood Cliffs, NJ: Prentice-Hall, 1994.

[24] P. A. Ioannou and J. Sun, *Robust Adaptive Control*. Englewood Cliffs, NJ: Prentice-Hall, 1996.

[25] W. Y. Wang, Y. G. Leu, and C. C. Hsu, "Robust adaptive fuzzy-neural control of nonlinear dynamical systems using generalized projection updated law and variable structure controller," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 31, no. 1, pp. 140–147, Feb. 2001.

[26] T. Takagi and M. Sugeno, "Fuzzy identification of systems and its applications to modeling and control," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-15, no. 1, pp. 116–132, Jan. 1985.

[27] R. Babuška, "Construction of fuzzy systems-interplay between precision and transparency," in *Proc. ESIT*, Aachen, Germany, 2000, pp. 445–452.

[28] U. Bodenhofer and P. Bauer, "Towards an axiomatic treatment of 'interpretability'," in *Proc. IIZUKA*, Iizuka, Japan, Oct. 2000, pp. 334–339.

[29] J. Espinosa and J. Vandewalle, "Constructing fuzzy models with linguistic integrity from numerical data-AFRELI algorithm," *IEEE Trans. Fuzzy Syst.*, vol. 8, no. 5, pp. 591–600, Oct. 2000.

[30] M. Setnes, R. Babuška, and H. B. Verbruggen, "Rule-based modeling: Precision and transparency," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 28, no. 1, pp. 165–169, Feb. 1998.

[31] P. Lindskog, "Fuzzy identification from a grey box modeling point of view," in *Fuzzy Model Identification: Selected Approaches*, H. Hellendoorn and D. Driankov, Eds. Berlin, Germany: Springer-Verlag, 1997.

[32] M. Kumar, R. Stoll, and N. Stoll, "Robust adaptive fuzzy identification of time-varying processes with uncertain data. Handling uncertainties in the physical fitness fuzzy approximation with real world medical data: An application," *Fuzzy Optim. Decis. Mak.*, vol. 2, no. 3, pp. 243–259, Sep. 2003.

[33] ——, "Robust solution to fuzzy identification problem with uncertain data by regularization. Fuzzy approximation to physical fitness with real world medical data: An application," *Fuzzy Optim. Decis. Mak.*, vol. 3, no. 1, pp. 63–82, Mar. 2004.

[34] ——, "Robust adaptive identification of fuzzy systems with uncertain data," *Fuzzy Optim. Decis. Mak.*, vol. 3, no. 3, pp. 195–216, Sep. 2004.

[35] T. L. Vincent and P. P. Khargonekar, "A class of nonlinear filtering problems arising from drifting sensor gains," *IEEE Trans. Autom. Control*, vol. 44, no. 3, pp. 509–520, Mar. 1999.

[36] C. V. Rao, J. B. Rawlings, and D. Q. Mayne, "Constrained state estimation for nonlinear discrete-time systems: Stability and moving horizon approximations," *IEEE Trans. Autom. Control*, vol. 48, no. 2, pp. 246–258, Feb. 2003.

[37] B. Hassibi, A. H. Sayed, and T. Kailath, "Linear estimation in Krein spaces—Part I: Theory," *IEEE Trans. Autom. Control*, vol. AC-41, no. 1, pp. 18–33, Jan. 1996.

[38] ——, "Linear estimation in Krein spaces—Part II: Applications," *IEEE Trans. Autom. Control*, vol. AC-41, no. 1, pp. 34–49, Jan. 1996.

[39] S. Haykin, A. H. Sayed, J. Zeidler, P. Yee, and P. Wei, "Adaptive tracking of linear time-variant systems by extended RLS algorithms," *IEEE Trans. Signal Process.*, vol. 45, no. 5, pp. 1118–1128, May 1997.

[40] A. H. Sayed, *Fundamentals of Adaptive Filtering*. New York: Wiley, 2003.

[41] A. P. Sage, *Optimum Systems Control*. Englewood Cliffs, NJ: Prentice-Hall, 1968.

[42] B. Hassibi and T. Kailath, "H^{∞} bounds for least-squares estimators," *IEEE Trans. Autom. Control*, vol. 46, no. 2, pp. 309–314, Feb. 2001.

[43] C. L. Lawson and R. J. Hanson, *Solving Least Squares Problems*. Philadelphia, PA: SIAM, 1995.

[44] T. R. Fortesque, L. S. Kershenbaum, and B. E. Ydstie, "Implementation of self tuning regulators with variable forgetting factors," *Automatica*, vol. 17, no. 6, pp. 831–835, 1981.

[45] L. E. Ghaoui and G. Calafiore, "Robust filtering for discrete-time systems with bounded noise and parametric uncertainty," *IEEE Trans. Autom. Control*, vol. 46, no. 7, pp. 1084–1089, Jul. 2001.

Mohit Kumar (A'05) received the B.Tech. degree in electrical engineering from the National Institute of Technology, Hamirpur, India, in 1999, the M.Tech. degree in control engineering from the Indian Institute of Technology, Delhi, in 2001, and the Ph.D. degree in electrical engineering from Rostock University, Rostock, Germany, in 2004.

His research interests include robust adaptive fuzzy identification, fuzzy logic in medicine, and robust adaptive control.

Regina Stoll received the diploma in medicine (Dip.-Med.), the degree of "Dr. Med." in occupational medicine, and the degree of "Dr. Med.habil" in occupational and sports medicine from Rostock University, Rostock, Germany, in 1980, 1984, and 2002, respectively.

She is the Head of the Institute of Occupational and Social Medicine, Rostock. Her research interests include occupational physiology, preventive medicine, and cardiopulmonary diagnostics.

Norbert Stoll received the diploma (Dip.-Ing.) in automation engineering, and the Ph.D. degree in measurement technology from Rostock University, Rostock, Germany, in 1979 and 1985, respectively.

Since 1994, he has been a Professor in measurement technology in the engineering faculty of Rostock University. His fields of interests include medical process measurement, lab automation, and smart systems and devices.