Available online at www.sciencedirect.com
SCIENCE DIRECT
automatica
Automatica 42 (2006) 711–721
www.elsevier.com/locate/automatica

An energy-gain bounding approach to robust fuzzy identification[☆]

Mohit Kumar[a,*], Norbert Stoll[b], Regina Stoll[c]

[a]Center for Life Science Automation, F.-Barnewitz-Str. 8, D-18119 Rostock, Germany
[b]Institute of Automation, Richard-Wagner-Str. 31, D-18119 Rostock-Warnemünde, Germany
[c]Institute of Occupational and Social Medicine, St.-Georg-Str. 108, D-18055 Rostock, Germany

Received 15 August 2005; received in revised form 19 October 2005; accepted 15 January 2006
Available online 3 March 2006

Abstract

A novel method for the robust identification of interpretable fuzzy models, based on the criterion that identification errors are least sensitive to data uncertainties and modelling errors, is suggested. The robustness of identification errors towards unknown disturbances (data uncertainties, modelling errors, etc.) is achieved by bounding (i.e. minimizing) the maximum possible value of energy-gain from disturbances to the identification errors. The solution of energy-gain bounding problem, being robust, shows an improved performance of the identification method. The flexibility of the proposed framework is shown by designing the variable learning rate identification algorithms in both deterministic and stochastic frameworks.

Keywords: Energy-gain; Robust fuzzy identification; H^{∞}-optimization; Interpretability

1. Introduction

During the last years, a large number of fuzzy identification techniques have been developed using ad hoc approaches, neural networks, genetic algorithms, clustering techniques, and Kalman filtering. The robustness issue of fuzzy identification has been addressed previously by Chen and Jain (1994), Hong, Harris, and Chen (2004), Johansen (1996), Kumar, Stoll, and Stoll (2003b, c, 2004a, b, 2006), Wang, Lee, Liu, and Wang (1997), Yu and Li (2004). However, there is still quite a need of studying robust fuzzy identification problem so as to meet all of the following requirements:

(1) Considering fuzzy identification to be an ill-posed problem (Burger, Engl, Haslinger, & Bodenhofer, 2002), the issue of data uncertainties and modelling errors should be mathematically taken into account, while identifying not only linear parameters (consequents) but also the nonlinear parameters (antecedents) of the fuzzy model.
(2) The identification procedure should be on-line.
(3) The identification procedure should not require a priori knowledge of upper bounds, statistics, and distribution of uncertainties and modelling errors.
(4) The identification procedure should preserve the interpretability (a key property) of the fuzzy models.

To the knowledge of author, simultaneously all of above requirements (in particular the first requirement) have not been met in the literature. Regularization has been suggested in Johansen (1996) for robust identification of linear fuzzy model parameters and Hong et al. (2004) suggests a regularized orthogonal least-squares algorithm combined with a D-optimality used for subspace-based rule selection for a linear-in-parameters fuzzy model. The identification of both linear and nonlinear fuzzy parameters from uncertain data together with interpretability consideration has been suggested in Burger et al. (2002), Kumar et al. (2004b) using regularization, and in Kumar et al. (2003c) using semidefinite programming (SDP) and second-order cone programming (SOCP). However,

---

[☆] This paper was not presented at any IFAC meeting. This paper was recommended for publication in revised form by Associate Editor John Schoukens under the direction of Editor Torsten Soederstroem.

[*] Corresponding author. Tel.: +49 381 4949956; fax: +49 381 4949952.
E-mail addresses: mohit.kumar@uni-rostock.de (M. Kumar), norbert.stoll@uni-rostock.de (N. Stoll), regina.stoll@uni-rostock.de (R. Stoll).

0005-1098/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.automatica.2006.01.013

their methods are off-line. Yu and Li (2004) proposes an on-line method for robust fuzzy identification based on *input-to-state* stability approach that does not require any priori knowledge of upper bounds, however, without addressing interpretability issue. Both, interpretability and robustness of an on-line fuzzy identification method, has been achieved in Kumar et al. (2003, 2006) by solving a min–max estimation problem. Although, the approach of Kumar et al. (2006) meets all of the four requirements, but the results in Kumar et al. (2006) have been derived using approximations. The aim of this paper is to provide a novel fuzzy identification method based on a robust criterion. Our method, for each of four requirements, suggests a mathematical criterion and all the criteria are studied in a unified framework. The main features of our approach can be summarized as follows:

(1) To design a robust identification method, we try to minimize the maximum possible value of energy-gain from disturbances to the identification errors.
(2) The maximum value of energy-gain (that will be minimized) is calculated over all possible finite disturbances without making any assumptions.
(3) An on-line solution of energy-gain-based fuzzy identification problem in a closed-form has been made possible by modelling the unknown process using a time-varying antecedents fuzzy model.
(4) The interpretability issue is addressed by constraining the identification of membership functions to a matrix inequality (Burger et al., 2002;Kumar et al., 2003a, Kumar, Stoll, & Stoll, 2003b, 2004a, b, 2006).

The first three features of our approach are new in context to nonlinear fuzzy model identification. The contribution of the paper with respect to the state of art is to meet simultaneously all the four requirements. To do so, we introduce an energy-gain bounding approach. To render robustness in the identification, one could argue for other criteria, e.g. robust least-squares (Kumar et al., 2003c, 2004b) and robust regularized least-squares estimation (Kumar et al., 2006). However, the energy-gain bounding approach should be preferred, since unlike Kumar et al. (2004b) and Kumar et al. (2003c) it provides an on-line solution without requiring the knowledge of upper bound on disturbances. Further, our approach will provide a framework to study the robust identification of nonlinear fuzzy models. Also, we will see that the solution of Kumar et al. (2006) (that was motivated using approximations) could be derived in our framework without approximations. Sugeno-type fuzzy inference systems combine simplicity with good analytical properties (Takagi & Sugeno, 1985), and allow qualitative insight into the relationships (Babuška, 2000; Bodenhofer & Bauer, 2000; Espinosa & Vandewalle, 2000; Setnes, Babuška, & Verbruggen, 1998). Therefore, we consider the identification of a Sugeno-type fuzzy model.

2. A Takagi–sugeno fuzzy model

Consider a zero-order Takagi–Sugeno fuzzy model ($F_s : X \to Y$) that maps n-dimensional input space ($X = X_1 \times X_2 \times \cdots \times X_n$) to one-dimensional real line. A rule of the model is represented as "If x_1 is A_1 and $\cdots$ and x_n is A_n then $y = c$". Here, $x_1, \ldots, x_n$ are the model input variables, y is the output variable. As regard, $A_1, \ldots, A_n$, they are the antecedents (if-part of the rule) and c is the consequent (then-part of the rule). The $A_1, \ldots, A_n$ are the linguistic terms (such as *low*, *high*) which are represented by fuzzy sets. Given a universe of discourse X_j, a fuzzy subset A_j of X_j is characterized by a mapping: $\mu_{A_j} : X_j \to [0, 1]$, where for $x_j \in X_j$, $\mu_{A_j}(x_j)$ can be interpreted as the degree or grade to which x_j belongs to A_j. This mapping is called as membership function of the fuzzy set. The constant c is a real number. Let us define, for jth input, P_j nonempty fuzzy subsets of X_j (represented by $A_{j1}, \ldots, A_{jP_j}$) such that for any $x_j \in X_j$,

$$\mu_{A_{j1}}(x_j) + \mu_{A_{j2}}(x_j) + \cdots + \mu_{A_{jP_j}}(x_j) = 1. \tag{1}$$

Now, the model could have maximum number of $K = \prod_{j=1}^{n} P_j$ rules. Let the ith rule of above rule-base is represented as R_i: If x_1 is A_{1i} and $\cdots$ and x_n is A_{ni} then $y = c_i$, where $A_{1i} \in \{A_{11}, \ldots, A_{1P_1}\}$, $A_{2i} \in \{A_{21}, \ldots, A_{2P_2}\}$ and so on. For a given input x, the *degree of fulfillment* of the ith rule, by modelling the logic operator 'and' using product, is given by $g_i(x) = \prod_{j=1}^{n} \mu_{A_{ji}}(x_j)$. The output of the fuzzy model to input vector $x \in X$ is computed by taking the weighted average of the output provided by each rule:

$$F_s(x) = \frac{\sum_{i=1}^{K} c_i g_i(x)}{\sum_{i=1}^{K} g_i(x)} = \frac{\sum_{i=1}^{K} c_i \prod_{j=1}^{n} \mu_{A_{ji}}(x_j)}{\sum_{i=1}^{K} \prod_{j=1}^{n} \mu_{A_{ji}}(x_j)}.$$

Since the membership functions satisfy (1), therefore $\sum_{i=1}^{K} \prod_{j=1}^{n} \mu_{A_{ji}}(x_j) = 1$ and hence

$$F_s(x) = \sum_{i=1}^{K} c_i \prod_{j=1}^{n} \mu_{A_{ji}}(x_j). \tag{2}$$

We characterize for any input x_j, the P_j different membership functions, by a knot sequence $\{\theta_{j1} < \theta_{j2} < \cdots < \theta_{jP_j}\}$, as shown in Fig. 1. Now, consider the problem of assigning two different memberships (say $\mu_{A_{ji}}$ and $\mu_{A_{j(i+1)}}$) to a point x_j such that $\theta_{ji} < x_j < \theta_{j(i+1)}$, based on following criterion: $[\mu_{A_{ji}}(x_j), \mu_{A_{j(i+1)}}(x_j)] = \arg\min_{[u_1, u_2; u_1 + u_2 = 1]} [u_1^2 (x_j - \theta_{ji})^2 + u_2^2 (x_j - \theta_{j(i+1)})^2]$. This results in

$$\mu_{A_{ji}}(x_j) = \frac{(x_j - \theta_{j(i+1)})^2}{(x_j - \theta_{ji})^2 + (x_j - \theta_{j(i+1)})^2}$$

and

$$\mu_{A_{j(i+1)}}(x_j) = \frac{(x_j - \theta_{ji})^2}{(x_j - \theta_{ji})^2 + (x_j - \theta_{j(i+1)})^2}.$$

Thus, P_j different membership functions for input x_j can be defined as follows:

$$\mu_{A_{j1}} = \begin{cases} 1, & x_j \leqslant \theta_{j1} \\ \dfrac{(x_j - \theta_{j2})^2}{(x_j - \theta_{j1})^2 + (x_j - \theta_{j2})^2}, & \theta_{j1} \leqslant x_j \leqslant \theta_{j2} \\ 0 & \text{otherwise} \end{cases}$$

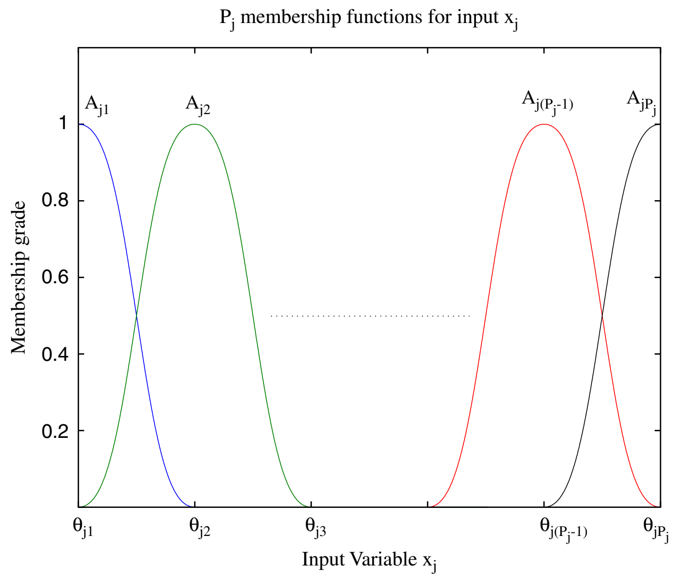

Fig. 1. The characterization of membership functions by a knot sequence.

$$\mu_{A_{j2}} = \begin{cases} \dfrac{(x_j-\theta_{j1})^2}{(x_j-\theta_{j1})^2+(x_j-\theta_{j2})^2}, & \theta_{j1}\leqslant x_j\leqslant\theta_{j2},\\ \dfrac{(x_j-\theta_{j3})^2}{(x_j-\theta_{j2})^2+(x_j-\theta_{j3})^2}, & \theta_{j2}\leqslant x_j\leqslant\theta_{j3},\ldots,\\ 0 & \text{otherwise,}\end{cases}$$

$$\mu_{A_{jP_j}} = \begin{cases} 1, & x_j\geqslant\theta_{jP_j},\\ \dfrac{(x_j-\theta_{j(P_j-1)})^2}{(x_j-\theta_{j(P_j-1)})^2+(x_j-\theta_{jP_j})^2}, & \theta_{j(P_j-1)}\leqslant x_j\leqslant\theta_{jP_j},\\ 0 & \text{otherwise.}\end{cases}$$

Fig. 1 shows the shape of above defined membership functions. Note that the membership functions fulfill (1) and jth input membership functions are described by a knot sequence $[\theta_{j1},\ldots,\theta_{jP_j}]$ such that $\theta_{j1}<\theta_{j2}<\cdots<\theta_{jP_j}$. We ensemble the knot sequences for all inputs by defining an L-dimensional (where $L=\sum_{j=1}^{n}P_j$) vector θ: $\theta=[\theta_{11}\cdots\theta_{1P_1}\theta_{21}\cdots\theta_{2P_2}\cdots\theta_{n1}\cdots\theta_{nP_n}]\in R^L$. Now, the *degree of fulfillment* of ith rule can be defined as a function of vector θ: $g_i(x,\theta)=\prod_{j=1}^{n}\mu_{A_{ji}}(x_j)$. If we define two K-dimensional (where $K=\prod_{j=1}^{n}P_j$) vectors $G(x,\theta)=\left[g_1(x,\theta)\cdots g_K(x,\theta)\right]^{\mathrm{T}}\in R^K$ and $\alpha=[c_1\cdots c_K]\in R^K$, then (2) can be rewritten as $F_{\mathrm{s}}(x)=G^{\mathrm{T}}(x,\theta)\alpha$. During the identification of membership functions, it is necessary that two consecutive knots must be sufficiently separated for the good interpretability of the fuzzymodel (Lindskog, 1997). That is, there must exist some ε_j such that $\theta_{j2}-\theta_{j1}\geqslant\varepsilon_j,\ldots,\theta_{jP_j}-\theta_{jP_{j-1}}\geqslant\varepsilon_j$, $\forall\ j=1,\ldots,n$. The above inequalities can be put together by defining a suitable matrix c and a vector h such that $c\theta\geqslant h$ (Burger et al., 2002; Kumar et al., 2003a, 2003b, 2004a, 2004b, 2006). Hence, a Takagi–Sugeno fuzzy model is characterized by $F_{\mathrm{s}}(x)=G^{\mathrm{T}}(x,\theta)\alpha, c\theta\geqslant h$.

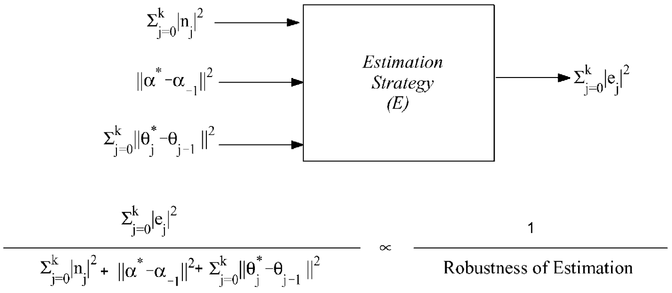

Fig. 2. Energy-gain from disturbances to estimation errors in fuzzy parameters estimation.

3. An energy-gain bounding approach

The classical approach to the fuzzy modelling of unknown processes is to assume that there exists an interpretable Sugeno-type fuzzy model, say (α^*,θ^*) with $c\theta^*\geqslant h$, for approximating the process. That is, at any time j, $y(j)=G^{\mathrm{T}}(x(j),\theta^*)\alpha^*+n_j$, where $x(j)$ is input vector, scalar $y(j)$ is output measurement, and n_j is measurement noise that includes also the modelling errors. Our approach, however, is to model the unknown process as

$$y(j)=G^{\mathrm{T}}(x(j),\theta^*_j)\alpha^*+n_j, \tag{3}$$

such that antecedents vector θ^*_j may vary with time index j. The point here is to model the unknown process using a time-varying antecedents fuzzy model. This will allow us to solve in closed-form the solution of formulated robust fuzzy identification problem in Theorem 2. However, it is expected that during the identification process, θ^*_j tends to converge with time-index j to a constant vector. The concern of this paper is to (1) find out an appropriate variation of antecedents vector θ^*_j, and (2) to estimate the parameter α^*, in presence of unknown deterministic disturbance n_j. At any time instant j, we are concerned with the robust estimation of fuzzy parameters (α^*,θ^*_j), say (α_j,θ_j) with $c\theta_j\geqslant h$, using input–output identification data sequence $\{x(j),y(j)\}$, without a priori knowledge of statistics and upper bound on the uncertainty signal n_j. The estimation error signal is given as $e_j=G^{\mathrm{T}}(x(j),\theta^*_j)\alpha^*-G^{\mathrm{T}}(x(j),\theta_j)\alpha_j$. Any method for computing $\{\alpha_j,\theta_j\}_{j=0}^{k}$ from identification data $\{x(j),y(j)\}_{j=0}^{k}$ (being referred as estimation strategy) will be considered *good* if it results in a small value of estimation errors *energy* (being measured as $\sum_{j=0}^{k}|e_j|^2$). The performance of an estimation strategy will be affected by three kind of unknown disturbances: (1) measurement noise n_j, (2) deviation of initialguess α_{-1} from true parameter α^*, (3) deviation of θ^*_j from its initial guess θ_{j-1}. Here, the initial guess about θ^*_j is taken equal to the estimate of θ^*_{j-1}, since θ^*_j tends to converge. Thus, to every estimation strategy, is associated a kind of mapping from disturbances to the estimation errors. This mapping, depending upon its robustness, will result to a gain of energy from disturbances to the estimation errors, as shown in Fig. 2. Now, to render robustness in the design of an estimation strategy against unknown disturbances, we try to minimize the value of an upper bound on energy-gain, along the line of

H^∞-optimal estimation (Hassibi, Sayed, & Kailath, 1996b, c). This, we call as energy-gain bounding approach:

$$\text{Minimize } \gamma,$$
$$\text{subject to}$$
$$\frac{\sum_{j=0}^{k} |G^{\mathrm{T}}(x(j), \theta_j^*)\alpha^* - G^{\mathrm{T}}(x(j), \theta_j)\alpha_j|^2}{\mu^{-1}\|\alpha^* - \alpha_{-1}\|^2 + \mu_\theta^{-1}\sum_{j=0}^{k}\|\theta_j^* - \theta_{j-1}\|^2 + \sum_{j=0}^{k}|n_j|^2} < \gamma^2,$$

for any vector α^* and for all sequences $\{\theta_j^*, c\theta_j^* \geqslant h\}_{j=0}^k, \{n_j\}_{j=0}^k$. Here, the parameter $\mu > 0$ reflects a priori knowledge as to how close α^* is to the initial guess α_{-1}, 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 $\{\theta_{j-1}\}_{j=0}^k$. The estimation strategy $\{\alpha_j, \theta_j\}_{j=0}^k$ is causal and $|n_j| < \infty$, for all $j = 0, \ldots, k$. Note that signal n_j is deterministic and no assumption about its nature has been made. We take, for simplicity, the initial guess α_{-1} equal to a null vector.

4. The sub-optimal solution

Given a scalar $\gamma \geqslant 0$, find a causal estimation strategy $\{\alpha_j, \theta_j, c\theta_j \geqslant h\}_{j=0}^k$ that achieves

$$\frac{\sum_{j=0}^{k} |G^{\mathrm{T}}(x(j), \theta_j^*)\alpha^* - G^{\mathrm{T}}(x(j), \theta_j)\alpha_j|^2}{\mu^{-1}\|\alpha^*\|^2 + \mu_\theta^{-1}\sum_{j=0}^{k}\|\theta_j^* - \theta_{j-1}\|^2 + \sum_{j=0}^{k}|n_j|^2} < \gamma^2, \tag{4}$$

for any vector α^* and for all sequences $\{\theta_j^*, c\theta_j^* \geqslant h\}_{j=0}^k, \{n_j\}_{j=0}^k$.

Lemma 1. *For* $\gamma \geqslant 1$, *a unique minimum exists for the problem*

$$\min_{\{\alpha^*\}} \begin{pmatrix} -\gamma^{-2}\sum_{j=0}^{k}|G^{\mathrm{T}}(x(j), \theta_j^*)\alpha^* - G^{\mathrm{T}}(x(j), \theta_j)\alpha_j|^2 \\ +\mu^{-1}\|\alpha^*\|^2 + \sum_{j=0}^{k}|y(j) - G^{\mathrm{T}}(x(j), \theta_j^*)\alpha^*|^2 \end{pmatrix},$$

and is given by

$$\sum_{j=0}^{k} \frac{[y(j) - G^{\mathrm{T}}(x(j), \theta_j^*)\widehat{\alpha}_j]^2}{1 + G^{\mathrm{T}}(x(j), \theta_j^*)P_j G(x(j), \theta_j^*)} - \sum_{j=0}^{k} \frac{[G^{\mathrm{T}}(x(j), \theta_j)\alpha_j - G^{\mathrm{T}}(x(j), \theta_j^*)\overline{\alpha}_j]^2}{\gamma^2 - G^{\mathrm{T}}(x(j), \theta_j^*)[P_j^{-1} + G(x(j), \theta_j^*)G^{\mathrm{T}}(x(j), \theta_j^*)]^{-1}G(x(j), \theta_j^*)}, \tag{5}$$

where $\widehat{\alpha}_0 = 0,\ P_0 = \mu I$,

$$\widehat{\alpha}_{j+1} = \widehat{\alpha}_j + \frac{P_j G(x(j), \theta_j^*)\left[y(j) - G^{\mathrm{T}}(x(j), \theta_j^*)\widehat{\alpha}_j\right]}{1 + G^{\mathrm{T}}(x(j), \theta_j^*)P_j G(x(j), \theta_j^*)} - \frac{P_j G(x(j), \theta_j^*)[G^{\mathrm{T}}(x(j), \theta_j)\alpha_j - G^{\mathrm{T}}(x(j), \theta_j^*)\overline{\alpha}_j]}{(1 + G^{\mathrm{T}}(x(j), \theta_j^*)P_j G(x(j), \theta_j^*))(\gamma^2 - G^{\mathrm{T}}(x(j), \theta_j^*)T_1)},$$

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

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

and

$$T_1 = [P_j^{-1} + G(x(j), \theta_j^*)G^{\mathrm{T}}(x(j), \theta_j^*)]^{-1}G(x(j), \theta_j^*).$$

Proof. The proof is based on following Theorem:

Theorem 1. *Consider a quadratic form*

$$\mathscr{J}_k(x_0, \{u_j\}_{j=0}^k, \{y_j\}_{j=0}^k) = x_0^{\mathrm{T}}\Pi_0^{-1}x_0 + \sum_{j=0}^{k}\begin{bmatrix} u_j \\ y_j - H_j x_j \end{bmatrix}^* \begin{bmatrix} Q_j & S_j \\ S_j^{\mathrm{T}} & R_j \end{bmatrix}^{-1} \times \begin{bmatrix} u_j \\ y_j - H_j x_j \end{bmatrix} \tag{6}$$

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,* $Q_j - S_j R_j^{-1} S_j^{\mathrm{T}} > 0$ *and* $[F_j G_j]$ *has full rank for all* j, *then the quadratic forms* (6) *will have a unique minimum, if and only if,*

$$P_j^{-1} + H_j^{\mathrm{T}} R_j^{-1} H_j > 0, \quad 0 \leqslant j \leqslant k, \tag{7}$$

$$P_{j+1} = F_j P_j F_j^{\mathrm{T}} + G_j Q_j G_j^{\mathrm{T}} - K_{p,j} R_{e,j} K_{p,j}^{\mathrm{T}}, \tag{8}$$

$P_0 = \Pi_0$, $R_{e,j} = R_j + H_j P_j H_j^{\mathrm{T}}$, $K_{p,j} = (F_j P_j H_j^{\mathrm{T}} + G_j S_j)R_{e,j}^{-1}$. *It also follows in the minimum case that* $P_{j+1} > 0$ *for all* $0 \leqslant j \leqslant k$. *Also, the minimum value of* $\mathscr{J}_k(x_0, \{u_j\}_{j=0}^k, \{y_j\}_{j=0}^k)$ *over* $(x_0, \{u_j\}_{j=0}^k)$ *is given by*

$$\sum_{j=0}^{k}(y_j - H_j\widehat{x}_j)^{\mathrm{T}} R_{e,j}^{-1}(y_j - H_j\widehat{x}_j), \tag{9}$$

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

Proof. This is the result of Theorem 6 and Lemma 13 in Hassibi, Sayed, and Kailath (1996a). □

It is easy to see that minimization problem

$$\min_{\{\alpha^*\}} \begin{pmatrix} -\gamma^{-2}\sum_{j=0}^{k}|G^{\mathrm{T}}(x(j), \theta_j^*)\alpha^* - G^{\mathrm{T}}(x(j), \theta_j)\alpha_j|^2 \\ +\mu^{-1}\|\alpha^*\|^2 + \sum_{j=0}^{k}|y(j) - G^{\mathrm{T}}(x(j), \theta_j^*)\alpha^*|^2 \end{pmatrix}$$

can be identified a special case of the quadratic form (6) by considering for all $0 \leqslant j \leqslant k$, $F_j = I, G_j = 0, x_j = \alpha^*, u_j = 0, \Pi_0 = \mu I, Q_j = I, S_j = 0$,

$$y_j = \begin{bmatrix} y(j) \\ G^{\mathrm{T}}(x(j), \theta_j)\alpha_j \end{bmatrix}, \quad H_j = \begin{bmatrix} G^{\mathrm{T}}(x(j), \theta_j^*) \\ G^{\mathrm{T}}(x(j), \theta_j^*) \end{bmatrix},$$
$$R_j = \begin{bmatrix} 1 & 0 \\ 0 & -\gamma^2 \end{bmatrix}.$$

Let $\widehat{\alpha}_j$ denotes a vector that corresponds to a variable $\widehat{x}_j$ in expression (9). Here, $\widehat{\alpha}_j$ should not be confused with the estimation strategy α_j. First, we show that a minimum exists for the problem for $\gamma \geqslant 1$ by checking the condition (7). That is, $\mu^{-1}\|\alpha^*\|^2 + \sum_{j=0}^{k}|y(j) - G^{\mathrm{T}}(x(j), \theta_j^*)\alpha^*|^2 - \gamma^{-2}\sum_{j=0}^{k}|G^{\mathrm{T}}(x(j), \theta_j^*)\alpha^* - G^{\mathrm{T}}(x(j), \theta_j)\alpha_j|^2$ will have a minimum over $\{\alpha^*\}$ for all $0 \leqslant j \leqslant k$, iff

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

where

$$P_0=\mu I,\, P_{j+1}=P_j-P_j[G(x(j),\theta_j^*)G(x(j),\theta_j^*)]T_2^{-1}\\ \times\begin{bmatrix}G^{\mathrm{T}}(x(j),\theta_j^*)\\ G^{\mathrm{T}}(x(j),\theta_j^*)\end{bmatrix}P_j,$$

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

The existence condition is equivalent to

$$P_j^{-1}+(1-\gamma^{-2})G(x(j),\theta_j^*)G^{\mathrm{T}}(x(j),\theta_j^*)>0, \tag{10}$$

$\forall j=0,\ldots,k$, and using matrix inversion lemma, it can be seen that $P_{j+1}^{-1}=P_j^{-1}+(1-\gamma^{-2})G(x(j),\theta_j^*)G^{\mathrm{T}}(x(j),\theta_j^*)$. At this end, consider the minimum eigenvalue of matrix P_j^{-1} by using the fact that $\lambda_{\min}(P_j^{-1})=\min_{u^{\mathrm{T}}u=1}u^{\mathrm{T}}P_j^{-1}u$. Therefore, for every j, we can determine a vector $u_0\varepsilon R^K$ with $u_0^{\mathrm{T}}u_0=1$ such that $\lambda_{\min}(P_j^{-1})=u_0^{\mathrm{T}}P_j^{-1}u_0=u_0^{\mathrm{T}}P_{j-1}^{-1}u_0+(1-\gamma^{-2})|G^{\mathrm{T}}(x(j-1),\theta_{j-1}^*)u_0|^2$. Assume that $\gamma\geqslant 1$. If $P_{j-1}^{-1}>0$ (i.e. $u_0^{\mathrm{T}}P_{j-1}^{-1}u_0>0$), then $\lambda_{\min}(P_j^{-1})>0$ and therefore $P_j^{-1}>0$. Since $P_0^{-1}=\mu^{-1}I>0$, therefore by induction $P_j^{-1}>0,\, j=0,\ldots,k$. Thus, for any nonzero vector $\alpha\varepsilon R^K$, $\alpha^{\mathrm{T}}P_j^{-1}\alpha+(1-\gamma^{-2})|G^{\mathrm{T}}(x(j),\theta_j^*)\alpha|^2>0$. Rewriting the above inequality, $\alpha^{\mathrm{T}}[P_j^{-1}+(1-\gamma^{-2})G(x(j),\theta_j^*)G^{\mathrm{T}}(x(j),\theta_j^*)]\alpha>0$. Since α is any nonzero vector, therefore $P_j^{-1}+(1-\gamma^{-2})G(x(j),\theta_j^*)G^{\mathrm{T}}(x(j),\theta_j^*)>0,\ \forall j=0,\ldots,k$. Hence, we see that for $\gamma\geqslant 1$, the existence condition (10) is satisfied and the minimum value of the function can be calculated using (9). For this consider

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

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

$$R_{e,j}^{-1}=\begin{bmatrix}1 & \frac{-G^{\mathrm{T}}P_jG}{1+G^{\mathrm{T}}P_jG}\\ 0 & 1\end{bmatrix}T_3^{-1}\begin{bmatrix}1 & 0\\ \frac{-G^{\mathrm{T}}P_jG}{1+G^{\mathrm{T}}P_jG} & 1\end{bmatrix}, \tag{11}$$

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

The minimum value is equal to $\sum_{j=0}^{k}T_4^{\mathrm{T}}R_{e,j}^{-1}T_4$,

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

From Theorem (1), $\widehat{\alpha}_0=0$, $\widehat{\alpha}_{j+1}=\widehat{\alpha}_j+P_j[G(x(j),\theta_j^*)G(x(j),\theta_j^*)]R_{e,j}^{-1}T_4$. By substituting the value of $R_{e,j}^{-1}$ from (11), the minimum value becomes (5). □

Theorem 2. *If any unknown physical process is modelled according to* (3) *by defining* $\theta_j^*=\theta_j$,

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

$$\Psi_j(\theta)=\frac{[y(j)-G^{\mathrm{T}}(x(j),\theta)\alpha_{j-1}]^2}{1+G^{\mathrm{T}}(x(j),\theta)P_jG(x(j),\theta)}+\mu_\theta^{-1}\|\theta-\theta_{j-1}\|^2,$$

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

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

$\alpha_{-1}=0$, $P_0=\mu I$, *then* $\{\alpha_j,\theta_j\}_{j=0}^{k}$ *is a solution of optimization problem* (4).

Proof. Define an indefinite quadratic form as $J_k(\alpha^*,\{\theta_j^*\}_{j=0}^{k},\{x(j),y(j)\}_{j=0}^{k},\{\alpha_j,\theta_j\}_{j=0}^{k})=\sum_{j=0}^{k}|y(j)-G^{\mathrm{T}}(x(j),\theta_j^*)\alpha^*|^2+\mu^{-1}\|\alpha^*\|^2-\gamma^{-2}\sum_{j=0}^{k}|G^{\mathrm{T}}(x(j),\theta_j^*)\alpha^*-G^{\mathrm{T}}(x(j),\theta_j)\alpha_j|^2+\mu_\theta^{-1}\sum_{j=0}^{k}\|\theta_j^*-\theta_{j-1}\|^2$. Noting $n_j=y(j)-G^{\mathrm{T}}(x(j),\theta_j^*)\alpha^*$, it can be seen that the sub-optimality condition (4) is satisfied, if and only if, $J_k(\alpha^*,\{\theta_j^*\}_{j=0}^{k},\{x(j),y(j)\}_{j=0}^{k},\{\alpha_j,\theta_j\}_{j=0}^{k})>0$, for any vector α^* and for all sequences $\{\theta_j^*,c\theta_j^*\geqslant h\}_{j=0}^{k}$, $\{y(j)\}_{j=0}^{k}$. Therefore, any sub-optimal causal estimation strategy $\{\alpha_j,\theta_j,c\theta_j\geqslant h\}_{j=0}^{k}$ that achieves a robustness level of γ, for a given fixed data sequence $\{x(j),y(j)\}_{j=0}^{k}$, must ensure that

$$\min_{\alpha^*,\{\theta_j^*,c\theta_j^*\geqslant h\}_{j=0}^{k}} J_k(\cdot)>0. \tag{12}$$

For a given sequence of antecedent parameters $\{\theta_j^*,c\theta_j^*\geqslant h\}_{j=0}^{k}$, we define

$$J_k^{\min}(\{\theta_j^*\}_{j=0}^{k},\{x(j),y(j)\}_{j=0}^{k},\{\alpha_j,\theta_j\}_{j=0}^{k})\\ =\min_{\{\alpha^*\}}J_k(\alpha^*,\{\theta_j^*\}_{j=0}^{k},\{x(j),y(j)\}_{j=0}^{k},\{\alpha_j,\theta_j\}_{j=0}^{k}).$$

Now, any sub-optimal estimation strategy $\{\alpha_j,\theta_j,c\theta_j\geqslant h\}_{j=0}^{k}$, must ensure that

$$\min_{\{\theta_j^*,c\theta_j^*\geqslant h\}_{j=0}^{k}} J_k^{\min}(\cdot)>0. \tag{13}$$

To design an estimation strategy based on (13), we first need to find the functional value of $J_k^{\min}(\cdot)$ by solving a deterministic quadratic form minimization problem

$$\min_{\{\alpha^*\}}\begin{pmatrix}-\gamma^{-2}\sum_{j=0}^{k}|G^{\mathrm{T}}(x(j),\theta_j^*)\alpha^*-G^{\mathrm{T}}(x(j),\theta_j)\alpha_j|^2\\ +\mu^{-1}\|\alpha^*\|^2+\sum_{j=0}^{k}|y(j)-G^{\mathrm{T}}(x(j),\theta_j^*)\alpha^*|^2\end{pmatrix},$$

using Lemma 1. Now, using Lemma 1, the functional value of $J_k^{\min}(\{\theta_j^*\}_{j=0}^{k},\{x(j),y(j)\}_{j=0}^{k},\{\alpha_j,\theta_j\}_{j=0}^{k})$, can be calculated as

$$
\begin{aligned}
&J_k^{\min}(\{\theta_j^*\}_{j=0}^k, \{x(j), y(j)\}_{j=0}^k, \{\alpha_j, \theta_j\}_{j=0}^k)\\
&\quad=\sum_{j=0}^{k}\frac{[y(j)-G^{\mathrm{T}}(x(j),\theta_j^*)\widehat{\alpha}_j]^2}{1+G^{\mathrm{T}}(x(j),\theta_j^*)P_jG(x(j),\theta_j^*)}\\
&\qquad-\sum_{j=0}^{k}\frac{\left[G^{\mathrm{T}}(x(j),\theta_j)\alpha_j-G^{\mathrm{T}}(x(j),\theta_j^*)\overline{\alpha}_j\right]^2}{\gamma^2-G^{\mathrm{T}}(x(j),\theta_j^*)T_1}\\
&\qquad+\mu_\theta^{-1}\sum_{j=0}^{k}\|\theta_j^*-\theta_{j-1}\|^2,
\end{aligned}
\tag{14}
$$

$T_1=[P_j^{-1}+G(x(j),\theta_j^*)G^{\mathrm{T}}(x(j),\theta_j^*)]^{-1}G(x(j),\theta_j^*)$. After calculating $J_k^{\min}$, we return to the original sub-optimal estimation problem (13). Therefore, all we have to do is to choose any causal estimation strategy $\{\alpha_j, \theta_j, c\theta_j \geqslant h\}_{j=0}^k$ that ensures inequality (13), where $J_k^{\min}(\{\theta_j^*\}_{j=0}^k,\{x(j),y(j)\}_{j=0}^k,\{\alpha_j,\theta_j\}_{j=0}^k)$ is given by (14). There may be more than one different estimation strategies which ensure inequality (13). To narrow our search and to pick up a simple and computationally cheap strategy, we put a constraint on the estimation strategy that $\alpha_j = A_\alpha\overline{\alpha}_j, \theta_j = A_\theta\theta_j^*, j = 0,\ldots,k$, 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 $J_k^{\min}(\cdot)$ w.r.t. $\{\theta_j^*, c\theta_j^* \geqslant h\}_{j=0}^k$, for a given choice of A_α and A_θ, is denoted by $\{\theta_{A_\theta}^{A_\alpha}(j)\}_{j=0}^k$, i.e. $\{\theta_{A_\theta}^{A_\alpha}(j)\}_{j=0}^k = \arg\min_{\{\theta_j^*, c\theta_j^*\geqslant h\}_{j=0}^k} T_5(\cdot)$, $T_5(\cdot) = J_k^{\min}(\{\theta_j^*\}_{j=0}^k, \{x(j), y(j)\}_{j=0}^k, \{A_\alpha\overline{\alpha}_j, A_\theta\theta_j^*\}_{j=0}^k)$.

Remark 1. Let us consider an example of computing the causal parameters $\{\theta_{A_\theta}^{A_\alpha}(j)\}_{j=0}^k$ when $A_\alpha = I$ and $A_\theta = I$. When $A_\alpha = I$ and $A_\theta = I$, then

$$
\{\theta_I^I(j)\}_{j=0}^k = \arg\min_{\{\theta_j^*, c\theta_j^*\geqslant h\}_{j=0}^k}\sum_{j=0}^{k}\Psi_j(\cdot),
\tag{15}
$$

$\Psi_j(\cdot)=[y(j)-G^{\mathrm{T}}(x(j),\theta_j^*)\widehat{\alpha}_j]^2/1+G^{\mathrm{T}}(x(j),\theta_j^*)P_jG(x(j),\theta_j^*) + \mu_\theta^{-1}\|\theta_j^*-\theta_{j-1}^*\|^2$, where θ_{-1}^* denotes the initial guess θ_{-1}, $\widehat{\alpha}_{j+1} = \widehat{\alpha}_j + \frac{P_jG(x(j),\theta_j^*)[y(j)-G^{\mathrm{T}}(x(j),\theta_j^*)\widehat{\alpha}_j]}{1+G^{\mathrm{T}}(x(j),\theta_j^*)P_jG(x(j),\theta_j^*)}$, $P_{j+1}=[P_j^{-1}+(1-\gamma^{-2})G(x(j),\theta_j^*)G^{\mathrm{T}}(x(j),\theta_j^*)]^{-1}$, $\widehat{\alpha}_0=0$, $P_0 = \mu I$. Since $\{\theta_I^I(j)\}_{j=0}^k$are causal, therefore $\theta_I^I(0) = \arg\min_\theta[\Psi_0(\theta), c\theta\geqslant h]$, $\Psi_0(\theta)=[y(0)-G^{\mathrm{T}}(x(0),\theta)\widehat{\alpha}_0]^2/1+G^{\mathrm{T}}(x(0),\theta)P_0G(x(0),\theta)+\mu_\theta^{-1}\|\theta-\theta_{-1}\|^2$, $\widehat{\alpha}_0=0$, $P_0=\mu I$. Now, the value $\theta_I^I(0)$ (and so the values $\widehat{\alpha}_1, P_1$) are fixed. Therefore, the estimation of $\theta_I^I(1)$ follows as $\theta_I^I(1) = \arg\min_\theta[\Psi_1(\theta), c\theta\geqslant h]$, $\Psi_1(\theta)=[y(1)-G^{\mathrm{T}}(x(1),\theta)\widehat{\alpha}_1]^2/1+G^{\mathrm{T}}(x(1),\theta)P_1G(x(1),\theta)+\mu_\theta^{-1}\|\theta-\theta_I^I(0)\|^2$, and so on follows the estimation of other parameters. Hence,the parameters sequence $\{\theta_I^I(j)\}_{j=0}^k$ can be recursively computed by solving $(k+1)$ minimization problems, i.e. for $j=0,\ldots,k$,

$$
\theta_I^I(j) = \arg\min_\theta\left[\Psi_j(\theta), c\theta\geqslant h\right],
\tag{16}
$$

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

$$
\widehat{\alpha}_{j+1}=\widehat{\alpha}_j+\frac{P_jG(x(j),\theta_I^I(j))[y(j)-G^{\mathrm{T}}(x(j),\theta_I^I(j))\widehat{\alpha}_j]}{1+G^{\mathrm{T}}(x(j),\theta_I^I(j))P_jG(x(j),\theta_I^I(j))},
\tag{17}
$$

$$
P_{j+1}^{-1}=P_j^{-1}+(1-\gamma^{-2})G(x(j),\theta_I^I(j))G^{\mathrm{T}}(x(j),\theta_I^I(j)),
\tag{18}
$$

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 sub-optimal (i.e. achieves a robustness level of $\gamma > 1$) if

$$
J_k^{\min}(\{\theta_{A_\theta}^{A_\alpha}(j)\}_{j=0}^k, \{x(j), y(j)\}_{j=0}^k, \{\alpha_j, \theta_j\}_{j=0}^k) > 0.
\tag{19}
$$

There may exist different estimation strategies which satisfy the above sub-optimal condition (i.e. inequality (19)). 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

$$
\begin{aligned}
&J_k^{\min}(\{\theta_{A_\theta}^I(j)\}_{j=0}^k, \{x(j), y(j)\}_{j=0}^k, \{\overline{\alpha}_j, \theta_{A_\theta}^I(j)\}_{j=0}^k)\\
&\quad=\left(\begin{array}{c}\displaystyle\sum_{j=0}^{k}\frac{[y(j)-G^{\mathrm{T}}(x(j),\theta_{A_\theta}^I(j))\widehat{\alpha}_j]^2}{1+G^{\mathrm{T}}(x(j),\theta_{A_\theta}^I(j))P_jG(x(j),\theta_{A_\theta}^I(j))}\\ +\mu_\theta^{-1}\sum_{j=0}^k\|\theta_{A_\theta}^I(j)-\theta_{j-1}\|^2\end{array}\right)>0,
\end{aligned}
$$

since $P_j > 0$, where $\widehat{\alpha}_0=0, \widehat{\alpha}_{j+1}=\widehat{\alpha}_j+P_jG(x(j),\theta_{A_\theta}^I(j))[y(j) - G^{\mathrm{T}}(x(j),\theta_{A_\theta}^I(j))\widehat{\alpha}_j]/1+G^{\mathrm{T}}(x(j),\theta_{A_\theta}^I(j))P_jG(x(j),\theta_{A_\theta}^I(j))$, and $\overline{\alpha}_j=\widehat{\alpha}_{j+1}$. The choice of operator A_θ that satisfies $A_\theta\theta_j^*=\theta_{A_\theta}^I(j)$, for $j=0,\ldots,k$, is still not clear. We have seen in Remark 1 that the causal parameters sequence $\{\theta_I^I(j)\}_{j=0}^k$ can be recursively computed using (16)–(18). Therefore, we motivate the choice $A_\theta = I$, by defining $\theta_j^*=\theta_I^I(j)$, i.e. we model the unknown process (see, (3)) as

$$
y(j)=G^{\mathrm{T}}(x(j),\theta_I^I(j))\alpha^*+n_j.
\tag{20}
$$

When $\theta_j=\theta_I^I(j)$ and $\alpha_j=\overline{\alpha}_j$, then $\alpha_j=\widehat{\alpha}_{j+1}$. Then, it follows from (16)–(18) that

$$
\theta_j=\arg\min_\theta[\Psi_j(\theta), c\theta\geqslant h],
\tag{21}
$$

$$
\Psi_j(\theta)=\frac{[y(j)-G^{\mathrm{T}}(x(j),\theta)\alpha_{j-1}]^2}{1+G^{\mathrm{T}}(x(j),\theta)P_jG(x(j),\theta)}+\mu_\theta^{-1}\|\theta-\theta_{j-1}\|^2,
$$

$$
\alpha_j=\alpha_{j-1}+\frac{P_jG(x(j),\theta_j)\left[y(j)-G^{\mathrm{T}}(x(j),\theta_j)\alpha_{j-1}\right]}{1+G^{\mathrm{T}}(x(j),\theta_j)P_jG(x(j),\theta_j)},
\tag{22}
$$

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

$$
\alpha_{-1}=0,\quad P_0=\mu I. \qquad \square
$$

5. The optimal solution

Now, our concern is to solve

$$\min_{\{(\alpha_j,\theta_j),c\theta_j \geqslant h\}_{j=0}^k} \quad \max_{\{\alpha^*,\{\theta_j^*\}_{j=0}^k,\{n_j\}_{j=0}^k\}} \mathscr{T}, \tag{23}$$

$$\mathscr{T} = \frac{\sum_{j=0}^k |G^{\mathrm{T}}(x(j),\theta_j^*)\alpha^* - G^{\mathrm{T}}(x(j),\theta_j)\alpha_j|^2}{\mu^{-1}\|\alpha^*\|^2 + \mu_\theta^{-1}\sum_{j=0}^k \|\theta_j^* - \theta_{j-1}\|^2 + \sum_{j=0}^k |n_j|^2}.$$

Theorem 3. *If any unknown physical process is modelled according to* (3) *by defining* $\theta_j^* = \theta_j$,

$$\theta_j = \arg\min_\theta [\Psi_j(\theta), c\theta \geqslant h], \tag{24}$$

$$\Psi_j(\theta) = \frac{[y(j) - G^{\mathrm{T}}(x(j),\theta)\alpha_{j-1}]^2}{1 + \mu\|G(x(j),\theta)\|^2} + \mu_\theta^{-1}\|\theta - \theta_{j-1}\|^2,$$

$$\alpha_j = \alpha_{j-1} + \frac{\mu G(x(j),\theta_j)[y(j) - G^{\mathrm{T}}(x(j),\theta_j)\alpha_{j-1}]}{1 + \mu\|G(x(j),\theta_j)\|^2}, \tag{25}$$

$\alpha_{-1} = 0$, *then* $\{\alpha_j, \theta_j\}_{j=0}^k$ *is a solution of optimization problem* (23).

Proof. To solve (23), we need to find a minimum possible value of γ (say γ_0) and the corresponding causal estimation strategy $\{(\alpha_j,\theta_j), c\theta_j \geqslant h\}_{j=0}^k$ that achieves

$$\frac{\sum_{j=0}^k |G^{\mathrm{T}}(x(j),\theta_j^*)\alpha^* - G^{\mathrm{T}}(x(j),\theta_j)\alpha_j|^2}{\mu^{-1}\|\alpha^*\|^2 + \mu_\theta^{-1}\sum_{j=0}^k \|\theta_j^* - \theta_{j-1}\|^2 + \sum_{j=0}^k |n_j|^2} < \gamma_0^2,$$

for any vector α^* and for all sequences $\{\theta_j^*, c\theta_j^* \geqslant h\}_{j=0}^k$, $\{n_j\}_{j=0}^k$. It can be seen that any possible value of $\gamma > 0$, that satisfies the existence condition (10) $P_j^{-1} + (1-\gamma^{-2})G(x(j), \theta_j^*)G^{\mathrm{T}}(x(j),\theta_j^*) > 0$, $\forall j = 0,\ldots,k$, cannot be less than 1. To see this, note that $P_j^{-1} = \mu^{-1}I + (1-\gamma^{-2})\sum_{i=0}^{j-1} G(x(i),\theta_i^*) G^{\mathrm{T}}(x(i),\theta_i^*)$, and so the existence condition simplifies to $\mu^{-1}I + (1-\gamma^{-2})\sum_{i=0}^{j} G(x(i),\theta_i^*)G^{\mathrm{T}}(x(i),\theta_i^*) > 0$, $\forall j = 0,\ldots,k$. The above inequality implies that γ cannot be less than 1. To verify this, first note that each element of vector $G(\cdot)$ corresponds to a normalized firing strength of a rule. That is, each element of G lies between 0 and 1 and sum of all its elements is equal to 1. Thus, $\|G(x(i),\theta_i^*)\|^2 \geqslant 1/$number of elements of vector G, and hence $\sum_{i=0}^\infty \|G(x(i),\theta_i^*)\|^2 = \infty$. Now, suppose that $\gamma < 1$, then for some large enough j, we must have $\sum_{i=0}^{j} |G_{\mathrm{p}}(x(i),\theta_i^*)|^2 > \frac{\mu^{-1}}{\gamma^{-2}-1}$, where $G_{\mathrm{p}}(x(i),\theta_i^*)$ is the any pth entry of vector $G(x(i),\theta_i^*)$. This implies that the pth diagonal entry of the matrix $\mu^{-1}I + (1-\gamma^{-2})\sum_{i=0}^{j} G(x(i),\theta_i^*)G^{\mathrm{T}}(x(i),\theta_i^*)$ is negative and hence the expression $\mu^{-1}I + (1-\gamma^{-2})\sum_{i=0}^{j} G (x(i),\theta_i^*)G^{\mathrm{T}}(x(i),\theta_i^*)$ cannot be positive-definite. Therefore, $\gamma_0 = 1$. Once γ_0 has been determined, the analysis made in the previous section i.e. expressions (21)–(22) can be used to find an optimal estimation strategy. For $\gamma_0 = 1$, $P_j = \mu I$, and hence the optimal estimation strategy is given as (24)–(25). □

Theorem 3 suggests an estimation strategy (24)–(25) which is exactly the same as suggested in Kumar et al. (2006), however, by solving a local min–max regularized least-squares estimation problem. Now, we have motivated it as a solution of the energy-gain bounding problem. The convergence and steady-state behavior of identification method (24)–(25) has been studied in Kumar et al. (2006). However, for the sake of completion, we outline some of the results. For this, define priori recursion error $\widetilde{e}_{\mathrm{a}}(j)$ and posteriori recursion error $\widetilde{e}_{\mathrm{p}}(j)$ as $\widetilde{e}_{\mathrm{a}}(j) = G^{\mathrm{T}}(x(j),\theta_j)\tilde{\alpha}_{j-1}$, $\widetilde{e}_{\mathrm{p}}(j) = G^{\mathrm{T}}(x(j),\theta_j)\tilde{\alpha}_j$, $\tilde{\alpha}_j = \alpha^* - \alpha_j$. It follows from (25) that $\tilde{\alpha}_j = \tilde{\alpha}_{j-1} - \frac{\widetilde{e}_{\mathrm{a}}(j) - \widetilde{e}_{\mathrm{p}}(j)}{\|G(x(j),\theta_j)\|^2} G(x(j),\theta_j)$. By taking squared norm of both sides,

$$\|\tilde{\alpha}_j\|^2 + \frac{|\widetilde{e}_{\mathrm{a}}(j)|^2}{\|G(x(j),\theta_j)\|^2} = \|\tilde{\alpha}_{j-1}\|^2 + \frac{|\widetilde{e}_{\mathrm{p}}(j)|^2}{\|G(x(j),\theta_j)\|^2}. \tag{26}$$

To study the convergence properties, assume that $n_j = 0$, then it follows from (25) that $\widetilde{e}_{\mathrm{p}}(j) = \frac{\widetilde{e}_{\mathrm{a}}(j)}{1 + \mu\|G(x(j),\theta_j)\|^2}$, and thus from (26), we obtain $\|\tilde{\alpha}_j\|^2 = \|\tilde{\alpha}_{j-1}\|^2 - \left(1 - \left|\frac{1}{1+\mu\|G(x(j),\theta_j)\|^2}\right|^2\right) \frac{|\widetilde{e}_{\mathrm{a}}(j)|^2}{\|G(x(j),\theta_j)\|^2}$. Above expression shows the convergence property in a sense that squared norm of estimation error vector (i.e. $\|\tilde{\alpha}_j\|^2$) is a nonincreasing function of time index j.

6. Energy-gain bounding approach with time-varying learning rate

In order to meet the requirements of fast convergence and low misadjustment, one may like to use a variable or time-varying learning rate (i.e. $\mu(j)$ and $\mu_\theta(j)$) in recursions (24)–(25). Let us, for simplicity, assume that ratio of antecedents learning rate to consequents learning rate at any time is constant, i.e. $\mu_\theta(j)/\mu(j) = s_\theta$, where $s_\theta > 0$ is a constant. Let us also define priori error $e_{\mathrm{a}}(j)$ and posteriori error $e_{\mathrm{p}}(j)$ as $e_{\mathrm{a}}(j) = y(j) - G^{\mathrm{T}}(x(j),\theta_j)\alpha_{j-1}$, $e_{\mathrm{p}}(j) = y(j) - G^{\mathrm{T}}(x(j),\theta_j)\alpha_j$.

Theorem 4. *If any unknown physical process is modelled according to* (3) *by defining* $\theta_j^* = \theta_j$,

$$\theta_j = \arg\min_\theta [\Psi_j(\theta), c\theta \geqslant h], \tag{27}$$

$$\Psi_j(\theta) = \frac{[y(j) - G^{\mathrm{T}}(x(j),\theta)\alpha_{j-1}]^2}{1 + \mu(j)\|G(x(j),\theta)\|^2} + (\mu_\theta(j))^{-1}\|\theta - \theta_{j-1}\|^2,$$

$$\alpha_j = \alpha_{j-1} + \frac{\mu(j) G(x(j),\theta_j)\left[y(j) - G^{\mathrm{T}}(x(j),\theta_j)\alpha_{j-1}\right]}{1 + \mu(j)\|G(x(j),\theta_j)\|^2}, \tag{28}$$

$\alpha_{-1} = 0$, *then* $\{\alpha_j, \theta_j\}_{j=0}^k$ *is a solution of following two optimization problems*:

(1)

$$\min_{\{(\alpha_j,\theta_j),c\theta_j \geqslant h\}_{j=0}^k} \quad \max_{\{\alpha^*,\{\theta_j^*\}_{j=0}^k,\{n_j\}_{j=0}^k\}} \mathscr{T}, \tag{29}$$

$$\mathscr{T} = \frac{\sum_{j=0}^k \mu(j)|G^{\mathrm{T}}(x(j),\theta_j^*)\alpha^* - G^{\mathrm{T}}(x(j),\theta_j)\alpha_j|^2}{\|\alpha^*\|^2 + s_\theta^{-1}\sum_{j=0}^k \|\theta_j^* - \theta_{j-1}\|^2 + \sum_{j=0}^k \mu(j)|n_j|^2}.$$

(2) *minimize* $\|\alpha_j - \alpha_{j-1}\|^2$ *subject to*

$$e_{\mathrm{p}}(j) = \frac{e_{\mathrm{a}}(j)}{1 + \mu(j)\|G(x(j), \theta_j)\|^2}.$$

Proof. The proof of first part of the theorem follows by replacing in (24)–(25), $y(j)$ by $\sqrt{\mu(j)}y(j)$, $G(x(j), \theta_j)$ by $\sqrt{\mu(j)}G(x(j), \theta_j)$, $G(x(j), \theta)$ by $\sqrt{\mu(j)}G(x(j), \theta)$, μ by 1, μ_θ by s_θ, since these replacements of variables in optimization problem (23), after substituting $n_j = y(j) - G^{\mathrm{T}}(x(j), \theta_j^*)\alpha^*$, $\theta_j^* = \theta_j$, would lead to the optimization problem (29). For the proof of second part, consider the optimization problem: minimize$\|\alpha_j - \alpha_{j-1}\|^2$ subject to $e_{\mathrm{p}}(j) = e_{\mathrm{a}}(j)/1 + \mu(j)\|G(x(j), \theta_j)\|^2$. Define $J_1 = \|\alpha_j - \alpha_{j-1}\|^2 + \lambda[e_{\mathrm{p}}(j) - e_{\mathrm{a}}(j)/1 + \mu(j)\|G(x(j), \theta_j)\|^2]$, where λ is a Lagrange multiplier. The partial derivatives of J_1, $\frac{\partial J_1}{\partial \alpha_j} = 2(\alpha_j - \alpha_{j-1}) - \lambda G(x(j), \theta_j)$, $\frac{\partial J_1}{\partial \lambda} = e_{\mathrm{p}}(j) - e_{\mathrm{a}}(j)/1 + \mu(j)\|G(x(j), \theta_j)\|^2$, are set equal to zero. This results, after solving both equations, the optimal value of α_j as $\alpha_j = \alpha_{j-1} + \mu(j)G(x(j), \theta_j)/1 + \mu(j)\|G(x(j), \theta_j)\|^2 e_{\mathrm{a}}(j)$, which is exactly the same as (28). □

We study the variable learning rate approach in both stochastic and deterministic framework through an example for each. As an example of variable learning rate design in a deterministic framework, consider the problem of incorporating a kind of *dead-zone* in the energy-gain bounding approach. That is, at every time index j, the learning rate $\mu(j)$ should be so chosen such that

$$e_{\mathrm{p}}(j) = \begin{cases} \dfrac{\tau}{|e_{\mathrm{a}}(j)|} e_{\mathrm{a}}(j) & \text{if } |e_{\mathrm{a}}(j)| > \tau, \\ e_{\mathrm{a}}(j) & \text{if } |e_{\mathrm{a}}(j)| \leqslant \tau, \end{cases}$$

where a positive constant τ is the dead-zone applied to the identification scheme. Theorem 4 indicates that the solution of energy-gain bounding problem will guarantee the following relation between $e_{\mathrm{p}}(j)$ and $e_{\mathrm{a}}(j)$: $e_{\mathrm{p}}(j) = e_{\mathrm{a}}(j)/1 + \mu(j)\|G(x(j), \theta_j)\|^2$. Thus, dead-zone can be incorporated in energy-gain bounding approach by choosing $\mu(j)$ as

$$\mu(j) = \begin{cases} \dfrac{1}{\|G(x(j), \theta_j)\|^2}\left[\dfrac{|e_{\mathrm{a}}(j)|}{\tau} - 1\right] & \text{if } |e_{\mathrm{a}}(j)| > \tau, \\ 0 & \text{if } |e_{\mathrm{a}}(j)| \leqslant \tau. \end{cases} \tag{30}$$

The learning rate (30) being dependent upon $|e_{\mathrm{a}}(j)|$ should provide a good compromise between convergence speed and misadjustment error. However, this may not be the optimal learning rate. To find an optimum learning rate, we consider the problem in a stochastic framework. Defining consequents-error vector $\tilde{\alpha}_j = \alpha^* - \alpha_j$, the recursion (28) can be written as $\tilde{\alpha}_j = \tilde{\alpha}_{j-1} - \mu(j)G(x(j), \theta_j)/1 + \mu(j)\|G(x(j), \theta_j)\|^2 e_{\mathrm{a}}(j)$. That is,

$$\|\tilde{\alpha}_j\|^2 = \|\tilde{\alpha}_{j-1}\|^2 - 2\frac{\mu(j)G^{\mathrm{T}}(x(j), \theta_j)\tilde{\alpha}_{j-1}e_{\mathrm{a}}(j)}{1 + \mu(j)\|G(x(j), \theta_j)\|^2} + \left[\frac{\mu(j)\|G(x(j), \theta_j)\|e_{\mathrm{a}}(j)}{1 + \mu(j)\|G(x(j), \theta_j)\|^2}\right]^2.$$

Taking expectations,

$$E\|\tilde{\alpha}_j\|^2 = E\|\tilde{\alpha}_{j-1}\|^2 - 2E\left[\frac{\mu(j)G^{\mathrm{T}}(x(j), \theta_j)\tilde{\alpha}_{j-1}e_{\mathrm{a}}(j)}{1 + \mu(j)\|G(x(j), \theta_j)\|^2}\right] + E\left[\frac{\mu(j)\|G(x(j), \theta_j)\|e_{\mathrm{a}}(j)}{1 + \mu(j)\|G(x(j), \theta_j)\|^2}\right]^2.$$

If we define

$$\Delta(\mu(j)) = 2E\left[\frac{\mu(j)G^{\mathrm{T}}(x(j), \theta_j)\tilde{\alpha}_{j-1}e_{\mathrm{a}}(j)}{1 + \mu(j)\|G(x(j), \theta_j)\|^2}\right] - E\left[\frac{\mu(j)\|G(x(j), \theta_j)\|e_{\mathrm{a}}(j)}{1 + \mu(j)\|G(x(j), \theta_j)\|^2}\right]^2,$$

then $E\|\tilde{\alpha}_j\|^2 = E\|\tilde{\alpha}_{j-1}\|^2 - \Delta(\mu(j))$. In Shin, Sayed, and Song (2004), an optimal value of step-size has been derived by maximizing $\Delta(\mu(j))$ at every time index j, since this will guarantee that expected value of consequents-error norm will undergo the largest decrease from iteration $(j-1)$ to j. That is, to solve $\mu^*(j) = \arg\max_{\mu(j)>0}\Delta(\mu(j))$. Let us define learning rate $\mu(j) > 0$ using a number $\overline{\mu}$ (where $0 < \overline{\mu} < 1$) such that $\mu(j) = \overline{\mu}/\|G(x(j), \theta_j)\|^2(1 - \overline{\mu})$, $0 < \overline{\mu} < 1$, and

$$\Delta(\overline{\mu}) = 2\overline{\mu}E\left[\frac{G^{\mathrm{T}}(x(j), \theta_j)\tilde{\alpha}_{j-1}e_{\mathrm{a}}(j)}{\|G(x(j), \theta_j)\|^2}\right] - \overline{\mu}^2E\left[\frac{|e_{\mathrm{a}}(j)|^2}{\|G(x(j), \theta_j)\|^2}\right].$$

We formulate our problem as $\mu^*(j) = \overline{\mu}^*/\|G(x(j), \theta_j)\|^2(1 - \overline{\mu}^*)$, $\overline{\mu}^* = \arg\max_{0<\overline{\mu}<1}\Delta(\overline{\mu})$. Substituting $e_{\mathrm{a}}(j) = G^{\mathrm{T}}(x(j), \theta_j)\tilde{\alpha}_{j-1} + n_j$, $\Delta(\overline{\mu})$ becomes

$$\Delta(\overline{\mu}) = 2\overline{\mu}E\left[\frac{|G^{\mathrm{T}}(x(j), \theta_j)\tilde{\alpha}_{j-1}|^2 + G^{\mathrm{T}}(x(j), \theta_j)\tilde{\alpha}_{j-1}n_j}{\|G(x(j), \theta_j)\|^2}\right] - \overline{\mu}^2E\left[\frac{|G^{\mathrm{T}}(x(j), \theta_j)\tilde{\alpha}_{j-1} + n_j|^2}{\|G(x(j), \theta_j)\|^2}\right].$$

Assuming that zero-mean noise sequence $\{n_j\}$ is independent, identically distributed and statistically independent of regression sequence $\{G(x(j), \theta_j)\}$, $\Delta(\overline{\mu})$ can be approximated as

$$\Delta(\overline{\mu}) \approx 2\overline{\mu}E\left[\frac{|G^{\mathrm{T}}(x(j), \theta_j)\tilde{\alpha}_{j-1}|^2}{\|G(x(j), \theta_j)\|^2}\right] - \overline{\mu}^2\left(E\left[\frac{|G^{\mathrm{T}}(x(j), \theta_j)\tilde{\alpha}_{j-1}|^2}{\|G(x(j), \theta_j)\|^2}\right] + \sigma_n^2E\left[\frac{1}{\|G(x(j), \theta_j)\|^2}\right]\right), \tag{31}$$

where $\sigma_n = E|n_j|^2$. Maximizing (31) leads to

$$\overline{\mu}^* = \frac{E\left[\frac{|G^{\mathrm{T}}(x(j),\theta_j)\tilde{\alpha}_{j-1}|^2}{\|G(x(j),\theta_j)\|^2}\right]}{E\left[\frac{|G^{\mathrm{T}}(x(j),\theta_j)\tilde{\alpha}_{j-1}|^2}{\|G(x(j),\theta_j)\|^2}\right] + \sigma_n^2E\left[\frac{1}{\|G(x(j),\theta_j)\|^2}\right]},$$

$$\mu^*(j) = \frac{E\left[\frac{|G^{\mathrm{T}}(x(j),\theta_j)\tilde{\alpha}_{j-1}|^2}{\|G(x(j),\theta_j)\|^2}\right]}{\sigma_n^2\|G(x(j), \theta_j)\|^2E\left[\frac{1}{\|G(x(j),\theta_j)\|^2}\right]}.$$

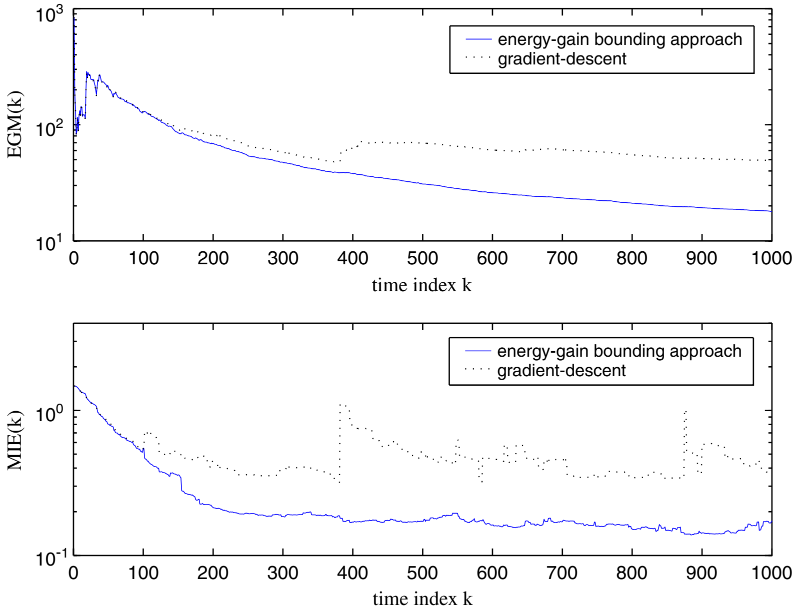

Fig. 3. Plot of $EGM(k)$ and $MIE(k)$ for energy-gain bounding approach and gradient-descent.

If we define a vector $p_j = G^{\mathrm{T}}(x(j),\theta_j)\tilde{\alpha}_{j-1}G(x(j),\theta_j)/\|G(x(j),\theta_j)\|^2$, then

$$\mu^*(j) = \frac{E\|p_j\|^2}{\sigma_n^2\|G(x(j),\theta_j)\|^2 E\left[\frac{1}{\|G(x(j),\theta_j)\|^2}\right]}.$$

Thus, to compute optimal learning rate $\mu^*(j)$, we first need to at least compute the term $E\|p_j\|^2$. To do this, we follow the approach of Shin et al. (2004) to estimate p_j as $\widehat{p}_j = \omega\widehat{p}_{j-1} + (1-\omega)e_{\mathrm{a}}(j)/\|G(x(j),\theta_j)\|^2 G(x(j),\theta_j)$, for a smoothing factor ω $(0<\omega<1)$, being motivated by the fact that $E[p_j] = E[e_{\mathrm{a}}(j)/\|G(x(j),\theta_j)\|^2 G(x(j),\theta_j)]$. Thus, we propose to estimate $\mu^*(j)$ as

$$\mu^*(j) = \frac{\|\widehat{p}_j\|^2}{C\|G(x(j),\theta_j)\|^2} \quad \text{where } C \text{ is a constant.} \tag{32}$$

7. Simulation studies

Our approach to the fuzzy identification of any unknown process $f(x)$ would be considered *good* if it results a small value of identification error defined as $IE(x,\alpha_j,\theta_j) = |f(x) - G^{\mathrm{T}}(x,\theta_j)\alpha_j|$. It was shown in Kumar et al. (2006) through different examples that estimation strategy (24)–(25) results a smaller value of identification error in comparison to other standard techniques including gradient-descent. Thus, we do not repeat the simulations for comparison with other techniques. For the sake of illustration of energy-gain bounding approach, consider first the fuzzy identification of a process $f(x) = 100x/(1+100x^2)$, $x \in [-2,2]$. The process was simulated by choosing x from a uniform distribution on the interval $[-2,2]$. The identification data is a sequence $\{x(j), f(x(j)) + \delta y_j\}$, where δy_j is a random noise, chosen from a uniform distribution on the interval $[-0.2, 0.2]$.

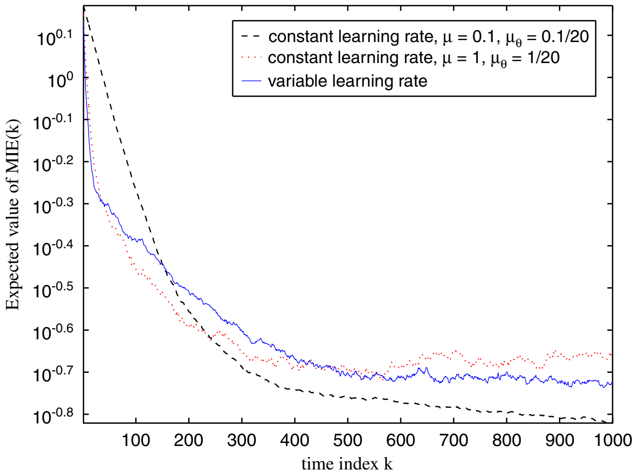

Fig. 4. Plot of the expected value of $MIE(k)$ for variable learning rate strategy.

To measure identification error, at any time index k, define mean value of identification error over interval $[-2,2]$ as $MIE(k) = (1/200)\sum_{l=1}^{200}|f(x^l) - G^{\mathrm{T}}(x^l,\theta_k)\alpha_k|$, where the points $\{x^l\}_{j=1}^{200}$ are uniformly distributed on $[-2,2]$. Also, we define an indirect measure of energy-gain, at any time index k, as $EGM(k) = \sum_{j=0}^{k}|f(x(j)) - G^{\mathrm{T}}(x(j),\theta_j)\alpha_j|^2/\sum_{j=0}^{k}|\delta y_j|^2$, assuming that modelling errors are small, i.e.$n_j \approx \delta y_j$. Let us consider the following estimation strategies: (1) constant learning rate, taking $\mu = 0.1$ and $\mu_\theta = 0.1/20$, (2) variable learning rate (30), taking $\tau = 2.2$ and $s_\theta = 0.0025$, (3) optimal learning rate (32), taking $C = 0.01$, $\omega = 0.99$, and $s_\theta = 0.025$. Let us choose a fuzzy model with eight membership functions. Initially, the membership functions are taken uniformly distributed in input range. The matrix c and vector h are defined such that two membership functions knots must be separated at least by a distance of 0.1. These estimation strategies can be implemented using a Gauss–Newton-based algorithm suggested in Kumar et al. (2006). The simulation results for strategy (24)–(25) and its comparison with gradient-descent have been shown in Fig. 3 by plotting the curves of $EGM(k)$ and $MIE(k)$. For a fair comparison, the gradient-descent step-size for linear parameters was 0.1 (equal to μ) and for nonlinear parameters was 0.1/20 (equal to μ_θ). The better performance of energy-gain bounding approach, as expected, could be seen in Fig. 3. Fig. 4 shows a plot of expected value of $MIE(k)$ for the variable and constant learning rate strategies. The simulation results are obtained by ensemble averaging over 100 independent trials. As seen from Fig. 4, the variable learning rate results in a compromise between convergence speed and misadjustment error. Finally, Fig. 5 shows the expected $MIE(k)$ curve (averaged over 100 independent trials), for the optimal and constant learning rate strategies. Fig. 5 clearly shows the best performance of optimal learning rate strategy in terms of fastconvergence and low misadjustment error. Now, we illustrate the effectiveness of our approach by considering a complex nonlinear function $f(x_1, x_2) = (1 - x_1x_2)\mathrm{e}^{-(x_1+x_2)^2} - \cos(4x_1x_2) + \log(1 + x_1x_2)$, $x_1 \in [-0.9, 0.9]$, $x_2 \in [-0.9, 0.9]$. The goal is to identify the function f using the identification data sequence $\{[x_1(j)x_2(j)]^{\mathrm{T}}, y(j)\}$ generated according to $y(j) = f(x_1(j), x_2(j)) + n_j$, where n_j is a random signal normally distributed with zero mean a variance of 0.01 and $\{x_1(j), x_2(j)\}$ are chosen from a uniform distribution on the interval $[-0.9, 0.9]$. Let us consider a fuzzy model that consists of nine different rules (i.e. three membership functions for each input). Again, the initial membership functions are taken uniformly distributed over each input range. The fuzzy model is identified using optimal learning rate with $C = 0.01$, $\omega = 0.99$, and $s_\theta = 0.025$. The mean identification error at time index k is defined as $MIE(k) = (1/100)\sum_{l=1}^{100} |f(x_1^l, x_2^l) - G^{\mathrm{T}}([x_1^l x_2^l]^{\mathrm{T}}, \theta_k)\alpha_k|$, where the points $\{[x_1^l x_2^l]^{\mathrm{T}}\}_{l=1,\dots,100}$ are uniformly distributed in the two-dimensional input space. The learning of model parameters has been shown in Fig. 6 by plotting $\|\alpha_k\|$ and$\|\theta_k - \theta_{-1}\|$ with k. Fig. 6 also shows the better performance of our approach than the standard gradient-descent technique (taking step-size equal to 0.1) in sense of mean identification error.

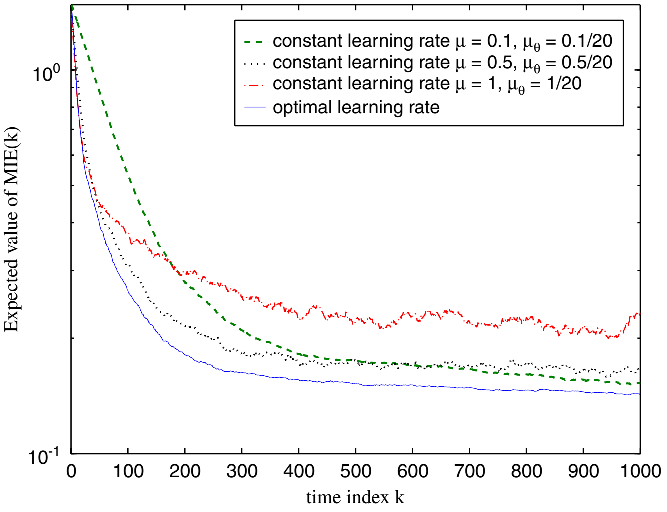

Fig. 5. Plot of the expected value of $MIE(k)$ for optimal learning rate strategy.

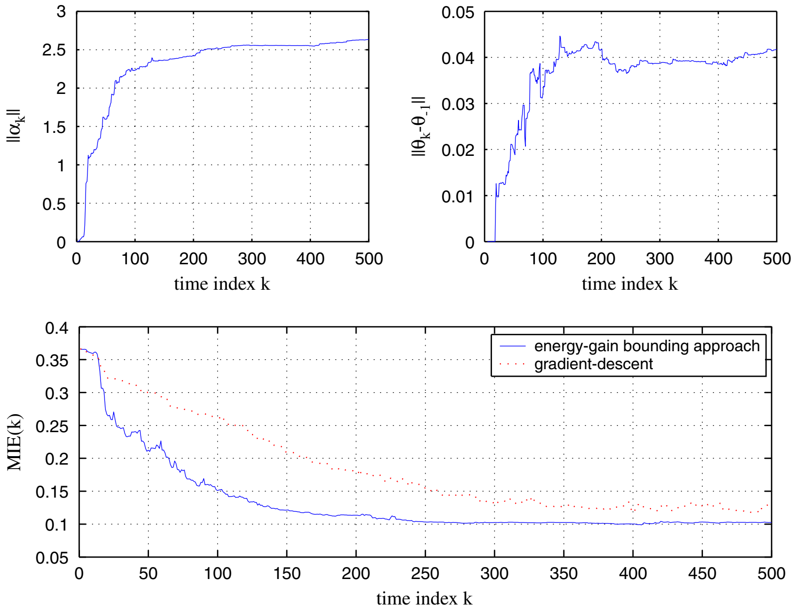

Fig. 6. The learning of fuzzy model parameters using noisy data.

8. Conclusion

This study has outlined a framework for on-line parameters identification of an interpretable fuzzy model in presence of data uncertainties and modelling errors without requiring a priori knowledge of upper bounds, statistics, and distribution of data uncertainties and modelling errors. The proposed approach has been illustrated through simulation studies.

References

Babuška, R. (2000). Construction of fuzzy systems-interplay between precision and transparency. In *Proceedings of ESIT 2000* (pp. 445–452). Germany: Aachen.

Bodenhofer, U., Bauer, P. (2000). Towards an axiomatic treatment of “interpretability”. In *Proceedings of IIZUKA2000* (pp. 334–339). Iizuka, 2000.

Burger, M., Engl, H. W., Haslinger, J., & Bodenhofer, U. (2002). Regularized data-driven construction of fuzzy controllers. *Journal of Inverse and Ill-posed Problems*, *10*, 319–344.

Chen, D. S., & Jain, R. C. (1994). A robust back propagation learning algorithm for function approximation. *IEEE Transactions on Neural Networks*, *5*, 467–479.

Espinosa, J., & Vandewalle, J. (2000). Constructing fuzzy models with linguistic integrity from numerical data-AFRELI algorithm. *IEEE Transactions on Fuzzy Systems*, *8*, 591–600.

Hassibi, B., Sayed, A. H., & Kailath, T. (1996a). Linear estimation in Krein spaces. I. Theory. *IEEE Transactions on Automatic Control*, *41*(1), 18–33.

Hassibi, B., Sayed, A. H., & Kailath, T. (1996b). Linear estimation in Krein spaces. II. Applications. *IEEE Transactions on Automatic Control*, *41*(1), 34–49.

Hassibi, B., Sayed, A. H., & Kailath, T. (1996c). H^{∞} optimality of the LMS algorithm. *IEEE Transactions on Signal Processing*, *44*(2), 267–280.

Hong, X., Harris, C. J., & Chen, S. (2004). Robust neurofuzzy rule base knowledge extraction and estimation using subspace decomposition combined with regularization and d-optimality. *IEEE Transactions on Systems, Man and Cybernetics B*, *34*(1), 598–608.

Johansen, T. (1996). Robust identification of Takagi–Sugeno–Kang fuzzy models using regularization. In *Proceedings of IEEE conference on Fuzzy Systems* (pp. 180–186). New Orleans, USA.

Kumar, M., Stoll, R., & Stoll, N., 2003a. Regularized adaptation of fuzzy inference systems. Modelling the opinion of a medical expert about physical fitness: An application. *Fuzzy Optimization and Decision Making, 2*.

Kumar, M., Stoll, R., & Stoll, N. (2003b). 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 Optimization and Decision Making*, *2*, 243–259.

Kumar, M., Stoll, R., & Stoll, N. (2003c). SDP and SOCP for outer and robust fuzzy approximation. In *Proceedings of the Seventh IASTED international conference on artificial intelligence and soft computing*, Banff, Canada, July 2003c.

Kumar, M., Stoll, R., & Stoll, N. (2004a). Robust adaptive identification of fuzzy systems with uncertain data. *Fuzzy Optimization and Decision Making*, *3*(3), 195–216.

Kumar, M., Stoll, R., & Stoll, N. (2004b). Robust solution to fuzzy identification problem with uncertain data by regularization. Fuzzy approximation to physical fitness with real world medical data: An application. *Fuzzy Optimization and Decision Making*, *3*(1), 63–82.

Kumar, M., Stoll, R., & Stoll, N. (2006). A robust design criterion for interpretable fuzzy models with uncertain data. *IEEE Transactions on Fuzzy Systems*, *14*(2), in press.

Lindskog, P. (1997). Fuzzy identification from a grey box modeling point of view. In H. Hellendoorn, & D. Driankov (Eds.), *Fuzzy model identification: Selected approaches* (pp. 1–5). Berlin, Germany: Springer.

Setnes, M., Babuška, R., & Verbruggen, H. B. (1998). Rule-based modeling: Precision and transparency. *IEEE Transactions on Systems, Man and Cybernetics—Part C: Applications and Reviews*, *28*(1), 165–169.

Shin, H. C., Sayed, A. H., & Song, W. J. (2004). Variable step-size nlms and affine projection algorithms. *IEEE Signal Processing Letters*, *11*(2),

Takagi, T., & Sugeno, M. (1985). Fuzzy identification of systems and its applications to modeling and control. *IEEE Transactions on Systems, Man, and Cybernetics*, *15*(1), 116–132.

Wang, W. Y., Lee, T. T., Liu, C. L., & Wang, C. H. (1997). Function approximation using fuzzy neural networks with robust learning algorithm. *IEEE Transactions on Systems, Man and Cybernetics B*, *27*, 740–747.

Yu, W., & Li, X. (2004). Fuzzy identification using fuzzy neural networks with stable learning algorithms. *IEEE Transactions on Fuzzy Systems*, *12*(3), 411–420.

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

He served as a research scientist in Institute of Occupational and Social Medicine, Rostock from 2001 to 2004. Currently, he is with the Center for Life Science Automation, Rostock. His research interests include robust adaptive fuzzy identification, fuzzy logic in medicine, and robust adaptive control.

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

He served as head of section analytical chemistry at the Academy of Sciences of GDR, Central Institute for Organic Chemistry till 1991. From 1992 to 1994, he was the associate director of Institute of Organic Catalysis, Rostock, Germany. Since 1994, he is a professor of measurement technology in the engineering faculty of Rostock University. From 1994 to 2000, he directed the Institution of Automation in Rostock University. He is also holding, since 2003, the position of vice president in Center for Life Science Automation, Rostock. His fields of interests include medical process measurement, lab automation, and smart systems and devices.

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, Germany in 1980, 1984, and 2002, respectively.

She is head of the Institute of Occupational and Social Medicine, Rostock, Germany. She is afaculty member in the medicine faculty and faculty associate in the College of Computer Science and Electrical Engineering of Rostock University. She also holds the adjunct faculty member position in the Industrial Engineering Department of North Carolina State University. Her research interests include occupational physiology, preventive medicine, and cardiopulmonary diagnostics.