International Journal of Sciences and Techniques of Automatic control & computer engineering
IJ-STA, Volume 7, N°1, April 2013, pp. 1908–1931.

# Development of parametric estimation methods for nonlinear stochastic systems described by Hammerstein mathematical models

**Samira KAMOUN**

Laboratory of Sciences and Techniques of Automatic Control and computer engineering (Lab–STA), National School of Engineering of Sfax, University of Sfax, B. P. W, 3038, Sfax, Tunisia.

kamounsamira@yahoo.fr

**Abstract.** *This paper treats the parametric estimation methods of nonlinear stochastic systems, which can be described by the discrete-time Hammerstein mathematical models. The dynamic linear block of the considered Hammerstein mathematical model is described by the ARMAX mathematical model, single-input single-output, with unknown slowly time-varying parameters. Two recursive parametric estimation methods are studied and compared. It is about the Recursive Extended Least Squares (RELS) method and the Recursive Approximated Maximum Likelihood (RAML) method. Different types of parametric estimation algorithms corresponding to the two methods are developed on the basis of the prediction error method. The convergence conditions and the techniques of the practical implementation of these algorithms are given. The results of a simulation study are included to illustrate the validity of the developed parametric estimation algorithms.*

**Key words:** *Nonlinear stochastic systems; Discrete-time Hammerstein mathematical model; RELS parametric estimation method; RAML parametric estimation method; Parametric estimation algorithms.*

## 1. Introduction

Over the past three decades, several studies dealing with the analysis, identification and control of nonlinear systems have been developed (see, e.g., Kamoun, 2003; Tlili and Mibar, 2007; Mâatoug *et al*., 2008; Hajji *et al*., 2008; Billings, 2013). The first step of the study of control law, which can apply to a nonlinear system, is the description of this system by a mathematical model. The parameters of this mathematical model must be estimated using an appropriate parametric estimation method (see, e.g., Kamoun, 2003; Billings, 2013).

The description of a nonlinear system by a mathematical model can present difficulties, because we must take account of the nature of the concerned non-linearity. Indeed, the nonlinear systems often have many different types. Several works relating to the description of nonlinear systems by various types of mathematical models were published in the literature (see, e.g., Billings, 1980;

This paper was recommended for publication in revised form by the Editor Staff.
Edition: CPU of Tunis, Tunisia, ISSN: 1737-7749

Kortmann and Unbehauen, 1988; Chen and Billings, 1989; Haber and Unbehauen, 1990; Haber and Keviczky, 1999; Kamoun, 2003; Billings, 2013).

We can classify mainly the mathematical models being able to describe the nonlinear systems in: state-space mathematical models, input-output mathematical models, mathematical models in series of functions and mathematical models in connected blocks.

We mainly consider two types of mathematical models in connected blocks, which can describe nonlinear systems. This is the Hammerstein mathematical model, which consists of a nonlinear static block followed by a linear dynamic block, and the Wiener mathematical model, which consists of a linear dynamic block followed by a static nonlinear block.

Several works were published in the literature, which treat the identification of nonlinear systems described by Hammerstein mathematical models (see, e.g., Eskinat *et al*., 1991; Rangan *et al*., 1995; Belforte and Gay, 2001) or Wiener mathematical models (see, e.g., Wigren, 1993; Kamoun, 2003). Note that different identification methods of nonlinear systems, which are described by other types of mathematical models, have been developed (Billings and Voon, 1986; Billings, 2013).

In this paper, we consider the nonlinear systems operating in a stochastic environment, which can be described by Hammerstein mathematical models. The Hammerstein mathematical model, which corresponds to a class of the mathematical models in connected blocks, consists of a static (also called zero memory) nonlinear block followed by a dynamic linear block. We suppose that the dynamic linear block of the nonlinear system is described by an input-output mathematical model of the type ARMAX (Auto-Regressive Moving Average with eXogenous input), linear, single-input single-output and with unknown slowly time-varying parameters.

Noting that several types of the Hammerstein mathematical models are used in the identification of the nonlinear systems, and this, by considering various configurations of the static nonlinear block (see, Haber and Keviczky, 1999; Vörös, 2002). In addition, different types of problems relating to the identification of the nonlinear systems described by the Hammerstein mathematical model were studied by several authors, while being based over various methods (see, e.g., Eskinat *et al*., 1991; Rangan *et al*., 1995; Al-duwaish Nazmul Karim, 1997; Li, 1999; Belforte and Gay, 2001; Kamoun, 2003; Billings, 2013).

The purpose of this study is to study recursive parametric estimation algorithms for the nonlinear systems operating in a stochastic environment, which can be described by a discrete-time Hammerstein mathematical model. We propose to formulate the parametric estimation problem by using the recursive methods, which are combined with the prediction error method. Thus, two recursive parametric estimation methods are analysed and compared. It is about the Recursive Extended Least Squares (RELS) method and the Recursive Approximated Maximum Likelihood (RAML) method. Different types of parametric estimation algorithms corresponding to the two methods are developed. The practical implementation of these algorithms is based of the knowledge of several experimental measures (couples of input-output) resulting from the considered nonlinear stochastic system.

The rest of this paper is organized as follows: In the Section 2, the description of a nonlinear stochastic system by the discrete-time Hammerstein mathematical model is considered. The formulation of the parametric estimation problem of nonlinear stochastic systems is treated in the Section 3, on the basis of the RELS and the RAML parametric estimation methods and by using the prediction error method. Parametric estimation algorithms are developed. The convergence conditions and the techniques of the practical implementation of these algorithms are given. A simulated example is treated in Section 4, which illustrate and valid the performances of the developed parametric estimation algorithms. Section 5 gives a conclusion.

## 2. Discrete-time Hammerstein mathematical model

The description of a nonlinear system by the Hammerstein mathematical model corresponds to a structure of mathematical model, which consists of a static nonlinear block followed by a dynamic linear block, such that shown in Figure 1.

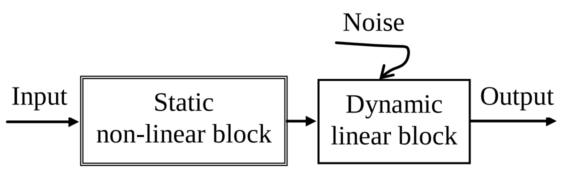


Figure 1. Structure of the Hammerstein mathematical model.

In all that will follow, we study the nonlinear stochastic systems that are described by a discrete-time Hammerstein mathematical model, where the dynamic linear block is described by the ARMAX mathematical model, single-input single-output, with unknown slowly time-varying parameters, but the structure parameters (order, delay) are well known.

The structure of the considered discrete-time Hammerstein mathematical model can be illustrated by Figure 2.

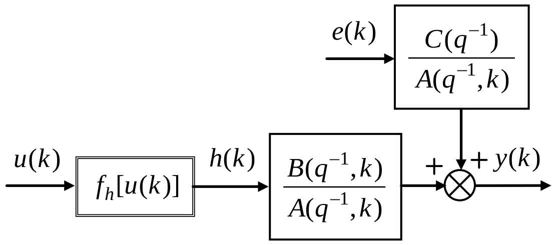


Figure 2. Structure of the considered discrete-time Hammerstein mathematical model.

The dynamic linear block of the considered Hammerstein mathematical model, as shown in Figure 2, is described by the following ARMAX mathematical model:

$$A(q^{-1},k)y(k) = B(q^{-1},k)h(k) + C(q^{-1})e(k) \tag{1}$$

where $h(k)$ and $y(k)$ are respectively the input and the output of the dynamic linear block at the discrete-time $k$, $e(k)$ represents the noise (together random disturbances) which can act on the output $y(k)$, and $A(q^{-1},k)$, $B(q^{-1},k)$ and $C(q^{-1})$ are polynomials, which are defined by:

$$A(q^{-1},k) = 1 + a_1(k)q^{-1} + \cdots + a_{n_A}(k)q^{-n_A} \tag{2}$$

$$B(q^{-1},k) = b_1(k)q^{-1} + \cdots + b_{n_B}(k)q^{-n_B} \tag{3}$$

and

$$C(q^{-1}) = 1 + c_1 q^{-1} + \cdots + c_{n_C} q^{-n_C} \tag{4}$$

where $n_A$, $n_B$ and $n_C$ represent the orders of the polynomials $A(q^{-1},k)$, $B(q^{-1},k)$ and $C(q^{-1})$, respectively.

We suppose that the noise sequence $\{e(k)\}$ is constituted by independent random variables with zero mean and variance $\sigma^2$.

Nothing the parameters intervening in the polynomials $A(q^{-1},k)$ and $B(q^{-1},k)$ are unknown slowly time-varying. However, the parameters intervening in the polynomial $C(q^{-1})$ are unknown, but constant. Also, the roots (in $q$) of the polynomial $C(q^{-1})$ are assumed to be strictly within the unit circle. This is related on the made assumption of the noise $e(k)$.

The static nonlinear block of the considered Hammerstein mathematical model is given by the following equation:

$$h(k) = f_h[u(k)] \tag{5}$$

in which $f_h[.]$ represents the nonlinear function.

The equation (5) can be approximated by a polynomial, such that:

$$h(k)=\sum_{r=1}^{p}\beta_r u^r(k)+\delta(k) \tag{6}$$

where $\delta(k)$ represents an approximation error of the nonlinear function, $\beta_r$, $r=1,\cdots,p$, are unknown parameters and $p$ indicates the degree of the non-linearity, which must be selected in a suitable way.

Let us note that the value of the approximation error $\delta(k)$ depends on the choice of the degree of the non-linearity $p$. Thus, if the value of the degree of the non-linearity $p$ is selected sufficiently large, then the value of this approximation error $\delta(k)$ becomes rather low. In this context, we suppose that the choice of the degree of the non-linearity $p$ is made in such way that the approximation error $\delta(k)$ can be neglected (i.e., $\delta(k)=0$).

From equations (1) and (6), which are relating to the linear dynamic block and the static nonlinear block of the Hammerstein mathematical model, respectively, and by neglecting the value of the approximation error $\delta(k)$, we can express the output $y(k)$ of the considered nonlinear system in the following form:

$$y(k)=-\sum_{i=1}^{n_A}a_i(k)y(k-i)+\sum_{j=1}^{n_B}\sum_{r=1}^{p}b_j(k)\beta_r u^r(k-j)+e(k)+\sum_{h=1}^{n_C}c_h e(k-h) \tag{7}$$

We suppose here that the polynomials $A(q^{-1},k)$, $B(q^{-1},k)$ and $C(q^{-1})$ of the ARMAX mathematical model (1) have the same order $n$ (i.e., $n=n_A=n_B=n_C$). This assumption does not constitute a restriction, but it was made for reasons of simplicity.

Taking into account the above assumption, the output $y(k)$ of the considered nonlinear stochastic system that is described by (7) can be expressed in the following developed form:

$$\begin{aligned}y(k)=&-a_1(k)y(k-1)-\cdots-a_n(k)y(k-n)+b_1(k)\beta_1u(k-1)\\&+b_2(k)\beta_1u(k-2)+\cdots+b_n(k)\beta_1u(k-n)+\cdots\ \cdots+\beta_p u^p(k-1)\\&+b_2(k)\beta_p u^p(k-2)+\cdots+b_n(k)\beta_p u^p(k-n)\\&+e(k)+c_1e(k-1)+\cdots+c_ne(k-n)\end{aligned} \tag{8}$$

## 3. Study of recursive parametric estimation methods

This section is relating to the parametric estimation of the nonlinear stochastic systems, which can be described by the discrete-time Hammerstein mathematical model (8). The formulation of this parametric estimation problem will be carried out while being based on the RELS parametric estimation method and the RAML parametric estimation method, by using the prediction error method and from the knowledge of the several measured values of the input and the output of the considered nonlinear system.

### 3.1. Formulation of the parametric estimation problem

This subsection is reserved to the formulation of the parametric estimation problem of the discrete-time Hammerstein mathematical model (8).

Let us notice that the structure of the discrete-time Hammerstein mathematical model (8) corresponds to a structure of a mathematical model being able to be made up of several inputs and only one output. This Hammerstein mathematical model contains the parameters $b_j(k)$, $j=1,\cdots,n$, which are nonlinear with respect to the parameters $\beta_r$, $r=1,\cdots,p$. However, this Hammerstein mathematical model is linear with respect to the parameters $a_i(k)$, $i=1,\cdots,n$. Let us add that there exists a certain redundancy of these parameters $b_j(k)$ and $\beta_r$. Consequently, we can affirm that the formulation of a parametric estimation scheme, in order to obtain estimated parameters with a minimum variance, presents serious difficulties, particularly in the practical implementation.

In addition, some of these difficulties can be surmounted if we consider the two following situations:

1. the knowledge of one of the parameters $b_j(k)$, $j=1,\cdots,n$. For example, let us suppose that the parameter $b_1(k)$ is well known at any discrete-time $k$. Thus, we can directly determine the estimated parameters $\hat{\beta}_r(k)$, $r=1,\cdots,p$. Consequently, we can easily determine the other parameters $b_2(k),\cdots,b_n(k)$, while basing ourselves on the knowledge of these estimated parameters $\hat{\beta}_r(k)$;
2. the knowledge of one of the parameters $\beta_r$, $r=1,\cdots,p$. Let us suppose, for example, that the parameter $\beta_1$ is known. Then, we can directly deduce from them the estimated parameters $\hat{b}_j(k)$, $j=1,\cdots,n$. While basing itself on the knowledge of these estimated parameters, we can determine easily the parameters $\beta_2,\cdots,\beta_n$.

For reasons of simplicity of the formulation of the posed parametric estimation problem, in order to develop the recursive parametric estimation algorithms on the basis of the RELS and the RAML methods, where their practical implementation can be ensured, we must consider one of the above situations.

Thus, we propose here to consider the first situation, where the parameter $b_1(k)$ intervening in the discrete-time Hammerstein mathematical model is supposed to be known time-invariant, such as: $b_1(k)=1,\ \forall k$.

By taking account of this assumption, we can define the output $y(k)$ of the considered nonlinear stochastic system in the following matrix form:

$$y(k)=\theta^T(k)\psi(k)+e(k) \tag{9}$$

where the true parameter vector $\theta(k)$ and the observation vector $\psi(k)$ are given by:

$$\begin{aligned}\theta^T(k) = [&a_1(k),\cdots,a_n(k),\beta_1,b_2(k)\beta_1,\cdots,b_n(k)\beta_1,\cdots\ \cdots,\\ &\beta_p,b_2(k)\beta_p,\cdots,b_n(k)\beta_p,c_1,\cdots,c_n]\end{aligned} \tag{10}$$

and

$$\begin{aligned}\psi^T(k) = [&-y(k-1),-\cdots,-y(k-n),u(k-1),u(k-2),\cdots,u(k-n),\cdots\ \cdots\\ &u^p(k-1),u^p(k-2),\cdots,u^p(k-n),e(k-1),\cdots,e(k-n)]\end{aligned} \tag{11}$$

Let us note that the dimension of these vectors $\theta(k)$ and $\psi(k)$ increases of the value $n$ when the degree of the non-linearity $p$ increases of the value 1.

We propose to use the two following notations:
$f_{21}(k)\ldots f_{n1}(k) = b_2(k)\beta_1 \ldots b_n(k)\beta_1$, $f_{2p}(k)\ldots f_{np}(k) = b_2(k)\beta_p \ldots b_n(k)\beta_p$.

Thus, the parameter vector $\theta(k)$ can be defined by:

$$\begin{aligned}\theta^T(k) = [&a_1(k),\cdots,a_n(k),\beta_1, f_{21}(k),\cdots, f_{n1}(k),\cdots\ \cdots,\beta_p\\ &f_{2p}(k),\cdots, f_{np}(k),c_1,\cdots,c_n]\end{aligned} \tag{12}$$

with: $f_{21}(k) = b_2(k)\beta_1$, etc.

The posed problem consists to estimate the parameters intervening in the discrete-time Hammerstein mathematical model (8), by using the RELS and the RAML parametric estimation methods, while basing itself on the prediction error method.

The prediction error method is based on the minimization of a certain quadratic criterion, which corresponds to the difference between the output of the nonlinear stochastic system and that predicted by the adjustable model.

The *a priori* predicted output $\hat{y}(k)$ by the adjustable model of the output $y(k)$ of the nonlinear stochastic system can be defined by the following expression:

$$\begin{aligned}\hat{y}(k) = &-\hat{a}_1(k-1)y(k-1)-\cdots-\hat{a}_n(k-1)y(k-n)+\hat{\beta}_1(k-1)u(k-1)\\ &+\hat{f}_{21}(k-1)u(k-2)+\cdots+ f_{n1}(k-1)u(k-n)+\cdots\ \cdots\\ &+\hat{\beta}_p(k-1)u^p(k-1)+\hat{f}_{2p}(k-1)u^p(k-2)+\cdots+\hat{f}_{np}(k-1)u^p(k-n)\\ &+\hat{c}_1(k-1)\varepsilon(k-1)+\cdots+\hat{c}_n(k-1)\varepsilon(k-n)\end{aligned} \tag{13}$$

or in the following compact form:

$$\hat{y}(k) = \hat{\theta}^T(k-1)\hat{\psi}(k) \tag{14}$$

with $\hat{\theta}(k-1)$ the estimated parameter vector at the discrete-time $k-1$ and $\hat{\psi}(k)$ corresponds to the prediction observation vector $\psi(k)$, where the noise sequence $\{e(k-h);\, h=1,\cdots,n\}$ is replaced by the *a priori* prediction error sequence $\{\varepsilon(k-h);\, h=1,\cdots,n\}$, such that:

$$\widehat{\psi}^T(k) = [-y(k-1), \cdots, -y(k-n), u(k-1), u(k-2), \cdots, u(k-n), \cdots \cdots \\ u^p(k-1), u^p(k-2), \cdots, u^p(k-n), \varepsilon(k-1), \cdots, \varepsilon(k-n)] \tag{15}$$

where $\varepsilon(k)$ denotes the *a priori* prediction error, which can be given by:

$$\varepsilon(k) = y(k) - \hat{\theta}^T(k-1)\widehat{\psi}(k) \tag{16}$$

The estimated parameter vector $\hat{\theta}(k)$ at the discrete-time $k$ can be defined by:

$$\hat{\theta}^T(k) = [\hat{a}_1(k), \cdots, \hat{a}_n(k), \hat{\beta}_1(k), \hat{f}_{21}(k), \cdots \\ \hat{f}_{n1}(k), \cdots \cdots, \hat{\beta}_p(k), \hat{f}_{2p}(k), \cdots, \hat{f}_{np}(k)] \tag{17}$$

The structure of the parametric estimation of the considered nonlinear stochastic system, by using the prediction error method, can be shown in the following Figure 3:

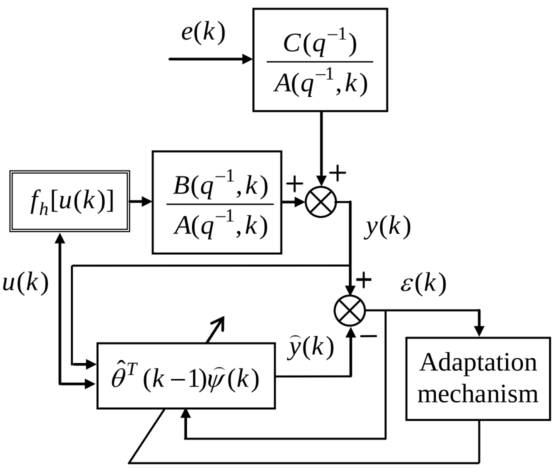


Figure 3. Structure of the parametric estimation of the considered non-linear stochastic system by using the prediction error method.

It important to indicate that the *a priori* prediction error $\varepsilon(k)$ corresponds to the best estimate of the noise $e(k)$, notably if the value of the discrete-time $k$ is sufficiently large (i.e., if $k \to \infty$, then $\varepsilon(\infty) \approx e(\infty)$).

Nothing that the adjustment of the estimated parameters of the adjustable model, at every discrete-time $k$, is ensured by the adaptation mechanism, while taking care to minimize the difference between the output $y(k)$ of the nonlinear stochastic system and the predicted output $\widehat{y}(k)$ of the adjustable model.

In the two following subsections, we will develop the RELS and the RAML parametric estimation methods, which can by applied to the considered nonlinear stochastic system that is described by the discrete-time Hammerstein mathematical model (8).

### 3.2. Recursive extended least squares method

This subsection is reserved to develop a parametric estimation algorithm on the basis of the RELS parametric estimation method, which is combined with the prediction error method. This algorithm must allow the estimate of the unknown slowly time-varying parameters intervening in the parameter vector $\theta(k)$, as described by (10) or (12).

By examining the discrete-time Hammerstein mathematical model (9), and taking into account that the noise sequence $\{e(k)\}$ is constituted by independent random variables with zero mean and variance $\sigma^2$, it results that the ordinary Recursive Least Squares (RLS) method can be used to estimate the parameter vector $\theta(k)$, which is defined by (10), while ensuring unbiased estimate. However, the implementation of the RLS parametric estimation method is not possible, since that the observation vector $\psi(k)$ that given by (11) contains the noise sequence $\{e(k-h); h=1,\cdots,n\}$, which is not measurable. To overcome this difficulty, we can use an extension of the RLS parametric estimation method. This corresponds to the RELS parametric estimation method. In this case, we must use the prediction observation vector $\widehat{\psi}(k)$, which is given by (15), in the RELS parametric estimation algorithm (see, Kamoun, 2003).

In fact, the RELS parametric estimation method corresponds to an extension of the RLS parametric estimation method, where the parameter vector and the observation vector are extended. In this direction, the extension corresponds to the estimate of the parameters $c_h$, $h=1,\cdots,n$, who have a connection with the noise sequence $\{e(k-h); h=1,\cdots,n\}$.

In the case of a nonlinear stochastic system with unknown time-invariant parameters, we can show easily that the RELS parametric estimation algorithm, which can estimate these parameters, is defined by:

$$\begin{aligned}
\hat{\theta}(k) &= \hat{\theta}(k-1) + P(k)\widehat{\psi}(k)\varepsilon(k) \\
P(k) &= P(k-1) - \frac{P(k-1)\widehat{\psi}(k)\widehat{\psi}^T(k)P(k-1)}{1+\widehat{\psi}^T(k)P(k-1)\widehat{\psi}(k)} \\
\varepsilon(k) &= y(k) - \hat{\theta}^T(k-1)\widehat{\psi}(k)
\end{aligned} \tag{18}$$

where $P(k)$ is an adaptation gain matrix, which corresponds to the variance-covariance matrix of the noise $e(k)$.

It is obvious that the RELS parametric estimation algorithm (18) cannot apply to a nonlinear stochastic system with unknown slowly time-varying parameter, as given by (9). This comes owing to the fact that the structure of the RELS parametric estimation algorithm (18) is not likely to follow the parametric variations of such system.

The principal element intervening in the structure of the RELS parametric estimation algorithm (18) is the adaptation gain matrix $P(k)$, which intervenes in the correction tem (i.e., $P(k)\widehat{\psi}(k)\varepsilon(k)$) of this algorithm. In fact, the computation procedure of the adaptation gain matrix $P(k)$, can condition the quality of the estimate (see, Kamoun, 2003). Thus, the trace of the adaptation gain matrix $P(k)$ decrease rather quickly towards a weak value (e.g., $\mathrm{tr}[P(0)=10000]$ and $\mathrm{tr}[P(20)=1]$, where $\mathrm{tr}[.]$ denotes the trace); consequently, the RELS parametric estimation algorithm (18) is not able to follow the time-variation of the system parameters for being likely to ensure a better quality of the estimate.

To solve the above problem in order to obtain a RELS parametric estimation algorithm, which can be applied to the considered nonlinear stochastic system as given by (9), we must use an appropriate computation procedure of the adaptation gain matrix of this algorithm. Several approaches of the computation procedure of the adaptation gain matrix intervening in the parametric estimation algorithms are proposed in the literature (see, e.g., Ljung and Gunnarsson, 1990; Guo and Ljung, 1995; Kamoun, 2003). Among these approaches, we will consider here that which is based on the introduction of a forgetting factor in the adaptation gain matrix.

The RELS parametric estimation method, which consists to introduce a forgetting factor into the adaptation gain matrix of the parametric estimation algorithm, makes it possible to improve the capacity of this matrix, wile ensuring best followed the slowly time-varying parameters of the considered nonlinear stochastic system, which is described by (9).

Consequently, the estimate of the parameter vector $\theta(k)$, as described by (10), can be ensured by using the following RELS parametric estimation algorithm with variable forgetting factor:

$$\begin{aligned}
&\hat{\theta}(k) = \hat{\theta}(k-1) + P(k)\widehat{\psi}(k)\varepsilon(k) \\
&P(k) = \frac{1}{\lambda(k)}\left[P(k-1) - \frac{P(k-1)\widehat{\psi}(k)\widehat{\psi}^T(k)P(k-1)}{\lambda(k)+\widehat{\psi}^T(k)P(k-1)\widehat{\psi}(k)}\right] \\
&\varepsilon(k) = y(k) - \hat{\theta}^T(k-1)\widehat{\psi}(k)
\end{aligned} \qquad (19)$$

where $\lambda(k)$ is a variable forgetting factor, such that: $0<\lambda(k)<1$.

The choice of the variable forgetting factor $\lambda(k)$ must be made in a suitable way. It can be chosen constant or time-varying parameter, and this, according the type of the considered application. We propose to compute the variable forgetting factor $\lambda(k)$ by using the following recursive equation:

$$\lambda(k) = \lambda_o \lambda(k-1) + \lambda^o (1-\lambda_o) \tag{20}$$

with: $0 < \lambda_o < 1$, $0 < \lambda^o < 1$.

In this case, we can show easily that:

$$\lim_{k \to \infty} \lambda(k) = \lambda^o \tag{21}$$

Nothing that the RELS parametric estimation algorithm (19) is based on the *a priori* prediction error $\varepsilon(k)$. However, in certain industrial applications, we may find it beneficial to use an *a posteriori* prediction error $\varepsilon_\circ(k)$, which is computed from the knowledge of the estimated parameters at the same time $k$, such that:

$$\varepsilon_\circ(k) = y(k) - \hat{\theta}^T(k)\hat{\psi}(k) \tag{22}$$

Thus, we can define the following *a posteriori* prediction observation vector $\hat{\psi}_\circ(k)$, which is based on the knowledge of the *a posteriori* prediction error sequence $\{\varepsilon_\circ(k-h); h = 1, \cdots, n\}$:

$$\begin{aligned} \hat{\psi}_\circ^{\,T}(k) = [&-y(k-1), \cdots, -y(k-n), u(k-1), u(k-2), \cdots, u(k-n), \cdots \cdots \\ & u^p(k-1), u^p(k-2), \cdots, u^p(k-n), \varepsilon_\circ(k-1), \cdots, \varepsilon_\circ(k-n)] \end{aligned} \tag{23}$$

The estimate of the parameter vector $\theta(k)$, as described by (10), can be also ensured by using the following RELS parametric estimation algorithm with variable forgetting factor (*a posteriori* version):

$$\begin{aligned} &\hat{\theta}(k) = \hat{\theta}(k-1) + P(k)\hat{\psi}_\circ(k)\varepsilon(k) \\ &P(k) = \frac{1}{\lambda(k)}\left[ P(k-1) - \frac{P(k-1)\hat{\psi}_\circ(k)\hat{\psi}_\circ^{\,T}(k)P(k-1)}{\lambda(k) + \hat{\psi}_\circ^{\,T}(k)P(k-1)\hat{\psi}_\circ(k)} \right] \\ &\varepsilon(k) = y(k) - \hat{\theta}^T(k-1)\hat{\psi}_\circ(k) \end{aligned} \tag{24}$$

We have studied the convergence conditions of the RELS parametric estimation algorithm (19) by using the ordinary differential equation approach, which is proposed initially by Ljung (1977). We can show that: if the vectors $\hat{\theta}(0)$ and $\hat{\psi}(k)$ are bounded, the adaptation gain matrix $P(k)$ is decreasing, the input $u(k)$ is a persistently and sufficiently exciting signal, the noise $e(k)$ consists of a sequence of an independent random variables with zero mean and variance $\sigma^2$ and the following condition is satisfied:

$$\frac{1}{C(q^{-1})} - \frac{1}{2} > 0 \tag{25}$$

then, the convergence of the RELS parametric estimation algorithm (19) is ensured.

### 3.3. Recursive approximated maximum likelihood method

We propose here to develop a parametric estimation algorithm, which can estimate the unknown slowly time-varying parameters of the nonlinear stochastic system that is described by the discrete-time Hammerstein mathematical model (9), on the basis of the RAML parametric estimation method.

According to Åström (1980), the maximum likelihood method was suggested initially by Gauss in 1809; then, it was developed by Ficher in 1912. This method consists in building a probability function, which depends on the experimental measurements (couples of input-output) and the unknown system parameters. The values of the unknown parameters are obtained as being the values that maximize this probability function.

The maximum likelihood method is used by Billings and Voon (1986) for estimating the parameters of the nonlinear stochastic systems that is described by the NARMAX mathematical model, where a no recursive parametric estimation algorithm is developed. The major inconvenient of this algorithm is that its implementation in the online estimation is not possible (particularly in an adaptive control scheme). To overcome this difficulty, we can use an approximation of the maximum likelihood method, which is called in the engineering literature ″approximated maximum likelihood method″.

In this subsection, we propose to use the approximated maximum likelihood method in order to develop a recursive parametric estimation algorithm, which can be applied to the considered nonlinear stochastic system.

The nonlinear stochastic system, which is described by the discrete-time Hammerstein mathematical model (9), can also to be described by the conditional distribution (or probability density function) of the output $y(k)$ given all the past input and output measurements. We assume that the probability distribution of the input-output data is known.

Let us define the following input and output measurement vectors, respectively:

$$U^T(k-1) = [u(1), \cdots, u(k-1)] \tag{26}$$

and

$$Y^T(k) = [y(1), \cdots, y(k)] \tag{27}$$

Thus, the conditional distribution of the output $y(k)$ can by define as follows, and this, on the basis of the all past input and output data:

$$p(y(k)/Y(k-1), U(k-1)) \tag{28}$$

where $p(.)$ denotes the probability.

The function (28) can be given in its predicted form, such that:

$$y(k) = f[Y(k-1), U(k-1)] + \varepsilon(k) \tag{29}$$

where the quantity $f[Y(k-1), U(k-1)]$ corresponds to the predicted value (or the mean-square error estimate) $\hat{y}(k)$ of the output $y(k)$, which can by give by:

$$\hat{y}(k) = \mathcal{E}[y(k)/Y(k-1), U(k-1)] \tag{30}$$

where the symbol $\mathcal{E}$ denotes the expectation.

The formulation of the posed parametric estimation problem will be carried out starting from the minimization of the prediction error $\varepsilon(k)$, while basing on the knowledge of the conditional distribution of the output $y(k)$ of the considered nonlinear stochastic system. In this case, the probability function can be expressed as being the product of the conditional densities of the prediction error sequence $\{\varepsilon(k)\}$.

The estimate of the parameter vector $\theta(k)$, defined by (10), by using the recursive approximated maximum likelihood method, consists of the minimization of the conditional probability of this parameter vector from the knowledge of all the values of the input and the output by considering a work horizon $M$ (i.e., $k = 1, \cdots, M$). This amounts writing: $p(\theta(k)/Y(M), U(M-1))$. The parameter vector $\theta(k)$ being deterministic, the use of the Bayes' rule makes it possible to affirm that: to maximize $p(\theta(k)/Y(M), U(M-1))$ is equivalent to maximize $p(y(k)/Y(M-1), U(M-1); \theta(k))$. The likelihood function $L(.)$ can be defined by:

$$L(\theta(k); Y(M), U(M-1)) = p(Y(M)/U(M-1), \theta(k)) \tag{31}$$

or

$$L(\theta(k); Y(M), U(M-1)) = \prod_{k=1}^{M} p(y(k)/Y(k-1), U(k-1); \theta(k)) \tag{32}$$

We already supposed that the noise $e(k)$, which acts on the considered nonlinear stochastic system, as described by the discrete-time Hammerstein mathematical (9), consists of a sequence of independent random variables with zero mean and variance $\sigma^2$. Thus, this noise $e(k)$ has a Gaussian density probability. Taking into account this assumption, we can write the likelihood function (31) in the following form:

$$L(\theta(k); Y(M), U(M-1)) = \frac{1}{[\sigma\sqrt{2\pi}]^M} \exp(-\frac{1}{2\sigma^2} \sum_{k=1}^{M} \varepsilon^2(k)) \tag{33}$$

By using the logarithm of this likelihood function, we can write:

$$\log L(\theta(k); Y(M), U(M-1)) = -\frac{1}{2\sigma^2} \sum_{k=1}^{M} \varepsilon^2(k) - M \log \sigma - \frac{M}{2} \log 2\pi \tag{34}$$

Consequently, the maximization of the likelihood function is reduced to the minimization of the following quadratic criterion $J(M, \theta(k))$:

$$J(M, \theta(k)) = \frac{1}{2} \sum_{k=1}^{M} \varepsilon^2(k) \tag{35}$$

We can show that the parametric estimation algorithm with variable forgetting factor, which permits to estimate the parameter vector $\theta(k)$, as described by (10), on the basis of the RAML parametric estimation method, is defined by:

$$\begin{aligned}
&\hat{\theta}(k) = \hat{\theta}(k-1) + P(k)\varphi(k)\varepsilon(k) \\
&P(k) = \frac{1}{\lambda(k)}\left[P(k-1) - \frac{P(k-1)\varphi(k)\varphi^T(k)P(k-1)}{\lambda(k)+\varphi^T(k)P(k-1)\varphi(k)}\right] \\
&\varepsilon(k) = y(k) - \hat{\theta}^T(k-1)\hat{\psi}(k)
\end{aligned} \tag{35}$$

where $\varphi(k)$ is a filtered observation vector, which is defined by:

$$\begin{aligned}
\varphi^T(k) = [&-\tilde{y}(k-1),\cdots,-\tilde{y}(k-n),\tilde{u}(k-1),\tilde{u}(k-2),\cdots,\tilde{u}(k-n),\cdots \cdots \\
&\tilde{u}^p(k-1),\tilde{u}^p(k-2),\cdots,\tilde{u}^p(k-n),\tilde{\varepsilon}(k-1),\cdots,\tilde{\varepsilon}(k-n)]
\end{aligned} \tag{37}$$

in that , the term $\tilde{y}(k)$ denotes a filtered signal at the discrete-time $k$, such that:

$$\tilde{y}(k) = \frac{y(k)}{\hat{C}(q^{-1},k-1)} \tag{38}$$

with $\hat{C}(q^{-1},k-1)$ corresponds to the estimated polynomial $C(q^{-1})$ at the discrete-time $k-1$.

We can write the filtered observation vector $\varphi(k)$ as follows:

$$\varphi(k) = \frac{\hat{\psi}(k)}{\hat{C}(q^{-1},k-1)} \tag{39}$$

We have studied the convergence conditions of the RAML parametric estimation algorithm (36) by using the ordinary differential equation approach. We can show that: if the vectors $\hat{\theta}(0)$ and $\hat{\psi}(k)$ are bounded, the adaptation gain matrix $P(k)$ is decreasing, the input $u(k)$ is a persistently and sufficiently exciting signal, the noise $e(k)$ consists of a sequence of an independent random variables with zero mean and variance $\sigma^2$ and the following condition is satisfied:

$$\frac{\hat{C}(q^{-1},k-1)}{C(q^{-1})} - \frac{1}{2} > 0 \tag{40}$$

then, the convergence of the RAML parametric estimation algorithm (36) is ensured.

It should be stressed that the convergence condition (40) of the RAML parametric estimation algorithm (36) can always be satisfied (in particular if the value of the discrete-time $k$ is sufficiently large), since the parameters intervening in the polynomial $\hat{C}(q^{-1},k-1)$ represent the best estimated values of the parameters intervening in the polynomial $C(q^{-1})$.

### 3.4. Techniques of the practical implementation of the developed algorithms

The practical implementation of the developed RELS and the RAML parametric estimation algorithms requires their initialisation, i.e., the choice of the initial conditions of the estimated parameters $\hat{\theta}(0)$ and the adaptation gain matrix $P(0)$. However, there do not exist rigorous methods for a better choice of these initial conditions. Thus, we can consider, for example, the following choice: $\hat{\theta}(0)=0$ and $P(0)=\alpha I$, with $\alpha$ a positive parameter and $I$ an identity matrix.

Let us note that in the estimate procedure of the parameters intervening in the discrete-time Hammerstein mathematical (9), by using the developed RELS or RAML parametric estimation algorithm, we can have a certain redundancy of the parameters $b_j(k)$, $j=2,\cdots,n$, and $\beta_r$, $r=1,\cdots,p$, intervening in the parameter vector $\theta(k)$, defined by (10), that must be estimated. In such a situation, we can choose a procedure allowing the selection of the values of these parameters, within the meaning of the minimization of the variance of the estimate.

We will propose, hereafter, a procedure allowing a better selection of the values of the parameters $b_j(k)$, $j=2,\cdots,n$, and $\beta_r$, $r=1,\cdots,p$. However, and before undertaking this procedure, we can state the following remarks, which are relating to the estimate of the parameters intervening in the vector $\theta(k)$, defined by (12), by using the RELS or the RAML parametric estimation algorithm:

1. the estimated parameters are unbiased, since the sequence $\{u(k),y(k);k=1,\cdots,M\}$ of the input $u(k)$ and the output $y(k)$ is not correlated with the noise sequence $\{e(k);k=1,\cdots,M\}$;
2. the estimated parameters $\hat{a}_i(k)$, $i=1,\cdots,n$, can be directly given;
3. a first whole of the estimated parameters $\hat{\beta}_r(k)$, $r=1,\cdots,p$, can be directly given;
4. a second whole of the parameters $\beta_r$, which is related to the parameters $b_j(k)$, $j=2,\cdots,n$, cannot be directly given, since these parameters must be estimated. Note that we can estimate all parameters $f_{jr}(k)=b_j(k)\beta_r$;
5. the values of the parameters $b_j(k)$, $j=2,\cdots,n$, which are related to the parameters $\beta_r$, $r=1,\cdots,p$, cannot be directly given. However, the determination of its parameters can be based on the values of the estimated parameters $\hat{f}_{jr}(k)$.

It is important to indicate that the parameters $f_{jr}(k)=b_j(k)\beta_r$, $j=2,\cdots,n$, $r=1,\cdots,p$, which are given in the parameter vector $\theta(k)$, as given by (12), contain a certain redundancy of the parameters $b_j(k)$ and $\beta_r$. Thus, there exists a redundancy of the estimated parameters intervening in $\hat{\theta}(k)$, as given by (17). In such a situation, we must establish a performance criterion (within the meaning of ensuring a better quality of the estimate) in order to select the best values of these estimated parameters.

Thus, we propose to determine the unknown values of the parameters $b_j(k)$, $j=2,\cdots,n$, and this, while basing ourselves on the values of the estimated parameters $\hat{\beta}_r(k)$, $r = 1, \cdots, p$, and $\hat{f}_{jr}(k)$. In this context, we can determine the value of the $r^{\text{th}}$ parameter, which is noted $b_{jr}(k)$, corresponding to the parameter $b_j(k)$. The parameter $b_{jr}(k)$ can be deduced from the estimated parameter $\hat{f}_{jr}(k)$ and the knowledge of the estimated parameter $\hat{\beta}_r(k)$. Thus, we can define the deduced parameter $b_{jr}(k)$ by be following equation:

$$b_{jr}(k) = \frac{\hat{\theta}[k:\ n+(r-1)n+j]}{\hat{\theta}[k:\ n+(r-1)n+1]} \tag{41}$$

where $\hat{\theta}[k:\ h]$ represents the $r^{\text{th}}$ component of the estimated parameter vector $\hat{\theta}(k)$, which is described by (17), and with $1 \le r \le p$ and $2 \le j \le n$.

In addition, it is advisable to mention that, so during the iterations the value of the component of the estimated parameter vector $\hat{\theta}(k)$ of the denominator of the equation (41) is cancelled (i.e., $\hat{\theta}[k:\ n+(r-1)n+1] = 0$), and then the bounded of the deduced parameter $b_{jr}(k)$ is not assured. To solve this problem, we can make a test on the value $\hat{\theta}[k:\ n+(r-1)n+1]$, and this, with each step of the discrete-time $k$, by proposing an appropriate solution allowing to exclude the situation where we can have: $\hat{\theta}[k:\ n+(r-1)n+1] = 0$. In such a situation, and as example, we can choose the following solution: if we will have $\hat{\theta}[k:\ n+(r-1)n+1] = 0$, then we take: $\hat{\theta}[k:\ n+(r-1)n+1] = 0.02$.

Let us add that the parameter $b_j(k)$ is $r$ time in the estimated parameter vector $\hat{\theta}(k)$, as given by (17). It follows that it is possible to determine this parameter $b_j(k)$ from $r = 1$, or $r = 2, \cdots,$ or $r = p$. In this case, we must retain a value of $r$ for which the good value of the parameter $b_j(k)$ is better. Moreover, we can retain the parameter $b_{j\text{m}}(k)$ corresponding to the good value of the parameter $b_j(k)$, which represents the statistical average value of the deduced parameters $b_{jr}(k)$, $r = 1, \cdots, p$. Thus, we can write the following expression:

$$b_{j\text{m}}(k) = \frac{1}{p} \sum_{r=1}^{p} \frac{\hat{\theta}[k:\ n+(r-1)n+j]}{\hat{\theta}[k:\ n+(r-1)n+1]} \tag{42}$$

Let us note that the parameter $\beta_r$, $1 \le r \le p$, which is related to the parameters $b_2(k), \cdots, b_n(k)$, is $j$ time intervening in the parameter vector $\theta(k)$, as defined by (10), that must be estimated. Thus, the determination of good value of the parameter $\beta_r$ can be made from $j = 2$, or $j = 3, \cdots,$ or $j = n$. In this case, we obtain a whole of values of the parameter $\beta_r$, from which we must retain the value that makes it possible to have the best quality of the estimate. The calculation of the good parameter $\beta_r$ is based on the knowledge of the statistical average value $b_{j\text{m}}(k)$, which corresponds to the good value of the parameter $b_j(k)$, such that:

$$\beta_r = \frac{\hat{\theta}[k:\ n+(r-1)n+j]}{b_{j\text{m}}(k)} \tag{43}$$

We must fix a performance criterion allowing selecting the best values of the parameters $\beta_r$, $r = 1,\cdots,p$, which can be related or not to the parameters $b_j(k)$, $j = 2,\cdots,n$.

It should be stressed that, in a general case of the discrete-time Hammerstein mathematical model (9), where the parameters $a_i(k)$ and $b_j(k)$ are given, such as: $i = 1,\cdots,n_A$ and $j = 2,\cdots,n_B$, with $n_B \le n_A$, then the value of the $r^{\text{th}}$ deduced parameter $b_{jr}(k)$ can be given by using the following equation:

$$b_{jr}(k) = \frac{\hat{\theta}[k:\ n_A + (r-1)n_B + j]}{\hat{\theta}[k:\ n_A + (r-1)n_B + 1]} \tag{44}$$

with $1 \le r \le p$ and $2 \le j \le n_B$.

## 4. Simulation results

This section is devoted to the parametric estimation of a nonlinear stochastic system being able to be described by the discrete-time Hammerstein mathematical model, where the dynamic linear and the static nonlinear blocks are defined by, respectively:

$$\begin{aligned} y(k) = &-a_1(k)y(k-1) - a_2(k)y(k-2) + h(k-1) + b_2(k)h(k-2) \\ &+ e(k) + c_1 e(k-1) \end{aligned} \tag{45}$$

and

$$h(k) = \beta_1 u(k) + \beta_2 u^2(k) \tag{46}$$

The output $y(k)$ of the considered nonlinear stochastic system can be described by:

$$\begin{aligned} y(k) = &-a_1(k)y(k-1) - a_2(k)y(k-2) + \beta_1 u(k-1) + b_2(k)\beta_1 u(k-2) \\ &+ \beta_2 u^2(k-1) + b_2(k)\beta_2 u^2(k-2) + e(k) + c_1 e(k-1) \end{aligned} \tag{47}$$

which can be rewritten in the following compact form?

$$y(k) = \theta^T(k)\psi(k) + e(k) \tag{48}$$

where the parameter vector $\theta(k)$ and the observation vector $\psi(k)$ are defined by:

$$\theta^T(k) = [a_1(k), a_2(k), \beta_1, b_2(k)\beta_1, \beta_2, b_2(k)\beta_2, c_1] \tag{49}$$

and

$$\psi^T(k) = [-y(k-1), -y(k-2), u(k-1), u(k-2), u^2(k-1), u^2(k-2), e(k-1)] \tag{50}$$

We can write the parameter vector $\theta(k)$, as given by (49), by:

$$\theta^T(k) = [a_1(k), a_2(k), \beta_1, f_{21}(k), \beta_2, f_{22}(k), c_1] \tag{51}$$

where the parameters $f_{21}(k)$ and $f_{22}(k)$ are given by: $f_{21}(k) = b_2(k)\beta_1$, $f_{22}(k) = b_2(k)\beta_2$.

We propose to estimate the parameter vector $\theta(k)$, as defined by (51) by using the RELS (19) and the RAML (36) parametric estimation algorithms. The relative data of the implementation (in numerical simulation) of these algorithms are given as follows:

1. the values of the parameters intervening in the mathematical model (47) are chosen, such as:
   $a_1(k) = -0.9200 + 0.0300\sin(3k)$, $a_2(k) = 0.4000 + 0.0200\cos(3k)$,
   $b_2(k) = 0.3600 + 0.0300\sin(2k)$, $\beta_1 = 0.3000$, $\beta_2 = 0.2400$, $c_1 = -0.3000$;
2. the input $u(k)$ applied to the considered nonlinear stochastic system is a pseudo-random binary sequence with level $[+1.2, -1.2]$ and length 1023;
3. the noise $e(k)$ acting on the considered system consists of a sequence of independent random variables with zero mean and variance $\sigma^2 = 0.0100$;
4. the initial conditions of the RELS and the RAML parametric estimation algorithms are chosen, such as: $\hat{\theta}(0) = 0$, $P(0) = 1000I$, with $I$ the identity matrix;
5. the variable forgetting factor $\lambda(k)$ being chosen, such as: $\lambda(k) = 0.9800, \forall k$;
6. the number of the measurements $M$ being chosen, such as: $M = 1, \cdots, 600$.

In the numerical simulations, the estimated parameter vector $\hat{\theta}(k)$ is defined by:

$$\hat{\theta}^T(k) = [\hat{a}_1(k), \hat{a}_2(k), \hat{\beta}_1(k), \hat{f}_{21}(k), \hat{\beta}_2(k), \hat{f}_{22}(k), \hat{c}_1(k)] \tag{52}$$

Let us notice that the determination of the parameter $b_2(k)$, which is related to the two parameters $\beta_1$ and $\beta_2$ is twice in this estimated parameter vector $\hat{\theta}(k)$. We will note by $b_{21}(k)$ the corresponding parameter of $b_2(k)$, which is deduced from the estimated parameters $\hat{\beta}_1(k)$ and $\hat{f}_{21}(k)$. In the same way, we will note by $b_{22}(k)$ the corresponding parameter of $b_2(k)$, which is deduced from the estimated parameters $\hat{\beta}_2(k)$ and $\hat{f}_{22}(k)$. We remark that these two deduced parameters $b_{21}(k)$ and $b_{22}(k)$ are related on the fourth and the sixth components of the estimated parameter vector $\hat{\theta}(k)$, as given by (52).

We define the estimation error $\delta_{b_2}(k)$ of the parameter $b_2(k)$ by:

$$\delta_{b_2}(k) = b_2(k) - b_{2\mathrm{m}}(k) \tag{53}$$

where $b_{2\mathrm{m}}(k)$ represents the average of the two deduced parameters $b_{21}(k)$ and $b_{22}(k)$, i.e., $b_{2\mathrm{m}}(k) = (b_{21}(k) + b_{22}(k))/2$.

To evaluate the computation quality of the parameter $b_2(k)$, we propose to calculate the variance of the estimation error $\delta_{b_2}(k)$. In this case, we can retain a performance criterion, which relates to the value of the variance of the error $\delta_{b_2}(k)$, in order to validate the calculated value of the parameter $b_2(k)$.

The validation of the global quality of the estimate of the parameters intervening in the discrete-time Hammestein mathematical model (47), by using the developed

RELS and the RAML algorithms, can be made by considering the following parametric distance $d(k)$:

$$d(k) = \left[ [\frac{a_1(k) - \hat{a}_1(k)}{a_1(k)}]^2 + [\frac{a_2(k) - \hat{a}_2(k)}{a_2(k)}]^2 + [\frac{\beta_1 - \hat{\beta}_1(k)}{\beta_1}]^2 + [\frac{\beta_2 - \hat{\beta}_2(k)}{\beta_2}]^2 + [\frac{f_{21}(k) - \hat{f}_{21}(k)}{f_{21}(k)}]^2 + [\frac{f_{22}(k) - \hat{f}_{22}(k)}{f_{22}(k)}]^2 + [\frac{c_1 - \hat{c}_1(k)}{c_1}]^2 \right]^{0.5} \quad (54)$$

Two numerical simulations are realised to estimate the parameters intervening in the discrete-time Hammestein mathematical model (47). In the first numerical simulation, the RELS parametric estimation (19) is used, where a party of the simulation results is given hereafter. Thus, the evolution curves of the estimated parameters $\hat{a}_1(k)$, $\hat{a}_2(k)$, $\hat{\beta}_1(k)$ and $\hat{\beta}_2(k)$ are given in Figure 4. We illustrate by the Figure 5 the evolution curves of the estimated parameters $\hat{f}_{21}(k)$, $\hat{f}_{22}(k)$ and $\hat{c}_1(k)$, and the parametric distance $d(k)$. The evolution curves of the deduced parameters $b_{21}(k)$ and $b_{22}(k)$, their average value $b_{2m}(k)$ and the estimation error $\delta_{b_2}(k)$ are shown in Figure 6.

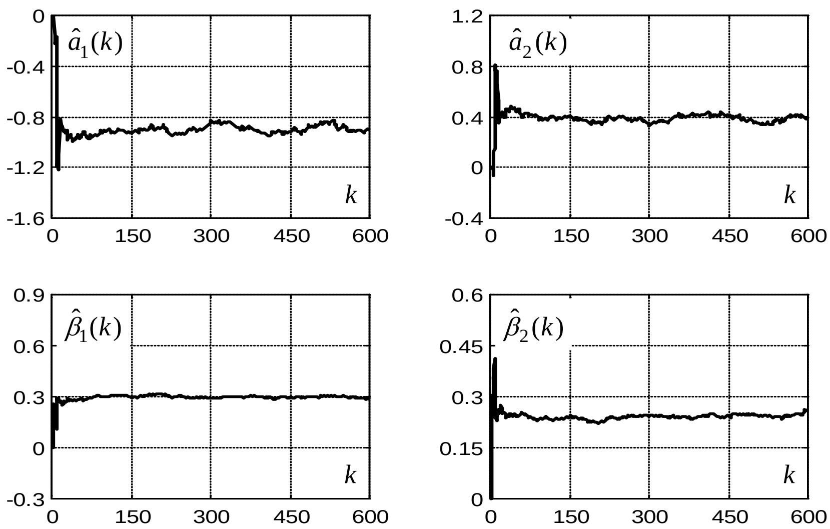


Figure 4. Curves of the estimated parameters $\hat{a}_1(k)$, $\hat{a}_2(k)$, $\hat{\beta}_1(k)$ and $\hat{\beta}_2(k)$.

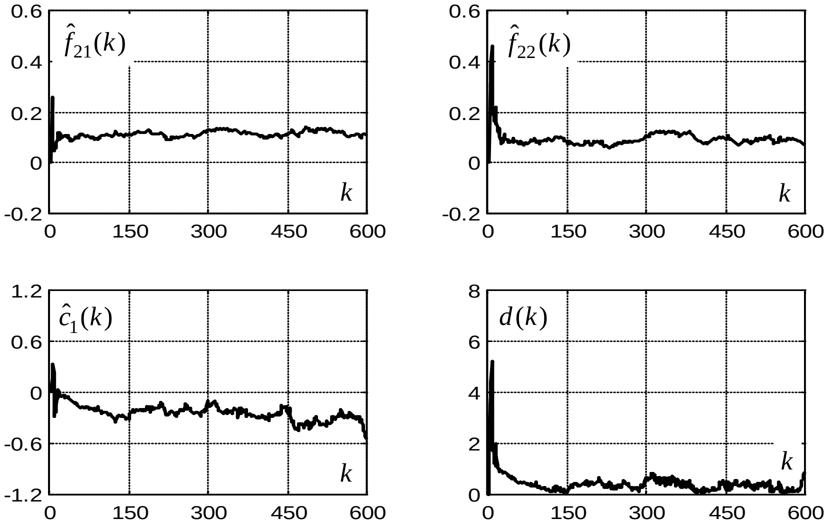


Figure 5. Curves of the estimated parameters $\hat{f}_{21}(k)$, $\hat{f}_{22}(k)$ and $\hat{c}_1(k)$, and the parametric distance $d(k)$.

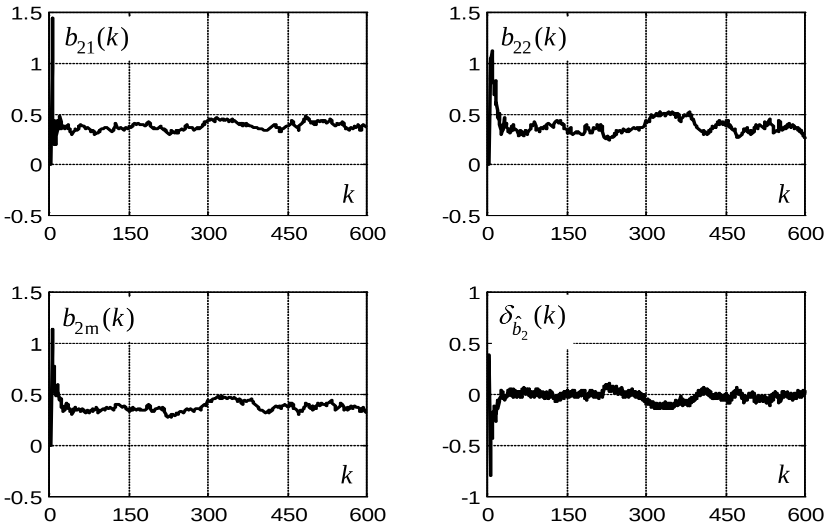


Figure 6. Curves of the deduced parameters $b_{21}(k)$ and $b_{22}(k)$, their average value $b_{2m}(k)$ and the estimation error $\delta_{\hat{b}_2}(k)$ of $b_2(k)$.

By examining the evolution curves of the different variables, which are represented by Figures. 4, 5 and 6, we can affirm that the quality of the estimate of the parameters intervening in the discrete-time Hammestein mathematical model (47) is good. In this direction, we notice that the parametric distance $d(k)$ decrease (on statistical average) towards a low value.

In the second numerical simulation, the RAML parametric estimation algorithm (36) is used. The obtained simulation results (no reported here) concerning the evolution curves of the same variables that are considered in the first simulation have shown the good quality of the estimate of the different parameters intervening in the discrete-time Hammestein mathematical model (47).

Nothing that the convergence speed of the estimated parameters $\hat{c}_i(k)$, $i = 1, \cdots, n$ is relatively slow that the other estimated parameters. This result is confirmed in several numerical simulations (non reported here) concerning the parametric estimation of nonlinear stochastic systems described by the discrete-time Hammerstein mathematical model, as given by (9), where the convergence speed of the estimated parameters $\hat{c}_i(k)$, $i = 1, \cdots, n$ is always slow that the other estimated parameters. Let us add that, if the sequence noise $e(k)$ does not have a Gaussian density probability, then the global quality of the estimate intervening of this discrete-time Hammerstein mathematical model is better for the RELS algorithm than for the RAML algorithm.

For the two considered numerical simulations, we presents in Table 1 the values of the statistical averages $\overline{m}_{\hat{a}_1}$, $\overline{m}_{\hat{a}_2}$, $\overline{m}_{\hat{\beta}_1}$, $\overline{m}_{\hat{\beta}_2}$, $\overline{m}_{\hat{b}_2\hat{\beta}_1}$, $\overline{m}_{\hat{b}_2\hat{\beta}_2}$, $\overline{m}_{\hat{c}_1}$, $\overline{m}_{\hat{b}_{21}}$ and $\overline{m}_{\hat{b}_{22}}$ of the estimated parameters $\hat{a}_1(k)$, $\hat{a}_2(k)$, $\hat{\beta}_1(k)$, $\hat{\beta}_2(k)$, $\hat{b}_2(k)\hat{\beta}_1(k)$, $\hat{b}_2(k)\hat{\beta}_2(k)$ and $\hat{c}_1(k)$. Also, we presents in Table 2 the values of the statistical averages $\overline{m}_{b_{21}}$, $\overline{m}_{b_{22}}$, $\overline{m}_d$ and $\overline{m}_\varepsilon$ of the deducted parameters $b_{21}(k)$ and $b_{22}(k)$, the parametric distance $d(k)$ and the *a priori* prediction error $\varepsilon(k)$, respectively, and the value of the variance $\sigma^2{}_\varepsilon$ of this prediction error. The computation of these different values is made for $k = 551, \cdots, 600$.

A comparison of the RELS and the RAML parametric estimation algorithms, which are used in the two numerical simulations, with respect to computation effort, convergence and computation time of each discrete-time $k$ is given in Table 3. They were programmed on a Pentium ®M process computer.

Table 1. Statistical averages $\overline{m}_{\hat{a}_1}$, $\overline{m}_{\hat{a}_2}$, $\overline{m}_{\hat{\beta}_1}$, $\overline{m}_{\hat{\beta}_2}$, $\overline{m}_{\hat{f}_{21}}$, $\overline{m}_{\hat{f}_{22}}$ and $\overline{m}_{\hat{c}_1}$ of the estimated parameters $\hat{a}_1(k)$, $\hat{a}_2(k)$, $\hat{\beta}_1(k)$, $\hat{\beta}_2(k)$, $\hat{f}_{21}(k)$, $\hat{f}_{22}(k)$ and $\hat{c}_1(k)$.

| Algorithm | $\overline{m}_{\hat{a}_1}$ | $\overline{m}_{\hat{a}_2}$ | $\overline{m}_{\hat{\beta}_1}$ | $\overline{m}_{\hat{\beta}_2}$ | $\overline{m}_{\hat{f}_{21}}$ | $\overline{m}_{\hat{f}_{22}}$ | $\overline{m}_{\hat{c}_1}$ |
|---|---|---|---|---|---|---|---|
| RELS | –0.9216 | 0.4050 | 0.3000 | 0.2528 | 0.1123 | 0.0907 | –0.3231 |
| RAML | –0.9204 | 0.4064 | 0.3002 | 0.2426 | 0.1096 | 0.0884 | –0.3152 |

Table 2. Statistical averages $\overline{\mathrm{m}}_{b_{21}}$, $\overline{\mathrm{m}}_{b_{22}}$, $\overline{\mathrm{m}}_{d}$ and $\overline{\mathrm{m}}_{\varepsilon}$ of the deducted parameters $b_{21}(k)$ and $b_{22}(k)$, the parametric distance $d(k)$ and the *a priori* prediction error $\varepsilon(k)$, respectively, and the value of the variance $\sigma^2{}_{\varepsilon}$ of this prediction error.

| Algorithm | $\overline{\mathrm{m}}_{b_{21}}$ | $\overline{\mathrm{m}}_{b_{22}}$ | $\overline{\mathrm{m}}_{d}$ | $\overline{\mathrm{m}}_{\varepsilon}$ | $\sigma^2{}_{\varepsilon}$ |
|---|---|---|---|---|---|
| RELS | 0.3743 | 0.3588 | 0.2566 | 0.0078 | 0.0154 |
| RAML | 0.3653 | 0.3644 | 0.2240 | 0.0053 | 0.0122 |

Table 3. Comparison of the RELS and the RAML parametric estimation algorithms.

| Algorithm | Computation effort | Convergence | Computation time at each discrete-time $k$ |
|---|---|---|---|
| RELS | small | medium | $1.8323\mathrm{e}-004$ sec. |
| RAML | larger than RELS | more reliable than RELS | $1.9868\mathrm{e}-004$ sec . |

It is important to note than globally, the convergence time of the recursive parametric estimation algorithms depend on the choice of the values on the initial conditions $\hat{\theta}(0)$ and $P(0)$. Indeed, if the initial condition of the estimated parameter vector $\hat{\theta}(0)$ is chosen near to the true parameter vector $\theta(k)$ and the value of the initial condition of the adaptation gain matrix is small (e.g., $P(0)=I$ ), then the estimated parameter vector $\hat{\theta}(k)$ converge rapidly to the parameter vector $\theta(k)$.

Let us recall that the *a priori* prediction error $\varepsilon(k)$ corresponds to the best estimate of the noise $e(k)$, which is not measurable. Thus, the computed value of the variance $\sigma^2{}_{\varepsilon}$ of this prediction error $\varepsilon(k)$, as given in Table 2, is very close to the variance of the considered noise $e(k)$. For example, the difference between these two variances being equal to 0.0054, by using the RELS parametric estimation algorithm (19).

In the same way, we notice that the differences between the values of the true parameters of the considered nonlinear stochastic system and the values of the statistical averages of their estimated are weak. Let us add that the computed value of the variance of the estimation error $\delta_{b_2}(k)$ of the parameter $b_2(k)$, by using the RELS parametric estimation algorithm (19), is equal to 0.0056. The value of this variance being very weak; this shows the good quality of estimate of the parameter $b_2(k)$.

Consequently, the obtained simulation results show well the performances which can ensure the developed RELS and RAML parametric estimation algorithms.

## 5. Conclusion

In this paper, the parametric estimation problem of nonlinear stochastic systems described by the discrete-time Hammerstein mathematical model is studied. We have considered the case of the discrete-time Hammerstein mathematical model, where the static nonlinear block is given by a nonlinear function and the dynamic linear block is described by the ARMAX mathematical model, single-input single-output, with unknown slowly time-varying parameters.

Two recursive parametric estimation methods are studied and compared in order to estimate the parameters intervening in the considered discrete-time Hammerstein mathematical model. It is about the recursive extended least squares method and the recursive approximated maximum likelihood method. The parametric estimation algorithms, which correspond, to these methods are developed on the basis of the prediction error method. The convergence conditions and the techniques of the practical implementation of these algorithms are given.

Numerical simulations have been treated in order to illustrate the performances and the effectiveness of the proposed parametric estimation algorithms.

## References

Al-duwaish, H. and M. Nazmul Karim, A new method for the identification of Hammerstein model, *Automatica*, vol. 33, pp. 1871–1875, 1997.

Åström, K. J., Maximum likelihood and prediction error methods, *Automatica*, vol. 16, pp. 551–574, 1980.

Belforte, G. and P. Gay, Discrete time Hammerstein model identification with unknown but bounded errors, *IEE Proceedings – Control Theory and Applications*, vol. 148, pp. 523–529, 2001.

Billings, S. A., Identification of non-linear systems: a survey, *IEE Proceedings Pt. D*, vol. 127, pp. 272–285, 1980.

Billings, S. A., Nonlinear System Identification NARMAX Methods in the Time, Frequency, and Spatio –Temporal Domains, Wiley, London, 2013.

Billings, S. A. and W. S. F. Voon, A prediction-error and stepwise-regression estimation algorithm, *International Journal of Control*, vol. 44, pp. 803–822, 1986.

Chen, S. and S. A. Billings, Representation of non-linear systems: the NARMAX model, *International Journal of Control*, vol. 49, pp. 1013–1032, 1989.

Eskinat, E. S.; H. Johnson and W. L. Luyben, Use of Hammerstein models in identification of non-linear systems, *AIChE Journal*, vol. 37, pp. 255–268, 1991.

Guo, L. and L. Ljung, Performance analysis of general tracking algorithms, *IEEE Transactions on Automatic Control*, vol. AC-40, pp. 1388–1402, 1995.

Haber, R. and L. Keviczky, Non-linear System Identification: Input-output Modelling Approach, Kluwer, Boston, 1999.

Haber, R. and U. Unbehauen, Structure identification of non-linear dynamic system – a survey on input/output approaches, *Automatica*, vol. 26, pp. 651–677, 1990.

Hajji, S.; M. M'Saad; M. Farza; A. Chaari and M. Kamoun, State feedback control design for a class of nonlinear systems, *International Journal of Sciences and Techniques of Automatic control & computer engineering IJ-STA*, volume 2, N° 2, pp. 748−767, 2008.

Kamoun, S., Contribution to the Identification and Adaptive Control of Complex Systems (In French), Ph.D. Degree in Automatic Control and Computer Engineering, National Engineering School of Sfax, University of Sfax, 2003.

Kortmann, K. and H. Unbehauen, Structure detection in the identification of non-linear systems, *RAIRO-APII*, vol. 22, pp. 5−25, 1988.

Li, H.-X., Identification of Hammerstein models using genetic algorithms, *IEE Proceedings – Control Theory and Applications*, vol. 146, pp. 599−504, 1999.

Ljung, L., On positive real transfer function and the convergence of some recursive schemes, *IEEE Transactions on Automatic Control*, vol. AC–22, pp. 539−550, 1977.

Ljung, L. and S. Gunnarsson, Adaptation and tracking in system identification – a survey, *Automatica*, vol. 26, pp. 7−22, 1990.

Mâatoug, T.; M. Farza; M. M'Saad; Y. Koubaa and M. Kamoun, Adaptive observer design for a class of nonlinear systems with coupled structures, *International Journal of Sciences and Techniques of Automatic control & computer engineering IJ-STA*, volume 2, N° 1, pp. 484−499, 2008.

Rangan, S.; G. Wolodkin and K. Poolla, New results for Hammerstein system identification, *Proceedings of the* 34*th* *IEEE Conference on Decision and Control*, New Orleans, pp. 697−702, 1995.

Tlili, S. and H. Mibar, A method for approximation and control of nonlinear systems, *International Journal on Sciences and Techniques of Automatic control & computer engineering* (*IJ-STA*), vol. 1, N° 1, pp. 1–11, 2007.

Wigren, T., Recursive prediction error identification using the nonlinear Wiener model, *Automatica*, vol. 29, pp. 1011–1025, 1993.

Vörös, J., Modeling and parametric identification of systems with multisegment piecewise-linear characteristics, *IEEE Transactions on Automatic Control*, vol. AC-47, pp. 184−188, 2002.