5

Polynomial Hammerstein models

In this chapter, we will review seven methods of the identification of Hammerstein systems which use Hammerstein models with a polynomial model of the nonlinear element. The reviewed methods use models of the linear dynamic system in the form of pulse transfer models [21, 27, 34, 63, 120, 161], or a Laguerre expansion of the impulse response [156]. Wiener system parameters are estimated with different procedures such as the ordinary least squares [27], iterative least squares [63, 120, 161], iterative correlation and steepest descent [156], prediction error [34], and pseudolinear regression [21] methods.

5.1 Noniterative least squares identification of Hammerstein systems

One of the best known identification methods for Hammerstein systems is based on the transformation of a nonlinear-in-parameters SISO identification problem into a linear-in-parameters MISO one and the application of least squares parameter algorithm. This method was originally proposed by Chang and Luus [27]. A short description of the method is given below.
First, we introduce a parallel polynomial Hammerstein model, i.e., a model with the linear part defined by the transfer function $\hat{B}(q^{-1})/\hat{A}(q^{-1})$:

$$\hat{y}(n) = \frac{\hat{B}(q^{-1})}{\hat{A}(q^{-1})}\hat{v}(n), \tag{5.1}$$

$$\hat{B}(q^{-1}) = \hat{b}_1 q^{-1} + \cdots + \hat{b}_{nb} q^{-nb}, \tag{5.2}$$

$$\hat{A}(q^{-1}) = 1 + \hat{a}_1 q^{-1} + \cdots + \hat{a}_{na} q^{-na}, \tag{5.3}$$

and the nonlinear part defined by a polynomial model:

$$\hat{v}(n) = \hat{f}\big(u(n)\big) = \hat{\mu}_1 u(n) + \hat{\mu}_2 u^2(n) + \ldots + \hat{\mu}_r u^r(n). \tag{5.4}$$

A. Janczak: Identification of Nonlinear Systems, LNCIS 310, pp. 143–157, 2005.

Without loss of generality, $\hat{\mu}_1$ can be assumed to be a unity, and $\hat{\mu}_2, \ldots, \hat{\mu}_r$ can be normalized accordingly. Then (5.1) can be expressed as

$$\hat{y}(n) = \frac{\hat{B}(q^{-1})}{\hat{A}(q^{-1})} u(n) + \sum_{k=2}^{r} \frac{\hat{B}(q^{-1})}{\hat{A}(q^{-1})} \hat{\mu}_k u^k(n). \tag{5.5}$$

Introducing $\hat{B}_1(q^{-1}) = \hat{B}(q^{-1}), \ldots, \hat{B}_k(q^{-1}) = \hat{\mu}_k \hat{B}(q^{-1})$ and $u_1(n) = u(n), \ldots, u_k(n) = u^k(n)$, (5.5) can be rewritten as

$$\hat{y}(n) = \sum_{k=1}^{r} \frac{\hat{B}_k(q^{-1})}{\hat{A}(q^{-1})} u_k(n). \tag{5.6}$$

The MISO Wiener model (Fig. 5.1) consists of r linear dynamic sub-models with a common denominator in each branch. Equivalently, (5.6) can be written as

$$\hat{y}(n) = \left[1 - \hat{A}(q^{-1})\right] \hat{y}(n) + \sum_{k=1}^{r} \hat{B}_k(q^{-1}) u_k(n). \tag{5.7}$$

By using the definitions of $\hat{A}(q^{-1})$ and $\hat{B}_k(q^{-1})$, we have

$$\hat{y}(n) = -\sum_{m=1}^{na} \hat{a}_m \hat{y}(n-m) + \sum_{k=1}^{r} \sum_{m=1}^{nb} \hat{b}_{km} u_k(n-m), \tag{5.8}$$

where $\hat{b}_{km}$ is the mth parameter of the polynomial $B_k(q^{-1})$. The parallel model (5.8) can be transformed into a linear-in-parameters series-parallel model using the available delayed system outputs $y(n-m)$ instead of the model outputs $\hat{y}(n-m)$:

$$\hat{y}(n) = -\sum_{m=1}^{na} \hat{a}_m y(n-m) + \sum_{k=1}^{r} \sum_{m=1}^{nb} \hat{b}_{km} u_k(n-m). \tag{5.9}$$

Introducing the parameter vector $\hat{\boldsymbol{\theta}}$ and the regression vector $\mathbf{x}(n)$:

$$\hat{\boldsymbol{\theta}} = \left[\hat{a}_1 \ldots \hat{a}_{na}\ \hat{b}_{11} \ldots \hat{b}_{1nb} \ldots \hat{b}_{r1} \ldots \hat{b}_{rnb}\right]^T, \tag{5.10}$$

$$\begin{aligned} \mathbf{x}(n) = \big[&- y(n-1) \ldots - y(n-na)\ u_1(n-1) \ldots u_1(n-nb) \ldots \\ &u_r(n-1) \ldots u_r(n-nb)\big]^T, \end{aligned} \tag{5.11}$$

we have

$$\hat{y}(n) = \mathbf{x}^T(n) \hat{\boldsymbol{\theta}}. \tag{5.12}$$

The minimization of the sum of N squared errors between the system output and the model output w.r.t. the regression vector $\hat{\boldsymbol{\theta}}$ gives

$$\hat{\boldsymbol{\theta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}, \tag{5.13}$$

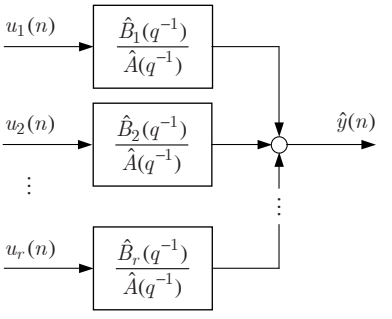

Fig. 5.1. Hammerstein model transformed into the MISO form

where

$$\mathbf{X} = \big[\mathbf{x}(1) \dots \mathbf{x}(N)\big]^T, \tag{5.14}$$

$$\mathbf{y} = \big[y(1) \dots y(N)\big]^T. \tag{5.15}$$

In this way, the parameters of the transformed MISO model (5.9) can be calculated, but our primary goal is to determine the parameters of the model defined by (5.1) and (5.4). The parameters $\hat{a}_m$ are elements of the vector $\hat{\boldsymbol{\theta}}$. Also, the parameters $\hat{b}_m$ are available directly as $\hat{b}_m = \hat{b}_{1m}$. The parameters $\hat{\mu}_k$ can be calculated from the remaining elements of $\hat{\boldsymbol{\theta}}$ but there is a certain amount of redundancy as

$$\hat{\mu}_k = \frac{\hat{b}_{mk}}{\hat{b}_{1k}},\ k = 2, \dots, r,\ m = 1, \dots, nb, \tag{5.16}$$

and nb different sets of parameters $\hat{\mu}$ can be calculated. Chang and Luus [27] suggested computing the root mean-square error (RMS) of the model output for all nb sets of $\hat{\mu}$ and accept the set that yields the least value for the RMS as a more reliable approach in comparison with computing the mean of the nb values.

5.2 Iterative least squares identification of Hammerstein systems

In the iterative least squares method of Narendra and Gallman [120], the parameters of the linear dynamic system and the nonlinear static element are updated separately and sequentially. This method utilizes an alternate adjustment of the parameters of the linear dynamic system and the nonlinear static element to minimize the sum of squared errors. The error $e(n)$ at the time n is

$$e(n) = y(n) - \left(- \sum_{m=1}^{na} \hat{a}_m y(n-m) + \sum_{m=1}^{nb} \hat{b}_m \hat{v}(n-m) \right), \tag{5.17}$$

where

$$\hat{v}(n) = \sum_{k=1}^{r} \hat{\mu}_k u^k(n). \tag{5.18}$$

Introducing $\hat{\boldsymbol{\theta}} = \left[\hat{\boldsymbol{\theta}}_a^T \ \hat{\boldsymbol{\theta}}_b^T\right]^T$ and $\mathbf{x}(n) = \left[\mathbf{x}_a^T(n) \ \mathbf{x}_b^T(n)\right]^T$, where

$$\hat{\boldsymbol{\theta}}_a = \left[\hat{a}_1 \dots \hat{a}_{na}\right]^T, \tag{5.19}$$

$$\hat{\boldsymbol{\theta}}_b = \left[\hat{b}_1 \dots \hat{b}_{nb}\right]^T, \tag{5.20}$$

$$\mathbf{x}_a(n) = \left[- y(n-1) \dots - y(n-na)\right]^T, \tag{5.21}$$

$$\mathbf{x}_b(n) = \left[\hat{v}(n-1) \dots \hat{v}(n-nb)\right]^T, \tag{5.22}$$

(5.17) can be written as

$$e(n) = y(n) - \mathbf{x}^T(n)\hat{\boldsymbol{\theta}} = y(n) - \mathbf{x}_a^T(n)\hat{\boldsymbol{\theta}}_a - \mathbf{x}_b^T(n)\hat{\boldsymbol{\theta}}_b. \tag{5.23}$$

Now, introducing the vectors $\hat{\boldsymbol{\mu}} = \left[\hat{\mu}_1 \dots \hat{\mu}_r\right]^T$ and $\mathbf{u}(n-1) = \left[u(n-1) \ u^2(n-1) \dots u^r(n-1)\right]^T$, the matrix $\mathbf{U}(n)$ can be defined:

$$\mathbf{U}(n) = \begin{bmatrix} \mathbf{u}^T(n-1) \\ \vdots \\ \mathbf{u}^T(n-nb) \end{bmatrix}. \tag{5.24}$$

From (5.18), it follows that the vector $\mathbf{x}_b(n)$ can be written as

$$\mathbf{x}_b(n) = \mathbf{U}(n)\hat{\boldsymbol{\mu}}, \tag{5.25}$$

and (5.23) becomes

$$e(n) = y(n) - \mathbf{x}_a^T(n)\hat{\boldsymbol{\theta}}_a - \left(\mathbf{U}(n)\hat{\boldsymbol{\mu}}\right)^T\hat{\boldsymbol{\theta}}_b. \tag{5.26}$$

In this iterative approach, parameter estimation is performed according to the following iterative procedure:

1. Based on N measurements of the system input and output signals and the assumed initial vector $\hat{\boldsymbol{\mu}}^{(1)}$, the parameter vector $\hat{\boldsymbol{\theta}}^{(1)} = [\hat{\boldsymbol{\theta}}_a^{(1)T} \ \hat{\boldsymbol{\theta}}_b^{(1)T}]^T$ of the model (5.23) is estimated:

$$\hat{\boldsymbol{\theta}}^{(1)} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}, \tag{5.27}$$

where

$$\mathbf{y} = \left[y(1) \dots y(N)\right]^T, \tag{5.28}$$

$$\mathbf{X} = \left[\mathbf{x}(1) \dots \mathbf{x}(N)\right]^T. \tag{5.29}$$

2. With $[\hat{\boldsymbol{\theta}}_a^{(1)T}\ \hat{\boldsymbol{\theta}}_b^{(1)T}]^T$, $\hat{\boldsymbol{\mu}}^{(2)}$ is calculated by minimizing the sum of squared errors for the model (5.26)

$$\hat{\boldsymbol{\mu}}^{(2)} = \big(\mathbf{U}^T\mathbf{U}\big)^{-1}\mathbf{U}^T(\mathbf{y} - \mathbf{X}_a\hat{\boldsymbol{\theta}}_a^{(1)}), \tag{5.30}$$

where

$$\mathbf{X}_a = \big[\mathbf{x}_a(1)\ldots\mathbf{x}_a(N)\big]^T, \tag{5.31}$$

$$\mathbf{U} = \big[\mathbf{U}^T(1)\hat{\boldsymbol{\theta}}_b^{(1)}\ldots\mathbf{U}^T(N)\hat{\boldsymbol{\theta}}_b^{(1)}\big]^T. \tag{5.32}$$

3. Using $\hat{\boldsymbol{\mu}}^{(2)}$, $\hat{\boldsymbol{\theta}}^{(2)} = [\hat{\boldsymbol{\theta}}_a^{(2)T}\ \hat{\boldsymbol{\theta}}_b^{(2)T}]^T$ is calculated according to the scheme used in Step 1, and the process is continued.

An obvious advantage of the iterative method is nonredundant model parameterization. The simulation study performed by Gallman [41] shows that such an iterative procedure gives significantly lower parameter variance and a slightly lower RMS error than the noniterative method of Chang and Luus. The iterative method applied to a variety of problems such as polynomial nonlinearity, a half-wave linear detector or saturating nonlinearity has proven to be very successful [120]. In spite of the reported successful experience, the method may not converge in some cases.

5.3 Identification of Hammerstein systems in the presence of correlated noise

A direct application of the method of Chang and Luus [27] for Hammerstein systems with correlated additive output disturbance results in asymptotically biased parameter estimates. To overcome this problem, an iterative least squares procedure, which estimates the parameters of the linear dynamic system, the parameters of the nonlinear static element and the parameters of a noise model, was proposed by Haist *et al.* [63]. Their approach is based on the assumption that the Hammerstein system with an output disturbance has the form

$$y(n) = \frac{B(q^{-1})}{A(q^{-1})}f\big(u(n)\big) + \frac{\varepsilon(n)}{D(q^{-1})}, \tag{5.33}$$

where

$$D(q^{-1}) = 1 + d_1q^{-1} + \cdots + d_{nd}q^{-nd}, \tag{5.34}$$

and $\varepsilon(n)$ is a zero-mean discrete-time white noise. Assume that the nonlinear element is described by the polynomial

$$f\big(u(n)\big) = \mu_1u(n) + \mu_2u^2(n) + \ldots + \mu_ru^r(n). \tag{5.35}$$

Denoting the undisturbed system output by $\bar{y}(n)$ and the additive disturbance at the system output by $\eta(n)$, we have

$$y(n) = \bar{y}(n) + \eta(n) \tag{5.36}$$

with

$$\bar{y}(n) = \sum_{k=1}^{r} \frac{B_k(q^{-1})}{A(q^{-1})} u_k(n), \tag{5.37}$$

$$\eta(n) = \frac{\varepsilon(n)}{D(q^{-1})}, \tag{5.38}$$

where $B_k(q^{-1}) = \mu_k B(q^{-1})$ and $u_k(n) = u^k(n)$ or, equivalently,

$$\begin{aligned} y(n) = &- \sum_{m=1}^{na} a_m \bar{y}(n-m) + \sum_{k=1}^{r} \sum_{m=1}^{nb} b_{km} u_k(n-m) \\ &- \sum_{m=1}^{nd} d_m \eta(n-m) + \varepsilon(n). \end{aligned} \tag{5.39}$$

Introducing the parameter vector $\boldsymbol{\theta}$ and the regression vector $\mathbf{x}(n)$:

$$\boldsymbol{\theta} = \left[a_1 \dots a_{na}\ b_{11} \dots b_{1nb} \dots b_{r1} \dots b_{rnb}\ d_1 \dots d_{nd}\right]^T, \tag{5.40}$$

$$\begin{aligned} \mathbf{x}(n) = \big[&- \bar{y}(n-1) \dots - \bar{y}(n-na)\ u_1(n-1) \dots u_1(n-nb) \dots \\ &u_r(n-1) \dots u_r(n-nb)\ - \eta(n-1) \dots - \eta(n-nd)\big]^T, \end{aligned} \tag{5.41}$$

we have

$$y(n) = \mathbf{x}^T(n)\boldsymbol{\theta} + \varepsilon(n). \tag{5.42}$$

Now, parameter updating can be made according to the following iterative procedure:

1. Since the sequences $\{\bar{y}(n)\}$ and $\{\eta(n)\}$ are not known initially, calculate the parameters of a deterministic part of the model using the noniterative procedure of Chang and Luus, described in Section 5.1. Generate an estimate of the sequence $\{\bar{y}(n)\}$ from (5.37) and calculate an estimate of the sequence $\{\eta(n)\}$ from (5.36) using the estimate of the sequence $\{\bar{y}(n)\}$.
2. Calculate improved parameter estimates from

$$\hat{\boldsymbol{\theta}} = \left[\sum_{i=1}^{N} \mathbf{x}(n)\mathbf{x}^T(n)\right]^{-1} \left[\sum_{i=1}^{N} \mathbf{x}(n) y(n)\right], \tag{5.43}$$

   where the estimated values of $\bar{y}(n)$ and $\eta(n)$ are used in $\mathbf{x}(n)$ instead of their unknown true values.
3. The improved parameter estimates yield another estimate of the sequence $\{\bar{y}(n)\}$. Also, the estimated sequence $\{\eta(n)\}$ is calculated again, from (5.36).
4. Continue the procedure from Step 2 until the change in the normalized RMS is less than some specified minimum.

To ensure numerical stability of the procedure, employing a stepping factor ϵ is suggested. With the stepping factor, the adopted value of the model parameter vector $\hat{\boldsymbol{\theta}}^{*(j+1)}$ at the iteration $j+1$ is calculated according to the formula

$$\hat{\boldsymbol{\theta}}^{*(j+1)} = \epsilon\, \hat{\boldsymbol{\theta}}^{(j+1)} + (1-\epsilon)\hat{\boldsymbol{\theta}}^{*(j)}, \tag{5.44}$$

where $\hat{\boldsymbol{\theta}}^{(j+1)}$ is the estimate of the parameter vector obtained from (5.43). As in the case of the method of Chang and Luus, it follows from (5.41) that there is some redundancy in parameter determination. In spite of its practically confirmed successful applications, no proof of the convergence of this method is available.

5.4 Identification of Hammerstein systems with the Laguerre function expansion

An example of identification methods which use orthonormal basis functions is the method proposed by Thathachar and Ramaswamy [156]. In this method, the nonlinear part of the Hammerstein model is represented by the polynomial (5.4), and the linear part – by a Laguerre expansion of its impulse response $h(n)$:

$$h(n) = \sum_{i=1}^{nl} A_i l_i(n), \tag{5.45}$$

where $l_i(n)$ is the ith Laguerre function, and A_i, $i = 1, \ldots, nl$, are the parameters. Discrete Laguerre functions are defined as

$$L_i(q^{-1}) = \sqrt{1-\exp(-2T)}\,\frac{\left(q^{-1}-\exp(-T)\right)^{i-1}}{\left(1-\exp(-T)q^{-1}\right)^{i}}, \text{ for } i = 1, 2, \ldots, \tag{5.46}$$

where T is the sampling period, and $L_i(q)$ is the positive half $\mathcal{Z}$ transform of $l_i(n)$:

$$L_i(q^{-1}) = \sum_{n=0}^{\infty} l_i(n) q^{-n}. \tag{5.47}$$

The functions $l_i(n)$ fulfill the condition of orthonormality:

$$\sum_{n=0}^{\infty} l_i(n) l_k(n) = \begin{cases} 0 \text{ for } i \neq k \\ 1 \text{ for } i = k \end{cases} \quad i, k \leqslant n, \tag{5.48}$$

and

$$l_i(n) = 0 \text{ for } n < 0 \text{ and all } i. \tag{5.49}$$

The identification problem is to determine the parameters $\hat{\mu}_k$, $k = 1, \ldots, r$, of the nonlinear part, and A_i, $i = 1, \ldots, nl$, of the linear part. The problem

is solved in an iterative manner, where each iteration comprises two steps. In the first step, the parameters $\hat{\mu}_k$ are kept constant at their values from the previous iteration and A_i are calculated by solving a set of linear equations. To start the identification procedure, some arbitrary initial values of $\hat{\mu}_k$ are assumed permitting the calculation of the output of the nonlinear part of the model.
The expressions for the parameters A_i can be derived from the input-output cross-correlation function of the linear part:

$$\left.\begin{aligned} A_1 &= \frac{C_{nl}}{B_{nl}} \\ A_s &= \frac{C_{nl-s+1} - \sum_{i=1}^{s-1} A_i B_{nl-s+i}}{B_{nl}} \\ A_{nl} &= \frac{C_1 - \sum_{i=1}^{nl-1} A_i B_i}{B_{nl}} \end{aligned}\right\}. \tag{5.50}$$

The coefficients $B_j,\ j = 1, \ldots, nl$, are calculated as the following averages:

$$B_j = \lim_{N\to\infty} \frac{1}{N+1} \sum_{n=0}^{N} u(n) f_j(n), \tag{5.51}$$

where $f_j(n)$ is the output of $L_j(q^{-1})$ excited by $\hat{f}\big(u(n)\big)$. In a similar way, the coefficients $C_j,\ j = 1, \ldots, nl$ are calculated as

$$C_j = \lim_{N\to\infty} \frac{1}{N+1} \sum_{n=0}^{N} u(n) y_{j-1}(n), \tag{5.52}$$

where $y_{j-1}(n)$ is the output of $L_{j-1}(q^{-1})$ excited by $y(n)$.
In the other step, the parameters A_i are kept constant while $\hat{\mu}_k$ are determined using an algorithm that adjusts $\hat{\mu}_k$ towards their optimal values minimizing the mean square error (MSE) between the model output and the system output. The minimization is carried out with gradient methods, i.e., steepest descent in the disturbance-free case, and stochastic approximation in the presence of an additive output disturbance. The MSE is defined as

$$J = \frac{1}{N} \sum_{n=1}^{N} \Big[\hat{\boldsymbol{\mu}}^T \sum_{m_1=0}^{\infty} h(m_1) \mathbf{u}(n-m_1) - y(n) \Big]^2, \tag{5.53}$$

where $\mathbf{u}(n) = \big[u(n)\ u^2(n) \ldots u^r(n)\big]^T$ and $\hat{\boldsymbol{\mu}}_i = \big[\hat{\mu}_1\ \hat{\mu}_2 \ldots \hat{\mu}_r\big]^T$. In the disturbance-free case, the parameters of the nonlinear element model are adjusted according to following rule:

$$\hat{\boldsymbol{\mu}}^{(new)} = \hat{\boldsymbol{\mu}}^{(old)} - \eta \frac{\partial J}{\hat{\boldsymbol{\mu}}}, \tag{5.54}$$

where η is the step size. In the presence of additive output disturbances, the iteration-dependent step size $\eta(n)$ is used instead of a constant one, which fulfills the following conditions:

$$\eta(n) \geqslant 0, \quad \sum_{k=1}^{\infty} \eta^2(n) < \infty, \quad \sum_{k=1}^{\infty} \eta(n) = \infty. \tag{5.55}$$

5.5 Prediction error method

In the prediction error method, the Hammerstein system is assumed to be given by

$$y(n) = \frac{B(q^{-1})}{A(q^{-1})} \sum_{k=1}^{r} \mu_k u^k(n) + \varepsilon(n), \tag{5.56}$$

where $\varepsilon(n)$ is a zero-mean white noise used to model disturbances. For the nonwhite case, the parameters of a noise filter should also be identified but for the sake of complexity this problem is not considered here. The one-step-ahead predictor for the system output has the form [34]:

$$\hat{y}(n|n-1) = \frac{\hat{B}(q^{-1})}{\hat{A}(q^{-1})} \sum_{k=1}^{r} \hat{\mu}_k u^k(n). \tag{5.57}$$

The parameter vector $\hat{\boldsymbol{\theta}} = \left[\hat{a}_1 \ldots \hat{a}_{na}\ \hat{b}_1 \ldots \hat{b}_{nb}\ \hat{\mu}_1 \ldots \hat{\mu}_r\right]^T$ should be estimated based on a set of N input-output measurements to minimize the sum of squared prediction errors

$$J = \frac{1}{N} \sum_{i=1}^{N} e^2(n, \hat{\boldsymbol{\theta}}), \tag{5.58}$$

where

$$e(n, \hat{\boldsymbol{\theta}}) = y(n) - \hat{y}(n|n-1). \tag{5.59}$$

The prediction error method requires the calculation of the gradient of (5.58) w.r.t. its parameters. For the Hammerstein model, partial derivatives of (5.58) are nonlinear in the parameters. Therefore, since a direct solution to the optimization problem cannot be found, iterative methods have to be used. The prediction error method is a second order gradient-based technique, in which the parameter vector $\hat{\boldsymbol{\theta}}$ is adjusted along the negative gradient of (5.58):

$$\hat{\boldsymbol{\theta}}^{(j)} = \hat{\boldsymbol{\theta}}^{(j-1)} - \eta_j H^{-1}(\hat{\boldsymbol{\theta}}^{(j-1)}) G(\hat{\boldsymbol{\theta}}^{(j-1)}), \tag{5.60}$$

where η_j is the step size, $H(\cdot)$ is the Hessjan of (5.58) or its approximation, and $G(\cdot)$ is the gradient of (5.58). The gradient of (5.58) can be computed as

$$G(\hat{\boldsymbol{\theta}}^{(j)}) = \frac{dJ}{d\hat{\boldsymbol{\theta}}} = \frac{2}{N}\sum_{i=1}^{N} e(n,\hat{\boldsymbol{\theta}})\frac{\partial e(n,\hat{\boldsymbol{\theta}})}{\partial \hat{\boldsymbol{\theta}}} = -\frac{2}{N}\sum_{i=1}^{N} e(n,\hat{\boldsymbol{\theta}})\boldsymbol{\psi}(n), \tag{5.61}$$

where

$$\begin{aligned}\boldsymbol{\psi}(n) = \Bigg[&\frac{\partial \hat{y}(n|n-1)}{\partial \hat{a}_1}\cdots\frac{\partial \hat{y}(n|n-1)}{\partial \hat{a}_{na}}\ \frac{\partial \hat{y}(n|n-1)}{\partial \hat{b}_1}\cdots\frac{\partial \hat{y}(n|n-1)}{\partial \hat{b}_{nb}}\\ &\frac{\partial \hat{y}(n|n-1)}{\partial \hat{\mu}_1}\cdots\frac{\partial \hat{y}(n|n-1)}{\partial \hat{\mu}_r}\Bigg]\end{aligned} \tag{5.62}$$

with

$$\frac{\partial \hat{y}(n|n-1)}{\partial \hat{a}_k} = -\frac{1}{\hat{A}(q^{-1})}\sum_{m=1}^{r}\frac{\hat{B}(q^{-1})}{\hat{A}(q^{-1})}\hat{\mu}_m u^m(n-k), \tag{5.63}$$

$$\frac{\partial \hat{y}(n|n-1)}{\partial \hat{b}_k} = \frac{1}{\hat{A}(q^{-1})}\sum_{m=1}^{r}\hat{\mu}_m u^m(n-k), \tag{5.64}$$

$$\frac{\partial \hat{y}(n|n-1)}{\partial \hat{\mu}_m} = \frac{\hat{B}(q^{-1})}{\hat{A}(q^{-1})}u^m(n). \tag{5.65}$$

The main drawback of using the Hessjan in (5.60) is that it requires second order derivatives. To avoid the calculation of second order derivatives, an approximate Hessjan can be used. In the Levenberg-Marquardt method, the approximate Hessjan is calculated as

$$H(\hat{\boldsymbol{\theta}}^{(j)}) = \frac{1}{N}\sum_{i=1}^{N}\frac{\partial e(n,\hat{\boldsymbol{\theta}})}{\partial \hat{\boldsymbol{\theta}}}\frac{\partial e^T(n,\hat{\boldsymbol{\theta}})}{\partial \hat{\boldsymbol{\theta}}} + \mu I = \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\psi}(n)\boldsymbol{\psi}^T(n) + \mu I, \tag{5.66}$$

where μ is a nonnegative small scalar and I is the identity matrix with an appropriate dimension. The iterative prediction algorithm comprises the following steps:

1. Start iterations with an initial estimate of the parameters $\hat{\boldsymbol{\theta}}^{(0)}$ and set $\hat{\mu}_1 = 1$.
2. Pick a small value for μ (a typical choice is 0.0001).
3. Compute $\hat{y}(n|n-1)$ and $J(\hat{\boldsymbol{\theta}}^{(j-1)})$.
4. Compute the gradient and the Hessjan through (5.61) – (5.66).
5. Update parameter estimates through (5.60), and calculate $J(\hat{\boldsymbol{\theta}}^{(j)})$.
6. If $J(\hat{\boldsymbol{\theta}}^{(j)}) > J(\hat{\boldsymbol{\theta}}^{(j-1)})$, decrease μ by a factor (say 10) and go to Step 4. $J(\hat{\boldsymbol{\theta}}^{(j)}) < J(\hat{\boldsymbol{\theta}}^{(j-1)})$, update solution and increase μ by a factor (say 10) and go to Step 2.

Having the rules for the calculation of the gradient derived, a recursive version of the algorithm can be implemented easily, see Equations (2.97) – (2.99).

5.6 Identification of MISO systems with the pseudolinear regression method

Consider a MISO Hammerstein model with nu inputs and the output disturbed additively by correlated measurement noise [22]:

$$y(n) = \sum_{j=1}^{nu} \frac{B_j(q^{-1})}{A_j(q^{-1})} f_j\big(u_j(n)\big) + \frac{C(q^{-1})}{D(q^{-1})}\varepsilon(n), \tag{5.67}$$

and the polynomial models of static nonlinearities

$$f_j\big(u_j(n)\big) = \mu_{j1}u_j(n) + \ldots + \mu_{jr_j}u_j^{r_j}(n), \tag{5.68}$$

where

$$A_j(q^{-1}) = 1 + a_{j1}q^{-1} + \ldots + a_{jna_j}q^{-na_j}, \tag{5.69}$$

$$B_j(q^{-1}) = 1 + b_{j1}q^{-1} + \ldots + b_{jnb_j}q^{-nb_j}, \tag{5.70}$$

$$C(q^{-1}) = 1 + c_1q^{-1} + \ldots + c_{nc}\,q^{-nc}, \tag{5.71}$$

$$D(q^{-1}) = 1 + d_1q^{-1} + \ldots + d_{nd}\,q^{-nd}. \tag{5.72}$$

Assume that $\varepsilon(n)$ is a zero-mean white noise, the model (5.67) is asymptotically stable, and the polynomial orders na_j, nb_j, r_j, nc, and nd are known. The model (5.67) can be transformed into the following equivalent form:

$$\bar{A}(q^{-1})y(n) = \sum_{j=1}^{nu} \bar{B}_j(q^{-1}) f_j\big(u_j(n)\big) + \frac{\bar{C}(q^{-1})}{D(q^{-1})}\varepsilon(n), \tag{5.73}$$

where

$$\bar{A}(q^{-1}) = \prod_{k=1}^{nu} A_j(q^{-1}) = 1 + \bar{a}_1q^{-1} + \ldots + \bar{a}_{na}q^{-na}, \tag{5.74}$$

$$\bar{B}_j(q^{-1}) = B_j(q^{-1}) \prod_{\substack{k=1 \\ k\neq j}}^{nu} A_k(q^{-1}) = \bar{b}_{j1}q^{-1} + \ldots + \bar{b}_{jnab_j}q^{-nab_j}, \tag{5.75}$$

$$\bar{C}(q^{-1}) = C(q^{-1})\bar{A}(q^{-1}) = 1 + \bar{c}_1q^{-1} + \ldots + \bar{c}_{nac}q^{-nac}, \tag{5.76}$$

with $na = na_1 + \ldots + na_{nu}$, $nab_j = na_1 + \ldots + na_{j-1} + nb_j + na_{j+1} + \ldots + na_{nu}$, $nac = na + nc$. Performing another transformation results in a model in the pseudolinear-in-parameters form:

$$\mathcal{A}(q^{-1})y(n) = \sum_{j=1}^{nu} \mathcal{B}_j(q^{-1}) f_j\big(u_j(n)\big) + \bar{C}(q^{-1})\varepsilon(n), \tag{5.77}$$

where

$$\mathcal{A}(q^{-1}) = \bar{A}(q^{-1})D(q^{-1}) = 1 + \alpha_1 q^{-1} + \ldots + \alpha_{n\alpha} q^{-n\alpha}, \tag{5.78}$$

$$\mathcal{B}_j(q^{-1}) = \bar{B}_j(q^{-1})D(q^{-1}) = \beta_{j1} q^{-1} + \ldots + \beta_{jn\beta_j} q^{-n\beta_j}, \tag{5.79}$$

with $n\alpha = na + nd$ and $n\beta_j = nab_j + nd$. Introducing the parameter vector $\boldsymbol{\theta}$,

$$\begin{aligned}\boldsymbol{\theta} = \big[&\alpha_1 \ldots \alpha_{n\alpha}\ \beta_{11}\mu_{11} \ldots \beta_{11}\mu_{1r_1} \ldots \beta_{1n\beta_1}\mu_{11} \ldots \beta_{1n\beta_1}\mu_{1r_1} \ldots \\ &\beta_{nu1}\mu_{nu1} \ldots \beta_{nu1}\mu_{nur_{nu}} \ldots \beta_{nun\beta_{nu}}\mu_{nu1} \ldots \beta_{nun\beta_{nu}}\mu_{nur_{nu}} \\ &\bar{c}_1 \ldots \bar{c}_{nac}\big]^T,\end{aligned} \tag{5.80}$$

and the vector $\mathbf{x}_0(n)$,

$$\begin{aligned}\mathbf{x}_0(n) = \big[&y(n-1) \ldots y(n-n\alpha)\ u_1(n) \ldots u_1^{r_1}(n) \ldots u_1(i-n\beta_1) \ldots \\ &u_1^{r_1}(i-n\beta_1) \ldots u_{nu}(n) \ldots u_{nu}^{r_{nu}}(n) \ldots u_{nu}(n-n\beta_1) \ldots \\ &u_{nu}^{r_{nu}}(n-n\beta_1)\ \ \varepsilon(n-1) \ldots \varepsilon(n-nac)\big]^T,\end{aligned} \tag{5.81}$$

we have

$$y(n) = \mathbf{x}_0^T(n)\boldsymbol{\theta} + \varepsilon(n). \tag{5.82}$$

Introduce the prediction error $e(n)$

$$e(n) = y(n) - \mathbf{x}^T(n)\hat{\boldsymbol{\theta}}(n-1), \tag{5.83}$$

where $\hat{\boldsymbol{\theta}}$ is the vector of adjustable parameters

$$\begin{aligned}\hat{\boldsymbol{\theta}} = \big[&\hat{\alpha}_1 \ldots \hat{\alpha}_{n\alpha}\ \hat{\beta}_{11}\hat{\mu}_{11} \ldots \hat{\beta}_{11}\hat{\mu}_{1r_1} \ldots \hat{\beta}_{1n\beta_1}\hat{\mu}_{11} \ldots \hat{\beta}_{1n\beta_1}\hat{\mu}_{1r_1} \ldots \\ &\hat{\beta}_{nu1}\hat{\mu}_{nu1} \ldots \hat{\beta}_{nu1}\hat{\mu}_{nur_{nu}} \ldots \hat{\beta}_{nun\beta_{nu}}\hat{\mu}_{nu1} \ldots \hat{\beta}_{nun\beta_{nu}}\hat{\mu}_{nur_{nu}} \\ &\hat{\bar{c}}_1 \ldots \hat{\bar{c}}_{nac}\big]^T,\end{aligned} \tag{5.84}$$

and $\mathbf{x}(n)$ is defined as

$$\begin{aligned}\mathbf{x}(n) = \big[&y(n-1) \ldots y(n-n\alpha)\ u_1(n) \ldots u_1^{r_1}(n) \ldots u_1(i-n\beta_1) \ldots \\ &u_1^{r_1}(i-n\beta_1) \ldots u_{nu}(n) \ldots u_{nu}^{r_{nu}}(n) \ldots u_{nu}(n-n\beta_1) \ldots \\ &u_{nu}^{r_{nu}}(n-n\beta_1)\ \ e(n-1) \ldots e(n-nac)\big]^T.\end{aligned} \tag{5.85}$$

The estimate $\hat{\boldsymbol{\theta}}(n)$ of $\boldsymbol{\theta}_0$, which minimizes the sum of squared errors between the system and model outputs, can be obtained with the recursive pseudolinear regression algorithm as follows:

$$\hat{\boldsymbol{\theta}}(n) = \hat{\boldsymbol{\theta}}(n-1) + \mathbf{K}(n)e(n), \tag{5.86}$$

$$\mathbf{K}(n) = \boldsymbol{P}(n)\mathbf{x}(n) = \frac{\boldsymbol{P}(n-1)\mathbf{x}(n)}{1 + \mathbf{x}^T(n)\boldsymbol{P}(n-1)\mathbf{x}(n)}, \tag{5.87}$$

$$\boldsymbol{P}(n) = \boldsymbol{P}(n-1) - \mathbf{K}(n)\mathbf{x}^T(n)\boldsymbol{P}(n-1). \tag{5.88}$$

5.7 Identification of systems with two-segment nonlinearities

In some cases, only polynomials of a higher order can approximate nonlinear characteristics adequately. An increase in the polynomial order r causes a multiple increase in the overall number of parameters in the linear-in-parameters model (5.12). Moreover, a single polynomial model can be inaccurate in a whole range of the system input signal. An alternative approach to single polynomial model solutions, discussed earlier, was proposed by Vörös [161]. In this method, a two-segment description of the nonlinear characteristic, composed of separate polynomial maps for positive and negative inputs, is used. The main motivation for such an approach is a better fit without increasing polynomial orders for some types of nonlinearities.
Assume that the nonlinear characteristic $\hat{f}(\cdot)$ is

$$\hat{v}(n) = \hat{f}\big(u(n)\big) = \begin{cases} \hat{f}_1\big(u(n)\big), \text{ if } u(n) > 0 \\ \hat{f}_2\big(u(n)\big), \text{ if } u(n) < 0 \end{cases}, \tag{5.89}$$

where

$$\hat{f}_1\big(u(n)\big) = \sum_{k=1}^{r} \hat{\mu}_{1k} u^k(n), \tag{5.90}$$

$$\hat{f}_2\big(u(n)\big) = \sum_{k=1}^{r} \hat{\mu}_{2k} u^k(n). \tag{5.91}$$

Introducing a switching sequence $\{g(n)\}$,

$$g(n) = g\big(u(n)\big) = \begin{cases} 0, \text{ if } u(n) > 0 \\ 1, \text{ if } u(n) < 0 \end{cases}, \tag{5.92}$$

(5.89) can be written as

$$\hat{v}(n) = \hat{f}_1\big(u(n)\big) + \big[\hat{f}_2\big(u(n)\big) - \hat{f}_1\big(u(n)\big)\big] g(n). \tag{5.93}$$

Then taking into account (5.90) and (5.91),

$$\hat{v}(n) = \sum_{k=1}^{r} \hat{\mu}_{1k} u^k(n) + \sum_{k=1}^{r} p_k u^k(n) g(n), \tag{5.94}$$

where

$$p_k = \hat{\mu}_{2k} - \hat{\mu}_{1k}. \tag{5.95}$$

The substitution of (5.94) into (5.1) results in a model which is nonlinear in the parameters. However, assuming $\hat{b}_1 = 1$, the model (5.1) can be expressed as

$$\hat{y}(n) = \hat{v}(n-1) + \big(\hat{B}(q^{-1}) - 1\big)\hat{v}(n) + \big(1 - \hat{A}(q^{-1})\big)\hat{y}(n). \tag{5.96}$$

Then substituting (5.94) only for $\hat{v}(n-1)$, the following pseudolinear-in-parameters model of the parallel type can be obtained:

$$\begin{aligned}\hat{y}(n) = &\sum_{k=1}^{r} \hat{\mu}_{1k} u^k(n-1) + \sum_{k=1}^{r} p_k u^k(n-1) g(n-1) \\ &+ \big(\hat{B}(q^{-1}) - 1\big)\hat{v}(n) + \big(1 - \hat{A}(q^{-1})\big)\hat{y}(n).\end{aligned} \tag{5.97}$$

Replacing $\hat{y}(n)$ with $y(n)$ transforms the model (5.97) into the series-parallel form:

$$\begin{aligned}\hat{y}(n) = &\sum_{k=1}^{r} \hat{\mu}_{1k} u^k(n-1) + \sum_{k=1}^{r} p_k u^k(n-1) g(n-1) \\ &+ \big(\hat{B}(q^{-1}) - 1\big)\hat{v}(n) + \big(1 - \hat{A}(q^{-1})\big)y(n)\end{aligned} \tag{5.98}$$

or

$$\hat{y}(n) = \mathbf{x}^T(n)\hat{\boldsymbol{\theta}}, \tag{5.99}$$

where the parameter vector is defined as

$$\hat{\boldsymbol{\theta}} = \big[\hat{\mu}_{11} \dots \hat{\mu}_{1r}\ p_1 \dots p_r\ \hat{b}_2 \dots \hat{b}_{nb}\ \hat{a}_1 \dots \hat{a}_{na}\big]^T, \tag{5.100}$$

and the regression vector is

$$\begin{aligned}\mathbf{x}(n) = \big[&u(n-1) \dots u^r(n-1)\ u(n-1)g(n-1) \dots u^r(n-1)g(n-1) \\ &\hat{v}(n-2) \dots \hat{v}(n-nb)\ -y(n-1) \dots -y(n-na)\big]^T.\end{aligned} \tag{5.101}$$

The model (5.98) is also of the pseudolinear form as $\hat{v}(n)$ is an unmeasurable variable which depends on the parameters of the nonlinear function $\hat{f}(.)$. Therefore, no noniterative algorithm can be applied to estimate $\hat{\boldsymbol{\theta}}$. Vörös proposed an iterative algorithm which uses the preceding parameter estimates of $\hat{\mu}_{1k}$, and p_k to estimate $\hat{v}(n)$. Denote with $\hat{v}^{(j)}(n)$, $\hat{\mu}_{1k}^{(j)}$, and $p_k^{(j)}$ the estimates of $\hat{v}(n)$, $\hat{\mu}_{1k}$ and p_k obtained at the step j:

$$\hat{v}^{(j)}(n) = \sum_{k=1}^{r} \hat{\mu}_{1k}^{(j)} u^k(n) + \sum_{k=1}^{r} p_k^{(j)} u^k(n) g(n). \tag{5.102}$$

Then the error to be minimized can be expressed as

$$e(n) = y(n) - \mathbf{x}^{(j)T}(n)\hat{\boldsymbol{\theta}}(j+1), \tag{5.103}$$

where $\mathbf{x}^{(j)}(n)$ is the regression vector with the estimates of $\hat{v}(n)$ calculated according to (5.102), and $\hat{\boldsymbol{\theta}}(j+1)$ is the $(j+1)$th estimate of $\hat{\boldsymbol{\theta}}$.
The iterative identification procedure can be divided into the following steps:

1. Using the regression vector $\mathbf{x}^{(j)}(n)$, minimize a proper criterion based on (5.103) to estimate $\hat{\boldsymbol{\theta}}(j+1)$.

2. Using (5.102), calculate $\hat{v}^{(j+1)}(n)$.
3. Repeat Steps 1 and 2 until the parameter estimates converge to constant values.

Testing the above identification procedure with differently shaped two-segment polynomial and exponential nonlinearities has revealed its good convergence properties, although the formal proof of the convergence is not available. Although the iterative procedure which uses the whole input-output data set was originally proposed by Vörös, the derivation of the sequential version of the method is straightforward.
A similar approach can also be used for the identification of discontinuous Hammerstein systems, i.e., systems with the nonlinear element described by a discontinuous function, or Wiener systems with this type of nonlinearity [160].

5.8 Summary

We have presented seven different methods of the identification of Hammerstein systems, which use Hammerstein models with a polynomial model of the nonlinear element. The iterative least squares method of Narendra and Gallman and the noniterative least squares method of Chang and Luus are the oldest and belong to the best known methods. Both of them have some drawbacks. While the convergence of the first one is not guaranteed, the drawbacks of the other one are parameter redundancy and a huge number of parameters in the case of high order models. The iterative method of Haist *et al.* [63] also uses a non-unique parameterization but, in contrast to a one-step method of Chang and Luus, its advantage are consistent parameter estimates in the case of correlated output noise. In the method of Thathachar and Ramaswamy, a Laguerre expansion is used to represent the linear part of the model. The parameters of the model are adjusted iteratively with a gradient based method. The parameters of a pulse transfer function model of the linear dynamic system and a polynomial model of the nonlinear element are estimated iteratively with the Levenberg-Marquardt method in the prediction error approach discussed by Eskinat *et al.* [34]. An alternative to the prediction error method is the pseudolinear regression method proposed by Boutayeb *et al.* Finally, the iterative method of Vörös allows one to identify Hammerstein systems using a two-segment polynomial model of the nonlinear element. The method uses a linear regression approach but a formal proof of convergence is not available.