4

Polynomial Wiener models

This chapter deals with polynomial Wiener models, i.e., models composed of a pulse transfer model of the linear dynamic system and a polynomial model of the nonlinear element or the inverse nonlinear element. A modified definition of the equation error and a modified series-parallel Wiener model are introduced. Assuming that the nonlinear element is invertible and the inverse nonlinear element can be described by a polynomial, the modified series-parallel Wiener model can be transformed into the linear-in-parameters form and its parameters can be calculated with the least squares method. Such an approach, however, results in inconsistent parameter estimates. As a remedy against this problem, an instrumental variables method is proposed with instrumental variables chosen as delayed system inputs and delayed and powered delayed outputs of the model obtained using the least squares method.

An alternative to this combined least squares-instrumental variables approach is the prediction method, in which the parameters of the noninverted nonlinear element are estimated, see [128] for the batch version and [85] for the sequential one. The pseudolinear regression method [86], being a simplified version of the prediction error method of lower computational requirements, is another effective technique for parameter estimation in Wiener systems disturbed additively by a discrete-time white noise.

This chapter is organized as follows: First, the least squares approach to the identification of Wiener systems based on the modified series-parallel model is introduced in Section 4.1. Two different cases of a Wiener system with and without the linear term are considered. Section 4.2 contains details of the recursive prediction error approach to the identification of polynomial Wiener systems. The pseudolinear regression method is discussed in Section 4.3. Finally, a brief summary is given in Section 4.4.

A. Janczak: Identification of Nonlinear Systems, LNCIS 310, pp. 117–141, 2005.

4.1 Least squares approach to the identification of Wiener systems

This section presents a least squares approach to the identification of polynomial Wiener systems. To transform the Wiener model into the linear-in-parameters form, the noninverted model of the linear dynamic system and the inverse model of the nonlinear element are used. The following assumptions are made about the identified Wiener system:

Assumption 4.1. The SISO Wiener system is

$$y(n) = f\left(\frac{B(q^{-1})}{A(q^{-1})}u(n) + \varepsilon(n)\right), \tag{4.1}$$

where

$$A(q^{-1}) = 1 + a_1 q^{-1} + \ldots + a_{na} q^{-na}, \tag{4.2}$$

$$B(q^{-1}) = b_1 q^{-1} + \ldots + b_{nb} q^{-nb}, \tag{4.3}$$

and $\varepsilon(n)$ is the additive disturbance.

Assumption 4.2. The polynomials $A(q^{-1})$ and $B(q^{-1})$ are coprime.

Assumption 4.3. The orders na and nb of the polynomials $A(q^{-1})$ and $B(q^{-1})$ are known.

Assumption 4.4. The linear dynamic system is casual and asymptotically stable.

Assumption 4.5. The input $u(n)$ has finite moments and is independent of $\varepsilon(k)$ for all n and k.

Assumption 4.6. The nonlinear function $f(\cdot)$ is defined on the interval $[a, b]$.

Assumption 4.7. The nonlinear function $f(\cdot)$ is invertible.

Assumption 4.8. The inverse nonlinear function $f^{-1}\big(y(n)\big)$ can be expressed by the polynomial

$$f^{-1}\big(y(n)\big) = \gamma_0 + \gamma_1 y(n) + \gamma_2 y(n)^2 + \cdots + \gamma_r y(n)^r \tag{4.4}$$

of a known order r.

The identification problem can be formulated as follows: Given the sequence of the Wiener system input and output measurements $\{u(n), y(n)\}$, $i = 1, \ldots, N$, estimate the parameters of the linear dynamic system and the inverse nonlinear element minimizing the following criterion:

$$J(n) = \frac{1}{2}\sum_{j=1}^{N}\big(y(n) - \hat{y}(n)\big)^2, \tag{4.5}$$

where $\hat{y}(n)$ is the output of the Wiener model.

4.1.1 Identification error

For a polynomial Wiener model, both its parallel and series-parallel forms are nonlinear functions of model parameters. Moreover, the series-parallel model contains not only a model of the nonlinear element but also its inverse [73] – Fig. 4.2. Consider the parallel model of the Wiener model given by

$$\hat{y}(n) = \hat{f}\left(\frac{\hat{B}(q^{-1})}{\hat{A}(q^{-1})}u(n)\right), \tag{4.6}$$

with

$$\hat{A}(q^{-1}) = 1 + \hat{a}_1 q^{-1} + \cdots + \hat{a}_{na} q^{-na}, \tag{4.7}$$

$$\hat{B}(q^{-1}) = \hat{b}_1 q^{-1} + \cdots + \hat{b}_{nb} q^{-nb}. \tag{4.8}$$

If $\hat{f}(\cdot)$ is invertible, (4.6) can be written as

$$\hat{f}^{-1}\big(\hat{y}(n)\big) = \frac{\hat{B}(q^{-1})}{\hat{A}(q^{-1})}u(n). \tag{4.9}$$

Assumption 4.9. The inverse nonlinear function $\hat{f}^{-1}(\cdot)$ has the form of a polynomial of the order r:

$$\hat{f}^{-1}\big(\hat{y}(n)\big) = \hat{\gamma}_0 + \hat{\gamma}_1 \hat{y}(n) + \hat{\gamma}_2 \hat{y}^2(n) + \cdots + \hat{\gamma}_r \hat{y}^r(n). \tag{4.10}$$

Assume also that the polynomial (4.10) contains the linear term, i.e., $\hat{\gamma}_1 \neq 0$. Then combining (4.10) and (4.9), the output of the model can be expressed as [80, 83]:

$$\hat{y}(n) = \frac{1}{\hat{\gamma}_1}\left(\frac{\hat{B}(q^{-1})}{\hat{A}(q^{-1})}u(n) - \Delta\hat{f}^{-1}\big(\hat{y}(n)\big)\right), \tag{4.11}$$

where

$$\Delta\hat{f}^{-1}\big(\hat{y}(n)\big) = \hat{\gamma}_0 + \hat{\gamma}_2 \hat{y}^2(n) + \hat{\gamma}_3 \hat{y}^3(n) + \cdots + \hat{\gamma}_r \hat{y}^r(n). \tag{4.12}$$

The model (4.11) can be written as

$$\hat{y}(n) = \big(1 - \hat{A}(q^{-1})\big)\hat{y}(n) + \frac{1}{\hat{\gamma}_1}\left(\hat{B}(q^{-1})u(n) - \hat{A}(q^{-1})\Delta\hat{f}^{-1}\big(\hat{y}(n)\big)\right). \tag{4.13}$$

Replacing $\hat{y}(n)$ by $y(n)$ on the r.h.s. of (4.13), the following modified series-parallel model can be obtained:

$$\hat{y}(n) = \big(1 - \hat{A}(q^{-1})\big)y(n) + \frac{1}{\hat{\gamma}_1}\left[\hat{B}(q^{-1})u(n) - \hat{A}(q^{-1})\Delta\hat{f}^{-1}\big(\hat{y}(n)\big)\right]. \tag{4.14}$$

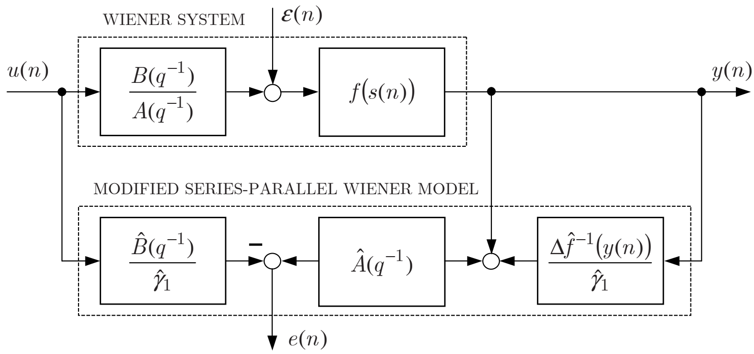

Fig. 4.1. Modified series-parallel Wiener model. The identification error definition for systems with the linear term

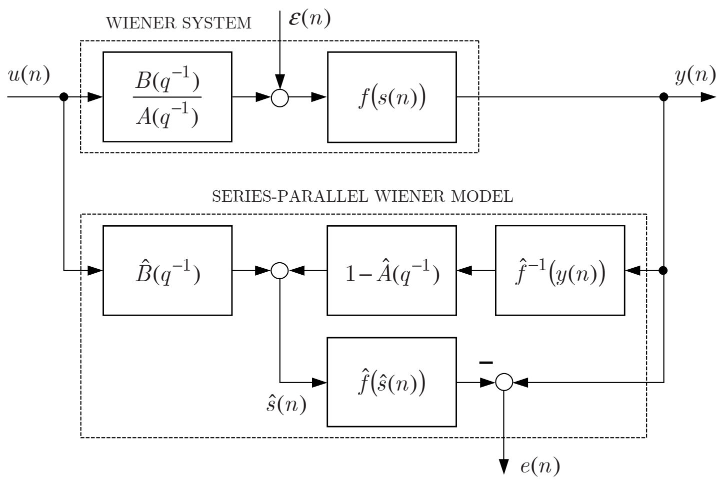

Fig. 4.2. Series-parallel Wiener model. The definition of the identification error

The modified series-parallel model, shown in Fig. 4.1, is different from the series-parallel model, which contains both the model of the nonlinear element and its inverse,

$$\hat{y}(n) = \hat{f}\Big[\big(1 - \hat{A}(q^{-1})\big)\hat{f}^{-1}\big(y(n)\big) + \hat{B}(q^{-1})u(n)\Big], \tag{4.15}$$

and the inverse series-parallel model

$$\hat{u}(n-1) = \frac{1}{\hat{b}_1}\Big[\big(\hat{b}_1 - \hat{B}(q^{-1})\big)u(n) + \hat{A}(q^{-1})\hat{f}^{-1}\big(\hat{y}(n)\big)\Big], \tag{4.16}$$

see Figs 4.2 – 4.3 for comparison. Applying (4.14), the following modified definition of the identification error can be introduced:

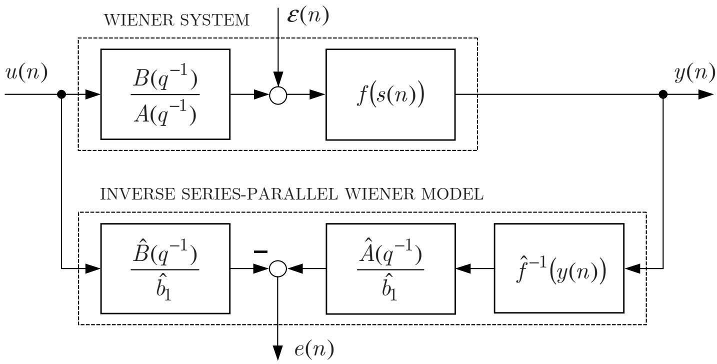

Fig. 4.3. Inverse series-parallel Wiener model. The definition of the identification error

$$e(n) = y(n) - \hat{y}(n) \tag{4.17}$$

$$e(n) = \hat{A}(q^{-1})y(n) - \frac{1}{\hat{\gamma}_1}\Big[\hat{B}(q^{-1})u(n) - \hat{A}(q^{-1})\Delta\hat{f}^{-1}\big(\hat{y}(n)\big)\Big]. \tag{4.18}$$

4.1.2 Nonlinear characteristic with the linear term

Assuming that the identified Wiener system has an invertible nonlinear characteristic with $\gamma_1 \neq 0$, we will express the modified series-parallel Wiener in the linear-in-parameters form.
Introduce the parameter vector $\hat{\boldsymbol{\theta}}$

$$\hat{\boldsymbol{\theta}} = \big[\hat{a}_1 \ldots \hat{a}_{na}\ \hat{\beta}_1 \ldots \hat{\beta}_{nb}\ \hat{\alpha}_{00}\ \hat{\alpha}_{20} \ldots \hat{\alpha}_{rna}\big]^T, \tag{4.19}$$

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

$$\begin{aligned}\mathbf{x}(n) = \big[&-y(n-1) \ldots -y(n-na)\ u(n-1) \ldots u(n-nb)\\ &1\ \ -y^2(n) \ldots -y^r(n-na)\big]^T,\end{aligned} \tag{4.20}$$

where

$$\hat{\beta}_k = \frac{\hat{b}_k}{\hat{\gamma}_1},\ k = 1, \ldots, nb, \tag{4.21}$$

$$\hat{\alpha}_{jk} = \begin{cases} \left(1 + \sum\limits_{m=1}^{na} \hat{a}_m\right)\dfrac{\hat{\gamma}_j}{\hat{\gamma}_1}, & k = 0,\ j = 0,\\ \dfrac{\hat{\gamma}_j}{\hat{\gamma}_1}, & k = 0,\ j = 2, 3, \ldots, r,\\ \hat{a}_k\dfrac{\hat{\gamma}_j}{\hat{\gamma}_1}, & k = 1, \ldots, na,\ j = 2, 3, \ldots, r. \end{cases} \tag{4.22}$$

Then the model (4.14) can be written as

$$\hat{y}(n) = \mathbf{x}^T(n)\hat{\boldsymbol{\theta}}. \tag{4.23}$$

Minimizing (4.5), the parameter vector $\hat{\boldsymbol{\theta}}$ can be calculated with the least squares (LS) or recursive least squares (RLS) method. Note that the number of parameters in (4.14) is $na+nb+r(na+1)$, while the number of parameters of $\hat{A}(q^{-1})$, $\hat{B}(q^{-1})$, and $\hat{f}(\cdot)$ is $na+nb+r+1$. Therefore, to obtain a unique solution, methods similar to these proposed in [34] for the identification of Hammerstein models can be employed.

4.1.3 Nonlinear characteristic without the linear term

Consider a Wiener system that fulfills the following assumptions:

- The linear term $\gamma_1 = 0$.
- The second order term $\gamma_2 \neq 0$.

In this case, the following modified series-parallel model can be defined (Fig. 4.4):

$$\hat{y}^2(n) = \big(1-\hat{A}(q^{-1})\big)y^2(n) + \frac{1}{\hat{\gamma}_2}\Big[\hat{B}(q^{-1})u(n) - \hat{A}(q^{-1})\Delta\hat{f}^{-1}\big(\hat{y}(n)\big)\Big]. \tag{4.24}$$

Now, the identification error can be defined as

$$\begin{aligned} e(n) &= y^2(n) - \hat{y}^2(n) \\ &= \hat{A}(q^{-1})y^2(n) - \frac{1}{\hat{\gamma}_2}\Big[\hat{B}(q^{-1})u(n) - \hat{A}(q^{-1})\Delta\hat{f}^{-1}\big(\hat{y}(n)\big)\Big]. \end{aligned} \tag{4.25}$$

Hence, (4.24) can be written in the following linear-in-parameters form:

$$\hat{y}^2(n) = \mathbf{x}^T(n)\hat{\boldsymbol{\theta}}, \tag{4.26}$$

with the parameter vector $\hat{\boldsymbol{\theta}}$,

$$\hat{\boldsymbol{\theta}} = \big[\hat{a}_1 \ldots \hat{a}_{na}\ \hat{\beta}_1 \ldots \hat{\beta}_{nb}\ \hat{\alpha}_{00}\ \hat{\alpha}_{30} \ldots \hat{\alpha}_{rna}\big]^T, \tag{4.27}$$

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

$$\begin{aligned} \mathbf{x}(n) = \big[&-y^2(n-1) \ldots -y^2(n-na)\ u(n-1) \ldots u(n-nb) \\ &1\ -y^3(n) \ldots -y^r(n-na)\big]^T, \end{aligned} \tag{4.28}$$

where

$$\hat{\beta}_k = \frac{\hat{b}_k}{\hat{\gamma}_2},\ k = 1, \ldots, nb, \tag{4.29}$$

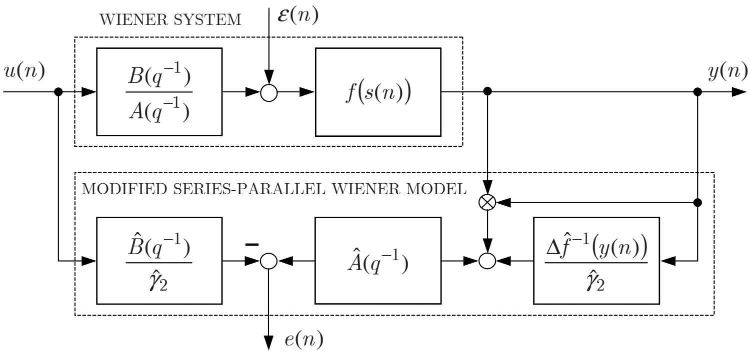

Fig. 4.4. Modified series-parallel model. The identification error definition for systems without the linear term

$$\hat{\alpha}_{jk} = \begin{cases} \left(1 + \sum_{m=1}^{na} \hat{a}_m\right) \frac{\hat{\gamma}_j}{\hat{\gamma}_2}, & k = 0,\ j = 0, \\ \frac{\hat{\gamma}_j}{\hat{\gamma}_2}, & k = 0,\ j = 3, 4, \ldots, r, \\ \hat{a}_k \frac{\hat{\gamma}_j}{\hat{\gamma}_2}, & k = 1, \ldots, na, j = 3, 4, \ldots, r. \end{cases} \tag{4.30}$$

As in the previous case, the parameter vector $\hat{\boldsymbol{\theta}}$ can be calculated with the least squares (LS) or recursive least squares (RLS) method minimizing the following criterion:

$$J = \frac{1}{2} \sum_{j=1}^{N} \left(y^2(n) - \hat{y}^2(n)\right)^2. \tag{4.31}$$

4.1.4 Asymptotic bias error of the LS estimator

Consider a polynomial Wiener system (4.1) – (4.4) that contains the linear term, i.e., $\gamma_1 \neq 0$, and the modified series-parallel Wiener model (4.23).We will show now that parameter estimates of the Wiener system obtained with the LS method are nonconsistent, i.e. asymptotically biased, even if the additive disturbance $\varepsilon(n)$ is

$$\varepsilon(n) = \frac{\epsilon(n)}{A(q^{-1})}, \tag{4.32}$$

where $\epsilon(n)$ is the discrete time white noise.

Theorem 4.1. Let $\hat{\boldsymbol{\theta}}$ denote the vector of parameter estimates, defined by (4.19), and $\boldsymbol{\theta}$ – the corresponding true parameter vector of the Wiener system, defined by (4.1) – (4.4).

Then the LS estimate of $\boldsymbol{\theta}$ is asymptotically biased, i.e., $\hat{\boldsymbol{\theta}}$ does not converge (with the probability 1) to the true parameter vector $\boldsymbol{\theta}$.

Proof: The output $y(n)$ of the Wiener system, defined by (4.1) and (4.32), is

$$y(n)=\big(1-A(q^{-1})\big)y(n)+\frac{1}{\gamma_1}\Big[B(q^{-1})u(n)-A(q^{-1})\Delta f^{-1}\big(\hat{y}(n)\big)+\epsilon(n)\Big]. \tag{4.33}$$

Introducing the true parameter vector $\boldsymbol{\theta}$,

$$\boldsymbol{\theta} = \big[a_1 \dots a_{na}\ \beta_1 \dots \beta_{nb}\ \alpha_{00}\ \alpha_{20} \dots \alpha_{rna}\big]^T, \tag{4.34}$$

where

$$\beta_k = \frac{b_k}{\gamma_1},\ \ k = 1, \dots, nb, \tag{4.35}$$

$$\alpha_{jk} = \begin{cases} \left(1 + \sum_{m=1}^{na} a_m\right)\dfrac{\gamma_j}{\gamma_1}, & k = 0,\ j = 0, \\ \dfrac{\gamma_j}{\gamma_1}, & k = 0,\ j = 2, 3, \dots, r, \\ a_k\dfrac{\gamma_j}{\gamma_1}, & k = 1, \dots, na,\ j = 2, 3, \dots, r, \end{cases} \tag{4.36}$$

the system output can be expressed as

$$y(n) = \mathbf{x}^T(n)\boldsymbol{\theta} + \frac{1}{\gamma_1}\,\epsilon(n). \tag{4.37}$$

The solution to the LS estimation problem is given by

$$\hat{\boldsymbol{\theta}} = \left[\frac{1}{N}\sum_{n=1}^{N}\mathbf{x}(n)\mathbf{x}^T(n)\right]^{-1}\left[\frac{1}{N}\sum_{n=1}^{N}\mathbf{x}(n)y(n)\right]. \tag{4.38}$$

From (4.37) and (4.38), it follows that the parameter estimation error $\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}$ is

$$\begin{aligned}\hat{\boldsymbol{\theta}}-\boldsymbol{\theta} &= \left[\frac{1}{N}\sum_{n=1}^{N}\mathbf{x}(n)\mathbf{x}^T(n)\right]^{-1}\left[\frac{1}{N}\sum_{n=1}^{N}\mathbf{x}(n)y(n)-\left(\frac{1}{N}\sum_{n=1}^{N}\mathbf{x}(n)\mathbf{x}^T(n)\right)\boldsymbol{\theta}\right] \\ &= \frac{1}{\gamma_1}\left[\frac{1}{N}\sum_{n=1}^{N}\mathbf{x}(n)\mathbf{x}^T(n)\right]^{-1}\left[\frac{1}{N}\sum_{n=1}^{N}\mathbf{x}(n)\epsilon(n)\right].\end{aligned} \tag{4.39}$$

Therefore, if $N \to \infty$,

$$\hat{\boldsymbol{\theta}}-\boldsymbol{\theta} \to \frac{1}{\gamma_1}\Big[\mathrm{E}\big(\mathbf{x}(n)\mathbf{x}^T(n)\big)\Big]^{-1}\Big[\mathrm{E}\big(\mathbf{x}(n)\epsilon(n)\big)\Big] \neq \mathbf{0}, \tag{4.40}$$

as $\mathrm{E}\big[y^2(n)\epsilon(n)\big] \neq 0, \dots,$ $\mathrm{E}\big[y^r(n)\epsilon(n)\big] \neq 0$, and thus $\mathrm{E}\big[\mathbf{x}(n)\epsilon(n)\big] \neq \mathbf{0}$.

Remark 4.1. In a similar way, it can also be shown that the parameter vector $\hat{\boldsymbol{\theta}}$ of the modified series-parallel Wiener model (4.24), calculated with the LS method, is asymptotically biased.

Remark 4.2. It can also be proved that asymptotically biased LS parameter estimates are obtained using other linear-in-parameter models which contain the inverse polynomial model of the nonlinear element. Examples of such models are the frequency sampling model [95, 96], the inverse Wiener model, and the model based on Laguerre filters [116].

4.1.5 Instrumental variables method

To obtain consistent parameter estimates, the regression vector $\mathbf{x}(n)$ should be uncorrelated with system disturbances. That is not the case if we use the modified series-parallel model, as the powered system outputs $y^2(n), \ldots, y^r(n)$ depend on $\epsilon(n)$. The instrumental variables method is a well-known remedy against such a situation. Applying the instrumental variables method, parameter estimation can be performed according to the following scheme:

1. Estimate the parameters of the system using the LS or the RLS method.
2. Simulate the model to determine the instrumental variables $\mathbf{z}(n)$.
3. Estimate the parameters of the system using the IV or the RIV method with the instrumental variables $\mathbf{z}(n)$.

The choice of instrumental variables is a vital design problem in any instrumental variables approach, see [149] for more details. Clearly, the best choice would be undisturbed powered system outputs, but these are not available for measurement. Instead, we can employ powered outputs of the model, or powered outputs of the linear dynamic model, calculated with the LS method, and define the instrumental variables as

$$\begin{aligned} \mathbf{z}(n) = \big[&- \hat{y}(n-1) \ldots - \hat{y}(n-na)\ u(n-1) \ldots u(n-nb) \\ &1\ - \hat{y}^2(n) \ldots - \hat{y}^r(n-na) \big]^T \end{aligned} \tag{4.41}$$

in the case of Wiener systems with the linear term, or

$$\begin{aligned} \mathbf{z}(n) = \big[&- \hat{y}^2(n-1) \ldots - \hat{y}^2(n-na)\ u(n-1) \ldots u(n-nb) \\ &1\ - \hat{y}^3(n) \ldots - \hat{y}^r(n-na) \big]^T \end{aligned} \tag{4.42}$$

in the case of Wiener systems without the linear term. The instrumental variables $\mathbf{z}(n)$ are uncorrelated with system disturbances:

$$E[\mathbf{z}(n)\varepsilon(n)] = \mathbf{0}. \tag{4.43}$$

4.1.6 Simulation example. Nonlinear characteristic with the linear term

A linear dynamic system given by the continuous-time polynomial pulse transfer function

$$G(s) = \frac{1}{6s^2 + 5s + 1} \tag{4.44}$$

was converted to discrete time, assuming a zero order hold on the input and the sampling interval 1s, leading to the following difference equation:

$$s(n) = 1.3231s(n–1) - 0.4346s(n-2) + 0.0635u(n–1) + 0.0481u(n-2). \tag{4.45}$$

The linear dynamic system was followed by a nonlinear element described by the function

$$f\big(s(n)\big) = 4\big(\sqrt[3]{0.75s(n)} - 1\big). \tag{4.46}$$

Therefore, the inverse nonlinear function $f^{-1}\big(y(n)\big)$ is a polynomial:

$$f^{-1}\big(y(n)\big) = \frac{4}{3} + y(n) + 0.25y^2(n) + \frac{1}{48}y^3(n). \tag{4.47}$$

The input sequence $\{u(n)\}$ consisted of 40000 pseudo-random numbers uniformly distributed in $(-5, 5)$. Parameter estimation was performed with both the LS method and the IV method, assuming that $r = 3$ and $\hat{\gamma}_1 = 1$. Additive system disturbances were given by $\varepsilon(n) = [1/A(q^{-1})]\,\epsilon(n)$ with $\{\epsilon(n)\}$ – a normally distributed pseudo-random sequence $\mathcal{N}(0, 0.1)$. This corresponds with the signal to noise ratio $SNR = \sqrt{\mathrm{var}(y(n) - \varepsilon(n))/\mathrm{var}(\varepsilon(n))} = 3.14$. The identification results, given in Tables 4.1 and 4.2 and illustrated in Figs 4.5 and 4.6, show a considerable improvement in IV parameter estimates in comparison with LS ones.

Table 4.1. Parameter estimates, $SNR = 3.14$

Parameter	True $\sigma_\varepsilon = 0$	LS $\sigma_\varepsilon = 0.1$	IV $\sigma_\varepsilon = 0.1$
a_1	−1.3231	−1.2803	−1.3292
a_2	0.4346	0.4158	0.4370
b_1	0.0635	0.0558	0.0636
b_2	0.0481	0.0423	0.0482
γ_0	1.3333	1.0764	1.3813
γ_1	1.0000	1.0000	1.0000
γ_2	0.2500	0.2473	0.2503
γ_3	0.0208	0.0199	0.0209

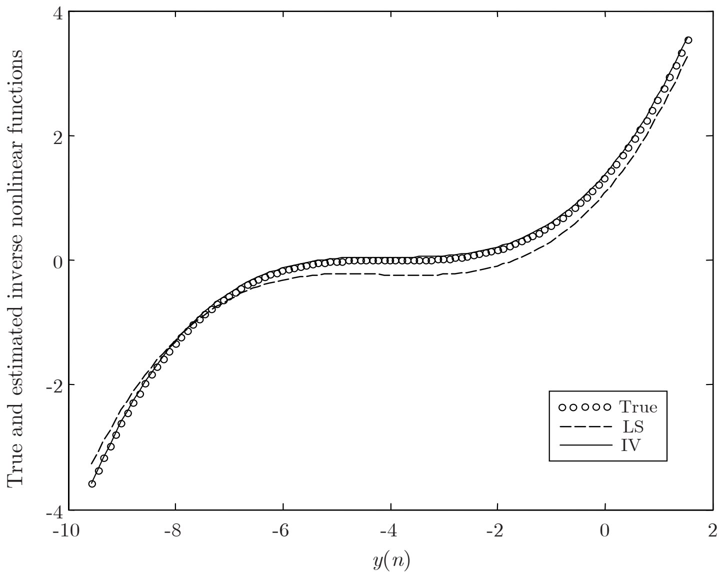

Fig. 4.5. Wiener system with the linear term. True $f^{-1}\big(y(n)\big)$ and estimated $\hat{f}^{-1}\big(y(n)\big)$ inverse nonlinear functions

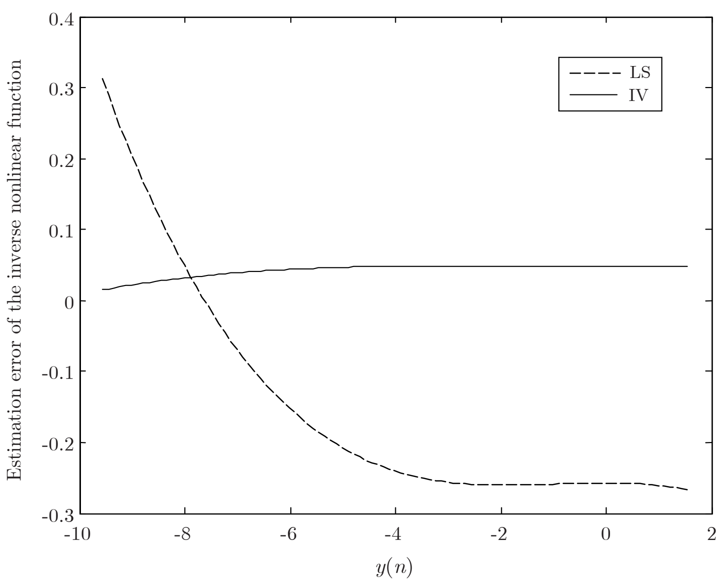

Fig. 4.6. Wiener system with the linear term. Estimation error $\hat{f}^{-1}\big(y(n)\big) - f^{-1}\big(y(n)\big)$.

Table 4.2. Comparison of estimation accuracy

Performance index	LS ($\sigma_\varepsilon = 0$)	LS	IV
$\frac{1}{4}\sum_{j=1}^{2}\left[(\hat{a}_j - a_j)^2 + (\hat{b}_j - b_j)^2\right]$	4.62×10^{-23}	5.70×10^{-4}	1.08×10^{-5}
$\frac{1}{3}\left[(\hat{\gamma}_0 - \gamma_0)^2 + \sum_{j=2}^{3}(\hat{\gamma}_j - \gamma_j)^2\right]$	1.13×10^{-21}	2.20×10^{-2}	7.66×10^{-4}
$\frac{1}{50}\sum_{i=1}^{50}\left[\hat{f}^{-1}\big(y(n)\big) - f^{-1}\big(y(n)\big)\right]^2$	3.78×10^{-21}	4.74×10^{-2}	1.91×10^{-3}

4.1.7 Simulation example. Nonlinear characteristic without the linear term

The linear dynamic system (4.45) and a nonlinear element defined by the function

$$f\big(s(n)\big) = \sqrt{\sqrt{s(n)} + 0.5}, \quad s(n) \geqslant 0 \tag{4.48}$$

were used in the example of a Wiener system without the linear term and a nonzero second order term. The inverse nonlinear function is a polynomial:

$$f^{-1}\big(y(n)\big) = 0.25 - y^2(n) + y^4(n), \quad y(n) \geqslant \sqrt{0.5}. \tag{4.49}$$

The input sequence $\{u(n)\}$ contained 50000 pseudo-random numbers uniformly distributed in $(1.5, 6)$. Additive system disturbances were $\varepsilon(n) = [1/A(q^{-1})]\,\epsilon(n)$ with $\{\epsilon(n)\}$ – a normally distributed pseudo-random sequence $\mathcal{N}(0, 0.025)$. As in the previous example, parameter estimation was performed using the LS method and the IV method and assuming: $r = 4$, $\hat{\gamma}_1 = \hat{\gamma}_3 = 0$, $\hat{\gamma}_2 = 1$. The identification results, given in Tables 4.3 and 4.4 and illustrated in Figs 4.7 and 4.8, confirm practical feasibility of the proposed approach.

Table 4.3. Parameter estimates, $SNR = 19.37$

Parameter	True $\sigma_\varepsilon = 0$	LS $\sigma_\varepsilon = 0.025$	IV $\sigma_\varepsilon = 0.025$
a_1	−1.3231	−1.4448	−1.2898
a_2	0.4346	0.6233	0.4107
b_1	0.0635	0.0014	0.0635
b_2	0.0481	0.0011	0.0481
γ_0	0.2500	1.2119	0.2187
γ_2	1.0000	1.0000	1.0000
γ_4	1.0000	0.2212	1.0005

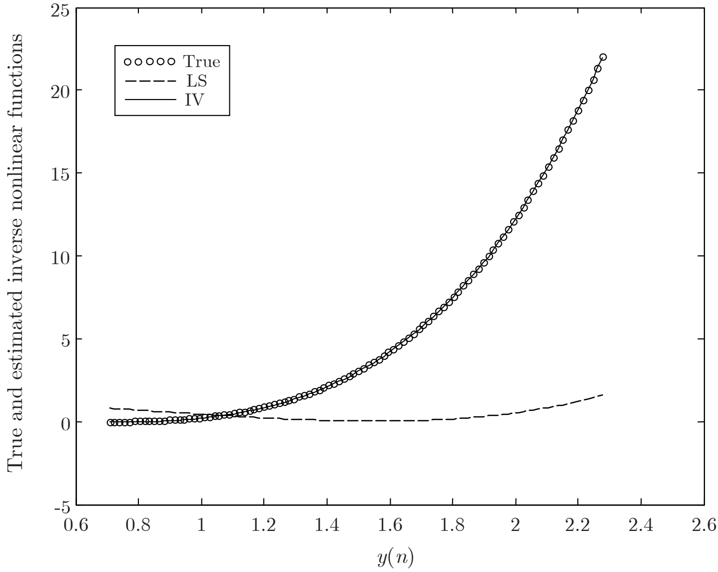

Fig. 4.7. Wiener system without the linear term. True $f^{-1}\big(y(n)\big)$ and estimated $\hat{f}^{-1}\big(y(n)\big)$ inverse nonlinear functions

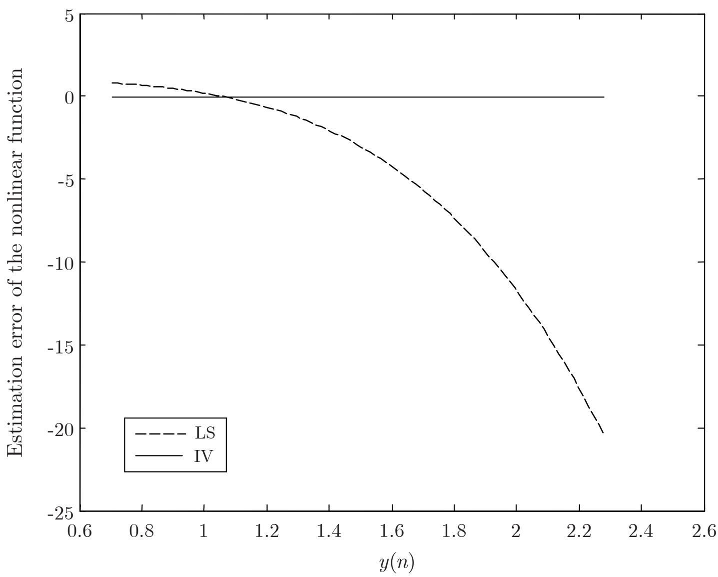

Fig. 4.8. Wiener system without the linear term. Estimation error $\hat{f}^{-1}\big(y(n)\big) - f^{-1}\big(y(n)\big)$

Table 4.4. Comparison of estimation accuracy

Performance index	LS ($\sigma_\varepsilon = 0$)	LS	IV
$\frac{1}{4}\sum_{j=1}^{2}\left[(\hat{a}_j - a_j)^2 + (\hat{b}_j - b_j)^2\right]$	3.85×10^{-16}	1.41×10^{-2}	4.20×10^{-4}
$\frac{1}{2}\left[(\hat{\gamma}_0 - \gamma_0)^2 + (\hat{\gamma}_4 - \gamma_4)^2\right]$	1.24×10^{-13}	7.66×10^{-1}	4.92×10^{-4}
$\frac{1}{50}\sum_{i=1}^{50}\left[\hat{f}^{-1}\big(y(n)\big) - f^{-1}\big(y(n)\big)\right]^2$	5.80×10^{-12}	6.16×10^{1}	7.58×10^{-4}

Although only one technique for instrumental variables generation is discussed and illustrated here, other known techniques can be considered as well. Contrary to the identification of the inverse Wiener model, an attractive feature of this approach is that the linear sub-system is not required to be minimum phase.

4.2 Identification of Wiener systems with the prediction error method

In the linear regression approach, described in Section 4.1, the class of identified systems is restricted by the assumption of invertibility of the nonlinear characteristic. This assumption is not necessary in the recursive prediction error method of Wigren [165], in which the nonlinear characteristic is approximated with a piecewise linear function.
This section presents an identification algorithm for Wiener systems which uses the recursive prediction error (RPE) approach with a polynomial model of the nonlinear element and a pulse transfer function model of the linear dynamic system – Fig. 4.9.

4.2.1 Polynomial Wiener model

Consider a discrete-time Wiener system (Fig. 4.9) composed of a SISO linear dynamic system in a cascade with a SISO nonlinear element. The output $y(n)$ of the Wiener system at the time n is

$$y(n) = f\big(s(n)\big) + \varepsilon(n), \tag{4.50}$$

where $f(\cdot)$ is the steady state characteristic, $\varepsilon(n)$ is the additive output disturbance, and $s(n)$ is the output of the linear dynamic system:

$$s(n) = \frac{B(q^{-1})}{A(q^{-1})}u(n) \tag{4.51}$$

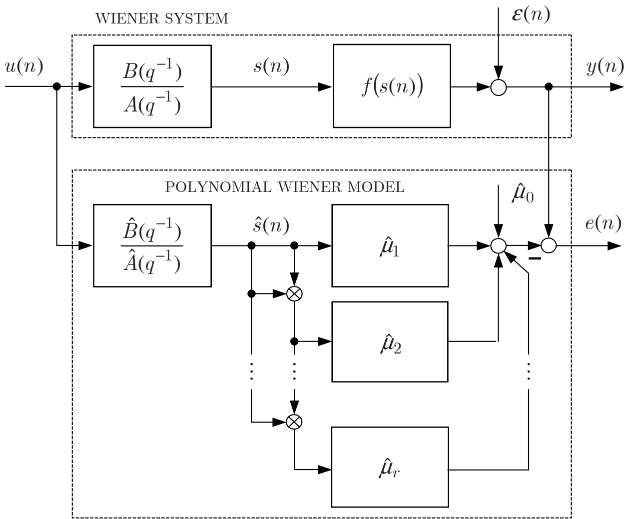

Fig. 4.9. Wiener system and its polynomial model

with

$$A(q^{-1}) = 1 + a_1 q^{-1} + \cdots + a_{na} q^{-na}, \tag{4.52}$$

$$B(q^{-1}) = b_1 q^{-1} + \cdots + b_{nb} q^{-nb}, \tag{4.53}$$

where $a_1, \ldots, a_{na}, b_1, \ldots, b_{nb}$ are the parameters of the linear dynamic system. Assume that the linear dynamic system is casual and asymptotically stable, and $f(\cdot)$ is a continuous function. Assume also that the polynomials $A(q^{-1})$ and $A(q^{-1})$ are coprime and $u(n)$ has finite moments and is independent of $\varepsilon(k)$ for all n and k. The steady state characteristic of the system can be approximated by a polynomial $\hat{f}(\cdot)$ of the order r:

$$\hat{f}\big(\hat{s}(n)\big) = \hat{\mu}_0 + \hat{\mu}_1 \hat{s}(n) + \hat{\mu}_2 \hat{s}^2(n) + \cdots + \hat{\mu}_r \hat{s}^r(n), \tag{4.54}$$

where $\hat{s}(n)$ is the output of the linear dynamic system model

$$\hat{s}(n) = \frac{\hat{B}(q^{-1})}{\hat{A}(q^{-1})} u(n) \tag{4.55}$$

with

$$\hat{A}(q^{-1}) = 1 + \hat{a}_1 q^{-1} + \cdots + \hat{a}_{na} q^{-na}, \tag{4.56}$$

$$\hat{B}(q^{-1}) = \hat{b}_1 q^{-1} + \cdots + \hat{b}_{nb} q^{-nb}, \tag{4.57}$$

where $\hat{a}_1, \ldots, \hat{a}_{na}, \hat{b}_1, \ldots, \hat{b}_{nb}$ are the parameters of the linear dynamic model. Therefore, the parameter vector $\hat{\boldsymbol{\theta}}$ of the model defined by (4.54) and (4.55) is

$$\hat{\boldsymbol{\theta}} = \left[\hat{a}_1 \ldots \hat{a}_{na} \; \hat{b}_1 \ldots \hat{b}_{nb} \; \hat{\mu}_0 \; \hat{\mu}_1 \ldots \hat{\mu}_r\right]^T. \tag{4.58}$$

4.2.2 Recursive prediction error method

The identification problem considered here can be formulated as follows: Given a set of input and output data $Z^N = \{(u(n), y(n)),\ k = 1, \ldots, N\}$ estimate the parameters of the Wiener system so that the predictions $\hat{y}(n|n-1)$ of the system output are close to the system output $y(n)$ in the sense of the following mean square error criterion:

$$J_N(\hat{\boldsymbol{\theta}}, Z^N) = \frac{1}{2N} \sum_{k=1}^{N} \big(y(n) - \hat{y}(n|n-1)\big)^2. \tag{4.59}$$

Given the gradient of the model output w.r.t. the parameter vector

$$\boldsymbol{\psi}(n) = \left[\frac{d\hat{y}(n|n-1)}{d\hat{\boldsymbol{\theta}}}\right]^T, \tag{4.60}$$

the RPE algorithm can be expressed by (2.97) – (2.99). In practice, it is useful to modify the criterion (4.59) with an exponential forgetting factor λ. The forgetting factor $\lambda \in [0, 1]$ and values close to 1 are commonly selected. To protect the algorithm from the so-called covariance blow-up phenomenon, other modifications of the algorithm may be useful that impose an upper bound on the eigenvalues of the matrix $\boldsymbol{P}(n)$ and are known as the constant trace and exponential forgetting and resetting algorithms [127].

4.2.3 Gradient calculation

The gradient $\boldsymbol{\psi}(n)$ of the model output w.r.t. to the model parameters is defined as

$$\boldsymbol{\psi}(n) = \left[\frac{\partial\hat{y}(n)}{\partial a_1} \cdots \frac{\partial\hat{y}(n)}{\partial a_{na}} \; \frac{\partial\hat{y}(n)}{\partial b_1} \cdots \frac{\partial\hat{y}(n)}{\partial b_{nb}} \, \frac{\partial\hat{y}(n)}{\partial\hat{\mu}_0} \; \frac{\partial\hat{y}(n)}{\partial\hat{\mu}_1} \cdots \frac{\partial\hat{y}(n)}{\partial\hat{\mu}_r}\right]^T. \tag{4.61}$$

Although the model given by (4.54) and (4.55) is a recurrent one due to the difference equation (4.55), the calculation of the gradient does not require much more computation than in the case of the pure static model. The only difference is in the calculation of partial derivatives of the model output w.r.t. the parameters of the linear dynamic model. This can be done with the sensitivity method solving by simulation the following set of linear difference equations [73, 74]:

$$\frac{\partial \hat{s}(n)}{\partial \hat{a}_k} = -\hat{s}(n-k) - \sum_{m=1}^{na} \hat{a}_m \frac{\partial \hat{s}(n-m)}{\partial \hat{a}_k}, \ k = 1, \ldots, na, \tag{4.62}$$

$$\frac{\partial \hat{s}(n)}{\partial \hat{b}_k} = u(n-k) - \sum_{m=1}^{na} \hat{a}_m \frac{\partial \hat{s}(n-m)}{\partial \hat{b}_k}, \ k = 1, \ldots, nb. \tag{4.63}$$

Hence, partial derivatives of the model output w.r.t. the parameters $\hat{a}_k$ and $\hat{b}_k$ can be calculated as

$$\frac{\partial \hat{y}(n)}{\partial \hat{a}_k} = \frac{\partial \hat{y}(n)}{\partial \hat{s}(n)} \frac{\partial \hat{s}(n)}{\partial \hat{a}_k}, \tag{4.64}$$

$$\frac{\partial \hat{y}(n)}{\partial \hat{b}_k} = \frac{\partial \hat{y}(n)}{\partial \hat{s}(n)} \frac{\partial \hat{s}(n)}{\partial \hat{b}_k}, \tag{4.65}$$

where

$$\frac{\partial \hat{y}(n)}{\partial \hat{s}(n)} = \hat{\mu}_1 + 2\hat{\mu}_2 \hat{s}(n) + \cdots + r\hat{\mu}_r \hat{s}^{r-1}(n). \tag{4.66}$$

Partial derivatives of the model output w.r.t. the parameters of the nonlinear element model are

$$\frac{\partial \hat{y}(n)}{\partial \hat{\mu}_j} = \hat{s}^j(n), \ j = 0, 1, \ldots, r. \tag{4.67}$$

Note that the derivation of the partial derivatives (4.62) and (4.63) is made under the assumption that the parameters of the linear dynamic model are time invariant. As this assumption is not true because of the sequential nature of the RPE algorithm, an approximate gradient is obtained. A more accurate evaluation of the gradient can be calculated using the truncated BPTT algorithm.

4.2.4 Pneumatic valve simulation example

The model of a pneumatic valve (2.100) – (2.101) was used in the simulation example. It was assumed that the system output $y(n)$ was additively disturbed by the zero-mean discrete white Gaussian noise $\varepsilon(n)$ with the standard deviation of $\sigma_\varepsilon = 0.005$ and 0.05:

$$y(n) = f\big(s(n)\big) + \varepsilon(n). \tag{4.68}$$

A sequence of 20000 pseudorandom numbers, uniformly distributed in $(0, 1)$, was used as the system input. Based on the simulated input-output data, the Wiener system was identified using the RPE algorithm. To compare estimation accuracy of the linear system parameters and the nonlinear function $f(\cdot)$, the indices (2.91) and (2.92) were used with $\{s(n)\}$ defined as a sequence of 100 linearly equally spaced values between $-\min\big(s(n)\big)$ and $\max\big(s(n)\big)$. The results shown in Tables 4.5 and 4.6 are illustrated in Figs 4.10 – 4.13.

In the example, the nonlinear characteristic is of an infinite order while a finite order model is estimated. In spite of the fact that the estimated parameters are different from the parameters of the polynomial approximating (2.101), the nonlinear characteristic is approximated quite well showing practical applicability of this approach.

Table 4.5. Parameter estimates

Parameter	Approximating polynomial	Estimated $\sigma_\varepsilon = 0.005$	Estimated $\sigma_\varepsilon = 0.05$
a_1	−1.4138	−1.4167	−1.4159
a_2	0.6065	0.6088	0.6092
b_1	0.1044	0.1059	0.0984
b_2	0.0833	0.0812	0.0899
μ_0	0.0010	0.0054	0.0595
μ_1	0.9530	1.0581	0.6871
μ_2	0.8149	−1.1304	−0.5716
μ_3	−11.651	−0.2255	0.1347
μ_4	34.749	0.8751	−0.0036
μ_5	−57.593	0.2007	−0.0121
μ_6	59.242	−0.3143	−0.0045
μ_7	−37.639	−0.3218	−0.0011
μ_8	13.5602	−0.0583	−0.0002
μ_9	−2.1210	0.2263	−0.0000

Table 4.6. Comparison of estimation accuracy, $s(j)$ – a pseudorandom sequence, uniformly distributed in $(\min\big(s(n)\big), \max\big(s(n)\big)$

Performance index	$\sigma_\varepsilon = 0.005$	$\sigma_\varepsilon = 0.05$
$\frac{1}{4}\sum_{j=1}^{2}\left[(\hat{a}_j - a_j)^2 + (\hat{b}_j - b_j)^2\right]$	2.95×10^{-6}	2.44×10^{-4}
$\frac{1}{100}\sum_{j=1}^{100}\left[\hat{f}\big(s(j)\big) - f\big(s(j)\big)\right]^2$	5.38×10^{-5}	1.87×10^{-3}

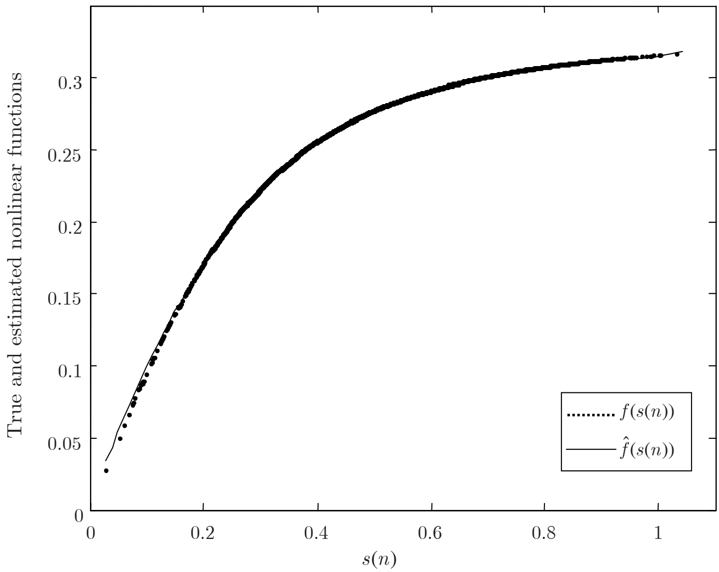

Fig. 4.10. True and estimated nonlinear functions ($\sigma_\varepsilon = 0.005$)

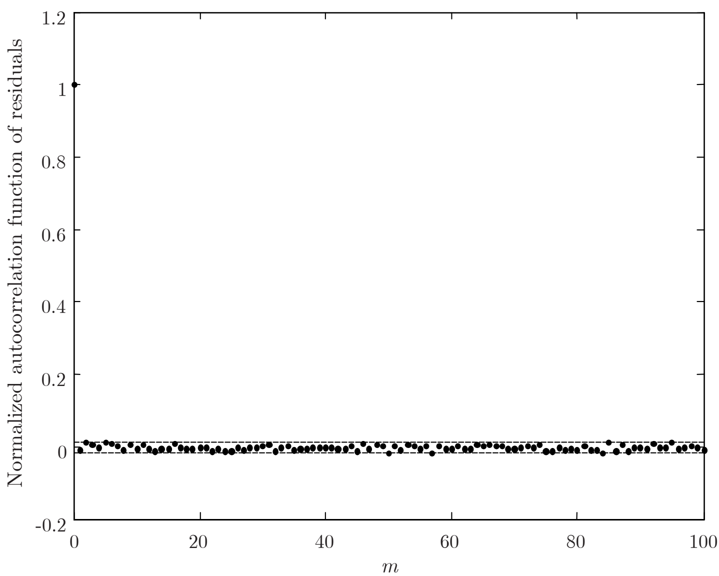

Fig. 4.11. Autocorrelation function of residuals and the 95% confidence interval ($\sigma_\varepsilon = 0.005$)

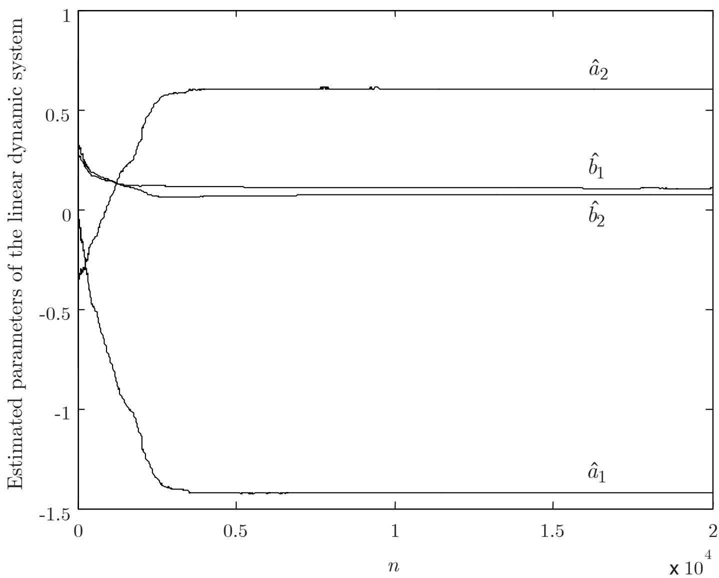

Fig. 4.12. Evolution of linear dynamic system parameter estimates ($\sigma_\varepsilon = 0.005$)

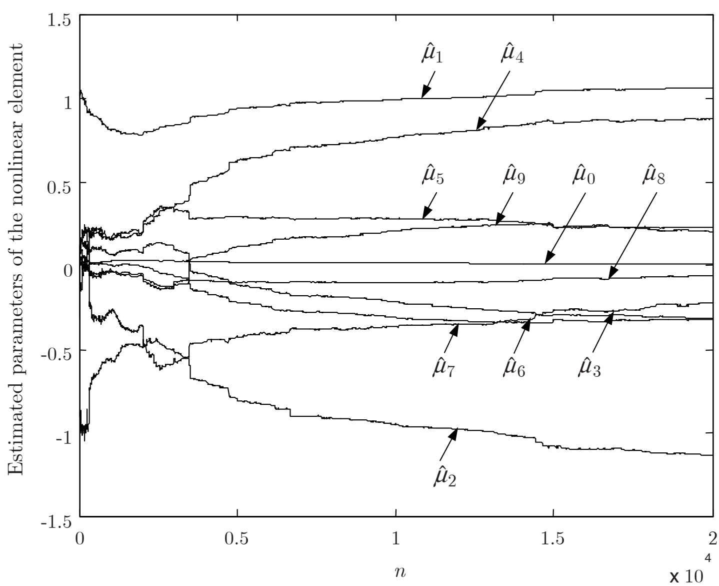

Fig. 4.13. Evolution of nonlinear element parameter estimates ($\sigma_\varepsilon = 0.005$)

4.3 Pseudolinear regression method

The pseudolinear regression approach to parameter estimation is based on the assumption that nonlinear components of a model can be neglected and the model can be treated as a linear-in-parameters one. For polynomial Wiener systems, we estimate the parameters of a pulse transform function and the parameters of nonlinear characteristic. It is assumed that the nonlinear characteristic contains the linear term. An obvious advantage of the pseudolinear regression approach, in comparison with other regression methods, is its low computational complexity and applicability for the identification of Wiener systems with noninvertible nonlinear characteristics. The pseudolinear regression method can be considered as a simplified prediction error approach, in which the exact gradient is replaced with an approximate one. Such a simplification reduces computational complexity of the method but may deteriorate its convergence rate.

4.3.1 Pseudolinear-in-parameters polynomial Wiener model

Consider the Wiener system (4.50) – (4.53) and its polynomial model (4.54) – (4.57). To derive the pseudolinear regression method, it is necessary to assume that the polynomial model $\hat{f}\big(\hat{s}(n)\big)$ has a nonzero linear term, i.e., $\hat{\mu}_1 \neq 0$. For convenience, we can assume that $\hat{\mu}_1 = 1$. Note that there is no loss of generality if we assume that $\hat{\mu}_1 = 1$, as the steady state gain of the linear dynamical model can be multiplied by $1/\hat{\mu}_1$. The polynomial model output can be written as

$$\begin{aligned}\hat{y}(n) = &- \hat{a}_1\hat{s}(n-1) - \cdots - \hat{a}_{na}\hat{s}(n-na) + \hat{b}_1 u(n-1) + \cdots \\ &+ \hat{b}_{nb}u(n-nb) + \hat{\mu}_0 + \hat{\mu}_2\hat{s}^2(n) + \cdots + \hat{\mu}_r\hat{s}^r(n) = \hat{\boldsymbol{\theta}}^T\boldsymbol{\psi},\end{aligned} \tag{4.69}$$

where

$$\hat{\boldsymbol{\theta}} = \big[\hat{\boldsymbol{\theta}}_l^T \ \hat{\mu}_0 \ \hat{\mu}_2 \ldots \hat{\mu}_r\big]^T, \tag{4.70}$$

$$\hat{\boldsymbol{\theta}}_l = \big[\hat{a}_1 \ldots \hat{a}_{na} \ \hat{b}_1 \ldots \hat{b}_{nb}\big]^T, \tag{4.71}$$

$$\boldsymbol{\psi}(n) = \big[\boldsymbol{\varphi}^T(n) \ 1 \ \hat{s}^2(n) \ldots - \hat{s}^r(n)\big]^T, \tag{4.72}$$

$$\boldsymbol{\varphi}(n) = \big[-\hat{s}(n-1) \ldots - \hat{s}(n-na) \ u(n-1) \ldots u(n-nb)\big]^T. \tag{4.73}$$

The model (4.69) is a linear function of the parameters $\hat{\mu}_0, \hat{\mu}_2, \ldots, \hat{\mu}_r$, but it is a nonlinear function of $\hat{a}_1$,. . .,$\hat{a}_{na}$, $\hat{b}_1$,. . .,$\hat{b}_{nb}$. This comes from the fact that both $\hat{s}^2(n), \ldots, \hat{s}^r(n)$ and $\hat{s}(n-1), \ldots, \hat{s}(n-na)$ depend on $\hat{\boldsymbol{\theta}}_l$.

4.3.2 Pseudolinear regression identification method

Minimization with respect to model parameters of the weighted cost function

$$J = \frac{1}{2}\sum_{j=1}^{N}\lambda^{N-j}\big(y(n) - \hat{y}(n)\big)^2 \tag{4.74}$$

results in the following identification algorithm:

$$\hat{\boldsymbol{\theta}}(n) = \hat{\boldsymbol{\theta}}(n-1) + \boldsymbol{K}(n)\, e(n), \tag{4.75}$$

$$e(n) = y(n) - \boldsymbol{\psi}^T(n)\hat{\boldsymbol{\theta}}(n-1), \tag{4.76}$$

$$\boldsymbol{K}(n) = \boldsymbol{P}(n)\boldsymbol{\psi}(n) = \frac{\boldsymbol{P}(n-1)\boldsymbol{\psi}(n)}{\lambda + \boldsymbol{\psi}^T(n)\boldsymbol{P}(n-1)\boldsymbol{\psi}(n)}, \tag{4.77}$$

$$\boldsymbol{P}(n) = \frac{1}{\lambda}\Big(\boldsymbol{P}(n-1) - \boldsymbol{K}(n)\boldsymbol{\psi}^T(n)\boldsymbol{P}(n-1)\Big), \tag{4.78}$$

where λ denotes the exponential forgetting factor.

4.3.3 Simulation example

The Wiener system described by the second order pulse transfer function

$$\frac{B(q^{-1})}{A(q^{-1})} = \frac{0.125q^{-1} - 0.025q^{-2}}{1 - 1.75q^{-1} + 0.85q^{-2}} \tag{4.79}$$

and the nonlinear characteristic (Fig. 4.16)

$$f\big(s(n)\big) = -0.25 + s(n) + s^2(n) - 0.5s^3(n) - 0.2s^4(n) + 0.2s^5(n) \tag{4.80}$$

was used in the numerical example. The system was excited with a sequence of 10000 pseudo-random numbers of uniform distribution in $(-1, 1)$. The system output was disturbed additively with another pseudo-random sequence of uniform distribution in $(-\sqrt{3\alpha}, \sqrt{3\alpha})$.
The identification results, obtained at $\alpha = 3.34{\times}10^{-5}$, $3.34{\times}10^{-3}$, $3.34{\times}10^{-1}$ and the forgetting factor $\lambda = 0.9995$, are summarized in Tables 4.7 and 4.8 and illustrated in Figs. 4.15 – 4.17.

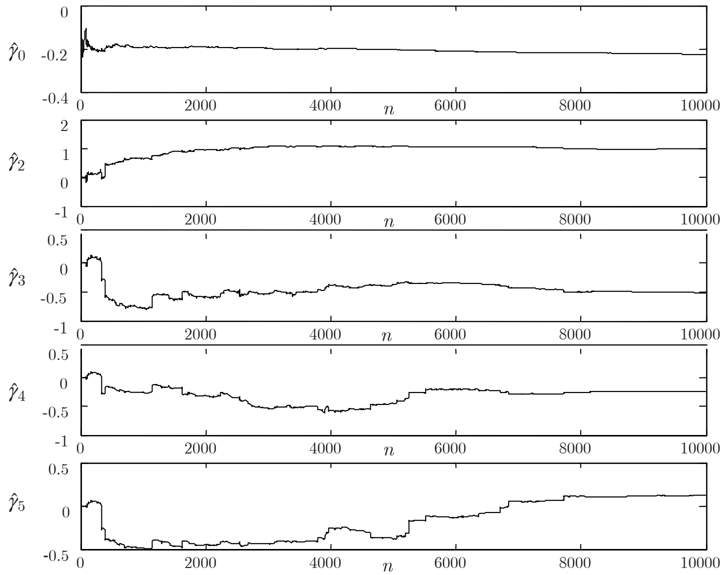

Fig. 4.14. Evolution of nonlinear element parameter estimates ($\alpha = 3.34 \times 10^{-3}$)

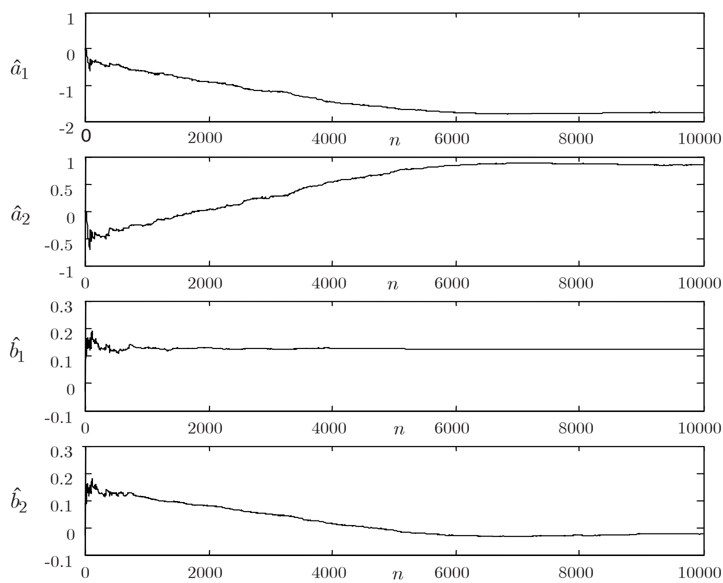

Fig. 4.15. Evolution of linear dynamic system parameter estimates ($\alpha = 3.34 \times 10^{-3}$)

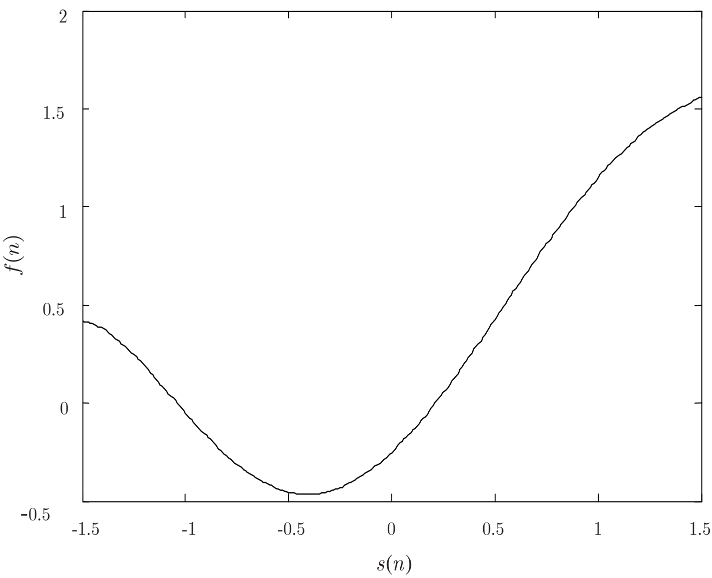

Fig. 4.16. Nonlinear element characteristic

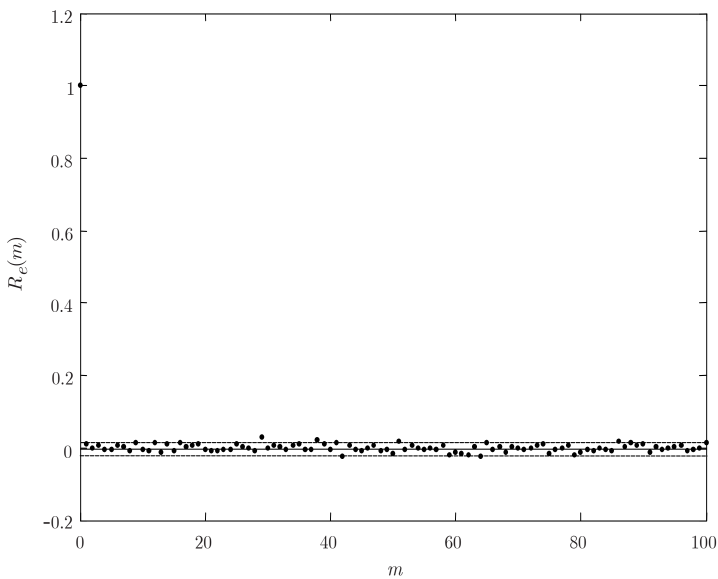

Fig. 4.17. Autocorrelation function of residuals and the 95% confidence interval ($\alpha = 3.3403 \times 10^{-3}$)

Table 4.7. Parameter estimates

Parameter α	True value	Estimated 3.34×10^{-5}	Estimated 3.34×10^{-3}	Estimated 3.34×10^{-1}
a_1	-1.7500	-1.7504	-1.7500	-1.7327
a_2	0.8500	0.8504	0.8501	0.8374
b_1	0.1250	0.1249	0.1242	0.1199
b_2	-0.0250	-0.0251	-0.0245	-0.0124
μ_0	-0.2500	-0.2499	-0.2500	-0.2454
μ_2	1.0000	1.0003	1.0053	1.0035
μ_3	-0.5000	-0.4992	-0.4913	-0.5493
μ_4	-0.2000	-0.2007	-0.2040	-0.2834
μ_5	0.1000	0.0991	0.0912	0.1493

Table 4.8. Comparison of estimation accuracy

Index	$\alpha = 3.34 \times 10^{-5}$	$\alpha = 3.34 \times 10^{-3}$	$\alpha = 3.34 \times 10^{-1}$
$\frac{1}{N}\sum_{i=1}^{N} e(n)^2$	3.3656×10^{-6}	3.3408×10^{-3}	3.3441×10^{-1}
$\sum_{j=0}^{5} (\hat{\mu}_j - \mu_j)^2$	2.3762×10^{-6}	1.9759×10^{-4}	1.2501×10^{-2}

4.4 Summary

In this chapter, it has been shown that the linear-in-parameters definition of the modified equation error makes it possible to use the linear regression approach to estimate the parameters of Wiener systems with an invertible nonlinear element. As such an approach results in inconsistent parameter estimates, a combined least squares-instrumental variables method has been proposed to overcome this problem. Contrary to linear regression approaches, prediction error methods make it possible to identify Wiener systems with both invertible and noninvertible nonlinear characteristics. Moreover, there is no problem of parameter redundancy, as the number of the estimated parameters is equal to the total number of the model parameters $na + nb + r + 1$. In comparison with the prediction error method, the pseudo-linear regression approach has lower computational requirements as it uses an approximate gradient instead of the true one.