Parameter Estimation of Wiener-Hammerstein Models*

Hosam E. EMARA-SHABAIK**, Mohammed S. AHMED**
and Khaled H. AL-AJMI**

Several algorithms, namely the Output Error (OE), the Equation Error (EE), the Prediction Error (PE), and the Instrumental Variable (IV), for parameter estimation of Wiener-Hammerstein models, are derived. Monte Carlo simulation results are provided showing the effectiveness, and comparing the performance of these algorithms.

Key Words: Wiener-Hammerstein Models, Parameter Estimation, Offline Methods

1. Introduction

Many physical systems exhibit nonlinear characteristics, and the assumption of linear behavior will only be valid locally around some operating conditions. As a result, quite often, the nonlinear effects must be considered in the model development of such systems. It is well known that model identification of nonlinear systems is quite involved as compared to linear model identification. In this regard the theory and practice of nonlinear system identification have not reached the same level of development as that of the linear systems identification.

A widely used model structure for nonlinear systems is the Wiener-Hammerstein model, Haber and Unbehauen⁽¹⁾. The Wiener-Hammerstein model, or the general model as sometimes referred to in the literature, is composed of a dynamic linear system in cascade with a static nonlinear element followed by another dynamic linear system. Figure 1 shows a system described by such a model where the static nonlinearity is represented by a finite degree

polynomial and the output is considered to be corrupted by correlated measurement noise.

One of the main advantages of this type of model is its representation by a finite number of parameters. Moreover, the model structure allows the extension of the estimation techniques available for linear systems.

The identification of Wiener-Hammerstein models, and their subclasses the Hammerstein and the Wiener models, have been a research subject of increasing importance which is being considered by many researchers. For example, Billings and Fakhouri have proposed in Ref.(2) a nonparametric identification of Wiener-Hammerstein model based on correlation analysis using inputs in the class of separable processes. Billings and Fakhouri also have considered in Ref.(3) the identification of the general

Fig. 1 A general Wiener-Hammerstein system with noisy measurements

* Received 12th October, 1999

** Systems Engineering Department, King Fahd University of Petroleum and Minerals (KFUPM), Dhahran 31261, Saudi Arabia. E-mail: shabaik@ccse.kfupm.edu.sa

model using compound input signals composed of uncorrelated, independent pseudo-random sequences with special properties. They have provided relations for estimating the impulse responses of the linear subsystems up to a constant multiplier using correlation methods. Following that step, they have applied least squares to estimate the coefficients of a polynomial representation of the static nonlinearity. Yoshine and Ishii⁽⁴⁾ have presented an identification technique of the general model based on Wiener kernel representation in a noise-free environment. Vandersteen et al.⁽⁵⁾ have presented a method for nonlinear estimation of the transfer functions of the two linear blocks in a Wiener-Hammerstein model using multi-sine excitation composed of large and small amplitudes. Considering the Wiener models, Kalafatis et al.⁽⁶⁾ have formulated the identification problem in terms of nonparametric frequency sampling filter (FSF) model. They have applied batch least squares, assuming noise-free conditions, to estimate the sampled discrete frequency response and the coefficients of a polynomial representing the inverse of the static nonlinearity.

In this paper, the identification of Wiener-Hammerstein models operating in noisy environment will be considered. For that purpose several parameter estimation algorithms are derived. The algorithms rely on batch processing of the available measured input/output data and produce offline parameter estimates. Offline methods are sometimes the only possibility for some applications, while for others they might be used to provide good initializations essential for the success of online techniques. The paper concludes with performance evaluation as well as comparative investigation, via simulation, of the developed algorithms.

2. Formulation of the Identification Problem

Consider a nonlinear system represented by the Wiener-Hammerstein model structure shown in Fig. 1 where the measured system output is corrupted by correlated noise. The mathematical description of the system model is given by:

$$A(q^{-1})y(t) = B(q^{-1})w(t) + F(q^{-1})e(t) \quad (1)$$

$$w(t) = g_1v(t) + g_2v^2(t) + \dots + g_mv^m(t) \quad (2)$$

$$C(q^{-1})v(t) = D(q^{-1})u(t) \quad (3)$$

Where, $A(q^{-1})$, $B(q^{-1})$, $C(q^{-1})$, $D(q^{-1})$, $F(q^{-1})$ are polynomials in the backward time shift operator (q^{-1}) with known orders ' n_a , n_b , n_c , n_d , n_f ' as follows.

$$\begin{aligned} A(q^{-1}) &= 1 + a_1q^{-1} + \dots + a_{n_a}q^{-n_a} \\ B(q^{-1}) &= 1 + b_1q^{-1} + \dots + b_{n_b}q^{-n_b} \\ C(q^{-1}) &= 1 + c_1q^{-1} + \dots + c_{n_c}q^{-n_c} \\ D(q^{-1}) &= 1 + d_1q^{-1} + \dots + d_{n_d}q^{-n_d} \\ F(q^{-1}) &= 1 + f_1q^{-1} + \dots + f_{n_f}q^{-n_f} \end{aligned} \quad (4)$$

The redundancy in the model parameterization is removed by considering $B(q^{-1})$, $D(q^{-1})$ as monic polynomials up to nonzero constant multipliers which are factored into the coefficients of the static nonlinearity in Eq.(2). Also, the order ' m ' of the polynomial representation of the static nonlinearity in Eq.(2) is known. The signals $y(t)$ and $u(t)$ are the system output and input at the time instant ' t '. These are the only accessible signals of the system. The noise signal $e(t)$ is an independent, identically distributed stationary white noise with zero mean. Furthermore, it is assumed that the linear subsystems are stable and time-invariant. The excitation signal $u(t)$ is considered to be stationary and persistently exciting of sufficient order. For nonlinear system identification, white Gaussian input, as well as uniformly distributed white input excitations⁽⁷⁾, or pseudo-random input signals of ' $m+1$ ' distinct levels⁽⁸⁾ will provide sufficient persistent excitation.

3. Offline Parameter Estimation Methods

The ($n_\theta \times 1$) vector of the unknown parameters of the Wiener-Hammerstein model is defined as,

$$\theta = (a_1 \dots a_{n_a} b_1 \dots b_{n_b} c_1 \dots c_{n_c} d_1 \dots d_{n_d} f_1 \dots f_{n_f} g_1 \dots g_m)^T$$

This unknown vector is to be estimated such that the following criteria function is minimized.

$$V_N(\theta) = \frac{1}{N} \sum_{t=1}^N \varepsilon^2(t, \theta) \quad (5)$$

where N is the number of available data point, and $\varepsilon(t, \theta)$ is an error, to be specified later on, which reflects the discrepancy between the actual system and the model. Let,

$$\phi(t, \theta) = - \left(\frac{\partial \varepsilon(t, \theta)}{\partial \theta} \right) \quad (6)$$

where $\phi(t, \theta)$ is of dimension $1 \times n_\theta$. The estimate θ must satisfy the following conditions,

$$V_N'(\theta) = - \frac{2}{N} \sum_{t=1}^N \varepsilon(t, \theta) \phi^T(t, \theta) = 0 \quad (7)$$

$$\begin{aligned} V_N''(\theta) &= \frac{2}{N} \sum_{t=1}^N \phi^T(t, \theta) \phi(t, \theta) \\ &- \frac{2}{N} \sum_{t=1}^N \varepsilon(t, \theta) \frac{\partial}{\partial \theta} \phi^T(t, \theta) \geq 0 \end{aligned} \quad (8)$$

Where, $V_N'(\theta)$, and $V_N''(\theta)$, are the first and the second derivatives of $V_N(\theta)$ with respect to θ . The above set of n_θ normal equations, given in Eq.(7), is nonlinear in the parameters. In practice such equations are to be solved iteratively. Applying the well-known Newton-Raphson iterations results in:

$$\theta^{(k+1)} = \theta^{(k)} - [V_N''(\theta^{(k)})]^{-1} [V_N'(\theta^{(k)})]^T \quad (9)$$

Here $\theta^{(k)}$ denotes the k th iteration parameters estimate. In order to insure the positive definiteness of the matrix $V_N''(\theta^{(k)})$ in Eq.(9), the second term on the right hand side of Eq.(8) is usually dropped resulting in the well-known Gauss-Newton iterations given by Eq.(10).

$$\theta^{(k+1)} = \theta^{(k)} + \alpha_k \left[\sum_{t=1}^N \psi^T(t, \theta^{(k)}) \psi(t, \theta^{(k)}) \right]^{-1} \times \left[\sum_{t=1}^N \varepsilon(t, \theta^{(k)}) \psi^T(t, \theta^{(k)}) \right] \quad (10)$$

The sequence of scalars α_k in Eq.(10) is used to control the step size and improve the convergence of the iterations.

The initial estimate $\theta^{(0)}$ will significantly influence the convergence of the iterations in Eq.(9) or (10). Therefore, any available a priori information must be considered in this regard. Convergence proofs of both of the above iterative schemes are given in Appendix A. To estimate the parameters of the Wiener-Hammerstein model, the iterations are applied in the context of the output errors (OE), the equation errors (EE), the prediction errors (PE), and the instrumental variables (IV) as given in the next sections.

3.1 The output error (OE) method

In the output error (OE) method the model parameters are estimated such that, the model output matches the actual measured system output as closely as possible in the sense of the criteria given by Eq. (5). Therefore,

$$\varepsilon(t, \theta) = e_{OE}(t, \theta)$$

Where the output error ' $e_{OE}(t, \theta)$ ' is generated as shown in Fig. 2. Formulas necessary for the implementation of this method are given in Appendix B.

3.2 The equation error (EE) method

The equation error algorithm for parameter estimation of the Wiener-Hammerstein model is obtained by minimizing the criteria function of Eq.(5) after making the following substitutions,

$$\varepsilon(t, \theta) = e_{EE}(t, \theta)$$

Where, the equation error ' $e_{EE}(t, \theta)$ ' is generated as shown in Fig. 3. The formulas necessary for implementing this algorithm are given in Appendix B.

Both the OE and the EE algorithms will produce biased estimate in the presence of non-white measurement noise. Such a situation can be treated either by estimating the noise dynamic model along with the system model in what is called the prediction error (PE) method, or alternatively by generating an output signal which is uncorrelated with the noise in an instrumental variable (IV) method.

3.3 The prediction error (PE) method

In the prediction error (PE) method both the system model parameters as well as the noise model parameters are to be estimated simultaneously. The noise model is incorporated to handle the situation of correlated measurement noise. The parameters estimates are generated in order to minimize the criteria of Eq.(5) after making the following substitution.

$$\varepsilon(t, \theta) = e_{PE}(t, \theta)$$

Where, the prediction error ' $e_{PE}(t, \theta)$ ' is generated as

Fig. 2 Output error method for the Wiener-Hammerstein model

Fig. 3 Equation error method for the Wiener-Hammerstein model

Fig. 4 Prediction error method for the Wiener-Hammerstein model in an EE context

shown in Fig. 4. The formulas necessary for the implementation of the method are given in Appendix B.

3.4 The instrumental variable (IV) method

The instrumental variables method for the identification of linear systems is as given in Ref.(9). Let $Z(t)$ be an $n_\theta \times 1$ vector generated such that it satisfies the following conditions:

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{t=1}^N Z(t) \psi_{EE}(t) \text{ be nonsingular}$$

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{t=1}^N e_{EE}(t) Z(t) = 0$$

Using such a vector, or instrumental variables, in the parameter estimation algorithm will remove the

bias due to the correlated noise. It is possible to generate the instrumental variables by passing the system-input signal through an auxiliary model whose output will be used to generate the vector $Z(t)$. In the case of nonlinear system identification, it is most appropriate to use a nonlinear auxiliary model. Therefore, the output of the auxiliary model can be generated in the same way as the model output used in the output error method of Section 3.1. Hence, the instrumental variable vector $Z(t)$ is taken as $\phi_{OE}^T(t)$ of the OE method. Therefore, the parameters estimate of the Wiener-Hammerstein model via an instrumental variable algorithm is obtained as follows,

$$\theta^{(k+1)} = \theta^{(k)} + \alpha_k \left[\sum_{t=1}^N \phi_{OE}^T(t, \theta^{(k)}) \phi_{EE}(t, \theta^{(k)}) \right]^{-1} \times \left[\sum_{t=1}^N e_{EE}(t, \theta^{(k)}) \phi_{OE}^T(t, \theta^{(k)}) \right]$$

4. Results

Dynamic systems which can be decomposed into linear dynamic subsystems interconnected through static nonlinear couplings are most suitable candidates to be represented by Wiener-Hammerstein models. Most often linear dynamic systems have finite, second or third, order dynamics. Also, static nonlinear couplings of up to a third order are common place in practice. Therefore, these general guidelines will be followed in designing cases to test the performance of the identification methods derived in Section 3. Two cases are considered in the numerical simulation. In the first case, up to second order dynamics and cubic static nonlinearity is considered. While in the second case up to third order dynamics with cubic nonlinearity is considered. The system input is taken as a white Gaussian sequence with zero mean and variance 0.5. The noise term $e(t)$ is taken as an independent and identically distributed zero-mean Gaussian sequence. The noise-to-signal ratio (NSR) is taken at the two levels of 10% and 50%. The measurements noise are generated as the output of an ARMA model. In case I, an ARMA(2, 1) model is used, while in case II an ARMA(2, 2) is used. Both will generate a correlated measurement noise to test the estimation algorithms under realistic noise conditions. Therefore, ten (10) parameters in the first case, and twelve (12) parameters in the second case will be estimated using the estimation algorithms developed in this paper. The true parameters of each case are as given below :

Case 1 :

$$\begin{aligned} \theta &= [a_1, a_2, b_1, c_1, c_2, d_1, f_1, g_1, g_2, g_3]^T \\ &= [-0.4, 0.25, 0.3, 0.5, -0.25, 0.7, 0.6, \\ &\quad 0.5, -1.8, 2.5]^T \end{aligned}$$

Table 1 Performance summary of the offline parameter estimation algorithms, case 1

Input Signal	NSR	Algorithm	MNE	ASD	AIFC
Gaussian input	10%	OE	0.1508	0.06036	20
		PE	0.0112	0.00237	10
		EE	0.1204	0.00307	10
		IV	0.0686	0.05298	16
	50%	OE	0.3047	0.26667	23
		PE	0.0413	0.00819	10
		EE	0.1970	0.06158	9
		IV	0.1119	0.13884	13

Table 2 Performance summary of the offline parameter estimation algorithms, case 2

Input Signal	NSR	Algorithm	MNE	ASD	AIFC
Gaussian input	10%	OE	0.1064	0.07077	42
		PE	0.0042	0.00382	9
		EE	0.1492	0.01028	13
		IV	0.0289	0.00469	26
	50%	OE	0.1288	0.04283	22
		PE	0.0600	0.04000	11
		EE	0.2230	0.08601	9
		IV	0.1129	0.14792	19

Case 2 :

$$\begin{aligned} \theta &= [a_1, a_2, b_1, c_1, c_2, c_3, d_1, d_2, f_1, f_2, g_1, g_3]^T \\ &= [-0.4, 0.25, 0.3, -0.6, -0.25, 0.15, \\ &\quad -0.55, -0.2, 0.6, -0.1, 1.0, 2.5]^T \end{aligned}$$

Data records of 1000 points for each case are generated. The initial parameters estimate $\theta^{(0)}$ is taken near the zero values, while the scalar ' α ' in the algorithm is set equal to 1. Monte Carlo simulations of 30 different runs in each case are used. In order to facilitate the comparison of the different estimation methods three different measures are used. The mean normalized error (MNE), as defined below, is computed for each method.

$$MNE = \frac{\|\theta_{\text{mean}} - \theta\|^2}{\|\theta\|^2}$$

Where θ_{mean} is the mean value of the estimates, at convergence, over the 30 runs. While the MNE is a composite measure of the bias in the parameter estimates, the average standard deviation (ASD) is used as a measure of their variance. The ASD is defined as below.

$$ASD = \frac{\sum_{i=1}^{n_\theta} \sigma_i}{n_\theta}$$

Where σ_i is the i th parameter standard deviation using the 30 Monte Carlo simulation runs. A third measure, the average number of iterations for convergence (AIFC), is also computed and used for comparison. Convergence is considered to be attained whenever the relative change in the parameters estimates is less than 2%. Tables 1 and 2 give the results in terms of the three measures for the different estimation methods derived in this paper. The Euclidean error norms of the parameter estimates of the four methods versus the number of iterations are provided in Figs. 5 and 6.

Fig. 5 Error Norm for different offline algorithms for case 1 using 10% NSR

Fig. 6 Error Norm for different offline algorithms for case 2 using 10% NSR

5. Discussions

The results in Table 1 and Table 2 and also those provided by Figs. 5 and 6 indicate the superiority of the PE algorithm. The PE method provides estimates with the smallest bias and variance among all the developed algorithms for different NSR. On the other hand, the equation error and the output error algorithms perform the least. The improved performance of the PE algorithm is due to the fact that it accounts for the dynamics of the noise in the output, a feature that is lacking in the OE and EE algorithms. For all the cases the instrumental variable algorithm provides better estimates than either the OE or the EE algorithms. This is the case since the IV has been developed to overcome the bias problem due to the correlated noise. This also, points to the effectiveness of the scheme used for generating the instrumental variables. However, the IV algorithm requires more computations than either the EE or the OE algorithms.

6. Conclusions

In this paper, four offline algorithms for the parameter estimation of the Wiener-Hammerstein models are derived. These are the Output Error (OE), Equation Error (EE), Prediction Error (PE), and Instrumental Variables (IV) algorithms. Monte Carlo simulations are performed. Results of the simulation indicate the effectiveness of the developed algorithms. Comparison of the simulation results indicates the superiority of the PE algorithm followed by the IV algorithm. The development of recursive parameter estimation techniques for the class of nonlinear

models considered in this paper is underway.

Acknowledgement

The authors wish to acknowledge the support provided by King Fahd University of Petroleum and Minerals.

Appendix A : Convergence Analysis

By ergodicity of $\varepsilon(t, \theta)$, one should note that $\lim_{N \rightarrow \infty} V_N(\theta) = V_\infty(\theta) = E\{\varepsilon^2(t, \theta)\}$

Further, the following assumptions are made in addition to the ones given in Section 2,

1. The second and third derivatives of $V_\infty(\theta)$, with respect to θ , are strictly finitely bounded and $V_\infty''(\theta)$ is of full rank.

2. $\Lambda(\theta^{(k)}) = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{t=1}^N \phi^T(t, \theta^{(k)}) \phi(t, \theta^{(k)})$ and $\Lambda'(\theta^{(k)})$ are strictly finitely bounded and $\Lambda(\theta^{(k)})$ is of full rank.

Now, consider the Newton-Raphson iteration, $\theta^{(k+1)} = \theta^{(k)} - [V_\infty''(\theta^{(k)})]^{-1} [V_\infty'(\theta^{(k)})]^T = g(\theta^{(k)})$

The iteration has a stationary point θ_0 for which $[V_\infty''(\theta_0)]^{-1} [V_\infty'(\theta_0)]^T = 0$

Since $V_\infty''(\theta_0)$ is assumed to have finite bound and is of full rank, then

$$V_\infty'(\theta_0) = 0$$

i.e. the stationary point satisfies the necessary condition for a minimum of $V_\infty(\theta)$. Note that at the stationary point θ_0 , the following is true,

$$g(\theta_0) = \theta_0 \text{ and } g'(\theta_0) = 0$$

Assuming $g''(\theta)$ is continuous and strictly bounded by 'M' on an open interval which contains θ_0 . Then

$$g(\theta^{(k)}) = g(\theta_0) + g'(\theta_0)(\theta^{(k)} - \theta_0)$$

$$\begin{aligned}
& + \frac{1}{2} (\theta^{(k)} - \theta_0)^T g''(\xi) (\theta^{(k)} - \theta_0) \\
& = \theta_0 + \frac{1}{2} (\theta^{(k)} - \theta_0)^T g''(\xi) (\theta^{(k)} - \theta_0), \\
& \xi \in (\theta_0, \theta^{(k)})
\end{aligned}$$

i.e.

$$\theta^{(k+1)} = \theta_0 + \frac{1}{2} (\theta^{(k)} - \theta_0)^T g''(\xi) (\theta^{(k)} - \theta_0)$$

Therefore,

$$\|\theta^{(k+1)} - \theta_0\| < \frac{1}{2} \|M\| \|\theta^{(k)} - \theta_0\|^2$$

And the iterations converge quadratically to θ_0 .

Next, consider the Gauss-Newton iterations,

$$\theta^{(k+1)} = \theta^{(k)} + \alpha_k \Lambda^{-1}(\theta^{(k)}) V_{\infty}'(\theta^{(k)}) = f(\theta^{(k)})$$

The iteration has a stationary point θ_0 for which

$$\Lambda^{-1}(\theta_0) V_{\infty}'(\theta_0) = 0$$

Since $\Lambda(\theta)$ is assumed to have a finite bound and is of full rank, then

$$V_{\infty}'(\theta_0) = 0$$

i.e. the stationary point satisfies the necessary condition for a minimum of $V_{\infty}(\theta)$. Note that at the stationary point θ_0 , the following is true in this case,

$$f(\theta_0) = \theta_0 \text{ and } f'(\theta_0) \neq 0$$

Assuming $f'(\theta)$ to be continuous and strictly bounded by 'D' on an open interval containing θ_0 , then

$$f(\theta^{(k)}) = f(\theta_0) + f'(\delta)(\theta^{(k)} - \theta_0),$$

$$\delta \in (\theta_0, \theta^{(k)})$$

i.e.

$$\theta^{(k+1)} = \theta_0 + f'(\delta)(\theta^{(k)} - \theta_0)$$

Therefore,

$$\|\theta^{(k+1)} - \theta_0\| < \|D\| \|\theta^{(k)} - \theta_0\|$$

And the iterations converge linearly to θ_0 . It is to be noticed that the step size α_k can be used to speed up the rate of convergence. $\square$

Appendix B : Implementation Formulas

Important formulas necessary for the implementation of the algorithms derived in this paper are given in the following table.

Formulas of the OE method	Formulas of the EE method	Formulas of the PE method
$A(q^{-1}) \frac{\partial y(t)}{\partial a_i} = -y(t-i)$	$\frac{\partial e_{EE}(t)}{\partial a_i} = y(t-i)$	$F(q^{-1}) \frac{\partial e_{PE}(t)}{\partial a_i} = y(t-i)$
$A(q^{-1}) \frac{\partial y(t)}{\partial b_i} = w(t-i)$	$\frac{\partial e_{EE}(t)}{\partial b_i} = -w(t-i)$	$F(q^{-1}) \frac{\partial e_{PE}(t)}{\partial b_i} = -w(t-i)$
$A(q^{-1}) \frac{\partial y(t)}{\partial c_i} = B(q^{-1}) \frac{\partial w(t)}{\partial c_i}$	$\frac{\partial e_{EE}(t)}{\partial c_i} = -B(q^{-1}) \frac{\partial w(t)}{\partial c_i}$	$F(q^{-1}) \frac{\partial e_{PE}(t)}{\partial c_i} = -B(q^{-1}) \frac{\partial w(t)}{\partial c_i}$
$\frac{\partial w(t)}{\partial c_i} = \sum_{l=1}^m l \times g_l v^{l-1}(t) \frac{\partial v(t)}{\partial c_i}$	$\frac{\partial w(t)}{\partial c_i} = \sum_{l=1}^m l \times g_l v^{l-1}(t) \frac{\partial v(t)}{\partial c_i}$	$\frac{\partial w(t)}{\partial c_i} = \sum_{l=1}^m l \times g_l v^{l-1}(t) \frac{\partial v(t)}{\partial c_i}$
$C(q^{-1}) \frac{\partial v(t)}{\partial c_i} = -v(t-i)$	$C(q^{-1}) \frac{\partial v(t)}{\partial c_i} = -v(t-i)$	$C(q^{-1}) \frac{\partial v(t)}{\partial c_i} = -v(t-i)$
$A(q^{-1}) \frac{\partial y(t)}{\partial d_i} = B(q^{-1}) \frac{\partial w(t)}{\partial d_i}$	$\frac{\partial e_{EE}(t)}{\partial d_i} = -B(q^{-1}) \frac{\partial w(t)}{\partial d_i}$	$F(q^{-1}) \frac{\partial e_{PE}(t)}{\partial d_i} = -B(q^{-1}) \frac{\partial w(t)}{\partial d_i}$
$\frac{\partial w(t)}{\partial d_i} = \sum_{l=1}^m l \times g_l v^{l-1}(t) \frac{\partial v(t)}{\partial d_i}$	$\frac{\partial w(t)}{\partial d_i} = \sum_{l=1}^m l \times g_l v^{l-1}(t) \frac{\partial v(t)}{\partial d_i}$	$\frac{\partial w(t)}{\partial d_i} = \sum_{l=1}^m l \times g_l v^{l-1}(t) \frac{\partial v(t)}{\partial d_i}$
$C(q^{-1}) \frac{\partial v(t)}{\partial d_i} = u(t-i)$	$C(q^{-1}) \frac{\partial v(t)}{\partial d_i} = u(t-i)$	$C(q^{-1}) \frac{\partial v(t)}{\partial d_i} = u(t-i)$
$A(q^{-1}) \frac{\partial y(t)}{\partial g_i} = B(q^{-1}) \frac{\partial w(t)}{\partial g_i}$	$\frac{\partial e_{EE}(t)}{\partial g_i} = -B(q^{-1}) \frac{\partial w(t)}{\partial g_i}$	$F(q^{-1}) \frac{\partial e_{PE}(t)}{\partial g_i} = -B(q^{-1}) \frac{\partial w(t)}{\partial g_i}$
$\frac{\partial w(t)}{\partial g_i} = v^i(t)$	$\frac{\partial w(t)}{\partial g_i} = v^i(t)$	$\frac{\partial w(t)}{\partial g_i} = v^i(t)$

References

- (1) Haber, R. and H. Unbehauen, Structure Identification of Nonlinear Dynamic Systems - A Survey on Input/Output Approaches, *Automatica*, Vol. 26, No. 4 (1990), pp. 651-677.
 - (2) Billings, S.A. and Fakhouri, S.Y., Identification of a Class of Nonlinear Systems Using Correlation Analysis, *IEE Part D*, Vol. 125 (1978), pp. 691-697.
 - (3) Billings, S.A. and Fakhouri, S.Y., Identification of Nonlinear Systems Using Correlation Analysis and Pseudorandom Inputs, *International Journal of Systems Science*, Vol. 11, No. 3 (1980), pp. 261-279.
 - (4) Yoshine, K. and Ishii, N., Nonlinear Analysis of a Linear-Non-Linear-Linear System, *International Journal of Systems Science*, Vol. 23, No. 4 (1992), pp. 623-630.
 - (5) Vandersteen, G., Rolain, Y. and Schoukens, J., Non-Parametric Estimation of The Frequency-Response Functions of The Linear Blocks of a Wiener-Hammerstein Model, *Automatica*, Vol. 33, No. 7 (1997), pp. 1351-1355.
 - (6) Kalafatis, A., Wang, L. and Cluett, W., Identification of Wiener-Type Nonlinear Systems in a Noisy Environment, *International Journal of Control*, Vol. 66, No. 6 (1997), pp. 923-941.
 - (7) Leontaritis, I.J. and Billings, S.A., Experimental Design and Identifiability for Nonlinear Systems, *International Journal of Systems Science*, Vol. 18, No. 1 (1987), pp. 189-202.
 - (8) Nowak, R. and Van Veen, B., Nonlinear System Identification with Pseudorandom Multilevel Excitation Sequences, *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Vol. 4 (1993), pp. 456-459.
 - (9) Soderstrom, T. and Stoica, P., *System Identification*, (1989), Prentice Hall.
-