Science**PG**
Science Publishing Group

Research/Technical Note

Identification of Nonlinear Model with General Disturbances

Wang Xiaoping[1], Yao Jie[1], Wang Jianhong[2], Liu Feifei[2]

[1]School of Mechanical and Electronic Engineering, Jingdezhen Ceramic Institute, Jingdezhen, China

[2]School of Electronic Engineering and Automation, Jiangxi University of Science and Technology, Ganzhou, China

Email address:
baixuexue3624@163.com (Yao Jie)

Abstract: The nonlinear model has a linear dynamic system following some static nonlinearity. The dominating approach to estimate the components of this model has been to minimize the error between the simulated and the measured outputs. For the special case of Gaussian input signals, we estimate the linear part of the Hammerstein model using the Bussgang's classic theorem. For the case with general disturbances, we derive the Maximum Likelihood method. Finally one simulation example is used to prove the efficiency of our theory.

Keywords: System Identification, Nonlinear System, Bussgang's Theorem, Maximum Likelihood Method

1. Introduction

Many nonlinear systems can be modeled by a Hammerstein model (linear time-invariant (LTI) block following some static nonlinear block), Wiener model (LTI block preceding some static nonlinear block), or Hammerstein-Wiener model (LTI block sandwiched by two nonlinear blocks).

The Hammerstein model is a special kind of nonlinear systems which has applications in many engineering problems and therefore, identification of Hammerstein models has been an active research topic for a long time. Existing methods in the literature can be roughly divided into six categories: the iterative method, the over-parameterization method, the stochastic method, the nonlinear least squares method, the separable least squares method and the blind method.

The paper focuses on the identification of Hammerstein models shown in Figure 1 which consist of a nonlinear memory less element followed by a linear dynamical system. The input signal is denoted by $u(t)$, the output signal by $y(t)$ and $x(t)$ denotes the intermediate, immeasurable signal. We will call $w(t)$ process noise and $e(t)$ measure noise, and assume that they are independent. Note that since G is a linear system, the process noise can equally well be applied anywhere after the nonlinearity with an additional filter.
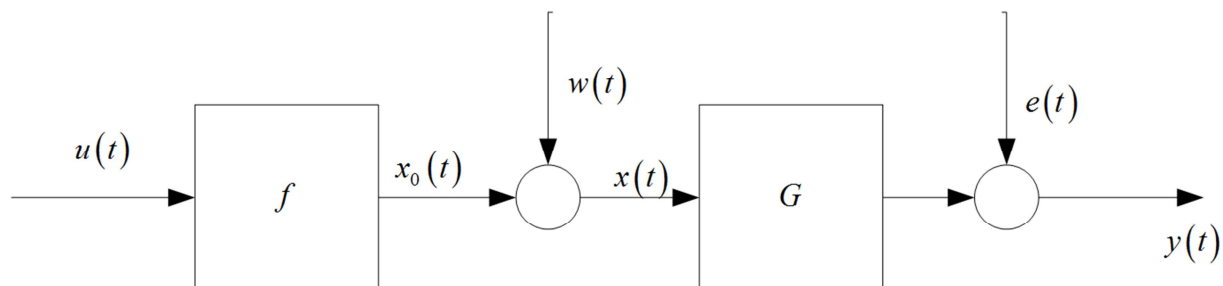

Figure 1. Hammerstein model.

The Hammerstein models can be described by the following equation:

$$\begin{cases} y(t) = G(q,\theta)x(t) + e(t) \\ x(t) = x_0(t) + w(t) \\ x_0(t) = f(u(t),\eta) \end{cases} \quad (1)$$

This paper will focuses on parametric models. We will assume f and G each belongs to a parameterized models class. Examples of such a model class may be polynomials, splines, or neural network for the nonlinear function f -in general a basis function expansion. The nonlinear f may also be a piecewise linear function, like a saturation or a dead-zone. Common model classes for G are FIR filters, rational transfer functions (OE models) or state space models, but also for example Laguerre filters may be used.

If the process noise w and the intermediate signal x are unknown, the parameterization of the Hammerstein models is not unique. Notice that in the characterization of the Hammerstein model shown in Figure 1. f and G are actually not unique. Any pair $(kf(u), G(z)/k)$ for some nonzero and finite constant k would produce identical input and output measurements. In other words, any identification scheme cannot distinguish between

$(kf(u), G(z)/k)$ and $(f(u), G(z))$. Therefore to get a unique parameterization, without loss of generality, one of the gains of $f(u)$ and $G(z)$ has to be fixed. (We may also need to scale the process noise variance with a factor k .)

Given input and output data, and model classes for f and G, we want to find the parameters θ and η that best match the data, measured as input u and output y from the system.

An estimate from a certain estimation method is said to be consistent if the parameters converge to their true values, when the number of data N tends to infinity.

To investigate the minimum of the approximative PEM criterion (2) we write the true system as

$$y(t) = G(q,\theta_0)f(u(t),\eta_0) + \tilde{w}(t) + e(t) \quad (4)$$

where

$$\tilde{w}(t) = G(q,\theta_0)w(t), \ E\tilde{w}(t) = G(q,\theta_0)Ew(t) = 0$$

$$E\tilde{w}^2(t) = G(q,\theta_0)Ew^2(t)G^*(q,\theta_0) = |G(q,\theta_0)|^2 \lambda_w \quad (5)$$

2. A Standard Method and Possible Bias Problems

Several different methods to identify Hammerstein models have been suggested in the literature. A common approach is to parameterize the linear and nonlinear block, and to estimate the parameters from data, by minimizing an error criterion.

If the process noise $w(t)$ in Figure 1 is disregarded or zero, a natural criterion is to minimize

$$V_N(\theta,\eta) = \frac{1}{N}\sum_{t=1}^{N}\left(y(t) - G(q,\theta)f(u(t),\eta)\right)^2 \quad (2)$$

This is a standard approach and has been used in several papers. If the process noise is indeed zero, this is the prediction error criterion. If the measurement noise is white and Gaussian, (2) is also the Maximum Likelihood Criterion and the estimate is the consistent.

While measurement noise e is discussed in several papers, few consider process noise w. Reference [10] is one exception where both the input and output are subject to noise. Consistency of the estimate method is, however, not discussed in that paper. It may seem reasonable to use an error criterion like (2) even in the case where there is process noise.

However $G(q,\theta)f(u(t),\eta)$ is not the true predictor in this case. We will name this method the approximate prediction error method, and we will show that the estimate obtained this way is not necessarily consistent.

Suppose that the true system can be described within the model class, i.e. there exist parameters (θ_0,η_0) such that.

$$y(t) = G(q,\theta_0)\left[f(u(t),\eta_0) + w(t)\right] + e(t) = G(q,\theta_0)f(u(t),\eta_0) + G(q,\theta_0)w(t) + e(t) \quad (3)$$

$$Ew^2(t) = \lambda_w, Ee^2(t) = \lambda_e \quad (6)$$

We may regard $\tilde{w}(t)$ as a (input-dependent) transformation of the process noise to the output. Stochastic properties, such as mean and variance of the process noise will typically be preserved in the transformation from $w(t)$ to $\tilde{w}(t)$.

Now insert the expression for y in Eq. (4) into the criterion (2):

$$\begin{aligned} V_N(\theta,\eta) &= \frac{1}{N}\sum_{t=1}^{N}\left[G(q,\theta_0)f(u(t),\eta_0) - G(q,\theta)f(u(t),\eta) + \tilde{w}(t) + e(t)\right]^2 \\ &= \frac{1}{N}\sum_{t=1}^{N}[G(q,\theta_0)f(u(t),\eta_0) - G(q,\theta)f(u(t),\eta)]^2 \\ &\quad + \frac{1}{N}\sum_{t=1}^{N}\left[\tilde{w}(t) + e(t)\right]^2 + \frac{2}{N}\sum_{t=1}^{N}[G(q,\theta_0)f(u(t),\eta_0) - G(q,\theta)f(u(t),\eta)]\left[\tilde{w}(t) + e(t)\right] \end{aligned} \quad (7)$$

Now assume all noises are ergodic, so that time averages tend to their mathematic expectations as N tends to infinity. Assume also that u is a (quasi)-stationary sequence, so that it also has well defined sample averages. Let E denote both mathematical expectation and averaging over time signals.

Using the fact that the measurement noise e is zero mean, and independent of the input u and the process noise w, means that several cross terms will disappear. The criterion then tends to

$$V_N(\theta,\eta) = \frac{1}{N}\sum_{t=1}^{N}\left[G(q,\theta_0)f(u(t),\eta_0) - G(q,\theta)f(u(t),\eta) + \tilde{w}(t) + e(t)\right]^2$$

$$= \frac{1}{N}\sum_{t=1}^{N}[G(q,\theta_0)f(u(t),\eta_0) - G(q,\theta)f(u(t),\eta)]^2 \qquad (8)$$

$$+ \frac{1}{N}\sum_{t=1}^{N}\left[\tilde{w}(t) + e(t)\right]^2 + \frac{2}{N}\sum_{t=1}^{N}[G(q,\theta_0)f(u(t),\eta_0) - G(q,\theta)f(u(t),\eta)]\left[\tilde{w}(t) + e(t)\right]$$

Note that the criterion has a quadratic form, and the true value (θ,η) will minimize the criterion

Using the partial derivation about the $\bar{V}_N(\theta,\eta)$, we can obtain the estimate values (θ,η)

$$\begin{cases} \dfrac{\partial \bar{V}_N(\theta,\eta)}{\partial \theta} = 0 \\ \dfrac{\partial \bar{V}_N(\theta,\eta)}{\partial \eta} = 0 \end{cases} \qquad (9)$$

3. Bussgang's Theorem and Its Implication for Hammerstein Models

We generalize the following Bussgang's theorem

Theorem 1. Let $y(t)$ be the stationary output from a static nonlinearity f with a stationary Gaussian input $u(t)$, i.e, $y(t) = f(u(t))$.

Assume that the expectations $E(y(t)) = E(u(t)) = 0$.

Then

$$R_{yu}(\tau) = b_0 R_u(\tau), \forall \tau \in \mathbb{Z} \qquad (10)$$

where

$$R_{yu}(\tau) = E(y(t)u(t-\tau)), R_u(\tau) = E(u(t-\tau)u(t)), b_0 = E(f'(u(t))),$$

Proof: Set

$$p(u) = \frac{1}{\sqrt{2\pi}}e^{-\frac{u^2}{2}}, \sigma_u^2 = \int_{-\infty}^{\infty}u^2 p(u)du = 1,$$

$$\bar{u} = \int_{-\infty}^{\infty}up(u)du = 0, \quad r = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}u_1 u_2 p(u_1,u_2,\tau)du_1 du_2$$

Using some knowledge of probability, we have

$$p(u_1,u_2,\tau) = \frac{1}{2\pi|B|^{\frac{1}{2}}}\exp\left\{-\frac{1}{2}(u_1,u_2)B^{-1}(u_1,u_2)^T\right\} \qquad (11)$$

where

$$B = \begin{pmatrix} 1 & r \\ r & 1 \end{pmatrix}, \qquad B^{-1} = \frac{1}{1-r^2}\begin{pmatrix} 1 & -r \\ -r & 1 \end{pmatrix}$$

Substituting B and B^{-1} into the right part of the $p(u_1,u_2,\tau)$

We can obtain

$$p(u_1,u_2,\tau) = \frac{1}{2\pi\sqrt{1-r^2}}e^{-\frac{u_1^2+u_2^2-2ru_1u_2}{2(1-r^2)}} \qquad (12)$$

Then

$$R_{yu}(\tau) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}u_1 f(u_2)p(u_1,u_2,\tau)du_1 du_2 = \frac{1}{2\pi\sqrt{1-r^2}}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}u_1 f(u_2)e^{-\frac{u_1^2+u_2^2-2ru_1u_2}{2(1-r^2)}}du_1 du_2 \qquad (13)$$

Applying the equality

$$\int_{-\infty}^{\infty} u_1 e^{-\frac{u_1^2}{2(1-r^2)}+\frac{ru_2}{1-r^2}u_1} du_1 = r\sqrt{2\pi(1-r^2)} u_2 e^{\frac{r^2 u_2^2}{2(1-r^2)}} \quad (14)$$

We insert (13) into (12) then

$$R_{yu}(\tau) = \frac{r}{\sqrt{2\pi}} \int_{-\infty}^{\infty} u_2 f(u_2) e^{-\frac{u_2^2}{2}} du_2 \quad (15)$$

Since we are dealing with stationary processes, the ensemble average and the time average can be equated, that is $r = R_u(\tau)$ and we assume

$$b_0 = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} u f(u) e^{-\frac{u^2}{2}} du$$

Then

$$R_{yu}(\tau) = b_0 R_u(\tau), \forall \tau \in \mathbb{Z}$$

To simplify b_0, we adopt the following fact that

$$E(h(z)z_i) = E(h'_{z_i}(z))$$

when z has a $N(0,I)$ distribution. This equality holes since

$$b_0 = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} u f(u) e^{-\frac{u^2}{2}} du = \frac{1}{\sqrt{2\pi}} \left[-f(u) e^{-\frac{u^2}{2}} \right]_{u=-\infty}^{\infty} + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f'(u) e^{-\frac{u^2}{2}} du$$

$$= 0 + E(f'(u)) = E(f'(u)) \quad (16)$$

The theorem 1 has thus been shown.

Let ϕ_{yu} and ϕ_u denote the z-transforms of R_{yu} and R_u, respectively. Provided that these transforms are well-defined, (10) can also be written as $\phi_{yu}(\omega) = b_0 \phi_u(\omega)$.

Bussgang's theorem has turned out to be very useful for the theory of Hammerstein and Wiener system identification. The reason for this is that Bussgang's theorem explains why it is possible to estimate the linear and nonlinear parts of a Hammerstein system separately when the input is Gaussian. It can be used to obtain a good estimate of the linear part of the model. It is interesting to note that the result applies also to our more general situation with process noise w.

Theorem 2. Consider the model structure defined by Figure 1. Assume that the input $u(t)$ and the process noise $w(t)$ are independent, Gaussian, stationary processes (not necessarily white). Assume that the measurement noise $e(t)$ is a stationary stochastic process, independent of u and w. It is

however not assumed that e is neither white nor Gaussian. Let $G(q,\theta)$ be an arbitrary transfer function parameterization with freely adjustable gain, such that $G(q,\theta_0) = G_0(q)$ (the true linear part of the system) for some parameter value θ_0. Let θ be estimated from u and y using an output error method, neglecting any possible of nonlinearity:

$$\hat{\theta}_N = \arg \min_{\theta} \sum_{t=1}^{N} (y(t) - G(q,\theta)u(t))^2 \quad (17)$$

Then

$$G(q,\hat{\theta}_N) \to b_0 G_0(q), \quad as \quad N \to \infty$$

Proof: define

$$x_0(t) = f(u(t)), \quad x(t) = x_0(t) + w(t), \quad y(t) = G_0(q)x(t) + e(t) \quad (18)$$

Then since e is independent of u the cross spectra between u and y will be: $\phi_{yu} = G_0(e^{iw})\phi_{xu}$, since u and w are Gaussian, then $x(t)$ is Gaussian, so Bussgang's theorem tells us that $\phi_{xu} = \phi_{x_0 u} = b_0 \phi_u$. Since $x_0(t) = f(u(t))$, the resulting conclusion is that

$$\phi_{yu}(\omega) = G_0(e^{iw}) b_0 \phi_u(\omega) \quad (19)$$

Now it is well known (see e.g. chapter 8 in Ljung (1999)) that $\hat{\theta}_N$ will converge to a value that minimizes

$$V(\theta) = E(y(t) - G(q,\theta)u(t))^2 = \frac{1}{2\pi} \int \left(\phi_y(w) - 2\mathrm{Re}\left[G(e^{-iw},\theta)\phi_{yu}(w) \right] + \left| G(e^{iw},\theta) \right|^2 \phi_u(w) \right) dw$$

Insert (19) into this expression, and replace the θ-independent term $\phi_y(w)$ with the θ-independent term $b_0^2 \left| G_0(e^{iw}) \right|^2 \phi_u(\omega)$. Minimizing $V(\theta)$ is thus the same as minimizing

$$W(\theta) = \int \left(b_0^2 \left| G_0(e^{iw}) \right|^2 \phi_u(\omega) - 2\operatorname{Re}\left[b_0 G(e^{-iw},\theta) G_0(e^{iw}) \phi_u(w) \right] + \left| G(e^{iw},\theta) \right| \phi_u(w) \right) dw$$

$$= \int \left| b_0 G_0(e^{iw}) - G(e^{iw},\theta) \right|^2 \phi_u(\omega) d\omega$$

which prove the theorem.

The theorem is a consequence of the fact that the best linear system approximation that relates u and y is proportional to the linear part G_0 of the true system.

Basically this means that an estimate of the linear system $G(q)$ will be consistent for many other common linear identification methods. Note that the gain of G cannot be estimated anyway, since a gain factor can be moved between G and f without affecting the input-output behavior.

4. An Example

Consider a generalized Hammerstein system $y(t) = G(q,\theta) f(u(t), u(t-1)) + w(t)$
where

$$G(q) = \frac{1}{1 + 0.6q^{-1} + 0.1q^{-2}} \qquad f(u(t), u(t-1)) = \arctan(u(t)) \cdot u(t-1)^2$$

And where $w(t)$ is white Gaussian noise with $E(w(t)) = 0, E(w(t)^2) = 1$

Let the input $u(t)$ be generated by linear filtering of a white Gaussian process $e(t)$ with

$$E(e(t)) = 0, E(e(t)^2) = 1$$

such that

$$u(t) = \frac{1 - 0.8q^{-1} + 0.1q^{-2}}{1 - 0.2q^{-1}} e(t).$$

And assume that $e(t)$ and $w(s)$ are independent for all t, s

This input signal has been used in an identification experiment where a data set consisting of 100000 measurements of $u(t)$ and $y(t)$ was collected. A linear output error model has been estimated from this data set and the result was:

$$\hat{G}_{OE}(q) = \frac{0.762 - 0.682q^{-1}}{1 + 0.613q^{-1} + 0.102q^{-2}}$$

As can be seen, the denominator of $\hat{G}_{OE}(q)$ is indeed close to the denominator of $G(q)$.

In Figure. 2, we compare the two output responses of the real system and the identification system. In Figure. 3, we draw the error curve between the two output responses. From these Figures, we can see that with the number of measured data pairs tend to infinity, the identification system can suit the real system and the error will tend to zero.
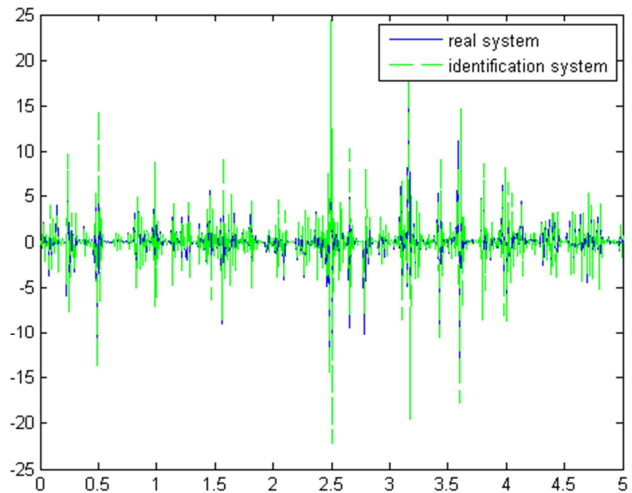

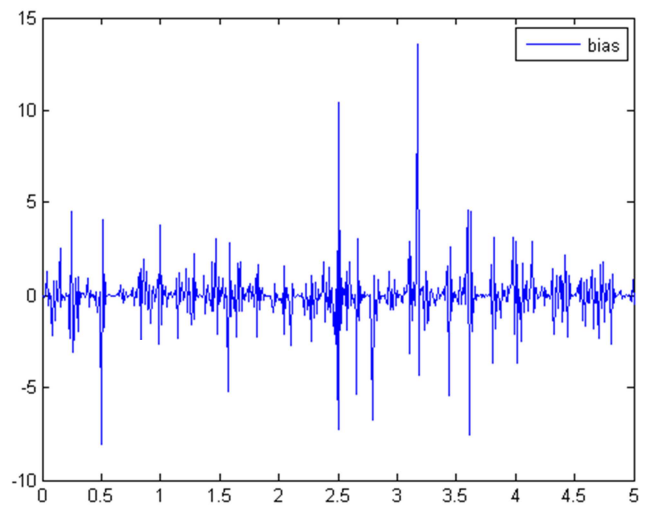Figure 2. The output responses of the real system and the identification system.

Figure 3. The error between the two output responses.

5. Maximum Likelihood Estimation

5.1. Derivation of the Likelihood Function for White Disturbances

The likelihood function is the probability density function (PDF) of the outputs $y^N = \{y(1), y(2) \cdots y(N)\}$ for given parameters θ and η. We shall also assume that the input sequence $u^N = \{u(1), u(2) \cdots u(N)\}$ is a given, deterministic sequence.(We condition the PDF wrt to this sequence, if it is described as a stochastic process.) Let $p_{y^N}(\theta, \eta, u^N)$ denote this PDF. For an observed data set y_*^N, the ML estimation is the one maximization the likelihood function:

$$(\hat{\theta}, \hat{\eta}) = \arg\max_{\theta, \eta} p_{y^N}(\theta, \eta, Z_*^N) \qquad (20)$$

where

$$Z_*^N = \{u^N, y_*^N\}$$

For the Hammerstein model (Figure 1), we first assume that the disturbance sequences $e(t)$ and $w(t)$ are white noises. This means that for given u^N, $y(t)$ will also be a sequence of independent variables. This in turn implies that the PDF of y^N will be the product of the PDF of $y(t)$, $t = 1, 2 \cdots N$. It is thus sufficient to derive the PDF of $y(t)$. To simplify notation we shall use $y(t) = y, \cdots x(t) = x$ for short.

To find the PDF, we introduce the intermediate signal x as a nuisance parameter. The PDF of y given x is basically a reflection of the PDF of e, since $y(t) = G(q, \theta)x(t) + e(t)$. It is easy to find if e is a white noise:

$$p_y(y/x) = p_e(y - G(q, \theta)x) \qquad (21)$$

where p_e is the PDF of e.

The same is true for the PDF of x given u^N if w is white noise,

Since $x(t) = f(u(t), \eta) + w(t)$. With given u^N and η, so

$$p_x(x/u^N, \theta) = p_w(x - f(u(t), \eta)) \qquad (22)$$

where p_w is the PDF of w.

Now by integrating over all $x \in R$, we then eliminate this immeasurable signal from the following equations:

$$p_y(\theta, \eta; Z_*^N) = \int_{x \in R} p_{x,y}(x, y / \theta, \eta; u^N) dx = \int_{x \in R} p_{y/x}(y / \theta, \eta, x; u^N) p_x(x / \theta, \eta; u^N) dx$$
$$= \int_{x \in R} p_e(y - G(q, \theta)x) p_w(x - f(u(t), \eta)) dx \qquad (23)$$

We now assume that the process noise $w(t)$ and the measurement noise $e(t)$ are Gaussian, with zero means and variances λ_w and λ_e respectively, i.e.

$$p_e(\varepsilon(t)) = \frac{1}{\sqrt{2\pi\lambda_e}} e^{-\frac{1}{2\lambda_e}\varepsilon^2(t)}, p_w(v(t)) = \frac{1}{\sqrt{2\pi\lambda_w}} e^{-\frac{1}{2\lambda_w}v^2(t)} \qquad (24)$$

For each time instant t. Since the noise is white, the joint likelihood is the product over all time instant, and thus

$$p_y(y^N / \theta, \eta; u^N) = \left(\frac{1}{2\pi\sqrt{\lambda_e\lambda_w}}\right)^N \prod_{t=1}^N \int_{-\infty}^\infty e^{-\frac{1}{2}E(t,\theta,\eta)} dx(t) = \left(\frac{1}{2\pi\sqrt{\lambda_e\lambda_w}}\right)^N \int_{x(1)=-\infty}^\infty \int_{x(2)=-\infty}^\infty \cdots \int_{x(N)=-\infty}^\infty e^{-\frac{1}{2}E(t,\theta,\eta)} dx^N \qquad (25)$$

where

$$E(t, \theta, \eta) = \frac{1}{\lambda_e}(y - G(q, \theta)x)^2 + \frac{1}{\lambda_w}(x - f(u(t), \eta))^2$$

Given data $Z_*^N = \{u_*^N, y_*^N\}$. We can calculate p_y and its gradients for each θ and η. This means that the ML criterion can be maximized numerically.

We may also note that each integral in (25) depends on $x(t)$ for only one time instant t, so they can be computed in parallel. If the noise covariance λ_w and λ_e are unknown, they can be just included among the parameters θ and η and their ML estimates are still obtained by (20).

5.2. Special Case: No Process Noise or no Measurement Noise

Most approaches suggested in the literature restrict the noise to either process noise or measurement noise. In this case, the likelihood function (25) is considerably simplified, and the criterion is reduced to something we recognize from other references.

First the case of no process noise, $\lambda_w = 0$. Since the only stochastic part is the measurement noise, we have

$$p_y\left(\theta,\eta;Z_*^N\right) = p_e\left(y(t)-G(q,\theta)x(t)\right) = p_e\left(y(t)-G(q,\theta)f\left(u(t),\eta\right)\right)$$

$$= \prod_{t=1}^{N}\frac{1}{\sqrt{2\pi\lambda_e}}\,e^{-\frac{1}{2\lambda_e}\left(y(t)-G(q,\theta)f\left(u(t),\eta\right)\right)^2} \tag{26}$$

Maximizing this is equivalent to minimizing the criterion

$$V_N\left(\theta,\eta\right) = \frac{1}{N}\sum_{t=1}^{N}\left(y(t)-G(q,\theta)f\left(u(t),\eta\right)\right)^2 \tag{27}$$

This is the prediction error criterion discussed before. It gives a consistent estimate if the condition of no process noise is satisfied. However if there is process noise, this criterion does not use the true prediction, and may be give biased estimates.

For systems with no measurement noise $\lambda_e = 0$ the criterion (25) forces $y(t) = G(q,\theta)x(t)$. So if G is invertible, the ML criterion becomes

$$p_y\left(\theta,\eta;Z_*^N\right) = p_w\left(G^{-1}(q,\theta)y(t)-f\left(u(t),\eta\right)\right) = \prod_{t=1}^{N}\frac{1}{\sqrt{2\pi\lambda_w}}\,e^{-\frac{1}{2\lambda_w}\left(G^{-1}(q,\theta)y(t)-f\left(u(t),\eta\right)\right)^2} \tag{28}$$

The maximum can be found by maximizing the logarithm, which reduces the problem to minimizing the criterion

$$V_N\left(\theta,\eta\right) = \frac{1}{N}\sum_{t=1}^{N}\left(G^{-1}(q,\theta)y(t)-f\left(u(t),\eta\right)\right)^2 \tag{29}$$

Remark. We should add the following theorem to describe the exist of the $G^{-1}(q,\theta)$

Theorem 3. Assume that the filter G is stable, and

$G^{-1}(q,\theta)=\frac{1}{G(q,\theta)}$. All that is needed is that the function $\frac{1}{G(q,\theta)}$ be analytic in $|q|\geq 1$; that is, it has no poles on or outside the unit circle. We could also phrase the condition as $G(q,\theta)$ must have no zeros on or outside the unit circle. This ties in very nicely with the spectral factorization result according to which for rational strictly positive spectra, we can always find a representation $G(q,\theta)$ with these properties.
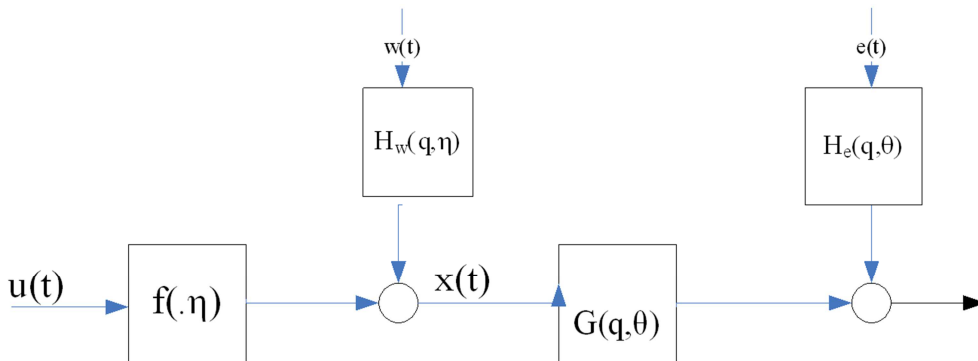
5.3. Colored Noise

Figure. 4. *Hammerstein model with colored noise.*

If the process and measurement noise are colored, we may represent the Hammerstein model as in Figure. 4. The following equations give the output:

$$\begin{aligned}
x(t) &= f\left(u(t),\eta\right)+H_w\left(q,\eta\right)w(t)\\
y(t) &= G(q,\theta)x(t)+H_e\left(q,\theta\right)e(t)
\end{aligned} \tag{30}$$

By using predictor form, we may write this as

$$\begin{aligned}
x\left(t,u^N,\eta\right) &= H_w^{-1}(q,\eta)f(u,\eta)+\left(1-H_w^{-1}(q,\eta)\right)x(t)+w(t)=\hat{x}\left(t/x^{t-1},u^N,\eta\right)+w(t)\\
y(t) &= H_e^{-1}(q,\theta)G(q,\theta)x(t)+\left(1-H_e^{-1}(q,\theta)\right)y(t)+e(t)\\
&= \hat{y}\left(t/y^{t-1},x^N,u^N,\theta\right)+e(t)
\end{aligned} \tag{31}$$

The only stochastic parts are e and w. For a given sequence x^N, the joint PDF of y^N is obtained in the stand way.

$$p_{y^N}\left(y^N / x^N\right) = \prod_{t=1}^{N} p_e\left(y(t) - \hat{y}\left(t / y^{t-1}, x^N, u^N, \theta\right)\right) \tag{32}$$

$$p_{x^N}\left(x^N\right) = \prod_{t=1}^{N} p_w\left(x(t) - \hat{x}\left(t / x^{t-1}, u^N, \eta\right)\right) \tag{33}$$

The likelihood function for y^N is obtained from (32) by integrating out the nuisance parameter x^N using its PDF (33):

$$p_{y^N}\left(y^N, \theta, \eta; u^N\right) = \int_{x(1) \in R}^{\infty} \int_{x(2) \in R}^{\infty} \cdots \int_{x(N) \in R}^{\infty} \prod_{t=1}^{N} p_e\left(H_e^{-1}(q,\theta) y(t) - H_e^{-1}(q,\theta) G(q,\theta) x(t)\right)$$
$$\times p_w\left(H_w^{-1}(q,\eta) x(t) - H_w^{-1}(q,\eta) f\left(u(t,\eta)\right)\right) \tag{34}$$

In the case e and w are Gaussian, we obtain

$$\prod_{t=1}^{N} p_e\left(H_e^{-1}(q,\theta) y(t) - H_e^{-1}(q,\theta) G(q,\theta) x(t)\right) \times p_w\left(H_w^{-1}(q,\eta) x(t) - H_w^{-1}(q,\eta) f\left(u(t,\eta)\right)\right)$$
$$= e^{-\frac{1}{2}\sum_{t=1}^{N} E(t,\theta,\eta)} \tag{35}$$

where

$$E(t,\theta,\eta) = \frac{1}{\lambda_e}\left(H_e^{-1}(q,\theta)\left(y(t) - G(q,\theta) x(t)\right)\right)^2 + \frac{1}{\lambda_w}\left(H_w^{-1}(q,\eta)\left(x(t) - f\left(u(t),\eta\right)\right)\right)^2 \tag{36}$$

The integration is a true multi-integral over all sequences x^N.

6. Conclusion

In this paper, one linear model with general disturbance is considered to be identified. The identification approach is the classical prediction error framework, where the error between the simulated and the measured outputs is needed to be minimized. Further more for the special case of Gaussian input signals, we estimate the linear part of the Hammerstein model using the Bussgang's classic theorem. We derive the maximum likelihood method to solve the identification problem with general disturbance.

References

[1] M C Campi, G Calafiore. "Interval predictor models: identification and reliability," Automatica, vol 45, no 2, pp. 382-393, 2009.

[2] M C Campi, M Vidyasagar. "Learning with prior information," IEEE Transaction on Automatic Control, vol 46, no 11, pp. 1682-1694, 2001.

[3] M Vidyasagar, Rajeeva L Karandikar. "A learning theory approach to system identification and stochastic adaptive control," Journal of Process Control, vol 18, no 3, pp. 421-430, 2008.

[4] M C Campi, P R Kumar. "Learning dynamical systems in a stationary environment," Systems & Control Letters, vol 34, no 3, pp. 125-132, 1998.

[5] G Calafiore, M C Campi. "Uncertain convex programs: randomized solutions and confidence levels," Mathematical Programming, vol 102, no 11, pp. 25-46, 2005.

[6] M Milanese, C Novara. "Set membership identification of nonlinear systems," Automatica, vol 40, no 6, pp. 957-975, 2004.

[7] T Alamo, J M Bravo, E F Camacho. "Guaranteed state estimation by zonotopes," Automatica, vol 41, no 6, pp. 1035-1043, 2005.

[8] J M Bravo, T Alamo, E F Camacho. "Bounded error identification of systems with time varying parameters," IEEE Transaction on Automatic Control, vol 51, no 7, pp. 1144-1150, 2006.

[9] J M Bravo, A Suarez, M Vasallo. "Slide window bounded error tome varying systems identification," IEEE Transaction on Automatic Control, vol 61, no 8, pp. 2282-2287, 2016.

[10] J M Bravo, T Alamo, M Vasallo. "A general framework for predictions based on bounding techniques and local approximation," IEEE Transaction on Automatic Control, vol 62, no 7, pp. 3430-3435, 2017.

[11] Marko Tanaskovic, Lorenzo Fagiano, Roy Smith. "Adaptive receding horizon control for constrained MIMO systems," Automatica, vol 50, no 12, pp. 3019-3029, 2014.

[12] Marko Tanaskovic, Lorenzo Fagiano, Carlo Novara. "Data driven control of nonlinear systems: an on line direct approach," Automatica, vol 75, no 1, pp. 1-10, 2017.