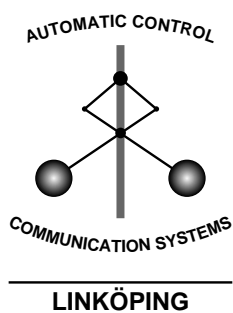


# Linear Models of Nonlinear FIR Systems with Gaussian Inputs

Martin Enqvist

Division of Automatic Control  
Department of Electrical Engineering  
Linköpings universitet, SE-581 83 Linköping, Sweden  
WWW: <http://www.control.isy.liu.se>  
E-mail: [maren@isy.liu.se](mailto:maren@isy.liu.se)

September 20, 2002



Report no.: LiTH-ISY-R-2462

Technical reports from the Control & Communication group in Linköping are available at <http://www.control.isy.liu.se/publications>.



# Linear Models of Nonlinear FIR Systems with Gaussian Inputs

Martin Enqvist

September 20, 2002

## Abstract

We show a result that can be viewed as a generalization of Bussgang's classic theorem about static nonlinearities with Gaussian inputs. This new result is used to characterize the best linear approximation of a nonlinear finite impulse response (NFIR) system with a Gaussian input. The best linear approximation is here defined as the causal LTI system that minimizes the expected squared prediction error. Furthermore, we discuss how this characterization can be used for structure identification and for identification of generalized Hammerstein and Wiener systems.

**Keywords:** System identification, Nonlinear FIR system, Gaussian input, Bussgang's theorem

## 1 Introduction

System identification deals with the problem of how to estimate a model of a dynamical system from measurements of the input and output signals. In practice, linear system models are very common and they are often used also when the true system is nonlinear. It is therefore interesting to understand how an estimated linear model depends on the properties of the true nonlinear system and of the input signal.

This question is hard to answer in general but it is possible to prove results for certain special cases. If the classes of systems and input signals are restricted to nonlinear finite impulse response (NFIR) systems with Gaussian inputs it is actually possible to characterize the best linear model completely.

## 2 Background

The following theorem is a classic result about Gaussian processes (cf. [4] for the original report and e.g. [9] for a more recent reference).

**Theorem 2.1 (Bussgang)** *Let  $y_t$  be the output from a static nonlinearity  $f$  with a Gaussian input  $u_t$ , i.e.  $y_t = f(u_t)$ . Assume that  $E\{y_t\} = E\{u_t\} = 0$ . Then*

$$R_{yu}(\tau) = E\{f'(u_t)\}R_u(\tau) \quad (1)$$

where  $R_{yu}(\tau) = E\{y_t u_{t-\tau}\}$  and  $R_u(\tau) = E\{u_t u_{t-\tau}\}$ .

Bussgang's theorem has turned out to be very useful for the theory of Hammerstein and Wiener system identification. (A Hammerstein systems consists of a static nonlinearity followed by an LTI system while a Wiener system is an LTI system followed by a static nonlinearity). The reason for this is that Bussgang's theorem explains why it is possible to estimate the linear and nonlinear parts of a Wiener or Hammerstein system separately when the input is Gaussian (cf. [1], [2], [5] and [6]).

### 3 A Generalization of Bussgang's Theorem

We will now show a result that can be viewed as a generalization of Bussgang's theorem.

**Theorem 3.1** *Let  $\tilde{x} = (x^T, v)^T = (x_1, x_2, \dots, x_N, v)^T$  be a jointly normally distributed random vector with zero mean and covariance matrix  $C$  with  $\det C \neq 0$ . Let  $f : \mathbb{R}^N \rightarrow \mathbb{R}$  be a differentiable function of  $x$  with  $\mathbb{E}\{f(x)\} = 0$  and let  $\varphi$  denote the probability density function of  $\tilde{x}$ . Assume that  $f \cdot \varphi$ ,  $f'_{\tilde{x}_i} \cdot \varphi$  and  $f \cdot \varphi'_{\tilde{x}_i}$ ,  $i = 1, \dots, (N + 1)$  all belongs to  $\mathcal{L}^1(\mathbb{R}^{N+1})$ . Then*

$$\mathbb{E}\{f(x)\tilde{x}\} = Cw \quad (2)$$

where

$$w = \begin{pmatrix} \mathbb{E}\{f'_{x_1}(x)\} \\ \mathbb{E}\{f'_{x_2}(x)\} \\ \vdots \\ \mathbb{E}\{f'_{x_N}(x)\} \\ 0 \end{pmatrix} \quad (3)$$

**Proof:** Factorize  $C$  as  $C = \tilde{M}\tilde{M}^T$  and define a new stochastic vector  $z$  as  $z = \tilde{M}^{-1}\tilde{x}$ . Then  $z$  is jointly normally distributed with zero mean and a covariance matrix that is equal to the identity matrix. Let  $M$  denote the matrix that is obtained from  $\tilde{M}$  by removing the last row. Then  $x = Mz$  and we get

$$\begin{aligned} \mathbb{E}\{f(x)\tilde{x}\} &= \tilde{M}\mathbb{E}\{f(x)\tilde{M}^{-1}\tilde{x}\} = \tilde{M}\mathbb{E}\{f(Mz)z\} = \\ &= \tilde{M} \begin{pmatrix} \mathbb{E}\left\{\frac{\partial f(Mz)}{\partial z_1}\right\} \\ \mathbb{E}\left\{\frac{\partial f(Mz)}{\partial z_2}\right\} \\ \vdots \\ \mathbb{E}\left\{\frac{\partial f(Mz)}{\partial z_{N+1}}\right\} \end{pmatrix} = \tilde{M}\tilde{M}^T \begin{pmatrix} \mathbb{E}\{f'_{x_1}(x)\} \\ \mathbb{E}\{f'_{x_2}(x)\} \\ \vdots \\ \mathbb{E}\{f'_{x_N}(x)\} \\ 0 \end{pmatrix} = \\ &= Cw \end{aligned} \quad (4)$$

(The third equality follows from the fact that  $\mathbb{E}\{h(z)z_i\} = \mathbb{E}\{h'_{z_i}(z)\}$  when  $z$  has an  $N(0, I)$  distribution).

□

Theorem 3.1 gives the following corollary.

**Corollary 3.1** Let  $y_t = f(u_t, u_{t-1}, \dots, u_{t-M_0})$  be an NFIR system with a stationary zero mean Gaussian process  $(u_t)_{t=-\infty}^{\infty}$  as input.

Form random vectors  $\omega_\sigma = (u_t, u_{t-1}, \dots, u_{t-M_0}, u_{t-\sigma})^T$ , with  $\sigma < 0$  or  $\sigma > M_0$ . Let  $C_\sigma$  and  $\varphi_\sigma$  denote the covariance matrices and joint probability density functions of these vectors, respectively. Assume that  $\det C_\sigma \neq 0$  for all  $\sigma < 0$  or  $\sigma > M_0$ .

Furthermore, assume that  $E\{y_t\} = 0$  and that  $f \cdot \varphi_\sigma$ ,  $f'_{u_{t-i}} \cdot \varphi_\sigma$ ,  $f \cdot \varphi'_{\sigma, u_{t-i}}$  and  $f \cdot \varphi'_{\sigma, u_{t-\sigma}}$ ,  $i = 0, \dots, M_0$  all belongs to  $\mathcal{L}^1(\mathbb{R}^{N+1})$  for all  $\sigma < 0$  or  $\sigma > M_0$ . Then

$$R_{yu}(\tau) = \sum_{k=0}^{M_0} b_k R_u(\tau - k) \quad \forall \tau \in \mathbb{Z} \quad (5)$$

where  $b_k = E\{f'_{u_{t-k}}(u_t, u_{t-1}, \dots, u_{t-M_0})\}$ ,  $R_{yu}(\tau) = E\{y_t u_{t-\tau}\}$  and  $R_u(\tau) = E\{u_t u_{t-\tau}\}$ .

**Proof:** Choose an arbitrary  $\sigma < 0$  or  $\sigma > M_0$  and let  $x = (u_t, u_{t-1}, \dots, u_{t-M_0})^T$  and  $v = u_{t-\sigma}$  in Theorem 3.1. Then equation (2) gives

$$\begin{aligned} E\{y_t \begin{pmatrix} u_t \\ u_{t-1} \\ \vdots \\ u_{t-M_0} \\ u_{t-\sigma} \end{pmatrix}\} &= \\ &= \begin{pmatrix} R_u(0) & R_u(1) & \dots & R_u(M_0) & R_u(\sigma) \\ R_u(1) & R_u(0) & \dots & R_u(M_0 - 1) & R_u(\sigma - 1) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ R_u(M_0) & R_u(M_0 - 1) & \dots & R_u(0) & R_u(\sigma - M_0) \\ R_u(\sigma) & R_u(\sigma - 1) & \dots & R_u(\sigma - M_0) & R_u(0) \end{pmatrix} w \quad (6) \end{aligned}$$

where  $w_{i+1} = E\{f'_{u_{t-i}}\}$  for  $i = 0, 1, \dots, M_0$  and  $w_{M_0+2} = 0$ . Equation (6) can be written more compactly as

$$R_{yu}(\tau) = \sum_{k=0}^{M_0} b_k R_u(\tau - k) \quad \tau = 0, 1, \dots, M_0 \quad \vee \quad \tau = \sigma \quad (7)$$

where  $b_k = E\{f'_{u_{t-k}}\}$ . As  $\sigma$  was arbitrary chosen, this relation must hold  $\forall \tau \in \mathbb{Z}$ . □

Let  $\Phi_{yu}$  and  $\Phi_u$  denote the z-transforms of  $R_{yu}$  and  $R_u$ , respectively. Provided that these transforms are well-defined, (5) can also be written as

$$\Phi_{yu}(z) = B(z)\Phi_u(z) \quad (8)$$

where  $B(z) = \sum_{k=0}^{M_0} b_k z^{-k}$ .

## 4 LTI-SOE:s of NFIR Systems

The generalization of Bussgang's theorem in the previous section can be used to characterize the "best" linear approximation of an NFIR system. We will here define this "best" linear approximation of an NFIR system to be the causal LTI system  $G_0$  that minimizes the expected squared prediction error,  $E\{(y_t - G(q)u_t)^2\}$ . We will call  $G_0$  the LTI second order equivalent (LTI-SOE) of the nonlinear system with respect to the input  $u_t$ .

Note that the LTI-SOE:s that we will consider here, unlike the general case, do not contain a noise model (cf. [8]). This implies that only past inputs  $u_t, u_{t-1}, \dots$  are used to predict the output  $y_t$ .

It can be shown that the LTI-SOE of a nonlinear system under this condition will be equal to a causal representation of the quotient  $\Phi_{yu}(z)/\Phi_u(z)$ , where  $\Phi_{yu}$  and  $\Phi_u$  are the z-transforms of  $R_{yu}$  and  $R_u$ , respectively (cf. [7] and [8]). However, from (8) we see that this quotient already is causal if the nonlinear system is an NFIR system with a Gaussian input.

Hence, the LTI-SOE of an NFIR system  $y_t = f(u_t, u_{t-1}, \dots, u_{t-M_0})$  with a Gaussian input  $u_t$  is the linear FIR system

$$G_0(z) = \frac{\Phi_{yu}(z)}{\Phi_u(z)} = \sum_{k=0}^{M_0} b_k z^{-k} \quad (9)$$

where  $b_k = E\{f'_{u_{t-k}}\}$ . Assume that the prediction error method is used to estimate an output error model from input output data that come from an NFIR system with a Gaussian input. It can then be shown that this model will converge to the LTI-SOE of the system when the number of measurements tends to infinity, provided that the model order is sufficiently high (cf. [8]).

In general, it is quite possible that the LTI-SOE of an NFIR system with a non-Gaussian input will have an infinite impulse response and it is usually hard to give a detailed characterization of it. However, as we have shown here, when the input is Gaussian the LTI-SOE is always an FIR system and the coefficients of this system can be characterized exactly by (9).

## 5 Geometric Interpretation

In many cases, it is possible to shed some light on a theoretical result by interpreting it in a geometrical framework. This can as matter of fact be done also in our case. For a fixed  $t$ , we can view the output  $y_t$  and the components of the input signal  $u_\tau, \tau \in \mathbb{Z}$  as vectors in an infinite dimensional inner-product space with the inner product  $\langle u, v \rangle = E\{uv\}$  (cf. [3]).

The LTI-SOE of the NFIR system will in this framework be the orthogonal projection of  $y_t$  into the linear subspace that is spanned by  $u_t, u_{t-1}, \dots, u_{t-\infty}$ . From (9) we can draw the conclusion that this projection actually lies in the finite dimensional linear subspace that is spanned by  $u_t, u_{t-1}, \dots, u_{t-M_0}$ .

## 6 Applications

The characterization (9) of the LTI-SOE of an NFIR system with a Gaussian input is not only theoretically interesting but can also be useful in some appli-

cations of system identification. We will here briefly discuss three such applied identification problems.

## 6.1 Structure Identification of NFIR Systems

The most obvious application of the result (9) is perhaps to use it for guidance when an NFIR system is to be identified. As (9) only is influenced by odd terms in the system we will here only consider odd NFIR systems.

When an odd NFIR system  $y_t = f(u_{t-n_k}, u_{t-n_k-1}, \dots, u_{t-n_k-M_0})$  is to be identified it is in general not obvious how the time delay  $n_k$  and order  $M_0$  should be estimated in an efficient way. However, if the input is Gaussian and sufficiently many measurements can be collected,  $n_k$  and  $M_0$  can both be obtained from an impulse response estimate. Such an estimate can be computed very efficiently by means of the least squares method.

Furthermore, if only a few of the input terms  $u_{t-n_k}, u_{t-n_k-1}, \dots, u_{t-n_k-M_0}$  enter the system in a nonlinear way it might be interesting to know which these terms are. If a nonlinear model of the system is desired this knowledge can be used to reduce the complexity of the proposed model. A coefficient  $b_j$  in (9) will be invariant of the input properties if the corresponding input term  $u_{t-j}$  only affects the system linearly while an input term that affects the system in a nonlinear way will have an input dependent  $b$ -coefficient in (9).

This fact makes it possible to extract information about which nonlinear terms that are present in the system simply by looking at the differences between FIR models that have been estimated with different Gaussian input signals. The coefficients that correspond to an input term that enters the system in a nonlinear way will be different in these estimates, provided that the covariance functions of the inputs are different.

## 6.2 Identification of Generalized Hammerstein Systems

In Section 2 we mentioned that Bussgang's theorem has been used to show important results concerning the identification of Hammerstein and Wiener systems. In principle, these results say that an estimated LTI model will converge to a scaled version of the linear part of a Hammerstein or Wiener system when the number of measurements tends to infinity, provided that the input is Gaussian. These results simplify the identification of Wiener and Hammerstein systems significantly.

Hence, it is interesting to investigate if the result (9) about the LTI-SOE:s of NFIR systems can be used to prove similar results for extended classes of systems. In this section we will study a type of systems that we will call generalized Hammerstein systems, while we in the next section will consider generalized Wiener systems.

More specifically, we will call a nonlinear system a generalized Hammerstein system if it consists of an NFIR system  $v_t = f(u_t, u_{t-1}, \dots, u_{t-M_0})$  followed by an LTI system  $y_t = G(q)v_t$ . If  $G$  is causal it can be written as

$$y_t = \sum_{k=0}^{\infty} g_k v_{t-k} \quad (10)$$

If we multiply both sides of Equation (10) with  $u_{t-\tau}$  and take the expectation we get

$$R_{yu}(\tau) = \sum_{k=0}^{\infty} g_k R_{vu}(\tau - k), \quad (11)$$

provided that a term by term calculation of the expectation is allowed. If all z-transforms are well-defined, Equation (11) can also be written as

$$\Phi_{yu}(z) = G(z)\Phi_{vu}(z) \quad (12)$$

If  $u$  is Gaussian and  $f$  is such that Corollary 3.1 can be applied we thus get

$$\Phi_{yu}(z) = G(z)B(z)\Phi_u(z) \quad (13)$$

where  $B(z) = \sum_{k=0}^{M_0} b_k z^{-k}$  and  $b_k = E\{f'_{u_{t-k}}\}$ .

Hence, the LTI-SOE of a generalized Hammerstein system with a Gaussian input will be  $G(z)B(z)$  and an estimated output error model will approach this system as the number of measurements tends to infinity. In particular, as  $B(z)$  is an FIR system, this shows that the denominator of the estimated model will approach the denominator of  $G$  if the degree of the model denominator polynomial is correct.

We will thus get consistent estimates of the poles of  $G$  despite the presence of the NFIR system. This result is verified experimentally in Example 6.1.

**Example 6.1** Consider a generalized Hammerstein system

$$y_t = G(q)f(u_t, u_{t-1}) + e_t \quad (14)$$

where

$$G(q) = \frac{1}{1 + 0.6q^{-1} + 0.1q^{-2}} \quad (15)$$

$$f(u_t, u_{t-1}) = \arctan(u_t) \cdot u_{t-1}^2 \quad (16)$$

and where  $e_t$  is white Gaussian noise with  $E\{e_t\} = 0$  and  $E\{e_t^2\} = 1$ .

Let the input  $u_t$  be generated by linear filtering of a white Gaussian process  $x_t$  with  $E\{x_t\} = 0$  and  $E\{x_t^2\} = 1$  such that

$$u_t = \frac{1 + q^{-1} + q^{-2}}{1 - 0.2q^{-1}} x_t \quad (17)$$

and assume that  $x_t$  and  $e_s$  are independent  $\forall t, s \in \mathbb{Z}$ .

This input signal has been used in an identification experiment where a data set consisting of 10000 measurements of  $u_t$  and  $y_t$  was collected. A linear output error model  $\hat{G}_{oe}$  with  $n_b = n_f = 2$  and  $n_k = 0$  has been estimated from this data set and the result was

$$\hat{G}_{oe} = \frac{1.13 + 2.61q^{-1}}{1 + 0.573q^{-1} + 0.0954q^{-2}} \quad (18)$$

As can easily be seen from (18), the denominator of  $\hat{G}_{oe}$  is indeed very close to the denominator of  $G$ . This is of course exactly what one would expect with the previous theoretical discussion in mind.



### 6.3 Identification of Generalized Wiener Systems

We will call a nonlinear system a generalized Wiener system if it consists of an LTI system  $n_t = G(q)u_t$  followed by an NFIR system  $y_t = f(n_t, \dots, n_{t-M_0})$ .

Assume that  $u_t$  is Gaussian and that the linear and nonlinear parts of the system are such that Theorem 3.1 can be applied with  $x = (n_t, n_{t-1}, \dots, n_{t-M_0})^T$  and  $v = u_{t-\tau}$  for any  $\tau \in \mathbb{Z}$ . (Note that  $n_t$  will be Gaussian as it is a linearly filtered Gaussian signal). The last row of Equation (2) then gives

$$R_{yu}(\tau) = \sum_{k=0}^{M_0} b_k R_{nu}(\tau - k) \quad (19)$$

where  $b_k = \mathbb{E}\{f'_{n_{t-k}}\}$ . Equation (19) can also be written as

$$\Phi_{yu}(\tau) = B(z)\Phi_{nu}(z) = B(z)G(z)\Phi_u(z) \quad (20)$$

and hence the LTI-SOE of a generalized Wiener system with a Gaussian input will be  $B(z)G(z)$ .

This implies, just as in the case with generalized Hammerstein systems, that consistent estimates of the poles of  $G$  can be obtained by estimating an output error model. The following example verifies this result for a particular generalized Wiener system.

---

**Example 6.2** Consider a generalized Wiener system consisting of the same linear and nonlinear blocks as the generalized Hammerstein system in Example 6.1 but with the linear block before the nonlinear, i.e.

$$y_t = f(n_t, n_{t-1}) + e_t \quad (21)$$

$$n_t = G(q)u_t \quad (22)$$

where

$$G(q) = \frac{1}{1 + 0.6q^{-1} + 0.1q^{-2}} \quad (23)$$

$$f(n_t, n_{t-1}) = \arctan(n_t) \cdot n_{t-1}^2 \quad (24)$$

and where  $e_t$  is white Gaussian noise with  $\mathbb{E}\{e_t\} = 0$  and  $\mathbb{E}\{e_t^2\} = 1$ .

Let the input  $u_t$  be generated in the same way as in Example 6.1 i.e.

$$u_t = \frac{1 + q^{-1} + q^{-2}}{1 - 0.2q^{-1}}x_t \quad (25)$$

where  $x_t$  is a white Gaussian process with  $\mathbb{E}\{x_t\} = 0$  and  $\mathbb{E}\{x_t^2\} = 1$  such that  $x_t$  and  $e_s$  are independent  $\forall t, s \in \mathbb{Z}$ .

An identification experiment has been performed on this generalized Wiener system with a realization of this  $u_t$  as input and 10000 measurements of  $u_t$  and  $y_t$  has been collected. A linear output error model  $\hat{G}_{oe}$  with  $n_b = n_f = 2$  and  $n_k = 0$  has been estimated from the measurements and the result was

$$\hat{G}_{oe} = \frac{1.01 + 0.874q^{-1}}{1 + 0.565q^{-1} + 0.0975q^{-2}} \quad (26)$$

From (26) we can see that the denominator of  $\hat{G}_{oe}$  is very close to the denominator of  $G$  also when the data has been generated by a generalized Wiener system.

---

## 7 Discussion

In the previous sections we have given a characterization of the LTI-SOE of an NFIR system with a Gaussian input. We have shown that this LTI-SOE will be an FIR system and described how the coefficients of this FIR system depends on the properties of the NFIR system and the input signal. Furthermore, we have also discussed some applications of these results in structure identification and identification of generalized Hammerstein and Wiener models.

The LTI-SOE will only depend on the odd terms of the NFIR system. This is due to the fact that the FIR coefficients are expectations of the partial derivatives of the NFIR system. However, a model that is estimated from a relatively small data set can be heavily influenced by the even terms of the nonlinear system. Hence, there is a need to investigate the influence of even nonlinearities further.

## References

- [1] J. S. Bendat. *Nonlinear Systems Techniques and Applications*. John Wiley & Sons, New York, 1998.
- [2] S. A. Billings and S. Y. Fakhouri. Identification of systems containing linear dynamic and static nonlinear elements. *Automatica*, 18(1):15–26, 1982.
- [3] P. J. Brockwell and R. A. Davis. *Time Series: Theory and Methods*. Springer, New York, 1987.
- [4] J. J. Bussgang. Crosscorrelation functions of amplitude-distorted Gaussian signals. Technical Report Technical report 216, MIT Laboratory of Electronics, 1952.
- [5] M. J. Korenberg. Identifying noisy cascades of linear and static nonlinear systems. In *Proc. 7th IFAC Symp. on Identification and System Parameter Identification*, pages 421–426, York, U.K., 1985.
- [6] P. Koukoulas and N. Kalouptsidis. Nonlinear system identification using Gaussian inputs. *IEEE Transactions on Signal Processing*, 43(8):1831–1841, August 1995.
- [7] L. Ljung. *System Identification: Theory for the User*. Prentice-Hall, Upper Saddle River, NJ, 2nd edition, 1999.
- [8] L. Ljung. Estimating linear time invariant models of non-linear time-varying systems. *European Journal of Control*, 7(2-3):203–219, Sept 2001. Semi-plenary presentation at the European Control Conference, Sept 2001.
- [9] A. Papoulis. *Probability, Random Variables and Stochastic Processes*. McGraw Hill, 2nd edition, 1984.