

# A Weighting Method for Approximate Nonlinear System Identification

Martin Enqvist

Division of Automatic Control

E-mail: [maren@isy.liu.se](mailto:maren@isy.liu.se)

10th October 2007

Report no.: LiTH-ISY-R-2829

Submitted to the 46th IEEE Conference on Decision and Control, New Orleans, Louisiana, USA

Address:

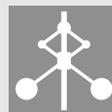
Department of Electrical Engineering

Linköpings universitet

SE-581 83 Linköping, Sweden

WWW: <http://www.control.isy.liu.se>

AUTOMATIC CONTROL  
REGLERTEKNIK  
LINKÖPINGS UNIVERSITET



## **Abstract**

Many approximation results in nonlinear system identification concern particular signal distributions. This seems to limit the applicability of these results to cases where the relevant signals have these distributions. However, by using a weighting method that modifies the cost function used in the identification method, the available approximation results can be used also for rather general classes of signal distributions. The purpose of this paper is to describe this weighting approach and to point at some interesting application areas within nonlinear system identification. In particular, it will be described how the impulse response of a Hammerstein system can be estimated consistently for an arbitrary input signal.

**Keywords:** Nonlinear systems, Approximation, System identification, Importance weighting

# A Weighting Method for Approximate Nonlinear System Identification

Martin Enqvist

2007-10-10

## Abstract

Many approximation results in nonlinear system identification concern particular signal distributions. This seems to limit the applicability of these results to cases where the relevant signals have these distributions. However, by using a weighting method that modifies the cost function used in the identification method, the available approximation results can be used also for rather general classes of signal distributions. The purpose of this paper is to describe this weighting approach and to point at some interesting application areas within nonlinear system identification. In particular, it will be described how the impulse response of a Hammerstein system can be estimated consistently for an arbitrary input signal.

## 1 Introduction

Given input and output data from a dynamical system, a model can be estimated using methods for system identification. For example, a model can be found by minimizing the mismatch between system and model output measured by some cost function. Although it might be possible to find a model structure that contains the true system in some applications, most model structures will only be able to give an approximate model. In particular, this is usually the case when the system is nonlinear. Hence, it is interesting to discuss approximation issues in nonlinear system identification.

A number of useful results about approximate models for some classes of nonlinear systems are available in the literature. Typically, these results concern particular combinations of system and model structures, input distributions and cost functions. For example, a number of results are available for linear models of nonlinear regression systems with Gaussian regressors and quadratic cost functions (Atalik and Utku, 1976; Bussgang, 1952; Cook, 1998; Enqvist and Ljung, 2005; Li, 1991, 1992; Li and Duan, 1989). Some of these results hold also for more general distributions, e.g., elliptical or separable distributions (Nuttall, 1958).

Another example is linear models of nonlinear systems with random multisine inputs and quadratic cost functions. These signals imply various useful properties of the linear approximations of systems with convergent Volterra series in general, and of nonlinear systems with special structures, like Wiener, Hammerstein and Wiener-Hammerstein systems, in particular (Pintelon and Schoukens, 2001; Schoukens et al., 1998, 2005). The main characteristic of a

random multisine signal is that its phases  $\psi_k$  are independent random variables with the property that

$$\mathbb{E}(e^{in\psi_k}) = 0, \quad \forall n \in \mathbb{Z} \setminus \{0\}.$$

For example, this property holds when the phases have a uniform distribution on  $[0, 2\pi]$ .

At first, it might seem that all results about approximate linear models for particular signal distributions are useless if the actual, measured signals cannot be guaranteed to come from a stochastic process with one of these distributions. However, this seems to be a rash conclusion. By using some techniques originally designed for random number generation, the approximation results for particular classes of signal distributions can be generalized. The main idea is to introduce weighting factors in the cost function. This idea has been used previously to generalize some results about dimension reduction in nonlinear regression models (Brillinger, 1991; Cook and Nachtsheim, 1994). Here, the main focus will be on some other problems in nonlinear system identification.

In Section 2, the weighting approach is formulated first for a generic identification problem using the methods from Beckman and McKay (1987), Brillinger (1991) and Cook and Nachtsheim (1994). Furthermore, a version of the algorithm from Cook and Nachtsheim (1994) for computing the weights based on measured data is described in Section 3. With this method, the weighting approach can be used also when the experimental signal distribution is unknown. In Section 4, it is described how the weighting method can be used to obtain consistent estimates of the impulse responses of Wiener or Hammerstein systems for arbitrary input signals. This section contains also a numerical example in which a Hammerstein system is studied. Furthermore, some other possible identification applications where the weighting method might be useful are described in Section 5. Finally, the use of the weighting method in deterministic identification problems and the relation between this method and the choice of cost function are discussed briefly in Section 6 and some conclusions are drawn in Section 7.

## 2 Weighted Cost Functions

Consider a generic system

$$y = f(x) + e, \quad (1)$$

where the input  $x$  and output  $y$  are random variables that can be measured and  $e$  is a disturbance that is independent of  $x$ . Assume that  $\dim(x) = m$  and  $\dim(y) = n$ . Let  $x_E$  denote the input vector when its probability density function is  $p_E$  and let  $y_E$  be the corresponding output from (1) and assume that a dataset  $(x_E(k), y_E(k))_{k=1}^N$  with  $N$  input and output measurements is available. Consider a generic model

$$\hat{y}_E(\theta) = g(x_E, \theta), \quad (2)$$

where  $\theta \in \mathbb{R}^p$  is a parameter vector, and a cost function

$$V_{N,w,E}(\theta) = \frac{1}{N} \sum_{k=1}^N w(x_E(k)) l(y_E(k) - g(x_E(k), \theta)), \quad (3)$$

where  $w$  and  $l$  are real-valued, non-negative functions from  $\mathbb{R}^m$  and  $\mathbb{R}^n$ , respectively. Based on the measurements, a parameter estimate  $\hat{\theta}_{N,w,E}$  can be obtained by minimizing  $V_{N,w,E}(\theta)$ , i.e.,

$$\hat{\theta}_{N,w,E} = \arg \min_{\theta \in \mathcal{D}} V_{N,w,E}(\theta),$$

where  $\mathcal{D} \subseteq \mathbb{R}^p$  is some set of parameters. Consider a particular input probability density function  $p_D$  that is such that the cost function converges to the mean-square error like

$$\lim_{N \rightarrow \infty} V_{N,1,D}(\theta) = \bar{V}_{1,D}(\theta) = \mathbb{E}(l(y_D - g(x_D, \theta))) \quad \text{w.p.1}$$

and, similarly, that

$$\lim_{N \rightarrow \infty} \hat{\theta}_{N,1,D} = \theta_{1,D}^* = \arg \min_{\theta \in \mathcal{D}} \bar{V}_{1,D}(\theta) \quad \text{w.p.1.}$$

Assume that the input distribution defined by  $p_D$  implies that  $\theta_{1,D}^*$  will have some useful properties, e.g., that it defines a natural approximation of the true system or that it reveals some interesting system properties. The most obvious way to obtain a consistent estimate of  $\theta_{1,D}^*$  is to perform an experiment where measurements  $y_D(k)$  and  $x_D(k)$  are collected and then to compute  $\hat{\theta}_{N,1,D}$ .

Of course, this is not an option if the system identification should be performed based on an *existing* dataset with some other input distribution. Furthermore, designing a new experiment in order to obtain the desired input distribution might be difficult in some cases, either because the distribution of  $x$  cannot be controlled directly by the user or because the desired distribution is unsuitable for the particular application.

Hence, it would be interesting to have a consistent estimator of  $\theta_{1,D}^*$  that is based on data with the input probability density function  $p_E$  instead of  $p_D$ . It turns out that there are a couple of related, but different, ways to construct such an estimator. A common feature of these approaches is that they are all based on different choices of weighting functions  $w$  in (3).

Without any prior knowledge, nothing can be said about a nonlinear system in regions where there is no data. Hence, the measured  $x_E(k)$  vectors must asymptotically be dense in the whole region where there should be  $x_D(k)$  vectors. More specifically, it is assumed that the support  $\mathcal{S}_D$  of  $p_D$  is a subset of the support  $\mathcal{S}_E$  of  $p_E$ , i.e., that

$$\mathcal{S}_D \subseteq \mathcal{S}_E. \quad (4)$$

Furthermore, it is assumed that

$$\frac{p_D(x)}{p_E(x)} \leq M, \quad \forall x \in \mathcal{S}_E,$$

where  $M \geq 1$  is a constant.

## 2.1 Weightings Based on Importance Sampling

The first estimator of  $\theta_{1,D}^*$  is based on a procedure known as *importance sampling*. Let

$$q(x) = \frac{p_D(x)}{p_E(x)}$$

and consider the estimator  $\hat{\theta}_{N,q,E}$ . Assume that this estimator converges to the parameter vector minimizing the mean-square error, i.e., that

$$\lim_{N \rightarrow \infty} \hat{\theta}_{N,q,E} = \theta_{q,E}^* = \arg \min_{\theta \in \mathcal{D}} \underbrace{\mathbb{E}(q(x_E)l(y_E - g(x_E, \theta)))}_{=\bar{V}_{q,E}(\theta)} \quad \text{w.p.1.}$$

Let  $p$  denote the probability density function of  $e$ . Since  $e$  and  $x$  are independent for any input distribution and  $\mathcal{S}_D \subseteq \mathcal{S}_E$ , it follows that

$$\begin{aligned} \bar{V}_{q,E}(\theta) &= \int_{e \in \mathbb{R}^n} \int_{x \in \mathcal{S}_E} q(x)l(f(x) + e - g(x, \theta))p_E(x)p(e) dx de \\ &= \int_{e \in \mathbb{R}^n} \int_{x \in \mathcal{S}_D} l(f(x) + e - g(x, \theta))p_D(x)p(e) dx de \\ &= \bar{V}_{1,D}(\theta) \end{aligned}$$

and thus that  $\theta_{q,E}^* = \theta_{1,D}^*$ . Hence,  $\hat{\theta}_{N,q,E}$  is a consistent estimator of  $\theta_{1,D}^*$ .

This approach was proposed in Beckman and McKay (1987) as a means to reuse data when estimating the expected value  $\mathbb{E}(h(v))$  of some function  $h$  of a random variable  $v$  for different distributions of  $v$ . This idea has also been used in Brillinger (1991) and Cook and Nachtsheim (1994), but with approximate weights instead of the exact ones above. These approximate weights can be obtained from Monte Carlo simulations based on  $p_D$  and do not require that  $p_E$  is known. In Cook and Nachtsheim (1994), it is discussed also how suitable choices of  $p_D$  can be found by removing some percentage of the original measurements and assuming an elliptical distribution on the minimum volume ellipsoid (MVE) containing the remaining measurements. A version of this approach is presented in Section 3.

## 2.2 Weightings Based on Rejection

The second approach that can be used for constructing a consistent estimator of  $\theta_{1,D}^*$  is based on the *rejection method* for random number generation. The underlying idea in this method is to choose binary weights  $w(x_E(k))$  randomly according to

$$w(x_E(k)) = \begin{cases} 1, & \text{with probability } \tilde{q}(x_E(k)), \\ 0, & \text{otherwise,} \end{cases} \quad (5)$$

where

$$\tilde{q}(x_E(k)) = \frac{p_D(x_E(k))}{Mp_E(x_E(k))}.$$

For example, the randomness can be generated by using  $N$  independent realizations  $v_k$  of a random variable  $v$  with uniform distribution on  $[0, 1]$ . Using these realizations, the weight selection in (5) can be rewritten as

$$w(x_E(k)) = \begin{cases} 1, & v_k \leq \tilde{q}(x_E(k)), \\ 0, & \text{otherwise.} \end{cases} \quad (6)$$

Let  $\preceq$  denote component-wise inequalities and let  $F_{x_D}$  denote the probability distribution function of  $x_D$ . The probability distribution function of  $x_E$  given

that  $w = w(x_E) = 1$  is

$$\begin{aligned}
F_{x_E|w=1}(z) &= P(x_E \preceq z | w = 1) \\
&= \frac{P(x_E \preceq z \cap w = 1)}{P(w = 1)} \\
&= \frac{P(x_E \preceq z \cap v \leq \tilde{q}(x_E))}{P(v \leq \tilde{q}(x_E))} \\
&= \frac{\int_{x \preceq z \cap x \in \mathcal{S}_E} P(v \leq \tilde{q}(x_E) | x_E = x) p_E(x) dx}{\int_{x \in \mathcal{S}_E} P(v \leq \tilde{q}(x_E) | x_E = x) p_E(x) dx} \\
&= \frac{\int_{x \preceq z \cap x \in \mathcal{S}_E} p_D(x) dx}{\int_{x \in \mathcal{S}_E} p_D(x) dx} = F_{x_D}(z).
\end{aligned}$$

Here, the fifth equality follows from the fact that

$$P(v \leq \tilde{q}(x_E) | x_E = x) = \frac{p_D(x)}{M p_E(x)}$$

and the last one holds since  $\mathcal{S}_D \subseteq \mathcal{S}_E$ . Hence, the  $x_E(k)$  elements with positive weights will have the desired distribution and

$$\begin{aligned}
\bar{V}_{w,E}(\theta) &= \int_{e \in \mathbb{R}^n} \int_{x \in \mathcal{S}_E} \int_{v \leq \tilde{q}(x)} l(f(x) + e - g(x, \theta)) p_E(x) p(e) dv dx de \\
&= \int_{e \in \mathbb{R}^n} \int_{x \in \mathcal{S}_E} \frac{p_D(x)}{M p_E(x)} l(f(x) + e - g(x, \theta)) p_E(x) p(e) dx de \\
&= \frac{1}{M} \int_{e \in \mathbb{R}^n} \int_{x \in \mathcal{S}_D} l(f(x) + e - g(x, \theta)) p_D(x) p(e) dx de \\
&= \frac{1}{M} \bar{V}_{1,D}(\theta).
\end{aligned}$$

Under the assumption that

$$\lim_{N \rightarrow \infty} \hat{\theta}_{N,w,E} = \theta_{w,E}^* = \arg \min_{\theta \in \mathcal{D}} \bar{V}_{w,E}(\theta) \quad \text{w.p.1,}$$

$\hat{\theta}_{N,w,E}$  will be a consistent estimator of  $\theta_{1,D}^*$ .

A well-known drawback with the rejection method is that it requires a large dataset if  $M$  is large. For example, assume that  $p_E(x) = p_D(x)$  everywhere except in a very small region where  $p_D(x_0) = M p_E(x_0)$  with  $M \gg 1$  for some  $x_0$ . In this case, most weights obtained from the rejection method will be zero while most weights based on importance sampling will be equal to one. Typically, this will result in a slower convergence of the rejection-based estimator.

The rejection method has also been discussed by Beckman and McKay (1987) in the context of estimating  $E(h(v))$  for different distributions.

### 3 Practical Issues

A couple of practical issues have to be solved before the weighting method from the previous section can be applied to an estimation problem.

### 3.1 Choice of Target Distribution

First, a procedure for finding a suitable target distribution, which is defined by the desired target probability density function  $p_D$ , is required. Since (4) is the only strict requirement on this distribution, there are many possible choices of  $p_D$ . In order to get accurate estimates, it is important to avoid including regions where there are few measurements in the support  $\mathcal{S}_D$  of  $p_D$ . Here, the aim has been to obtain an elliptical  $p_D$ . Hence, its support has been chosen as an ellipsoid

$$\mathcal{S}_D = \{x \in \mathbb{R}^m \mid (x - \hat{m}_x)^T (s_e \hat{C}_x)^{-1} (x - \hat{m}_x) \leq 1\}, \quad (7)$$

where  $\hat{m}_x$  and  $\hat{C}_x$  are estimates of the mean and covariance matrix of  $x$ , respectively, and where  $s_e$  is a scaling factor that can be varied in order to cover a certain fraction of the measurements.

An alternative to this approach is to use an estimate of the *minimum volume ellipsoid* (MVE) for the support of  $p_D$  (Cook and Nachtsheim, 1994). The MVE is the ellipsoid with smallest volume that contains a specified fraction of the measurements. Several estimators for the MVE have been proposed in literature (Poston et al., 1997; Titterton, 1975). However, for computational complexity reasons, this approach has not been used here.

When the support of  $p_D$  has been determined, there are still many possible choices of an elliptical target probability density function. Any elliptically distributed random variable  $z$  with mean  $m_z$  and covariance matrix  $C_z$  can be written in the form (Cook and Nachtsheim, 1994)

$$z = rBu + m_z, \quad (8)$$

where  $C_z = BB^T$ ,  $u$  is uniformly distributed on the surface of the unit hypersphere with dimension equal to that of  $z$  and  $r$  is a positive scalar-valued random variable that is independent of  $u$ . This implies that in order to characterize an elliptically distributed random variable,  $m_z$ ,  $C_z$  and  $r$  need to be determined. Here, in the design of a  $z$  with probability density function  $p_D$ , the estimated mean and covariance matrix of the  $x$  measurements have been used as  $m_z$  and  $C_z$ , respectively. Furthermore, the empirical probability density function of the lengths of the standardized  $x$  vectors in  $\mathcal{S}_D$  has been used as probability density function for the random variable  $r$ , following the approach suggested in Cook and Nachtsheim (1994). The standardization of the  $x$  vectors has been made using the estimated mean and covariance matrix and the empirical probability density function has been estimated from a histogram with  $n_h$  bins.

### 3.2 Monte Carlo Weightings

A problem with the weighting method as it was presented in Section 2 is that it requires knowledge of the experimental probability density function  $p_E$ . Of course, it is possible to estimate the distribution of  $x$  from measured data, but a more convenient solution might be to use a Monte Carlo method for obtaining approximate weights directly from data (Cook and Nachtsheim, 1994).

Consider a set  $(x(k))_{k=1}^N$  of  $x$  measurements and let

$$d(x(k)) = \{x \in \mathbb{R}^m \mid \|x - x(k)\| \leq \|x - x(j)\|, j = 1, \dots, N\}$$

denote the *Dirichlet cell* around  $x(k)$ . Using these regions, approximate weights

$$\hat{q}(x(k)) = \int_{\tilde{x} \in d(x(k))} p_D(\tilde{x}) d\tilde{x} \quad (9)$$

can be calculated for  $k = 1, 2, \dots, N$ . Besides an insignificant scaling factor, the weights  $\hat{q}(x(k))$  will be accurate approximations of the exact weights  $q(x(k))$  for large  $N$ . Instead of solving the integral in (9) analytically, which can be hard, it can be solved approximately using Monte Carlo simulations. By generating  $P$  independent realizations of a random variable with probability density function  $p_D(x)$  and counting the number of occurrences in each  $d(x(k))$ , the weights can be obtained. This is the method for weight generation that has been used here and it is described in more detail in Cook and Nachtsheim (1994).

## 4 Block-Oriented Systems

Previously, the weighting method has been applied to a number of problems concerning dimension reduction in nonlinear regression analysis and general simulation problems (Beckman and McKay, 1987; Cook and Nachtsheim, 1994). However, there seems to be some applications concerning nonlinear system identification that remains to be investigated. One such application is estimation of impulse responses in block-oriented systems and this application will be described here, while some other identification applications are presented in the next section.

Block-oriented systems are systems that consist of cascaded blocks of linear time-invariant (LTI) subsystems and static nonlinearities. The two simplest forms of block-oriented systems are Hammerstein systems

$$y(t) = G(q)f(u(t)) + w(t)$$

and Wiener systems

$$y(t) = f(v(t)) + w(t), \quad v(t) = G(q)u(t).$$

Here,  $q$  denotes the shift operator,  $qu(t) = u(t + 1)$  and  $w(t)$  is an output disturbance.

It is well-known that the linear part of a Hammerstein or a Wiener system can be estimated without compensating for the nonlinearity at the input or output when the input signal is Gaussian (Billings and Fakhouri, 1978, 1982; Bussgang, 1952; Hunter and Korenberg, 1986). This result can be generalized to signals with elliptical distributions in the Wiener case and to separable signals in the Hammerstein case (Nuttall, 1958), but it does not hold for all input signals.

However, the weighting method provides a way to handle also arbitrary input signals. Consider a Wiener or Hammerstein system where the impulse response of the LTI subsystem is  $g(k)$ . A linear least-squares (LLS) estimator  $\hat{\theta}_{LLS}$  of this impulse response can be written

$$\hat{\theta}_{LLS} = \arg \min_{\theta} \frac{1}{N} \sum_{t=n_b}^N (y(t) - \varphi(t)^T \theta)^2,$$

where  $\varphi(t) = (u(t) \ u(t-1) \ \dots \ u(t-n_b+1))^T$  and where element  $i$  in  $\theta$  corresponds to  $g(i-1)$ .

In the case of an elliptical input signal,  $\hat{\theta}_{LLS}$  is a consistent estimator of  $cg(k)$ , where  $c$  is some constant, but in the general case, this estimator is biased. However, by compensating for the deviation from an elliptical distribution using the weighting method, a consistent weighted LLS impulse response estimator can be obtained also for an arbitrary input signal. The advantages of the weighting method in this application are illustrated in the following example.

---

**Example 4.1**

---

Consider a Hammerstein system

$$y(t) = G(q)f(u(t)) + w(t),$$

where

$$G(q) = \frac{3}{1 - 0.5q^{-1}}, \quad f(x) = x^3 - 0.7x - 0.6 \operatorname{sign}(x)$$

and where  $w(t)$  is a white Gaussian disturbance signal with  $E(w(t)) = 0$  and standard deviation 0.25. The input signal to this system is generated as

$$u(t) = (1 + 0.5q^{-1})e(t),$$

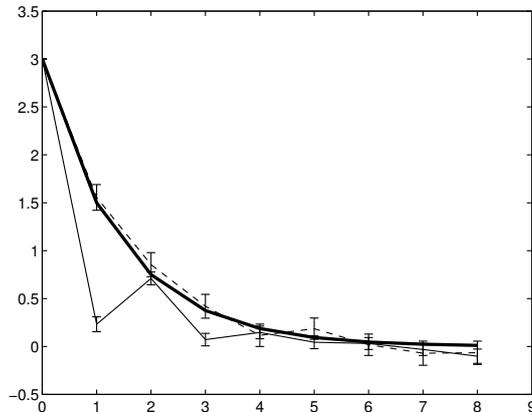
where  $e(t)$  is a white noise signal with uniform distribution on the interval  $[-1, 1]$ . Without the measurement noise in the system, this input signal gives an output signal with standard deviation 2.49, i.e., the signal to noise ratio is 10 (20 dB).

The response of this system to a particular realization of the input signal has been simulated and a dataset with 20 000 input and output measurements has been collected. The LLS impulse response estimate has been computed from this dataset and the result is shown in Fig. 1, together with the true impulse response. Furthermore, confidence intervals corresponding to two standard deviations are also presented there. The standard deviations of the impulse response estimator have been estimated from 100 Monte Carlo simulations of the complete experiment, where both the input and noise realizations have been varied. As can be seen in Fig. 1, the bias in the LLS estimate is quite large. Since this bias depends on the nonlinearity in the system and the non-elliptical input signal distribution, it will not decrease much if the number of measurements is increased.

The weighting method based on importance sampling and the approach described in Section 3 has also been applied to the dataset. The support of  $p_D$  has been chosen as in (7) using  $s_e = 6$ . In this case, 14% of the original  $\varphi$  measurements were in the ellipsoid. An elliptical random variable has been designed using (8). Here, the distribution of the  $r$  variable was estimated using a histogram with  $n_h = 20$  bins.

The weights were computed from  $P = 10\,000$  realizations of the elliptically distributed random variable. The resulting weighted LLS impulse response estimate is shown in Fig. 1 together with confidence intervals. As can be seen there, the weighting method has reduced the bias considerably.  $\square$

---



**Figure 1:** The impulse response of the linear part of the Hammerstein system in Example 4.1 (thick solid), the linear least-squares estimate (thin solid) and the weighted linear least-squares estimate (dashed) with confidence intervals corresponding to two times the standard deviation. For comparison reasons, the estimates have been scaled such that the direct term is equal to 3.

## 5 Other Possible Applications

The fact that it is hard to control signal distributions in nonlinear systems is a common problem in system identification. In this section, three identification and modeling problems in which the weighting method might be applied will be described.

### 5.1 NARX Systems

Consider an NARX system

$$y(t) = f(\varphi(t)) + e(t), \quad (10)$$

where

$$\varphi(t) = \begin{pmatrix} -y(t-1) \\ \vdots \\ -y(t-n_a) \\ u(t-n_k) \\ \vdots \\ u(t-n_k-n_b+1) \end{pmatrix},$$

$e(t)$  is a white noise signal that is independent of  $\varphi(s)$  for  $s \leq t$  and  $u(t)$  is the input signal. The weighting method can be applied directly to this system in order to facilitate the estimation of a suitable approximate model.

### 5.2 Systems with Periodic Inputs

Consider a nonlinear system with a random  $P$ -periodic input  $u(t)$ , i.e., an input with the property that  $u(t-P) = u(t)$ . Assume that the system output  $y(t)$

consists of two parts,  $y(t) = y_{nf}(t) + e(t)$ , where  $y_{nf}(t)$  is the noise-free output of the system, which is determined completely by the input signal, and  $e(t)$  is a disturbance that is independent of the input signal. Assume that the system has the property that a periodic input eventually results in a periodic noise-free output  $y_{nf}(t)$ , and that the input signal has been applied since  $t = -\infty$  such that all nonperiodic transients have disappeared.

Let  $\varphi(t) = (u(t) \ u(t-1) \ \dots \ u(t-P+1))^T$  such that the nonlinear system can be written

$$y(t) = f_{TD}(\varphi(t)) + e(t) \quad (11)$$

for some function  $f_{TD}$ . The weighting method can be applied directly to data from this system, but it might be better to consider the equivalent frequency domain description of the system in some cases.

The frequency domain version of (11) can be written

$$\mathbf{Y} = f_{FD}(\mathbf{U}) + \mathbf{E}, \quad (12)$$

where  $\mathbf{Y}$ ,  $\mathbf{U}$  and  $\mathbf{E}$  are vectors containing the components of the discrete Fourier transforms (DFTs) of the signals  $y(t)$ ,  $u(t)$  and  $e(t)$  for  $t = 0, 1, \dots, P-1$ , respectively. The frequency domain description is particularly useful if the input contains only a small number of frequency components, i.e., if most of the components  $U(n)$  of  $\mathbf{U}$  are zero. In this case, it might be better to view the output as a function of the amplitudes  $A_{n_i}$  and phases  $\psi_{n_i}$  of the signal components at the excited frequencies. By using these stochastic variables for defining the input signal, the frequency domain description of the system can be written

$$\mathbf{Y} = f_{AP}(\mathbf{A}, \mathbf{\Psi}) + \mathbf{E}, \quad (13)$$

where  $\mathbf{A}$  and  $\mathbf{\Psi}$  are vectors containing the  $A_{n_i}$  and  $\psi_{n_i}$  variables, respectively. The weighting method can be applied to the system (13) in order to compensate for an undesirable distribution of the amplitudes and phases. The distribution of  $\mathbf{\Psi}$  is particularly important when linear approximations of the nonlinear system (13) are estimated (Pintelon and Schoukens, 2001; Schoukens et al., 1998). If  $\mathbf{A}$  and  $\mathbf{\Psi}$  are independent, it is possible to use the weighting method just for  $\mathbf{\Psi}$ .

### 5.3 Systems Written on State-Space Form

Consider a system written on state-space form

$$\begin{aligned} x(t+1) &= f(x(t), u(t)) + w(t), \\ y(t) &= h(x(t), u(t)) + v(t), \end{aligned}$$

where  $w(t)$  and  $v(t)$  are system and measurement noises, respectively. Assume that the states  $x(t)$  can be measured as well as the input and output signals and consider the problem of estimating the functions  $f$  and  $h$ . This can be quite a challenging problem for a multiple input multiple output (MIMO) system with many state variables.

If the vector  $z(t) = (x(t) \ u(t))^T$  could be guaranteed to have a Gaussian, or at least an elliptical, distribution, some methods for dimension reduction could be applied as an attempt to reduce the complexity of the modeling problem

(Li, 1991, 1992). However, the distribution of  $z(t)$  is not elliptical in most applications. In these cases, the weighting method might be used for extending the applicability of the dimension reduction methods.

## 6 Discussion

Obviously, the main advantage with the weighting method is that the conditions on the signal distributions might be relaxed in some estimation problems. For example, the weighting method might be particularly useful for approximate identification of systems in closed-loop or of cascade connected nonlinear systems, since it is hard to control the input distribution in both these cases. The main limitation of the weighting method is that very large datasets are required if the input dimension ( $\dim(x)$  in (1)) is high and some of the weights are large.

On a conceptual level, the weighting method can be a convenient tool when linking approximation results in a stochastic framework to the ones obtained with a deterministic setting (Mäkilä, 2004, 2005, 2006; Mäkilä and Partington, 2003, 2004). It is obvious that the stochastic and deterministic frameworks are closely linked since, for a fixed dataset, it does not matter whether the input is viewed as purely deterministic or as a realization of a stochastic process. Any model estimated from the dataset using a particular numerical method will be the same no matter which theoretical framework is used, and the different viewpoints are just ways to describe how the dataset has been generated and how the results should be interpreted. The weighting that is used to obtain a consistent estimator of the optimal parameters for a different signal distribution in the stochastic framework can be used in exactly the same way in the deterministic framework. Hence, all stochastic approximation results that can be obtained by the use of special signal distributions can be obtained also in a deterministic framework without introducing a concept similar to the probability density function. The only necessary deterministic condition is that the signal components are dense in the region of interest.

In most estimation problems, the choice of cost function is determined completely by the statistical properties of the measurement noise. However, many estimation frameworks do not consider approximation issues, i.e., they assume that the true system (1) can be described exactly by the model (2) for some parameter vector  $\theta = \theta_0$  such that  $f(x) = g(x, \theta_0)$  for all  $x \in \mathbb{R}^m$ . This is often a rather unrealistic assumption and in general, the model residuals contain an input-dependent term coming from the unmodeled part of  $f$  for all values of  $\theta$ . This term acts like an additional noise term that causes the parameter estimate to vary with the input realization even when there is no measurement noise  $e$  in the true system (1). It is hard to motivate why the cost function should be determined only by the measurement noise in these cases, and the weighting method can be viewed as a way to take into account also the approximation errors.

## 7 Conclusions

The weighting method, which is based either on importance sampling or the rejection method, seems to give solutions to and insight into a number of differ-

ent approximation problems in nonlinear system identification. In particular, it relaxes conditions on the input distributions needed to guarantee that the approximations will have some desirable properties. Here, it has been shown that the weighting method can be used to obtain consistent estimates of the impulse responses of Hammerstein systems for arbitrary input signals. However, further work is required in order to investigate the usefulness of the method in the other identification problems that have been mentioned here.

## 8 Acknowledgments

Most of the results that are presented in this paper were obtained during a post-doc year at Vrije Universiteit Brussel in Brussels, Belgium. This work has been supported by the FWO-Vlaanderen, the Flemish community (Concerted action ILiNoS) and the Belgian government (IAP-V/22).

## References

- T. S. Atalik and S. Utku. Stochastic linearization of multi-degree-of-freedom non-linear systems. *Earthquake Engineering and Structural Dynamics*, 4:411–420, 1976.
- R. J. Beckman and M. D. McKay. Monte Carlo estimation under different distributions using the same simulation. *Technometrics*, 29(2):153–160, 1987.
- S. A. Billings and S. Y. Fakhouri. Theory of separable processes with applications to the identification of nonlinear systems. *Proceedings of the IEE*, 125(9):1051–1058, 1978.
- S. A. Billings and S. Y. Fakhouri. Identification of systems containing linear dynamic and static nonlinear elements. *Automatica*, 18(1):15–26, 1982.
- D. R. Brillinger. Sliced inverse regression for dimension reduction: Comment. *Journal of the American Statistical Association*, 86(414):333, 1991.
- J. J. Bussgang. Crosscorrelation functions of amplitude-distorted Gaussian signals. Technical Report 216, MIT Research Laboratory of Electronics, Cambridge, Massachusetts, 1952.
- R. D. Cook. Principal Hessian directions revisited. *Journal of the American Statistical Association*, 93(441):84–94, 1998.
- R. D. Cook and C. J. Nachtsheim. Reweighting to achieve elliptically contoured covariates in regression. *Journal of the American Statistical Association*, 89(426):592–599, 1994.
- M. Enqvist and L. Ljung. Linear approximations of nonlinear FIR systems for separable input processes. *Automatica*, 41(3):459–473, 2005.
- I. W. Hunter and M. J. Korenberg. The identification of nonlinear biological systems: Wiener and Hammerstein cascade models. *Biological Cybernetics*, 55:135–144, 1986.

- K.-C. Li. Sliced inverse regression for dimension reduction. *Journal of the American Statistical Association*, 86(414):316–327, 1991.
- K.-C. Li. On principal Hessian directions for data visualization and dimension reduction: Another application of Stein’s lemma. *Journal of the American Statistical Association*, 87(420):1025–1039, 1992.
- K.-C. Li and N. Duan. Regression analysis under link violation. *The Annals of Statistics*, 17(3):1009–1052, 1989.
- P. M. Mäkilä. On optimal LTI approximation of nonlinear systems. *IEEE Transactions on Automatic Control*, 49(7):1178–1182, 2004.
- P. M. Mäkilä. LTI modelling of NFIR systems: near-linearity and control, LS estimation and linearization. *Automatica*, 41(1):29–41, 2005.
- P. M. Mäkilä. On robustness in control and LTI identification: Near-linearity and non-conic uncertainty. *Automatica*, 42(4):601–612, 2006.
- P. M. Mäkilä and J. R. Partington. On linear models for nonlinear systems. *Automatica*, 39(1):1–13, 2003.
- P. M. Mäkilä and J. R. Partington. Least-squares LTI approximation of nonlinear systems and quasistationarity analysis. *Automatica*, 40(7):1157–1169, 2004.
- A. H. Nuttall. Theory and application of the separable class of random processes. Technical Report 343, MIT Research Laboratory of Electronics, Cambridge, Massachusetts, 1958.
- R. Pintelon and J. Schoukens. *System Identification: A Frequency Domain Approach*. IEEE Press, New York, 2001.
- W. L. Poston, E. J. Wegman, C. E. Priebe, and J. L. Solka. A deterministic method for robust estimation of multivariate location and shape. *Journal of Computational and Graphical Statistics*, 6(3):300–313, 1997.
- J. Schoukens, T. Dobrowiecki, and R. Pintelon. Parametric and nonparametric identification of linear systems in the presence of nonlinear distortions - A frequency domain approach. *IEEE Transactions on Automatic Control*, 43(2):176–190, 1998.
- J. Schoukens, R. Pintelon, T. Dobrowiecki, and Y. Rolain. Identification of linear systems with nonlinear distortions. *Automatica*, 41(3):491–504, 2005.
- D. M. Titterton. Optimal design: Some geometrical aspects of d-optimality. *Biometrika*, 62:313–320, 1975.

	<b>Avdelning, Institution</b> Division, Department  Division of Automatic Control Department of Electrical Engineering	<b>Datum</b> Date  2007-10-10
	<b>Språk</b> Language  <input type="checkbox"/> Svenska/Swedish <input checked="" type="checkbox"/> Engelska/English  <input type="checkbox"/> _____	<b>Rapporttyp</b> Report category  <input type="checkbox"/> Licentiatavhandling <input type="checkbox"/> Examensarbete <input type="checkbox"/> C-uppsats <input type="checkbox"/> D-uppsats <input checked="" type="checkbox"/> Övrig rapport <input type="checkbox"/> _____
<b>URL för elektronisk version</b>  <a href="http://www.control.isy.liu.se">http://www.control.isy.liu.se</a>		LiTH-ISY-R-2829
<b>Titel</b> A Weighting Method for Approximate Nonlinear System Identification Title		
<b>Författare</b> Martin Enqvist Author		
<b>Sammanfattning</b> Abstract  <p>Many approximation results in nonlinear system identification concern particular signal distributions. This seems to limit the applicability of these results to cases where the relevant signals have these distributions. However, by using a weighting method that modifies the cost function used in the identification method, the available approximation results can be used also for rather general classes of signal distributions. The purpose of this paper is to describe this weighting approach and to point at some interesting application areas within nonlinear system identification. In particular, it will be described how the impulse response of a Hammerstein system can be estimated consistently for an arbitrary input signal.</p>		
<b>Nyckelord</b> Keywords      Nonlinear systems, Approximation, System identification, Importance weighting		