IFAC

Blind Maximum Likelihood Identification of Hammerstein Systems

L. Vanbeylen, R. Pintelon, J. Schoukens

Vrije Universiteit Brussel, dept. ELEC, Pleinlaan 2, B1050 Brussels, Belgium (e-mail:laurent.vanbeylen@vub.ac.be)

Abstract: This paper is about the identification of discrete-time Hammerstein systems from output measurements only (blind identification). Assuming that the unobserved input is white Gaussian noise, that the static nonlinearity is invertible, and that the output is observed without errors, a Gaussian maximum likelihood estimator is constructed. Its asymptotic properties are analyzed and the Cramér-Rao lower bound is calculated. In practice, the latter can be computed accurately without using the strong law of large numbers. A two-step procedure is described that allows to find high quality initial estimates to start up the iterative Gauss-Newton based optimization scheme. The paper includes the illustration of the method on a simulation example.

1. INTRODUCTION: BLIND NONLINEAR MODELLING

Despite their structural simplicity – namely a static nonlinearity followed by a linear time invariant dynamic system followed (see Fig. 1) – Hammerstein nonlinear model structures have been effective in several application areas, where linear modelling has failed: e.g. microwave and RF technology (Greblicki, 1996; Prakriya and Hatzinakos, 1997), chemical processes (Eskinat *et al.*, 1991) and biology (Hunter and Korenberg, 1986). They can also be used in model predictive control (Wang and Henriksen, 1994; Zhu, 2000).
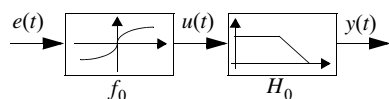

Fig. 1: Block diagram of a Hammerstein system. A static nonlinearity f_0 and a linear dynamic block H_0 are cascaded.

The Hammerstein model fits inside the family of block oriented nonlinear structures, which have been extensively studied over the past few decades (Billings and Fakhouri, 1982; Schoukens *et al.*, 2003). Most of these identification approaches assume the availability of both input and output measurements of the system. However, in a real-world situation – as frequently in sensor and measurement applications, such as operational modal analysis (Peeters and De Roeck, 2001) – one often does not have access to the system input. In this case, blind identification of the nonlinear system becomes the only option. Without measurements of the input, and with little prior knowledge and hypotheses about the system and the input, our goal is to identify the parameters of the static nonlinear and the linear dynamic block. Blind identification (closely related to blind inversion and equalization) for linear systems has a long history with many theoretical results and applications, especially in telecommunications. Comprehensive reviews can be found in (Abed-Meraim *et al.*, 1997) and (Tong and Perreau, 1998). In the linear case (i.e. without nonlinearity), the problem is also known as classical ARMA time-series modelling (e.g. Ljung,

1999). However, the present knowledge on blind identification of nonlinear systems is rather limited; the table below gives an overview and comparison of the most recently available blind identification methods for nonlinear systems.

Model (approach)	Assumptions + properties of the method
Volterra-Hammerstein (Kalouptsidis and Koukoulas, 2005): cumulant-based	• Input: Gaussian and white • Robust to low-order moving-average output noise • Consistent • Parametric
Hammerstein-Wiener (Bai, 2002)	• Input: zero-order-hold (piecewise constant) • Oversampled output • Invertible output nonlinearity admitting a polynomial representation • Not robust to output noise • Parametric
Wiener-Hammerstein (Prakriya and Hatzinakos, 1995): polyspectral slices	• Input: circularly symmetric Gaussian • Polynomial nonlinearity • First linear system has minimum phase • Robust to circular symmetric output noise • Nonlinearity is not identified • Nonparametric
Wiener (Taleb *et al.*, 2001): mutual information	• Input: non-Gaussian iid • Invertible nonlinearity and filter • Not robust to output noise • Quasi-nonparametric
Wiener (Vanbeylen *et al.*, 2007a): maximum likelihood	• Input: Gaussian and white • Invertible nonlinearity and filter • Not robust to output noise • Consistent and efficient • Parametric

In this paper, similarly as was done for Wiener systems – which have a similar structure as Hammerstein systems but blocks in reverse order – in (Vanbeylen *et al.*, 2007a), a maximum likelihood estimator (MLE) is proposed, assuming a white unobserved Gaussian input signal, errorless output observations (this assumption is restrictive but necessary for the analysis), and an invertible nonlinearity. The major advantages of the maximum likelihood approach are the consistency (convergence to the true value as the number of data tend to infinity), the asymptotic normality, the asymptotic

unbiasedness and the asymptotic efficiency (asymptotically the smallest variance) of the estimates. To the knowledge of the authors, no MLE based method is currently available for blind identification of Hammerstein systems.

The main contributions of this paper are the following:

(i) generation of high quality initial estimates for both the linear and the nonlinear part of a Hammerstein system
(ii) presentation of the Gaussian MLE for Hammerstein systems
(iii) calculation of the Cramér-Rao lower bound
(iv) verification of the theoretical results by means of a simulation example.

The method is similar to the Wiener case discussed in (Vanbeylen *et al.*, 2007a). This explains that this exposition on Hammerstein systems follows approximately the same lines as the MLE for Wiener systems.

2. ASSUMPTIONS AND CHOICES

The assumptions and choices are very close to those of the Wiener case. For the sake of brevity, we briefly mention them without discussion (Vanbeylen *et al.*, 2007a).

2.1. Class of discrete-time Hammerstein systems considered

Hammerstein systems are defined as a cascade of a static (memoryless) nonlinearity and a linear time-invariant dynamic system (LTI) H_0 (Fig. 1). In general, the static nonlinearity can be characterized by its mapping function from e (input signal) to u (intermediate signal): $u = f_0(e)$. The LTI system can be characterized by its transfer function $H_0(z)$.

Assumption 1: *(the class of discrete-time Hammerstein systems considered)*

1.a $u = f_0(e)$ is a monotonic, hence bijective function.
1.b Moreover, the function f_0 may not increase too fast: $\log|f_0(e)| = O(|e|^{2-\varepsilon})$ as $e \to \pm\infty$ with $\varepsilon > 0$.
1.c Its derivative $\partial f_0/\partial e$ is bounded almost everywhere. It is zero on a countable set, such that for every point of the set, there exists a differentiability order s such that $\partial^r f_0/\partial e^r = 0 \quad \forall r : 1 \le r < s$ at that point, and $0 < |\partial^s f_0/\partial e^s| < \infty$ over an open interval containing the point.
1.d $H_0(z)$ is a causal, stable and inversely stable monic transfer function.

2.2. Parameterization

The parameter vectors characterizing the linear and the nonlinear part are respectively denoted by θ_L and θ_{NL}. The nonlinear function is parameterized inversely: $e = g(u, \theta_{NL})$. On the other hand, the LTI is parameterized in the numerator and denominator coefficients of its transfer function $H(z, \theta_L)$:

$$H(z, \theta_L) = \frac{C(z, \theta_L)}{D(z, \theta_L)} = \frac{1 + \sum_{r=1}^{n_c} c_r z^{-r}}{1 + \sum_{r=1}^{n_d} d_r z^{-r}} \tag{1}$$

with $\theta_L^T = [c_1, c_2, ..., c_{n_c}, d_1, d_2, ..., d_{n_d}]$.

Moreover, the 'inverse' nonlinear function g is assumed to be twice differentiable w.r.t. θ_{NL}. Note that the inverse

parameterization enables a direct inversion (or equalization) of the Hammerstein system by means of a Wiener system formed by H^{-1} in series with g. (This could be useful for, e.g., calibration applications.) Finally, we define a global parameter vector by stacking the parameters of both subblocks onto each other:

$$\theta^T = [\theta_L^T, \theta_{NL}^T] \tag{2}$$

2.3. Stochastic framework

Assumption 2: *(stochastic framework)*

2.a The unknown, unobserved input $e(t)$ is zero mean, white Gaussian noise with unknown variance $\lambda_0 > 0$.
2.b The output $y(t)$ is known exactly (i.e. observed without errors).

3. THE GAUSSIAN MAXIMUM LIKELIHOOD ESTIMATOR

3.1. The negative log-likelihood function

Theorem 1: Under Assumptions 1 and 2, the conditional Gaussian negative log-likelihood function of the observations

$$y_N^T = y^T = [y(0), ..., y(N-1)] \tag{3}$$

given the model parameters θ and input variance λ, is:

$$L(y|\theta, \lambda) = \frac{N}{2}\log 2\pi + \frac{N}{2}\log\lambda + \tag{4}$$

$$\sum_{t=0}^{N-1} \frac{(g(H^{-1}(q, \theta_L)y(t), \theta_{NL}))^2}{2\lambda} - \sum_{t=0}^{N-1} \log|g'(H^{-1}(q, \theta_L)y(t), \theta_{NL})|$$

with q the forward shift operator ($qx(t) = x(t+1)$), with $(...)'$ denoting the first order partial derivative w.r.t. the first argument of the function: i.e. $g'(x, y) = \frac{\partial g}{\partial x}(x, y)$.

Here conditional means "given the initial conditions of the LTI part". However, asymptotically, the conditional MLE equals the true MLE.

Proof: Follow the same lines as in Appendix A of (Vanbeylen *et al.*, 2007a). The proof is based on the classical expressions of the log-likelihood for ARMA models and on the transformation formula of probability density functions through nonlinear mappings.

It remains possible to eliminate the input variance λ analytically from the log-likelihood cost function, by setting $\partial L(y|\theta, \lambda)/\partial\lambda = 0$, and solving for λ. This yields the following result:

$$\lambda = \frac{1}{N}\sum_{t=0}^{N-1}(g(H^{-1}(q, \theta_L)y(t), \theta_{NL}))^2 \tag{5}$$

and the likelihood-based cost function boils down to the sum-of-squares expression

$$V(y, \theta) = g_N^2(y, \theta)\sum_{t=0}^{N-1}(g(H^{-1}(q, \theta_L)y(t), \theta_{NL}))^2 \tag{6}$$

with $g_N(y, \theta) = \exp\left(-\frac{1}{N}\sum_{t=0}^{N-1}\log|g'(H^{-1}(q, \theta_L)y(t), \theta_{NL})|\right)$.

Stated differently, the cost function $V(y, \theta)$ can be rewritten as a sum of squares of residuals $\varepsilon(t, \theta)$:

$$\varepsilon(t, \theta) = g(H^{-1}(q, \theta_L)y(t), \theta_{NL})g_N(y, \theta) \qquad (7)$$

The sum-of-squares property is useful, since it allows to use efficient Gauss-Newton based routines for the optimization.

At this point, it is also interesting to note that the first two differences arise in contrast to the Wiener case: first, the factor g_N no longer only depends on the data and the parameter estimates of the nonlinear part; also θ_L pops up in the equations, and g_N becomes a function of the full parameter vector θ. This results in slightly more complicated expressions for the Jacobian matrix. Second, in contrast to the Wiener case, we prefer to stay in the time domain to perform the calculations, since it is the natural domain in which the reconstructed input signal $g(H^{-1}(q, \theta_L)y(t), \theta_{NL})$ is found.

3.2. Identifiability

As it is the case for the blind identification of Wiener systems, the actual parameterization of the Hammerstein is not identifiable as such. It can be seen that e.g. the displacement of a gain factor from the input variance to the linear or nonlinear block would result in the same output signal. This motivates the need of introducing a set of identifiability conditions.

Assumption 3: *(identifiability conditions)*

  3.a $H_0(z, \theta_L)$ has no common pole-zero pairs.

  3.b $H_0(z, \theta_L) = \dfrac{C_0(z, \theta_L)}{D_0(z, \theta_L)}$ is monic (this means that the constants in the C_0 and D_0 polynomials are 1).

  3.c $g_0'(u_1, \theta_{NL}) = 1$ for some $u_1 \in \mathbb{R}$.

3.3. Result

As a consequence of the preceding paragraphs, the following holds:

Result: Under Assumptions 1 and 2, the maximum likelihood estimator $\hat{\theta}$ of the system parameters minimizes the cost function (6), subject to the constraints given in Assumption 3. Since the cost function has been written as a sum-of-squares, its minimizer

$$\hat{\theta} = \arg\min_{\theta} V(y, \theta) \qquad (8)$$

can be calculated in a numerical stable way via the classical Gauss-Newton based iterative schemes (Fletcher, 1991). The most likely input variance $\hat{\lambda}$ corresponding to this parameter set $\hat{\theta}$ can be found by evaluating (5).

3.4. Asymptotic properties

To study the asymptotic $(N \to \infty)$ properties of the MLE, standard assumptions are needed for the true model, the parametric model and the excitation.

Assumption 4: *(consistency)*

  4.a The true model is within the proposed model set.

  4.b The normalized cost function $V(y_N, \theta)/N$ has continuous second order derivatives w.r.t. θ in a compact (i.e. closed and bounded) set Θ_r for any N, infinity included. The compact set Θ_r is constructed such that it contains a unique global minimum of $V(y_N, \theta)/N$, which is an interior point (i.e. not on the boundaries) of Θ_r.

  4.c There exists an N_0 such that for any $N \geq N_0$, infinity included, the Hessian of the expected value of the normalized cost function $V(\theta) = \mathbb{E}\{V(y_N, \theta)/N\}$, subject to the constraints given in Assumption 3, is regular (i.e. invertible) at the unique global minimizer of $V(\theta)$ in Θ_r.

Under Assumptions 1-4, and since the MLE satisfies the standard conditions, amongst others, (i) the likelihood function is based on iid Gaussian random variables, (ii) the number of model parameters does not increase with the amount of data N, $\hat{\theta}$ is strongly consistent, asymptotically efficient, asymptotically unbiased, and asymptotically normally distributed (Caines, 1988, p. 317, theorem 3.2).

4. GENERATION OF INITIAL ESTIMATES

As already mentioned, an iterative algorithm can be used to find the MLE. However, the generation of initial estimates is still necessary. This is done in two steps: in a first step, the LTI part H is estimated. This allows to calculate an estimate of the intermediate signal $u(t)$ (Fig. 1), from which – in a second step – a nonparametric estimate of the nonlinearity f is generated.

4.1. Linear part

Making use of Assumptions 1.b and 2.a, it follows that the intermediate signal is identically and independently distributed and has bounded moments of all orders. The initial estimates $\hat{\theta}_L$ can therefore be found by applying the blind ARMA frequency domain estimator (noise model of Pintelon and Schoukens, 2006) or classical time domain algorithms (Ljung, 1999) to the observed output signal $y(t)$. After inverse filtering of $y(t)$ with $H^{-1}(q, \hat{\theta}_L)$, one finds an estimate of the intermediate signal $\hat{u}$. The order of H should be selected as to whiten the power spectrum of the reconstructed signal $\hat{u}$.

4.2. Nonlinear part

Like in the Wiener case, it is possible to come up with a nonparametric estimate of the nonlinearity. The method is based on the increasing property of the nonlinear function g. As a consequence, it preserves order relationships, and for every pair (e_0, u_0) satisfying $e_0 = g(u_0)$, we have that

$$F_E(e_0) = \mathbb{P}(E \leq e_0) = \mathbb{P}(U \leq u_0) = F_U(u_0) \qquad (9)$$

with E and U representing the random processes corresponding to e and u, and F_E and F_U representing their corresponding cumulative distribution functions. Since the distribution of U is known (besides an unknown variance scaling factor), by equating the corresponding quantiles using the empirical distribution function of $\hat{u}$, it becomes possible to reconstruct (i.e. generate an estimate of) the values of the input signal (within a scaling factor). Hence we have a nonparametric estimate of the nonlinearity.

Afterwards, it is possible to calculate the variances of the reconstructed input at each time instant, which are used as weights in a weighted least squares step to find an initial value of $\hat{\theta}_{NL}$ to start the actual MLE optimization with.

5. MINIMIZATION OF THE COST FUNCTION

Gauss-Newton based iterative algorithms for optimizing sum-of-squares cost functions require use of the Jacobian matrix, defined as the partial derivatives of the residuals $\varepsilon(t, \theta)$ (7) w.r.t. the parameters: $J_{ti}(y, \theta) = \partial \varepsilon(t, \theta) / \partial \theta_i$. Its calculation is straightforward, and, hence, left out of this paper. In the general case, the practical implementation requires the calculation of all mixed second order derivatives $\partial^2 g(u, \theta_{NL}) / (\partial u \partial \theta_i)$, which can be quite a tedious task. Therefore, we propose a simple parameterization for the nonlinearity.

5.1. Parameterization of the nonlinearity

To simplify the computations, a model that is linear-in-the-parameters is used to describe the nonlinearity $g(y)$:

$$e = g(u, \theta_{NL}) = \boldsymbol{f}(u)^T \theta_{NL} \qquad (10)$$

with $\theta_{NL}^T = [\theta_{NL, 1}, ..., \theta_{NL, M}]$ the parameter vector of the nonlinear block, and with $\boldsymbol{f}(u)^T = [f_1(u), ..., f_M(u)]$ the vector of basis functions. Moreover, this linear parameterization simplifies, considerably, the weighted least squares computation in the initial estimates algorithm; the problem is linear-in-the-parameters, and an iterative procedure is avoided.

5.2. Gauss-Newton procedure

The equations of J_{ti} make it possible to compute the Jacobian matrix from the data and the parameters, and to minimize the cost function $V(y, \theta)$. This minimization is performed by using a Levenberg-Marquardt algorithm. It yields the most likely parameter vector $\hat{\theta}$, given the output observations y, if the global minimum is found.

5.3. Practical implementation of the constraint

In practice, for the model (10), the constraint on the nonlinearity is not implemented as stated in Assumption 3.c, but by setting the two-norm of the nonlinear parameter vector in the linear parameterization to unity:

$$\sum_m \theta_{NL, m}^2 = \|\theta_{NL}\|_2^2 = 1 \qquad (11)$$

This norm-one constraint avoids fixing a particular coefficient to unity. The latter has the disadvantage that it could be detrimental to the numerical conditioning (of the Jacobian of the Gauss-Newton scheme) if the true value of the coefficient is zero. When performing the optimization, this constraint between the coefficients θ_{NL} is taken into account at each step of the Gauss-Newton procedure, by using the pseudo-inverse of

$$J_{\text{re}}(y, \theta) = \begin{bmatrix} \text{Re}(J(y, \theta)) \\ \text{Im}(J(y, \theta)) \end{bmatrix} \qquad (12)$$

for solving the normal equations, and then imposing (11) on the updated parameter vector.

6. CRAMÉR-RAO LOWER BOUND

6.1. Theoretic expressions

Since we know from subsection 3.4 that the estimator is asymptotically $(N \to \infty)$ efficient – which means the asymptotic covariance matrix equals the Cramér-Rao lower bound (CRB) – its calculation allows us to compute the uncertainties on the estimated model parameters. Note that the full model structure is overparameterized if no constraints are imposed (subsection 3.2), and that the negative log-likelihood function (4) remains invariant under the transformation

$$\lambda \to 1 , \; g \to g / \sqrt{\lambda} \qquad (13)$$

which preserves the statistical properties of the output observations. As follows from (Pintelon *et al.*, 1999), exactly the same result would have been obtained if another choice had been made to make the model structure identifiable (like e.g. gain of the nonlinearity left free, combined with $\lambda = 1$ and a monic H). Although the obtained Fisher information matrix (FIM) depends on this choice, the uncertainty bounds on invariants, like the statistical properties of the observations or the pole/zero positions, calculated via a certain chosen CRB, are independent of the specific choice of the imposed constraints. Hence, without any loss of generality, the computations of the FIM – which is exactly the inverse of the CRB matrix – may be performed by imposing $\lambda = 1$, thus simplifying the calculations considerably. In practice, transformation (13) (i.e. $\hat{\theta}_{NL}$ is replaced by $\hat{\theta}_{NL} / \sqrt{\hat{\lambda}}$) is applied before calculating the FIM with the expressions in the theorem given below.

Theorem 2: For the Gaussian MLE given in (8), and under the above Assumptions 1-3, and for polynomial nonlinearities of the form (10), asymptotically, the elements of the Fisher information matrix are given by:

(a) for $\theta_i \in \theta_L, \theta_j \in \theta_L$, $\text{Fi}_{i,j} = \mathbb{E}[(g')^2 + (g''/g')^2] \times$

$$\left(N \mu_U^2 H_0^{-2} \frac{\partial H_0}{\partial \theta_i} \frac{\partial H_0}{\partial \theta_j} + 2\sigma_U^2 \left(\sum_{k=0}^{N-1} \text{Re}(H_k^{-1} \frac{\partial H_k}{\partial \theta_i}) \text{Re}(H_k^{-1} \frac{\partial H_k}{\partial \theta_j}) \right) \right)$$

(b) for $\theta_i \in \theta_L, \theta_j \in \theta_{NL}$,

$$\text{Fi}_{i,j} = -N \mu_U H_0^{-1} \frac{\partial H_0}{\partial \theta_i} \cdot \mathbb{E}[f_j g' + f_j' g''/g'^2]$$

(c) for $\theta_i \in \theta_{NL}, \theta_j \in \theta_{NL}$, $\text{Fi}_{i,j} = N \cdot \mathbb{E}[f_i f_j + f_i' f_j'/(g')^2]$

Each of these symbols is a compact notation where, for notational simplicity, the dependencies on the parameters have been dropped. In these expressions, $\mu_U = \mathbb{E}[U]$, $\sigma_U^2 = \mathbb{E}[(U - \mu_U)^2]$, and $H_k = H(\exp(j2\pi k/N), \theta_L)$.
Proof: see Appendix A.

First, it is interesting to note that – in contrast to the Wiener case – the obtained CRB matrix can be calculated without having recourse to the strong law of large numbers to approximate the mathematical expectations. Since the expectation operators do not contain frequency domain expressions like H, no leakage effects can occur either. Second, it is seen that, in the case of an antisymmetric nonlinearity, μ_U is zero, and, therefore, the statistical coupling between the linear and nonlinear parts disappears.

6.2. Practical calculation

In practice, the mathematical expectations are calculated by numerical integration of the considered expressions weighted by the pdf of U denoted by $f_U(u)$. The integration limits are chosen as the values of u which correspond with $e = \pm 5\sqrt{\hat{\lambda}}$. The pdf $f_U(u)$ is found by differentiation of (9) w.r.t. u_0:

$$f_U(u) = f_E(g(u, \hat{\theta}_{NL}))|g'(u, \hat{\theta}_{NL})| \qquad (14)$$

$$\text{with } f_E(e) = \frac{1}{\sqrt{2\pi\hat{\lambda}}}\exp\left(-\frac{e^2}{2\hat{\lambda}}\right) \qquad (15)$$

The values of μ_U and σ_U^2 are calculated similarly by numerical integration.

7. OUTPUT NOISE

An important topic to bear in mind is the fact that the observations are assumed to be errorless (Assumption 2.b). In practice, it is highly likely that the presence of measurement noise will have an impact on the estimates. E.g. consider the particular case of a purely static nonlinear Hammerstein system without linear part ($H = 1$). In general, the noiseless output will be non-Gaussian. The effect of adding an independent noise source at the output will result in the convolution of both pdf's. Since it is the nonlinear function which determines the change of pdf from Gaussian (cf. equation (14)), it becomes clear that the estimates of the nonlinear function will be affected in this sense. Moreover, in the case of Gaussian output noise, the pdf of the noise corrupted output will look closer to Gaussian, and the 'strength' of the nonlinearity will be underestimated.

To summarize, as a consequence of output noise (i.e. the violation of Assumption 2.b), the method will still work, but a bias will pop up in the estimates. It can be shown that, asymptotically, this bias has the property to be – in a first order approximation – proportional to the variance of the additional unmodelled noise source.

Nevertheless, it is possible to formulate the MLE problem with Gaussian output noise, but it is not feasible at this moment to solve it. This is left for future research.

It is interesting to notice that the model is suited as a noise model as well, for the analysis of non-Gaussian and coloured residuals.

8. SIMULATION RESULTS

8.1. Setup: presentation of the example

In this section, the results of the method applied to a simulation example are discussed. The underlying Hammerstein system used in the simulations was given by a fifth-order polynomial inverse nonlinearity with powers of u as basis functions $f_m(u)$:

$$\begin{aligned}g_0(u) &= \theta_{NL,1} + \theta_{NL,2}u + \theta_{NL,3}u^3 + \theta_{NL,4}u^5 \\ &= 2 + u + 4u^3 + 2u^5\end{aligned} \qquad (16)$$

and a fourth order filter with transfer function

$$H_0(z^{-1}) = \frac{1 + 0z^{-1} - 0.49z^{-2} + 0.01z^{-3} + 0.065z^{-4}}{1 + 0.3676z^{-1} + 0.88746z^{-2} + 0.52406z^{-3} + 0.55497z^{-4}} \qquad (17)$$

A 500-run Monte-Carlo simulation with $N = 8192$ points each was performed, and the resulting parameter estimates were analyzed.

8.2. Results and validation

It was first noticed that, as expected, the MLE improves (i.e. lowers) the variances of the estimated parameters, compared to the initial estimates. Also some statistical hypothesis tests were performed on the Monte Carlo data at the 5% significance level:

- Kolmogorov-Smirnov: could not detect significant deviations from normality (12 out of 12 parameters).
- Deviation of the mean from the true value: for 11 out of the 12 parameters, no significant deviations could be detected.
- Sample variance vs. Cramér-Rao lower bound: could not detect significant (variance) deviations on any of the 12 components of the parameter vector. The preceding facts are shown graphically for the parameter $\theta_{NL,3}$ in Fig. 2.
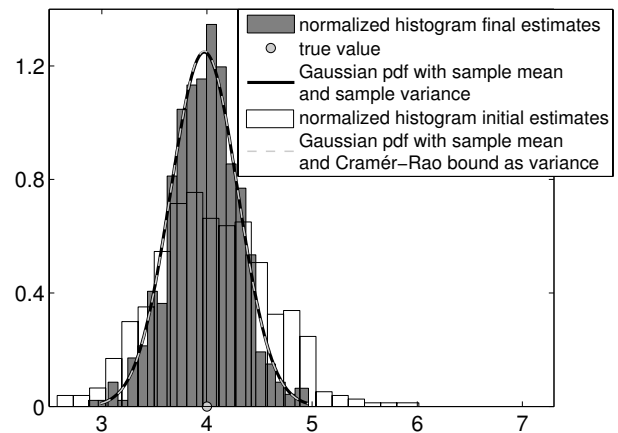

Fig. 2: Normalized histogram of the initial and final estimates for $\theta_{NL,3}$, illustrating the asymptotic unbiasedness, normality, and efficiency.

The other parameters give similar results. For this simulation example, the maximal ratio between the standard deviations of initial and final estimates is 1.86. This illustrates the high quality of the initial estimates.

Finally, the cross-validation of the results is performed (after the MLE optimization) by calculation, for a randomly selected realization, of:

- the pdf of the intermediate signal $f_U(u)$ based on the estimated nonlinear parameters $\hat{\theta}_{NL}$ and the estimated input variance $\hat{\lambda}$, as given in (14)-(15),
- and, the normalized histogram of the data sequence $\hat{u}$ obtained by inverse filtering of the output with the estimated parameters of the linear part $\hat{\theta}_L$.

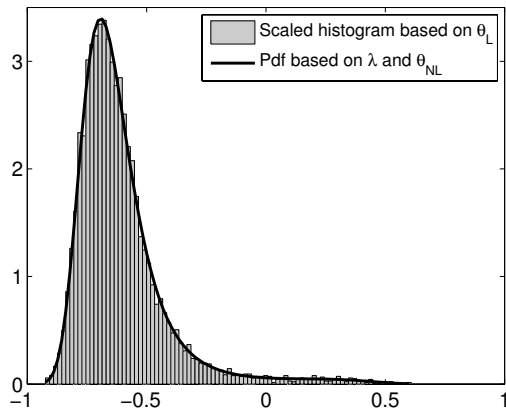Both are shown in Fig. 3, from which the excellent agreement between both can be seen.

Fig. 3: Cross-validation: calculation of the intermediate signal in two ways.

This cross-validation provides a practical way to the user to check whether the Hammerstein model with white Gaussian input is appropriate to model the signal provided to the algorithm.

9. CONCLUSIONS

Following the lines of the Wiener case, a maximum likelihood procedure for a blind identification of Hammerstein systems has been handled. Like in the Wiener case, the cost function can be rewritten as a sum of squares. Furthermore, a two-step algorithm for generating high-quality initial estimates has been presented.

APPENDIX A. FISHER INFORMATION MATRIX

First λ is set to unity in (4), and the second derivative is calculated. It is easily found that

$$\frac{\partial^2 L(\boldsymbol{y}|\theta,1)}{\partial\theta_i\partial\theta_j} = \sum_{t=0}^{N-1}\frac{\partial g}{\partial\theta_i}\frac{\partial g}{\partial\theta_j} + g\frac{\partial^2 g}{\partial\theta_i\partial\theta_j} - g'^{-1}\frac{\partial^2 g'}{\partial\theta_i\partial\theta_j} + g'^{-2}\frac{\partial g'}{\partial\theta_i}\frac{\partial g'}{\partial\theta_j} \quad (A1)$$

with $g = g(H^{-1}(q,\theta_L)y(t),\theta_{NL})$.

From this equation and parameterization (10), the FIM in case (c) follows immediately. Cases (a) and (b) are found after some very lengthy – but not so complicated – calculations, which are available as a technical note (Vanbeylen, 2007b).

ACKNOWLEDGEMENTS

REFERENCES

Abed-Meraim, K., W. Qiu and Y. Hua (1997). Blind system identification. *Proc. IEEE* **85**(8), 1310-1322.

Bai, E.W. (2002). Blind approach to the Hammerstein-Wiener model identification. *Automatica* **38**(6), 967-979.

Billings, S.A. and S.Y. Fakhouri (1982). Identification of systems containing linear dynamic and static nonlinear elements. *Automatica* **18**(1), 15-26.

Caines, P.E. (1988). *Linear stochastic systems*. Wiley, New York.

Eskinat, E., S.H. Johnson and W.L. Luyben (1991). Use of the Hammerstein models in identification of nonlinear systems. *Am. Inst. Chem. Eng. J.* **37**(2), 255-268.

Fletcher, R. (1991). *Practical methods of optimization*. Wiley, New York, 2nd edition.

Greblicki, W. (1996). Nonlinearity estimation in Hammerstein systems based on ordered observations. *IEEE Trans. Signal Processing* **44**(5), 1224-1233.

Hunter, I.W. and M.J. Korenberg (1986). The identification of nonlinear biological systems: Wiener and Hammerstein cascade models. *Biol. Cybern.* **55**, 135-144.

Kalouptsidis, N. and P. Koukoulas (2005). Blind identification of Volterra-Hammerstein systems. *IEEE Trans. Signal Processing* **53**(8), 2777-2787.

Ljung, L. (1999). *System identification - Theory for the user.* Prentice-Hall, Upper Saddle River, N.J., 2nd edition.

Peeters, B. and G. De Roeck (2001). Stochastic system identification for operational modal analysis: a review. *J. Dyn. Syst., Meas., Control* **123**(4), 659-667.

Pintelon, R., J. Schoukens, G. Vandersteen and Y. Rolain (1999). Identification of invariants of (over)parameterized models: finite sample results. *IEEE Trans. Automat. Contr.* **44**(5), 1073-1077.

Pintelon, R. and J. Schoukens (2006). Box-Jenkins identification revisited - part I: theory. *Automatica* **42**(1), 63-75.

Prakriya, S. and D. Hatzinakos (1995). Blind identification of LTI-ZMNL-LTI nonlinear channel models. *IEEE Trans. Signal Process.* **43**(12), 3007-3013.

Prakriya, S. and D. Hatzinakos (1997). Blind identification of linear subsystems of LTI-ZMNL-LTI models with cyclostationary inputs. *IEEE Trans. Signal Processing* **45**(8), 2023-2036.

Schoukens, J., J. G. Nemeth, P. Crama, Y. Rolain and R. Pintelon (2003). Fast approximate identification of nonlinear systems. *Automatica* **39**(7), 1267-1274.

Taleb, A., J. Solé and C. Jutten (2001). Quasi-nonparametric blind inversion of Wiener systems. *IEEE Trans. Signal Processing* **49**(5), 917-924.

Tong, L. and S. Perreau (1998). Multichannel blind identification: from subspace to maximum likelihood methods. *Proc. IEEE* **86**(10), 1951-1968.

Vanbeylen, L., R. Pintelon and J. Schoukens (2007a). Blind maximum likelihood identification of Wiener systems. In *Proc. European Control Conf.* (pp. 4625-4632), Kos, Greece.

Vanbeylen, L. (2007b). Calculation of the Fisher information matrix for blind ML Hammerstein identification. *Technical note*, http://wwwir.vub.ac.be/elec/Papers%20 on%20web/Papers/LaurentVBeylen/FisherHamm.pdf.

Wang, W. and R. Henriksen (1994). Generalized predictive control of nonlinear systems of the Hammerstein form. *Modeling Identif. and Control* **15**(4), 253-262.

Zhu, Y. (2000). Hammerstein model identification for control using ASYM. *Internat. Journ. Control* **73**(18), 1692-1702.