PARAMETER ESTIMATION IN PARTITIONED NONLINEAR STOCHASTIC MODELS

Ola Markusson and Håkan Hjalmarsson
Dept. of Signals, Sensors, & Systems
Royal Institute of Technology (KTH)
SE-100 44 Stockholm, SWEDEN
Tel: +46 8 790 7422, Fax: +46 8 790 7329
ola.markusson@s3.kth.se hjalmars@s3.kth.se

ABSTRACT

The application of maximum likelihood estimation and prediction error methods on dynamical systems require that it is possible to compute the innovations of the systems model which more or less implies that it must be possible to invert the system model. For nonlinear stochastic models this can be very difficult. In this contribution it is shown how this can be done very efficiently for a very rich class of nonlinear models by way of exact linearization. The method is illustrated on two non-trivial examples.

1 INTRODUCTION

Identification of nonlinear systems has attracted a lot of interest, see e.g. [7] and [12], for an overview of identification of nonlinear black-box models. Models of nonlinear stochastic systems are, for example, estimated using *a priori* knowledge about the system along with experimental data, so called grey-box identification, in [2]. Methods using measured input and output data for the identification of nonlinear stochastic models are described by, e.g, Bendat [1] and Chen [3].

In this contribution we consider estimation of nonlinear stochastic systems, i.e. systems where the inputs are white but not measured. It is well-known that the maximum likelihood (ML) principle provides estimates with nice statistical properties, e.g. asymptotic efficiency can be shown to hold under very general assumptions. For a linear time-discrete ARMA model

$$y(t) = \frac{C(q)}{A(q)} w(t)$$

where $y(t)$ and $w(t)$ are the output and white input respectively and where $C(q)$ and $A(q)$ are unknown polynomials in q^{-1}, the backward shift operator $q^{-1}w(t) = w(t-1)$, ML estimation amounts to inverting the model

$$e(t) = \frac{A(q)}{C(q)} y(t)$$

and minimizing the input e in some suitable norm which depends on the probability density function (PDF) of w.

ML estimation of nonlinear stochastic systems also requires that the system model is inverted. However, for nonlinear systems this is often a non-trivial task. The main contribution of this paper is to show that there is a very rich class of nonlinear dynamic models where the inversion can be easily performed.

This paper will start with a short introduction to inversion followed by a description of the class of models we are considering. Conditions for stability and initial conditions will be presented. Thereafter, a summary of the ML method will follow where emphasis will be made on the way we are calculating the likelihood. Finally, numerical examples and parameter estimation results will be presented.

2 INVERSION

Let us first give some background to inversion of nonlinear systems. The main condition for invertibility is a one-to-one relationship between the input and the output. Inversion of continuous time nonlinear systems and the relation to exact linearization is considered in, e.g., [5]. Discrete time systems are less commonly commented on, although in [8] an inversion algorithm is shown.

For a continuous time system, the *traditional* way of linearization by feedback is based on taking the derivative of the output function a number of times. In the discrete time case we will instead consider different time-shifts of the output. Let us illustrate the inversion technique by a simple example, see [8] for further details. Consider

$$x_1(t+1) = \alpha_1 x_2(t) \tag{1}$$

$$x_2(t+1) = \alpha_2 x_2(t) + \alpha_3 u^3(t) \tag{2}$$

$$y(t) = x_1(t) \tag{3}$$

where $x_{1,2}$ are the states, u is the input, y is the output and α_i are parameters. We can calculate u from the output by considering time-shifted values of y

$$y(t+1) = x_1(t+1) = \alpha_1 x_2(t) \tag{4}$$

$$y(t+2) = \alpha_1 x_2(t+1) = \alpha_1\alpha_2 x_2(t) + \alpha_1\alpha_3 u^3(t). \tag{5}$$

Now, we can calculate $u(t)$ from (5) as

$$u(t) = \left(\frac{1}{\alpha_1\alpha_3}[y(t+2) - \alpha_1\alpha_2 x_2(t)]\right)^{1/3}. \tag{6}$$

However, since $x_2(t)$ is included in (6) the state has to be updated which can be done as

$$x_2(t+1) = (1 - \frac{1}{\alpha_3})\alpha_2 x_2(t) + \frac{1}{\alpha_1\alpha_3} y(t+2). \tag{7}$$

Equation(7) is the result of inserting (6) into (2). Together, (6) and (7) can be used to compute u from y. Notice that we need $y(t+2)$ to compute $u(t)$ and, hence, the filter is noncausal. For off-line identification purposes, however, this is not a problem.

3 PARTITIONED NONLINEAR MODELS

We will now consider a class of models where the inverse can be implemented in a more direct way than described in the previous section. We consider the class of models where the nonlinear system, P, can be partitioned into a linear and a nonlinear part, see Figure 1. Thus, we can write the system P as

$$P = L + N \tag{8}$$

where L is the linear operator and N is the nonlinear operator. It can be noted that both L and N can be either static or dynamic, independent of each other. This model structure is naturally imposed by the Volterra functional representation of nonlinear systems but may also reflect the structure of actual systems. In [1], examples of partitioned models are shown.

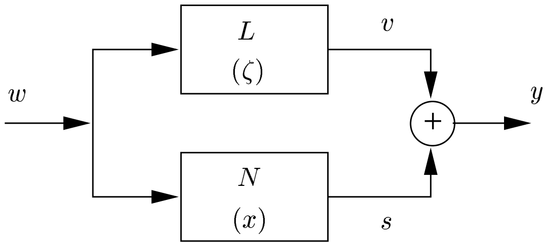

Figure 1: Block diagram of the parallel system. The states of L and N are represented by ζ and x. The outputs of the linear and the nonlinear part are represented by v and s, respectively.

As stated by Doyle et. al. [4], we can also write P as

$$P = L(I + L^{-1}N) \tag{9}$$

provided that the inverse of L exists. Furthermore, the inverse of P can be expressed as

$$P^{-1} = (I + L^{-1}N)^{-1}L^{-1} \tag{10}$$

provided that the inverse of $(I + L^{-1}N)$ exists. The existence of an inverse of the linear operator L is very easy to analyze and implement but it may not be easy to comment on the general existence of the inverse of $(I + L^{-1}N)$, [4]. However, if the above inverses exist then the inverse of P can be implemented as shown in Figure 2. As seen in Figure 2, the system can be inverted without explicitly inverting the whole system. It is sufficient to invert the linear part and filter the data through the nonlinear feedback system in Figure 2. This fact can now be explored in ML estimation.

Remark: In some special cases L can represent a simple and easily inverted nonlinear system.

4 STABILITY

Assuming invertibility there are two major concerns about the inverse system in Figure 2, namely stability and exponential forgetting of unknown initial conditions. Since, in general it is impossible to get consistent estimates of the initial states it is necessary that transients due to unknown initial die out sufficiently fast. Exponential forgetting is important for the convergence of the ML estimator.

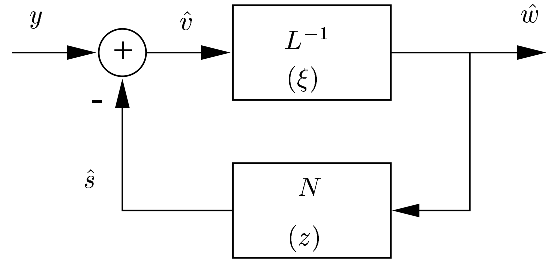

Figure 2: Realization of the nonlinear model inverse. The states of L^{-1} and N are represented by ξ and z. The output of the nonlinear part is represented by $\hat{s}$ and the input to L^{-1} by $\hat{v}$.

4.1 The Small Gain Approach

The stability of nonlinear systems is dealt with in terms of ℓ_p stability where $p \in [1, \infty]$. First we will restate some definitions and stability results given in [13].

Definition: Let S denote the linear space of all sequences $\{x_i\}_{i \leq 0}$, then for $p \in [1, \infty)$ we can define

$$\ell_p = \left\{ x \in S : \sum_{i=0}^{\infty} |x(i)|^p < \infty \right\} \text{and} \tag{11}$$

$$\ell_\infty = \{x \in S : x \text{ is a bounded sequence}\}. \tag{12}$$

Further on, ℓ_p is a subspace of S for each $p \in [1, \infty]$ and we can define the norms

$$\|x\|_p = \left[\sum_{i=0}^{\infty} |x(i)|^p \right]^{1/p}, \quad \forall x \in \ell_p \tag{13}$$

Now suppose X is a set and R is a binary relation on X. Then, according to [13], $x \in X$ is related to $y \in X$ if the ordered pair $(x, y) \in R$. R is ℓ_p stable if

$$(x, y) \in R, \quad x \in \ell_p \quad \Rightarrow y \in \ell_p \tag{14}$$

and R is ℓ_p stable with finite gain and zero bias if it is ℓ_p stable and if there exists a finite constant γ_p such that

$$(x, y) \in R, \quad x \in \ell_p \quad \Rightarrow \|y\|_p \leq \gamma_p \|x\|_p \tag{15}$$

where the ℓ_p gain of R is defined as

$$\gamma_p(R) = \inf\{\gamma_p : \text{provided (15) holds}\} \tag{16}$$

The small gain stability approach, [13], is providing a necessary condition for the inverse system to be stable. In this case we consider the system in Figure 2 and suppose that both L^{-1} and N are causal and ℓ_p stable with finite gains and zero bias. We also denote $\gamma_{p1} = \gamma_p(L^{-1})$ and $\gamma_{p2} = \gamma_p(N)$. Under these circumstances the system in Figure 2 is ℓ_p stable, by the small gain stability approach, if

$$\gamma_{p1}\gamma_{p2} < 1. \tag{17}$$

It should be noted that this condition can be quite conservative in some cases.

4.2 Dependence of Unknown Initial Conditions

When a sequence $y^M = (y(1), ..., y(M))$ is used to compute an estimate of the input sequence, the states of the inverse system are unknown. Therefore, the output will not be the exact inverse of the input for all samples. Here, we will briefly describe how the conditions for when the transients will die out boil down to the stability conditions of a linear time-varying (LTV) system.

To analyze the dependence of the unknown initial conditions of the inverse model we consider the differences

$$\rho(t) = \zeta(t) - \xi(t); \quad \Delta(t) = z(t) - x(t) \tag{18}$$

$$\sigma(t) = s(t) - r(t); \quad \delta(t) = \hat{w}(t) - w(t) \tag{19}$$

where $\zeta(t)$, $x(t)$ and $\xi(t)$, $z(t)$ are the states of the linear and the nonlinear system for the forward and the inverse system. The input to the forward system and the output of the inverse system are denoted $w(t)$ and $\hat{w}(t)$. Finally, the output of the nonlinear parts are denoted by $s(t)$ and $\hat{s}(t)$, see also Figure 1 and 2 where the states of each block are given within parenthesis. In this section we assume that the parameters of the forward and the inverse system are the same such that all the differences in (18) and (19) should go to zero.

It can be shown, although omitted here due to space limitations, that the state equations of (8) and (9) (see also Figure 1 and 2) can be used to describe a dynamical system based on the differences Δ and ρ as

$$\Delta(t+1) = \Gamma(\Delta(t), \rho(t)) \tag{20}$$

$$\rho(t+1) = \Sigma(\Delta(t), \rho(t)). \tag{21}$$

Hence, we get a system which has no driving term but typically has nonzero initial states due to unknown initial conditions. Clearly, we want the differences due to unknown initial conditions to vanish. Thus, we want the origin to be a stable equilibrium of the system. A problem is that it may be hard to analyze the stability of (20) and (21) since they are nonlinear functions.

However, to specify the conditions for a uniformly exponentially stable system we can use the following fact: If the linearized system has a uniformly exponentially stable equilibrium at the origin, then the nonlinear system is uniformly exponentially stable at the origin, [13].

We can describe the linearized system by

$$\Delta(t+1) = \Gamma_{\Delta(t)}(0,0)\Delta(t) + \Gamma_{\rho(t)}(0,0)\rho(t) \tag{22}$$

$$\rho(t+1) = \Sigma_{\Delta(t)}(0,0)\Delta(t) + \Sigma_{\rho(t)}(0,0)\rho(t) \tag{23}$$

where Γ_x and Σ_x denotes the derivative with respect to x, ($x = \Delta(t)$ or $\rho(t)$). Thus, the studied class of models need a stable linearization (22, 23). Exponential forgetting, important for the ML estimator, means that the initial errors are forgotten with a factor λ^t where $0 \leq \lambda < 1$. This is implied by an exponentially stable system (22, 23). It should be noted that (22) and (23) represents a linear time varying (LTV) system and the stability of such a LTV system can be analyzed by regular methods [11]. However, the important thing here is the condition of an exponentially stable linearization of the nonlinear system (22, 23) to obtain vanishing initial errors and a converging ML estimator.

Remark: If the system N is a linear time invariant system then the linearized system, of course, reduces to a linear time invariant system.

5 MAXIMUM LIKELIHOOD ESTIMATION

The maximum likelihood (ML) identification method is a very general identification method. Evaluating the likelihood exactly is possible when the model is linear in the noises, see [9], or when the measurements are invertible with respect to the input noises, [2].

The principle of the ML method is based on treating the observed variable $y^M = (y(1), \ldots, y(M))$ as a random variable with a probability density function (PDF) $\phi_y(y \mid \theta)$. The probability of y thus depends on a parameter vector θ. To estimate θ from an observation of y the method chooses a θ which maximizes the likelihood

$$L(\theta \mid y) = \phi_y(y \mid \theta) \tag{24}$$

for one sample of y. In general, the measured output y is not independent from sample to sample. However, if we consider M observations, then, by repeated use of the multiplication rule for conditioned probabilities, we get

$$L(\theta \mid y) = \phi_y(y(M) \mid y^{M-1}) ... \phi_y(y(2) \mid y(1))\phi_y(y(1)), \tag{25}$$

where y^{M-1} denotes all observations from $y(1)$ up to $y(M-1)$. As shown in [6], ϕ_y can be computed as

$$\phi_y(y \mid \theta) = \phi_w(W(y)) \left| \frac{d[W(y)]}{dy} \right|, \tag{26}$$

in the scalar case, where ϕ_w is the PDF of the input noise w and $|\cdot|$ denotes determinant. In (26), W denotes the inverse of the relation between the input and the output. An important part of this paper is the use of the inverse to calculate the likelihood.

We can now rewrite (25) as

$$L(\theta \mid y) = \prod_{i=1}^{M} \phi_w(W(y(i))) \left| \frac{d[W(y(i))]}{dy(i)} \right|. \tag{27}$$

Maximizing the likelihood is the same as minimizing the negative logarithm of $L(\theta \mid y)$ and the cost function, assuming Gaussian inputs, becomes

$$V(\theta) = \sum_{i=1}^{M} \left(\frac{1}{2}\hat{w}(i)^2 - \log_e \left| \frac{d[W(y(i))]}{dy(i)} \right| \right). \tag{28}$$

Finally, by minimizing $V(\theta)$ we get the ML estimate of θ as

$$\hat{\theta} = \arg\min_{\theta} V(\theta) \tag{29}$$

Notice the log term which does not appear for models where the inverse is computed as

$$w(t) = y(t) - f(y(t-1), y(t-2), \ldots).$$

6 NUMERICAL EXAMPLES

We will show two different numerical examples. In these examples the parameters of the models were estimated and the inverse of these systems were implemented using the scheme given in Figure 2. One hundred input realizations were used for each example, each containing 1000 samples ($M = 1000$).

First, consider a discrete time system with a first order transfer function in parallel with an exponential nonlinearity and a Gaussian input as shown in Figure 3. Using this scheme the cost function is computed as

$$V(\theta) = \sum_{i=1}^{M} \left(\frac{1}{2}\hat{w}(i)^2 - \log_e \left| \frac{1}{1-(1-d\hat{w}(i))ce^{-d\hat{w}(i)}} \right| \right) \tag{30}$$

where $\hat{w}$ is computed according to Figure 2.

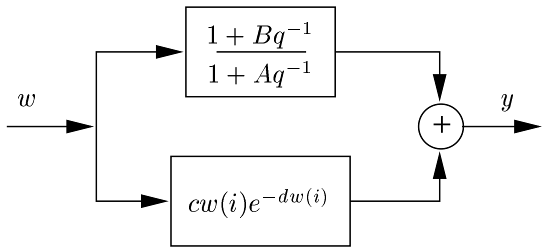

Figure 3: Block diagram of the first numerical example.

The four parameters, $\theta = (A, B, c, d)^T$ of the system were simultaneously estimated using the above described ML method. The results are shown in Table 1 where the actual parameter value is shown as the true value θ_0. The mean of the 100 parameter estimates are presented as $mean(\hat{\theta})$ and the estimated standard deviation is denoted $std(\hat{\theta})$.

Table 1: *Parameter estimation results.*

	True value θ_0	$mean(\hat{\theta})$	$std(\hat{\theta})$
A	0.90	0.90	0.03
B	0.60	0.60	0.06
c	0.70	0.71	0.05
d	0.20	0.20	0.04

The second example has a dynamic nonlinear part although the linear part is the same as in the first example. The nonlinear part is described by

$$x(t+1) = ax(t) + \frac{e(t)}{1+be^2(t)} \tag{31}$$

$$y(t) = x(t) + ce(t)e^{-de^2(t)}, \tag{32}$$

thus, $\theta = (a, b, c, d, A, B)^T$ were estimated, results are presented in Table 2, using the cost function

$$V = \sum_{i=1}^{M} \left(\frac{1}{2}\hat{w}(i)^2 - \log_e \left| \frac{1}{1+(1-2cd\hat{w}(i))ce^{-d\hat{w}^2(i)}} \right| \right). \tag{33}$$

It can be concluded that the estimation procedure works very well for these models which at first glance appear very difficult to apply ML estimation on. A similar method, but without special notice taken to partitioned models, has also been successfully implemented on a nonlinear model generating EEG-like activity [10].

Table 2: *Parameter estimation results.*

	True value θ_0	$mean(\hat{\theta})$	$std(\hat{\theta})$
a	0.80	0.79	0.03
b	1.00	1.01	0.20
c	0.70	0.71	0.06
d	0.20	0.20	0.03
A	0.90	0.89	0.02
B	0.60	0.59	0.04

7 CONCLUSION

In this contribution we have shown how ML estimation can be applied to invertible nonlinear models. A class of partitioned nonlinear models has been considered. The stability of the inverse of such systems were given by the criterion of the small gain approach and the conditions concerning initial conditions boiled down to the stability of a linear time-varying system. The ML criterion was presented for our application and finally the method was tested on numerical examples.

8 REFERENCES

[1] J.S. Bendat *Nonlinear System Analysis and Identification from Random Data*, John Wiley and Sons, 1990.

[2] T. Bohlin. *Interactive System Identification: Prospects and Pitfalls* , Springer-Verlag, 1991.

[3] H.W. Chen "Modeling and Identification of Parallel Nonlinear Systems - Structural Classification and Parameter Estimation Methods". *Proceedings of the IEEE*, 83:39–66, 1995.

[4] F.J. Doyle III, B.A. Ogunnaike and R.K. Pearson "Nonlinear Model-based Control Using Second-order Volterra Models". *Automatica*, 31:697–714, 1995.

[5] A. Isidori *Nonlinear Control Systems, Third edition*, Springer, 1995.

[6] A.H. Jazwinski *Stochastic Processes and Filtering Theory*, Academic Press, 1970.

[7] A. Juditsky, H. Hjalmarsson, A. Benveniste, B. Delyon, L. Ljung, J. Sjöberg and Q. Zhang "Nonlinear Black-box Modeling in System Identification: Mathematical Foundations" *Automatica*, 31:1725–1750, 1995.

[8] U. Kotta "Right inverse of a discrete time non-linear system", *International Journal of Control*, 51:1–9, 1990.

[9] L. Ljung *System Identification, Theory for the user*, Prentice-Hall, 1987.

[10] O. Markusson and T. Bohlin "Identification of a nonlinear EEG generating model", *IEEE Workshop on Nonlinear Signal and Image Processing*, Michigan, USA, 1997.

[11] W.J. Rugh *Linear System Theory*, Prentice-Hall, 1996.

[12] J. Sjöberg, Q. Zhang, L. Ljung, A. Benveniste, B. Delyon, P-Y. Glorennec, H. Hjalmarsson and A. Juditsky. "Nonlinear Black-box Modeling in System Identification: a Unified Overview". *Automatica*, 31:1691–1724, 1995.

[13] M. Vidyasagar *Nonlinear Systems Analysis*, Prentice-Hall, 1993.