PARAMETER AND DERIVATIVE ESTIMATION FOR NONLINEAR CONTINUOUS-TIME SYSTEM IDENTIFICATION

Marc Niethammer [*] **Patrick H. Menold** [*]
Frank Allgöwer [*]

[*] *Institut für Systemtheorie technischer Prozesse, Universität Stuttgart,*
{niethammer, menold, allgower}@ist.uni-stuttgart.de

Abstract: In this note we investigate parameter identification for nonlinear continuous-time SISO systems, based on input/output data, where the output is assumed to be noise corrupted. Continuous-time identification requires (i) the estimation of derivatives of the output (which is nontrivial due to the influence of noise), a suitable measure thereof or a way to avoid them and (ii) methods for parameter estimation based on the measured data. Problem (i) is addressed by reviewing modulating functions and the concept of delayed state variable filters; furthermore a high-gain observer is introduced as an approach to provide the necessary derivative information. Its performance is investigated with respect to the modulating function and the delayed state variable filter approaches. Least-squares methods are assessed for continuous-time nonlinear identification (problem (ii)). It is shown that parameter identification based on modulating functions and a standard least-squares method does not guarantee bias-free estimates for some systems. Whereas ordinary (or weighted) least-squares is sufficient for parameter identification by means of modulating functions it is not for the delayed state variable filter and the high-gain observer approaches (due to dependencies between error terms). Requirements on least-squares methods for nonlinear continuous-time system identification are discussed and a solution for bilinear systems is given. The importance of an appropriate least-squares method is underlined by parameter identification for a simulated bilinear example system.

Keywords: high-gain observer, delayed state variable filter, modulating functions, least-squares techniques, parameter estimation

1. INTRODUCTION

Usually system identification is performed in discrete time (using difference equations). However, besides the apparent advantage of discrete descriptions for computer implementations, there are also good reasons to leave this track and to perform identification in continuous-time (CT): Models derived from first order principles are given in CT, and parameters are usually physically meaningful in this domain. However, the disadvantage of CT identification lies in the occurrence of derivative terms, especially in combination with noisy measurements and nonlinearities. The problem is thus twofold:

- P1. To determine the derivatives, suitable measures thereof or to avoid them completely (under the influence of noise).
- P2. Given the structure of the nonlinear system, to estimate parameters based on (possibly) noisy data in a sound way.

This paper will consider nonlinear continuous-time SISO systems, affine in their parameters $\boldsymbol{p}$. The system input u and its derivatives are assumed to be known and noise-free, the output y_n to be known but noisy. Sec. 2 describes the transformation of a nonlinear system to observability normal form, which will be used for the identification and represents the system in input-output form. Sec. 3 describes the modulating function, the delayed state variable filter and the high-gain observer approaches to handle the derivative problem. Sec. 4 investigates least-squares algorithms with respect to their suitability for parameter identification in conjunction with the methods mentioned above and Sec. 5 applies the methods to a simple bilinear system. Only parameter identification for CT systems will be considered.

2. PROBLEM STATEMENT

A nonlinear SISO system is given by

$$\begin{aligned} \dot{\boldsymbol{x}} &= \boldsymbol{f}(\boldsymbol{x}, u, \boldsymbol{p}_1) \\ y &= h(\boldsymbol{x}, u, \boldsymbol{p}_2). \end{aligned} \tag{1}$$

$\boldsymbol{x}$ is the state vector, y and u are the scalar output and input respectively. $\boldsymbol{f}(\cdot)$ is a nonlinear vector field, $h(\cdot)$ a nonlinear function and $\boldsymbol{p}_1$, $\boldsymbol{p}_2$ are constant parameter vectors. Eq. (1) can be transformed to observability normal form (Zeitz, 1990)

$$\dot{\boldsymbol{x}}^* = [x_2^*,\ \dots,\ x_n^*,\ f_n^*(\boldsymbol{x}^*, \boldsymbol{u}^*, \boldsymbol{p})]^T$$
$$y = x_1^*, \qquad (2)$$

where $\boldsymbol{u}^* := (u, \dot{u}, \dots, u^{(n)})^T$, $\boldsymbol{p} = [\boldsymbol{p}_1^T,\ \boldsymbol{p}_2^T]^T$, by means of its invertible observability map

$$\left(y\ \dots\ y^{(n-1)}\right)^T =: v(\boldsymbol{x}, \overline{\boldsymbol{u}}, \boldsymbol{p}) =: \left(x_1^*\ \dots\ x_{n-1}^*\right)^T,$$

with $\overline{\boldsymbol{u}} := (u, \dot{u}, \dots, u^{(n-1)})^T$ (note, that we assume that f_n^* is affine in its parameters $\boldsymbol{p}$ in this paper). It is easy to see that this leads to

$$y^{(n)} = f_n^*(y, \dot{y}, \dots, y^{(n-1)}, \boldsymbol{u}^*, \boldsymbol{p}). \qquad (3)$$

Note that a bilinear SISO system

$$\dot{\boldsymbol{x}} = A\boldsymbol{x} + \boldsymbol{b}u + E\boldsymbol{x}u \qquad (4)$$
$$y = \boldsymbol{c}^T\boldsymbol{x} + du,$$

will be affine in its states $\boldsymbol{x}^*$ under this transformation.

CT parameter identification in the context of this paper is then splitted in the following subtasks:

S1. Decide upon a system structure (1) (e.g. derived from first order principles) and transform it to observability normal form (3).
S2. Excite the system with a suitable input function u and measure the output y_n.
S3. Estimate measures of the derivatives $\hat{y}^{(i)}$, $i = 0(1)n$ of the noisy output y_n or avoid them by an appropriate method (see Sec. 3).
S4. Estimate sound parameters $\boldsymbol{p}$ based on the data generated in S3 (see Sec. 4).
S5. Validate the identified parameters using a validation data set.

3. DERIVATIVES

Literature for CT identification is mainly concerned with linear systems (Sinha and Rao, 1991; Unbehauen and Rao, 1987). Only few methods to cope with the derivative problem can be applied to nonlinear systems: This section reviews the modulating function approach, the delayed state variable filter (a direct supplement of the state variable filter concept for linear systems) and introduces the high-gain observer for use with the CT identification.

3.1 *Modulating functions*

A classical way to deal with the derivative problem (Shinbrot, 1957; Unbehauen and Rao, 1998) is to avoid the computation of the derivatives of the noisy output y completely: the modulating functions (MFs) approach. Given the SISO system (3), where $\boldsymbol{p}$ is the vector of parameters to be determined, multiplication by the MF ϕ_j and integration over the time interval $[0, T]$ yields

$$\int_0^T y^{(n)}\phi_j \mathrm{d}t = \int_0^T f_n^*(y, \dot{y}, \dots, y^{(n-1)}, \boldsymbol{u}^*, \boldsymbol{p})\phi_j \mathrm{d}t. \qquad (5)$$

Choosing the MFs such that

$$\phi_j^{(i)}(t) = 0, i = 0(1)n, t \in \{0, T\} \qquad (6)$$

and integration by parts of (5) results in

$$\int_0^T y\phi_j^{(n)} \mathrm{d}t = \int_0^T \overline{f}_n(\phi_j, \dot{\phi}_j, \dots, \phi_j^{(n-1)}, \boldsymbol{u}^*, \boldsymbol{p}, y)\mathrm{d}t. \qquad (7)$$

Note that the influence of initial and end conditions was eliminated by (6) and that (7) no longer depends on derivatives of y. The use of N different modulating functions ($N >> dim(\boldsymbol{p})$) leads to a set of equations which can be solved for $\boldsymbol{p}$. They are applicable within the class specified by the nonlinear input-output differential operator model (Pearson, 1992):

$$\sum_{i=0}^{n_1}\sum_{k=1}^{n_2} g_i(\boldsymbol{p})F_{ik}(u, y)P_{ik}(\frac{\mathrm{d}}{\mathrm{d}t})E_k(u, y) = 0, \qquad (8)$$

and thus not for any nonlinear system. $P_{ik}(\frac{\mathrm{d}}{\mathrm{d}t})$ are polynomials of order n in the differential operator $\frac{\mathrm{d}}{\mathrm{d}t}$, $F_{ik}(u, y)$ and $E_k(u, y)$ are functions of the output y and the input u; $g_i(\boldsymbol{p})$ are functions depending on the parameter vector $\boldsymbol{p}$, where $g_0 = 1$. If (as usually done) an ordinary (OLS) or a weighted least-squares (WLS) approach is used to estimate the parameters $\boldsymbol{p}$ from the set of N equations (Unbehauen and Rao, 1998), the MF cannot guarantee bias-free estimates for the complete class (8), instead it has to be true (for systems in form (3)) that

$$\mathcal{E}[\overline{f}_n(\phi_j, \dot{\phi}_j, \dots, \phi_j^{(n-1)}, \boldsymbol{u}^*, \boldsymbol{p}, y_n)] = \overline{f}_n(\phi_j, \dot{\phi}_j, \dots, \phi_j^{(n-1)}, \boldsymbol{u}^*, \boldsymbol{p}, y),$$

where $\mathcal{E}[x]$ is the expected value of x, y_n is the noisy measurement of y, and y is the exact output value. E.g. terms of the form $y^{2m}f_u(\boldsymbol{u}^*)$, $m \in \mathbb{N}^+$ (where f_u is an arbitrary function in $\boldsymbol{u}^*$) will introduce bias, because the expected value of the integral term (using $y_n = y + \Delta_+$, $\mathcal{E}[\Delta_+] = 0$) is

$$\int_0^T \mathcal{E}[(y + \Delta_+)^{2m}]f_u(u^*)\phi_n \mathrm{d}t \neq \int_0^T y^{2m}f_u(u^*)\phi_n \mathrm{d}t.$$

This fact is widely overlooked in the MF literature.

3.2 *Delayed state variable filter*

The output y_Γ of the delayed state variable filter (DSVF) $\frac{1}{\Gamma}$ as proposed by (Tsang and Billings, 1994) and depicted in Fig. (1), is a lowpass filtered version of its input y_n. The derivatives $y_\Gamma^{(i)}$ of y_Γ are given as the state variables of the DSVF. To avoid the computation of the derivatives $y^{(i)}$ of y ($y_n^{(i)}$ respectively), one would like to replace them by the state variables of the DSVF in Eq. (3), resulting in

$$y_\Gamma^{(n)} = f_n^*(y_\Gamma, \dot{y}_\Gamma, \dots, y_\Gamma^{(n-1)}, \boldsymbol{u}_\Gamma^*, \boldsymbol{p}). \qquad (9)$$

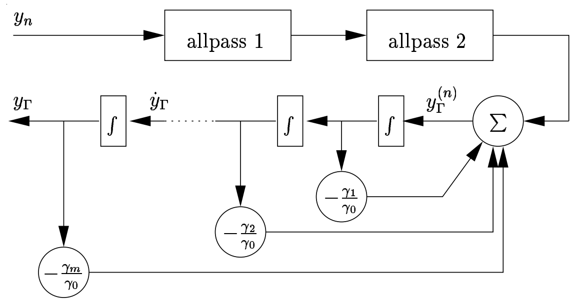

Fig. 1. Delay-equalized Butterworth filter

However, this is only allowed, if the filter $\frac{1}{\Gamma}$ commutes with the nonlinearity f_n^*, i.e. if

$$\frac{1}{\Gamma}y^{(n)} = \frac{1}{\Gamma}f_n^*(y, \dot{y}, \ldots, y^{(n-1)}, \boldsymbol{u}^*, \boldsymbol{p})$$

is equivalent to (9). The DSVF is thus designed as an approximate dead time element (which fulfills the commutation requirement up to a chosen cutoff frequency ω_c) and realized by a delay equalized Butterworth filter: a Butterworth filter with its input prefiltered by two allpass filters, so to guarantee constant group delay and thus dead time behavior. Note, that the DSVF behaves like an integrator and will thus be sensitive to low frequency noise. As known, a Gaussian input signal will result in a Gaussian output signal (with possibly different standard deviation and amplitude), since the DSVF is a linear system (the same is true for the HG-observer, see below).

3.3 *High-gain observer*

A high-gain (HG) observer for the nonlinear system (2) as proposed by (Tornambè, 1992) is given by

$$\dot{\hat{\boldsymbol{x}}} = \begin{pmatrix} 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & 1 \\ 0 & 0 & 0 & \ldots & 0 \end{pmatrix} \hat{\boldsymbol{x}} + \begin{pmatrix} p_1 k \\ p_2 k^2 \\ \vdots \\ p_{n-1}k^{n-1} \\ p_n k^n \end{pmatrix} \cdot (y_l - \hat{y})$$

$$\hat{y} = \hat{x}_1, \; \hat{\boldsymbol{x}} = [\hat{x}_1, \ldots, \hat{x}_n]^T, \qquad (10)$$

where k is the observer gain, $\hat{y}$ is the observer output, $\hat{\boldsymbol{x}}$ is the observer state and y_l the lowpass filtered y_n. Note that the observer does not incorporate an approximation of the nonlinearity $f_n^*(\boldsymbol{x}^*, \boldsymbol{u}^*)$ of Equation (2). The parameters $\{p_i\}$ $i \in [1, n]$ have to ensure that

$$p(\lambda) = \lambda^n + \lambda^{n-1}p_1 + \cdots + \lambda p_{n-1} + p_n$$

is a Hurwitz polynomial with distinct eigenvalues (note that the poles of the HG-observer are given by these eigenvalues multiplied by the observer gain k. A large k will thus lead to a faster, but more noise-sensitive HG-observer). The HG observer then allows for the computation of estimates for y and its derivatives. For the identification problem described below it is

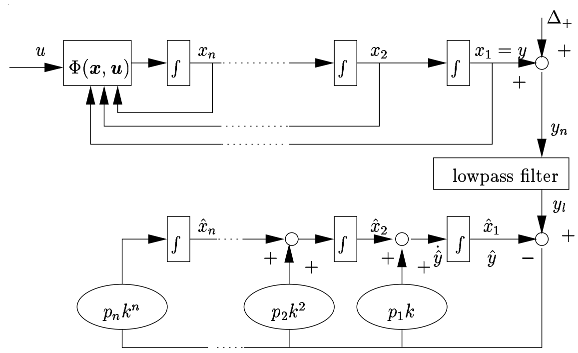

Fig. 2. High-gain observer

crucial to determine the value of the n^{th} order derivative of the output as well. To do so one can simply augment the HG-observer (10) by an additional state. The overall structure stays the same, but the observer system order is increased by one ($n \to n+1$). However this imposes even stronger conditions on the class of nonlinearity the HG-observer can handle. Whereas for the regular (without extended state) HG-observer the nonlinearity f_n^* has to be bounded for all states $\boldsymbol{x}$ and for all inputs u ($|f_n^*| < \mu$), the state extension requires the derivative of the nonlinearity f_n^* with respect to time to be bounded (if the state is extended by two it's the second derivative, etc. ...). The proof stays essentially the same as the one given in (Tornambè, 1992) without state extension. Since the HG-observer computes its states by approximate differentiation it is very sensitive to high-frequency noise. Thus, it is desirable to filter the noisy output y_n with a lowpass filter before using it with the HG observer (see Fig. (2)). The cutoff frequency for the lowpass filter should be chosen higher than the bandwidth of the system to be identified (so not to change the actual output of the system). Of course, only high-frequency noise will be eliminated in this way.

4. LEAST-SQUARES METHODS

We assume that the system to be identified is affine in its parameters $\boldsymbol{p}$. Any system (3) that fulfills this condition can be decomposed into

$$\underbrace{\boldsymbol{f}^{*^T}(y(t), \dot{y}(t), \ldots, y^{(n-1)}(t), \boldsymbol{u}^*(t))}_{\boldsymbol{a}^T(t)}\,\boldsymbol{p} = \underbrace{y^{(n)}(t) + k(y(t), \dot{y}(t), \ldots, y^{(n-1)}(t), \boldsymbol{u}^*(t))}_{b(t)} \qquad (11)$$

for any time t, where $\boldsymbol{f}^*$ and k are arbitrary nonlinear functions in their respective arguments. Combining N time instants $t_i, i = 1(1)N$ of Eq. (11) leads to

$$A\boldsymbol{p} = \boldsymbol{b}, \qquad (12)$$

where $\boldsymbol{p}$ are the parameters to be identified and $A_{ij} = a_j(t_i)$ (the j^{th} element of vector $\boldsymbol{a}$ at time

$t = t_i$) and $b_i = b(t_i)$ (the i^{th} element of $\boldsymbol{b}$ is $b(t_i)$). If A and $\boldsymbol{b}$ contain the estimated and noise corrupted value $\hat{y} = y - \Delta y_1$ and its derivatives $\hat{y}^{(i)} = y^{(i)} - \Delta y_{i-1}$, Eq. (12) will not be exactly fulfilled ($A\boldsymbol{p} \approx \boldsymbol{b}$) and thus

$$\underbrace{(A + \Delta A)}_{\hat{A}}\boldsymbol{p} = \underbrace{\boldsymbol{b} + \Delta\boldsymbol{b}}_{\hat{\boldsymbol{b}}}, \tag{13}$$

where ΔA and $\Delta\boldsymbol{b}$ are error terms.

4.1 *The linear case*

Given the parameter estimation problem (13) the covariance matrix $C = \mathcal{E}(E^T E)$, with $E = [\Delta A; \Delta\boldsymbol{b}] = [\Delta E_1, \ldots, \Delta E_m]$ is of crucial importance. It describes the dependency between the individual error terms, and is thus important for deciding which least-squares method to choose. In the context of CT system identification, where derivatives of a noisy output $y_n = y + \Delta_n$ (or measures thereof) are determined, rather than measured, the errors of each of the columns of E will generally be dependent on the errors of all the other columns. This amounts to a fully occupied covariance matrix C. Least-squares methods require $\mathcal{E}(\Delta E_i) = 0$, which is fulfilled for any linear system, with $\mathcal{E}(\Delta_+) = 0$ and derivatives (or measures thereof) generated by linear filtering. However neither the ordinary-least squares (OLS), nor the total least squares (TLS) method are able to cope with a fully occupied covariance matrix C. OLS assumes an error free A, and an erroneous $\boldsymbol{b}$. Only then is $\boldsymbol{p} = A^+\boldsymbol{b}$ (where $A^+ = (A^T A)^{-1} A^T$) a sound estimate. TLS assumes independently and identically distributed ΔE_i, which amounts to $C = \sigma^2 I$ (where σ is a constant and I an appropriately sized identity matrix). The TLS solution is then computed by solving the minimization problem (van Huffel and Vandevalle, 1991)

$$[\hat{A}; \hat{\boldsymbol{b}}][\boldsymbol{p}^T; -1] = \boldsymbol{0}, \ \underset{[\hat{A};\hat{\boldsymbol{b}}]}{min} \|[A; \boldsymbol{b}] - [\hat{A}; \hat{\boldsymbol{b}}]\|_F, \tag{14}$$

where $\|X\|_F$ is the Frobenius norm of the matrix X. Eq. (14) is solved by a singular value decomposition that facilitates a sound rank reduction (if necessary) of $[A; \boldsymbol{b}]$ so that $[\hat{A}; \hat{\boldsymbol{b}}][\boldsymbol{p}^T; -1] = \boldsymbol{0}$ becomes feasible. Based on the TLS approach, generalized TLS (gTLS) uses a coordinate transform

$$[A; \boldsymbol{b}] R^{-1} [\boldsymbol{p}^T; -1]^T = \boldsymbol{0}, \quad with \ \ C = R^T R$$

to recast the parameter estimation problem, s.t. the covariance matrix in the transformed coordinates becomes $C' = I$. Thus TLS can be applied to solve for $\boldsymbol{p}$. gTLS leads to a sound solution for any linear system, with any regular covariance matrix C and is thus a highly suitable least-squares method for CT parameter identification for linear systems. In practice the transformation is not performed in such a way, because of numerical problems. Instead the solution is computed via the generalized singular value decomposition (gSVD), which is numerically more robust (van Huffel and Vandevalle, 1989).

4.2 *Nonlinear case*

All the least-squares methods reviewed in this paper so far (OLS, TLS, gTLS) were based on the underlying assumption of a stationary noise process (Δy_i) overlaid on $\hat{y}$ and its derivatives. For linear systems the columns of E will represent stationary noise processes (each column of $[A; \boldsymbol{b}]$ is made up of a linear combination in $\hat{y}$, u and their respective derivatives). But for nonlinear systems this assumption breaks down. In general the following is true:

F1. The noise processes can no longer be regarded as zero mean.
F2. The noise processes are non-stationary.

As a simple example to illustrate F1. and F2. just take the expected value of the square of the random variable x_t; with $\mathcal{E}[x_t] = 0, \mathcal{E}[x_t^2] = \sigma^2$. The mean value of this square term is simply the variance of the random variable x_t, which is nonzero. Cross product terms (e.g. $x_1 x_2$ or $u x_4$) have time-variable probability density functions. This is easy to see by considering the product of a deterministic signal u (e.g. the input signal) and a random signal x_t. The mean value $\mathcal{E}[u x_t] = u\mathcal{E}[x_t]$ at time t is dependent on the current value of $u(t)$, and the variance $\mathcal{E}[u^2 x_t^2] = u^2 \mathcal{E}[x_t^2]$ on the current value of $u(t)^2$, thus violating the stationarity assumption (note, that the mean value of a zero mean signal multiplied with a deterministic signal stays zero). A general nonlinear estimation problem is the following: Find the parameters $\boldsymbol{p}$ which minimize

$$\Xi = \sum_{i=1}^{N} \begin{pmatrix} \Delta y_1(i) \\ \vdots \\ \Delta y_{n+1}(i) \end{pmatrix}^T Q \begin{pmatrix} \Delta y_1(i) \\ \vdots \\ \Delta y_{n+1}(i) \end{pmatrix} \tag{15}$$

subject to the constraint

$$\boldsymbol{f}^{*^T}(\hat{y} + \Delta y_1, \ldots, \hat{y}^{(n-1)} + \Delta y_n, \boldsymbol{u}^*)\boldsymbol{p} = \tag{16}$$
$$\hat{y}^{(n)} + \Delta y_{n+1} + k(\hat{y} + \Delta y_1, \ldots, \hat{y}^{(n-1)} + \Delta y_n), \boldsymbol{u}^*)$$

for all time $t_i, i = 1(1)N$, where $Q \geq 0$ is a general positive semi-definite weighting matrix. The goal is thus to find a solution for the parameters $\boldsymbol{p}$ that is maximally close to the estimated $\hat{y}$ and its derivatives with respect to Eq. (15). The general solution for the nonlinear problem is involved (a dynamic optimization problem), however for bilinear systems Eqns. (15,16) can be cast to a relatively simple static nonlinear optimization problem without constraints, of the form

$$L^* = min \sum_{i=1}^{N} \frac{\boldsymbol{p'}^T Z(i) \boldsymbol{p'}}{\boldsymbol{p'}^T N(i) \boldsymbol{p'}}, \tag{17}$$

using calculus of variations (see (Han *et al.*, 1996) for the discrete-time version). Here, $Z(i)$ and $N(i)$ are matrices depending on the input u and its derivatives at time instant t_i, $\boldsymbol{p'}$ is the parameter vector and $\boldsymbol{u}^*$ is assumed to be known exactly. If $Q = C^{-1}$, where C^{-1} is the covariance matrix of $[\Delta y_1, \ldots, \Delta y_{n+1}]^T$, this approach is called extended generalized total least squares (eGTLS). It it the counterpart to gTLS for linear systems, which decouples the noise processes Δy_i by means of the transformation matrix C^{-1}, which can be approximated by simulation.

5. BILINEAR EXAMPLE

To demonstrate the usefulness of the previously described methods let's consider the bilinear system

$$\begin{aligned} \dot{x}_1 &= (\mu_a - \mu_e)x_1 + ux_{10} - ux_1 \\ \dot{x}_2 &= \mu_e x_1 - \mu_g x_2 - ux_2 \\ y &= x_2, \end{aligned} \tag{18}$$

with states x_1, x_2, input u, known constant x_{10} and the unknown parameters μ_a, μ_e and μ_g to be determined. Transformation to observability normal form yields

$$\underbrace{(p_2^* \; p_1^* \; -1)}_{a^T} \begin{pmatrix} y \\ \dot{y} \\ \ddot{y} \end{pmatrix} + \underbrace{p_3^* x_{10}}_{b^T} u + \underbrace{(p_1^* \; -2)}_{c_{(0)}^T} u \begin{pmatrix} y \\ \dot{y} \end{pmatrix} + \underbrace{(-1 \; 0)}_{c_{(1)}^T} \dot{u} \begin{pmatrix} y \\ \dot{y} \end{pmatrix} + \underbrace{(-1 \; 0)}_{c_{(2)}^T} u^2 \begin{pmatrix} y \\ \dot{y} \end{pmatrix} = 0 \tag{19}$$

which is equivalent to form (16) ($p_1^* = \mu_a - \mu_e - \mu_g$, $p_2^* = \mu_g(\mu_a - \mu_e)$, $p_3^* = \mu_e$).

5.1 *Results of the example*

Identification of system (19) was performed using the input u and the noisy output y_n as given in Fig. (3) [1]. The input was composed of four sine waves with frequencies 0.25 Hz, 0.5 Hz, 0.75 Hz, 1.0 Hz and (randomly chosen) phases 0.1367 rad, 5.8046 rad, 3.9749 rad, 0.4202 rad respectively; each with unit amplitude. The amplitude of the resulting function was subsequently scaled to lie within the interval $[10-7.5, 10+7.5]$. The noise to signal ratio of the output y_n was 20% (Gaussian white noise over all frequencies, zero mean).

All poles of the HG observer were at -10 with observer gain $k = 50$. For the high-gain observer the noisy output signal y_n was lowpass filtered prior to calculating the derivatives (Remez lowpass, passband cutoff frequency $f_p = 20$ Hz, stopband cutoff frequency $f_s = 25$ Hz, passband deviation $\delta_p = 1e-4$, stopband deviation $\delta_s = 1e-3$). Identification was performed by eGTLS and for comparison with OLS.

An 8^{th} order Butterworth filter with cutoff frequency $f_p = 20$ Hz was used for the delayed state variable filter, together with two allpass filters (see (Tsang and Billings, 1994) for details on this filter, and the coefficients for the Butterworth and the two allpass filters). Because of its inherent lowpass characteristic y_n was not lowpass filtered, before applying the DSVF.

[1] For details on the parameter identification of system (19) via eGTLS see (Niethammer, 2000), via MFs see Appendix A.

	p_1^*	p_2^*	p_3^*
exact	-4	-4	5
Hartley mod fcn. (LS)	-3.92	-4.58	5.02
HG observer (OLS)	31.65	-115.62	-2.445
HG observer (eGTLS)	-3.68	-4.72	4.94
DSVF (OLS)	4.20	-36.33	3.48
DSVF (eGTLS)	-2.87	-7.56	4.77

Table 1. Estimated parameters.

The high (compared to the frequency content of u and y) cutoff frequency $f_p = 20$ Hz was chosen to assess the robustness of the algorithms with respect to noise. A lower cutoff frequency will improve the performance of the filters, particularly for the HG observer (since the calculation of derivatives is especially sensitive to high-frequency noise).

Hartley MFs of order $m = 3$

$$\phi_j = \sum_{r=0}^{m} (-1)^r \binom{m}{r} \, cas(m + j - r)\omega_0 t, j \in \mathbb{Z},$$

with $cas\ t = \cos t + \sin t$ and $\omega_0 = \frac{2\pi}{T}$ (Unbehauen and Rao, 1998) were used for the MF approach (modulating frequency indices: $j = -9(1)9$). y_n was not filtered, and the identification was performed using OLS.

Fig. (3) shows the estimated output $\hat{y}$ for the HG observer and the DSVF. Table (1) summarizes the identification results for the different methods based on y_n and u (where u and its derivatives are assumed to be known exactly) of Fig. (3). For parameter identification via eGTLS and DSVF/HG-observer, 9 different initial conditions were used for the optimization algorithm ($\{p_1^*, p_2^*, p_3^*\}$ and $\{p_1^* \pm 2, p_2^* \pm 2, p_3^* \pm 2\}$). Any of these optimization runs converged to the same solution. Clearly the parameters identified by the MF approach are closest to the correct one. OLS failed for the HG-observer and the DSVF, but eGTLS produced reasonable results. Fig (4) shows the deviations for the eGTLS solution for the DSVF and the HG-observer, and for the OLS solution for the MF approach, from benchmark values generated for validation of the identified parameters (different input u; approximate orders of magnitude (after startup transient): $y \in [0, 0.175]$, $\dot{y} \in [-0.125, 0.125]$, $\ddot{y} \in [-0.8, 0.5]$). Modulating functions and the HG-observer lead to excellent results comparable in quality. Both clearly outperform the DSVF in this example.

6. CONCLUSIONS

In this paper MFs and the DSVF were reviewed for CT identification. The HG-observer was introduced in this context. LS methods were revisited and their applicability with respect to nonlinear system identification was discussed. The MF approach turned out (as expected) to perform

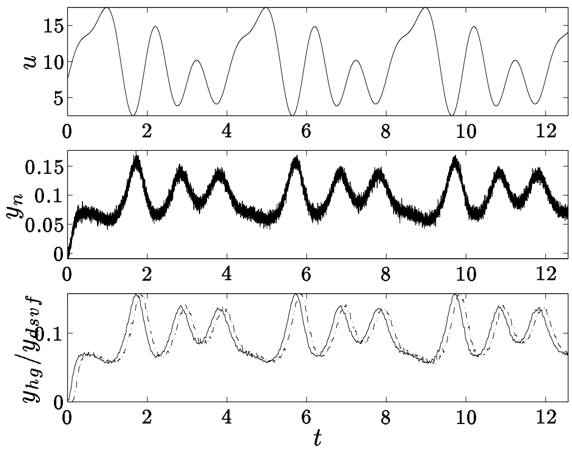

Fig. 3. Input u and noisy output y_n used for the identification. y_{hg} (solid line, HG obvserver), y_{dsvf} (dash-dotted line, DSVF) : estimated output $\hat{y}$.

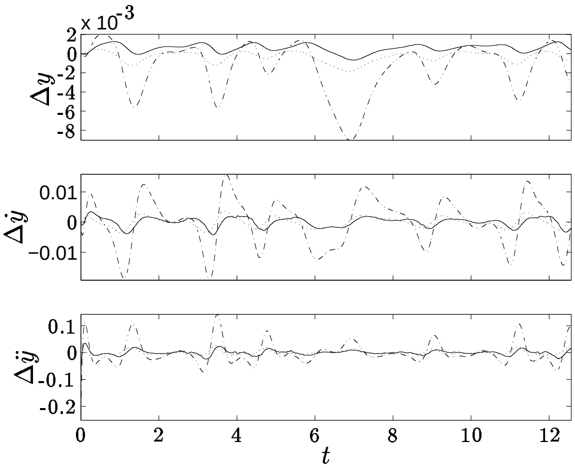

Fig. 4. Deviations from the exact values of y, $\dot{y}$, $\ddot{y}$ (Δy, $\Delta \dot{y}$, $\Delta \ddot{y}$) , for the identified systems excited with the validation input u. Dotted lines: results for the MF approach (Hartley MFs). Dash-dotted lines: DSVF. Solid lines: HG observer.

best for the identification of a bilinear example system under noise influence. However, with the HG-observer (eGTLS) approach leading to only slightly worse results. The DSVF (eGTLS) approach was clearly outperformed by the MF and the HG-observer (eGTLS) approaches.

All three methods seem to have their suitable application area. The main disadvantage of MFs being that they are confined to a certain system class (which has to permit integration by parts), whereas DSVF and HG-observer are not. However, MFs are easy to apply, and computationally inexpensive (no optimization is required, OLS or WLS are sufficient). This bilinear example showed a superiority of the HG-observer over the DSVF approach. It is expected that this superiority will depend on the choice of the input function $u(t)$, since the HG-observer is sensitive to high-frequency, whereas the DSVF is sensitive to low-frequency noise. Further research should be directed towards, and should try to exploit this phenomenon.

7. REFERENCES

Han, S., J. Kim and K. Sung (1996). Extended Generalized Total Least Squares Method for the Identification of Bilinear Systems. *IEEE Transactions on Signal Processing* **44**(4), 1015–1018.

Niethammer, M. (2000). An approach towards nonlinear continuous-time system identification. Master's thesis. Universität Stuttgart.

Pearson, A. E. (1992). Explicit parameter identification for a class of nonlinear I/O differential operator models. In: *In Proc. of 31st IEEE Conference on Decision and Control '92, Tucson, Arizona.*

Shinbrot, M. (1957). On the Analysis of Linear and Nonlinear Systems. *Transactions of the ASME* **79**, 547–552.

Sinha, N. K. and G. P. Rao (1991). *Identification of Continuous-Time Systems, Methodology and Computer Implementation.* Vol. 7 of *Microprocessor-based systems engineering.* Kluwer Academic Publishers.

Tornambè, A. (1992). High-gain observers for non-linear systems. *International Journal of Systems Sciences* **23**(9), 1475–1489.

Tsang, K. M. and S. A. Billings (1994). Identification of continuous time nonlinear systems using delayed state variable filters. *International Journal of Control* **60**(2), 159–180.

Unbehauen, H. and G. P. Rao (1987). *Identification of continuous systems.* Vol. 10 of *North-Holland Systems and Control Series.* North-Holland.

Unbehauen, H. and G. P. Rao (1998). A review of identification in continuous-time systems. *Annual Reviews in Control* **22**, 145–171.

van Huffel, S. and J. Vandevalle (1989). Analysis and Properties of the Generalized Total Least Squares Problem $AX \approx B$ when some or all Columns in A are Subject to Error. *SIAM Journal of Matrix Analysis and Applications* **10**(3), 294–315.

van Huffel, S. and J. Vandevalle (1991). *The Total Least Squares Problem, Computational Aspects and Analysis.* Frontiers in Applied Mathematics. The Society for Industrial and Applied Mathematics (SIAM).

Zeitz, M. (1990). Canonical forms for nonlinear systems. In: *Nonlinear Control Systems Design. Selected Papers from the IFAC Symposium* (A. Isidori, Ed.). Pergamon. pp. 33–38.

Appendix A. MF APPROACH FOR THE EXAMPLE SYSTEM

Multiplication of Eq. (19) with the MF ϕ_j and integration over the interval $[0, T]$ yields

$$\int_0^T (\ddot{y} + 2u\dot{y} + \dot{u}y + u^2 y)\phi_j \mathrm{d}t = \int_0^T (p_1^*(\dot{y} + uy) + p_2^* y + p_3^* u x_{10})\phi_j \mathrm{d}t,$$

and integration by parts (and applying (6)) results in

$$\int_0^T y\ddot{\phi}_j - y\dot{u}\phi_j - 2yu\dot{\phi}_j + yu^2\phi_j \mathrm{d}t = p_2^* \int_0^T y\phi_j \mathrm{d}t + p_1^*(-\int_0^T y\dot{\phi}_j + yu\phi_j \mathrm{d}t) + p_3^* \int_0^T u x_{10}\phi_j \mathrm{d}t.$$

Note that any dependency on the derivatives of y was removed. Evaluating the integrals for different modulating functions ϕ_j, results in a equation system for p_1^*, p_2^*, and p_3^*. This can be solved by OLS.