

Third-order cumulants based methods for continuous-time errors-in-variables model identification [★]

Stéphane Thil ^a, Hugues Garnier ^a, Marion Gilson ^a

^aCentre de Recherche en Automatique de Nancy – Nancy-Université, CNRS,
Faculté des Sciences et Techniques, BP 239 – 54506 Vandœuvre-lès-Nancy Cedex – France.

Abstract

In this paper, the problem of identifying stochastic linear continuous-time systems from noisy input/output data is addressed. The input of the system is assumed to have a skewed probability density function, whereas the noises contaminating the data are assumed to be symmetrically distributed. The third-order cumulants of the input/output data are then (asymptotically) insensitive to the noises, that can be colored and/or mutually correlated. Using this noise-cancellation property two computationally simple estimators are proposed. The usefulness of the proposed algorithms is assessed through a numerical simulation.

Key words: System identification; errors-in-variables; continuous-time systems; higher-order statistics.

1 Introduction

System identification is an established field in the area of system analysis and control. It aims at determining mathematical models for dynamical systems based on observed inputs and outputs. Many different solutions have been presented for system identification of linear dynamic systems from noise-corrupted output measurements (Söderström and Stoica 1989, Ljung 1999). When the input is also affected by noise, the problem becomes more difficult and two cases arise: closed-loop system identification and ‘errors-in-variables’ system identification.

The main difficulty in the former case is due to the correlation between the disturbances and the control signal, induced by the loop. Several alternatives are available to cope with this problem (Söderström and Stoica 1989, Ljung 1999). Some particular versions of these methods have been developed more recently in the area of control-relevant identification as *e.g.* the two-stage, the coprime factor, the dual-Youla, the tailor-made instrumental-variable methods. An overview of these recent developments can be found in (Gevers 1993, Forsell and

Ljung 1999, Van den Hof 1998, Gilson and Van den Hof 2005).

Models where uncertainties or measurement noises are present on both input and output observations but not necessarily linked by a feedback loop are usually called ‘errors-in-variables’ (EIV) models. This type of models plays an important role when the identification purpose is the determination of the inner laws that describe the process, rather than the prediction of its future behavior. Numerous scientific disciplines use such errors-in-variables models, including time series modeling, array signal processing for direction-of-arrival estimation, blind channel equalization, multivariate calibration in analytical chemistry, image processing, or environmental modeling (Van Huffel 1997, Van Huffel 2002).

Many methods have been proposed to solve the EIV problem for discrete-time models, see (Söderström 2006) for an overview. Nonetheless, in many areas of science and engineering, the identified dynamic models should be physically meaningful. As a result, there is a need for modeling approaches that are able to yield directly from the sampled data efficiently parameterized (parsimonious) continuous-time models that have clear physical interpretations. Although dynamical systems in the physical world are native to continuous-time (CT) domain, the attention in the system identification community was almost completely focused on the discrete-time model identification techniques until recently. The last few years has indeed witnessed considerable development in CT approaches to system identification from sampled data (see *e.g.* Garnier *et al.* (2003), Larsson *et*

[★] This paper was not presented at any IFAC meeting. Corresponding author S. Thil. Tel. +33 (0)383 684 461. Fax. +33 (0)383 684 462.

Email addresses: stephane.thil@cran.uhp-nancy.fr
(Stéphane Thil), hugues.garnier@cran.uhp-nancy.fr
(Hugues Garnier), marion.gilson@cran.uhp-nancy.fr
(Marion Gilson).

al. (2006), Rao and Unbehauen (2006)). A software package, the CONTSID toolbox for Matlab, including the majority of the time-domain methods proposed over the last thirty years, is now available (Garnier *et al.* 2006, Garnier *et al.* 2007).

Continuous-time model identification in an EIV framework is a relatively unexplored area. The first attempts have been very recently proposed (Mahata and Garnier 2006, Söderström *et al.* 2006), assuming the noises contaminating the data to be white. The whiteness of the noises allows not only to simplify the algorithms, but to rule out identifiability problems as well. Indeed, without any further assumptions on the signal and noise models, it is well-known that the general EIV model is not uniquely identifiable from second order statistics (Anderson and Deistler 1984, Agüero and Goodwin 2006). EIV systems suffer from this lack of identifiability, and it is thus of interest to study alternative approaches based on higher-order statistics. Several identification methods using higher-order statistics have been proposed for discrete-time EIV models (see *e.g.* Friedlander and Porat (1990), Tugnait (1992), Delopoulos and Giannakis (1994), Tugnait and Ye (1995)). However, to the authors' best knowledge, CT EIV system identification using higher-order statistics is an unbroken field.

This paper presents a third-order cumulant based method for CT system identification in an EIV framework, inspired by the work on discrete-time system identification of Anderson and Giannakis (1996). Two simple estimators are proposed. They can be applied in various noise situations, including the case of colored and/or correlated noises on input and output of the system. The cornerstones of the proposed solution are the use of third-order cumulants and state-variable filtering.

The paper is organized in the following way. The identification problem is formulated in Section 2. The main definitions and properties of higher-order statistics used in the proposed approach are then recalled in Section 3. Two higher-order statistics-based algorithms for continuous-time EIV models are presented in Section 4. A numerical study is presented in Section 5 before concluding in Section 6.

2 Problem statement

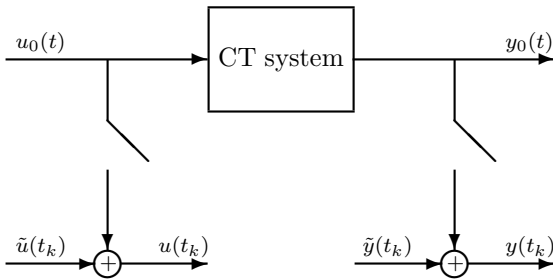


Fig. 1. CT system with noisy input/output data

Consider a CT linear time-invariant system represented in Figure 1. The noise-free input and output signals are related by

$$y_0(t) = G_0(p)u_0(t) = \frac{B_0(p)}{A_0(p)}u_0(t) \quad (1)$$

where p is the differential operator and $G_0(p)$ is the transfer function of the true system. We assume that $u_0(t)$ and $y_0(t)$ are sampled at time-instants $\{t_k\}_{k=1}^N$, not necessarily uniformly spaced. The sampled signals are both contaminated by discrete-time noise sequences, denoted as $\tilde{u}(t_k)$ and $\tilde{y}(t_k)$ respectively. The measured input and output signals are therefore given by

$$u(t_k) = u_0(t_k) + \tilde{u}(t_k) \quad (2)$$

$$y(t_k) = y_0(t_k) + \tilde{y}(t_k) \quad (3)$$

The data-generating CT system is thus characterized by

$$\mathcal{S} : \begin{cases} y_0(t) = G_0(p)u_0(t) = \frac{B_0(p)}{A_0(p)}u_0(t) \\ u(t_k) = u_0(t_k) + \tilde{u}(t_k) \\ y(t_k) = y_0(t_k) + \tilde{y}(t_k) \end{cases} \quad (4)$$

It is then parameterized as follows

$$\mathcal{G} : \begin{cases} G(p, \boldsymbol{\theta}) = B(p, \boldsymbol{\theta})/A(p, \boldsymbol{\theta}) \\ A(p, \boldsymbol{\theta}) = a_0 + a_1p + \dots + a_{n_a-1}p^{n_a-1} + p^{n_a} \\ B(p, \boldsymbol{\theta}) = b_0 + b_1p + \dots + b_{n_b}p^{n_b} \end{cases} \quad (5)$$

with $\boldsymbol{\theta}^T = [a_0 \dots a_{n_a-1} \ b_0 \dots b_{n_b}]$ and $n_a \geq n_b$.

In addition to the aforesaid conditions, we assume that

- A1. $A_0(s) \neq 0$ for $\Re(s) \geq 0$ and $A_0(s)$, $B_0(s)$ are coprime (s is the Laplace variable);
- A2. the polynomial degrees n_a and n_b are *a priori* known;
- A3. the noise-free signal $u_0(t_k)$ is a zero-mean stationary stochastic process such that its third-order cumulants are non-zero. Its probability density function (pdf) cannot therefore be symmetric;
- A4. $\tilde{u}(t_k)$ and $\tilde{y}(t_k)$ are stationary, zero-mean random variables which have a symmetric probability density function, and are independent of $u_0(t_k)$ and $y_0(t_k)$.

REMARK 1

Assumption A1 aggregates the traditional assumptions of stability and observability of the system. We also suppose that the system (1) has been in operation for a time long enough so that the output y_0 is third-order wide-sense stationary, i.e. its moments of order up to three are invariant to any time shift.

Assumption A2 means that only the system belongs to the model class, a situation denoted as $G_0 \in \mathcal{G}$ (Ljung 1999). Assumptions A3 and A4 are linked to properties of higher-order statistics, and will be justified in the sequel. Note that, except A4, there is no other assumption on the noises. They can thus be white or colored, and mutually correlated or not. ■

The identification problem can then be formulated as follows: given N samples of noisy input/output data $\{u(t_k), y(t_k)\}_{k=1}^N$, directly estimate the continuous-time parameter vector $\boldsymbol{\theta}$.

3 Definition and properties of third-order cumulants

The identification technique developed in this paper is based on higher-order statistics (HOS) (Brillinger 1981, Mendel 1991, Lacoume *et al.* 1997). The main definitions and properties used in the proposed approach are recalled in this section.

The third-order cumulant of a real-valued, zero-mean stationary random process $x(t_k)$ is defined as

$$\begin{aligned} C_{xxx}(\tau_1, \tau_2) &= \text{Cum}[x(t_k), x(t_k + \tau_1), x(t_k + \tau_2)] \quad (6) \\ &= E[x(t_k)x(t_k + \tau_1)x(t_k + \tau_2)] \quad (7) \end{aligned}$$

The cumulants of order higher than two have many properties amongst which we only recall those used in the proposed estimation scheme.

Let $\mathbf{x} = [x(t_1), \dots, x(t_n)]^T$ and $\mathbf{y} = [y(t_1), \dots, y(t_n)]^T$ be two random vectors.

P1. Multilinearity: the cumulants are linear with respect to each of their arguments. If $\alpha_i, \beta_j, \gamma_k$ are scalars

$$\begin{aligned} \text{Cum} \left[\sum_i \alpha_i x(t_i), \sum_j \beta_j x(t_j), \sum_k \gamma_k x(t_k) \right] \\ = \sum_{i,j,k} \alpha_i \beta_j \gamma_k \text{Cum}[x(t_i), x(t_j), x(t_k)] \end{aligned}$$

P2. Additivity: if \mathbf{x} and \mathbf{y} are *independent*, the cumulant of their sum equals the sum of their cumulants

$$\begin{aligned} \text{Cum}[x(t_1) + y(t_1), \dots, x(t_n) + y(t_n)] \\ = \text{Cum}[x(t_1), \dots, x(t_n)] + \text{Cum}[y(t_1), \dots, y(t_n)] \end{aligned}$$

P3. The third-order cumulant of a random variable with a symmetric pdf (consequently including Gaussian distributions) is equal to zero.

Now the relevance of assumptions A3 and A4 is pointed out: if the third-order cumulant of the available input signal (2) is considered, using properties P2 and P3 yields

$$C_{uuu}(\tau_1, \tau_2) = C_{u_0 u_0 u_0}(\tau_1, \tau_2) + C_{\tilde{u}\tilde{u}\tilde{u}}(\tau_1, \tau_2) \quad (8)$$

$$= C_{u_0 u_0 u_0}(\tau_1, \tau_2) \quad (9)$$

In the same manner this result is obviously true for the third-order cumulant of any combination of input/output signals. The use of third-order cumulants therefore allows to get naturally rid of the noise contaminating the input/output data, under assumptions A3-A4.

4 HOS-based methods for continuous-time models

PROPOSITION 1

The third-order cross-cumulant between the measured input/output signals satisfies

$$C_{uyu}(\tau_1, \tau_2) = \frac{B(p, \boldsymbol{\theta})}{A(p, \boldsymbol{\theta})} C_{uuu}(\tau_1, \tau_2) \quad (10)$$

where the differential operator p stands for $\frac{\partial}{\partial \tau_1}$.

PROOF

See Appendix A. ■

This result is the starting point of the identification methods: the differential equation of the system is also satisfied by the third-order cumulants. If the input/output signals are contaminated by noises with symmetric pdf, then equation (10) only involves noise-free terms.

In practice, however, the cumulants are estimated from N samples of input/output data $\{u(t_k), y(t_k)\}_{k=1}^N$, replacing mathematical expectation by sample averages

$$\begin{aligned} \hat{C}_{x_1 x_2 x_3}(\tau_1, \tau_2) \\ = \frac{1}{N - \mu} \sum_{k=1}^{N-\mu} x_1(t_k) x_2(t_k + \tau_1) x_3(t_k + \tau_2) \quad (11) \end{aligned}$$

where $\mu = \max(\tau_1, \tau_2)$ and for generality purposes x_i represents either u or y . Hence, with the available data we get

$$\hat{C}_{uyu}(\tau_1, \tau_2) = \frac{B(p, \boldsymbol{\theta})}{A(p, \boldsymbol{\theta})} \hat{C}_{uuu}(\tau_1, \tau_2) + \varepsilon(\tau_1, \tau_2, \boldsymbol{\theta}) \quad (12)$$

$$\varepsilon(\tau_1, \tau_2, \boldsymbol{\theta}) = \frac{B(p, \boldsymbol{\theta})}{A(p, \boldsymbol{\theta})} \tilde{C}_{uuu}(\tau_1, \tau_2) - \tilde{C}_{uyu}(\tau_1, \tau_2) \quad (13)$$

where $\tilde{C}_{x_1 x_2 x_3} = C_{x_1 x_2 x_3} - \hat{C}_{x_1 x_2 x_3}$ denotes the estimation error of $C_{x_1 x_2 x_3}$. Since the cumulant estimates (11) are unbiased and consistent (see *e.g.* Porat and Friedlander (1989)), it holds that

$$\lim_{N \rightarrow \infty} \varepsilon(\tau_1, \tau_2, \boldsymbol{\theta}) = 0 \quad \text{w.p. 1} \quad (14)$$

4.1 Cumulant-based LS approach

Equation (12) can be rewritten as a linear regression

$$\hat{C}_{uyu}^{(n_a)}(\tau_1, \tau_2) = \hat{\boldsymbol{\Phi}}^T(\tau_1, \tau_2) \boldsymbol{\theta} + A(p, \boldsymbol{\theta}) \varepsilon(\tau_1, \tau_2, \boldsymbol{\theta}) \quad (15)$$

where

$$\widehat{C}_{uyu}^{(j)}(\tau_1, \tau_2) \triangleq \frac{\partial^j}{\partial \tau_1^j} \widehat{C}_{uyu}(\tau_1, \tau_2) \quad (16)$$

and the regression vector is given by

$$\widehat{\Phi}^T(\tau_1, \tau_2) = \begin{bmatrix} -\widehat{C}_{uyu}^{(0)}(\tau_1, \tau_2) \dots -\widehat{C}_{uyu}^{(n_a-1)}(\tau_1, \tau_2) \\ \widehat{C}_{uuu}^{(0)}(\tau_1, \tau_2) \dots \widehat{C}_{uuu}^{(n_b)}(\tau_1, \tau_2) \end{bmatrix} \quad (17)$$

Define then the error function

$$e(\tau_1, \tau_2, \boldsymbol{\theta}) = \widehat{C}_{uyu}^{(n_a)}(\tau_1, \tau_2) - \widehat{\Phi}^T(\tau_1, \tau_2) \boldsymbol{\theta} \quad (18)$$

$$= B(p, \boldsymbol{\theta}) \tilde{C}_{uuu}(\tau_1, \tau_2) - A(p, \boldsymbol{\theta}) \tilde{C}_{uyu}(\tau_1, \tau_2) \quad (19)$$

Minimizing the following cost function with respect to the parameter vector $\boldsymbol{\theta}$

$$V(\tau_2, \boldsymbol{\theta}, M) = \frac{1}{M} \sum_{\tau_1=0}^{M-1} \frac{1}{2} e^2(\tau_1, \tau_2, \boldsymbol{\theta}) \quad (20)$$

leads to the tocls estimator (Third-Order Cumulants based Least Squares algorithm), namely

$$\widehat{\boldsymbol{\theta}}_{\text{tocls}}(\tau_2, M) = \begin{bmatrix} \frac{1}{M} \sum_{\tau_1=0}^{M-1} \widehat{\Phi}(\tau_1, \tau_2) \widehat{\Phi}^T(\tau_1, \tau_2) \\ \frac{1}{M} \sum_{\tau_1=0}^{M-1} \widehat{\Phi}(\tau_1, \tau_2) \widehat{C}_{uyu}^{(n_a)}(\tau_1, \tau_2) \end{bmatrix}^{-1} \quad (21)$$

The ‘noise’ in the right-hand side of the error function (19) is composed of the estimation errors \tilde{C}_{uuu} and \tilde{C}_{uyu} filtered by A and B . It is thus not expected to be white, and the tocls estimator is in turn not expected to be unbiased, from a finite data set. However, as N tends towards infinity, the sample averages (11) tend toward the third-order cumulants with probability one. Since (10) only involves noise-free terms and is solved in a least squares sense under the assumption $G_0 \in \mathcal{G}$ we have

$$\lim_{N \rightarrow \infty} \widehat{\boldsymbol{\theta}}_{\text{tocls}}(\tau_2, M) = \boldsymbol{\theta}_0 \quad \text{w.p.1} \quad (22)$$

where $\boldsymbol{\theta}_0$ is the ‘true’ parameter vector, *i.e.* the parameter vector such that $G(p, \boldsymbol{\theta}_0) = G_0(p)$.

The role of the user parameter M that appears in the cost function (20) consists in preventing the estimation of cumulants with large time-lags. We could indeed calculate that cost function with all the available data, *i.e.* setting the upper bound of the sum in (20) to N . However, the cumulants are estimated by sample averages, and as the time-lags increase these estimates will be calculated with less data, thus becoming less reliable. Note that it is required that $M \geq n_a + n_b + 1$ so that the matrix to be inverted in (21) is non-singular.

The influence of M is illustrated with the help of a numerical example in Section 5.

REMARK 2

Several choices of cost functions are possible; (20) may notably be replaced by

$$V_2(\boldsymbol{\theta}, M_1, M_2) = \frac{1}{M_1 M_2} \sum_{\tau_1=0}^{M_1-1} \sum_{\tau_2=0}^{M_2-1} \frac{1}{2} e^2(\tau_1, \tau_2, \boldsymbol{\theta}) \quad (23)$$

leading to the estimator

$$\widehat{\boldsymbol{\theta}}_{V_2}(M_1, M_2) = \begin{bmatrix} \frac{1}{M_1 M_2} \sum_{\tau_1=0}^{M_1-1} \sum_{\tau_2=0}^{M_2-1} \widehat{\Phi}(\tau_1, \tau_2) \widehat{\Phi}^T(\tau_1, \tau_2) \\ \frac{1}{M_1 M_2} \sum_{\tau_1=0}^{M_1-1} \sum_{\tau_2=0}^{M_2-1} \widehat{\Phi}(\tau_1, \tau_2) \widehat{C}_{uyu}^{(n_a)}(\tau_1, \tau_2) \end{bmatrix}^{-1} \quad (24)$$

By definition the third-order cumulants are two-dimensional, in the sense that they depend on two variables (here τ_1 and τ_2). From the available data we can thus derive a matrix constituted of the third-order cumulants calculated at different points. The second estimator $\widehat{\boldsymbol{\theta}}_{V_2}(M_1, M_2)$ can then be interpreted as the solution of the minimization of the error function (19) considering a $M_1 \times M_2$ submatrix of the aforementioned matrix, while the first estimator $\widehat{\boldsymbol{\theta}}_{\text{tocls}}(\tau_2, M)$ is obtained considering only a ‘slice’ of that matrix, *i.e.* a vector; hence the dependence on τ_2 . These slices are useful in applications of cumulants, as they allow a trade-off between accuracy and computational load. In numerical simulations, the second estimator $\widehat{\boldsymbol{\theta}}_{V_2}(M_1, M_2)$ did not give a significant improvement, while being more time-consuming. Lastly, the slice should always be chosen in order to maximize the number of data points from which the cumulants are estimated, and in the simulation example to follow we chose to set $\tau_2 = 0$. ■

4.2 Cumulant-based iterative LS approach

Equations (12)-(13) define an output error given by

$$\varepsilon(\tau_1, \tau_2, \boldsymbol{\theta}) = \widehat{C}_{uyu}(\tau_1, \tau_2) - \frac{B(p, \boldsymbol{\theta})}{A(p, \boldsymbol{\theta})} \widehat{C}_{uuu}(\tau_1, \tau_2) \quad (25)$$

$$= \frac{B(p, \boldsymbol{\theta})}{A(p, \boldsymbol{\theta})} \tilde{C}_{uuu}(\tau_1, \tau_2) - \tilde{C}_{uyu}(\tau_1, \tau_2) \quad (26)$$

This error is nonlinear in the parameters. A non-linear optimization scheme can therefore be used to estimate the parameter vector. However, to keep the algorithm simple and tractable, a second approach based on an iterative LS procedure is proposed. This approach has been previously suggested for discrete-time model identification in (Anderson and Giannakis 1996), and is extended here to continuous-time model identification. The preceding output error (25) can be rewritten as

$$\begin{aligned}
\varepsilon(\tau_1, \tau_2, \boldsymbol{\theta}) &= \frac{1}{A(p, \boldsymbol{\theta})} \left(A(p, \boldsymbol{\theta}) \widehat{C}_{uyu}(\tau_1, \tau_2) - B(p, \boldsymbol{\theta}) \widehat{C}_{uuu}(\tau_1, \tau_2) \right) \\
&= A(p, \boldsymbol{\theta}) \widehat{C}_{uyu,f}(\tau_1, \tau_2) - B(p, \boldsymbol{\theta}) \widehat{C}_{uuu,f}(\tau_1, \tau_2) \quad (27)
\end{aligned}$$

where $\widehat{C}_{uuu,f}$ and $\widehat{C}_{uyu,f}$ denote the cumulants filtered by $1/A(p, \boldsymbol{\theta})$. An equation error is thus obtained, linear in the parameters, which can be estimated by the least squares method.

Since $A(p, \boldsymbol{\theta})$ is unknown, the idea is to proceed in an iterative fashion to transform the equation error (27) into the output error (25). Let $\widehat{\boldsymbol{\theta}}^i$ be the estimate of $\boldsymbol{\theta}$ at the i th iteration. At each iteration, $\widehat{\boldsymbol{\theta}}^{i+1}$ is given by a least squares estimate, using the cumulants filtered by $1/A(p, \widehat{\boldsymbol{\theta}}^i)$. Insomuch as the parameters converge to a constant value, we have

$$\frac{A(p, \widehat{\boldsymbol{\theta}}^{i+1})}{A(p, \widehat{\boldsymbol{\theta}}^i)} \longrightarrow 1 \quad \text{and} \quad \frac{B(p, \widehat{\boldsymbol{\theta}}^{i+1})}{A(p, \widehat{\boldsymbol{\theta}}^i)} \longrightarrow \frac{B(p, \widehat{\boldsymbol{\theta}}^{i+1})}{A(p, \widehat{\boldsymbol{\theta}}^{i+1})}$$

Hence, the equation error (27) tends toward the output error (25). The convergence of this algorithm is very fast: typically a few iterations are enough (Walter and Pronzato 1997).

The equation error (27) can be rewritten as a linear regression

$$\varepsilon(\tau_1, \tau_2, \boldsymbol{\theta}) = \widehat{C}_{uyu,f}^{(n_a)}(\tau_1, \tau_2) - \widehat{\boldsymbol{\Phi}}_f^T(\tau_1, \tau_2) \boldsymbol{\theta} \quad (28)$$

where the regression vector is

$$\widehat{\boldsymbol{\Phi}}_f^T(\tau_1, \tau_2) = \left[-\widehat{C}_{uyu,f}^{(0)}(\tau_1, \tau_2) \quad \dots \quad -\widehat{C}_{uyu,f}^{(n_a-1)}(\tau_1, \tau_2) \right. \\
\left. \widehat{C}_{uuu,f}^{(0)}(\tau_1, \tau_2) \quad \dots \quad \widehat{C}_{uuu,f}^{(n_b)}(\tau_1, \tau_2) \right] \quad (29)$$

Minimizing the following criterion at each iteration

$$V_i(\tau_2, \boldsymbol{\theta}, M) = \frac{1}{M} \sum_{\tau_1=0}^{M-1} \frac{1}{2} \varepsilon_i^2(\tau_1, \tau_2, \boldsymbol{\theta}) \quad (30)$$

where ε_i is given by

$$\begin{aligned}
\varepsilon_i(\tau_1, \tau_2, \boldsymbol{\theta}) &= A(p, \boldsymbol{\theta}) \frac{\widehat{C}_{uyu}(\tau_1, \tau_2)}{A(p, \boldsymbol{\theta}^i)} - B(p, \boldsymbol{\theta}) \frac{\widehat{C}_{uuu}(\tau_1, \tau_2)}{A(p, \boldsymbol{\theta}^i)} \quad (31)
\end{aligned}$$

leads to the tocils estimator (Third-Order Cumulants based Iterative Least Squares algorithm)

$$\begin{aligned}
\widehat{\boldsymbol{\theta}}_{\text{tocils}}(\tau_2, M) &= \left[\frac{1}{M} \sum_{\tau_1=0}^{M-1} \widehat{\boldsymbol{\Phi}}_f(\tau_1, \tau_2) \widehat{\boldsymbol{\Phi}}_f^T(\tau_1, \tau_2) \right]^{-1} \\
&\quad \left[\frac{1}{M} \sum_{\tau_1=0}^{M-1} \widehat{\boldsymbol{\Phi}}_f(\tau_1, \tau_2) \widehat{C}_{uyu,f}^{(n_a)}(\tau_1, \tau_2) \right] \quad (32)
\end{aligned}$$

REMARK 3

This iterative method is greatly inspired by the algorithm of Steiglitz and McBride (1965). There are however a major difference: the Steiglitz-McBride algorithm uses the iterative LS estimator on measured I/O data rather than on the cumulants, being therefore applicable under quite restrictive assumptions (white noise on output) as it is pointed out in (Stoica and Söderström 1981). The proposed method uses that procedure on the third-order cumulants, and since they are insensitive to symmetrically distributed noises, the same restrictions do not apply. ■

5 Handling of the cumulant time-derivatives: the SVF approach

The time-derivatives of the cumulants are needed to build up the regression vectors (17), (29). These can be obtained *e.g.* by using the traditional state variable filtering (SVF) approach, whose basics are quickly recalled in the sequel.

The SVF approach allows to reconstruct the time-derivatives of a signal by passing it into the following filter bank (Young 1964, Young 1981)

$$F_n(p) = p^n \left(\frac{\lambda}{p + \lambda} \right)^{n_a}, \quad 0 \leq n \leq n_a \quad (33)$$

where λ is a user parameter accounting for the filter cut-off frequency. Intuitively, it can be chosen in order to emphasize the frequency band of interest and generally, it should be chosen close to, or larger than the bandwidth of the system to be identified.

The filter bank outputs provide the time-derivatives of the signal in the bandwidth of interest

$$x_f^{(n)}(t) = F_n(p)x(t) = [f_n \star x](t) \quad (34)$$

where \star stands for the convolution operator and $f_n(t)$ denotes the impulse response of the linear time-invariant filter $F_n(p)$

$$f_n(t) = \mathcal{L}^{-1} [F_n(s)] \quad (35)$$

with \mathcal{L} symbolizing the Laplace transform.

The time-derivatives of the cumulants are now computed using this SVF approach. As a result of their multilinearity property, a formula for cumulant filtering can be derived (Lacoume *et al.* 1997, Nikias and Petropulu 1993)

$$\begin{aligned}
C_{xxx,f}^{(n)}(\tau_1, \tau_2) &= F_n(p)C_{xxx}(\tau_1, \tau_2) \quad (36) \\
&= \int_{\mathbb{R}} f_n(\tau) E[x(t)x(t + \tau_1 - \tau)x(t + \tau_2)] d\tau \\
&= E \left[x(t) \left(\int_{\mathbb{R}} f_n(\tau) x(t + \tau_1 - \tau) d\tau \right) x(t + \tau_2) \right] \\
&= E \left[x(t) x_f^{(n)}(t + \tau_1) x(t + \tau_2) \right] \quad (37)
\end{aligned}$$

We thus obtain

$$C_{xxx,f}^{(n)}(\tau_1, \tau_2) = C_{xx_f^{(n)}x}(\tau_1, \tau_2) \quad (38)$$

Consequently, the use of the SVF approach, that is the application on both sides of equation (10) of the filters $\{F_n(s)\}_{n=0}^{n_a}$, allows to transfer the cumulant time-derivative estimation problem to an input/output signal time-derivative estimation problem, which constitutes a well-known task in continuous-time model identification (Garnier *et al.* 2003).

In the tocls algorithm, the basic filter bank (33) is applied to the input/output signals. Their time-derivatives in the bandwidth of interest are therefore obtained, from which the cumulants time-derivatives are computed. In the tocils algorithm however, the basic filter bank (33) is applied to the input/output signals only at the initialization stage (or equivalently: the tocils algorithm is initialized with the tocls estimate). In the subsequent iterations, as an estimate $\hat{\theta}_{\text{tocils}}^i(\tau_2, M)$ is available, the following filter bank is used

$$F_n(p) = \frac{p^n}{A(p, \hat{\theta}_{\text{tocils}}^i(\tau_2, M))}, \quad 0 \leq n \leq n_a \quad (39)$$

Note that this filter bank allows to realize both the filtering by $1/A(p, \hat{\theta}_{\text{tocils}}^i(\tau_2, M))$ and the differentiation in one step only.

The main steps of both algorithms are given below.

The tocls algorithm

The tocls estimation approach can be summarized by the following algorithm

- (1) Design a bank of n_a -order SVF filters

$$F_n(p) = p^n \left(\frac{\lambda}{p + \lambda} \right)^{n_a}, \quad 0 \leq n \leq n_a$$

Generate the time-derivatives $u_f^{(n)}(t_k)$ and $y_f^{(n)}(t_k)$ of both observed input and output signals as

$$\begin{cases} u_f^{(n)}(t_k) = F_n(p)u(t_k) \\ y_f^{(n)}(t_k) = F_n(p)y(t_k) \end{cases}, \quad 0 \leq n \leq n_a$$

- (2) Calculate the third-order cumulants and their time-derivatives for $0 \leq n \leq n_a$

$$\begin{aligned} \hat{C}_{uuu,f}^{(n)}(\tau_1, \tau_2) &= \frac{1}{N - \mu} \sum_{k=1}^{N-\mu} u(t_k)u_f^{(n)}(t_k + \tau_1)u(t_k + \tau_2) \end{aligned}$$

$$\begin{aligned} \hat{C}_{uyu,f}^{(n)}(\tau_1, \tau_2) &= \frac{1}{N - \mu} \sum_{k=1}^{N-\mu} u(t_k)y_f^{(n)}(t_k + \tau_1)u(t_k + \tau_2) \end{aligned}$$

where $\mu = \max(\tau_1, \tau_2)$;

- (3) Build up the regression vector as

$$\hat{\Phi}_f^T(\tau_1, \tau_2) = \begin{bmatrix} -\hat{C}_{uyu,f}^{(0)}(\tau_1, \tau_2) & \dots & -\hat{C}_{uyu,f}^{(n_a-1)}(\tau_1, \tau_2) \\ \hat{C}_{uuu,f}^{(0)}(\tau_1, \tau_2) & \dots & \hat{C}_{uuu,f}^{(n_b)}(\tau_1, \tau_2) \end{bmatrix}$$

Calculate the tocls estimator (21)

$$\hat{\theta}_{\text{tocls}}(\tau_2, M) = \begin{bmatrix} \frac{1}{M} \sum_{\tau_1=0}^{M-1} \hat{\Phi}_f(\tau_1, \tau_2) \hat{\Phi}_f^T(\tau_1, \tau_2) \\ \frac{1}{M} \sum_{\tau_1=0}^{M-1} \hat{\Phi}_f(\tau_1, \tau_2) \hat{C}_{uyu,f}^{(n_a)}(\tau_1, \tau_2) \end{bmatrix}^{-1}$$

The tocils algorithm

Only the major steps of the algorithm are given below.

- (1) Initialization stage.

The first estimate $\hat{\theta}_{\text{tocils}}^1(\tau_2, M)$ is given by the tocls algorithm

$$\hat{\theta}_{\text{tocils}}^1(\tau_2, M) = \hat{\theta}_{\text{tocls}}(\tau_2, M)$$

- (2) $(i+1)$ th step. As an estimate $\hat{\theta}_{\text{tocils}}^i(\tau_2, M)$ is available, execute the same steps as in the tocls algorithm with

$$F_n(p) = \frac{p^n}{\hat{A}_i(p, \hat{\theta}_{\text{tocils}}^i(\tau_2, M))}$$

to get $\hat{\theta}_{\text{tocils}}^{i+1}(\tau_2, M)$.

- (3) Go back to step 2 above until the difference between two successive estimates becomes small enough.

REMARK 4

The general EIV model is not identifiable from second order statistics. Approaches based on them thus require additional assumptions on the noise models and/or the system to provide consistent parameter estimates. The algorithms presented above also require an additional assumption, on the noise-free input: it must have a skewed distribution. The advantage of the proposed methods is that, without any change, they can handle white or colored noise affecting the system (as long as they are symmetrically distributed, which is a natural assumption). Assuming the noise-free input to have a skewed distribution is certainly restrictive. However, indications on the 'applicability' of the methods can be obtained, such as the magnitude third-order cumulants calculated from the noisy data, or the conditioning of the matrix to be inverted in (21). ■

6 Numerical example

This section is divided into three parts, where some aspects of the algorithms are studied with the help of Monte Carlo simulations of $n_{mc} = 100$ runs.

In the first part, the performance of the proposed algorithms are analyzed and compared to those obtained with the *srvc* and *ivsvf*¹ methods (Young 1970, Young and Jakeman 1980, Young *et al.* 2007), both available in the *CONTSID* toolbox². In the second part, the influence of the user parameter M on the quality of the estimates is assessed. In the final part, the influence of the SVF cut-off frequency on the proposed methods is studied. In each part, the following system is considered

$$G_0(p) = \frac{p-1}{p^2+2p+1} \quad (40)$$

The noise-free input signal u_0 is obtained as the output of the filter (q stands for the shift operator)

$$H(q) = 1 - 0.2q^{-1} + 0.3q^{-2} \quad (41)$$

driven by a zero-mean chi-square distributed (with two degrees of freedom) independent identically distributed sequence. The noises contaminating the input/output signals are defined as

$$\tilde{u}(t_k) = (1 + 2q^{-1} - q^{-2}) e(t_k) \quad (42)$$

$$\tilde{y}(t_k) = \frac{\alpha}{1 + 0.8q^{-1}} \tilde{u}(t_k) \quad (43)$$

where $e(t_k)$ is a zero-mean uniformly distributed white noise and α is a constant. The variance of $e(t)$ and α are then adjusted so that the signal-to-noise ratio (SNR) is equal to 5 dB on both input and output, with

$$\text{SNR} = 10 \log_{10} (P_{x_0}/P_{\tilde{x}}) \quad (44)$$

where $P_{\tilde{x}}$ represents the average power of the zero-mean additive noise while P_{x_0} denotes the average power of the noise-free signal fluctuations.

The noises are thus colored and mutually correlated.

6.1 Performance analysis

We first focus on the performance of the proposed algorithms and compare them with the performance of *ivsvf* and *srvc*. Although these two methods are not specifically developed for EIV model identification, they show a robustness (to some degree) to the input noise. This phenomenon can be explained by the filtering inherent to these methods, since applying a linear transformation to the data has been shown to have a regularization effect (Moussaoui *et al.* 2005).

¹ The acronyms stand for *Simplified Refined Instrumental Variable for Continuous-time models* and *Instrumental Variable based State Variable Filter*.

² Available at <http://www.uhp-nancy.fr/contsid/>.

HOS-based methods are known to require a large number of data to achieve a good accuracy (Mendel 1991). To assess the influence of the data length, the algorithms have been applied for $N = 1000$ and $N = 5000$.

Table 1 contains the mean and standard deviation of the estimates, as well as the normalized root mean square error, defined as

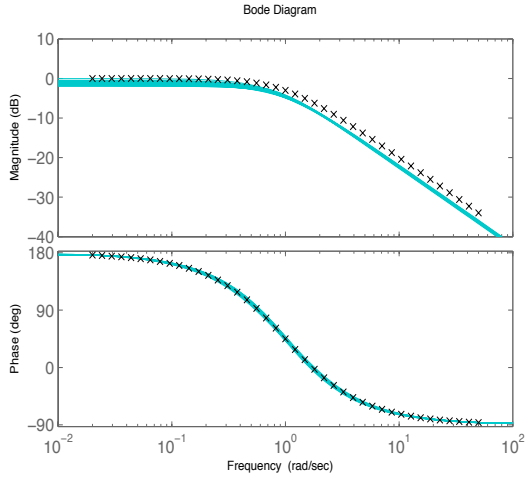
$$\text{NRMSE} \triangleq \sqrt{\frac{1}{n_{mc}} \sum_{j=1}^{n_{mc}} \frac{\|\hat{\theta}_j - \theta\|^2}{\|\theta\|^2}} \quad (45)$$

where $\|\cdot\|$ is the Euclidian norm and $\hat{\theta}_j$ is the parameter vector estimate obtained at the j th run of the Monte Carlo simulation. The other entries of Table 1 are λ , M and $\#it$, respectively the cut-off frequency of the SVF, the user parameter of the proposed methods and the mean of the iteration number.

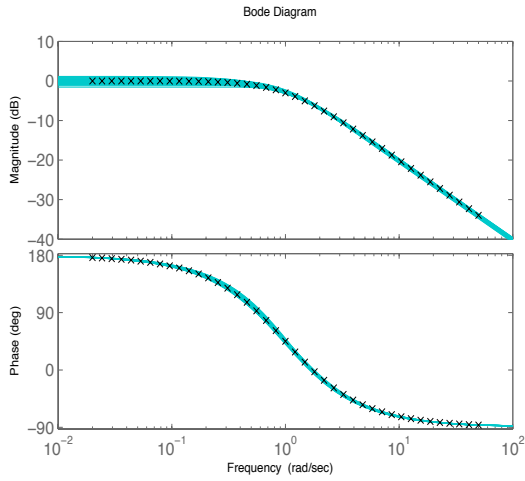
Table 1 shows as expected that the *ivsvf* and the *srvc* methods give biased estimates. As the number of data increases, the bias of both methods does not significantly decrease (and even slightly increases), confirming that they do not deliver consistent estimates in an EIV context, in contrast to the proposed algorithms. Indeed in both *tocls* and *tocils* the bias is reduced. This reduction of the bias is made at the expense of a higher variance: it is well-known that HOS-based estimators deliver a higher variance than those based on second-order statistics (Delopoulos and Giannakis 1994). This is confirmed here: although remaining low, the variance is roughly doubled. When $N = 1000$, the proposed algorithms give results similar to the *srvc* method in terms of NRMSE. However when $N = 5000$ both *tocls* and *tocils* give very good estimates. Finally, when comparing the two proposed algorithms, we see that *tocils* gives better estimates than *tocls* in terms of bias (thus indicating that the error minimized by the *tocils* method is 'whiter' than the one minimized by the *tocls* method), but exhibits a slightly larger variance. The Bode diagrams of the true and estimated models, plotted in Figure 2, corroborate this analysis.

	λ	M	N	$a_1 = 2$	$a_0 = 1$	$b_1 = 1$	$b_0 = -1$	NRMSE	$\#it$
ivsvf	3	-	1000	1.884 ± 0.100	0.907 ± 0.088	0.696 ± 0.024	-0.973 ± 0.061	14.2%	-
			5000	1.842 ± 0.055	0.941 ± 0.040	0.656 ± 0.013	-0.971 ± 0.025	13.8%	-
srvc	-	-	1000	1.807 ± 0.123	0.973 ± 0.075	0.773 ± 0.033	-0.785 ± 0.065	13.1%	7.1
			5000	1.900 ± 0.072	0.943 ± 0.034	0.765 ± 0.018	-0.821 ± 0.028	11.6%	5.2
tocls	3	50	1000	1.861 ± 0.191	0.989 ± 0.124	0.942 ± 0.055	-0.969 ± 0.113	10.7%	-
			5000	1.939 ± 0.087	0.992 ± 0.064	0.982 ± 0.030	-0.987 ± 0.042	4.8%	-
tocils	-	50	1000	1.932 ± 0.219	0.977 ± 0.139	0.954 ± 0.068	-0.980 ± 0.147	11.2%	3.5
			5000	1.992 ± 0.104	0.992 ± 0.065	0.994 ± 0.037	-0.999 ± 0.056	4.9%	2.9

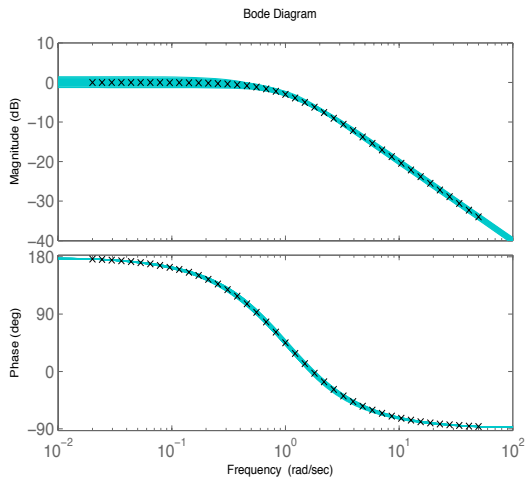
Table 1
Mean and standard deviation of the parameter estimates.



(a) srivc algorithm



(b) tocls algorithm



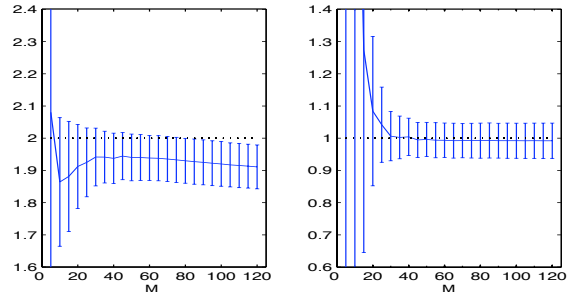
(c) tocls algorithm

Fig. 2. Bode diagrams of the true ('x') and estimated models ($N = 5000$ data points, $M = 50$).

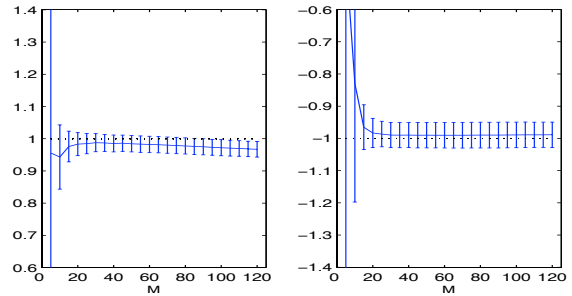
6.2 Influence of the user parameter M

To study the influence of the user parameter M on the quality of the estimates, the proposed algorithms have been applied for different values of M , the number of available data being $N = 5000$.

The mean and standard deviation of the parameters estimates as a function of M are represented in Fig. 3 for the tocls method and in Fig. 4 for the tocls method. In Fig. 5 the normalized root mean square error is plotted as a function of M . It can be seen that for very small values of M the estimates are bad, since the cumulants are then estimated from only a few samples. However, as M increases they become accurate, resulting in an accurate estimation of the parameter vector. Note that as M becomes larger a very small bias appears for both the tocls and tocls methods, the latter seeming more sensitive. The NRMSE plot (Fig. 5) summarizes this phenomenon: it first swiftly decreases, stabilizes, and then slowly increases for large values of M . The 'best' values of M are dependent on the system and the experimental conditions in a way that remains to be determined. However, the two methods are not too sensitive to this user parameter M , as long as it is not chosen too small.



(a) Denominator parameters: $a_1 = 2, a_0 = 1$



(b) Numerator parameters: $b_1 = 1, b_0 = -1$

Fig. 3. tocls algorithm – Mean and standard deviation of the parameter estimates as a function of the user parameter M .

6.3 Influence of the user parameter λ

In this paragraph the robustness of the algorithms with respect to the SVF cut-off frequency is studied. The value of M is set to 50 and the proposed algorithms have been applied with $0.4 \leq \lambda \leq 4.9$, the number of available data being $N = 5000$. The results of the simulation

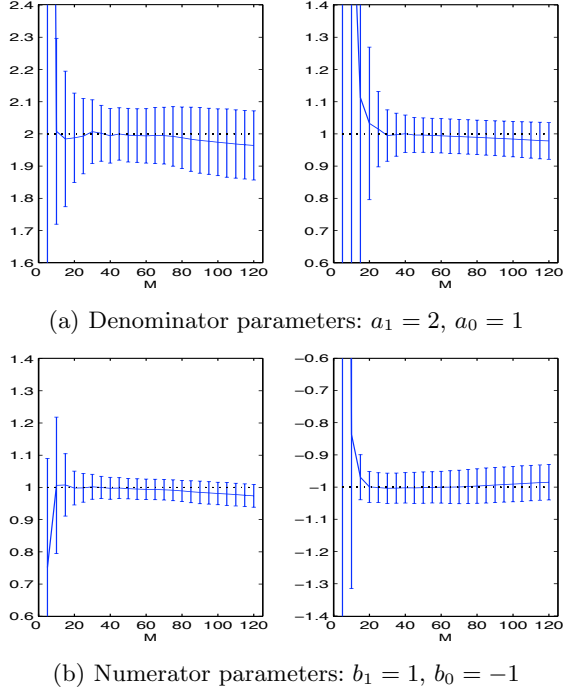


Fig. 4. `tocils` algorithm – Mean and standard deviation of the parameter estimates as a function of the user parameter M .

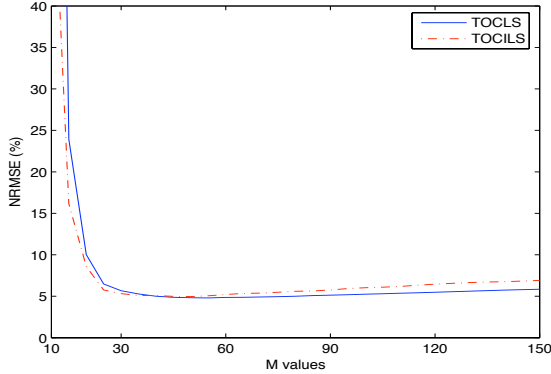


Fig. 5. NRMSE as a function of the user parameter M

are given in Table 2.

It can be seen that this parameter affects the performance of the `tocls` method: as may be expected, the closer the cut-off value of the filter to the one of the system, the smaller the NRMSE. As for the `tocils` method, since λ is only used for the initialization, it only influences the number of iterations before convergence. The closer the initialisation to the true value of the parameters, the quicker the convergence.

7 Conclusion

Continuous-time model identification in an errors-in-variables framework is a relatively unexplored area, and this paper presents an attempt to solve this problem.

	λ	$a_1 = 2$	$a_0 = 1$	$b_1 = 1$	$b_0 = -1$	NRMSE	#it
<code>tocls</code>	0.4	2.010 ± 0.123	0.999 ± 0.061	1.004 ± 0.106	-0.989 ± 0.062	6.5%	–
	1.3	1.998 ± 0.098	0.999 ± 0.056	0.995 ± 0.033	-0.990 ± 0.053	4.6%	–
	2.2	1.976 ± 0.093	0.998 ± 0.056	0.989 ± 0.027	-0.986 ± 0.050	4.4%	–
	3.1	1.940 ± 0.096	1.000 ± 0.055	0.981 ± 0.028	-0.978 ± 0.050	5.0%	–
	4.0	1.884 ± 0.102	1.006 ± 0.055	0.969 ± 0.031	-0.964 ± 0.050	6.4%	–
	4.9	1.806 ± 0.110	1.019 ± 0.056	0.951 ± 0.035	-0.945 ± 0.052	8.8%	–
<code>tocils</code>	0.4						2.7
	1.3						2.3
	2.2	2.002 ± 0.102	0.999 ± 0.057	0.997 ± 0.040	-0.990 ± 0.055	4.8%	2.8
	3.1						3.0
	4.0						3.1
	4.9						3.1

Table 2
Results for $M = 50$ and $0.4 \leq \lambda \leq 4.9$.

Two new methods to consistently identify a continuous-time model in an EIV framework have been presented. They are based on the use of third-order cumulants and their noise-cancellation property. The system is assumed to be excited by an input with a skewed distribution and contaminated by symmetrically distributed noises. As no other assumption is made on the measurement noises, they can be colored and even mutually correlated. That case cannot be handled by the few existing continuous-time EIV methods, since the general EIV model is not identifiable (without further assumptions) by methods based on second order statistics. Numerical simulations have illustrated the performance of the proposed methods in the case of colored noises. Future developments will investigate the use of fourth-order cumulants, that allow the use of an excitation signal with a non-skewed distribution.

A Proof of proposition 1

Let $g(t)$ be the impulse response of the system (5). As a result of their multilinearity property, a formula for cumulant filtering can be derived (Lacoume *et al.* 1997, Nikias and Petropulu 1993). Consider the third-order cross-cumulant between the measured input and the measured output $u(t)$ and $y(t)$. Then

$$C_{uyu}(\tau_1, \tau_2) = C_{u_0y_0u_0}(\tau_1, \tau_2) \quad (\text{A.1})$$

$$= E[u_0(t)y_0(t+\tau_1)u_0(t+\tau_2)] \quad (\text{A.2})$$

$$= E\left[u_0(t)\left(\int_{\mathbb{R}} g(\tau)u_0(t+\tau_1-\tau)d\tau\right)u_0(t+\tau_2)\right] \quad (\text{A.3})$$

$$= \int_{\mathbb{R}} g(\tau)E[u_0(t)u_0(t+\tau_1-\tau)u_0(t+\tau_2)]d\tau \quad (\text{A.4})$$

$$= \int_{\mathbb{R}} g(\tau)C_{u_0u_0u_0}(\tau_1-\tau, \tau_2)d\tau \quad (\text{A.5})$$

$$= G(p)C_{u_0u_0u_0}(\tau_1, \tau_2) \quad (\text{A.6})$$

$$= G(p)C_{uuu}(\tau_1, \tau_2) \quad (\text{A.7})$$

which completes the proof.

Note that since the integration in (A.5) is made with respect to the first cumulant argument, the operator p stands for $\frac{\partial}{\partial \tau_1}$ in the compact notation of (A.6).

References

- Agüero, J.C. and G.C. Goodwin (2006). Identifiability of errors-in-variables dynamic systems. In: *14th IFAC Symposium on System Identification*. Newcastle, Australia. pp. 196–201.
- Anderson, B.D.O. and M. Deistler (1984). Identifiability in dynamic errors-in-variables models. *Journal of Time Series Analysis* **5**(1), 1–13.
- Anderson, J.M.M. and G.B. Giannakis (1996). Noisy input/output system identification using cumulants and the Steiglitz-McBride algorithm. *IEEE Transactions on Signal Processing* **44**(4), 1021–1024.
- Brillinger, D.R. (1981). *Time Series, Data Analysis and Theory*. Holden Day.
- Delopoulos, A. and G.B. Giannakis (1994). Consistent identification of stochastic linear systems with noisy input-output data. *Automatica* **30**(8), 1271–1294.
- Forssell, U. and L. Ljung (1999). Closed-loop identification revisited. *Automatica* **35**(7), 1215–1241.
- Friedlander, B. and B. Porat (1990). Asymptotically optimal estimation of MA and ARMA parameters of non-Gaussian processes from high-order moments. *IEEE Transactions on Automatic Control* **35**(1), 27–35.
- Garnier, H., M. Gilson and O. Cervellin (2006). Latest developments for the Matlab CONTSID toolbox. In: *14th IFAC Symposium on System Identification*. Newcastle, Australia. pp. 714–719.
- Garnier, H., M. Gilson, T. Bastogne and M. Mensler (2007). *Identification of continuous-time models from sampled data*. Chap. The CONTSID toolbox for Matlab: a software support for data-based continuous-time modeling. H. Garnier and L. Wang ed.. Springer-Verlag. To appear.
- Garnier, H., M. Mensler and A. Richard (2003). Continuous-time model identification from sampled data. Implementation issues and performance evaluation. *International Journal of Control* **76**(13), 1337–1357.
- Gevers, M. (1993). *Essays on Control: perspectives in the theory and its applications*. Chap. Towards a joint design of identification and control?, pp. 111–152. H.L. Trentelman and J.C. Willems ed.. Birkhauser Boston.
- Gilson, M. and P. Van den Hof (2005). Instrumental variable methods for closed-loop system identification. *Automatica* **41**(2), 241–249.
- Lacoume, J.-L., P.-O. Amblard and P. Comon (1997). *Statistiques d'ordre supérieur pour le traitement du signal*. Masson.
- Larsson, E.K., M. Mossberg and T. Söderström (2006). An overview of important practical aspects of continuous-time ARMA system identification. *Circuits, Systems, and Signal Processing* **25**(1), 17–46.
- Ljung, L. (1999). *System Identification: Theory for the User*. 2nd ed.. Prentice Hall.
- Mahata, K. and H. Garnier (2006). Identification of continuous-time errors-in-variables models. *Automatica* **49**(9), 1470–1490.
- Mendel, J.M. (1991). Tutorial on high-order statistics (spectra) in signal processing and system theory: theoretical results and some applications. *Proceedings of the IEEE* **79**(3), 278–305.
- Moussaoui, S., D. Brie and A. Richard (2005). Regularization aspects in continuous-time model identification. *Automatica* **41**(2), 197–208.
- Nikias, C.L. and A.P. Petropulu (1993). *Higher-order spectra analysis*. Signal Processing Series. Prentice Hall. Englewood Cliffs.
- Porat, B. and B. Friedlander (1989). Performance analysis of parameter estimation algorithms based on high-order moments. *International Journal of Adaptive Control and Signal Processing* **3**, 191–229.
- Rao, G.P. and H. Unbehauen (2006). Identification of continuous-time systems. *IEEE Proceedings - Control Theory and Applications* **153**(2), 185–220.
- Söderström, T. (2006). Errors-in-variables methods in system identification. In: *14th IFAC Symposium on System Identification*. Newcastle, Australia. pp. 1–19. Plenary Session.
- Söderström, T. and P. Stoica (1989). *System identification*. Series in Systems and Control Engineering. Prentice Hall.
- Söderström, T., E. Larsson, K. Mahata and M. Mossberg (2006). Using continuous-time modeling for errors-in-variables identification. In: *14th IFAC Symposium on System Identification*. Newcastle, Australia. pp. 428–433.
- Steiglitz, K. and L.E. McBride (1965). A technique for the identification of linear systems. *IEEE Transactions on Automatic Control* **43**(10), 2386–2397.
- Stoica, P. and T. Söderström (1981). The Steiglitz-McBride identification algorithm revisited – convergence analysis and accuracy aspects. *IEEE Transactions on Automatic Control* **26**(3), 712–717.
- Tugnait, J.K. (1992). Stochastic system identification with noisy input using cumulant statistics. *IEEE Transactions on Automatic Control* **37**(4), 476–485.
- Tugnait, J.K. and Y. Ye (1995). Stochastic system identification with noisy input-output measurements using polyspectra. *IEEE Transactions on Automatic Control* **40**(4), 670–683.
- Van den Hof, P. (1998). Closed-loop issues in system identification. *Annual Reviews in Control* **22**, 173–186.
- Van Huffel, S. (1997). *Recent advances in total least squares techniques and errors-in-variables modeling*. Society for Industrial and Applied Mathematics.
- Van Huffel, S. (2002). *Total least squares and errors-in-variables modeling*. Kluwer Academic Publishers.
- Walter, E. and L. Pronzato (1997). *Identification of Parametric Models from Experimental Data*. Springer.
- Young, P.C. (1964). In flight dynamic checkout - a discussion. *IEEE Transactions on Aerospace* **2**, 1106–1111.
- Young, P.C. (1970). An instrumental variable method for real-time identification of a noisy process. *Automatica* **6**(2), 271–287.
- Young, P.C. (1981). Parameter estimation for continuous-time models - a survey. *Automatica* **17**(1), 23–39.
- Young, P.C. and A.J. Jakeman (1980). Refined instrumental variable methods of time-series analysis: Part III, extensions. *International Journal of Control* **31**, 741–764.
- Young, P.C., H. Garnier and M. Gilson (2007). *Identification of continuous-time models from sampled data*. Chap. Refined instrumental variable for identifying continuous-time hybrid Box-Jenkins models. H. Garnier and L. Wang ed.. Springer-Verlag. To appear.