Proceedings of the 45th IEEE Conference on Decision & Control
Manchester Grand Hyatt Hotel
San Diego, CA, USA, December 13-15, 2006

The Module Structure of ARMAX Systems

Ivan Markovsky, Jan C. Willems, and Bart De Moor

***Abstract*— We consider ARMAX system representations and identification problems. Identifiability conditions in terms of the correlation function of the process are given. One of the conditions is persistency of excitation of an input component of the process and another one is a rank condition for a pair of Hankel matrices.**

We study the linear combinations of the process and its shifts that produce a process independent of the input. The set of all such linear combinations, called the orthogonalizers, has a module structure and under identifiability conditions completely specifies the deterministic part of the ARMAX system. Computing a module basis for the orthogonalizers is a deterministic identification problem.

We propose an ARMAX identification algorithm, which has three steps: first compute the deterministic part of the system via the orthogonalizers, then the AR part, which also has a module structure, and finally the MA part.

***Index Terms*— System identification, ARMAX representation, orthogonalizers, annihilators, module structure.**

I. Introduction

A. Stochastic processes and correlation functions

Consider a vector w of $\mathtt{w}$ real valued, jointly Gaussian, zero mean, stationary, ergodic stochastic processes on $\mathbb{Z}$. In what follows, *process* means a vector stochastic processes that satisfies the above conditions. The correlation function of w

$$R_{ww} := \big(\ldots, R_{ww}(0), \ldots, R_{ww}(t), \ldots\big)$$

is defined as

$$R_{ww}(t) := \mathbf{E}\big(w(t)w^\top(0)\big) \in \mathbb{R}^{\mathtt{w}\times\mathtt{w}},$$

where $\mathbf{E}$ is the expectation operator. To avoid unessential complications, we assume that $R_{ww}(0) > 0$.

A realization of w is denoted by w_d ("d" stands for "data"). Due to the ergodicity assumption, R_{ww} can be expressed in terms of an infinite realization w_d as

$$R_{ww}(t) = \lim_{T\to\infty} \frac{1}{T} \sum_{\tau=1}^{T} w_\mathrm{d}(t+\tau) w_\mathrm{d}^\top(\tau).$$

If the realization w_d is over a finite interval

$$w_\mathrm{d} = \big(w_\mathrm{d}(1), \ldots, w_\mathrm{d}(T)\big),$$

we can only compute a finite sample estimate of R_{ww}, e.g., the standard biased estimator

$$\hat{R}_{ww}(t) := \frac{1}{T} \sum_{\tau=1}^{T-t} w_\mathrm{d}(t+\tau) w_\mathrm{d}^\top(\tau),$$
$$\hat{R}_{ww}(-t) = \hat{R}_{ww}^\top(t), \quad \text{for } t = 0, \ldots, T-1. \quad (1)$$

Alternative methods for computing $\hat{R}_{ww}$ from w_d are described in [1, Chapter 2] in the context of nonparameteric spectral estimation.

A process ε is called *white* if all the $\varepsilon(t)$'s are independent and normalized if $\mathbf{E}\big(\varepsilon(0)\varepsilon^\top(0)\big) = I$. Independence of two random vectors or processes is denoted by $\perp$.

B. ARMAX systems

We use a notion of a stochastic system, which is analogous to the behavioral definition of a deterministic dynamical system as a collection of trajectories (see, e.g., [2]). Roughly speaking, a stochastic system $\mathscr{B}$ is a collection of processes w. (In this paper, a process means real valued, jointly Gaussian, zero mean, stationary, ergodic stochastic processes on $\mathbb{Z}$.)

Denote by σ the shift operator

$$\sigma(f)(t) = f(t+1)$$

and consider the difference equation

$$Y(\sigma)y + U(\sigma)u = E(\sigma)\varepsilon, \quad \text{(ARMAX)}$$

with Y, U, and E real polynomial matrices; and Y square and nonsingular. The stochastic system consisting of all processes

$$w := \mathrm{col}(u,y), \quad \dim(u) = \mathtt{m}, \quad \dim(y) = \mathtt{p},$$
$$\mathtt{m} + \mathtt{p} = \mathtt{w} := \dim(w), \quad (2)$$

satisfying (ARMAX) with ε white normalized process that is independent of u is called an ARMAX (autoregressive moving average exogenous) system with parameters (Y,U,E). Under generic conditions u is free, which means that for all u there are y and $\varepsilon \perp u$ that satisfy (ARMAX). Therefore, u is an exogenous input and y is an endogenous output. The triples (Y,U,E) and (Y',U',E') are equivalent if they define the same system.

The polynomial matrix R is left prime if any factorization $R = FR'$ with F square implies that F is unimodular, i.e., $\det(F)$ is a nonzero constant. Equivalently, R is left prime if $R(\lambda)$ is full row rank for all $\lambda \in \mathbb{C}$. Each polynomial matrix R of full row rank can be factored as $R = FR'$ with F square and R' left prime. Using this factorization we can refine the representation (ARMAX) as follows:

$$A(\sigma)\big(P(\sigma)y + Q(\sigma)u\big) = M(\sigma)\varepsilon, \quad \text{(ARMAX')}$$

where A and P are square and nonsingular, A is Schur, and $R := \begin{bmatrix} P & Q \end{bmatrix}$ is left prime. The polynomial matrix

- A represents the AR (autoregressive)-part,
- M the MA (moving average)-part, and
- $R = \begin{bmatrix} P & Q \end{bmatrix}$ the X (exogenous)-part.

A full version of this paper, with included proofs, is in preparation.

All authors are with the Electrical Engineering Department, K.U.Leuven, Kasteelpark Arenberg 10, B-3001 Leuven, Belgium.

The rational function $G := P^{-1}Q$ is called the transfer function of the deterministic part of the ARMAX system. Note, however, that we do not require G to be proper, i.e., we do not impose causality.

C. Representability, identifiability, and estimation

Consider a model class $\mathscr{M}$, e.g., all ARMAX systems. Given an observed *infinite* realization w_{d} of a process w (or equivalently given its correlation function R_{ww}), the following questions arise.

- *Representability:* When is there a system $\mathscr{B} \in \mathscr{M}$, such that $w \in \mathscr{B}$?
- *Identifiability:* When is a representation $\mathscr{B}$ of w unique in $\mathscr{M}$?
- *Identification:* If w is identifiable, obtain a representation of $\mathscr{B}$ from w_{d}.

In a more practical situation when the given realization

$$w_{\mathrm{d}} = \big(w_{\mathrm{d}}(1), \ldots, w_{\mathrm{d}}(T)\big)$$

of w is *finite*, the *exact* stochastic identification problem, stated above, becomes an *approximate* stochastic identification problems. Assuming that an ARMAX process w is identifiable, the following estimation problem and related question are of interest.

- *Estimation:* Assuming that w is identifiable, find a representation of an estimate $\hat{\mathscr{B}}$ of $\mathscr{B}$ from w_{d}.
- *Consistency:* Study the asymptotic behavior of the estimate $\hat{\mathscr{B}}$, as $T \to \infty$.

In this paper, we approach the ARMAX identification and estimation problems by first computing the X-part of the ARMAX system, then the AR-part, and finally the MA-part, i.e., we consider algorithms that decompose the original

$$w_{\mathrm{d}} \mapsto (P, Q, A, M)$$

problem into three sequential subproblems as follows:

$$w_{\mathrm{d}} = (u_{\mathrm{d}}, y_{\mathrm{d}}) \mapsto (P, Q) \mapsto A \mapsto M.$$

First we consider the case when w_{d} is infinite and then propose a modification for the more difficult finite time case.

The basic idea, developed in the paper, is that the finite linear combinations of the rows of the Hankel matrix composed of w_{d}

$$W := \begin{bmatrix} w_{\mathrm{d}}(1) & w_{\mathrm{d}}(2) & w_{\mathrm{d}}(3) & \cdots & w_{\mathrm{d}}(t) & \cdots \\ w_{\mathrm{d}}(2) & w_{\mathrm{d}}(3) & w_{\mathrm{d}}(4) & \cdots & w_{\mathrm{d}}(t+1) & \cdots \\ w_{\mathrm{d}}(3) & w_{\mathrm{d}}(4) & w_{\mathrm{d}}(5) & \cdots & w_{\mathrm{d}}(t+2) & \cdots \\ \vdots & \vdots & \vdots & & \vdots & \end{bmatrix} \tag{3}$$

that are orthogonal to the rows of the Hankel matrix composed of inputs u_{d}

$$U := \begin{bmatrix} u_{\mathrm{d}}(1) & u_{\mathrm{d}}(2) & u_{\mathrm{d}}(3) & \cdots & u_{\mathrm{d}}(t) & \cdots \\ u_{\mathrm{d}}(2) & u_{\mathrm{d}}(3) & u_{\mathrm{d}}(4) & \cdots & u_{\mathrm{d}}(t+1) & \cdots \\ u_{\mathrm{d}}(3) & u_{\mathrm{d}}(4) & u_{\mathrm{d}}(5) & \cdots & u_{\mathrm{d}}(t+2) & \cdots \\ \vdots & \vdots & \vdots & & \vdots & \end{bmatrix} \tag{4}$$

determine $R = \begin{bmatrix} P & Q \end{bmatrix}$. However, there are infinite number of such "orthogonalizing" linear combinations. The questions occur: What structure do they have in order to be generated by a finite number of them: the rows of $R = \begin{bmatrix} P & Q \end{bmatrix}$? How we can limit the number of rows? What algorithms can be used for actually computing an estimate of R from a finite realization w_{d}?

II. Conditions for representability and identifiability

Given a process $w = (u, y)$ with a correlation function R_{ww}, we want to determine under what conditions w can be represented as a solution of a stochastic difference equation (ARMAX') with ε a white normalized process that is independent of u and A Schur, i.e., under what conditions $w \in \mathscr{B}$, where $\mathscr{B}$ is an ARMAX system. This is the ARMAX representability problem. The related problem: Given a process $w = (u, y)$ with a correlation function R_{ww}, determine under what conditions (and in what sense) an ARMAX representation $\mathscr{B}$ of w is unique is the ARMAX identifiability problem.

Define the partitioning of R_{ww} conformable with (2)

$$R_{ww} = \begin{bmatrix} R_{uu} & R_{uy} \\ R_{yu} & R_{yy} \end{bmatrix} \begin{matrix} \mathtt{m} \\ \mathtt{p} \end{matrix}, \qquad \begin{matrix} \mathtt{m} & \mathtt{p} \end{matrix}.$$

A key condition for ARMAX representability turns out to be the notion of a rank increment of a Hankel matrix composed of R_{uu} with respect to a Hankel matrix composed of R_{yu}. The rank increment of A with respect to B, where A and B have the same number of columns, is

$$\operatorname{rank}\operatorname{inc}(A, B) := \operatorname{rank}\big(\operatorname{col}(A, B)\big) - \operatorname{rank}(B).$$

The notion of rank increment is well defined for (two sided, see (3) and (4)) infinite matrices as well. Let $A_{k,q}$ denotes the submatrix of A formed of the first k block rows and the first q block columns. The rank increment of the (two sided) infinite matrix A with respect to the is (two sided) infinite matrix B is

$$\lim_{k,q\to\infty} \Big(\operatorname{rank}\big(\operatorname{col}(A_{kq}, B_{kq})\big) - \operatorname{rank}(A_{kq}) \Big).$$

A process $w = (u, y)$ with a correlation function R_{ww} is ARMAX representable if and only if

$$\operatorname{rank}\operatorname{inc}\left(\begin{bmatrix} R_{uu}(1) & R_{uu}(2) & \cdots \\ R_{uu}(2) & R_{uu}(3) & \cdots \\ \vdots & \vdots & \end{bmatrix}, \begin{bmatrix} R_{yu}(1) & R_{yu}(2) & \cdots \\ R_{yu}(2) & R_{yu}(3) & \cdots \\ \vdots & \vdots & \end{bmatrix} \right)$$

is finite.

For identifiability we need, in addition, the assumption that the input u is persistently exciting of a sufficiently high order. A process u with a correlation function R_{uu} is persistently exciting of order k if the Toeplitz matrix

$$\begin{bmatrix} R_{uu}(0) & R_{uu}^{\top}(1) & \cdots & R_{uu}^{\top}(k) \\ R_{uu}(1) & R_{uu}(0) & \cdots & R_{uu}^{\top}(k-1) \\ \vdots & \vdots & & \vdots \\ R_{uu}(k) & R_{uu}(k+1) & \cdots & R_{uu}(0) \end{bmatrix}$$

is full rank.

A process $w=(u,y)$ with a correlation function R_{ww} is ARMAX identifiable if

$$\operatorname{rank}\operatorname{inc}\left(\begin{bmatrix} R_{uu}(1) & R_{uu}(2) & \cdots \\ R_{uu}(2) & R_{uu}(3) & \cdots \\ \vdots & \vdots & \end{bmatrix}, \begin{bmatrix} R_{yu}(1) & R_{yu}(2) & \cdots \\ R_{yu}(2) & R_{yu}(3) & \cdots \\ \vdots & \vdots & \end{bmatrix}\right)$$

is finite and if, in addition, u is persistently exciting of any order.

Note 1. If upper bounds `n` and `l` for the order and the lag of the deterministic part of the ARMAX system are a priori given, the persistency of excitation assumption can be relaxed to

"u is persistently exciting of order `n+l`",

as in the deterministic case, see [3]. However, in this paper "persistently exciting" means persistently exciting of any order.

III. The module of orthogonalizers

Let $\mathbb{R}[\xi]^n$ denotes the set of all n-dimensional vector polynomials with real coefficients in the indeterminate ξ. The set $\mathbb{R}[\xi]^n$ has the structure of a module over the ring $\mathbb{R}[\xi]$. A submodule of $\mathbb{R}[\xi]^n$ is a subset of $\mathbb{R}[\xi]^n$ that is also a module, e.g., the submodule $p_1v_1+\cdots+p_kv_k$ generated by $v_1,\ldots,v_k\in\mathbb{R}[\xi]^n$, where $v_1,\ldots,v_k\in\mathbb{R}[\xi]$. In fact, every submodule of $\mathbb{R}[\xi]^n$ is of this (finitely generated) form. The minimal number of generators is by called the dimension of the module. A submodule is called *slim* if it does not strictly contain another submodule of the same dimension. This is equivalent to the set of generators $V:=\begin{bmatrix} v_1 & \cdots & v_k\end{bmatrix}$ of the submodule being right prime.

The importance of modules in systems theory stems from the fact that

> submodules are in a one-to-one relation with LTI systems and slim submodules are in a one-to-one relation with controllable LTI systems.

With some abuse of notation, we will occasionally view a vector $v\in\mathbb{R}^{(l+1)\mathrm{w}}$ as a polynomial

$$v(\xi):=v_0+v_1\xi+\cdots+v_l\xi^l\in\mathbb{R}^{\mathrm{w}}[\xi]$$

of degree (at most) l, and vise versa. Using this convention, it is easy to see that although the left kernel of the Hankel matrix W, see (3), might be infinite dimensional over $\mathbb{R}$, actually it is a module over $\mathbb{R}[\xi]$, so is finitely generated.

Lemma 1. *The left kernel of a two sided infinite Hankel matrix has a module structure.*

Proof: Consider the matrix W in (3) and let $\mathscr{A}$ denote its left kernel. We need to show that if $u,v\in\mathscr{A}$ then

1) $u+v\in\mathscr{A}$, and
2) $\alpha u\in\mathscr{A}$, for all $\alpha\in\mathbb{R}[\xi]$.

Item 1 is trivial and item 2 follows from $\sigma^t u\in\mathscr{A}$, for all $t\in\mathbb{N}$, which is a simple consequence of the Hankel structure. □

An element of the left kernel of W is called an *annihilator* of the the Hankel matrix. Consider a process w with a partition $w=(u,y)$. An orthogonalizer of w with respect to u is a polynomial $n\in\mathbb{R}^{\mathrm{w}}[\xi]$ that makes the process $n(\sigma)w$ independent of u.

Example 1. The rows of $R=\begin{bmatrix} P & Q\end{bmatrix}$ are orthogonalizers of the ARMAX process w with parameters (P,Q,A,M). To see this, note that

$$A(\sigma)\big(P(\sigma)y+Q(\sigma)u\big)=M(\sigma)\varepsilon \implies P(\sigma)y+Q(\sigma)u=\sum_{t=-\infty}^{\infty}H(t)\sigma^t\varepsilon,$$

where the elements of H are the Markov parameters of the system $A(\sigma)a=M(\sigma)\varepsilon$, viewed as a system with input ε and output a. Now, since $\varepsilon\perp u$, this implies that

$$P(\sigma)y+Q(\sigma)u\perp u.$$

Moreover, every element of the module generated by the rows of R is an orthogonalizer of w. The basic question is: Are these the complete set of orthogonalizer?

The orthogonalizers of w with respect to u have a slim module structure. The main result of the paper is the following one: Consider an identifiable ARMAX process $w=\operatorname{col}(u,y)$ with parameters (P,Q,A,M). The module of the orthogonalizers of w with respect to u is generated by the rows of $R=\begin{bmatrix} P & Q\end{bmatrix}$.

IV. Identification algorithm: infinite-time case

Our main result, stating that the module of the orthogonalizers is generated by R, suggests an algorithm for the computation of the deterministic part of the ARMAX system, i.e., an algorithm that realizes the mapping $w_{\mathrm{d}}\mapsto R$. We need to compute the module of the orthogonalizers of w_{d}. This turns out to be a deterministic identification problem. In this section, we consider the infinite-time case. In the following section, we adapt the algorithm for the finite-time case.

A. X-part: $w_{\mathrm{d}}\mapsto(P,Q)$

We aim to find a module basis for the linear combinations of W (see (3)) that are orthogonal to U (see (4)). This question is equivalent to the question of computing a module basis for the left kernel of $WU^\top$. Observe that

$$WU^\top=\begin{bmatrix} R_{wu}(0) & R_{wu}^\top(1) & R_{wu}^\top(2) & \cdots & R_{wu}^\top(t) & \cdots \\ R_{wu}(1) & R_{wu}(0) & R_{wu}^\top(1) & \cdots & R_{wu}^\top(t-1) & \cdots \\ R_{wu}(2) & R_{wu}(1) & R_{wu}(0) & \cdots & R_{wu}^\top(t-2) & \cdots \\ \vdots & \vdots & \vdots & & \vdots & \end{bmatrix}.$$

Computing a module basis for the left kern of a Hankel or Toeplitz matrix is a deterministic identification problem, see [4, Section 8.5], so that

> computing the orthogonalizers is a deterministic identification problem for the correlation function R_{wu}.

There are many algorithms developed for solving this problem, e.g., the subspace algorithms based on state construction [5] or computation of an observability matrix [6].

How many correlation coefficients are actually needed in order to compute a set of generators of this left kernel? In other words can we limit the number of rows and columns of $WU^\top$? Suppose that upper bounds $\mathtt{n}$ and $\mathtt{l}$ are given for the order and the lag of the deterministic part of the ARMAX system. Using the result of [3], we have that if u is persistently exciting of order $\mathtt{n}+\mathtt{l}$, then the left kernel of the Hankel matrix

$$\begin{bmatrix} R_{wu}(-\mathtt{l}-\mathtt{n}) & \cdots & R_{wu}(0) & \cdots & R_{wu}(\mathtt{n}) \\ R_{wu}(-\mathtt{l}-\mathtt{n}+1) & \cdots & R_{wu}(1) & \cdots & R_{wu}(\mathtt{n}+1) \\ \vdots & & \vdots & & \vdots \\ R_{wu}(\mathtt{n}) & \cdots & R_{wu}(\mathtt{l}) & \cdots & R_{wu}(\mathtt{l}+\mathtt{n}) \end{bmatrix} \tag{5}$$

determines all orthogonalizing polynomials of degree at most $\mathtt{l}$ and is therefore equal to the module generated by R. Hence it determines R uniquely. Exploiting symmetry, this means that $T = \mathtt{n}+\mathtt{l}+1$ correlation coefficients are sufficient.

B. AR-part: $\big(w_{\mathrm{d}}, (P,Q)\big) \mapsto A$

Once R is determined, we consider the ARMA identification problem

$$A(\sigma)a = M(\sigma)\varepsilon, \quad \text{where} \quad a := R(\sigma)w,$$

with a realization $a_{\mathrm{d}} := R(\sigma)w_{\mathrm{d}}$. Let R_{aa} be the correlation function of a. Then the process $a = R(\sigma)w$ is an ARMA process and the set of annihilators of R_{aa} is the module generated by the rows of A.

Therefore, we can determine the AR-part of the ARMAX system by computing a module basis of the left kernel of the two sided infinite Hankel matrix composed of R_{aa}. As in the previous subsection, however, knowing an upper bound $\mathtt{l}$ for the degree of A, we can consider a finite Hankel matrix

$$\begin{bmatrix} R_{aa}(1) & R_{aa}(2) & \cdots & R_{aa}(\mathtt{l}+1) \\ R_{aa}(2) & R_{aa}(3) & \cdots & R_{aa}(\mathtt{l}+2) \\ \vdots & \vdots & & \vdots \\ R_{aa}(\mathtt{l}+1) & R_{aa}(\mathtt{l}+2) & \cdots & R_{aa}(2\mathtt{l}+1) \end{bmatrix} \tag{6}$$

which is again a deterministic identification problem and can be solved by standard algorithms.

C. MA-part: $(w_{\mathrm{d}}, P, Q, A) \mapsto M$

Once A is determined, we consider the MA identification problem

$$m = M(\sigma)\varepsilon, \quad \text{where} \quad m := A(\sigma)a,$$

with a realization $m_{\mathrm{d}} := A(\sigma)a_{\mathrm{d}}$. The process m is an MA process, so that standard MA identification methods can be employed for computing M. We use the efficient MA identification method proposed in [1, Section 3.6].

D. Summary of the algorithm

If w_{d} is a realization of an identifiable ARMAX process $w \in \mathscr{B}$, where the deterministic and autoregressive parts of $\mathscr{B}$ have lags less than or equal to $\mathtt{l}$, the system $\hat{\mathscr{B}}$ computed by Algorithm 1 with inputs w_{d} and $\mathtt{l}$ is equal to the true data generating system $\mathscr{B}$, i.e., $\hat{\mathscr{B}} = \mathscr{B}$.

Algorithm 1 ARMAX identification: infinite-time case.

Input: Time series $w_{\mathrm{d}} = (u_{\mathrm{d}}, y_{\mathrm{d}})$ and upper bound $\mathtt{l}$ for the degrees of the X and AR parts.

1: Compute the first $\mathtt{l}+\mathtt{n}+1$ correlation coefficients of w_{d}.
2: Compute a module basis $\hat{R}$ for the left kernel of (5).
3: Let $a_{\mathrm{d}} := \hat{R}(\sigma)w_{\mathrm{d}}$.
4: Compute the first $2\mathtt{l}+1$ correlation coefficients of a_{d}.
5: Compute a module basis $\hat{A}$ for the left kernel of (6).
6: Let $m_{\mathrm{d}} := \hat{A}(\sigma)a_{\mathrm{d}}$.
7: Compute the parameter $\hat{M}$ of an MA system for m_{d}.

Output: ARMAX system $\hat{\mathscr{B}}$ defined by the parameters $(\hat{P}, \hat{Q}, \hat{A}, \hat{M})$, where $\begin{bmatrix} \hat{P} & \hat{Q} \end{bmatrix} := \hat{R}$.

V. Identification algorithm: finite-time case

In the infinite time case, under the identifiability assumption, we can recover the true data generating system exactly. Of course, this is no longer possible in the finite-time case. The question of main interest, considered in the literature, is the *consistency* property of the estimate produced by an algorithm: Does the finite-time estimate converge to the true system as the time horizon goes to infinity? For a fixed time horizon, however, the identification problem necessarily involves approximation. Estimators that achieve statistically optimal approximation are called *efficient*. Another point of view of finite-time ARMAX identification problem is the bias–variance decomposition of the approximation error.

In this section, we are not aiming at an optimal finite-time approximation, i.e., the proposed algorithm will not be efficient. We are looking instead at a heuristic adaptation of the exact (in the infinite-time case) Algorithm 1 for the finite-time case. This is similar to the application of exact deterministic subspace algorithms for approximate deterministic and/or stochastic identification problems.

A straightforward finite-time version of Algorithm 1 is obtained by replacing the computation of

- the true correlation coefficients on steps 1 and 4 by the standard biased estimates,
- a module basis of an exact left kernel on steps 2 and 5 by module basis of an approximate left kernel, computed, e.g., by the SVD, and
- the exact MA model on step 7 by an approximate MA model, computed, e.g., by the polynomial time algorithms of [7] or [8].

The quality of the correlation estimates, however, affects the accuracy of the parameter estimates and the question:

Which estimates $\hat{R}_{ww}$, $\hat{R}_{aa}$ yield optimal efficiency?

is non trivial.

The matrices W and U defined in (3) and (4) are redefined for a finite realization w_{d} as

$$W := \begin{bmatrix} w_{\mathrm{d}}(\mathtt{n}+1) & w_{\mathrm{d}}(\mathtt{n}+2) & \cdots & w_{\mathrm{d}}(T-\mathtt{l}-\mathtt{n}) \\ w_{\mathrm{d}}(\mathtt{n}+2) & w_{\mathrm{d}}(\mathtt{n}+3) & \cdots & w_{\mathrm{d}}(T-\mathtt{l}-\mathtt{n}+1) \\ \vdots & \vdots & & \vdots \\ w_{\mathrm{d}}(\mathtt{n}+\mathtt{l}+1) & w_{\mathrm{d}}(\mathtt{n}+\mathtt{l}+2) & \cdots & w_{\mathrm{d}}(T-\mathtt{l}+1) \end{bmatrix}$$

and

$$U := \begin{bmatrix} u_{\mathrm{d}}(1) & u_{\mathrm{d}}(2) & \cdots & u_{\mathrm{d}}(T-\mathtt{l}-2\mathtt{n}) \\ u_{\mathrm{d}}(2) & u_{\mathrm{d}}(3) & \cdots & \\ \vdots & \vdots & & \vdots \\ u_{\mathrm{d}}(\mathtt{l}+2\mathtt{n}+1) & u_{\mathrm{d}}(\mathtt{l}+2\mathtt{n}+2) & \cdots & u_{\mathrm{d}}(T) \end{bmatrix},$$

Our experience is that on step 2 it is better to compute the left kernel of the matrix $WU^\top$ instead of the Hankel matrix composed of the standard biased correlation estimates $\hat{R}_{wu}$. A possible explanation for the superior results obtained from $WU^\top$ is that the standard biased estimator (1) implicitly extends w_{d} with zeros (rendering w_{d} infinite), which changes the data. For small sample size, the change of the data due to the extension gives inferior estimates compared to the computation from $WU^\top$, where the data is not extended. For example, in deterministic identification problem, i.e., $\mathrm{var}(\varepsilon) = 0$, the left kernel of $WU^\top$ gives exact result, while the approach using (1) gives biased result. The same observation is made in [1, pages 98–99], where it is supported by a statistical argument (bias vs variance tradeoff).

Similarly on step 5, we replace the Hankel matrix formed from the standard biased correlation estimates $\hat{R}_{aa}$ (see (6)) by the matrix obtained from the product

$$\begin{bmatrix} a_{\mathrm{d}}(\mathtt{l}+1) & a_{\mathrm{d}}(\mathtt{l}+2) & \cdots & a_{\mathrm{d}}(T-\mathtt{l}-1) \\ \vdots & \vdots & & \vdots \\ a_{\mathrm{d}}(2) & a_{\mathrm{d}}(3) & \cdots & a_{\mathrm{d}}(T-2\mathtt{l}+2) \\ a_{\mathrm{d}}(1) & a_{\mathrm{d}}(2) & \cdots & a_{\mathrm{d}}(T-2\mathtt{l}+1) \end{bmatrix} \begin{bmatrix} a_{\mathrm{d}}(\mathtt{l}+2) & a_{\mathrm{d}}(\mathtt{l}+3) & \cdots & a_{\mathrm{d}}(2\mathtt{l}+2) \\ a_{\mathrm{d}}(\mathtt{l}+3) & a_{\mathrm{d}}(\mathtt{l}+4) & \cdots & a_{\mathrm{d}}(2\mathtt{l}+3) \\ \vdots & \vdots & & \vdots \\ a_{\mathrm{d}}(T-\mathtt{l}) & a_{\mathrm{d}}(T-\mathtt{l}+1) & \cdots & a_{\mathrm{d}}(T) \end{bmatrix}, \quad (7)$$

which corresponds for estimation of R_{aa} without extending the data with zeros.

Algorithm 2 ARMAX identification: finite-time case.

Input: Time series $w_{\mathrm{d}} = (u_{\mathrm{d}}, y_{\mathrm{d}})$, upper bound $\mathtt{l}$ for the degrees of the X and AR parts.

1: Compute a module basis $\hat{R}$ for an approximate left kernel of $WU^\top$, using the SVD.
2: Let $a_{\mathrm{d}} := \hat{R}(\sigma) w_{\mathrm{d}}$.
3: Compute a module basis $\hat{A}$ for the approximate left kernel of (7), using the SVD.
4: Let $m_{\mathrm{d}} := \hat{A}(\sigma) a_{\mathrm{d}}$.
5: Compute the parameter $\hat{M}$ of an approximate MA system for m_{d}, using the method of [7].

Output: ARMAX system $\hat{\mathscr{B}}$ defined by the parameters $(\hat{P}, \hat{Q}, \hat{A}, \hat{M})$, where $\begin{bmatrix} \hat{P} & \hat{Q} \end{bmatrix} := \hat{R}$.

If $w_{\mathrm{d}} \in (\mathbb{R}^{\mathtt{w}})^T$ is a realization of an identifiable ARMAX process $w \in \mathscr{B}$, where the deterministic and autoregressive parts of $\mathscr{B}$ have lags less than or equal to $\mathtt{l}$, Algorithm 2 with inputs w_{d} and $\mathtt{l}$ yields a consistent estimator for the true data generating system $\mathscr{B}$.

VI. Simulation example

The simulation example, presented in this section, is only meant to illustrate the proposed identification algorithm. A comparison with alternative algorithms for ARMAX identification is in order and will be reported elsewhere.

The polynomials P, Q, A, and M are selected as follows:

$$P(\xi) = 1 - 0.8713\xi - 1.539\xi^2 + 1.371\xi^3 + 0.6451\xi^4 - 0.5827\xi^5,$$

$$Q(\xi) = 1 - 1.2\xi + 0.6\xi^2 + 0.7\xi^3,$$

$$A(\xi) = 1 + \xi + 0.5\xi^2, \quad \text{and} \quad M(\xi) = 1 + 0.5\xi.$$

The inputs u and e are white, normally distributed processes with zero-mean and variances 1 and 0.2, respectively. The initial condition, under which y is obtained from u and e is a random vector. The time horizon for the simulation is $T = 1000$ and the whole simulated time series $w_{\mathrm{d}} = (u_{\mathrm{d}}, y_{\mathrm{d}})$ is used for estimation.

We apply Algorithm 2 with $\mathtt{l}$ equal to the true system lag. The experiment is repeated $N = 5$ times with different realizations of u and e in each run. Let $\hat{P}^{(k)}$ and $\hat{Q}^{(k)}$ be the estimates of the polynomials P and Q, respectively, obtained on the kth repetition of the experiment.

The poles and zeros of the deterministic part of the true and identified ARMAX systems are shown in the left plot of Figure 1. The spread of the estimates around the true values indicates the bias and variance of the estimates. Similarly, the right plot in Figure 1 shows the zeros of the AR part of the true and estimated systems.

Figure 2, left shows the bode plots of deterministic parts of the true and estimated ARMAX systems. Figure 2, right shows the bode plots of the transfer functions M/P and $\hat{M}^{(1)}/\hat{P}^{(1)}$.

VII. Conclusions

We gave an ARMAX representability condition for a stochastic process: the rank increment of a pair of infinite Hankel matrices, composed of the process's correlation function, should be finite. As in the deterministic case, a key additional condition for identifiability is the persistency of excitation of an input component of the process.

We introduced a notion called orthogonalizer. An orthogonalizer of an input/output partitioned process is a linear combination of the process and its shifts that produces a process orthogonal to the input. The set of orthogonalizers has a module structure, i.e., it is finitely generated. Under the identifiability conditions, a module basis for the set of orthogonalizers is given by the rows of the polynomial matrix R, determining the deterministic part of the ARMAX system. This result suggests an algorithm for computing R from data.

The proposed ARMAX identification algorithm is based on the decomposition of the problem into three sequential subproblems: compute first the X-part, then the AR-part, and finally the MA-part. The deterministic part of the system is computed, using the orthogonalizers. This turns out to

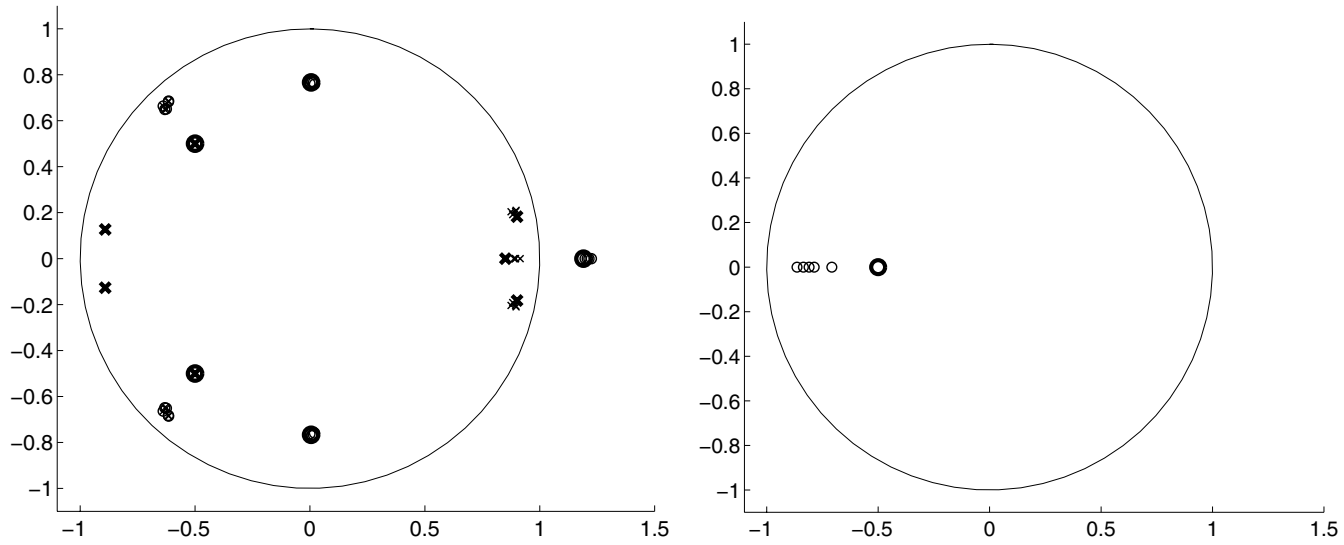

Fig. 1. Left: roots of P (solid crosses), Q (solid circles) $\hat{P}^{(k)}$ (tiny crosses), and $\hat{Q}^{(k)}$ (tiny circles), for $k = 1, \ldots, 5$. Right: roots of M (solid circles) and $\hat{M}^{(k)}$ (tiny circles), for $k = 1, \ldots, 5$.

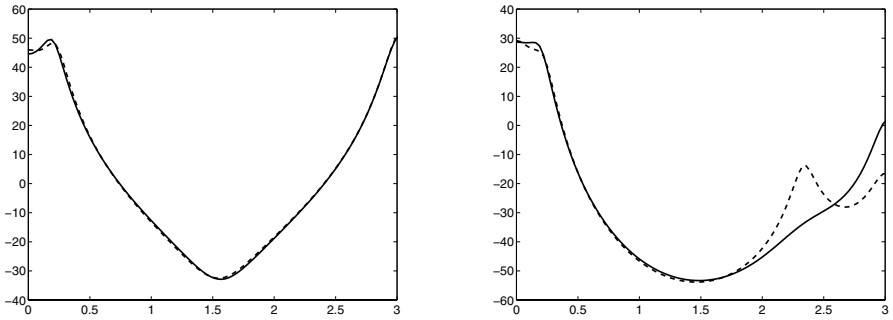

Fig. 2. Left: Bode plots of Q/P (solid line) and $\hat{Q}^{(1)}/\hat{P}^{(1)}$ (dashed line). Right: Bode plots of M/P (solid line) and $\hat{M}^{(1)}/\hat{P}^{(1)}$ (dashed line).

be a classic deterministic identification problem for the correlation function R_{wu}: compute a module basis for the left annihilators of the Hankel matrix composed of R_{wu}. Computing the deterministic part, reduces the ARMAX identification problem to an ARMA identification problem. The second subproblem—find the AR-part—is again a deterministic identification problem. Computing the AR-part, further reduces the ARMAX identification problem to an MA identification problem.

We presented a heuristic modification of the infinite sample size ARMAX identification algorithm for the finite sample size case, where an approximation is needed. Each of the three steps of the modified algorithm computes an approximate result. The first and the second step use an SVD approximation of the left kernel of an appropriate Hankel matrix and the third step is based on a convex optimization method for MA system identification.

VIII. Acknowledgments

This research is supported by the Belgian Federal Government under the DWTC program Interuniversity Attraction Poles, Phase V, 2002–2006, Dynamical Systems and Control: Computation, Identification and Modelling, by the KUL Concerted Research Action (GOA) MEFISTO–666, and by several grants en projects from IWT-Flanders and the Flemish Fund for Scientific Research; HPC-EUROPA (RII3-CT-2003-506079), with the support of the European Community — Research Infrastructure Action under the FP6 "Structuring the European Research Area" Program.

References

[1] P. Stoica and R. Moses, *Spectral Analysis of Signals*. Prentice Hall, 2005.

[2] J. C. Willems, "Paradigms and puzzles in the theory of dynamical systems," *IEEE Trans. Automat. Control*, vol. 36, pp. 259–294, 1991.

[3] J. C. Willems, P. Rapisarda, I. Markovsky, and B. D. Moor, "A note on persistency of excitation," *Control Lett.*, vol. 54, pp. 325–329, 2005.

[4] I. Markovsky, J. C. Willems, B. De Moor, and S. Van Huffel, *Exact and Approximate Modeling of Linear Systems: A Behavioral Approach*, Monographs on Math. Modeling and Comp., no. 11, SIAM, 2006.

[5] P. Van Overschee and B. De Moor, *Subspace Identification for Linear Systems: Theory, Implementation, Applications*. Boston: Kluwer, 1996.

[6] M. Verhaegen and P. Dewilde, "Subspace model identification, Part 1: The output-error state-space model identification class of algorithms," *Int. J. Control*, vol. 56, pp. 1187–1210, 1992.

[7] P. Stoica, T. McKelvey, and J. Mari, "MA estimation in polynomial time," *IEEE Trans. Signal Process.*, vol. 48, pp. 1999–2012, 2000.

[8] B. Alkire and L. Vandenberghe, "Convex optimization problems involving finite autocorrelation sequences," *Mathematical Programming, Series A*, vol. 93, no. 3, pp. 331–359, 2002.