Impulse response identification from input/output binary measurements

Mathieu Pouliquen, Eric Pigeon, Olivier Gehan, Abdelhak Goudjil, Romain Auber

Impulse Response Identification from Input/Output Binary Measurements ⋆

Mathieu Pouliquen [a], Eric Pigeon [a], Olivier Gehan [a], Abdelhak Goudjil [a], Romain Auber [a,b]

[a] Normandie Univ, UNICAEN, ENSICAEN, LAC, 14000 Caen, France.

[b] Bodycap,3 rue du Dr Laennec, 14000 Caen, France.

Abstract

In this paper, we study the identification of Finite Impulse Response systems in a particular context: data on the input and the output are obtained with one-bit quantizers, the thresholds of quantizers can be different from zero. A three-step identification algorithm is proposed from these binary-valued measurements. This algorithm is based on the normal distribution of the input and noises. The algorithm is appropriately analyzed: it is shown to be asymptotically unbiased, its asymptotic variance is also expressed. Numerical simulations are provided to demonstrate the effectiveness of the proposed algorithm even in presence of noise and to validate the analysis.

Key words: Identification of FIR, input and output binary-valued measurements

1 Introduction

1.1 The considered identification problem and prior works

In this paper we are interested in the identification of dynamical systems using binary measurements. This context is justified by the fact that sometimes it is difficult to obtain high resolution data. This is the case when no high resolution sensor is available (does not exist or too expensive) or when it is not possible to transmit high resolution data (limited bandwidth for instance) or when the use of binary data allows to preserve memory and battery capacities (on a small wireless connected device for instance) or when we want to analyze categorical data (detected/not-detected for instance). We consider here the extreme case where we use one-bit quantization both on the input and the output. Such a situation occurs when we do not want or can not interfere with the system. The only available information is the fact that samples are lower or higher than a threshold of quantization. Note that this threshold can be different from zero.

The identification of dynamical systems using binary measurements on the output has already been studied and several methods have been proposed. These methods address the identification problem in different ways: some of them use a periodic input signal ([14]), others use the knowledge of the noise distribution function ([14], [16], [5], etc.) and some approaches are based on a specific identification criteria ([3], [10], [11], etc.). These previous solutions are dedicated to the identification of Finite Impulse Response (FIR) systems but there also exist some solutions for the identification of Infinite Impulse Response (IIR) systems (see for instance [13]). It might be noticed that all of these methods require high resolution data on the input signal, consequently they are not adapted to our framework. There are less solutions in the case of one-bit quantization both on the input and the output. In [6] only the case of a threshold equal to zero is considered. [2] deals with the identification of FIR linear systems assuming that both input and output measurements are subjected to quantization. Two programming techniques based algorithms are presented therein. Considering the identification with only binary measurements on the input and output signals, [15] and [9] proposed algorithms for the identification of a gain system. They are extended in [7] and [8] where it is assumed that the thresholds of the one-bit quantizers can be adapted.

1.2 Contributions and paper outline

In this paper, we present an alternative for the identification of dynamical systems using binary measurements both on

⋆ This paper was not presented at any IFAC meeting. Corresponding author M. Pouliquen.

Email address: mathieu.pouliquen@unicaen.fr (Mathieu Pouliquen).

the input and the output. We investigate here the case where thresholds can be different from zero. The cornerstone of the proposed algorithm is the estimation, in a first time, of the correlation function of the input and the correlation function between the input and the output. These estimates are then used, in a second time, for the estimation of the parameters. This paper continues the development of the algorithm proposed in [1], this later algorithm being dedicated to the identification of Auto-Regressive models for time-series from one-bit quantized observation sequences.

The organization of the paper is given below. In section 2, we formally state the problem under consideration, we introduce some notations and assumptions. Section 3 is divided into two parts: in subsection 3.1 we present our solution, in subsection 3.2 an asymptotic analysis of this solution is provided. Section 4 illustrates the proposed method with a numerical example and confirms the analysis. Section 5 concludes the paper. Proofs are given in Appendix.

2 Notation and problem formulation

Consider a discrete-time linear system whose dynamic is given by $y_t = G(q)u_t$ where u_t is the input and y_t is the output. $G(q)$ is a FIR system of order n. It is defined by its impulse response $\{g_k\}_{k\in[0;n]}$ as follows: $G(q) = \sum_{k=0}^{n} g_k q^{-k}$ with q^{-1} the backward shift operator such that $q^{-1}u_t = u_{t-1}$. u_t and y_t are not known. As depicted on Fig. 1, the unique information about u_t and y_t is given by x_t and z_t as follows: $x_t = \mathbf{Q}_{C_u}(u_t + v_t^u)$, $z_t = \mathbf{Q}_{C_y}(y_t + v_t^y)$ where v_t^u and v_t^y are additive noises assumed to be wide sense stationary. $\mathbf{Q}_C(.)$ is the operator such that $\mathbf{Q}_C(a_t) = \begin{cases} 1 & \text{if } \frac{a_t}{\sigma_a} \geq C \\ 0 & \text{if } \frac{a_t}{\sigma_a} < C \end{cases}$ where C is a constant relative threshold which can be different from zero and $\sigma_a^2 = \mathscr{E}\{a_t^2\}$ is the variance of a_t. Thresholds C_u and C_y can be chosen independently if needed, however we use $C = C_u = C_y$ in the following for simplicity of presentation and without lose of generality.

Objective: The objective is, using N samples of $\{x_t\}$ and $\{z_t\}$, to estimate the parameter vector $\theta \in \mathbb{R}^{n+1}$ defined by $\theta^T = \begin{pmatrix} g_0 & g_1 & \cdots & g_n \end{pmatrix}^T$.

In this paper, we use the following notations: $\rho_{ab}(i) = \mathscr{E}\{a_t b_{t-i}\}$ and $\overline{\rho_{ab}}(i) = \frac{\rho_{ab}(i)}{\sigma_a \sigma_b}$ are respectively the correlation function of lag i between $\{a_t\}$ and $\{b_t\}$ and the normalized correlation function of lag i between $\{a_t\}$ and $\{b_t\}$. The following assumptions complete the description of the problem:

Assumption 1: u_t is a stationary sequence with normal distribution with a zero mean.
Assumption 2: v_t^u and v_t^y are zero mean white noises, with known variances $\sigma_{v^u}^2$ and $\sigma_{v^y}^2$, uncorrelated to each other and uncorrelated with u_t and y_t.

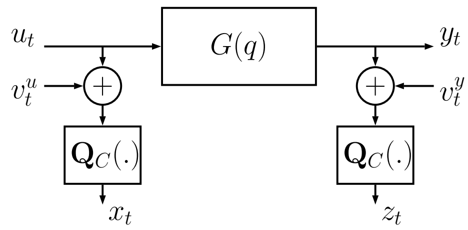

Fig. 1. The considered linear system with binary measurements on the input and the output.

Assumption 3: Variances σ_u^2 and σ_y^2 are known.

Remark 1 *Assumption 3 is a normalization assumption. Such a normalization assumption is usual in system identification using binary measurements on the output. This normalization assumption can take several forms: knowledge of the static gain for instance or knowledge of g_0 or assumption on θ such that $\|\theta\|_2 = 1$.*

3 Identification algorithm

3.1 Identification algorithm

It is well known that, if the input is a quasi-stationary sequence, then $\overline{\rho_{yu}}(i) = \frac{\sigma_u}{\sigma_y} \sum_{k=0}^{n} g_k \overline{\rho_{uu}}(i-k)$. It follows that it is possible to estimate parameters $\{g_k\}_{k\in[0;n]}$ if $\rho_{yu}(i)$ and $\rho_{uu}(i)$ are known for $i \in [0;n]$. The main difficulty here is the fact that samples of $\{u_t\}$ and $\{y_t\}$ are unknown and then $\rho_{yu}(i)$ and $\rho_{uu}(i)$ can not been directly estimated. In order to overcome this difficulty we propose the following three-step algorithm.

*3.1.1 **Step 1**: Estimation of $\rho_{zx}(i)$ and $\rho_{xx}(i)$*

Define $\widehat{\rho_{zx}}(i)$ the estimate of the correlation function $\rho_{zx}(i)$ and $\widehat{\rho_{xx}}(i)$ the estimate of the correlation function $\rho_{xx}(i)$. Using N samples of $\{x_t\}$ and $\{z_t\}$ it is possible to express $\widehat{\rho_{zx}}(i)$ and $\widehat{\rho_{xx}}(i)$ for $i \geq 0$ as follows: $\widehat{\rho_{zx}}(i) = \frac{1}{N-i}\sum_{t=i+1}^{N} z_t x_{t-i}$, $\widehat{\rho_{xx}}(i) = \frac{1}{N-i}\sum_{t=i+1}^{N} x_t x_{t-i}$.

*3.1.2 **Step 2**: Estimation of $\overline{\rho_{yu}}(i)$ and $\overline{\rho_{uu}}(i)$*

Consider first $\rho_{zx}(i)$. $\rho_{zx}(i)$ can be written as $\rho_{zx}(i) = \mathscr{P}_r\left\{\frac{y_t+v_t^y}{\sigma_{y+v^y}} \geq C, \frac{u_{t-i}+v_{t-i}^u}{\sigma_{u+v^u}} \geq C\right\}$ where $\sigma_{u+v^u} = \sqrt{\sigma_u^2 + \sigma_{v^u}^2}$ and $\sigma_{y+v^y} = \sqrt{\sigma_y^2 + \sigma_{v^y}^2}$. It follows that $\rho_{zx}(i)$ corresponds to the proportion of points $(y_t + v_t^y; u_{t-i} + v_{t-i}^u)$ such that $\frac{y_t+v_t^y}{\sigma_{y+v^y}} \geq C$ and $\frac{u_{t-i}+v_{t-i}^u}{\sigma_{u+v^u}} \geq C$. This proportion depends on $\overline{\rho_{(y+v^y)(u+v^u)}}(i)$. From the fact that $\{y_t + v_t^y\}$ and $\{u_t + v_t^u\}$ are normally distributed then this proportion, denoted $P_C\left(\overline{\rho_{(y+v^y)(u+v^u)}}(i)\right)$ in the following, can be expressed as follows: $P_C\left(\overline{\rho_{(y+v^y)(u+v^u)}}(i)\right) =$

$\frac{1}{2\pi\sqrt{1-\overline{\rho_{(y+v^y)(u+v^u)}}(i)^2}}\int_C^{+\infty}\int_C^{+\infty}\psi((y_t+v_t^y),(u_{t-i}+v_{t-i}^u))\,d(y_t+v_t^y)d(u_{t-i}+v_{t-i}^u)$ where $\psi((y_t+v_t^y),(u_{t-i}+v_{t-i}^u))$ is the following function

$$\psi(y_t,u_{t-i})=e^{-\frac{(y_t+v_t^y)^2+(u_{t-i}+v_{t-i}^u)^2-2\overline{\rho_{(y+v^y)(u+v^u)}}(i)(y_t+v_t^y)(u_{t-i}+v_{t-i}^u)}{2\left(1-\overline{\rho_{(y+v^y)(u+v^u)}}(i)^2\right)}}. \tag{1}$$

$P_C\left(\overline{\rho_{(y+v^y)(u+v^u)}}(i)\right)$ is a continuous monotone strictly increasing function of $\overline{\rho_{(y+v^y)(u+v^u)}}(i)$, it is then possible to define the function $P_C^{-1}(.)$ such that $P_C^{-1}(P_C(a))=a$. Define $\widehat{\overline{\rho_{(y+v^y)(u+v^u)}}}(i)$ the estimate of the normalized correlation $\overline{\rho_{(y+v^y)(u+v^u)}}(i)$. The second step of the algorithm consists in computing $\widehat{\overline{\rho_{(y+v^y)(u+v^u)}}}(i)$ from $\widehat{\overline{\rho_{(y+v^y)(u+v^u)}}}(i)=P_C^{-1}\left(\widehat{\rho_{zx}}(i)\right)$. Currently, for $C\neq 0$, there is no analytical expression for $P_C^{-1}(.)$, consequently in practice $\widehat{\overline{\rho_{(y+v^y)(u+v^u)}}}(i)$ is computed minimizing the criterion

$$\widehat{\overline{\rho_{(y+v^y)(u+v^u)}}}(i)=\text{ARGMIN}_{\overline{\rho_{(y+v^y)(u+v^u)}}(i)}\left|\widehat{\rho_{zx}}(i)-P_C\left(\overline{\rho_{(y+v^y)(u+v^u)}}(i)\right)\right|. \tag{2}$$

Then $\widehat{\overline{\rho_{yu}}}(i)$ can be computed as follows:

$$\widehat{\overline{\rho_{yu}}}(i)=\frac{\sigma_{y+v^y}\sigma_{u+v^u}}{\sigma_y\sigma_u}\widehat{\overline{\rho_{(y+v^y)(u+v^u)}}}(i). \tag{3}$$

Similarly, $\rho_{xx}(i)$ can be expressed using $\overline{\rho_{(u+v^u)(u+v^u)}}(i)$. $\widehat{\overline{\rho_{(u+v^u)(u+v^u)}}}(i)$ can then be computed minimizing the criterion

$$\widehat{\overline{\rho_{(u+v^u)(u+v^u)}}}(i)=\text{ARGMIN}_{\overline{\rho_{(u+v^u)(u+v^u)}}(i)}\left|\widehat{\rho_{xx}}(i)-P_C\left(\overline{\rho_{(u+v^u)(u+v^u)}}(i)\right)\right|. \tag{4}$$

and then

$$\widehat{\overline{\rho_{uu}}}(i)=\frac{\sigma_{u+v^u}^2}{\sigma_u^2}\widehat{\overline{\rho_{(u+v^u)(u+v^u)}}}(i)-\frac{\sigma_{v^u}^2}{\sigma_u^2}\delta(i). \tag{5}$$

where $\delta(i)$ the discrete-time impulse function.

3.1.3 *Step 3: Estimation of* $\{g_k\}_{k\in[0;n]}$

Denote $\mathbf{M}\left(\overline{\rho_{yu}}(i)\right)\in\mathbb{R}^{(n+1)\times(n+1)}$ and $\mathbf{N}\left(\overline{\rho_{uu}}(i)\right)\in\mathbb{R}^{n+1}$ as follows:

$$\mathbf{M}\left(\overline{\rho_{uu}}(i)\right)=\begin{pmatrix}1 & \overline{\rho_{uu}}(1) & \dots & \overline{\rho_{uu}}(n)\\ \overline{\rho_{uu}}(1) & 1 & \dots & \overline{\rho_{uu}}(n-1)\\ \vdots & \vdots & \vdots & \vdots\\ \overline{\rho_{uu}}(n) & \overline{\rho_{uu}}(n-1) & \dots & 1\end{pmatrix};\ \mathbf{N}\left(\overline{\rho_{yu}}(i)\right)=\begin{pmatrix}\overline{\rho_{yu}}(0)\\ \overline{\rho_{yu}}(1)\\ \dots\\ \overline{\rho_{yu}}(n)\end{pmatrix}. \tag{6}$$

θ can be expressed as follows: $\mathbf{N}\left(\overline{\rho_{yu}}(i)\right)=\frac{\sigma_u}{\sigma_y}\mathbf{M}\left(\overline{\rho_{uu}}(i)\right)\theta$. The third step consists in computing the estimate $\widehat{\theta}_N$ as follows:

$$\widehat{\theta}_N=\frac{\sigma_y}{\sigma_u}\mathbf{M}\left(\widehat{\overline{\rho_{uu}}}(i)\right)^{-1}\mathbf{N}\left(\widehat{\overline{\rho_{yu}}}(i)\right). \tag{7}$$

The proposed algorithm is summarized in Algorithm 1. $\widehat{\rho_{zx}}(i)$, $\widehat{\rho_{xx}}(i)$, $\widehat{\overline{\rho_{yu}}}(i)$ and $\widehat{\overline{\rho_{uu}}}(i)$ depend on N. For the sake of simplicity, we omit this dependence on N in the notation.

Algorithm 1.

input: n, $\{x_t\}_{t\in[1;N]}$, $\{z_t\}_{t\in[1;N]}$

1- Compute $\widehat{\rho_{zx}}(i)$ and $\widehat{\rho_{xx}}(i)$.

2- Compute $\widehat{\overline{\rho_{(y+v^y)(u+v^u)}}}(i)$ and $\widehat{\overline{\rho_{(u+v^u)(u+v^u)}}}(i)$ from (2) and (4) and then compute $\widehat{\overline{\rho_{yu}}}(i)$ and $\widehat{\overline{\rho_{uu}}}(i)$ from (3) and (5).

3- Compute $\widehat{\theta}_N$ from (7).

Table 1
Algorithm 1: a batch algorithm using $\{x_t\}_{t\in[1;N]}$ and $\{z_t\}_{t\in[1;N]}$

Remark 2 *For $C=0$, from a result presented in [12], it can be shown that $\overline{\rho_{(u+v^u)(u+v^u)}}(i)=cos\left(\pi\left(1-2\rho_{xx}(i)\right)\right)$.*

3.2 *Analysis and comments*

An analysis of the asymptotical behavior of Algorithm 1 is proposed in this section. Theorem 1 below shows that the proposed algorithm is asymptotically unbiased.

Theorem 1 *Consider assumptions of section 2. Algorithm 1 is such that*

$$\lim_{N\to\infty}\widehat{\theta}_N=\theta. \tag{8}$$

The proof of this theorem is given in appendix A. We now present some theorems in order to characterize asymptotically the estimate $\widehat{\theta}_N$. Theorem 2 below establishes the asymptotic variance of $\widehat{\rho_{zx}}(i)$ and $\widehat{\rho_{xx}}(i)$.

Theorem 2 *Consider assumptions of section 2. Denote* $\widehat{\rho_{zx}}^T=\left(\widehat{\rho_{zx}}(0)\ \cdots\ \widehat{\rho_{zx}}(n)\right)$ *and* $\widehat{\rho_{xx}}^T=\left(\widehat{\rho_{xx}}(1)\ \cdots\ \widehat{\rho_{xx}}(n)\right)$.

- *The asymptotic distribution of $\widehat{\rho_{zx}}$ is $\mathcal{N}\left(\rho_{zx},\frac{1}{N}\Sigma^{zx}\right)$ where Σ^{zx} denotes a $(n+1)\times(n+1)$ matrix. The element at the $(i+1)^{th}$ line and $(j+1)^{th}$ column, with $i\in[0,n]$ and $j\in[0,n]$, is denoted $\sigma_{i,j}^{zx}$ and defined by*

$$\sigma_{i,j}^{zx}=\sum_{h=-\infty}^{\infty}\left(\mathscr{E}\left\{z_tx_{t-i}z_{t+h}x_{t+h-j}\right\}-\mathscr{E}\left\{z_tx_{t-i}\right\}\mathscr{E}\left\{z_tx_{t-j}\right\}\right). \tag{9}$$

- *The asymptotic distribution of $\widehat{\rho_{xx}}$ is $\mathcal{N}\left(\rho_{xx},\frac{1}{N}\Sigma^{xx}\right)$ where Σ^{xx} denotes a $n\times n$ matrix. The element at the i^{th} line and j^{th} column, with $i\in[1,n]$ and $j\in[1,n]$, is denoted $\sigma_{i,j}^{xx}$ and defined by*

$$\sigma_{i,j}^{xx}=\sum_{h=-\infty}^{\infty}\left(\mathscr{E}\left\{x_tx_{t-i}x_{t+h}x_{t+h-j}\right\}-\mathscr{E}\left\{x_tx_{t-i}\right\}\mathscr{E}\left\{x_tx_{t-j}\right\}\right). \tag{10}$$

- *The covariance of $\widehat{\rho_{zx}}$ and $\widehat{\rho_{xx}}$ is $\frac{1}{N}\Sigma_{i,j}^{zx/xx}$, a $(n+1)\times n$ matrix. The element at the $(i+1)^{th}$ line and j^{th} column,* *with $i \in [0,n]$ and $j \in [1,n]$, is denoted $\sigma_{i,j}^{zx/xx}$ and defined by*

$$\sigma_{i,j}^{zx/xx} = \sum_{h=-\infty}^{\infty} \left(\mathscr{E}\{z_t x_{t-i} x_{t+h} x_{t+h-j}\} - \mathscr{E}\{z_t x_{t-i}\}\mathscr{E}\{x_t x_{t-j}\}\right). \quad (11)$$

The proof of this theorem is given in appendix B. Let us notice that $\mathscr{E}\{z_t x_{t-i} z_{t+h} x_{t+h-j}\} = \mathscr{P}_r\{\frac{y_t+v_t^y}{\sigma_{y+v^y}} \geq C, \frac{u_{t-i}+v_{t-i}^u}{\sigma_{u+v^u}} \geq C, \frac{y_{t+h}+v_{t+h}^y}{\sigma_{y+v^y}} \geq C, \frac{u_{t+h-j}+v_{t+h-j}^u}{\sigma_{u+v^u}} \geq C\}$ can be expressed with a quadruple integral using $\overline{\rho_{(y+v^y)(u+v^u)}}(i)$ and $\overline{\rho_{(u+v^u)(u+v^u)}}(i)$ for different lags i. It follows that Σ^{zx} can be computed if the previous correlation function are known. The same applies for Σ^{xx} and $\Sigma^{zx/xx}$. Theorem 3 below focuses on the asymptotic variance of $\widehat{\rho_{yu}}(i)$ and $\widehat{\rho_{uu}}(i)$.

Theorem 3 *Consider assumptions of section 2. Denote $\widehat{\rho_{yu}}^T = \left(\widehat{\rho_{yu}}(0) \cdots \widehat{\rho_{yu}}(n)\right)$, $\widehat{\rho_{uu}}^T = \left(\widehat{\rho_{uu}}(1) \cdots \widehat{\rho_{uu}}(n)\right)$ and $P_C^{-1'}(a) = \frac{dP_C^{-1}(a)}{da}$.*

- *The asymptotic distribution of $\widehat{\rho_{yu}}$ is $\mathscr{N}\left(\overline{\rho_{yu}}, \frac{1}{N}\Sigma^{yu}\right)$ where Σ^{yu} denotes a $(n+1)\times(n+1)$ matrix. The element at the $(i+1)^{th}$ line and $(j+1)^{th}$ column, with $i \in [0,n]$ and $j \in [0,n]$, is denoted $\sigma_{i,j}^{yu}$ and defined by*

$$\sigma_{i,j}^{yu} = \frac{\sigma_{y+v^y}^2 \sigma_{u+v^u}^2}{\sigma_y^2 \sigma_u^2} \left(P_C^{-1'}(\rho_{zx}(i))\right) \sigma_{i,j}^{zx} \left(P_C^{-1'}(\rho_{zx}(j))\right). \quad (12)$$

- *The asymptotic distribution of $\widehat{\rho_{uu}}$ is $\mathscr{N}\left(\overline{\rho_{uu}}, \frac{1}{N}\Sigma^{uu}\right)$ where Σ^{uu} denotes a $n \times n$ matrix. The element at the i^{th} line and j^{th} column, with $i \in [1,n]$ and $j \in [1,n]$, is denoted $\sigma_{i,j}^{uu}$ and defined by*

$$\sigma_{i,j}^{uu} = \frac{\sigma_{u+v^u}^4}{\sigma_u^4} \left(P_C^{-1'}(\rho_{xx}(i))\right) \sigma_{i,j}^{xx} \left(P_C^{-1'}(\rho_{xx}(j))\right). \quad (13)$$

- *The covariance of $\widehat{\rho_{yu}}$ and $\widehat{\rho_{uu}}$ is $\frac{1}{N}\Sigma_{i,j}^{yu/uu}$, a $(n+1)\times n$ matrix. The element at the $(i+1)^{th}$ line and j^{th} column, with $i \in [0,n]$ and $j \in [1,n]$, is denoted $\sigma_{i,j}^{yu/uu}$ and defined by*

$$\sigma_{i,j}^{yu/uu} = \frac{\sigma_{y+v^y}\sigma_{u+v^u}^3}{\sigma_y \sigma_u^3} \left(P_C^{-1'}(\rho_{zx}(i))\right) \sigma_{i,j}^{zx/xx} \left(P_C^{-1'}(\rho_{xx}(j))\right). \quad (14)$$

The proof of this theorem is given in appendix C. Theorem 4 below establishes the asymptotic variance of $\widehat{\theta}_N$.

Theorem 4 *Consider assumptions of section 2. Algorithm 1 is such that the asymptotic distribution of $\widehat{\theta}_N$ is $\mathscr{N}\left(\theta, \frac{1}{N}\Sigma^{\theta}\right)$ where Σ^{θ} denotes the $(n+1)\times(n+1)$ matrix*

$$\begin{aligned}\Sigma^{\theta} = \quad & \textstyle\sum_{i\in[0;n],j\in[0;n]} \left(\frac{\partial\theta}{\partial\overline{\rho_{yu}}(i)}\right) \sigma_{i,j}^{yu} \left(\frac{\partial\theta}{\partial\overline{\rho_{yu}}(j)}\right)^T \\ & + \textstyle\sum_{i\in[1;n],j\in[1;n]} \left(\frac{\partial\theta}{\partial\overline{\rho_{uu}}(i)}\right) \sigma_{i,j}^{uu} \left(\frac{\partial\theta}{\partial\overline{\rho_{uu}}(j)}\right)^T \\ & + \textstyle\sum_{i\in[0;n],j\in[1;n]} \left(\frac{\partial\theta}{\partial\overline{\rho_{yu}}(i)}\right) \sigma_{i,j}^{yu/uu} \left(\frac{\partial\theta}{\partial\overline{\rho_{uu}}(j)}\right)^T \\ & + \textstyle\sum_{i\in[1;n],j\in[0;n]} \left(\frac{\partial\theta}{\partial\overline{\rho_{uu}}(i)}\right) \sigma_{i,j}^{uu/yu} \left(\frac{\partial\theta}{\partial\overline{\rho_{yu}}(j)}\right)^T\end{aligned}. \quad (15)$$

The proof of this theorem is given in appendix D. Theorem 5 below provides the mean square convergence rate of Algorithm 1.

Theorem 5 *Consider assumptions of section 2. Defining $\|.\|_2$ as the norm 2, Algorithm 1 is such that*

$$\mathscr{E}\left\{\|\widehat{\theta}_N - \theta\|_2^2\right\} = \mathscr{O}\left(\frac{1}{N}\right). \quad (16)$$

The proof of this theorem is given in appendix E. Notice that terms $\frac{\partial\theta}{\partial\overline{\rho_{yu}}(i)}$ and $\frac{\partial\theta}{\partial\overline{\rho_{uu}}(i)}$ in Theorem 4 can be analytically evaluated with $\mathbf{M}(\overline{\rho_{uu}}(i))$, $\mathbf{N}(\overline{\rho_{yu}}(i))$ and θ. It is then possible to use these theorems so as to compute an a posteriori estimate of the variance via the following process:

- Compute $\widehat{\theta}_N$ with Algorithm 1.
- Compute $\sigma_{i,j}^{zx}$, $\sigma_{i,j}^{xx}$ and $\sigma_{i,j}^{zx/xx}$ from (9), (10) and (11) using $\widehat{\rho_{(y+v^y)(u+v^u)}}(i)$ and $\widehat{\rho_{(u+v^u)(u+v^u)}}(i)$.
- Compute $\sigma_{i,j}^{yu}$, $\sigma_{i,j}^{uu}$ and $\sigma_{i,j}^{yu/uu}$ from (12), (13) and (14).
- Compute Σ^{θ} from (15) using $\widehat{\theta}_N$.

From these theorems we can conclude that:

- The asymptotic variance of $\widehat{\theta}_N$ depends on N: the variance is lower with a higher N.
- The asymptotic variance of $\widehat{\theta}_N$ also depends on C and this can be evaluated from (12), (13) and (14). It can be shown that $P_C^{-1'}(\rho_{zx}(i)) = \frac{1}{P_C'(\overline{\rho_{yu}}(i))}$ with $P_C'(a) = \frac{dP_C(a)}{da}$. By depicting $P_C'(\overline{\rho_{yu}}(i))$ it can be seen that variance is more important for high $|C|$.
- The asymptotic variance of $\widehat{\theta}_N$ depends on the noise level. This can be seen in (9), (10) and (11) from the fact that $\sigma_{i,j}^{zx}$, $\sigma_{i,j}^{xx}$ and $\sigma_{i,j}^{zx/xx}$ depends on σ_{v^y} and σ_{v^u}. It is quite difficult to precisely interpret the impact of σ_{v^y} and σ_{v^u}. Some numerical simulations proposed in section 4 show that the variance increases for low Signal-to-Noise Ratio.

Remark 3 *The estimation of θ and Σ^{θ} require the knowledge of the variance of the noise, the input and the output. About θ, this dependency justifies the normalization assumption 3. About Σ^{θ}, the previous variances may significantly influence the accuracy of the estimate of Σ^{θ}, especially since it is often difficult to have precise values of these variances in practical uses.*

4 Numerical examples

This section provides some numerical results to illustrate performance of the proposed method. Experimental data are generated according to section 2. The system is the following FIR system of order $n=3$: $G(q) = 0.8219 + 0.5011q^{-1} +$

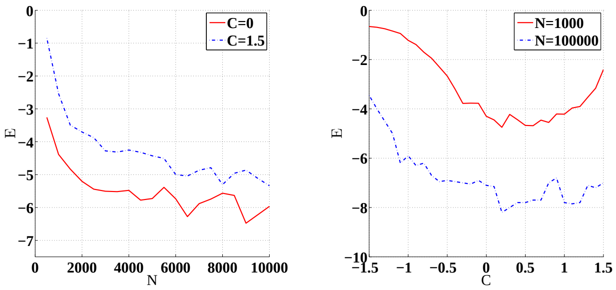

Fig. 2. $E = log_{10}\left(\|\theta - mean(\widehat{\theta}_N)\|\right)$ as a function of N and C.

N	400	800	1200	1600	2000
E (SNR=100dB)	-3.09	-3.93	-4.52	-4.65	-4.98
E (SNR=5dB)	-1.9	-2.90	-3.53	-4.02	-4.20

Table 2
$E = log_{10}\left(\|\theta - mean(\widehat{\theta}_N)\|\right)$ as a function of N with $C = 0$, for SNR= $100dB$ and SNR= $5dB$.

$0.2516q^{-2} + 0.1003q^{-3}$. The input sequence u_t is a zero mean random sequence with normal distribution and $\sigma_u = 1$. Noises v_t^u and v_t^y are zero mean Gaussian random signal with the same standard deviation chosen in order to test different values of the Signal-to-Noise Ratio (SNR). In a first experiment we investigate the influence of N. Two Monte Carlo simulations, for $C = 0$ and $C = 1.5$, are carried out with 100 runs, $v_t^u = v_t^y = 0$ and for N from 500 to 10000. Performance of the algorithm is evaluated by means of the size of the parameter error vector $E = log_{10}\left(\|\theta - mean(\widehat{\theta}_N)\|\right)$. Fig. 2 left presents E as a function of N. It appears that performance increases for high N and depends on C. In a second experiment we investigate the influence of C. Two Monte Carlo simulations are carried out with 100 runs for $N = 50000$ and $N = 100000$, in both cases with $v_t^u = v_t^y = 0$. Fig. 2 right presents E as a function of C. It appears that performance degrades for high $|C|$. In a third experiment we investigate the influence of SNR. Two Monte Carlo simulations, for SNR= $100dB$ and SNR= $5dB$, are carried out with 100 runs and for N from 200 to 2000. The threshold is $C = 0$. Tab. 2 presents E as a function of N. These results show that: (1-) the proposed algorithm is asymptotically unbiased in presence of noise, (2-) the proposed algorithm has better performance for high SNR. About the variance, Fig. 3 presents, without and with noise, the experimental value of $N\sum_{i\in[1;n]}\sigma^2_{\widehat{\theta}_N(i)}$ as a function of C and its theorical value obtained with Theorem 4, i.e. $trace\left(\Sigma^\theta\right)$. The correspondence between experimental and theorical results confirms analysis of subsection 3.2.

5 Conclusion and Future work

In this paper we have proposed an identification algorithm for FIR systems using binary measurements both on the input and the output. The proposed algorithm is an offline

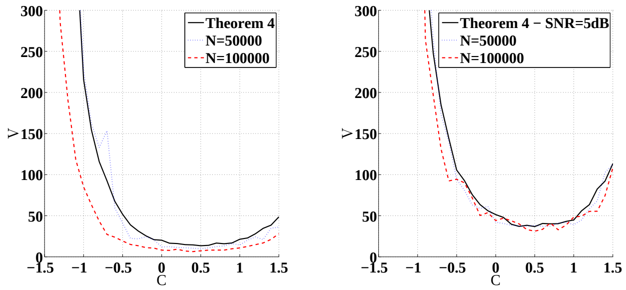

Fig. 3. $V = N\sum_{i\in[1;n]}\sigma^2_{\widehat{\theta}_N(i)}$ as a function of C for $N = 50000$ and $N = 100000$ without and with noise.

algorithm. It is organized as follows: first, the correlation function of the input and the correlation function between the input and the output are estimated. Second, the parameters of the model are computed. The proposed algorithm is shown to be asymptotically unbiased. Simulation results are included to validate the proposed identification algorithm and to confirm the variance analysis. Some first extensions of the algorithm can be the following: the proposition of an online version of the algorithm, the extension to IIR systems, the extension to other distribution on the input signal.

A Proof of Theorem 1

From the fact that u_t, v_t^u and v_t^y are stationary sequences, x_t and z_t are also stationary sequences. It follows that for all i we have $\lim_{N\to\infty}\widehat{\rho_{xx}}(i) = \mathscr{E}\{x_t x_{t-i}\} = \rho_{xx}(i)$ and $\lim_{N\to\infty}\widehat{\rho_{zx}}(i) = \mathscr{E}\{z_t x_{t-i}\} = \rho_{zx}(i)$ which means that $\widehat{\rho_{xx}}(i)$ is an asymptotically unbiased estimation of $\rho_{xx}(i)$ and $\widehat{\rho_{zx}}(i)$ is an asymptotically unbiased estimation of $\rho_{zx}(i)$ for all i. $\widehat{\overline{\rho_{uu}}}(i)$ and $\widehat{\overline{\rho_{yu}}}(i)$ are directly computed from $P_C^{-1}\left(\widehat{\rho_{xx}}(i)\right)$ and $P_C^{-1}\left(\widehat{\rho_{zx}}(i)\right)$. $P_C^{-1}(.)$ is a continuous function, it follows $\lim_{N\to\infty}\widehat{\overline{\rho_{uu}}}(i) = \overline{\rho_{uu}}(i)$ and $\lim_{N\to\infty}\widehat{\overline{\rho_{yu}}}(i) = \overline{\rho_{yu}}(i)$., then we have $\lim_{N\to\infty}\mathbf{M}\left(\widehat{\overline{\rho_{uu}}}(i)\right) = \mathbf{M}\left(\overline{\rho_{uu}}(i)\right)$ and $\lim_{N\to\infty}\mathbf{N}\left(\widehat{\overline{\rho_{yu}}}(i)\right) = \mathbf{N}\left(\overline{\rho_{yu}}(i)\right)$. It follows from (7) that $\widehat{\theta}_N$ satisfies $\lim_{N\to\infty}\widehat{\theta}_N = \theta$ which concludes the proof.

B Proof of Theorem 2

From proof of Theorem 1 the estimation of ρ_{zx} is asymptotically unbiased. The covariance on the estimate $\widehat{\rho_{zx}}$ is a $(n+1)\times(n+1)$ matrix. In this matrix, the element at the $(i+1)^{th}$ line and $(j+1)^{th}$ column is denoted $Cov_{zx}(i,j)$ and is defined by $Cov_{zx}(i,j) = \mathscr{E}\left\{\left(\widehat{\rho_{zx}}(i) - \rho_{zx}(i)\right)\left(\widehat{\rho_{zx}}(j) - \rho_{zx}(j)\right)\right\}$. We have

$$Cov_{zx}(i,j) = \frac{1}{(N-i)(N-j)}\sum_{t=i+1}^{N}\sum_{t'=j+1}^{N}\left(\mathscr{E}\left\{z_t x_{t-i} z_{t'} x_{t'-j}\right\} - \rho_{zx}(i)\rho_{zx}(j)\right) \quad \text{(B.1)}$$

From [4] we have asymptotically

$$\lim_{N\to\infty} NCov_{zx}(i,j) = \sum_{h=-\infty}^{\infty}\left(\mathscr{E}\left\{z_t x_{t-i} z_{t+h} x_{t+h-j}\right\} - \rho_{zx}(i)\rho_{zx}(j)\right) \quad \text{(B.2)}$$

The asymptotic covariance on the estimate $\widehat{\rho_{zx}}$ is then a matrix $\frac{1}{N}\Sigma^{zx}$ where Σ^{zx} is a matrix with elements $\sigma_{i,j}^{zx}$ given by (9). Similarly, it is possible to characterize the asymptotic distribution of $\widehat{\rho_{xx}}$ and the covariance between $\widehat{\rho_{zx}}$ and $\widehat{\rho_{xx}}$.

C Proof of Theorem 3

From proof of Theorem 1 the estimation of $\overline{\rho_{yu}}$ is asymptotically unbiased. $\widehat{\overline{\rho_{yu}}}(i)$ is computed from $\widehat{\overline{\rho_{yu}}}(i) = \frac{\sigma_{y+v^y}\sigma_{u+v^u}}{\sigma_y\sigma_u}P_C^{-1}\left(\widehat{\rho_{zx}}(i)\right)$ where $P_C^{-1}(.)$ is continuously differentiable. Using a first order Taylor approximation on $P_C^{-1}(.)$ it follows that the asymptotic covariance of $\widehat{\rho_{yu}}$ is a $(n+1)\times(n+1)$ matrix where the $(i+1)^{th}$ line and $(j+1)^{th}$ column corresponds to $\mathscr{E}\left\{\left(\widehat{\rho_{yu}}(i)-\rho_{yu}(i)\right)\left(\widehat{\rho_{yu}}(j)-\rho_{yu}(j)\right)\right\} = \frac{1}{N}\frac{\sigma_{y+v^y}^2\sigma_{u+v^u}^2}{\sigma_y^2\sigma_u^2}(P_C^{-1'}(\rho_{zx}(i)))\sigma_{i,j}^{zx}(P_C^{-1'}(\rho_{zx}(j)))$. Similarly, it is possible to characterize the asymptotic distribution of $\widehat{\overline{\rho_{uu}}}$ and the covariance between $\widehat{\overline{\rho_{yu}}}$ and $\widehat{\overline{\rho_{uu}}}$.

D Proof of Theorem 4

From Theorem 1 the estimation of θ is asymptotically unbiased. Moreover $\widehat{\theta}_N$ is computed with $\widehat{\theta}_N = \frac{\sigma_y}{\sigma_u}\mathbf{M}\left(\widehat{\overline{\rho_{uu}}}(i)\right)^{-1}\mathbf{N}\left(\widehat{\overline{\rho_{yu}}}(i)\right)$. From the fact that θ is a continuous differentiable function of $\overline{\rho_{yu}}$ and $\overline{\rho_{uu}}$, it follows from Taylor's theorem that $\widehat{\theta}_N-\theta$ can be written has

$$\begin{aligned}\widehat{\theta}_N-\theta = & \textstyle\sum_{i\in[0;n]}\left(\frac{\partial\theta}{\partial\overline{\rho_{yu}}(i)}\right)\left(\widehat{\overline{\rho_{yu}}}(i)-\overline{\rho_{yu}}(i)\right)+\sum_{i\in[1;n]}\left(\frac{\partial\theta}{\partial\overline{\rho_{uu}}(i)}\right)\left(\widehat{\overline{\rho_{uu}}}(i)-\overline{\rho_{uu}}(i)\right)\\ & +\|\widehat{\overline{\rho}}-\overline{\rho}\|_2^2\varepsilon(\widehat{\overline{\rho}}-\overline{\rho})\end{aligned} \tag{D.1}$$

where $\overline{\rho}^T = \left(\overline{\rho_{yu}}^T\ \overline{\rho_{uu}}^T\right)\in\mathbb{R}^{2n+1}$, $\|.\|_2$ is the norm 2 and $\varepsilon(.)$ a function from $\mathbb{R}^{2n+1}$ to $\mathbb{R}^{n+1}$ such that $\lim_{\overline{\rho}\to\infty}\varepsilon(\overline{\rho})=0$. From Theorem 3 we know $\mathscr{E}\left\{\left(\widehat{\overline{\rho_{yu}}}(i)-\overline{\rho_{yu}}(i)\right)\left(\widehat{\overline{\rho_{yu}}}(j)-\overline{\rho_{yu}}(j)\right)\right\}$, $\mathscr{E}\left\{\left(\widehat{\overline{\rho_{uu}}}(i)-\overline{\rho_{uu}}(i)\right)\left(\widehat{\overline{\rho_{uu}}}(j)-\overline{\rho_{uu}}(j)\right)\right\}$, $\mathscr{E}\left\{\left(\widehat{\overline{\rho_{yu}}}(i)-\overline{\rho_{yu}}(i)\right)\left(\widehat{\overline{\rho_{uu}}}(j)-\overline{\rho_{uu}}(j)\right)\right\}$ and $\mathscr{E}\left\{\left(\widehat{\overline{\rho_{uu}}}(i)-\overline{\rho_{uu}}(i)\right)\left(\widehat{\overline{\rho_{yu}}}(j)-\overline{\rho_{yu}}(j)\right)\right\}$. Moreover, from the proof of Theorem 1 we have $\lim_{N\to\infty}\widehat{\overline{\rho}}=\overline{\rho}$, it follows that $\mathscr{E}\left\{\left(\widehat{\theta}_N-\theta\right)\left(\widehat{\theta}_N-\theta\right)^T\right\}=\frac{1}{N}\Sigma^\theta$ with

$$\begin{aligned}\Sigma^\theta = & \textstyle\sum_{i\in[0;n],j\in[0;n]}\left(\frac{\partial\theta}{\partial\overline{\rho_{yu}}(i)}\right)\sigma_{i,j}^{yu}\left(\frac{\partial\theta}{\partial\overline{\rho_{yu}}(j)}\right)^T\\ & +\textstyle\sum_{i\in[1;n],j\in[1;n]}\left(\frac{\partial\theta}{\partial\overline{\rho_{uu}}(i)}\right)\sigma_{i,j}^{uu}\left(\frac{\partial\theta}{\partial\overline{\rho_{uu}}(j)}\right)^T\\ & +\textstyle\sum_{i\in[0;n],j\in[1;n]}\left(\frac{\partial\theta}{\partial\overline{\rho_{yu}}(i)}\right)\sigma_{i,j}^{yu/uu}\left(\frac{\partial\theta}{\partial\overline{\rho_{uu}}(j)}\right)^T\\ & +\textstyle\sum_{i\in[1;n],j\in[0;n]}\left(\frac{\partial\theta}{\partial\overline{\rho_{uu}}(i)}\right)\sigma_{i,j}^{uu/yu}\left(\frac{\partial\theta}{\partial\overline{\rho_{yu}}(j)}\right)^T\end{aligned} \tag{D.2}$$

where $\sigma_{i,j}^{yu}$, $\sigma_{i,j}^{uu}$, $\sigma_{i,j}^{yu/uu}$ and $\sigma_{i,j}^{uu/yu}$ are given in Theorem 3.

E Proof of Theorem 5

Theorem 5 is a consequence of Theorem 4. From Theorem 4 we have $\mathscr{E}\left\{\left(\widehat{\theta}_N-\theta\right)\left(\widehat{\theta}_N-\theta\right)^T\right\}=\frac{1}{N}\Sigma^\theta$. It follows that $\mathscr{E}\left\{\|\widehat{\theta}_N-\theta\|_2^2\right\}=\frac{1}{N}trace(\Sigma^\theta)$. This gives (16).

References

[1] R. Auber, M. Pouliquen, E. Pigeon, M. M'Saad, O. Gehan, P.A. Chapon, and S. Moussay. Identification of ar time-series based on binary data. *IET Signal processing, DOI: 10.1049/iep-spr.2019.0152*, 2019.

[2] V. Cerone, D. Piga, and D. Regruto. Fixed-order FIR approximation of linear systems from quantized input and output data. *Systems and Control Letters*, 62:1136–1142, 2013.

[3] E. Colinet and J. Juillard. A weighted least-squares approach to parameter estimation problems based on binary measurements. *IEEE Transactions on Automatic Control*, 55(1):148–152, 2010.

[4] A.V. Dandawate and G.B. Giannakis. Asymptotic properties and covariance expressions of kth-order sample moments and cumulants. *Proceedings of 27th Asilomar Conference on Signals, Systems and Computers, Pacific Grove*, 1993.

[5] J. Guo, L.Y. Wang, G. Yin, Y. Zhao, and J.F. Zhang. Identification of Wiener systems with quantized inputs and binary-valued output observations. *Automatica*, 78:280–286, 2017.

[6] V. Krishnamurthy. Estimation of quantized linear errors-in-variables models. *Automatica*, 31:1459–1464, 1995.

[7] A.S. Leong, E. Weyer, and G.N. Nair. On the identification of FIR systems with binary input and output observations. *IEEE Conference on Decision and Control, Las Vegas, 2016*.

[8] A.S. Leong, E. Weyer, and G.N. Nair. Identification of FIR systems with binary input and output observations. *Technical report*, 2018.

[9] Y. Lian, Z. Luo, E. Weyer, and G.N. Nair. Parameter estimation with binary observations of input and output signals. *Australian Control Conference, Newcastle - Australia, 2016*.

[10] M. Pouliquen, T. Menard, E. Pigeon, O. Gehan, and A. Goudjil. Recursive system identification algorithm using binary measurements. *European Conference on Control, Aalborg*, 2016.

[11] M. Pouliquen, E. Pigeon, O. Gehan, and A. Goudjil. Identification using binary measurements for IIR systems. *IEEE Transactions on Automatic Control*, 2019.

[12] W. F. Sheppard. On the application of the theory of error to cases of normal distribution and normal correlation. *Philosophical Transactions of the Royal Society A*, 192:101–167, 1899.

[13] L.Y. Wang, G.G. Yin, and J. Zhang. Joint identification of plant rational models and noise distribution functions using binary-valued observations. *Automatica*, 42(4):535 – 547, 2006.

[14] L.Y. Wang, J.F. Zhang, and G. Yin. System identification using binary sensors. *IEEE Transactions on Automatic Control*, 48(11):1892–1907, 2003.

[15] K. You, E. Weyer, and G.N. Nair. Identification of a gain system with binary input and output measurements. *IEEE Conference on Decision and Control, Osaka, 2015*.

[16] Y. Zhao, L.Y Wang, G.G. Yin, and J.F. Zhang. Identification of Wiener systems with binary-valued output observations. *Automatica*, 43(10):1752–1765, 2007.