Available online at www.sciencedirect.com

ScienceDirect

IFAC-PapersOnLine 48-28 (2015) 455–460

Bayesian kernel-based system identification with quantized output data

Giulio Bottegal * **Gianluigi Pillonetto** ** **Håkan Hjalmarsson** *

* *Automatic Control Lab and ACCESS Linnaeus Centre, School of Electrical Engineering, KTH Royal Institute of Technology, Stockholm, Sweden (e-mail: {bottegal; hjalmars}@kth.se)*

** *Department of Information Engineering, University of Padova, Padova, Italy (e-mail: giapi@dei.unipd.it)*

Abstract: In this paper we introduce a novel method for linear system identification with quantized output data. We model the impulse response as a zero-mean Gaussian process whose covariance (kernel) is given by the recently proposed stable spline kernel, which encodes information on regularity and exponential stability. This serves as a starting point to cast our system identification problem into a Bayesian framework. We employ Markov Chain Monte Carlo (MCMC) methods to provide an estimate of the system. In particular, we show how to design a Gibbs sampler which quickly converges to the target distribution. Numerical simulations show a substantial improvement in the accuracy of the estimates over state-of-the-art kernel-based methods when employed in identification of systems with quantized data.

1. INTRODUCTION

Identification of systems from quantized data finds applications in a wide range of areas such as communications, networked control systems, bioinformatics (see e.g. [Bae and Mallick, 2004] and [Wang et al., 2010]).

From a system identification perspective, identification of systems having quantized output data constitutes a challenging problem. The presence of a quantizer cascaded to a dynamic system causes loss of information on the system dynamics and standard system identification techniques, such as least-squares or prediction error method (PEM) [Ljung, 1999], [Söderström and Stoica, 1989], may give poor performances. For this reason, in recent years several techniques for system identification from quantized data have been proposed in a series of papers. Some of these methods are specifically tailored for identification of systems with binary measurements [Wang et al., 2003], [Wang et al., 2006], and are possibly implemented in a recursive fashion [Guo and Zhao, 2013], [Jafari et al., 2012]. Other methods, e.g. [Colinet and Juillard, 2010], exploit the knowledge of a dithering signal to improve the identification performances. Specific input design techniques are studied in [Godoy et al., 2014], [Casini et al., 2011] and [Casini et al., 2012]. Methods for handling general types of quantization of data have been proposed recently [Godoy et al., 2011], [Chen et al., 2012b]. In these contributions, the problem of identifying a linear dynamic system with quantized data is posed as a likelihood problem. In particular, in [Chen et al., 2012b] authors exploit the recently proposed Bayesian kernel-based formulation of the linear dynamic system identification problem (see [Pillonetto et al., 2014] for a survey).

Similarly to [Chen et al., 2012b], the starting point of this paper is the formulation of the problem of identifying a linear dynamic systems with quantized data using a Bayesian approach. We model the impulse response of the unknown system as a realization of a Gaussian random process. Such a process has zero mean and its covariance matrix (in this context also called a *kernel*) is given by the recently introduced *stable spline kernels*, [Pillonetto and De Nicolao, 2010], [Pillonetto et al., 2011], [Bottegal and Pillonetto, 2013] which are specifically designed for linear system identification purposes. The structure of this type of kernels depends on two *hyperparameters*, namely a *scaling* parameter and a *shaping* parameter. Tuning these parameters can be seen as a model selection step. In the standard setting (i.e. when there is no quantizer), kernel hyperparameters are chosen as those maximizing the marginal likelihood of the output measurements, obtained by integrating out the dependence on the system. Once the hyperparameters are chosen, the impulse response of the system is computed as the minimum mean-square Bayes estimate given the observed input/output data (see e.g. [Pillonetto and De Nicolao, 2010], [Chen et al., 2012a]).

A key assumption in kernel-based methods is that the output data and the system admit a joint Gaussian description. Such an assumption does not hold with quantized data and we need to think of a different approach. In this paper we propose a solution based on Markov Chain Monte Carlo (MCMC) techniques [Gilks et al., 1996]. To this end, we define a target probability density; the estimate of the system can be obtained by drawing samples from it. Such a probability density is function of the following random variables: 1) the (unavailable) non-quantized output of the linear system, 2) the scaling hyperparameter of the

⋆ This work was supported by the European Research Council under the advanced grant LEARN, contract 267381, the Swedish Research Council under contract 621-2009-4017 and the European Union Seventh Framework Programme [FP7/2007-2013] under grant agreement no. 257462 HYCON2 Network of excellence, by the MIUR FIRB project RBFR12M3AC - Learning meets time: a new computational approach to learning in dynamic systems

2405-8963 © 2015, IFAC (International Federation of Automatic Control) Hosting by Elsevier Ltd. All rights reserved.
Peer review under responsibility of International Federation of Automatic Control.
10.1016/j.ifacol.2015.12.170

kernel, 3) the unknown measurement noise variance, 4) the impulse response of the system. The main contribution of this paper is to show how to design a Gibbs sampler [Gilks et al., 1996] by exploiting the knowledge of the conditional densities of the target distribution. The main advantage of the Gibbs sampler is that it does not require any rejection criterion of the generated samples and quickly converges to the target distribution. Note that MCMC-based approaches have recently gained popularity in system identification [Ninness and Henriksen, 2010], [Lindsten et al., 2013], [Bottegal et al., 2014].

The paper is organized as follows. In the next section, we introduce the problem of the identification of dynamic systems from quantized data. In Section 3, we give a Bayesian description of the variables entering the system. In Section 4, we describe the proposed method for identification. Section 5 shows some simulations to assess the performances of the proposed method. Some conclusions end the paper.

2. PROBLEM STATEMENT

We consider the following linear time-invariant BIBO stable output error system

$$z_t = (g * u)_t + v_t\,, \tag{1}$$

where $\{g_t\}$, $t \in \mathcal{T}$ is the impulse response characterizing the unknown system, which is fed by the input $\{u_t\}$, $t \in \mathcal{T}$. The set $\mathcal{T}$ corresponds to either $\mathbb{R}^+$ or $\mathbb{Z}^+$, depending on whether the system is continuous-time or discrete-time. The output z_t is corrupted by the additive white Gaussian noise v_t, which has zero mean and unknown variance σ^2, and measured at the time instants $t \in \mathcal{I}$. If the system is continuous-time, then $\mathcal{I}$ can represent any non-uniform sampling, whereas in the discrete-time case we shall consider $\mathcal{I} \equiv \mathbb{Z}^+$ (i.e., uniform sampling). For ease of exposition, in this paper we shall derive our algorithm in the discrete-time case only; the extension to the continuous-time is quite straightforward (see e.g. [Wahba, 1990], [Pillonetto and De Nicolao, 2010]). Actually, the

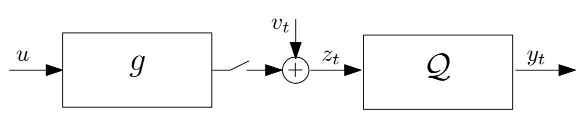

Fig. 1. *Block scheme of the system identification scenario.*

output z_t is not directly measurable, and only a quantized version is available, namely

$$y_t = \mathcal{Q}[z_t]\,, \tag{2}$$

where $\mathcal{Q}$ is a known map (our quantizer) of the type

$$\mathcal{Q}[x] = p_k \qquad \text{if } x \in (q_{k-1},\, q_k]\,, \tag{3}$$

with $p_k \in \{p_1, \ldots, p_Q\}$ and $q_k \in \{q_0, \ldots, q_Q\}$ being known (and typically $q_0 = -\infty$ and $q_Q = \infty$).

Remark 2.1. A particular and well-studied case is the binary quantizer, defined as

$$\mathcal{Q}[x] = \begin{cases} -1 & \text{if } x < C \\ 1 & \text{if } x \geq C \end{cases}\,. \tag{4}$$

It is well-known that a condition on the threshold to guarantee identifiability of the system is $C \neq 0$. In fact, when $C = 0$, the system can be determined up to a scaling factor [Godoy et al., 2011].

Without loss of generality, let us assume the system to be strictly causal, i.e. $g_0 = 0$. We assume that N input-output data samples $y_1, \ldots, y_N$, $u_0, \ldots, u_{N-1}$ are collected during an experiment. We formulate our system identification problem as the problem of estimating the impulse response g for n time instants, namely obtain $\{g_t\}_{t=1}^n$. Recall that, if n is sufficiently large, these samples can be used to approximate the dynamics of the systems with arbitrary accuracy [Ljung and Wahlberg, 1992]. Introducing the vector notation

$$g := \begin{bmatrix} g_1 \\ \vdots \\ g_n \end{bmatrix},\; y := \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix},\; z := \begin{bmatrix} z_1 \\ \vdots \\ z_N \end{bmatrix},\; v := \begin{bmatrix} v_1 \\ \vdots \\ v_N \end{bmatrix}$$

$$U = \begin{bmatrix} u_0 & 0 & & \cdots & 0 \\ u_1 & u_0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \ddots & \vdots \\ u_{N-2} & u_{N-3} & \cdots & u_{N-n+1} & 0 \\ u_{N-1} & u_{N-2} & \cdots & \cdots & u_{N-n} \end{bmatrix} \in \mathbb{R}^{N \times n},$$

the input-output relation for the available samples can be written

$$\begin{aligned} z &= Ug + v \\ y_t &= \mathcal{Q}[z_t] \quad , t = 1, \ldots, N \end{aligned}$$

so that our estimation problem can be cast in, say, a "linear regression plus quantization" form.

3. BAYESIAN MODELS FOR THE QUANTITIES OF INTEREST

3.1 Establishing a prior for the system

In this paper we cast the system identification problem into a Bayesian framework. Our starting point is the setting of a proper prior on g. Following a Gaussian regression approach [Rasmussen and Williams, 2006], we model g as a zero-mean Gaussian random vector, i.e. we assume the following probability density function for g:

$$p(g|\lambda,\, \beta) \sim \mathcal{N}(0,\, \lambda K_\beta)\,, \tag{5}$$

where K_β is a covariance matrix whose structure depends on the value of the *shaping hyperparameter* β and $\lambda \geq 0$ is the *scaling hyperparameter*. In this context, K_β is usually called a *kernel* and determines the properties of the realizations of g. In this paper, we choose K_β from the family of *stable spline kernels* [Pillonetto and De Nicolao, 2010], [Pillonetto et al., 2011]. Such kernels are specifically designed for system identification purposes and give clear advantages compared to other standard kernels [Bottegal and Pillonetto, 2013], [Pillonetto and De Nicolao, 2010] (e.g. the Gaussian kernel or the Laplacian kernel, see [Schölkopf and Smola, 2001]). In this paper we make use of the *first-order stable spline kernel* (or *TC kernel* in [Chen et al., 2012a]). It is defined as

$$\{K_\beta\}_{i,j} := \beta^{\max(i,j)} \quad , 0 \leq \beta < 1\,, \tag{6}$$

The above kernel is parameterized by β, which regulates the decay velocity of the generated impulse responses.

3.2 Bayesian description of the non-quantized output

Since we have assumed Gaussian distribution of the noise v, the joint distribution of the vectors z and g, given values