Preprints of the 16th IFAC Symposium on System Identification
The International Federation of Automatic Control
Brussels, Belgium. July 11-13, 2012

On Identification of Linear Systems with Quantized and Intermittent Information

Damián Marelli * **Minyue Fu** *,**

* *School of Electrical Engineering and Computer Science, University of Newcastle, University Drive, Callaghan, NSW 2308, Australia.*
** *Department of Control Science and Engineering, Zhejiang University, 388 Yuhangtang Road, Hangzhou, Zhejiang Province, 310058, P. R. China.*

Abstract: In this paper, we consider a number of technical problems associated with identification of linear systems using quantized and intermittent information. More specifically, we study asymptotic properties for identification of such systems, including strong consistency, asymptotic normality and asymptotic efficiency, and determine their relationship with quantization errors and the packet losses. Furthermore, we discuss how to design quantizers to improve these asymptotic properties. Some open problems will also be proposed.

Keywords: System identification, quantized signals, quantizer design, communication networks, asymptotic analysis, statistical analysis.

1. INTRODUCTION

This paper is concerned with identification of linear systems using quantized and intermittent information. This setting is motivated by networked control systems where sensor measurements and actuator signals are transmitted over a digital network. In contrast with traditional system identification problems, limited communication capacity imposed by network resource constraints means that it is essential to understand the tradeoff between communication resources and identification performance. More specifically, we need to know the adverse effect of undesirable properties of network communications, such as quantization errors, packet losses and transmission delays, on the key properties of system identification. Conversely, we need to know how much communication resources, expressed through communication network design parameters such as data rate, packet loss rate and latency distribution, are needed to deliver a given measure of system identification performance. In this paper, we will focus on data rate and packet loss rate constraints only, but it should be recognized that the key concepts and ideas discussed in this paper can be readily generalized to treat transmission delays as well.

System identification using quantized information has been a rich area in recent years; see e.g., Wang et al. [2003, 2010] and references therein. In Wang et al. [2003], a comprehensive treatment on quantized system identification is presented for single-input-single-output linear discrete-time stable systems. Using periodic input signals, the computational complexity and the impact of disturbances and unmodeled dynamics on the identification accuracy is also studied in Wang et al. [2003]. This work has been extended to various other system models such as rational models, Wiener systems and Hammerstein systems Wang et al. [2010]. Although these identification algorithms are shown to be optimal in the sense of asymptotically achieving the Cramr-Rao lower bound Wang and Yin [2007], the assumption on periodic inputs makes the identification algorithm inappropriate for tracking control applications. The periodic input assumption is dropped in Godoy et al. [2010], where an algorithm has been proposed for identifying moving-average (MA) models using quantized output, under the maximum likelihood criterion. In Marelli et al. [2011], identification of auto-regressive moving-average (ARMA) models has been considered using both quantized and intermittent measurements. Special care was taken for ARMA models because, in contrast with MA models, the loss of a measurement cannot be dealt with by simply removing the corresponding output error equation from the identification recursions.

The goal of this paper is to provide conditions to guarantee certain asymptotic properties for system identification based on quantized and intermittent data. These properties are *strong consistency* (whether the estimated parameters will converge to the true parameters with probability one), *asymptotic normality* (whether the estimation error will converge to a normal distribution and which will be the value of the error covariance) and *asymptotic efficiency* (whether the asymptotic error covariance matrix will approach the Cramr-Rao bound). For doing so, we aim to generalize our analysis in Marelli et al. [2011], which was done considering an ARMA model, and assuming that the input signal is known, and the output noise is white. To this end, we state in Lemma 1, a set of technical conditions on the data and the statistical model, which guarantee the above asymptotic properties. These conditions are given in a very general setting, so that they can be used to guarantee asymptotic properties for a wide range of system identification problems. We then show how to use this

[1] *Email addresses:* Damian.Marelli@newcastle.edu.au (Damin Marelli), Minyue.Fu@newcastle.edu.au (Minyue Fu).

result to derive the conditions obtained in Marelli et al. [2011] for the problem described above.

A unique feature in networked control systems is that the data rate constraint leaves the flexibility for the choice of quantizer. Thus, not only the quantization effect on system identification needs to be understood, a much more interesting and technically challenging problem is how to jointly design the quantizer and the corresponding estimation algorithm to minimize the estimation error. The main challenge lies in the fact that the unknown parameters are inaccessible to the design of an optimal quantizer. For example, to estimate θ under binary quantization of $y = \theta + v$, where v is a Gaussian random variable with zero mean, an optimal quantizer to minimize the mean square error (MSE) is to simply place the quantizer threshold at θ Ribeiro and Giannakis [2006]. However, such threshold selection is impractical since θ is not available at the estimator side. Instead, appropriate adaptive schemes are needed for the selection of quantization thresholds. Another major difficulty in quantizer design is the possible temporal correlation of the signal to be quantized. Although this difficulty exists in the problem of state estimation with quantized information (see, e.g., Sinopoli et al. [2004], Schenato et al. [2007]) and many effective algorithms exist for treating this problem, quantizer design for system identification is more involved because, in this case, the temporal correlation typically depends on the system parameters to be identified. A key question we are interested in is how to select the quantization thresholds and quantized values in a recursive manner so that the asymptotic properties of parameter estimation are optimized.

The rest of the paper is organized as follows: In Section 2, we formulate the technical problems with regard to the asymptotic properties described above. This formulation is done in a very general setting, and we describe three system identification problems which fall into this setting. Section 3 provides conditions to guarantee the asymptotic properties. These conditions are rather technical because they are stated in the most general setting. Using this result, in Section 4.1 we provide a complete solution for asymptotic conditions for the problem studied in Marelli et al. [2011]. These conditions are valid for a given quantization scheme and packet loss model. Section 4.2 then studies quantizer design. A dynamic quantizer is used, and quantization thresholds and quantized values are computed iteratively using the probability density function of the system measurements conditioned on the estimated parameters. This design turns out to be able to achieve asymptotic efficiency. Finally, in Section 5 we give concluding remarks, and we discus the key technical difficulties for using Lemma 1 in more challenging identification problems.

2. PROBLEM DESCRIPTION

Let $y(t) \in \mathbb{R}^q$, $t \in \mathbb{N}$, be a vector random process. Let p_θ, $\theta \in \mathcal{D} \subseteq \mathbb{R}^p$, denote a set of probability density functions (PDF's) for $y(t)$, $t \in \mathbb{N}$, and $\theta_\star \in \mathcal{D}$ denote the vector describing the 'true' statistics of $y(t)$, $t \in \mathbb{N}$. Let $\gamma(t)$, $t \in \mathbb{N}$, be a sequence of Bernoulli random variables with parameter $\lambda(t)$ (i.e., $\mathbb{P}(\gamma(t) = 1) = \lambda(t)$). The variables $\gamma(t)$ do not need to be statistically independent, but they are independent of $y(t)$. Let $\mathcal{Q}_t$, $t \in \mathbb{N}$, be a sequence of quantizers and $z(t) = \gamma(t)\mathcal{Q}_t[y(t)]$. Consider the maximum likelihood estimator

$$\hat{\theta}_N = \arg\max_{\theta \in \mathcal{D}} p_\theta\left(Z_N\right), \tag{1}$$

where $Z_N = \{z(1), \cdots, z(N)\}$. We aim to provide conditions for strong consistency, i.e.,

$$\hat{\theta}_N \overset{\text{w.p.1}}{\to} \theta_\star,$$

asymptotic normality, i.e.,

$$\sqrt{N}(\hat{\theta}_N - \theta_\star) \overset{\text{in dist.}}{\to} \mathcal{N}(0, \Sigma),$$

for some (positive-definite) matrix Σ, and asymptotic efficiency, i.e.,

$$\Sigma = \bar{\mathcal{I}}_Z^{-1}(\theta_\star),$$

where

$$\bar{\mathcal{I}}_Z(\theta) = \lim_{N \to \infty} \frac{1}{N} \mathcal{I}_{Z_N}(\theta),$$

with

$$\mathcal{I}_{Z_N}(\theta) = \mathcal{E}_\theta \left\{ \frac{\partial}{\partial \theta} \log p_\theta\left(Z_N\right) \frac{\partial}{\partial \theta'} \log p_\theta\left(Z_N\right) \right\}$$

denoting the Fisher information of Z_N, and $\mathcal{E}_\theta$ denoting expectation under the PDF p_θ, i.e., for all maps $f : \mathbb{R} \to \mathbb{R}$ and $\theta, \phi \in \mathcal{D}$, $\mathcal{E}_\theta\left\{f(p_\phi(Z_N))\right\} = \int p_\theta(Z_N) f(p_\phi(Z_N))\, dZ_N$.

Since the asymptotic covariance Σ depends on the quantization scheme $\mathcal{Q}_t$, $t \in \mathbb{N}$, we also aim to study the optimal scheme in the sense of minimizing the trace of Σ.

The setting stated above describes a general maximum likelihood estimation problem based on quantized data. We list below some system identification problems which fall into this general description. To this end, we assume that $g(q, \theta)$ is a parametric linear time-invariant single-input-single-output model, where $\theta \in \mathcal{D}$ is the parameter vector and q is the forward-shift operator, i.e., $qx(t) = x(t + 1)$.

(P1) **Known input and quantized white output:** In this case

$$\begin{aligned} y(t) &= g(q, \theta)u(t) + v(t), \\ z(t) &= \mathcal{Q}_t[y(t)], \end{aligned}$$

where the noise $v(t)$ is white and the input $u(t)$ is known. This scheme is the simplest to analyze because the quantized samples $z(t)$ are statistically independent.

(P2) **Unknown input and quantized output:** The scheme is similar to the one above, with the only difference in that $u(t)$ is unknown. Hence, $u(t)$ is modeled as a random process, and therefore the quantized samples $z(t)$ are correlated.

(P3) **Quantized input and output:** In this case

$$\begin{aligned} w(t) &= g(q, \theta)u(t) + v(t), \\ y(t) &= [w(t), u(t)], \\ z(t) &= \mathcal{Q}_t[y(t)], \end{aligned}$$

where $u(t)$ and $v(t)$ are random processes. In this case, the observations $z(t)$ are not only correlated but also vector-valued.

In Section 4.1 we use the result of Section 3 to provide conditions to guarantee the desired asymptotic properties for problem (P1), and in Section 5 we discuss the technical difficulties for doing the same with problems (P2) and (P3).

3. A SET OF SUFFICIENT CONDITIONS FOR THE DESIRED ASYMPTOTIC PROPERTIES

In this section we state a set of sufficient conditions for guaranteeing the asymptotic properties described in Section 2. Conditions of this kind are provided in Ljung [1999, Theorems 8.3 and 9.1] for the prediction (quadratic) error identification criterion. Since we are concerned with the maximum likelihood criterion, we derive our conditions using generic asymptotic results, like those in Gourieroux and Monfort [1996, Properties 24.2 and 24.16] and Gallant and White [1988, Theorems 3.3 and 5.1]. We state these conditions in a general setting, in the sense that they are valid for any maximum likelihood estimation problem, including (P1)-(P3) as listed earlier.

To study the problem described in Section *2*, we write (1) as

$$\hat{\theta}_N = \arg\max_{\theta\in\mathcal{D}} \Xi_N(\theta), \tag{2}$$

$$\Xi_N(\theta) = \frac{1}{N}\log p_\theta\left(Z_N\right). \tag{3}$$

Let $\bar{\Xi}(\theta) = \lim_{N\to\infty}\mathcal{E}_{\theta_\star}\left\{\Xi_N(\theta)\right\}$ and

$$\dot{\Xi}_N(\theta) = \frac{\partial}{\partial\theta}\Xi_N(\theta), \quad \bar{\dot{\Xi}}(\theta) = \lim_{N\to\infty}\mathcal{E}_{\theta_\star}\left\{\dot{\Xi}_N(\theta)\right\},$$

$$\ddot{\Xi}_N(\theta) = \frac{\partial^2}{\partial\theta\partial\theta'}\Xi_N(\theta), \; \bar{\ddot{\Xi}}(\theta) = \lim_{N\to\infty}\mathcal{E}_{\theta_\star}\left\{\ddot{\Xi}_N(\theta)\right\},$$

whenever the limits above exist. Then, we have the following result:

Lemma 1. Suppose;

(L1) $\mathcal{D}$ is compact (i.e., closed and bounded),
(L2) $\bar{\Xi}(\theta)$, $\bar{\dot{\Xi}}(\theta)$ and $\bar{\ddot{\Xi}}(\theta)$ exist, for all $\theta\in\mathcal{D}$,
(L3) $\arg\max_{\theta\in\mathcal{D}}\bar{\Xi}(\theta) = \{\theta_\star\}$, i.e., $\bar{\Xi}(\theta)$ is maximized at the unique value $\theta_\star$,
(L4) $\theta_\star\in\text{interior}(\mathcal{D})$,
(L5) $\sup_{N\in\mathbb{N}}\sup_{\theta,\phi\in\mathcal{D}}\left\|\ddot{\Xi}_N(\theta)-\ddot{\Xi}_N(\phi)\right\|\|\theta-\phi\|^{-1}\overset{\text{w.p.1}}{<}\infty$,
(L6) for each $\theta\in\mathcal{D}$, $\Xi_N(\theta)\overset{\text{w.p.1}}{\to}\bar{\Xi}(\theta)$, $\dot{\Xi}_N(\theta)\overset{\text{w.p.1}}{\to}\bar{\dot{\Xi}}(\theta)$ and $\ddot{\Xi}_N(\theta)\overset{\text{w.p.1}}{\to}\bar{\ddot{\Xi}}(\theta)$,
(L7) $\sqrt{N}\dot{\Xi}_N(\theta_\star)\overset{\text{in dist.}}{\to}\mathcal{N}(0,C)$, for some (positive definite) matrix C.

Then, strong consistency and asymptotic normality hold with

$$\Sigma = \bar{\mathcal{I}}_Z(\theta_\star)^{-1}C\bar{\mathcal{I}}_Z(\theta_\star)^{-1}. \tag{4}$$

Sketch of proof. We split the proof in steps:

(Step 1) Condition (L5) is required so that the convergences in (L6) are uniform on θ. To see this fix $i,j\in\{1,\cdots,p\}$ and define $f_N(\theta) = \left[\ddot{\Xi}_N(\theta)-\bar{\ddot{\Xi}}(\theta)\right]_{i,j}$, where $[\cdot]_{i,j}$ denotes the i,j-th entry of a matrix. For all $\theta,\phi\in\mathcal{D}$,

$$|f_N(\theta)-f_N(\phi)| \le \left|\left[\ddot{\Xi}_N(\theta)-\ddot{\Xi}_N(\phi)\right]_{i,j}\right| + \left|\left[\bar{\ddot{\Xi}}(\theta)-\bar{\ddot{\Xi}}(\phi)\right]_{i,j}\right|.$$

Then, from (L5) it is not difficult to verify that there exists $M>0$ such that

$$|f_N(\theta)-f_N(\phi)| \overset{\text{w.p.1}}{\le} M\|\theta-\phi\|.$$

Hence, from Davidson [1994, Theorem 21.10], $f_N(\theta)$ is strongly stochastically equicontinuous. Then, from Davidson [1994, Theorem 21.8], $[\ddot{\Xi}_N(\theta)]_{i,j}\overset{\text{w.p.1}}{\to}[\bar{\ddot{\Xi}}(\theta)]_{i,j}$, uniformly on θ, which shows the uniform convergence of $\ddot{\Xi}_N(\theta)$. The uniform convergences of $\Xi_N(\theta)$ and $\dot{\Xi}_N(\theta)$ then follow from Rudin [1976, Theorem 7.17].

(Step 2) Consider the set (in the underlying probability space), where $\dot{\Xi}_N(\theta)$ and $\ddot{\Xi}_N(\theta)$ converge uniformly on θ. Then, on this set, we have

$$\begin{aligned}\bar{\ddot{\Xi}}(\theta_\star) &= \lim_{N\to\infty}\mathcal{E}_{\theta_\star}\left\{\frac{\partial^2}{\partial\theta\partial\theta'}\Xi_N(\theta_\star)\right\}\\ &= \lim_{N\to\infty}\frac{\partial^2}{\partial\theta\partial\theta'}\Xi_N(\theta_\star)\\ &\overset{(a)}{=} \frac{\partial^2}{\partial\theta\partial\theta'}\lim_{N\to\infty}\Xi_N(\theta_\star)\\ &= \bar{\ddot{\Xi}}(\theta_\star)\\ &\overset{(b)}{<} 0,\end{aligned}$$

where (a) follows from two applications of Rudin [1976, Theorem 7.17], and (b) follows from (L3).

(Step 3) The existence of $\dot{\Xi}_N(\theta)$ implies that $\Xi_N(\theta)$ is continuous. This, together with (L1), (L3) and the uniform convergence of $\Xi_N(\theta)$, gives the conditions for strong convergence stated in Gourieroux and Monfort [1996, Property 24.2].

(Step 4) From condition (L5), it follows that $\ddot{\Xi}_N(\theta)$ is continuous. This, together with (L1), (L4), (L7), Step 2 and the uniform convergence of $\ddot{\Xi}_N(\theta)$, gives the conditions for asymptotic normality stated in Gourieroux and Monfort [1996, Property 24.16], with Σ given by (4).

Corollary 2. If, in addition to (L1)-(L5), $\sqrt{N}\dot{\Xi}_N(\theta_\star)\overset{\text{in dist.}}{\to}\mathcal{N}(0,\bar{\mathcal{I}}_Z(\theta_\star))$, then asymptotic efficiency also holds.

4. IDENTIFICATION OF AN ARMA MODEL WITH KNOWN INPUT AND QUANTIZED WHITE OUTPUT

In this section we study problem (P1) described in Section 2. We assume that

$$g(q,\theta) = \frac{b_0+\cdots+b_nq^{-n}}{1+a_1q^{-1}+\cdots+a_mq^{-m}},$$

where $\theta = [b_0,\cdots,b_n,a_1,\cdots,a_m]^T$, and define $x(t,\theta) = g(q,\theta)u(t)$. In Section 4.1 we use Lemma 1 to provide conditions to guarantee the asymptotic properties described in Section 2. Then, in Section 4.2 we use this result to derive the optimal quantization scheme in the sense of minimizing the asymptotic error covariance.

4.1 Asymptotic Properties

Definition 3. A time-varying quantizer is a sequence of maps $\mathcal{Q}_t : \mathbb{R} \to \mathcal{V}_t$, $t \in \mathbb{Z}$, from $\mathbb{R}$ to the set $\mathcal{V}_t = \{v_{t,1}, \cdots, v_{t,K_t}\} \subseteq \mathbb{R}$ of K_t quantization levels, for some given K_t. It is said to be *uniformly bounded* if there exists $M \geq 0$, such that, $|b_{t,k}| < M$, for all $t \in \mathbb{Z}$ and $k = 1, \cdots, K_t - 1$, where $b_{t,k}$ is defined by $[b_{t,k-1}, b_{t,k}] = \mathcal{Q}^{-1}[v_{t,k}]$.

Theorem 4. Suppose;

(T1) The set $\mathcal{D}$ is compact,
(T2) $\theta_\star \in \text{interior}(\mathcal{D})$,
(T3) the model $g(q, \theta)$ is such that its poles have magnitudes smaller than or equal to $1 - \epsilon$, for some $\epsilon > 0$,
(T4) the input $u(t)$ is such that
  (T4a) $\sup_{t \in \mathbb{N}} u(t) < \infty$,
  (T4b) $\lim_{N \to \infty} \frac{1}{N} \sum_{t=1}^{N} u(t)$ exists,
  (T4c) $\lim_{N \to \infty} \frac{1}{N} \sum_{t=1}^{N} \left(g(q,\theta)u(t) - g(q,\theta_\star)u(t)\right)^2 = 0$ holds for $\theta \in \mathcal{D}$, if and only if $\theta = \theta_\star$,
(T5) the quantizer $\mathcal{Q}_t$ is uniformly bounded,
(T6) the noise $v(t)$ is white and $v(t) \sim \mathcal{N}(0, \sigma^2)$,
(T7) the random process $\lambda(t)$ is ergodic, i.e.,

$$\frac{1}{N} \sum_{t=1}^{N} \lambda(t) \overset{\text{w.p.1}}{\to} \bar{\lambda} \triangleq \lim_{N \to \infty} \frac{1}{N} \sum_{t=1}^{N} \mathcal{E}\{\lambda(t)\}.$$

Then strong consistency, asymptotic normality and asymptotic efficiency hold with

$$\Sigma = \frac{\sigma^2}{\bar{\lambda}} \left(\lim_{N \to \infty} \frac{1}{N} \sum_{t=1}^{N} \mu(t) \psi(t, \theta_\star) \psi'(t, \theta_\star) \right)^{-1}, \quad (5)$$

$$\psi(t, \theta) = \frac{\partial}{\partial \theta} g(q, \theta) u(t), \quad (6)$$

$$\mu(t) = \frac{\bar{\sigma}^2(t)}{\sigma^2}, \quad (7)$$

$$\bar{\sigma}^2(t) = \mathcal{E}_{\theta_\star} \left\{ (\bar{y}(t, \theta_\star) - g(q, \theta) u(t))^2 \right\}, \quad (8)$$

$$\bar{y}(t, \theta) = \mathcal{E}_\theta \{ y(t) | z(t), \gamma(t) \}. \quad (9)$$

Sketch of proof. The proof consists in showing that Conditions (T1)-(T7) imply Conditions (L1)-(L7) of Lemma 1. To this end, notice that, since $z(1), z(2), \cdots$ are independent, we have

$$\Xi_N(\theta) = \frac{1}{N} \sum_{t=1}^{N} \xi(t, \theta), \quad (10)$$

where

$$\xi(t, \theta) = \log p_\theta(z(t)). \quad (11)$$

We split the proof in seven steps, after noting that Conditions (L1) and (L4) are the same as (T1) and (T2).

(Step 1) Using Cappé et al. [2005, Proposition 10.1.4], with some algebraic steps we can show that

$$\begin{aligned} \frac{\partial}{\partial \theta} \xi(t, \theta) &= \frac{\partial}{\partial \theta} \int \log p_\theta (z(t), y(t))\, p_\theta (y(t)|z(t))\, dy(t) \\ &= \frac{1}{\sigma^2} (\bar{y}(t, \theta) - x(t, \theta)) \psi(t, \theta), \end{aligned} \quad (12)$$

where

$$\psi(t, \theta) = \phi(q, \theta) u(t)$$
$$\phi(q, \theta) = \left[\frac{\Omega_n^T(q)}{A(q, \theta)}, \frac{q^{-1} B(q, \theta) \Omega_{m-1}^T(q)}{A^2(q, \theta)} \right]^T.$$

and $\Omega_n(q) = [1, q^{-1}, \cdots, q^{-n}]^T$. Using this, Condition (L2) follows from (T4b).

(Step 2) From (10)-(11), (L5) holds if

$$\sup_{N \in \mathbb{N}} \sup_{\theta, \phi \in \mathcal{D}} \left\| \ddot{\xi}(t, \theta) - \ddot{\xi}(t, \phi) \right\| \|\theta - \phi\|^{-1} \overset{\text{w.p.1}}{<} \infty, \quad (13)$$

where $\ddot{\xi}(t, \theta) = \frac{\partial^2}{\partial \theta \partial \theta'} \xi(t, \theta)$. Condition (13) in turn can be verified using (T3), (T4a), (T5) and (T6).

(Step 3) Since the PDF of $z(t)$ depends on θ only via $x(t, \theta)$, we define $\tilde{p}(z(t)|x)$ such that $p_\theta(z(t)) = \tilde{p}(z(t)|x(t, \theta))$. Let
$f_t(d) = \mathcal{E}_{\theta_\star} \{ \log \tilde{p}(z(t)|x(t, \theta_\star)) - \log \tilde{p}(z(t)|x(t, \theta_\star) + d) \}$.
Then, it is straightforward to verify that,

$$\arg\max_{\theta \in \mathcal{D}} \overline{\Xi}(\theta) = \arg\min_{\theta \in \mathcal{D}} H(\theta). \quad (14)$$

where

$$\begin{aligned} H(\theta) &= \lim_{N \to \infty} \mathcal{E}_{\theta_\star} \{ \Xi_N(\theta_\star) - \Xi_N(\theta) \} \\ &= \lim_{N \to \infty} \frac{1}{N} \sum_{t=1}^{N} f_t (x(t, \theta) - x(t, \theta_\star)). \end{aligned} \quad (15)$$

Now, in view of (T3) and (T4a), $|x(t, \theta) - x(t, \theta_\star)|$ is bounded. Then, from (T6), we can find $\epsilon > 0$ such that $f_t (x(t, \theta) - x(t, \theta_\star)) > \epsilon (x(t, \theta) - x(t, \theta_\star))^2$, for all t. Hence, (L3) follows from (T4c), (14) and (15).

(Step 4) Since $\xi(t, \theta)$ are independent, so are its derivatives with respect to θ. Hence, (L6) follows from Rajchman's strong law of large numbers Chung [2001, Theorem 5.1.2]. To this end, the uniform boundedness of second moments need to be verified. This follows from (T3), (T4a) and (T5).

(Step 5) From, (T3) and (T4a), for each $y \in \mathbb{R}^p$, it can be verified that $y' \frac{\partial}{\partial \theta} \xi(t, \theta_\star)$ satisfies Lyapunov's condition Klenke [2007, Definition 15.40]. Then, in view of Klenke [2007, Definition 15.40] and the Lindeberg-Feller central limit theorem Klenke [2007, Theorem 15.43], and since $\frac{\partial}{\partial \theta} \xi(t, \theta_\star)$ are independent and have zero mean,

$$\frac{1}{\sqrt{N}} \sum_{t=1}^{N} y' \frac{\partial}{\partial \theta} \xi(t, \theta_\star) \overset{\text{in dist.}}{\to} \mathcal{N}(0, y' C y).$$

where $C = \lim_{N \to \infty} \frac{1}{N} \sum_{t=1}^{N} \mathcal{E}_{\theta_\star} \left\{ \frac{\partial}{\partial \theta} \xi(t, \theta_\star) \frac{\partial}{\partial \theta'} \xi(t, \theta_\star) \right\}$. Then, (L7) follows from Davidson [1994, eq. (25.28)].

(Step 6) In view of the steps above, strong convergence and asymptotic normality hold. To see that asymptotic efficiency also does, notice that, since $\frac{\partial}{\partial \theta} \xi(t, \theta_\star)$ are independent,

$$\begin{aligned} C &= \lim_{N \to \infty} \frac{1}{N} \mathcal{E}_{\theta_\star} \left\{ \sum_{t=1}^{N} \frac{\partial}{\partial \theta} \xi(t, \theta_\star) \sum_{s=1}^{N} \frac{\partial}{\partial \theta'} \xi(s, \theta_\star) \right\} \\ &= \lim_{N \to \infty} \frac{1}{N} \mathcal{E}_{\theta_\star} \left\{ \frac{\partial}{\partial \theta} \log p_{\theta_\star}(Z_N) \frac{\partial}{\partial \theta'} \log p_{\theta_\star}(Z_N) \right\} \\ &= \bar{\mathcal{I}}_Z(\theta_\star). \end{aligned}$$

Hence, $\Sigma = \bar{\mathcal{I}}_Z^{-1}(\theta_\star)$.

(Step 7) Finally, (5)-(9) follow from (12).

The asymptotic covariance given in the theorem above depends on the known input $u(t)$. The next corollary considers the case where the known input is a realization of a random process.

Corollary 5. If the conditions in Theorem 4 hold, $u(t)$ is the realization of a wide-sense stationary and ergodic random process, then the result of Theorem 4 holds with

$$\Sigma \overset{\text{w.p.1}}{=} \frac{\sigma^2}{\bar{\lambda}} \mathcal{E}_u \left\{ \mu(t) \psi(t, \theta_\star) \psi^T(t, \theta_\star) \right\}^{-1},$$

where $\mathcal{E}_u\{\cdot\}$ denotes the expectation taken with respect to $u(t)$.

Proof. See Marelli et al. [2011].

4.2 Optimum Quantization Scheme

The result of the Theorem 4 is valid for any uniformly bounded quantizer. This section studies how to choose the quantizer to minimize the asymptotic covariance Σ.

We have that

$$\begin{aligned} \bar{\sigma}^2(t) &= \mathcal{E}\left\{ \left(\bar{\mathcal{Q}}_t\left[y(t)\right] - x(t,\theta_\star) \right)^2 \right\} \\ &= \mathcal{E}\left\{ \left(\tilde{\mathcal{Q}}_t\left[y(t) - x(t,\theta_\star)\right] \right)^2 \right\}, \end{aligned} \quad (16)$$

where $\bar{\mathcal{Q}}_t$ and $\tilde{\mathcal{Q}}_t$ are the quantizers defined by

$$\begin{aligned} \bar{\mathcal{Q}}_t\left[y\right] &= \mathcal{E}\left\{ y | y \in \mathcal{Q}_t^{-1}[v_k] \right\}, \text{ if } y \in \mathcal{Q}_t^{-1}[v_k] \\ \tilde{\mathcal{Q}}_t\left[y\right] &= \bar{\mathcal{Q}}_t\left[y + x(t,\theta_\star)\right] - x(t,\theta_\star), \end{aligned}$$

and for each $t \in \mathbb{N}$ and $k = 1, \cdots, K_t$, $\mathcal{Q}_t^{-1}[v_k] = \{y \in \mathbb{R}^q : \mathcal{Q}_t[y] = v_k\}$. From (16), $\bar{\sigma}^2(t)$ can be interpreted as the power of the quantized version of $\tilde{y}(t) = y(t) - x(t,\theta_\star)$, obtained from the time-varying quantizer $\tilde{\mathcal{Q}}_t$.

It is clear that Σ is minimized if $\mathcal{Q}_t$ is chosen so that $\bar{\sigma}^2(t)$ is minimized for each $t \in \mathbb{N}$. In view of Gersho and Gray [1991, eq. (6.2.14)], this is equivalent to minimizing the quantization error of $\tilde{y}(t)$. This is achieved by choosing $\tilde{\mathcal{Q}}_t$ to be a Lloyd-Max quantizer for the density of $\tilde{y}(t)$, and choosing

$$\mathcal{Q}_t[y] = \tilde{\mathcal{Q}}_t\left[y(t) - x(t,\theta_\star)\right] + x(t,\theta_\star). \quad (17)$$

Moreover, if this results in $\bar{\sigma}^2(t) = \bar{\sigma}^2$ being independent of t, then,

$$\Sigma = \frac{\sigma^2}{\bar{\lambda}\mu} \left(\lim_{N\to\infty} \frac{1}{N} \sum_{t=1}^{N} \psi(t,\theta_\star)\psi'(t,\theta_\star) \right)^{-1}, \quad (18)$$

with $\mu = \bar{\sigma}^2/\sigma^2$.

Remark 6. The equation above differs from the classical result of system identification Ljung [1999] in the factor $1/\bar{\lambda}\mu$, which accounts for the effect of the quantizer and packet drops. The coefficient μ states the inverse ratio between the power of the signal $\tilde{y}(t)$ and the power of the signal obtained after quantizing $\tilde{y}(t)$ using an optimum Lloyd-Max quantizer. Hence, μ monotonically increases and tends to 1 as the number of quantization levels tends to infinity.

In practice, it is not possible to choose the quantizer $\mathcal{Q}_t$ as in (17), because this requires the knowledge of the true parameters $\theta_\star$. A practical workaround is to replace $\theta_\star$by the previous estimate $\hat{\theta}_{t-1}$, i.e.,

$$\mathcal{Q}_t[y] = \tilde{\mathcal{Q}}_t\left[y(t) - x(t,\hat{\theta}_{t-t})\right] + x(t,\hat{\theta}_{t-1}).$$

Assuming that the arrival of each packet is acknowledged by the receiver, $\hat{\theta}_{t-1}$ is known at both ends. A question that naturally arises then is whether the minimum value (18) of the asymptotic covariance matrix can still be achieved in this case. The answer turns out to be positive. The proof of this claim can be found in Marelli et al. [2011].

5. CONCLUDING REMARKS

We studied conditions to guarantee asymptotic properties for identification of linear systems using quantized and intermittent information. More specifically, we studied strong consistency, asymptotic normality and asymptotic efficiency. In Lemma 1 we stated a number of technical conditions in a rather general setting. We then showed how to use this lemma to provide conditions to guarantee asymptotic properties for problem (P1) in Section 2, and we discussed how to design quantizers to improve the asymptotic estimation error covariance.

A key property of problem (P1) is that the quantized samples $z(t)$ are statistically independent. This permits writing $\Xi_N(\theta)$ as a Cesro summation of $\xi(t,\theta)$, as in (10)-(11). Then, we can replace (L5) by (13), i.e., a condition on the joint PDF $p_\theta(Z_N)$ by one on the PDF $p_\theta(z(t))$ of each sample. Also, the independence of $z(t)$ results in the independence of $\xi(t,\theta)$ and its derivatives. Hence, (L6) can be guaranteed using a version of the strong law of large numbers for independent variables. Similarly, (L7) can be guaranteed using a version of the central limit theorem for independent variables. These versions essentially require that samples have uniformly bounded moments of certain order. In problems (P2) and (P3), the samples $z(t)$ are not independent. Hence, the remarks above do not apply, and more technical effort is required to address their analysis. This is an interesting topic for future research.

REFERENCES

D. Cappé, E. Moulines, and T. Rydén. *Inference in Hidden Markov Models.* Springer, 2005.

Kai Lai Chung. *A course in probability theory.* Academic Press Inc., San Diego, CA, third edition, 2001.

James Davidson. *Stochastic Limit Theory: An Introduction for Econometricicans (Advanced Texts in Econometrics).* Oxford University Press, USA, 12 1994.

A.R. Gallant and H. White. *A unified theory of estimation and inference for nonlinear dynamic models.* Basil Blackwell New York, 1988.

A. Gersho and R.M. Gray. *Vector quantization and signal compression.* Kluwer Academic Publishers Norwell, MA, USA, 1991.

Boris Godoy, Graham Goodwin, Juan Agüero, Damián Marelli, and Torbjörn Wigren. Identification of FIR systems having quantized output data. *to appear in Automatica*, 2010.

Christian Gourieroux and Alain Monfort. *Statistics and Econometric Models: Volume 2.* Cambridge University Press, 1996.

Achim Klenke. *Probability Theory: A Comprehensive Course (Universitext)*. Springer, 12 2007.

Lennart Ljung. *System Identification: Theory for the User*. Prentice Hall, Upper Saddle River, NJ, second edition, 1999.

Damián Marelli, Keyou You, and Minyue Fu. Identification of arma models using intermittent and quantized output observations. submitted to Automatica, 2011.

A. Ribeiro and G. Giannakis. Bandwidth-Constrained Distributed Estimation for Wireless Sensor Networks-Part I:Gaussian Case. *IEEE Transactions on Signal Processing*, 54(3):1131–1143, 2006.

W. Rudin. *Principles of mathematical analysis*. McGraw-Hill, New York, 1976.

L. Schenato, B. Sinopoli, M. Franceschetti, K. Poolla, and S. Sastry. Foundations of control and estimation over lossy networks. *Proceedings of the IEEE*, 95(1):163–187, 2007.

B. Sinopoli, L. Schenato, M. Franceschetti, K. Poolla, M. Jordan, and S. Sastry. Kalman filtering with intermittent observations. *IEEE Transactions on Automatic Control*, 49(9):1453–1464, 2004.

L. Wang and G. Yin. Asymptotically efficient parameter estimation using quantized output observation. *Automatica*, 43(7):1178–1191, 2007.

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

L. Wang, G. Yin, J. Zhang, and Y. Zhao. *System identification with quantized observations*. Birkhäuser Boston, 2010.