Automatica 49 (2013) 360–369

Contents lists available at SciVerse ScienceDirect

Automatica

journal homepage: www.elsevier.com/locate/automatica

Identification of ARMA models using intermittent and quantized output observations☆

Damián Marelli [a,1], Keyou You [b], Minyue Fu [a,c]

[a] School of Electrical Engineering and Computer Science, University of Newcastle, University Drive, Callaghan, NSW 2308, Australia

[b] Department of Automation, Tsinghua University, Beijing, 100084, PR China

[c] Department of Control Science and Engineering, Zhejiang University, 388 Yuhangtang Road, Hangzhou, Zhejiang Province, 310058, PR China

ARTICLE INFO

Article history:
Received 11 April 2011
Received in revised form
28 May 2012
Accepted 4 September 2012
Available online 8 December 2012

Keywords:
Identification methods
Network-based computing systems
ARMA model
Finite-level quantization
Packet dropout

ABSTRACT

This paper studies system identification of ARMA models whose outputs are subject to finite-level quantization and random packet dropouts. Using the maximum likelihood criterion, we propose a recursive identification algorithm, which we show to be strongly consistent and asymptotically normal. We also propose a simple adaptive quantization scheme, which asymptotically achieves the minimum parameter estimation error covariance. The joint effect of finite-level quantization and random packet dropouts on identification accuracy are exactly quantified. The theoretical results are verified by simulations.

1. Introduction

System identification of plants with quantized observations is significant in understanding the modeling capacity for systems with limited sensor information, and the trade off between communication resources and identification performance (Wang, Zhang, & Yin, 2003). This work is concerned with the identification of autoregressive moving average (ARMA) models whose quantized outputs must be communicated through a digital noisy channel. A motivating example is given by a sensor and an estimator communicating over wireless channels with limited resources in terms of bandwidth and transmission power. By modeling the packet dropout process as an independent and identically distributed (i.i.d.) Bernoulli process, this paper aims to quantify the joint effect of finite-level quantization and packet dropouts on the identification accuracy of ARMA models. The key difference of quantized identification from the classical identification problem is that the estimator is no longer able to access the original analog amplitude (unquantized) observations. Especially under aggressive quantization, the discrete-valued observations supply limited information on system outputs, and hence introduce new challenges in system modeling, identification and control. In addition, channel errors, e.g., packet dropouts, further induce information loss that influences identification accuracy.

Recently, research on quantized identification/estimation constitutes a vast body of literature, see e.g., Wang, Yin, Zhang, and Zhao (2010); Xiao, Ribeiro, Luo, and Giannakis (2006) and references therein. In Xiao et al. (2006) and the references therein, various quantized estimation algorithms are developed in the context of wireless sensor networks. In Wang et al. (2010), a comprehensive treatment on quantized system identification is presented for single-input-single-output linear discrete time-invariant stable systems. Based on periodic inputs, they study the computational complexity and the impact of disturbances and unmodeled dynamics on the identification accuracy. In the same spirit, various models such as rational models, Wiener systems and Hammerstein systems have been studied as well. Although their identification algorithms are shown to be optimal in the sense of asymptotically achieving the Cramer–Rao lower bound (Wang et al., 2010), the assumption on periodic inputs makes the identification algorithm inappropriate for tracking control applications. Moreover, input design is of essential importance in system identification to

☆ The material in this paper was partially presented at the 16th IFAC Symposium on System Identification (SYSID 2012), July 11–13, 2012, Brussels, Belgium. This paper was recommended for publication in revised form by Associate Editor Johan Schoukens under the direction of Editor Torsten Söderström.

E-mail addresses: Damian.Marelli@newcastle.edu.au (D. Marelli), youky@tsinghua.edu.cn (K. You), Minyue.Fu@newcastle.edu.au (M. Fu).

[1] Tel.: +61 2 4921 7845; fax: +61 2 49216993.

0005-1098/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.automatica.2012.11.020

provide sufficient probing capacities to guarantee the existence of a consistent estimator. The periodic input assumption is dropped in Godoy, Goodwin, Aguero, Marelli, and Wigren (2011), and an algorithm has been proposed for identifying moving average (MA) models using quantized output, under the maximum likelihood criterion. The present work extends the work in Godoy et al. (2011) in three ways: (1) We study the identification of ARMA models using not only quantized but also intermittent output measurements. A difficulty in treating ARMA models with lost measurements is that, as opposed to MA models, it is not possible to simply remove the corresponding output error equations from the identification recursions. Hence, in order to use standard estimation algorithms, these missing measurements need to be properly estimated. By doing so, the problem can be stated in an iterative weighted linear least squares form. (2) We carry out an asymptotic stochastic analysis where we provide conditions for strong consistency and asymptotic normality of the estimated parameters. As in Godoy et al. (2011), we also drop the periodic assumption on the input signal. We show that, to ensure consistency, input signals are required to be *persistently exciting* (of a certain order) (Lennart, 1999). (3) We propose a simple adaptive quantization scheme that asymptotically achieves the minimum parameter estimation error covariance. Notice that the quantizer design is not addressed in Godoy et al. (2011), where it is assumed to be static and given as a prior.

The fundamental problem of system identification with quantized observations is the joint design of 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 & Giannakis, 2006). However, such threshold selection is impractical since θ is not available at the estimator side. It is known that the estimation performance is very sensitive to the choice of the quantizer threshold. Motivated by this, an interesting quantizer threshold selection scheme is proposed in Papadopoulos, Wornell, and Oppenheim (2002). It consists of periodically applying a set of thresholds with equal frequencies, hoping that some thresholds are close to the unknown parameter. To asymptotically approach the minimum MSE, Fang and Li (2008) constructs an adaptive quantization involving delta modulation with variable stepsize. However, the on-line optimal stepsize is obtained through a maximum likelihood estimation process lacking a recursive form. This paper proposes a simple adaptive quantizer and the corresponding recursive identification algorithm to asymptotically approach the minimum parameter estimation error covariance. Our scheme exploits the fact that quantizing innovations requires fewer bits than quantizing observations; see Fu and de Souza (2009); You, Xie, Sun, and Xiao (2008).

Another detrimental factor impairing the identification performance is the dropouts of quantized observations. To the best of our knowledge, there is no work to date to quantify the joint effect of finite-level quantization and packet dropouts, on identification performance. As in Sinopoli et al. (2004), we model the packet dropout process by an i.i.d. process. Then, the basic problem is how to deal with the loss of a packet. For the state estimation of dynamical systems with intermittent un-quantized observations, it is well understood that the optimality of the Kalman filter still holds (Sinopoli et al., 2004). Although some stability conditions are derived for the corresponding estimation error covariance matrices (Sinopoli et al., 2004; You, Fu, & Xie, 2011), the performance degeneration due to packet dropout is still unclear. Our framework is closely related to that of You and Xie (2010), which focuses on the stabilizability problem by accounting for the joint effect of finite-level quantization and packet dropouts. In this paper we develop an asymptotically optimal algorithm for system identification, which shows that the joint effect of finite-level quantization and packet dropouts on the identification performance can be exactly quantified by the packet dropout rate and the number of quantization levels.

The rest of the paper is organized as follows. In Section 2 we describe the system identification problem. In Section 3 we describe the proposed identification method, using the maximum likelihood criterion. In Section 4 we provide conditions for strong consistency and asymptotic normality of the estimated parameter vector. In Section 5 we study the optimum quantization scheme that minimizes the covariance of the parameter estimation error, and we propose an adaptive quantization scheme that asymptotically achieves this minimum covariance. In Section 6 we present some simulation results. Concluding remarks are given in Section 7.

2. Problem description

Consider the following ARMA model

$$x(t) = \frac{B(q)}{A(q)}u(t), \qquad y(t) = x(t) + w(t),$$
$$z(t) = \gamma_t \mathcal{Q}_t[y(t)],$$

where the input $u(t)$ is modeled by either a deterministic signal or a random process and $w(t)$ is a sequence of independent and identically distributed (i.i.d.) samples with normal distribution $\mathcal{N}(0, \sigma^2)$. The noisy output $y(t)$ is quantized by a time-varying K-level scalar quantizer $\mathcal{Q}_t : \mathbb{R} \to \{v_{t,1}, \ldots, v_{t,K}\}$, $t \in \mathbb{Z}$ (which accounts for non-stationary quantization schemes), defined by the quantization intervals $[b_{t,k-1}, b_{t,k}] = \mathcal{Q}_t^{-1}[v_{t,k}]$, $k = 1, \ldots, K$, with $b_{t,0} = -\infty$ and $b_{t,K} = \infty$, for all $t \in \mathbb{Z}$. The quantized values are then transmitted through an unreliable communication network whose packet dropouts are modeled by a sequence γ_t of i.i.d. Bernoulli random variables with parameter λ (i.e., $\mathbb{P}(\gamma_t = 1) = \lambda$).

To simplify the notation, we combine the effects of quantization and packet dropouts in a single quantizer $\breve{\mathcal{Q}}_t$ defined by

$$\breve{\mathcal{Q}}_t[y] = \begin{cases} \mathcal{Q}_t[y], & \gamma_t = 1; \\ 0, & \gamma_t = 0. \end{cases}$$

When a packet is lost, $\breve{\mathcal{Q}}_t$ can be interpreted as a quantizer with a single quantization interval spanning $[-\infty, \infty]$, which is mapped to zero. Denote $U_N = \{u(t) : t = 1, \ldots, N\}$, $Z_N = \{z(t) : t = 1, \ldots, N\}$, $A(q) = 1 + a_1 q^{-1} + \cdots + a_m q^{-m}$, $B(q) = b_0 + \cdots + b_n q^{-n}$ and $\theta_\star = [b_0, \ldots, b_n, a_1, \ldots, a_m]^T$ (the superscript ^{T} denotes vector/matrix transpose). Let $r = m + n + 1$. For each $\theta \in \mathbb{R}^r$, we denote the parametric versions of $A(q)$ and $B(q)$ by $A(q, \theta)$ and $B(q, \theta)$, respectively. Denote the probability density function (PDF) of $w(t)$ by $f_W(w) = (2\pi\sigma^2)^{-1/2} \exp(-\frac{w^2}{2\sigma^2})$, and its cumulative distribution function by $F_W(w) = \int_{-\infty}^{w} f_W(\tilde{w})\, d\tilde{w}$.

Our goal is to estimate $\theta_\star$, given the knowledge of U_N and Z_N. We do so using the maximum likelihood (ML) criterion.

3. Maximum likelihood estimation

Using the ML criterion, an estimate $\hat{\theta}_N$ up to sample time N is obtained by[2]

$$\hat{\theta}_N \in \arg\max_\theta \; p_\theta(Z_N | U_N) = \arg\max_\theta \; l(\theta | U_N, Z_N), \tag{1}$$

[2] Notice that, for any function $f(\theta)$, the symbol $\arg\max_\theta f(\theta)$ denotes a set whose elements all maximize $f(\theta)$.

where p_θ denotes the probability distribution given θ, and

$$l(\theta|U_N, Z_N) = \log p_\theta\,(Z_N|U_N) \tag{2}$$

is the log-likelihood function of θ, given the knowledge of U_N and Z_N.

In Section 3.1 we propose an on-line algorithm for solving (1), based on the expectation maximization (EM) method. The advantage of this method is that we can derive an algorithm which does not require an initialization. However, the EM method is known to suffer from a slow convergence rate (Cappé, Moulines, & Rydén, 2005, pp. 358–359). Toward this end, the estimate obtained after a few EM iterations can be used to initialize a gradient search method (Fletcher, 1987). Thus, we propose in Section 3.2 an on-line algorithm based on the quasi-Newton method. The purpose of using the EM method is to provide a good initial estimate for the quasi-Newton method.

Since the input signal U_N is always known, we remove it from the list of conditioning variables in expressions like (1) and (2) for notational simplicity.

3.1. EM-based on-line estimation method

Before deriving the required method, we give a brief intuitive explanation for the EM method. For a detailed presentation see McLachlan and Krishnan (2008). A drawback of (1) is that $\log p_\theta\,(Z_N)$ is difficult to maximize with respect to θ due to the induction of a nonlinear quantizer. Suppose that the output $Y_N = \{y(t) : t = 1, \ldots, N\}$ before the quantizer was known. Note that it would be much easier to maximize $\log p_\theta\,(Y_N)$. Since Y_N is unavailable, we replace $\log p_\theta\,(Z_N)$ by the average of $\log p_\theta\,(Z_N, Y_N) = \log p_\theta\,(Z_N|Y_N) + \log p_\theta\,(Y_N)$ over all possible values of Y_N. To do this average we use the conditional distribution $p_{\hat\theta}(Y_N|Z_N)$ of Y_N given the observations Z_N, and some previous estimate $\hat\theta$ of θ. This leads to the EM method, which solves the ML problem (1) using the following iterative procedure:

$$\hat\theta_N^{(i)} \in \arg\max_\theta\ Q_N(\theta, \hat\theta_N^{(i-1)}), \tag{3}$$

$$Q_N(\theta, \hat\theta) = \int \log p_\theta\,(Z_N, Y_N)\, p_{\hat\theta}\,(Y_N|Z_N)\, dY_N. \tag{4}$$

The iterations (3)–(4) permit computing $\hat\theta_N$, for a fixed N. To obtain an on-line algorithm, we can compute one iteration for each new available sample. Doing so we obtain the following iterative algorithm:

$$\hat\theta_N \in \arg\max_\theta\ Q_N(\theta, \hat\theta_{N-1}). \tag{5}$$

Our next step is to find a closed-form expression for each iteration in (5). To this end, we have the following result.

Lemma 1. *The function $Q_N(\cdot,\cdot)$ in (4) is given by*

$$\begin{aligned} Q_N(\theta, \hat\theta) = &-\frac{N}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{t=1}^{N}(\bar y(t,\hat\theta) - x(t,\theta))^2 \\ &- \frac{1}{2\sigma^2}\sum_{t=1}^{N}(\overline{y^2}(t,\hat\theta) - \bar y^2(t,\hat\theta)), \end{aligned}$$

where $\overline{y^2}(t,\theta) = \mathcal{E}_\theta\left\{y^2(t)|z(t)\right\}$ and[3]

$$\bar y(t,\theta) = \mathcal{E}_\theta\left\{y(t)|z(t)\right\}. \tag{6}$$

[3] If a and b are random variables, and $f(\cdot)$ is a function, the conditional expectation $\mathcal{E}_\theta\{f(a)|b\}$ of $f(a)$ given b is defined by $\mathcal{E}_\theta\{f(a)|b\} = \int f(a)p_\theta(a|b)da$.

Remark 2. Notice that (6) can be computed by

$$\begin{aligned} \bar y(t,\theta) &= \int_{\breve{\mathcal{Q}}_t^{-1}[z(t)]} y(t)p_\theta(y(t))dy(t) \\ &= \frac{1}{2}\left[\operatorname{erf}\left(\frac{b(t)-x(t,\theta)}{\sqrt{2\sigma^2}}\right) - \operatorname{erf}\left(\frac{a(t)-x(t,\theta)}{\sqrt{2\sigma^2}}\right)\right], \end{aligned}$$

where $\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}dt$ denotes the error function and $[a(t), b(t)] = \breve{\mathcal{Q}}_t^{-1}[z(t)]$ denotes the quantization cell corresponding to $z(t)$.

Combining (5) with Lemma 1, we obtain that

$$\hat\theta_N \in \arg\min_\theta \sum_{t=1}^{N}(\bar y(t,\hat\theta_{N-1}) - x(t,\theta))^2. \tag{7}$$

The procedure in (7) requires tuning the ARMA model parameters θ to fit $\bar y(t,\hat\theta_{N-1})$ at each iteration. For a fixed N, a number of recursive formulas are available for doing so Pintelon, Guillaume, Rolain, Schoukens, and Van hamme (1994). In order to obtain an algorithm which does not require initialization, we use the iterative weighted linear least squares algorithm (Pintelon et al., 1994; Steiglitz & McBride, 1965). Again, to obtain an on-line algorithm we compute one iteration for each new available sample. This results in the following iterations, which can be initialized by choosing $\hat\theta_0$ so that $B(q,\hat\theta_0) = A(q,\hat\theta_0) = 1$ and

$$\hat\theta_N \in \arg\min_\theta \sum_{t=1}^{N}\left(A(q,\theta)\frac{\bar y(t,\hat\theta_{N-1})}{A(q,\hat\theta_{N-1})} - B(q,\theta)\frac{u(t)}{A(q,\hat\theta_{N-1})}\right)^2.$$

Then, $\hat\theta_N$ can be computed as follows:

$$\begin{aligned} \hat\theta_N = &\left(\sum_{t=1}^{N}\tilde\phi(t,\hat\theta_{N-1})\tilde\phi^T(t,\hat\theta_{N-1})\right)^\dagger \\ &\times\left(\sum_{t=1}^{N}\tilde\phi(t,\hat\theta_{N-1})\tilde y(t,\hat\theta_{N-1})\right), \end{aligned} \tag{8}$$

where the superscript † denotes the Moore–Penrose pseudoinverse (Ben-Israel & Greville, 2003), and

$$\tilde y(t,\theta) = \frac{\bar y(t,\theta)}{A(q,\theta)}, \tag{9}$$

$$\begin{aligned} \tilde\phi(t,\theta) = &\frac{1}{A(q,\theta)}[u(t),\ldots,u(t-n), \\ &-\bar y(t-1,\theta),\ldots,-\bar y(t-m,\theta)]^T. \end{aligned} \tag{10}$$

By following the steps in Lennart (1999, Sec. 11), (8) can be written in a recursive form:

$$\hat\theta_N = \hat\theta_{N-1} + L_N\left(\tilde y(N,\hat\theta_{N-1}) - \tilde\phi^T(N,\hat\theta_{N-1})\hat\theta_{N-1}\right), \tag{11}$$

with

$$L_N = \frac{P_{N-1}\tilde\phi(N,\hat\theta_{N-1})}{1+\tilde\phi^T(N,\hat\theta_{N-1})P_{N-1}\tilde\phi(N,\hat\theta_{N-1})}, \tag{12}$$

$$P_N = P_{N-1} - \frac{P_{N-1}\tilde\phi(N,\hat\theta_{N-1})\tilde\phi^T(N,\hat\theta_{N-1})P_{N-1}}{1+\tilde\phi^T(N,\hat\theta_{N-1})P_{N-1}\tilde\phi(N,\hat\theta_{N-1})}. \tag{13}$$

3.2. Quasi-Newton-based on-line estimation method

For a fixed N, the ML problem (1) can be solved using a gradient search algorithm. For this purpose we use a quasi-Newton method. As above, to obtain an on-line algorithm we compute one quasi-Newton iteration for each new available sample. This gives the

following iterations:

$$\hat{\theta}_{N+1} = \hat{\theta}_N - \mu_N T_N g_N, \tag{14}$$

where the scalar μ_N denotes the stepsize at iteration N, the matrix T_N denotes an approximation of the inverse of the Hessian of $l(\theta|Z_N)$ at $\hat{\theta}_N$, and the vector g_N denotes the gradient of $l(\theta|Z_N)$ at $\hat{\theta}_N$, i.e.,

$$g_N = \frac{\partial}{\partial\theta} l(\theta|Z_N)\Big|_{\hat{\theta}_N}. \tag{15}$$

We choose to compute T_N using the Broyden–Fletcher–Goldfarb–Shanno (BFGS) formula (Fletcher, 1987), which is initialized by $T_{N_0} = I$ (N_0 denotes the first sample after the EM iterations switch to quasi-Newton iterations) and proceed as follows:

$$T_{N+1} = T_N + \left(1 + \frac{q_N^T T_N q_N}{s_N^T q_N}\right)\frac{s_N s_N^T}{s_N^T q_N} - \frac{s_N q_N^T T_N + T_N q_N s_N^T}{s_N^T q_N}, \tag{16}$$

with $s_N = \hat{\theta}_{N+1} - \hat{\theta}_N$ and $q_N = g_{N+1} - g_N$.

Also, the stepsize parameter μ_N is obtained from a line search algorithm. We use the backtracking algorithm described in Boyd and Vandenberghe (2004), which is formed *by sub-iterations* of the *main iterations* (14). Let $\alpha = 0.01$ and $\beta = 0.5$. At sub-iteration i, the stepsize is updated using $\mu_{N,i} = \beta\mu_{N,i-1}$, starting from the initial value $\mu_{N,1} = 1$. The sub-iterations are stopped when

$$l(\hat{\theta}_N - \mu_{N,i} T_N g_N|Z_N) < l(\hat{\theta}_N|Z_N) + \alpha\mu_{N,i} g_N^T s_N. \tag{17}$$

To implement the iterations (14), we need to provide expressions for $l(\theta|Z_N)$ and its gradient. Since $w(t)$ is a sequence of independent random variables, we have that

$$l(\theta|Z_N) = \sum_{t=1}^{N} \log p_\theta\left(z(t)\right). \tag{18}$$

Lemma 3. *The gradient of the log-likelihood function $l(\theta|Z_N)$ is given by*

$$\frac{\partial}{\partial\theta} l(\theta|Z_N) = \frac{1}{\sigma^2}\sum_{t=1}^{N}\left(\bar{y}(t,\theta) - x(t,\theta)\right)\dot{x}(t,\theta),$$

where

$$\dot{x}(t,\theta) \triangleq \frac{\partial}{\partial\theta} x(t,\theta) = \phi(q,\theta)u(t), \tag{19}$$

$$\phi(q,\theta) = \left[\frac{\Omega_n^T(q)}{A(q,\theta)}, \frac{q^{-1}B(q,\theta)\Omega_{m-1}^T(q)}{A^2(q,\theta)}\right]^T, \tag{20}$$

and $\Omega_n(q) = [1, q^{-1}, \ldots, q^{-n}]^T$.

3.3. Summary of the identification algorithm

The identification algorithm described above can be summarized in this subsection. Choose an initial estimate $\hat{\theta}_0$ so that $B(q,\hat{\theta}_0) = A(q,\hat{\theta}_0) = 1$, and choose a number N_0 of samples at which the on-line estimation algorithm switches from the EM-based method to the quasi-Newton-based method. Then, at sample time N, regardless of whether the sample $z(N)$ is received or not, if $N < N_0$.

(1) Use (6) to compute $\bar{y}(N,\hat{\theta}_{N-1})$;
(2) Use (9) and (10) to compute $\tilde{y}(N,\hat{\theta}_{N-1})$ and $\tilde{\phi}(N,\hat{\theta}_{N-1})$;
(3) Compute $\hat{\theta}_N$ using (11)–(13).

Otherwise, if $N \geq N_0$, set the initial value of the iterations (16) as $T_{N_0} = I$. Then do:

(1) Use (15) and Lemma 3 to compute the gradient g_N;
(2) Use (17) and (18) to compute the stepsize μ_N;
(3) Compute $\hat{\theta}_N$ using (14);
(4) Compute T_N using (16).

4. Asymptotic analysis

In this section we study the statistical properties of the estimate $\hat{\theta}_N$ when the number N of samples tends to infinity. Before presenting the results, we introduce the following technical definitions.

Definition 4. For $1 \leq p < \infty$, the p-th (absolute) moment of the random variable x is defined by $\|x\|_p = \mathcal{E}\{|x|^p\}^{1/p}$. For $p = \infty$, the infinite moment of x is defined by $\|x\|_\infty = \inf\{M > 0 : x < M, \text{w.p.1}\}$. A sequence $x(t)$, $t \in \mathbb{Z}$ of random variables is said to have uniformly bounded p-th moments if there exists $M_x > 0$ such that $\|x(t)\|_p \leq M_x$, for all $t \in \mathbb{Z}$.

Definition 5. A time-varying quantizer $\mathcal{Q}_t : \mathbb{R} \to \{v_{t,1}, \ldots, v_{t,K}\}$, $t \in \mathbb{Z}$, (as defined in Section 2) is said to be bounded if there exists $M_{\mathcal{Q}} > 0$, such that, for each $t \in \mathbb{Z}$, there exists $k \in \{1, \ldots, K-1\}$ satisfying $|b_{t,k}| < M_{\mathcal{Q}}$.

4.1. Strong consistency

In this subsection we provide conditions to guarantee that the estimate $\hat{\theta}_N$ converges to the true parameter vector $\theta_\star$, as the number of samples tends to infinity.

Theorem 6. *Let $\mathcal{D} \subset \mathcal{R}$ be a compact set containing the true parameter vector $\theta_\star$, and such that, for all $\theta \in \mathcal{D}$, the roots of $A(q,\theta)$ have magnitudes smaller than or equal to $1-\epsilon$, for some $\epsilon > 0$. Let $u(t)$ be bounded (i.e., there exists $M_u > 0$ such that $|u(t)| < M_u$, for all $t \in \mathbb{Z}$) and such that $\lim_{N\to\infty}\frac{1}{N}\sum_{t=1}^{N}(x(t,\theta) - x(t,\theta_\star))^2 = 0$ holds for $\theta, \theta_\star \in \mathcal{D}$ if and only if $\theta = \theta_\star$. Let also $\mathcal{Q}_t$ be bounded and $\lambda > 0$. If for each N, $\hat{\theta}_N$ is constrained to belong to $\mathcal{D}$, i.e., $\hat{\theta}_N \in \arg\max_{\theta\in\mathcal{D}} l(\theta|Z_N)$, then*

$$\lim_{N\to\infty}\hat{\theta}_N \overset{\text{w.p.1}}{=} \theta_\star.$$

Remark 7. A strong consistency result for system identification without output quantization or packet dropouts can be found in Lennart (1999, Th. 8.3). When applied to the setting described in Section 2, this result states that consistency holds under the same assumptions as those in Theorem 6 (except for the boundedness of $\mathcal{Q}_t$ and $\lambda > 0$). More precisely, our assumption on $\mathcal{D}$ is equivalent to a condition called *uniform stability* in Lennart (1999). Also, in view of Lennart (1999, Th. 13.1), our assumption on the input signal is equivalent to the assumption called *persistently exciting of order r* in Lennart (1999), which is in turn equivalent to the condition called *informative enough*, required by Lennart (1999, Th. 8.3). Hence, Theorem 6 essentially states that placing a (bounded) quantizer at the output does not alter strong consistency, provided that a non-vanishing fraction of observations is received.

4.2. Asymptotic normality

In this section we state conditions under which the random vector $\hat{\theta}_N - \theta_\star$ converges in distribution to a multivariate normal vector. The main result of this section is stated in Theorem 10 below.

Lemma 8. *Let $F = \lim_{N\to\infty}\frac{1}{N}F_N$, where $F_N = \mathcal{E}_{\theta_\star}\left\{\frac{\partial}{\partial\theta} l(\theta|Z_N)\big|_{\theta_\star} \frac{\partial}{\partial\theta^T} l(\theta|Z_N)|_{\theta_\star}\right\}$ denotes the Fisher information matrix. Then $F = \frac{\lambda}{\sigma^2}\Phi_\mu$, where*

$$\Phi_\mu = \lim_{N\to\infty}\frac{1}{N}\sum_{t=1}^{N}\mu(t)\dot{x}(t,\theta_\star)\dot{x}^T(t,\theta_\star), \tag{21}$$

with $\mu(t) = \frac{\bar{\sigma}^2(t)}{\sigma^2}$, $\bar{\sigma}^2(t) = \mathcal{E}_{\theta_\star}\left\{\left(\check{y}(t,\theta_\star) - x(t,\theta_\star)\right)^2\right\}$ and $\check{y}(t,\theta) = \mathcal{E}_\theta\{y(t)|\mathcal{Q}_t[y(t)]\}$. Also, F is strictly positive-definite.

Remark 9. Notice that $\check{y}(t,\theta)$ in Lemma 8 denotes the conditional mean of $y(t)$ given the knowledge of its quantized value, regardless of whether this value was received or not. Hence $\mu(t)$ accounts only for the effect of quantization and does not depend on packet dropouts.

Theorem 10. *If the conditions in Theorem 6 hold, and $\theta_\star$ lies in the interior of $\mathcal{D}$, then*

$$\sqrt{N}\left(\hat{\theta}_N - \theta_\star\right) \overset{\text{in dist.}}{\rightarrow} \mathcal{N}(0, C), \tag{22}$$

where $\overset{\text{in dist.}}{\rightarrow}$ denotes convergence in distribution and

$$C = F^{-1} = \frac{\sigma^2}{\lambda}\Phi_\mu^{-1}. \tag{23}$$

Remark 11. A result similar to that in Theorem 10, but for system identification without output quantization or packet dropouts, and with the estimate obtained using the prediction (quadratic) error criterion, is given in Lennart (1999, Theorem 9.1). This criterion is shown to be equivalent to the ML criterion when the noise has normal distribution (Lennart, 1999, Theorem 9.1). Notice that the asymptotic covariance (23) reduces to their result if $\lambda = 1$ (i.e., without packet dropouts) and $\mu(t) = 1$ (i.e., without quantization).

The result in Theorem 10 states that one of the conditions required for $\sqrt{N}\left(\hat{\theta}_N - \theta_\star\right)$ to converge to a normal distribution, is that the (deterministic) input $u(t)$ is bounded. Hence, the asymptotic covariance C depends on the particular input $u(t)$, via Φ_μ. If instead of a deterministic signal, the input $u(t)$ is a bounded random process, the result still holds, and the value of C is given in the next corollary.

Corollary 12. *If the conditions in Theorem 10 hold, $u(t)$ is the realization of a wide-sense stationary and ergodic[4] random process with uniformly bounded ∞-th moment, and $\mathcal{Q}_t$ is built as a time-invariant function of $u(\tau)$, $\tau \leq t$, then the result of Theorem 10 holds with*

$$\Phi_\mu \overset{\text{w.p.1}}{=} \mathcal{E}_u\left\{\mu(t)\dot{x}(t,\theta_\star)\dot{x}^T(t,\theta_\star)\right\}, \tag{24}$$

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

Remark 13. Corollary 12 requires $u(t)$ to be ergodic (Klenke, 2008). This is a technical condition that is satisfied by any signal composed of i.i.d. samples. By Birkhoff's ergodic theorem (Klenke, 2008), the ergodic property continues to hold for the output of a linear time-invariant stable filter with any ergodic sequence as its input.

5. Optimum quantizer design

5.1. Optimum time-varying quantization scheme

The result in Theorem 10 can be used to choose the quantization thresholds $b_{t,k}$, $k = 0,\ldots,K$ of the quantizer $\mathcal{Q}_t[\cdot]$ to minimize the asymptotic error covariance C. This is equivalent to choosing the thresholds so that $\bar{\sigma}^2(t)$ is maximized. We have that

$$\begin{aligned}\bar{\sigma}^2(t) &= \mathcal{E}\left\{\left(\bar{\mathcal{Q}}_t\left[y(t,\theta_\star)\right] - x(t,\theta_\star)\right)^2\right\} \\ &= \mathcal{E}\left\{\left(\tilde{\mathcal{Q}}_{t,x(t,\theta_\star)}\left[w(t)\right]\right)^2\right\},\end{aligned} \tag{25}$$

[4] An ergodic process here refers to a random process such that all events which are invariant under time-shifts have probability either one or zero (Gray & Davisson, 2010, Section 6.14).

Table 1
Coefficient μ for different number K of quantization levels, when $w(t)$ has normal distribution.

K	2	3	4	5	6	7	8
μ	0.6366	0.8089	0.8825	0.9201	0.9420	0.9560	0.9655

(notice that the expectation operations above are independent of $\theta_\star$, hence the subindex $_{\theta_\star}$ is omitted to simplify the notation) with $\bar{\mathcal{Q}}_t$ and $\tilde{\mathcal{Q}}_{t,x}$ being the quantizers defined by

$$\bar{\mathcal{Q}}_t[y] = \mathcal{E}\{y | y \in [b_{k-1}, b_k]\}, \quad \text{if } y \in [b_{k-1}, b_k]; \tag{26}$$

$$\tilde{\mathcal{Q}}_{t,x}[w] = \bar{\mathcal{Q}}_t[x+w] - x. \tag{27}$$

Hence, $\bar{\sigma}^2(t)$ can be interpreted as the power of the quantized version of the noise $w(t)$ obtained from the time-varying quantizer $\tilde{\mathcal{Q}}_{t,x(t,\theta_\star)}$. Then, the quantization thresholds $b_{t,k}$, $k = 0,\ldots,K$ need to be chosen so that the quantizer $\tilde{\mathcal{Q}}_{t,x(t,\theta_\star)}$ maximizes $\bar{\sigma}^2(t)$. This is equivalent to minimizing the power $\mathcal{E}\{w_e^2(t)\}$ of the quantization error $w_e(t) = w(t) - \tilde{\mathcal{Q}}_{t,x(t,\theta_\star)}[w(t)]$, since from Gersho and Gray (1991, Eq. (6.2.14)), we have that

$$\bar{\sigma}^2(t) = \sigma^2 - \mathcal{E}\{w_e^2(t)\}. \tag{28}$$

Hence, the optimal choice for the quantization thresholds of $\tilde{\mathcal{Q}}_{t,x(t,\theta_\star)}$ is given by those of a Lloyd–Max quantizer (Max, 1960) which is designed using the PDF of $w(t)$. From (27), and since the quantization thresholds of $\bar{\mathcal{Q}}_t$ equal those of $\mathcal{Q}_t$, we have that $\bar{\sigma}^2(t)$ is maximized by choosing, for each $k = 0,\ldots,K$,

$$b_{t,k} = \tilde{b}_k + x(t,\theta_\star) \tag{29}$$

with $\tilde{b}_k = \frac{\tilde{w}_{k-1}+\tilde{w}_k}{2}$ and $\tilde{w}_k = \mathcal{E}\left\{w(t) | w(t) \in [\tilde{b}_{k-1}, \tilde{b}_k]\right\}$.

Using this quantizer design, the following corollary of Theorem 10, which we state without proof, immediately follows.

Corollary 14. *If the conditions in Theorem 10 hold, and the quantization thresholds b_k, $k = 0,\ldots,K$ of $\mathcal{Q}_t$ are chosen according to (29), then*

$$C = \frac{\sigma^2}{\lambda\mu}\Phi^{-1}, \tag{30}$$

where

$$\Phi = \lim_{N\to\infty}\frac{1}{N}\sum_{t=1}^{N}\dot{x}(t,\theta_\star)\dot{x}^T(t,\theta_\star), \tag{31}$$

with $\mu = \bar{\sigma}^2/\sigma^2$, $\bar{\sigma}^2 = \mathcal{E}\left\{\tilde{\mathcal{Q}}^2[w(t)]\right\}$, and $\tilde{\mathcal{Q}}$ being a Lloyd–Max quantizer designed in an off-line manner using the PDF of $w(t)$.

Remark 15. Notice that Eq. (30) differs from the classical result of system identification Lennart (1999, Eq. (9.17)) in the factor $1/\mu\lambda$, where μ accounts for the effect of the quantizer and λ accounts for that of the packet dropouts. The coefficient μ states the (inverse of the) relationship between the power of the noise $w(t)$, and the power of the signal obtained after quantizing $w(t)$ using an optimum Lloyd–Max quantizer. Hence, μ tends to 1 as the number of quantization levels K tends to infinity. This can be seen in Table 1, which shows the dependence of μ on K, for the case when $w(t)$ has normal distribution.

Remark 16. Notice that if $\lambda = 1$, θ is a scalar, $x(t,\theta) = \theta$ and $\mathcal{Q}_t$ is a one-bit quantizer defined by $\mathcal{Q}_t = 1$ if $y > \tau$ and 0 otherwise, for all $t \in \mathbb{N}$, then as shown in Appendix C, (23) reduces to

$$C = \frac{F_W(\tau-\theta_\star)(1-F_W(\tau-\theta_\star))}{f_W^2(\tau-\theta_\star)}. \tag{32}$$

This is in agreement with the result derived in Ribeiro and Giannakis (2006). It is clear that the optimal quantizer that minimizes C is obtained by choosing $\tau = \theta_\star$, in which case $C = \pi\sigma^2/2$. This is in agreement with (29). However, they fail to give a practical quantizer to asymptotically approach this minimum error covariance. Actually, the best of our knowledge suggests that there is no simple quantizer in the literature to asymptotically approach the minimum error covariance. It is worth mentioning that the asymptotically optimal quantizer given in Fang and Li (2008) requires solving a maximum likelihood estimation at each time step.

5.2. Adaptive quantization scheme for system identification

Unfortunately, the optimal quantizer (29) in Section 5.1 requires the knowledge of the unknown true parameter vector $\theta_\star$. A practical workaround is to replace $\theta_\star$ by the estimate $\hat{\theta}_{t-1}$ obtained at the previous sample-time $t-1$. 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 C can still be achieved in this case. Precisely, whether Corollary 14 holds if (29) is replaced by

$$b_{t,k} = \tilde{b}_k + x(t, \hat{\theta}_{t-1}). \tag{33}$$

The answer is positive and is stated in the following corollary.

Corollary 17. *If the conditions in Theorem 10 hold, and the quantization thresholds $b_{t,k}, k = 0, \ldots, K$ of $\mathcal{Q}_t$ are chosen according to (33), then the asymptotic error covariance matrix C is given by (30).*

6. Simulations

6.1. Comparison with the prediction error criterion

The identification method derived in Section 3 uses the maximum likelihood (ML) criterion to account for the presence of the output quantizer and packet dropouts. In this section we compare the performance of this method, with that of the prediction (quadratic) error (PE) method, described in Lennart (1999, Sec. 7.2). We assume that $\lambda = 1$, so that no packet is lost. Then, the PE method ignores the presence of the quantizer and estimates the parameters θ to minimize the power of the difference between the quantized samples $z(t)$ and their value predicted using the input signal $u(t)$ and θ. To do so we use the iterative weighted linear least squares algorithm (Pintelon et al., 1994; Steiglitz & McBride, 1965) to obtain an initial set of parameters, which is then used to initialize a quasi-Newton search method (Fletcher, 1987). Since the PE method is not able to identify the system's gain when using a one-bit quantization scheme, we normalize the identified systems for doing the comparison.

We generate the input signal $u(t)$ using an i.i.d. random process. The distribution of each sample $u(t)$ is obtained from a $\mathcal{N}(0, \nu^2)$ distribution, by truncating it to the interval $[-4\nu^2, 4\nu^2]$, and ν^2 is computed so that the power of $x(t)$ equals unity. We use a time-invariant (i.e., $\mathcal{Q}_t = \mathcal{Q}$, for each $t \in \mathbb{N}$) quantizer, designed using the Lloyd–Max algorithm applied to the distribution $\mathcal{N}(0, 1)$. The noise power is $\sigma^2 = 0.1$, and the true system is given by $B(q)/A(q) = 1/(1 - 1.764q^{-1} + 0.81q^{-2})$.

In Fig. 1 we compare the trace $\mathrm{Tr}\{C_N/N\}$ of the covariance C_N/N of $\left(\hat{\theta}_N - \theta_\star\right)$, for different values of N, and for one- and two-bits quantization schemes (i.e., $K = 2$ and $K = 4$). For each N, we compute $\mathrm{Tr}\{C_N/N\}$ using 1000 Monte Carlo runs. We see that the proposed ML method outperforms the PE method, specially for a one-bit quantization scheme.

6.2. Convergence comparison for different quantizers

We now evaluate how the choice of the output quantizer affects the evolution of the trace $\mathrm{Tr}\{C_N/N\}$ of the covariance C_N/N of $\left(\hat{\theta}_N - \theta_\star\right)$. To this end, we choose the input signal, the true system and the noise power σ^2 as in Section 6.1, and set $\lambda = 0.9$. We consider three different two-bit quantizers (i.e., $K = 4$). The first quantizer $\mathcal{Q}_s$ is stationary and is designed as described in Section 6.1. The second quantizer $\mathcal{Q}_o$ is the optimal quantizer given by (29), and the third $\mathcal{Q}_a$ is the adaptive (asymptotically optimal) quantizer given by (33). Normalization, as done in the experiment in Section 6.1, is not done in this comparison.

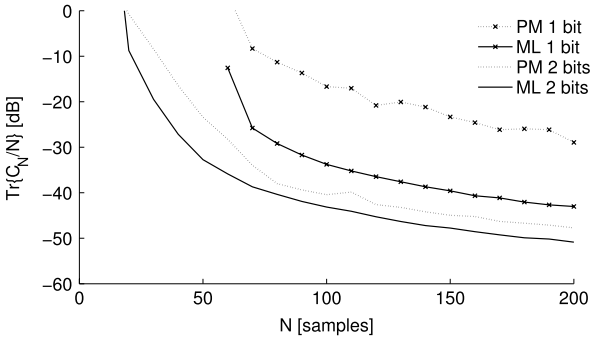

Fig. 1. Comparison between the proposed ML identification method, and the classic PE method.

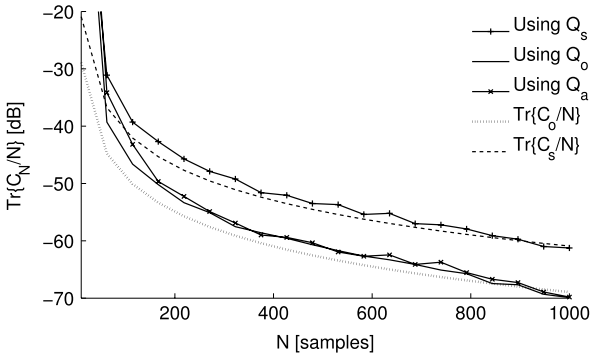

Fig. 2. Comparison of convergence rates for different output quantizers.

As in Section 6.1, for each quantization scheme we compute $\mathrm{Tr}\{C_N/N\}$ using 1000 Monte Carlo runs. The result is shown in Fig. 2, which also shows the theoretical asymptotic values $\mathrm{Tr}\{C_s/N\}$ and $\mathrm{Tr}\{C_o/N\}$ (obtained using (23) and (24)), corresponding to $\mathcal{Q}_s$ and $\mathcal{Q}_o$, respectively. We see how the use of $\mathcal{Q}_o$ leads to an asymptotic value of $\mathrm{Tr}\{C_N/N\}$ which is about 8 dB smaller than the one resulting from $\mathcal{Q}_s$. Also, the adaptive quantizer $\mathcal{Q}_a$ approaches the optimal value $\mathrm{Tr}\{C_o/N\}$ in the limit.

7. Conclusion

In this paper we have studied the problem of system identification for ARMA models subject to noisy digital communication constraints. We have proposed a simple adaptive quantizer, and the corresponding recursive identification algorithm, to address the joint effect of finite-level quantization and packet dropouts on the identification accuracy. The proposed algorithm is shown to be optimal in the sense of asymptotically achieving the minimum parameter estimation error covariance. Simulation results are included to validate the identification algorithm.

Appendix A. Proofs of results in Section 3

Proof of Lemma 1. Since the samples $w(t)$ are statistically independent, we have $\log p_\theta(Z_N, Y_N) = \sum_{t=1}^{N} \log p_\theta(z(t), y(t)|\theta)$.

Also, $p_{\hat\theta}(Y_N|Z_N) = \prod_{t=1}^N p_{\hat\theta}(y(t)|z(t))$. Then, since $(y(t), z(t))$ and $(y(s), z(s))$ are statistically independent whenever $t \neq s$, we have

$$Q_N(\theta,\hat\theta) = \int \sum_{t=1}^N \log p_\theta(z(t),y(t)) \prod_{s=1}^N p_{\hat\theta}(y(s)|z(s))dY_N$$
$$= \sum_{t=1}^N \int \log p_\theta(z(t),y(t)) p_{\hat\theta}(y(t)|z(t))dy(t).$$

Now, $p_\theta(z(t),y(t)) = p_\theta(z(t)|y(t))p_\theta(y(t))$, with $p_\theta(z(t)|y(t)) = 1$ if $y \in \breve{\mathcal{Q}}_t^{-1}[z(t)]$ and zero otherwise. Hence, $p_\theta(z(t),y(t)) = 0$ if and only if $p_{\hat\theta}(y(t)|z(t)) = 0$. From the convention $0 \times \infty = 0$, it follows that $\int \log p_\theta(z(t),y(t))p_{\hat\theta}(y(t)|z(t))dy(t) = \int \log p_\theta(y(t))p_{\hat\theta}(y(t)|z(t))dy(t)$. Hence, since $p_\theta(y(t)) \sim \mathcal{N}(x(t,\theta),\sigma^2)$, it follows that

$$Q_N(\theta,\hat\theta) = \sum_{t=1}^N \int \log p_\theta(y(t))p_{\hat\theta}(y(t)|z(t))dy(t)$$
$$= -\frac{N}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{t=1}^N \int (y(t)-x(t,\theta))^2 \times p_{\hat\theta}(y(t)|z(t))dy(t).$$

Then, the result follows since

$$\int (y(t)-x(t,\theta))^2 p_{\hat\theta}(y(t)|z(t))dy(t)$$
$$= \int y^2(t)p_{\hat\theta}(y(t)|z(t))dy(t) - 2x(t,\theta)\int y(t)p_{\hat\theta}(y(t)|z(t))dy(t) + x^2(t,\theta)$$
$$= \overline{y^2}(t,\hat\theta) - 2x(t,\theta)\bar y(t,\hat\theta) + x^2(t,\theta)$$
$$= \overline{y^2}(t,\hat\theta) - \bar y^2(t,\hat\theta) + (\bar y(t,\hat\theta) - x(t,\theta))^2. \quad \square$$

Lemma 18. *The following equality holds*

$$\frac{\partial}{\partial\theta}\log p_\theta(z(t)) = \frac{1}{\sigma^2}(\bar y(t,\theta) - x(t,\theta))\dot x(t,\theta). \tag{A.1}$$

Proof. Suppose that $\gamma_s = 1$ if $s = t$ and 0 otherwise. Then $Q_N(\tilde\theta,\theta) = \log p_\theta(z(t))$. From Lemma 1 we have

$$\frac{\partial}{\partial\tilde\theta}Q_N(\tilde\theta,\theta)\Big|_\theta = \frac{1}{\sigma^2}(\bar y(t,\theta) - x(t,\theta))\dot x(t,\theta).$$

Now, since the roots of a polynomial depend continuously on its coefficients (Tyrtyshnikov, 1997, Th. 3.9.1), $x(t,\theta)$ and $\dot x(t,\theta)$ are continuous functions of θ. From this, it is straightforward to verify that $p_\theta(z(t))$, $\bar y(t,\theta)$ and $Q_N(\tilde\theta,\theta)$ are continuous in θ. Thus, we can apply Proposition 10.1.4 in Cappé et al. (2005) to obtain $\frac{\partial}{\partial\theta}\log p_\theta(z(t)) = \frac{\partial}{\partial\tilde\theta}Q_N(\tilde\theta,\theta)\Big|_\theta$, and the result follows. $\square$

Proof of Lemma 3. It follows immediately from (18) and Lemma 18.

Appendix B. Proofs of results in Section 4

In this appendix we express the problem (1) as

$$\hat\theta_N = \arg\max_{\theta\in\mathcal{D}} N\Xi_N(\theta) \quad \text{with } \Xi_N(\theta) = \frac{1}{N}\log p_\theta(Z_N).$$

Since the samples $z(t)$, $t \in \mathbb{N}$ are independent, we have

$$\Xi_N(\theta) = \frac{1}{N}\sum_{t=1}^N \xi(t,\theta), \tag{B.1}$$

where

$$\xi(t,\theta) = \log p_\theta(z(t)). \tag{B.2}$$

Define $\overline{\Xi}(\theta) = \lim_{N\to\infty}\mathcal{E}_{\theta_\star}\{\Xi_N(\theta)\}$ and

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

We also define $\dot\xi(t,\theta) = \frac{\partial}{\partial\theta}\xi(t,\theta)$ and $\ddot\xi(t,\theta) = \frac{\partial^2}{\partial\theta\partial\theta'}\xi(t,\theta)$.

Lemma 19. *Let $0 \le p_k \le 1$, for $k = 1,\dots,K$. Then $\sum_{k=1}^K p_k\log^2 p_k \le Ke^{-2}$.*

Proof. Let $V = \sum_{k=1}^K p_k\log^2 p_k$. Then, $\frac{\partial V}{\partial p_k} = \log^2 p_k + 2\log p_k$, for each $k = 1,\dots,K$. These derivatives equal zero whenever $p_k = e^{-2}$. Also, $\frac{\partial^2 V}{\partial p_k^2}\Big|_{e^{-2}} < 0$. Hence, the result follows by considering $p_k = e^{-2}$ for all $k = 1,\dots,K$. $\square$

Lemma 20. *Under the assumptions of Theorem 6, $x(t,\theta)$ is uniformly bounded (i.e., there exists $M > 0$ such that $|x(t,\theta)| < M$, for all $t \in \mathbb{N}$ and $\theta \in \mathcal{D}$), and so is $\dot x(t,\theta)$. Also $\|\xi(t,\theta)\|_2$, $\|\dot\xi(t,\theta)\|_2$ and $\|\ddot\xi(t,\theta)\|_2$ are uniformly bounded, as well as $\left\|\frac{\partial}{\partial\theta_i}\ddot\xi(t,\theta)\right\|_\infty$, for all $i = 1,\dots,r$.*[5]

Proof. The uniform boundedness of $x(t,\theta)$ and $\dot x(t,\theta)$ follow from the boundedness of $u(t)$ and the assumption on the roots of $A(q,\theta)$. Also, the uniform boundedness of $\|\xi(t,\theta)\|_2$ follows from Lemma 19

$$\|\xi(t,\theta)\|_2^2 = \mathcal{E}_{\theta_\star}\left\{\log^2 p_\theta(z(t))\right\}$$
$$= \lambda\sum_{k=1}^K p_\theta(v_{t,k})\log^2 p_\theta(v_{t,k}) \le \lambda Ke^{-2}.$$

From Gersho and Gray (1991, Eq. (6.2.14)) we obtain that

$$\|\bar y(t,\theta)\|_2^2 \le \|y(t,\theta)\|_2^2 \le \sigma^2 \sup_{t,\theta}|x(t,\theta)|.$$

Hence, the uniform boundedness of $\|\dot\xi(t,\theta)\|_2$ follows from (B.2) and Lemma 18. Now,

$$\bar y(t,\theta) = x(t,\theta) + \frac{\int_{a-x(t,\theta)}^{b-x(t,\theta)} w f_W(w)\,dw}{\int_{a-x(t,\theta)}^{b-x(t,\theta)} f_W(w)\,dw}$$
$$= x(t,\theta) + \frac{G_W(b-x(t,\theta)) - G_W(a-x(t,\theta))}{F_W(b-x(t,\theta)) - F_W(a-x(t,\theta))}, \tag{B.3}$$

where $[a,b] = \breve{\mathcal{Q}}_t[z(t)]$ and $G_W(w)$ denote the anti-derivative of $wf_W(w)$. Although technically tedious, it is straightforward to verify from (B.3) that the first two derivatives of $\bar y(t,\theta)$ with respect to θ have uniformly bounded absolute values. Hence, the uniform boundedness of $\|\ddot\xi(t,\theta)\|_2$ and $\left\|\frac{\partial}{\partial\theta_i}\ddot\xi(t,\theta)\right\|_\infty$, $i = 1,\dots,r$, follow from (B.2) and Lemma 18. $\square$

Lemma 21. *Under the assumptions of Theorem 6, for all $N \in \mathbb{N}$, the functions $\Xi_N(\theta)$, $\dot\Xi_N(\theta)$ and $\ddot\Xi_N(\theta)$ are continuous on $\mathcal{D}$. Also, $\Xi_N(\theta) \to \bar\Xi(\theta)$, $\dot\Xi_N(\theta) \to \bar{\dot\Xi}(\theta)$ and $\ddot\Xi_N(\theta) \to \bar{\ddot\Xi}(\theta)$ with probability one and uniformly on $\theta \in \mathcal{D}$ (strong uniform convergence).*

[5] For a random variable x, $\|x\|_2 = \mathcal{E}_{\theta_\star}\{x^2\}^{1/2}$ and $\|x\|_\infty = \inf\{c \ge 0 : |x| \le c \text{ w.p.1}\}$.

Proof. We split the proof into three steps:

(Step 1). From Lemma 20, the third derivative of $\xi(t,\theta)$ with respect to θ is uniformly bounded. Hence,

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

In view of (B.1), we have that

$$\sup_{N\in\mathbb{N}}\sup_{\theta,\phi\in\mathcal{D}}\left\|\ddot{\Xi}_N(\theta)-\ddot{\Xi}_N(\phi)\right\|\|\theta-\phi\|^{-1}<\infty. \tag{B.4}$$

This in turn implies that $\ddot{\Xi}_N(\theta)$ is continuous, and so are $\Xi_N(\theta)$ and $\dot{\Xi}_N(\theta)$.

(Step 2). From Lemma 20, $\|\xi(t,\theta)\|_2$, $\left\|\dot{\xi}(t,\theta)\right\|_2$ and $\left\|\ddot{\xi}(t,\theta)\right\|_2$ are uniformly bounded. Since these variables are also statistically independent, Rajchman's strong law of large numbers (Chung, 2001, Theorem 5.1.2) asserts that, for all $\theta\in\mathcal{D}$, $\Xi_N(\theta)\overset{\text{w.p.1}}{\to}\overline{\Xi}(\theta)$, $\dot{\Xi}_N(\theta)\overset{\text{w.p.1}}{\to}\overline{\dot{\Xi}}(\theta)$ and $\ddot{\Xi}_N(\theta)\overset{\text{w.p.1}}{\to}\overline{\ddot{\Xi}}(\theta)$.

(Step 3). Fix $i,j\in\{1,\dots,p\}$ and define $f_N(\theta)=[\ddot{\Xi}_N(\theta)-\overline{\ddot{\Xi}}(\theta)]_{i,j}$, where $[\cdot]_{i,j}$ denotes the i,j-th entry of a matrix. For all $\theta,\phi\in\mathcal{D}$,

$$\begin{aligned}|f_N(\theta)-f_N(\phi)|\le&\left|\left[\ddot{\Xi}_N(\theta)-\ddot{\Xi}_N(\phi)\right]_{i,j}\right|\\&+\left|\left[\overline{\ddot{\Xi}}(\theta)-\overline{\ddot{\Xi}}(\phi)\right]_{i,j}\right|.\end{aligned}\tag{B.5}$$

Consider the event (i.e., the set in the underlying probability space) where $\ddot{\Xi}_N(\theta)\overset{\text{w.p.1}}{\to}\overline{\ddot{\Xi}}(\theta)$ and $\ddot{\Xi}_N(\phi)\overset{\text{w.p.1}}{\to}\overline{\ddot{\Xi}}(\phi)$. On this event, we have that

$$\begin{aligned}&\sup_{\theta,\phi\in\mathcal{D}}\left\|\overline{\ddot{\Xi}}(\theta)-\overline{\ddot{\Xi}}(\phi)\right\|\|\theta-\phi\|^{-1}\\&=\sup_{\theta,\phi\in\mathcal{D}}\lim_{N\to\infty}\left\|\ddot{\Xi}_N(\theta)-\ddot{\Xi}_N(\phi)\right\|\|\theta-\phi\|^{-1}\\&\le\sup_{N\in\mathbb{N}}\sup_{\theta,\phi\in\mathcal{D}}\left\|\ddot{\Xi}_N(\theta)-\ddot{\Xi}_N(\phi)\right\|\|\theta-\phi\|^{-1}<\infty,\end{aligned}\tag{B.6}$$

where the last inequality follows from (B.4). Putting (B.4) and (B.6) into (B.5), we have that there exists $M>0$ such that $|f_N(\theta)-f_N(\phi)|\le M\|\theta-\phi\|$. Hence, from Davidson (1994, Theorem 21.10), $f_N(\theta)$ is strongly stochastically equi-continuous. Then, from Davidson (1994, Theorem 21.8), $[\ddot{\Xi}_N(\theta)]_{i,j}\overset{\text{w.p.1}}{\to}[\overline{\ddot{\Xi}}(\theta)]_{i,j}$, uniformly in θ, which shows the strong uniform convergence of $\ddot{\Xi}_N(\theta)$. The strong uniform convergence of $\Xi_N(\theta)$ and $\dot{\Xi}_N(\theta)$ then follows from Rudin (1976, Theorem 7.17) and (Step 3). □

Lemma 22. *Under the assumptions of Theorem 6, we have that*

$$\arg\max_{\theta\in\mathcal{D}}\overline{\Xi}(\theta)=\{\theta_\star\}, \tag{B.7}$$

i.e., $\overline{\Xi}(\theta)$ is maximized only at $\theta_\star$. Also, $\overline{\ddot{\Xi}}(\theta_\star)$ is non-singular.

Proof. We split the proof into four steps:

(Step 1). 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

$$\begin{aligned}f_t(d)&=D_{KL}\left(\tilde{p}(z(t)|x(t,\theta_\star))\parallel\tilde{p}(z(t)|x(t,\theta_\star)+d)\right)\\&=\mathcal{E}_{\theta_\star}\left\{\log\frac{\tilde{p}(z(t)|x(t,\theta_\star))}{\tilde{p}(z(t)|x(t,\theta_\star)+d)}\right\}\end{aligned}$$

where $D_{KL}(p\parallel q)$ denotes the Kullback–Leibler distance between the probability distributions p and q (Cover & Thomas, 1999). Then we have

$$\arg\max_{\theta\in\mathcal{D}}\overline{\Xi}(\theta)=\arg\min_{\theta\in\mathcal{D}}H(\theta). \tag{B.8}$$

where

$$\begin{aligned}H(\theta)&=\lim_{N\to\infty}\mathcal{E}_{\theta_\star}\{\Xi_N(\theta_\star)-\Xi_N(\theta)\}\\&=\lim_{N\to\infty}\frac{1}{N}\sum_{t=1}^{N}f_t\left(x(t,\theta)-x(t,\theta_\star)\right).\end{aligned}\tag{B.9}$$

From the positivity of $D_{KL}(\cdot\|\cdot)$, it follows that the minimum of $H(\theta)$ equals zero, and is attained when $\theta=\theta_\star$. It remains to show that $H(\theta)=0$ if and only if $\theta=\theta_\star$.

(Step 2). For $-\infty<\alpha<\beta<\infty$, define

$$g(\alpha,\beta)=\left(\frac{f_W(\beta)-f_W(\alpha)}{F_W(\beta)-F_W(\alpha)}\right)^2-\frac{f'_W(\beta)-f'_W(\alpha)}{F_W(\beta)-F_W(\alpha)}.$$

From numerical evaluation we can verify that, for each $A>0$ there exists $B>0$ such that $g(\alpha,\beta)\ge B$, whenever either $|\alpha|\le A$ or $|\beta|\le B$.

(Step 3). It is straightforward to verify that,

$$\frac{\partial^2}{\partial d^2}f_t(d)=\lambda\sum_{k=1}^{K}p_{\theta_\star}(z(t)=v_{t,k})g(b_{t,k-1}+d,b_{t,k}+d).$$

Now, from Lemma 20 we have that $|x(t,\theta)-x(t,\theta_\star)|$ is bounded. Since the quantizer is bounded, and $\lambda>0$, from Step 2 it follows that there exists $\epsilon>0$ such that $\frac{\partial^2}{\partial d^2}f_t(d)\Big|_{d=x(t,\theta)-x(t,\theta_\star)}>\epsilon$, for all t. Hence, $f_t\left(x(t,\theta)-x(t,\theta_\star)\right)>\epsilon\left(x(t,\theta)-x(t,\theta_\star)\right)^2$, for all t. Then, (B.7) follows from our assumption on $u(t)$, (B.8) and (B.9).

(Step 4). For the second part, consider the event (i.e., the set in the underlying probability space), where $\dot{\Xi}_N(\theta)$ and $\ddot{\Xi}_N(\theta)$ converge uniformly on θ. Then, on this event, we have that

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

where (a) follows from two applications of Rudin (1976, Theorem 7.17), and (b) follows from (B.7). □

Proof of Theorem 6. From Lemma 21 we have that, for each $N\in\mathbb{N}$, $\Xi_N(\theta)$ is a continuous function of θ, and with probability one, $\Xi_N(\theta)\to\bar{\Xi}(\theta)$ uniformly. Also, from Lemma 22, $\bar{\Xi}(\theta)$ attains its maximum at $\theta_\star$ only. This, together with the compactness of $\mathcal{D}$ gives the conditions for strong convergence stated in Gourieroux and Monfort (1996, Property 24.2). □

Proof of Lemma 8. We split the proof in two steps:

(Step 1). From (6), it can be readily verified that

$$\mathcal{E}_{\theta_\star}\{\bar{y}(t,\theta_\star)\}=\mathcal{E}_{\theta_\star}\{y(t,\theta_\star)\}=x(t,\theta_\star). \tag{B.10}$$

Hence, from (A.1), we have that $\mathcal{E}_{\theta_\star}\left\{\dot{\xi}(t,\theta_\star)\right\}=0$, for all t. Since the random variables $\dot{\xi}(t,\theta_\star)$ are also statistically independent, it follows from Lemma 3 that

$$\begin{aligned}F_N&=N^2\mathcal{E}_{\theta_\star}\left\{\dot{\Xi}_N(\theta_\star)\dot{\Xi}_N^T(\theta_\star)\right\}\\&=\sum_{t=1}^{N}\mathcal{E}_{\theta_\star}\left\{\dot{\xi}(t,\theta_\star)\dot{\xi}^T(t,\theta_\star)\right\}\\&=\frac{1}{\sigma^4}\sum_{t=1}^{N}\mathcal{E}_{\theta_\star}\left\{(\bar{y}(t,\theta_\star)-x(t,\theta_\star))^2\right\}\dot{x}(t,\theta_\star)\dot{x}^T(t,\theta_\star).\end{aligned}\tag{B.11}$$

Now, if $\gamma_t = 0$, we have that $\bar{y}(t,\theta_\star) = x(t,\theta_\star)$. Hence, $\bar{y}(t,\theta_\star) - x(t,\theta_\star) = \gamma_t\left(\check{y}(t,\theta_\star) - x(t,\theta_\star)\right)$. The expression of F follows from that $F_N = \frac{\lambda}{\sigma^4}\sum_{t=1}^{N}\mathcal{E}_{\theta_\star}\{(\check{y}(t,\theta_\star) - x(t,\theta_\star))^2\}\dot{x}(t,\theta_\star)\dot{x}^T(s,\theta_\star)$.

(Step 2). From Gourieroux and Monfort (1995, Property 3.8) we have that

$$\overline{\ddot{\Xi}}(\theta_\star) = \lim_{N\to\infty}\frac{1}{N}\mathcal{E}_{\theta_\star}\left\{\ddot{\Xi}_N(\theta_\star)\right\} = -\lim_{N\to\infty}\frac{1}{N}F_N = -F.$$

Then, $F > 0$ follows from Lemmas 8 and 22. □

Lemma 23. *From the assumptions of Theorem 10, we have that*

$$\sqrt{N}\dot{\Xi}_N(\theta_\star) \overset{\text{in dist.}}{\to} \mathcal{N}(0, F^{-1}).$$

Proof. We split the proof into four steps:

(Step 1). Let $\vartheta(t) = \dot{\xi}(t,\theta_\star)$. For each $y \in \mathbb{R}^r$, define $\vartheta_y(t) = y^T\vartheta(t)$ and $P_{y,N} = \frac{1}{N}\sum_{t=1}^{N}\mathcal{E}\left\{\vartheta_y(t)\vartheta_y^T(t)\right\} = y^T P_N y$, where $P_N = \frac{1}{N}\sum_{t=1}^{N}\mathcal{E}\{\vartheta(t)\vartheta(t)\}$.

(Step 2). Let $\delta > 0$, and define

$$L_y = \lim_{N\to\infty}\left(NP_{y,N}\right)^{-1-\delta/2}\left(\frac{1}{N}\sum_{t=1}^{N}\mathcal{E}\{|\vartheta_y(t)|^{2+\delta}\}\right).$$

From Lemma 20, we have that there exists $M > 0$ such that $\|\vartheta(t)\|_{2+\delta} < M$ for all t. Hence, from Minkowsky's inequality (Rudin, 2006),

$$\left\|\vartheta_y(t)\right\|_{2+\delta} \le \sum_{i=1}^{r}|y_i|\,\|\vartheta(t)\|_{2+\delta} \le M\|y\|_1. \tag{B.12}$$

Moreover,

$$P_{y,N} = y^T P_N y \ge \underline{\text{eig}}(P_N)\,\|y\|_2^2, \tag{B.13}$$

where $\underline{\text{eig}}(P_N)$ denotes the smallest eigenvalue of P_N. In view of (B.11) we have that $P_N = \frac{1}{N}F_N$. Hence, $\lim_{N\to\infty}P_N = F > 0$. In addition, it follows from Tyrtyshnikov (1997, Th. 3.9.1) that the eigenvalues of a matrix depend continuously on its entries. Then, we have that

$$\lim_{N\to\infty}\underline{\text{eig}}(P_N) = \underline{\text{eig}}(\lim_{N\to\infty}P_N) = \underline{\text{eig}}(F) > 0. \tag{B.14}$$

By (B.12)–(B.14), it follows that $L_y = 0$.

(Step 3). From (B.2), (A.1) and (B.3), we have that $\mathcal{E}\{\vartheta(t)\} = 0$, hence $\mathcal{E}\left\{\vartheta_y(t)\right\} = 0$. Since $L_y = 0$, for each y, the sequence $\vartheta_y(t)$, $t \in \mathbb{N}$, satisfies the Lyapunov condition (Klenke, 2008, Definition 15.40). By the Lindeberg–Feller central limit theorem (Klenke, 2008, Theorem 15.43), it follows that $\frac{1}{\sqrt{NP_{N,y}}}\sum_{t=1}^{N}\vartheta_y(t) \overset{\text{in dist.}}{\to} \mathcal{N}(0,1)$.

(Step 4). Since $\lim_{N\to\infty}P_{N,y} = \lim_{N\to\infty}y^T P_N y = y^T F y$, we have that, $\frac{1}{\sqrt{N}}\sum_{t=1}^{N}\vartheta_y(t) \overset{\text{in dist.}}{\to} \mathcal{N}(0, y^T F y)$. From (B.1), we can write $\sqrt{N}\dot{\Xi}_N(\theta_\star) = \frac{1}{\sqrt{N}}\sum_{t=1}^{N}\vartheta(t)$. Then, the result follows from Davidson (1994, Eq. (25.28)). □

Proof of Theorem 10. From Lemma 21 we have that, for each $N \in \mathbb{N}$, $\ddot{\Xi}_N(\theta)$ is a continuous function of θ, and with probability one, $\ddot{\Xi}_N(\theta) \to \overline{\ddot{\Xi}}(\theta)$ uniformly. Also, from Lemma 22, $\overline{\ddot{\Xi}}(\theta_\star)$ is non-singular. This, together with Lemma 23, the compactness of $\mathcal{D}$ and $\theta_\star$ being in the interior of $\mathcal{D}$, gives the conditions for strong convergence stated in Gourieroux and Monfort (1996, Property 24.16). □

Proof of Corollary 12. From (28) we have that $\left\|\bar{\sigma}^2(t)\right\|_\infty \le \sigma^2$ (notice that now $\bar{\sigma}^2(t)$ is a random variable because $u(t)$ is a random process). Also, from Lemma 20, there exists M such that $|\dot{x}_i(t,\theta_\star)| < M$, for all t and all i. Hence, we have that

$$\left\|\mu(t)\dot{x}_i(t,\theta_\star)\dot{x}_j(t,\theta_\star)\right\|_1 \le M^2 < \infty.$$

Then, since $u(t)$ is ergodic and wide-sense stationary, and $\mathcal{Q}_t$ is obtained as a time-invariant function of $u(t)$, the result follows by applying Birkhoff's ergodic theorem (Klenke, 2008) to (21). □

Appendix C. Proofs of results in Section 5

Proof of (32). The proof consists of two steps:

(Step 1). Since $x(t,\theta) = \theta$, we have

$$\begin{aligned}\bar{\sigma}^2(t) &= p(y(t)\le\tau)\left(\mathcal{E}\{y(t)|y(t)\le\tau\} - \theta_\star\right)^2 \\ &\quad + p(y(t)>\tau)\left(\mathcal{E}\{y(t)|y(t)>\tau\} - \theta_\star\right)^2 \\ &= \frac{(1-F_W(\tau-\theta_\star))\left(\int_{-\infty}^{\tau-\theta_\star}wf_W(w)dw\right)^2}{F_W(\tau-\theta_\star)(1-F_W(\tau-\theta_\star))} \\ &\quad + \frac{F_W(\tau-\theta_\star)\left(\int_{\tau-\theta_\star}^{\infty}wf_W(w)dw\right)^2}{F_W(\tau-\theta_\star)(1-F_W(\tau-\theta_\star))}.\end{aligned}$$

Now, since $\int_{-\infty}^{\infty}wf_W(w)d(w) = \mathcal{E}\{w(t)\} = 0$, we have that $\int_{\tau-\theta_\star}^{\infty}wf_W(w)d(w) = -\int_{-\infty}^{\tau-\theta_\star}wf_W(w)d(w) = \sigma^2 f_W(\tau-\theta_\star)$. Hence,

$$\bar{\sigma}^2(t) = \frac{\sigma^4 f_W^2(\tau-\theta_\star)}{F_W(\tau-\theta_\star)(1-F_W(\tau-\theta_\star))}. \tag{C.1}$$

(Step 2). Since $x(t,\theta) = \theta$, we have that $\dot{x}(t,\theta) = 1$. Then, from (21) and (C.1),

$$\Phi_\mu = \lim_{N\to\infty}\frac{1}{N}\sum_{t=1}^{N}\frac{\bar{\sigma}^2(t)}{\sigma^2} = \frac{\sigma^2 f_W^2(\tau-\theta_\star)}{F_W(\tau-\theta_\star)(1-F_W(\tau-\theta_\star))}. \tag{C.2}$$

Finally, the result follows by putting (C.2) into (23) with $\lambda = 1$. □

Proof of Corollary 17. The proof consists of three steps:

(Step 1). From (33) we have that $\bar{\mathcal{Q}}_{t,x(t,\hat{\theta}_t)} = \mathcal{Q}_0$, for all $t \in \mathbb{Z}$, where $\mathcal{Q}_0$ is a Lloyd–Max quantizer adapted to the PDF of $w(t)$. Then, from (27) we have that $\bar{\mathcal{Q}}_t[y] = \mathcal{Q}_0\left[y - x(t,\hat{\theta}_{t-1})\right] + x(t,\hat{\theta}_{t-1})$, and Theorem 10 holds with

$$\begin{aligned}\bar{\sigma}^2(t) &= \mathcal{E}\left\{(\bar{y}(t,\theta_\star) - x(t,\theta_\star))^2\right\} \\ &= \mathcal{E}\left\{\left(\bar{\mathcal{Q}}_t\left[x(t,\theta_\star) + w(t)\right] - x(t,\theta_\star)\right)^2\right\} \\ &= \mathcal{E}\left\{\left(\mathcal{Q}_0\left[w(t) - \tilde{x}(t,\hat{\theta}_{t-1})\right] + \tilde{x}(t,\hat{\theta}_{t-1})\right)^2\right\} \\ &= \mathcal{E}\left\{\left(\check{\mathcal{Q}}_{\tilde{x}(t,\hat{\theta}_{t-1})}[w(t)]\right)^2\right\},\end{aligned}$$

where $\tilde{x}(t,\hat{\theta}_{t-1}) = x(t,\hat{\theta}_{t-1}) - x(t,\theta_\star)$ and $\check{\mathcal{Q}}_x[w] = \mathcal{Q}_0[w-x] + x$.

(Step 2). From Theorem 10, we have that $\hat{\theta}_t \to \theta_\star$ with probability one. Hence, from Chung (2001, Theorem 4.1.2) $\hat{\theta}_t \to \theta_\star$ in probability. Now, our assumption on the roots of $A(q,\theta)$ assures that $\tilde{x}(t,\theta)$ is a continuous function of θ. Then, from Lehmann (2010, Theorem 2.1.4), $\tilde{x}(t,\hat{\theta}_t) \overset{\text{in pr.}}{\to} 0$, where $\overset{\text{in pr.}}{\to}$ denotes convergence in probability. Now it is easy to check that $x \mapsto \mathcal{E}\left\{\left(\check{\mathcal{Q}}_x[w(t)]\right)^2\right\}$ is a continuous function of $x \in \mathbb{R}$. Hence,

$$\mathcal{E}_{w(t)}\left\{\left(\check{\mathcal{Q}}_{\tilde{x}(t,\hat{\theta}_{t-1})}[w(t)]\right)^2\right\} \overset{\text{in pr.}}{\to} \mathcal{E}_{w(t)}\left\{(\mathcal{Q}_0[w(t)])^2\right\} \tag{C.3}$$

(notice that the left-hand side of (C.3) is a function of the random variable $\tilde{x}(t, \hat{\theta}_{t-1})$, which is independent of $w(t)$). Then,

$$\bar{\sigma}^2(t) = \mathcal{E}_{\hat{\theta}_{t-1}} \left\{ \mathcal{E}_{w(t)} \left\{ \left(\breve{\mathcal{Q}}_{\tilde{x}(t,\hat{\theta}_{t-1})} [w(t)] \right)^2 \right\} \right\} \rightarrow \mathcal{E}_{w(t)} \left\{ (\mathcal{Q}_0 [w(t)])^2 \right\} = \bar{\sigma}^2. \quad \text{(C.4)}$$

Hence, $\Phi_\mu = \frac{\bar{\sigma}^2}{\sigma^2} \Phi = \mu \Phi$, and the result follows. □

References

Ben-Israel, A., & Greville, T. N. E. (2003). *Generalized inverses: theory and applications*: *vol. 15*. Springer Verlag.

Boyd, S., & Vandenberghe, L. (2004). *Convex optimization*. Cambridge Univ Pr.

Cappé, D., Moulines, E., & Rydén, T. (2005). *Inference in hidden Markov models*. Springer.

Chung, K. L. (2001). *A course in probability theory*. San Diego, CA: Academic Press Inc.

Cover, T. M., & Thomas, J. A. (1999). *Elements of information theory*. Wiley-India.

Davidson, J. (1994). *Stochastic limit theory: an introduction for econometricians*. USA: Oxford University Press.

Fang, J., & Li, H. (2008). Distributed adaptive quantization for wireless sensor networks: from delta modulation to maximum likelihood. *IEEE Transactions on Signal Processing*, *56*(10), 5246–5257.

Fletcher, R. (1987). *Practical methods of optimization*. John Wiley & Sons, Inc.

Fu, M., & de Souza, C. E. (2009). State estimation for linear discrete-time systems using quantized measurements. *Automatica*, *45*(12), 2937–2945.

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

Godoy, B. I., Goodwin, G. C., Aguero, J. C., Marelli, D., & Wigren, T. (2011). On identification of fir systems having quantized output data. *Automatica*, *47*(9), 1905–1915.

Gourieroux, Christian, & Monfort, Alain (1995). *Statistics and econometric models: Volume 1*. Cambridge University Press.

Gourieroux, Christian, & Monfort, Alain (1996). *Statistics and econometric models: Volume 2*. Cambridge University Press.

Gray, R. M., & Davisson, L. D. (2010). *An introduction to statistical signal processing*. Cambridge University Press, p. 2.

Klenke, A. (2008). *Probability theory: a comprehensive course*. Springer Verlag.

Lehmann, E. L. (2010). *Elements of large-sample theory*. New York: Springer, p. 2.

Lennart, L. (1999). *System identification: theory for the user*. Upper Saddle River, NJ: PTR Prentice Hall.

Max, J. (1960). Quantization for minimum distortion. *IEEE Transactions on Information Theory*, *6*(2), 7–12.

McLachlan, G. J., & Krishnan, T. (2008). *The EM algorithm and extensions* (2nd ed.). Wiley-Interscience.

Papadopoulos, H. C., Wornell, G. W., & Oppenheim, A. V. (2002). Sequential signal encoding from noisy measurements using quantizers with dynamic bias control. *IEEE Transactions on Information Theory*, *47*(3), 978–1002.

Pintelon, R., Guillaume, P., Rolain, Y., Schoukens, J., & Van hamme, H. (1994). Parametric identification of transfer functions in the frequency domain. *IEEE Transactions on Automatic Control*, *39*(11), 2245–2260.

Ribeiro, A., & Giannakis, G. (2006). Bandwidth-constrained distributed estimation for wireless sensor networks-Part I: Gaussian case. *IEEE Transactions on Signal Processing*, *54*(3), 1131–1143.

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

Rudin, W. (2006). *Real and complex analysis*. McGraw-Hill Education.

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

Steiglitz, K., & McBride, L. (1965). A technique for the identification of linear systems. *IEEE Transactions on Automatic Control*, *AC-10*(10), 461–464.

Tyrtyshnikov, E. E. (1997). *A brief introduction to numerical analysis*. Birkhauser.

Wang, L., Yin, G., Zhang, J., & Zhao, Y. (2010). *System identification with quantized observations*. Boston: Birkhäuser.

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

Xiao, J., Ribeiro, A., Luo, Z, & Giannakis, G. (2006). Distributed compression-estimation using wireless sensor networks. *IEEE Signal Processing Magazine*, *23*(4), 27–41.

You, K., Fu, M., & Xie, L. (2011). Mean square stability for Kalman filtering with Markovian packet losses. *Automatica*, *47*(12), 2647–2657.

You, K., & Xie, L. (2010). Minimum data rate for mean square stabilization of discrete LTI systems over lossy channels. *IEEE Transactions on Automatic Control*, *55*(10), 2373–2378.

You, K., Xie, L., Sun, S., & Xiao, W. 2008. Multiple-level quantized innovation Kalman filter. In: Proc. 17th IFAC World Congress, pp. 1420–1425.

Damián Marelli received his Bachelors Degree in electronics engineering from the Universidad Nacional de Rosario, Argentina in 1995, a Ph.D. degree in electrical engineering and a Bachelor (Honous) degree in mathematics from the University of Newcastle, Australia in 2003. From 2004 to 2005 he held a postdoctoral position at the Laboratoire d'Analyse Topologie et Probabilités, CNRS / Université de Provence, France. Since 2006 he is Research Academic at the School of Electrical Engineering and Computer Science at the University of Newcastle, Australia. In 2007 he received a Marie Curie Postdoctoral Fellowship, hosted at the Faculty of Mathematics, University of Vienna, Austria, and in 2010 he received a Lise Meitner Senior Fellowship, hosted at the Acoustics Research Institute of the Austrian Academy of Sciences. His main research interests include signal processing and communications.

Keyou You was born in Jiangxi Province, China, in 1985. He received the B.S. degree in statistical science from Sun Yat-sen (Zhongshan) University, Guangzhou, China, in 2007 and the Ph.D. degree in electrical and electronic engineering from Nanyang Technological University, Singapore, in 2012.

He was with the ARC Center for Complex Dynamic Systems and Control, University of Newcastle, Australia, as a visiting scholar from May 2010 to July 2010, and with the Sensor Network Lab at Nanyang Technological University as a Research Fellow from June 2011 to June 2012. Since July 2012, he has been with the Department of Automation, Tsinghua University, China as a Lecturer. His current research interests include control and estimation of networked systems, distributed control and estimation over complex networks, and sensor networks.

Dr. You won the Guan Zhaozhi best paper award at the 29th Chinese Control Conference, Beijing, China, in 2010.

Minyue Fu received his Bachelor's Degree in electrical engineering from the University of Science and Technology of China, Hefei, China, in 1982, and M.S. and Ph.D. degrees in electrical engineering from the University of Wisconsin-Madison in 1983 and 1987, respectively. From 1983 to 1987, he held a teaching assistantship and a research assistantship at the University of Wisconsin-Madison. He worked as a Computer Engineering Consultant at Nicolet Instruments, Inc., Madison, Wisconsin, during 1987. From 1987 to 1989, he served as an Assistant Professor in the Department of Electrical and Computer Engineering, Wayne State University, Detroit, Michigan. He joined the Department of Electrical and Computer Engineering, University of Newcastle, Australia, in 1989. Currently, he is a Chair Professor in Electrical Engineering and Head of the School of Electrical Engineering and Computer Science. In addition, he was a Visiting Associate Professor at the University of Iowa in 1995–1996, and a Senior Fellow/Visiting Professor at Nanyang Technological University, Singapore, 2002. He holds a Qian-ren Professorship at Zhejiang University, China. He is a Fellow of the IEEE. His main research interests include control systems, signal processing and communications. He has been an Associate Editor for the IEEE Transactions on Automatic Control, Automatica and Journal of Optimization and Engineering.