

IDENTIFICATION OF ARMA MODELS USING INTERMITTENT AND QUANTIZED OUTPUT OBSERVATIONS

Damián Marelli[†], Keyou You[‡] and Minyue Fu^{†, §}

[†] University of Newcastle, Australia

[‡]Nanyang Technological University, Singapore

[§]Zhejiang University, China

ABSTRACT

This paper studies system identification of ARMA models whose outputs are subject to finite-level quantization and random packet dropouts. A simple adaptive quantizer and the corresponding recursive identification algorithm are proposed and shown to be optimal in the sense of asymptotically achieving the minimum mean square estimation error. The joint effects of finite-level quantization and random packet dropouts on identification accuracy are exactly quantified. The theoretic results are verified by simulations.

Index Terms— System identification, ARMA models, quantization, packet dropouts.

1. INTRODUCTION

System identification of plants with quantized observations is significant in understanding modeling capacity for systems with limited sensor information, and the tradeoff between communication resources and identification performance [1]. This work is concerned with identification of ARMA models whose outputs is 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 exactly quantify the joint effects of finite-level quantization and packet dropout on the identification accuracy of ARMA models.

Recently, research on quantized identification/estimation bears a vast body of literature, see, e.g., [1, 2] and references therein. In [2], various quantized estimation schemes are reviewed in the context of wireless sensor networks. In [1], a comprehensive treatment on quantized identification is presented for single-input-single-output linear discrete time-invariant stable systems. Based on full rank periodic inputs and empirical measure, they examine the optimal identification errors, time complexity and impacts of disturbances and unmodeled dynamics on identification accuracy. Although their identification algorithms are shown to be optimal in the sense of asymptotically achieving the well-known Cramér-Rao lower bound [1], the assumption on periodic inputs makes the identification algorithm inappropriate for tracking control design. To ensure the identifiability property, this work requires input signals to be *informative enough* [3].

The fundamental problem of system identification with quantized observations consists of 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 an unknown parameter 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 [4]. It is known that the choice of the quantizer threshold is very sensitive to estimation performance. An interesting threshold selection scheme, by periodically applying a set of thresholds with equal frequencies, is proposed in [5], hoping that some thresholds are close to the unknown parameter. To asymptotically approach the minimum MSE, the authors in [6] construct an adaptive quantizer 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 MSE.

Another detrimental factor impairing the identification performance is the dropouts of the quantized observations. To the best of our knowledge, there is no work to date to quantify the joint effects of finite-level quantization and packet dropouts on identification performance. As in [7], the packet dropout process is modeled by an i.i.d. process. The basic problem is how to do compensation to derive an optimal identification algorithm if a packet dropout occurs. In the case of a Kalman filter, it is well understood that its optimality still holds under random packet dropouts of the observations, and a large packet dropout rate may lead to unstable mean estimation error covariance matrices [8, 9, 10]. An optimal technique for system identification is developed in this paper, which shows that the effect of packet dropouts on the identification performance can be explicitly expressed in terms of the packet dropout rate. As expected, the result reveals that the smaller the dropout rate, the better the resulting performance. The proofs are omitted and will appear in a journal version.

2. PROBLEM DESCRIPTION

Consider the following ARMA model

$$\begin{aligned}x(t) &= \frac{B(q)}{A(q)}u(t) \\y(t) &= x(t) + w(t) \\z(t) &= \gamma_t \mathcal{Q}_t(y(t)),\end{aligned}$$

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 distribution $\mathcal{N}(0, \sigma^2)$. The noisy output $y(t)$ is quantized by a time-varying K -level scalar quantizer $\mathcal{Q}_t : \mathbb{R} \rightarrow \{v_{t,1}, \dots, v_{t,K}\}$, $t \in \mathbb{Z}$, (which accounts for

non-stationary quantization schemes), defined by the quantization intervals $[b_{t,k-1}, b_{t,k}] = \mathcal{Q}^{-1}[v_{t,k}]$, $k = 1, \dots, K$, 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., $\mathbb{P}(\gamma_t = 1) = \lambda$).

For a fixed $N \in \mathbb{N}$, we denote $\mathcal{I} = \{1 \leq t \leq N : \gamma_t = 1\}$ the set of sample-times at which measurements are available. To simplify the notation we make the dependence of \mathcal{I} on N implicit. Define $U_N = \{u(t) : t = 1, \dots, N\}$, $Z_{\mathcal{I}} = \{z(t) : t \in \mathcal{I}\}$, $A(q) = 1 + a_1q^{-1} + \dots + a_mq^{-m}$, $B(q) = b_0 + \dots + b_nq^{-n}$ and $\theta_* = [b_0, \dots, b_n, a_1, \dots, a_m]^T$ (the superscript T denotes matrix transpose). Let $r = m + n + 1$. Then, 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.

The problem consists in estimating θ_* , given the knowledge of U_N and $Z_{\mathcal{I}}$. We do so using the maximum likelihood (ML) criterion.

3. MAXIMUM LIKELIHOOD ESTIMATION

By using the ML criterion, an estimate $\hat{\theta}_N$ up to sample time N is obtained by¹

$$\hat{\theta}_N \in \arg \max_{\theta} p(Z_{\mathcal{I}}|U_N, \theta) = \arg \max_{\theta} l(\theta|U_N, Z_{\mathcal{I}}), \quad (1)$$

where

$$l(\theta|U_N, Z_{\mathcal{I}}) = \log p(Z_{\mathcal{I}}|U_N, \theta) \quad (2)$$

is the log-likelihood function of θ , given the knowledge of U_N and $Z_{\mathcal{I}}$.

In Section 3.1 we propose an on-line method for solving (1), based on the expectation maximization (EM) algorithm. The advantage of this method is that it does not require an initialization. However, the EM algorithm is known to suffer from a slow convergence rate [11, pp. 358-359]. To cope with this, the estimate obtained after a few EM iterations is then used to initialize a gradient search algorithm [12]. To this end, we describe in Section 3.2 an on-line algorithm based on the quasi-Newton method.

Notation 1 We assume that the input signal U_N is known. Hence, to simplify the notation, we remove it from the list of conditioning variables in expressions like (1) and (2).

3.1. EM-based On-line Estimation Method

The EM algorithm solves the ML problem (1) using the following iterative procedure:

$$\hat{\theta}_N^{(i)} \in \arg \max_{\theta} Q(\theta, \hat{\theta}_N^{(i-1)}) \quad (3)$$

where the choice for $\hat{\theta}_N^{(i)}$ is arbitrary if the $\arg \max_{\theta}$ set is not a singleton, and $Q(\theta, \hat{\theta})$ is defined by

$$Q(\theta, \hat{\theta}) = \int \log p(Z_{\mathcal{I}}, Y_N|\theta) p(Y_N|Z_{\mathcal{I}}, \hat{\theta}) dY_N. \quad (4)$$

The iterations (3)-(4) permit computing $\hat{\theta}_N$, for a fixed N [11, Sec. 10.1]. To obtain an adaptive algorithm, we compute one iteration for each new available sample. Doing so we obtain the following

¹Notice that, for any function $f(\theta)$, the symbol $\arg \max_{\theta} f(\theta)$ denotes a set of θ values that maximize $f(\theta)$.

iterations:

$$\hat{\theta}_N \in \arg \max_{\theta} Q(\theta, \hat{\theta}_{N-1}). \quad (5)$$

While details are omitted, it can be shown that (5) becomes

$$\hat{\theta}_N \in \arg \min_{\theta} \sum_{t=1}^N \left(\bar{y}(t, \hat{\theta}_{N-1}) - x(t, \theta) \right)^2, \quad (6)$$

where

$$\bar{y}(t, \hat{\theta}) = \begin{cases} \mathcal{E} \left\{ y(t)|z(t), \hat{\theta} \right\}, & t \in \mathcal{I} \\ x(t, \hat{\theta}), & t \notin \mathcal{I} \end{cases} \quad (7)$$

and $\mathcal{E} \left\{ y(t)|z(t), \hat{\theta} \right\} = \int y(t)p(y(t)|z(t), \hat{\theta})dy(t)$.

The procedure in (6) requires tuning, at each iteration, the ARMA model parameters θ to fit $\bar{y}(t, \hat{\theta}_{N-1})$. For a fixed N , a number of recursive formulas are available for doing so [13]. We use the iterative weighted linear least squares algorithm [13]. Again, to obtain an adaptive algorithm, we compute one iteration for each new available sample. This results in the following iterations:

$$\begin{aligned} \hat{\theta}_N \in \arg \min_{\theta} \sum_{t=1}^N \left(\frac{A(q, \theta)\bar{y}(t, \hat{\theta}_{N-1})}{A(q, \hat{\theta}_{N-1})} - \frac{B(q, \theta)u(t)}{A(q, \hat{\theta}_{N-1})} \right)^2 \\ = \arg \min_{\theta} \sum_{t=1}^N \left(\bar{y}(t, \hat{\theta}_{N-1}) - \tilde{\phi}^T(t, \hat{\theta}_{N-1})\theta \right)^2, \end{aligned} \quad (8)$$

where

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

and $\Omega_n(q) = [1, q^{-1}, \dots, q^{-n}]^T$. Finally, an on-line version of (8), can be obtained using the recursive least-squares algorithm [3, Sec. 11].

3.2. Newton search algorithm

For a fixed N , the ML problem (1) can be solved using a gradient search algorithm. In this work we use a quasi-Newton method, whose adaptive version is given by the following iterations:

$$\hat{\theta}_N = \hat{\theta}_{N-1} - \mu_N T_N g_N. \quad (9)$$

The scalar μ_N denotes the step-size at iteration i , the matrix T_N denotes an approximation to the inverse of the Hessian of $l(\theta|Z_{\mathcal{I}})$ at $\hat{\theta}_{N-1}$ and the vector g_N denotes the gradient of $l(\theta|Z_{\mathcal{I}})$ at $\hat{\theta}_{N-1}$. To compute T_N we use the Broyden-Fletcher-Goldfarb-Shanno (BFGS) formula [12]. Also, the step-size parameter μ_N is obtained from a linear search algorithm where starting from the initial value $\mu_{N,1} = 1$, the value of $\mu_{N,i}$ is halved at each $i \in \mathbb{N}$ until

$$l(\hat{\theta}_{N-1} - \mu_{N,i} T_N g_N | Z_{\mathcal{I}}) < l(\hat{\theta}_{N-1} | Z_{\mathcal{I}}),$$

or a maximum number of sub-iterations is reached.

In order to implement the iterations (9), we need to provide an expression for $l(\theta|Z_{\mathcal{I}})$ and its gradient. Since $w(t)$ is a sequence of independent random variables, we have that

$$l(\theta|Z_{\mathcal{I}}) = \sum_{t \in \mathcal{I}} \log \int_{\mathcal{Q}_t^{-1}[z(t)]} p(y(t)|\theta) dy(t). \quad (10)$$

Also, it can be shown that,

$$\frac{\partial}{\partial \tilde{\theta}} l(\tilde{\theta} | Z_I) \Big|_{\theta} = \frac{1}{\sigma^2} \sum_{t \in \mathcal{I}} (\bar{y}(t, \theta) - x(t, \theta)) \psi(t, \theta),$$

where

$$\psi(t, \theta) = \frac{\partial}{\partial \tilde{\theta}} x(t, \tilde{\theta}) \Big|_{\theta} = \phi(q, \theta) u(t), \quad (11)$$

$$\phi(q, \theta) = \left[\frac{\Omega_n^T(q)}{A(q, \theta)}, \frac{B(q, \theta) q^{-1} \Omega_{m-1}^T(q)}{A^2(q, \theta)} \right]^T. \quad (12)$$

4. ASYMPTOTIC ANALYSIS

In this section we study the statistical properties of the estimate $\hat{\theta}_N$, when the number N of sample times tends to infinity. The next theorem provides conditions to guarantee that the estimate $\hat{\theta}_N$ converges to the true parameter vector θ_* , as the number of samples tends to infinity.

Theorem 1 *Let $\mathcal{D} \subset \mathbb{R}$ be a compact set containing the true parameter vector θ_* , 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 $0 < \epsilon < 1$. Let \mathcal{Q}_t be such that there exists $M_{\mathcal{Q}} > 0$ satisfying that $|b_{t,k}| < M_{\mathcal{Q}}$, for all $t \in \mathbb{Z}$ and $k = 1, \dots, K - 1$. Let $u(t)$ be bounded and such that $\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{t=1}^N (x(t, \theta) - x(t, \theta_*))^2 = 0$ holds if and only if $\theta = \theta_*$. If for each N , $\hat{\theta}_N$ is constrained to belong to \mathcal{D} , i.e., $\hat{\theta}_N \in \arg \max_{\theta \in \mathcal{D}} l(\theta | Z_I)$, then $\lim_{N \rightarrow \infty} \hat{\theta}_N = \theta_*$ w.p.1.*

Remark 1 *A strong consistency result for system identification without output quantization can be found in [3, Section 8]. When applied to the setting described in Section 2, this result states that consistency holds under the same assumptions as those in Theorem 1 (except for the condition on \mathcal{Q}_t). More precisely, our assumption on \mathcal{D} is equivalent to a condition called uniform stability in [3], and our assumption on the input signal is equivalent to the assumption called informative enough in [3]. Hence, Theorem 1 essentially states that placing a quantizer and having packet dropouts at the output does not alter strong consistency.*

The next theorem states conditions under which the random vector $\hat{\theta}_N - \theta_*$ converges in distribution to a multivariate normal vector.

Theorem 2 *Let*

$$\Phi_{\mu} = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{t=1}^N \mu(t) \psi(t, \theta_*) \psi^T(t, \theta_*), \quad (13)$$

where $\mu(t) = \frac{\bar{\sigma}^2(t)}{\sigma^2}$ and $\bar{\sigma}^2(t) = \mathcal{E} \{ (\bar{y}(t, \theta_*) - x(t, \theta_*))^2 \}$. If the conditions in Theorem 1 hold, and θ_* lies in the interior of \mathcal{D} , then Φ is invertible and

$$\sqrt{N} (\hat{\theta}_N - \theta_*) \rightarrow \mathcal{N}(0, C) \text{ in distribution}, \quad (14)$$

with

$$C = \frac{\sigma^2}{\lambda} \Phi_{\mu}^{-1}. \quad (15)$$

The asymptotic covariance C stated in Theorem 2 depends, via Φ_{μ} , on the particular input signal $u(t)$ used for identification. If the input is assumed to be a wide-sense stationary, ergodic and bounded random process, then the result is still valid with

$$\Phi_{\mu} = \mathcal{E}_u \left\{ \mu(t) \psi(t, \theta_*) \psi^T(t, \theta_*) \right\}, \quad (16)$$

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

5. QUANTIZER DESIGN

5.1. Optimum Time-Varying Quantization Scheme

The result in Theorem 2 can be used to choose the boundaries $b_{t,k}$, $k = 0, \dots, K$ of the quantizer \mathcal{Q}_t to minimize the asymptotic error covariance C . Notice that this is equivalent to choosing the boundaries so that $\bar{\sigma}^2(t)$ is maximized. We have that

$$\begin{aligned} \bar{\sigma}^2(t) &= \mathcal{E} \left\{ (\bar{\mathcal{Q}}_t [y(t, \theta_*)] - x(t, \theta_*))^2 \right\} \\ &= \mathcal{E} \left\{ \left(\bar{\mathcal{Q}}_{t,x(t, \theta_*)} [w(t)] \right)^2 \right\}, \end{aligned} \quad (17)$$

with $\bar{\mathcal{Q}}_{t,x}$ being the quantizer defined by

$$\begin{aligned} \bar{\mathcal{Q}}_{t,x} [w] &= \bar{\mathcal{Q}}_t [x + w] - x \\ \bar{\mathcal{Q}}_t [y] &= \mathcal{E} \{ y | y \in [b_{k-1}, b_k] \}, \text{ if } y \in [b_{k-1}, b_k]. \end{aligned} \quad (18)$$

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 $\bar{\mathcal{Q}}_{t,x(t, \theta_*)}$. Then, the boundaries $b_{t,k}$, $k = 0, \dots, K$ need to be chosen so that the quantizer $\bar{\mathcal{Q}}_{t,x(t, \theta_*)}$ maximizes $\bar{\sigma}^2(t)$. It can be shown that this is asymptotically achieved when, for each $k = 0, \dots, K$,

$$b_{t,k} = \tilde{b}_k + x(t, \theta_*), \quad (19)$$

where \tilde{b}_k , $k = 0, \dots, K$ are the boundaries of a Lloyd's quantizer designed using the PDF of $w(t)$. By using this design, the expression for C in Theorem 2 becomes

$$C = \frac{\sigma^2}{\lambda \mu} \Phi^{-1}, \quad (20)$$

where

$$\Phi = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{t=1}^N \psi(t, \theta_*) \psi^T(t, \theta_*), \quad (21)$$

$$\mu = \frac{\bar{\sigma}^2}{\sigma^2} \text{ and } \bar{\sigma}^2 = \mathcal{E} \left\{ \bar{\mathcal{Q}}^2 [w(t)] \right\}.$$

Remark 2 *Notice that equation (20) differs from the classical result of system identification [3, 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. Thus, the effects of finite-level quantization and packet drops on identification accuracy are exactly quantified.*

If $\lambda = 1$, θ is a scalar, $x(t, \theta) = \theta$ and \mathcal{Q}_t is a one-bit quantizer, it follows from (19) that the optimal quantization threshold equals θ . Then, it can be verified that $C = \frac{\pi}{2} \sigma^2$, which is in agreement with the result derived in [4], by setting the quantizer threshold as the true parameter. Notice that they fail to give a practical quantizer to asymptotically approach this minimum MSE.

5.2. Adaptive Quantization Scheme for System Identification

In Section 5.1 we pointed out that the asymptotic covariance matrix C is minimized when the boundaries $b_{t,k}$, $k = 0, \dots, K$ of the quantizer \mathcal{Q}_t are chosen according to (19). The problem with (19) is that it requires the knowledge of the true parameter vector θ_* , which is unavailable to the quantizer design. A practical workaround is to replace θ_* 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 value of the asymptotic covariance matrix C can still be achieved in this case. Or in other words, whether the result in (20) still holds if (19) is replaced by

$$b_{t,k} = \tilde{b}_k + x(t, \hat{\theta}_{t-1}). \quad (22)$$

The answer turns out to be positive, but its proof is omitted.

6. SIMULATIONS

In this section we evaluate the evolution of the trace $\text{Tr}\{C_N\}$ of the covariance C_N of $\sqrt{N}(\hat{\theta}_N - \theta_*)$. To this end we assume that the input signal $u(t)$ is an i.i.d. random process. The distribution of each sample 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 consider three different two-bit quantizers (i.e., $K = 4$). The first quantizer \mathcal{Q}_s is stationary and designed using Lloyd's algorithm applied to the distribution $\mathcal{N}(0, 1)$. The second quantizer \mathcal{Q}_o is the optimal quantizer given by (19), and the third \mathcal{Q}_a is the adaptive (asymptotically optimal) quantizer given by (22). The true system is given by

$$\frac{B(q)}{A(q)} = \frac{1}{1 - 1.764q^{-1} + 0.81q^{-2}}.$$

Also, the noise power is $\sigma^2 = 0.1$ and the packet drop rate is $\lambda = 0.9$. For each quantization scheme, we compute $\text{Tr}\{C_N\}$ using 1000 Monte Carlo runs. The result is shown in Figure 1, which also shows the theoretical asymptotic values $\text{Tr}\{C_s\}$ and $\text{Tr}\{C_o\}$ (obtained using (15) and (16)), 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 $\text{Tr}\{C_N\}$ which is about 8 dB smaller than the one resulting from \mathcal{Q}_s . Also, the adaptive quantizer \mathcal{Q}_a achieves the optimal value $\text{Tr}\{C_o\}$ in the limit.

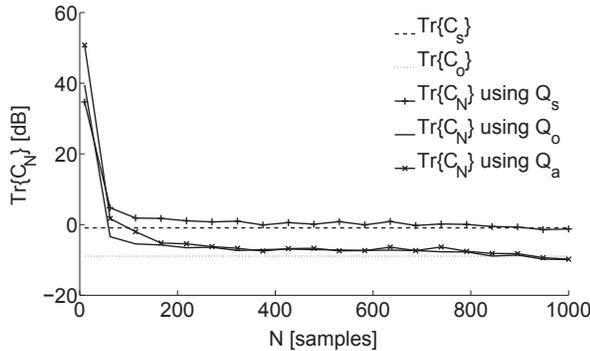


Fig. 1. Comparison of convergence rates for different output quantizers.

7. CONCLUSION

Motivated by identification problem over noisy digital communication channels, this paper has proposed a simple adaptive quantizer and the corresponding recursive identification algorithm to address the joint effects of finite-level quantization and packet dropouts on identification accuracy. The proposed algorithm is shown to be optimal in the sense of asymptotically achieving the minimum MSE. Simulation results are included to validate the identification algorithm.

8. REFERENCES

- [1] L. Wang, G. Yin, J. Zhang, and Y. Zhao, *System identification with quantized observations*, Birkhäuser Boston, 2010.
- [2] J. Xiao, A. Ribeiro, Z. Luo, and G. Giannakis, "Distributed compression-estimation using wireless sensor networks," *IEEE Signal Processing Magazine*, vol. 23, no. 4, pp. 27–41, 2006.
- [3] Lennart Ljung, *System Identification: Theory for the User*, Prentice Hall, Upper Saddle River, NJ, second edition, 1999.
- [4] A. Ribeiro and G. Giannakis, "Bandwidth-Constrained Distributed Estimation for Wireless Sensor Networks-Part I: Gaussian Case," *IEEE Transactions on Signal Processing*, vol. 54, no. 3, pp. 1131–1143, 2006.
- [5] H.C. Papadopoulos, G.W. Wornell, and A.V. Oppenheim, "Sequential signal encoding from noisy measurements using quantizers with dynamic bias control," *IEEE Transactions on Information Theory*, vol. 47, no. 3, pp. 978–1002, 2002.
- [6] J. Fang and H. Li, "Distributed adaptive quantization for wireless sensor networks: From delta modulation to maximum likelihood," *IEEE Transactions on Signal Processing*, vol. 56, no. 10, pp. 5246–5257, 2008.
- [7] K. You and L. Xie, "Minimum data rate for mean square stabilization of discrete LTI systems over lossy channels," *IEEE Transactions on Automatic Control*, vol. 55, no. 10, pp. 2373–2378, 2010.
- [8] B. Sinopoli, L. Schenato, M. Franceschetti, K. Poolla, M. Jordan, and S. Sastry, "Kalman filtering with intermittent observations," *IEEE Transactions on Automatic Control*, vol. 49, no. 9, pp. 1453–1464, 2004.
- [9] K. You, M. Fu, and L. Xie, "Necessary and sufficient conditions for stability of Kalman filtering with Markovian packet losses," *submitted to Proc. 18th IFAC World Congress, Milano, Italy*, 2010.
- [10] Y. Mo and B. Sinopoli, "A characterization of the critical value for Kalman filtering with intermittent observations," in *47th IEEE Conference on Decision and Control*, 2008, pp. 2692–2697.
- [11] D. Cappé, E. Moulines, and T. Rydén, *Inference in Hidden Markov Models*, Springer, 2005.
- [12] R. Fletcher, *Practical Methods for optimization*, John Wiley & Sons, Inc., 1987.
- [13] R. Pintelon, P. Guillaume, Y. Rolain, J. Schoukens, and H. Van hamme, "Parametric identification of transfer functions in the frequency domain," *IEEE Transactions on Automatic Control*, vol. 39, no. 11, pp. 2245 – 2260, 1994.