Parameter estimation in state-space models having quantized output data

Damián E. Marelli, Boris I. Godoy and Graham C. Goodwin

Abstract—In this paper, we describe an algorithm to estimate the parameters of a linear system in state-space form having quantized measurements. The estimation is carried out by posing the problem in the maximum likelihood framework and solving the problem by using the Expectation Maximization (EM) algorithm to maximize the likelihood function. The a posteriori probability density function found in the EM algorithm is calculated using Sequential Montecarlo Methods, known as particle filters.

I. INTRODUCTION

The problem of estimating the parameters in a linear system having quantized data has been of recent interest and it is technically challenging [1], [2].

The present paper presents an algorithm to estimate the parameters of a system in state-space form using quantized measurements. The current paper extends the results presented in [?] by considering a more general system. One advantage of considering the system in state-space form is, for example, the handling of multivariable data. The configuration to be studied is shown in Fig. 1.

Fig. 1. FIR system with quantized output.

In certain applications, one is forced to work with quantized data transmitted through a communication channel. For example, in Networked Control Systems (NCSs) and in sensor networks, data is typically quantized to minimize communication resource utilization [3]. The motivation for the configuration under study as given in Fig. 1 matches these areas. It may be used, for example, to describe the problem of identifying a linear plant, whose outputs are quantized and which are then transmitted through a communication channel.

In this paper, we pose the problem of estimating the parameters of a linear system in a Maximum Likelihood (ML) framework. One problem in the use of the ML framework is that the resulting likelihood function may be difficult to compute. This drawback can be overcome by

using the Expectation Maximization (EM) algorithm [4]. This algorithm can be used as an iterative method to find a local maximum of the likelihood function.

One difficulty in using the EM algorithm for our problem is the high complexity associated with the computation of the conditional probability density function (pdf) of the hidden variable, X , given the (quantized) measurements, Z . To address this issue we approximate this function using Sequential Montecarlo Methods (SMM) - known as particle filters. Particle Filters have been widely used in the context of system identification [5].

The remainder of the paper is organized as follows: In Section II, we summarize the problem of interest. In Section III, we present the Expectation Maximization (EM) algorithm. In Section IV, we show how to solve the ML problem by using the EM algorithm. In Section VI, we show a simple example of using the new algorithm presented in this paper. Finally, in Section VII, we present final remarks on the current work.

II. STATEMENT OF THE PROBLEM

We focus on the problem of estimating the parameters of a state-space system having quantized measurements. The scheme we study in the present paper considers the model:

$$x_{n+1} = Ax_n + Bu_n + v_n \quad (1)$$

$$y_n = Cx_n + Du_n + w_n \quad (2)$$

$$z_n = Q[y_n] \quad (3)$$

where $u_n \in \mathbb{R}^p$, $x_n \in \mathbb{R}^q$, $y_n \in \mathbb{R}^r$ and $z_n \in \mathbb{R}^r$. Also, $x_1 \sim \mathcal{N}(\mu, P)$, $v_n \sim \mathcal{N}(0, R)$, $w_n \sim \mathcal{N}(0, S)$, for all $n \in \mathbb{Z}$ forms a set of independent random variables.

Notation 1: For notational convenience we write (1)-(2) as

$$\xi_n = \Gamma \zeta_n + \eta_n$$

with $\xi_n = [x_{n+1}^T, y_n^T]^T$, $\zeta_n = [x_n^T, u_n^T]^T$, $\eta_n = [v_n^T, w_n^T]^T$ and

$$\Gamma = \begin{bmatrix} A & B \\ C & D \end{bmatrix}.$$

We also write

$$\Sigma = \begin{bmatrix} R & 0 \\ 0 & S \end{bmatrix},$$

and $U^N = \{u_1, \dots, u_N\}$, $X^N = \{x_1, \dots, x_N\}$, $Y^N = \{y_1, \dots, y_N\}$, $Z^N = \{z_1, \dots, z_N\}$, $\xi^N = \{\xi_1, \dots, \xi_N\}$, $\zeta^N = \{\zeta_1, \dots, \zeta_N\}$.

B.I. Godoy (corresponding author), D. E. Marelli, G.C. Goodwin are with the ARC Centre of Excellence for Complex Dynamic Systems and Control, The University of Newcastle, Australia. Emails: damian.marelli, boris.godoy@newcastle.edu.au; juan.aguero, graham.goodwin@newcastle.edu.au

III. THE MAXIMUM LIKELIHOOD CRITERION

Suppose we know the input u_n and the output z_n up to time N , and we want to use this knowledge to find an estimate θ^N of some set of parameters θ . Using the maximum likelihood criterion this is achieved as follows:

$$\hat{\theta} = \arg \max_{\theta} p(Z^N | \theta). \quad (4)$$

In our particular case, the joint p.d.f for the data Z^N is given by

$$p(Z^N | \theta) = \prod_{n=1}^N p(z_n | \theta). \quad (5)$$

Hence, the ML estimator satisfies:

$$\hat{\theta} = \arg \max_{\theta \in \Omega} L(\theta) = \arg \max_{\theta \in \Omega} \sum_{n=1}^N \log p(z_n | \theta). \quad (6)$$

The function $p(z_n | \theta)$ can be obtained recursively as:

$$p(z_n | \theta) = p(z_n | y_n, \theta) p(x_{n+1}, y_n | x_n, \theta) \quad (7)$$

We note that $(x_{n+1}, y_n | x_n, \theta) \sim \mathcal{N}(\Gamma \zeta_n, \Pi)$. Therefore, the pdf of $(x_{n+1}, y_n | x_n, \theta)$ can be expressed as

$$p(x_{n+1}, y_n | x_n, \theta) = \frac{1}{\sqrt{2\pi |\Pi|^{q+r}}} e^{-\frac{1}{2}(\xi_n - \Gamma \zeta_n)^T \Pi^{-1} (\xi_n - \Gamma \zeta_n)}. \quad (8)$$

Also, we have that the pdf for $(z_n | y_n, \theta)$ is given by

$$p(z_n | y_n, \theta) = \begin{cases} 1, & z_n = \mathcal{Q}[y_n] \\ 0, & z_n \neq \mathcal{Q}[y_n] \end{cases}. \quad (9)$$

Therefore, $p(z_n | \theta)$ is given by

$$\begin{aligned} p(z_n | \theta) &= \frac{1}{\sqrt{2\pi |\Pi|^{q+r}}} \int \int p(z_n | y_n, \theta) \\ &\quad \exp \left\{ -\frac{1}{2}(\xi_n - \Gamma \zeta_n)^T \Pi^{-1} (\xi_n - \Gamma \zeta_n) \right\} dx_n dy_n \\ &= \frac{1}{\sqrt{2\pi |\Pi|^{q+r}}} \int_{y_n: p(z_n | y_n, \theta)=1} \int \\ &\quad \exp \left\{ -\frac{1}{2}(\xi_n - \Gamma \zeta_n)^T \Pi^{-1} (\xi_n - \Gamma \zeta_n) \right\} dx_n dy_n. \end{aligned} \quad (10)$$

Usually, the likelihood function (10) is maximized using gradient-based methods or similar methods. However, for the case of quantized outputs, the likelihood function cannot easily be optimized because it requires the evaluation of several integrals. A more suitable and easier way to optimize the ML function is by using the Expectation Maximization algorithm [6].

IV. THE EXPECTATION MAXIMIZATION ALGORITHM

The EM algorithm utilizes [7] a *hidden*¹ variable, say X_N . The basic idea of the EM algorithm is to iterate by first finding an estimate of X_N , say $\hat{X}_N^i$ using a trial value

¹The terminology arises from the statistics literature.

of θ , say $\hat{\theta}_i$, and then maximizing a quantity related to $p(Y_N, \hat{X}_N^i | \theta)$ with respect to θ .

To develop the idea, we begin by considering the joint density function for Y_N and X_N . After taking logarithms, we obtain

$$\log p(Y_N | \theta) = \log p(Y_N, X_N | \theta) - \log p(X_N | Y_N, \theta). \quad (11)$$

Assume that, we have an estimate of θ , say $\hat{\theta}_i$, then, the likelihood function $L(\theta)$ can be expressed as

$$L(\theta) = \log p(Y_N | \theta) = E \{ \log [p(Y_N | \theta)] | Y_N, \hat{\theta}_i \}.$$

Using (11), we obtain

$$L(\theta) = \mathbf{Q}(\theta, \hat{\theta}_i) - \mathbf{H}(\theta, \hat{\theta}_i),$$

where $\mathbf{Q}$ and $\mathbf{H}$ are defined as

$$\mathbf{Q}(\theta, \hat{\theta}_i) := E \{ \log [p(Y_N, X_N | \theta)] | Y_N, \hat{\theta}_i \}$$

$$\mathbf{H}(\theta, \hat{\theta}_i) := E \{ \log [p(X_N | Y_N, \theta)] | Y_N, \hat{\theta}_i \}.$$

The EM algorithm may then be summarised as follows [7]:

- 1) Choose an initial estimate $\hat{\theta}_0 \in \Omega$, where Ω is a constraint set in the parameter space. Then, for $i = 0, 1, \dots$
- 2) E-step: Evaluate $p(X_N | Y_N, \hat{\theta}_i)$. Then, compute the auxiliary function $\mathbf{Q}(\theta, \hat{\theta}_i)$ which is the expected value (given the observed data Y_N and the previous estimate $\hat{\theta}_i$) of the complete data log-likelihood:

$$\mathbf{Q}(\theta, \hat{\theta}_i) = E_{X_N} \{ \log [p(X_N, Y_N | \theta)] | Y_N, \hat{\theta}_i \}, \quad (12)$$

- 3) M-step: Set $\hat{\theta}_{i+1} = \arg \max_{\theta \in \Omega} \mathbf{Q}(\theta, \hat{\theta}_i)$.
- 4) Go to step 2, and continue until convergence.

Under quite general conditions, the EM algorithm can be proven to converge to a stationary point of the likelihood function [7].

Notice that direct application of the EM algorithm to quantized data estimation is computationally expensive due to the need to evaluate the conditional distribution of the hidden variable given the (quantized) measurements for the current value of the parameter estimates.

V. APPLICATION OF THE EM ALGORITHM

The expectation maximization (EM) method recursively solves (4) as follows:

$$\theta_i^N = \arg \max_{\theta} Q(\theta, \theta_{i-1}^N) \quad (13)$$

$$Q(\theta, \hat{\theta}) = \mathcal{E} \left\{ \log p(X^{N+1}, Y^N, Z^N | \theta) | Z^N, \hat{\theta} \right\} \quad (14)$$

We find expressions for (13) and (14) below. To this end we split our analysis into four steps. The first three steps are independent of the choice of θ , while the fourth considers the case $\theta = \Gamma$.

1) **Computing** $p(X^{N+1}, Y^N, Z^N | \theta)$: We have that

$$\begin{aligned}
& p(X^{N+1}, Y^N, Z^N | \theta) \\
&= \prod_{n=1}^N p(x_{n+1}, y_n, z_n | X^n, Y^{n-1}, Z^{n-1}, \theta) \\
&= \prod_{n=1}^N p(x_{n+1}, y_n, z_n | x_n, \theta) \\
&= \prod_{n=1}^N p(z_n | x_n, x_{n+1}, y_n, \theta) p(x_{n+1}, y_n | x_n, \theta) \\
&= \prod_{n=1}^N p(z_n | y_n, \theta) p(x_{n+1}, y_n | x_n, \theta)
\end{aligned}$$

The last expression can be calculated by using equations (8) and (9).

2) **Computing** $\log p(X^{N+1}, Y^N, Z^N | \theta)$: We have that

$$\begin{aligned}
& -\frac{2}{N} \log p(X^{N+1}, Y^N, Z^N | \theta) \\
&= \frac{1}{N} \sum_{n=1}^N \log p(z_n | y_n) + \frac{1}{N} \sum_{n=1}^N \log p(x_{n+1}, y_n | x_n, \theta) \\
&= \frac{1}{N} \sum_{n=1}^N \log p(z_n | y_n) + \\
&\log(2\pi |\Pi|^{q+r}) + \frac{1}{N} \sum_{n=1}^N (\xi_n - \Gamma \zeta_n)^T \Pi^{-1} (\xi_n - \Gamma \zeta_n) \\
&= \frac{1}{N} \sum_{n=1}^N \log p(z_n | y_n) + \log(2\pi |\Pi|^{q+r}) + \\
&T_r \left\{ \Pi^{-1} \frac{1}{N} \sum_{n=1}^N (\xi_n - \Gamma \zeta_n) (\xi_n - \Gamma \zeta_n)^T \right\}
\end{aligned} \tag{15}$$

3) **Computing** $Q(\theta, \hat{\theta})$: For all $\hat{\theta}$, all $n = 1, \dots, N$, and all y_n satisfying $p(y_n | Z^N, \hat{\theta}) \neq 0$, we have that $p(z_n | y_n) = 1$. Then,

$$\begin{aligned}
& -\frac{2}{N} Q(\theta, \hat{\theta}) \\
&= \mathcal{E} \left\{ -2 \log p(X^{N+1}, Y^N, Z^N | \theta) | Z^N, \hat{\theta} \right\} \\
&= \frac{1}{N} \sum_{n=1}^N \mathcal{E} \left\{ \log p(z_n | y_n) | Z^N, \hat{\theta} \right\} \\
&+ \log(2\pi |\Pi|^{q+r}) + T_r \left\{ \Pi^{-1} \frac{1}{N} \sum_{n=1}^N \mathcal{E} \{ (\xi_n - \Gamma \zeta_n) \right. \\
&\quad \left. (\xi_n - \Gamma \zeta_n)^T | Z^N, \hat{\theta} \} \right\} \\
&= \log(2\pi |\Pi|^{q+r}) \\
&+ T_r \left\{ \Pi^{-1} [\Phi - \Psi \Gamma^T - \Gamma \Psi^T + \Gamma \Sigma \Gamma^T] \right\}
\end{aligned} \tag{16}$$

where

$$\Phi = \frac{1}{N} \sum_{n=1}^N \mathcal{E} \left\{ \xi_n \xi_n^T | Z^N, \hat{\theta} \right\}, \tag{17}$$

$$\Psi = \frac{1}{N} \sum_{n=1}^N \mathcal{E} \left\{ \xi_n \zeta_n^T | Z^N, \hat{\theta} \right\}, \tag{18}$$

$$\Sigma = \frac{1}{N} \sum_{n=1}^N \mathcal{E} \left\{ \zeta_n \zeta_n^T | Z^N, \hat{\theta} \right\}. \tag{19}$$

4) **Computing** $\arg \max_{\theta} Q(\theta, \hat{\theta})$ in the case $\theta = \Gamma$: We have that

$$\begin{aligned}
& \arg \max_{\theta} Q(\theta, \hat{\theta}) \\
&= \arg \min_{\Gamma} T_r \left\{ \Pi^{-1} [\Phi - \Psi \Gamma^T - \Gamma \Psi^T + \Gamma \Sigma \Gamma^T] \right\} \\
&= \arg \min_{\Gamma} T_r \left\{ \Pi^{-1} [(\Gamma - \Psi \Sigma^{-1}) \Sigma (\Gamma - \Psi \Sigma^{-1})^T + \right. \\
&\quad \left. + \Phi - \Psi \Sigma^{-1} \Psi] \right\} \\
&= \arg \min_{\Gamma} T_r \left\{ \Pi^{-1/2} (\Gamma - \Psi \Sigma^{-1}) \Sigma (\Gamma - \Psi \Sigma^{-1})^T \Pi^{-1/2} \right\}
\end{aligned}$$

which is clearly minimized by

$$\theta = \Gamma = \Psi \Sigma^{-1}. \tag{20}$$

A. *Computing Ψ and Σ*

In order to compute (20) we need to compute Ψ and Σ from (18) and (19). This requires the knowledge of the statistics of ξ_n and ζ_n conditioned to the quantized observations Z^N , which in turn requires the knowledge of $p(X^{N+1}, Y^N | Z^N, \hat{\theta})$. We address this problem below. From now on, to simplify the notation we remove $\hat{\theta}$ from the conditioning variables.

1) *Recursive computation of $p(X^{N+1}, Y^N | Z^N)$* : We have that

$$\begin{aligned}
p(X^{N+1}, Y^N | Z^N) &= p(Y^N | X^{N+1}, Z^N) \\
&= p(X^{N+1} | Z^N) \\
&= p(Y^N | X^N, Z^N) \\
&= p(X^{N+1} | Z^N),
\end{aligned} \tag{21}$$

and $p(X^{N+1} | Z^N)$ can be recursively computed as follows:

$$\begin{aligned}
p(X^N | Z^N) &= p(X^N | Z^{N-1}, z_N) \\
&= \frac{p(X^N, z_N | Z^{N-1})}{p(z_N | Z^{N-1})} \\
&= \frac{p(z_N | X^N, Z^{N-1})}{p(z_N | Z^{N-1})} p(X^N | Z^{N-1}) \\
&= \frac{p(z_N | x_N)}{p(z_N | Z^{N-1})} p(X^N | Z^{N-1})
\end{aligned} \tag{22}$$

and

$$\begin{aligned}
p(X^{N+1} | Z^N) &= p(x_{N+1}, X^N | Z^N) \\
&= p(x_{N+1} | X^N, Z^N) p(X^N | Z^N) \\
&= p(x_{N+1} | x_N) p(X^N | Z^N).
\end{aligned} \tag{23}$$

2) *Approximating $p(X^{N+1}, Y^N | Z^N)$ using particles:* The idea consists of approximating the distribution $p(X^{N+1}, Y^N | Z^N)$ by a weighted sum of impulses. We detail below a recursive method for doing so.

Suppose we have the following approximation of $p(X^N | Z^{N-1})$,

$$p(X^N | Z^{N-1}) \simeq \frac{1}{I} \sum_{i=1}^I \delta(X^N - X_i^N),$$

from (23), we have that

$$\begin{aligned} p(X^N | Z^N) &\simeq \frac{p(z_N | x_N)}{p(z_N | Z^{N-1})} \frac{1}{I} \sum_{i=1}^I \delta(X^N - X_i^N) \\ &= \frac{1}{I} \sum_{i=1}^I w_{N,i} \delta(X^N - X_i^N) \end{aligned} \quad (24)$$

where for each $i = 1, \dots, I$,

$$\begin{aligned} w_{N,i} &\propto p(z_N | x_{N,i}) \\ &= \int p(z_N | x_{N,i}, y_N) p(y_N | x_{N,i}) dy_N \\ &= \int p(z_N | y_N) p(y_N | x_{N,i}) dy_N \\ &= \int_{S_N} p(y_N | x_{N,i}) dy_N, \end{aligned} \quad (25)$$

with $S_N = \mathcal{Q}^{-1}[z_N]$ and $p(y_N | x_{N,i}) \sim \mathcal{N}(Cx_{N,i} + Du_N, S)$. We then obtain a new set of particles $\tilde{X}_i^N$, $i = 1, \dots, N$, by drawing them from the distribution (24). By doing so (24) becomes

$$p(X^N | Z^N) \simeq \frac{1}{I} \sum_{i=1}^I \delta(X^N - \tilde{X}_i^N). \quad (26)$$

To complete the iteration, using (23), we write

$$p(X^{N+1} | Z^N) \simeq \frac{1}{I} \sum_{i=1}^I \delta(X^{N+1} - X_i^{N+1}),$$

with

$$X_i^{N+1} = \{\tilde{X}_i^N, x_{N+1,i}\} \quad (27)$$

and, for each $i = 1, \dots, I$,

$$x_{N+1,i} \sim \mathcal{N}(Ax_{N,i} + Bu_N, R). \quad (28)$$

Finally, to compute $p(X^{N+1}, Y^N | Z^N)$ At each iteration notice that

$$p(Y^N | X^N, Z^N) = \prod_{n=1}^N p(y_n | x_n, z_n).$$

Hence, from (21), we can write where

$$\{X_i^{N+1}, Y_i^N\} = \{X_i^{N+1}, Y_i^{N-1}, y_{N,i}\} \quad (29)$$

and,

$$y_{N,i} \sim p(y_N | x_N, z_N). \quad (30)$$

Remark 1: Equations (29)-(28) mean that the particles $\{X_i^{N+1}\}$ and $\{X_i^{N+1}, Y_i^N\}$, $i = 1, \dots, I$ can be iteratively

generated by adding, at iteration N , the extra components generated by (30) and (28), to the available particles $\{X_i^N\}$ and $\{X_i^{N+1}, Y_i^{N-1}\}$, respectively.

Remark 2: The recursive equations above permit an on-line implementation of the estimate $p(X^N, Y^N | Z^N)$, as N increases. To avoid the estimation of the whole $\{X_i^N, Y_i^N\}$ for large N , it can be reduced to that of the last L samples, where the *smoothing lag* L is chosen longer than the impulse response of the system to be identified.

3) *Computing Ψ and Σ :*

Notation 2: For each $i = 1, \dots, I$, we denote $\xi_i^N = \{\xi_{1,i}, \dots, \xi_{N,i}\}$ and $\zeta_i^N = \{\zeta_{1,i}, \dots, \zeta_{N,i}\}$, with $\xi_{n,i} = [x_{n+1,i}^T, y_{n,i}^T]^T$ and $\zeta_{n,i} = [x_{n,i}^T, u_n^T]^T$.

Using (??) we have that

$$\begin{aligned} \Psi &= \frac{1}{N} \sum_{n=1}^N \mathcal{E} \{ \xi_n \zeta_n^T | Z^N \} \\ &= \frac{1}{N} \sum_{n=1}^N \int \xi_n \zeta_n^T p(x_{n+1}, y_n | Z^N) d(x_{n+1}, y_n) \\ &= \frac{1}{N} \sum_{n=1}^N \frac{1}{I} \sum_{i=1}^I \xi_{n,i} \zeta_{n,i}^T, \end{aligned}$$

and

$$\begin{aligned} \Sigma &= \frac{1}{N} \sum_{n=1}^N \mathcal{E} \{ \zeta_n \zeta_n^T | Z^N \} \\ &= \frac{1}{N} \sum_{n=1}^N \int \zeta_n \zeta_n^T p(x_n | Z^N) dx_n \\ &= \frac{1}{N} \sum_{n=1}^N \frac{1}{I} \sum_{i=1}^I \zeta_{n,i} \zeta_{n,i}^T. \end{aligned}$$

VI. EXAMPLE

Consider the following system in a state-space representation

$$\begin{aligned} x_{k+1} &= \begin{bmatrix} -0.5 & 9 \\ 0 & -0.4 \end{bmatrix} x_k + \begin{bmatrix} -1 \\ 2 \end{bmatrix} u_k + w_k \\ y_k &= [2 \quad 0.5] x_k + v_k, \end{aligned} \quad (31)$$

where w_k and v_k are gaussian

The simulation conditions are set as follows:

- The initial value for the parameters is chosen from a subspace approach, considering the input and quantized output.
- The number of iterations for the EM algorithm is chosen to be 1000
- The number of particles is chosen 500
- The number of quantization bits is chosen equal to 3.

VII. CONCLUSIONS

REFERENCES

- [1] L. Wang, J. Zhang, and G.G. Yin. System identification using binary sensors. *IEEE Trans. on Aut. Control*, 48(11):1892–1907, 2003.
- [2] F. Gustafsson and R. Karlsson. Statistical results for system identification based on quantized observations. *Automatica*, (45):2794–2801, 2009.

- [3] G. N. Nair, F. Fagnani, S. Zampieri, and R. J. Evans. Feedback control under data rate constraints: An overview. *Proceeding of the IEEE*, 95(1):108–137, 2007.
- [4] M. Sanjeev Arulampalam, S. Maskell, N. Gordon, and T. Clapp. A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking. *IEEE Trans. on Signal Process.*, 50(2):174–188, 2002.
- [5] N. Kantas, A. Doucet, S.S. Singh, and J.M. Maciejowsky. An overview of Sequential Monte Carlo Methods for parameter estimation in general state-space models. In *15th IFAC Symposium on System Identification*, Saint-Malo, France, 2009.
- [6] G.J. McLachlan and T. Krishnan. *The EM algorithm and extensions*. Wiley Interscience, 2008.
- [7] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Statistical Society, Series B*, 39(1):1–38, 1977.