INFORMATION THEORETIC QUANTISER DESIGN FOR DECENTRALISED ESTIMATION OF HIDDEN MARKOV MODELS

Subhrakanti Dey, Ferdinando A. Galati

Department of Electrical and Electronics Engineering
The University of Melbourne, Parkville, Victoria 3010, Australia.

ABSTRACT

Quantiser design for a nonlinear filter is considered in the context of a decentralised estimation system with communication constraints. The filter is based on quantised outputs of a discrete-time, two-state Hidden Markov Model (HMM) as measured by two remote sensor nodes. The optimal quantisation scheme is obtained by maximising the mutual information between the quantised meaurements and the hidden Markov states. Filter performance is measured in terms of the probability of estimation error and is investigated through simulation for HMM's with both independent and correlated white Gaussian noise in the measurements. The performance of the filter based on continuous, unquantised signals provides a benchmark for the performance of the filter based on quantised measurements. Therefore a method for computing the probability of estimation error directly for the continuous filter is also presented.

1. INTRODUCTION

In Communications systems, there often arises the need to quantise data before further processing or transmission, for example, in bandwith limited communications channels such as might be found in decentralised estimations systems like the one shown in Figure 1. The system consists of two remote processing nodes and a fusion center. In this work, X_k is a discrete time, two-state Markov chain, and the fusion center estimates the hidden state X_k based on quantised inputs $\bar{Y}_k^j$ obtained from the continuous sensor outputs $Y_k^j = X_k + V_k^j, j = 1, 2$ up to time k. The noises V_k^j are asssumed to be sequences of independent identically distributed (*i.i.d*) random variables. The fact that Hidden Markov Processes can model many nonlinear systems well provides the motivation for studying them in this context.

Quantiser design for linear decentralised estimation systems has been investigated by many authors, for example [1, 2, 3]. For detecting deterministic signals in additive white Gaussian noise, it has been shown [4] that using identical binary sensors is asymptotically optimal as the number of observations per sensor goes to infinity. For nonlinear filters, quantiser design based on minimising the information loss characterised by means of a functional central limit theorem, has been investigated in [5]. A more direct approach for HMM's, based on maximising the mutual information between the quantised measurement and the hidden Markov

state was investigated in [6], and it is this work which is generalised here in the context of decentralised estimation.
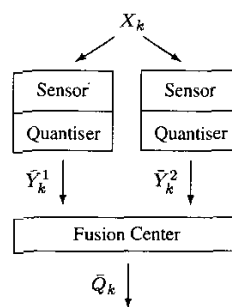

Fig. 1. Decentralised Estimation System.

In particular, the methods for quantiser design, and for directly computing the estimation error are generalised to the case of two sensors (Section 3). Also, through simulation studies (Section 4) the following three questions are addressed. (1) What happens to the performance of the continuous filter when the noise level in one of the sensors becomes large? (2) If one sensor has a restriction on its available bandwidth, and thus, on the number of its quantisation levels, then can any significant improvements in the performance of the quantised filter be achieved by adding a second sensor with higher noise level, but perhaps with more available bandwidth? (3) In the case of independent noise, is there a simple relationship between the quantisation scheme obtained jointly for the two sensors and the quantisations schemes obtained individually for each sensor?

Simulations were performed for HMM's with white Gaussian noise. With regard to questions (1) it was observed in the case of correlated noise, that there exists a critical signal to noise ratio beyond which filter performance improves as the signal to noise ratio increases. In regard to question (2) it was observed again in the case of correlated noise that in general, improvements in filter performance cannot be guaranteed simply by the addition of a second sensor.

2. MODEL DEFINITION

In the following, all stochastic processes are defined on the probability space $(\Omega, \mathcal{F}, P)$. Consider a discrete time, two-state Markov Chain $\{X_k\}$ with state space $S \doteq \{x_1, x_2\}$, and transition probability matrix $A = (a_{ij})$, where $a_{ij} = P(X_k = x_j \mid X_{k-1} = x_i)$

for $i, j = 1, 2$. Without loss of generality it is assumed that $x_i = (-1)^i$ for $i = 1, 2$. To simplify the computations it is also assumed that $\{X_k\}$ has symmetric transition matrix $A = \begin{bmatrix} 1-\alpha & \alpha \\ \alpha & 1-\alpha \end{bmatrix}$, where $0 < \alpha < 1$. The sensor output at time k, is given by $Y_k = (Y_k^1, Y_k^2)$, where $Y_k^j = X_k + V_k^j$, for $j = 1, 2$. The noises $\{V_k^1\}$ and $\{V_k^2\}$ are, respectively, sequences of independent identically distributed (i.i.d) random variables with known densities, which are also independent of $\{X_k\}$.

An M level quantisation scheme for $\{Y_k^1\}$ is defined using the partition of the real line $-\infty = \ell_0^1 < \ell_1^1 < \cdots < \ell_M^1 = \infty$ as follows. For each $k = 0, 1, \ldots$, the quantised sensor $\bar{Y}_k^1$ is defined by $\bar{Y}_k^1 = m$, for the unique $m \in \{1, 2, \ldots, M\}$ satisfying $\ell_{m-1}^1 < Y_k^1 \leq \ell_m^1$. An N level quantisation scheme for $\{Y_k^2\}$ is defined analogously.

Define $\mathcal{Y}_k = \sigma(Y_0, \ldots, Y_k)$ and $\bar{\mathcal{Y}}_k = \sigma(\bar{Y}_0, \ldots, \bar{Y}_k)$. Let $\widehat{X}_k$ denote the filtered estimate of X_k at time k. Then for continuous measurements, $\widehat{X}_k \doteq \mathrm{sign}(Q_k)$ where $Q_k \doteq \mathbb{E}[X_K|\mathcal{Y}_k]$. For quantised measurements $\widehat{X}_k \doteq \mathrm{sign}(\bar{Q}_k)$, where $\bar{Q}_k \doteq \mathbb{E}[X_k|\bar{\mathcal{Y}}_k]$. Standard definitions of mutual information and conditional entropy, as used in [6], are used in this work.

3. QUANTISER DESIGN

Let $f_{Y|X}(y_k|x_i)$ denote the conditional density of Y_k given $X_k = x_i$. Then the conditional probability distribution of $\bar{Y}_k$ given $X_k = x_i$ is given by:

$$P_{\bar{Y}|X}(m, n|x_i) = \int_{\ell_{m-1}^1}^{\ell_m^1} \int_{\ell_{n-1}^2}^{\ell_n^2} f_{Y|X}(y^1, y^2|x_i)dy^2 dy^1 \quad (1)$$

Let $\Pi_{k|k}$ denote the conditional probability vector for the filter observation q_k when contiuous measurements are used, where $\Pi_{k|k}(i) = P(X_k = x_i|Y_k = y_k, \ldots, Y_0 = y_0)$, for $i = 1, 2$. Define $\bar{\Pi}_{k|k}$ by analogy when quantised signals are used. Then these vectors are updated by the recursive relations [7]:

$$\Pi_{k+1|k+1} = \frac{1}{1_2' B_{k+1} \Pi_{k|k}} B_{k+1} A \Pi_{k|k} \quad (2)$$

$$\bar{\Pi}_{k+1|k+1} = \frac{1}{1_2' \bar{B}_{k+1} \bar{\Pi}_{k|k}} \bar{B}_{k+1} A \bar{\Pi}_{k|k} \quad (3)$$

where $B_k = \mathrm{diag}[f_{Y|X}(y_k|x_1), f_{Y|X}(y_k|x_2)]$, $1_2'$ is a (1×2) vector of ones, $\bar{B}_k = \mathrm{diag}[P_{\bar{Y}|X}(m_k, n_k|x_1), P_{\bar{Y}|X}(m_k, n_k|x_2)]$. It follows from the definition of q_k, that $q_k = \Pi_{k|k}(2) - \Pi_{k|k}(1)$. Hence, using (2), the iterative update equation for q_k is given by:

$$q_k = g(y_k, q_{k-1}) = \frac{C_k^1 - C_k^2 + (C_k^1 + C_k^2)\beta q_{k-1}}{C_k^1 + C_k^2 + (C_k^1 - C_k^2)\beta q_{k-1}} \quad (4)$$

where $\beta = 1 - 2\alpha$, $C_k^1 = f_{Y|X}(y_k|x_1)$, $C_k^2 = f_{Y|X}(y_k|x_2)$.

The method used to determine the quantisation levels $\ell_m^1, 1 \leq m \leq M$ and $\ell_n^2, 1 \leq n \leq N$ is based on maximising the mutual information between the Markov state X_k and the quantised observation sequence $\{\bar{Y}_0, \bar{Y}_1, \ldots, \bar{Y}_k\}$. The justification for this approach [6] is based on the fact that intuitively, since a filter performs information processing on the observation sequence, the filter performance can be improved by maximising the amount of information in the observation sequence about the state [8]. It follows from the definitions of mutual information and conditional entropy that the mutual information is maximised by minimising the conditional entropy $H(X_k|\bar{Y}_0, \ldots, \bar{Y}_k)$ which is bounded

above by $H(X_k|\bar{Y}_k)$. In the absence of a closed form expression for $H(X_k|\bar{Y}_0, \ldots, \bar{Y}_k)$, the bound $H(X_k|\bar{Y}_k)$ is minimised. Clearly the resulting conditional entropy $H^*(X_k|\bar{Y}_0, \ldots, \bar{Y}_k)$ after minimisation can be no greater than the minimum $H^*(X_k|\bar{Y}_k)$.

Due to the underlying simplicity of the model, it was seen [6] that the bound $H(X_k|\bar{Y}_k)$ is reasonably tight and good filter performance can be obtained. As model complexity increases, the method can be extended to use progressively tighter bounds $H(X_k|\bar{Y}_k, \bar{Y}_{k-1})$ etc with a corresponding increase in computational complexity.

Using the defintion of conditional entropy and Bayes' Rule we have

$$H(X_k|\bar{Y}_k) = -\sum_{i=1}^2 \sum_{m=1}^M \sum_{n=1}^N P_{X,\bar{Y}}(x_i, m, n) \log P_{X|\bar{Y}}(x_i|m, n)$$

$$= -\frac{1}{2} \sum_{i=1}^2 \sum_{m=1}^M \sum_{n=1}^N P_{\bar{Y}|X}(m, n|x_i) \log \zeta(m, n|i)$$

$$\zeta(m, n|i) = \frac{P_{\bar{Y}|X}(m, n|x_i)}{P_{\bar{Y}|X}(m, n|x_1) + P_{\bar{Y}|X}(m, n|x_2)}$$

Direct computation of the continuous filter error requires the conditional density $S_i(q, \bar{q}) = f_{Q_k|X, Q_{k-1}}(q|x_i, \bar{q})$, whose existence is guaranteed under the following assumptions. For each $\bar{q} \in (-1, 1)$ define $g_{\bar{q}}: \mathbb{R}^2 \to \mathbb{R}$ by $g_{\bar{q}}(y^1, y^2) = g(y^1, y^2, \bar{q})$. If there exists a family of functions $f_{\bar{q}}: \mathbb{R}^2 \to \mathbb{R}, \bar{q} \in (-1, 1)$ such that each $G_{\bar{q}} \doteq (g_{\bar{q}}, f_{\bar{q}})$ has continuously differentiable inverse $H_{\bar{q}}: (-1, 1) \times \mathbb{R} \to \mathbb{R}$, then $S_i(q, \bar{q})$ exists and is given by:

$$S_i(q, \bar{q}) = \int_{-\infty}^\infty f_{Y|X}(H_{\bar{q}}(q, z)|x_i)|J_{H_{\bar{q}}}(q, z)| dz$$

where $J_{H_{\bar{q}}}(q, z)$ is the Jacobian of $H_{\bar{q}}$. Given $S_i(q, \bar{q})$, the estimation error (PFE) is computed using the steady state solutions of a discrete approximation to the recursive equation [6]:

$$f_{k+1}^i(q) = \sum_{j=1}^2 a_{ij} \int_{-1}^1 S_i(q, \bar{q}) f_k^j(\bar{q}) d\bar{q} \quad (5)$$

4. SIMULATIONS

Simulations were performed for HMM's having Gaussian noises $V_0^j \sim \mathcal{N}(0, \sigma_j)$, $j = 1, 2$ and correlation coefficient r. The conditional density of Y_k given $X_k = x_i$ is given by:

$$f_{Y|X}(y^1, y^2|x_i) = \frac{\sqrt{\kappa}}{2\pi\sigma_1\sigma_2} \exp\left\{-\frac{\kappa}{2}(u_i^2 - 2r u_i v_i + v_i^2)\right\}$$

where $u_i = (y^1 - x_i)/\sigma_1$, $v_i = (y^2 - x_i)/\sigma_2$ and $\kappa = 1/(1 - r^2)$. For the correlated noise case, using an appropriate change of variables, the integral in (1) can be written in terms of the Standard Normal Bivariate Integral:

$$\Phi(h, k, r) = \frac{\sqrt{\kappa}}{2\pi} \int_{-\infty}^h \int_{-\infty}^k \exp\left\{-\frac{\kappa}{2}(u^2 - 2r uv + v^2)\right\} du dv$$

which was computed using a Gauss quadrature method. Moreover with $f_{\bar{q}}(y^1, y^2) \doteq y^2$ for each $\bar{q} \in (-1, 1)$, the conditional density $S_i(q, \bar{q})$ was found to be:

$$S_i(q, \bar{q}) = \frac{1}{|\delta_1|\pi\sigma_1\sigma_2(1 - q^2)} \sqrt{\frac{\kappa\pi}{a}} \exp\left\{\frac{b^2}{4a} - c\right\} \quad (6)$$
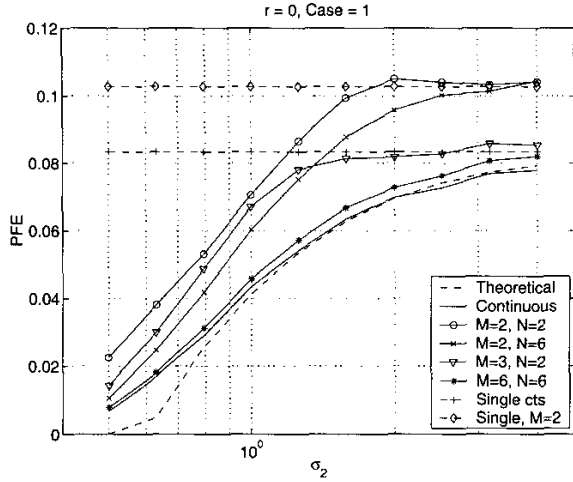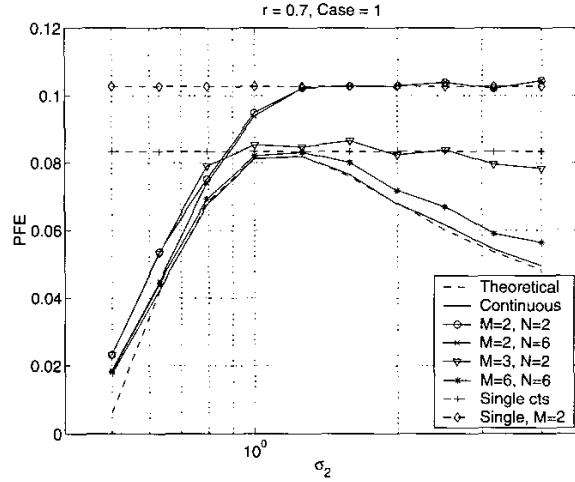
Fig. 2. Simulated PFE (Case 1, $r = 0$).

Fig. 3. Simulated PFE (Case 1, $r = 0.7$).

where

$$a = \frac{\kappa}{2} \left\{ \left(\frac{\delta_2}{\sigma_1 \delta_1} \right)^2 + \frac{2r\delta_2}{\sigma_1 \sigma_2 \delta_1} + \frac{1}{\sigma_2^2} \right\}$$

$$b = \kappa \left\{ \frac{\delta_2}{\sigma_1^2 \delta_1} \left(x_i - \frac{1}{\delta_1} \log(\gamma) \right) \right.$$
$$\left. - \frac{r}{\sigma_1 \sigma_2 \delta_1} \left(\log(\gamma) + (\delta_2 - \delta_1) x_i \right) - \frac{2}{\sigma_2^2} x_i \right\}$$

$$c = \frac{\kappa}{2} \left\{ \frac{1}{\sigma_1^2} \left(1 - \frac{x_i}{\delta_1} \log(\gamma) \right)^2 \right.$$
$$\left. - \frac{2r}{\sigma_1 \sigma_2} \left(1 - \frac{x_i}{\delta_1} \log(\gamma) \right) + \frac{1}{\sigma_2^2} \right\}$$

$$\delta_1 = \frac{2\kappa}{\sigma_1} \left(\frac{r}{\sigma_2} - \frac{1}{\sigma_1} \right), \quad \delta_2 = \frac{2\kappa}{\sigma_2} \left(\frac{r}{\sigma_1} - \frac{1}{\sigma_2} \right)$$

$$\gamma = \frac{(1+q)(1 - \beta\bar{q})}{(1-q)(1 + \beta\bar{q})}$$

The intergal in (5) was computed numerically by discretising the domain which reduced (5) to a matrix recursion. It was shown [6] that there exists a unique solution to the matrix recursion which, by continuity, converges to a solution of (5).

All simulations were obtained from data sets using 100,000 points, averaged over 5 sets of 20,000 points each. Optimisation was performed using a sequential simplex search routine (Matlab's optimisation toolbox). The transition probability matrix for the Markov Chain was $A = \left[\begin{smallmatrix} 0.8 & 0.2 \\ 0.2 & 0.8 \end{smallmatrix} \right]$. Simulations involving three values of the correlation coefficient, $r = 0, 0, 3, 0.7$ were performed for the following four noise level cases. Cases 1 to 3 had σ_1 fixed at 0.8, 1.6 and 2.5 respectively, while σ_2 was sampled logarithmically in the range (0.5, 4). In Case 4, $\sigma_1 = \sigma_2$, and σ_2 was sampled as in cases 1 to 3. For each (σ_1, σ_2) pair, a theroetical computation of the PFE for the continuous filter, and a simulation based computation of the PFE for the continuous filter was performed. Also, a quantisation scheme was calculated, and simulation based computation of the PFE for the associated quantised

filter was performed for every combination of quantistion levels in the ranges $2 \leq M \leq 6$ and $2 \leq N \leq 6$. Further simulations involving $r = 0.1, 0.2, \ldots, 0.8$ were performed for the continuous sensor filter only to further investigate the behaviour of the filter under large differences in the values of σ_1 and σ_2.

The typical behaviour of the filters is illustrated in figures 2 and 3. It was seen in almost all cases, that there was good fit between the theoretical filter error curve obtained from equation 5 and the simulated error curve obtained using continuous measurements. The discrepancies for small σ values are attributed partly to limits obtained by machine precision, and partly to errors introduced by discretisation of the integral in equation 5.

A surprising result seen in cases 1 to 3, is how filter performance improves as σ_2 increases beyond some critical value. This result was investigated further through simulations for the continuous filter for $r = 0.1, 0.2, \ldots, 0.8$, with the results for case 1 shown in figure 4. Results indicate that the worst filter performance (maximum PFE) occurs at the critical value $\sigma_2 = \sigma_1 / r$. Furthermore, this value corresponds to the performance value of a single sensor filter (with noise level σ_1). It should be stressed that this result applies to $r > 0$ only.

This behaviour can be explained as follows. The correlation coefficient, r, between two real valued signals indicates the degree to which both signals tend to have the same ($r > 0$) or opposite ($r < 0$) sign. The correlation coefficient introduces a bias into the filter which enhances filter performance. However, the performance of the filter also suffers from reduced independence in the two signals. Thus for a given amount of correlation, the loss in performance of the filter due to reduced independence and increasing noise levels is offset by the gain in performance from improved noise sign estimation.

With regard to question 1, the limiting behaviour of the continuous filter for large σ_2 can be seen from the recursive update equations 2 and 4 for the filter. As $\sigma_2 \to \infty$, the resulting equations approach the case of a single sensor filter with noise level $\sigma = \sigma_1 / \sqrt{\kappa}$. Simulations for $\sigma_2 > 4$ were performed for case 1 to further investigate this behaviour, however, in the case of correlated noise, as seen in figure 5, filter performance does not ap-
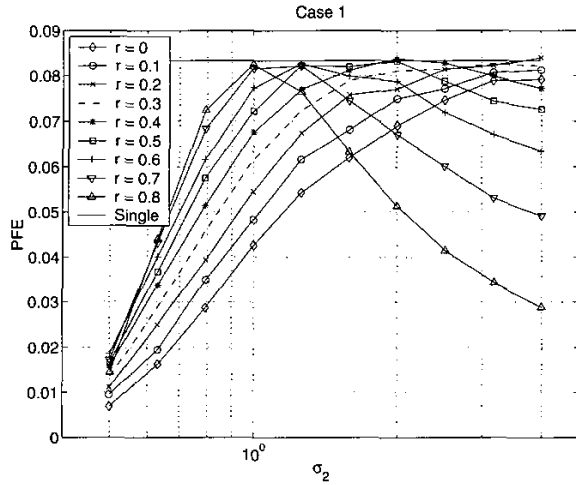
Fig. 4. Continuous filter: Simulated PFE (Case 1).

Fig. 5. Continuous filter: Simulated PFE (Case 1, large σ_2).

proach that of the single continuous filter as σ_2 becomes large, but approaches a lower value instead. This result can be explained by the fact that even at large noise levels, the bias introduced by the correlation is still present in the filter.

In the simulations involving quantised filters, it was seen in almost all cases that the performance of the quantised filter could be made to approach the performance of the continuous filter with relatively few numbers of quantisation levels, indicating the usefulness of the mutual information based approach to quantisation. In general, the presense of correlation in the signals meant a reduction in the amount of mutual information between the signals and the Markov state, which resulted in reduced filter performance. This was especially noticable for low numbers of quantisation levels.

Question 2 can be addressed from the simulations involving cases 1 to 3. Firstly, it was observed that, at worst, the addition of a second sensor made no change at all to filter performance, while at best, the filter could be made to perform better than the corresponding single continuous filter (with noise level σ_1).

In the independent noise case, filter performance degraded as σ_2 increased in all cases. In the correlated noise case, for σ_2 in the range $\sigma_1 \leq \sigma_2 \leq \sigma_1/r$, improvements in filter performance were achievable with the addition of a second sensor in all cases. However, the level of improvement decreased as σ_2 increased until the critical value was reached, at which point no improvement in filter performance was obtained regardless of the number of quantisation levels used. As σ_2 increased beyond the critical value, improvements in filter performance with the addition of a second sensor began to increase again with the exception of the case $M = 2$, in which no further improvements were observed, regardless of the number of quantisation levels for the second sensor.

Perhaps the most important observation that can be made in the case of correlated noise is that improvements in filter performance are no longer guaranteed by the addition of a second signal.

With regard to question 3 it was observed that quantisation schemes obtained for each sensor from the joint distribution were dependent on the noise levels in both sensors. For a given sensor, the corrsponding single sensor quantisation scheme was ap-
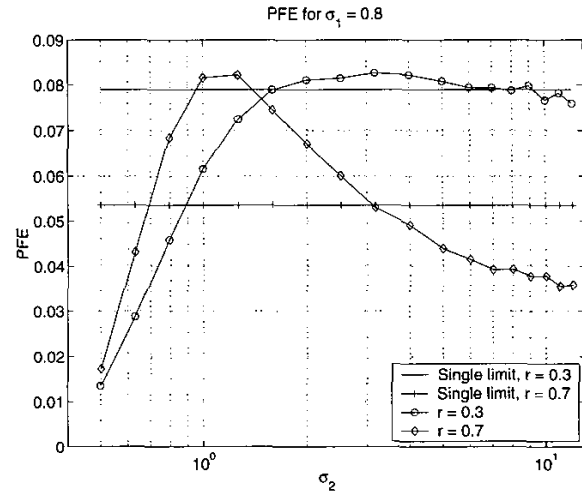
proached only as noise levels in the other sensor increased without bound.

5. REFERENCES

[1] Wai-Man Lam and Amy R. Reibman, "Design of quantizers for decentralized estimation systems," *IEEE Tarnsactions on Communications*, vol. 41, no. 11, November 1993.

[2] Zhi-Quan Luo and john N. Tsitsiklis, "Data fusion with minimal communication," *IEEE Transactions on Information Theory*, vol. 40, no. 5, September 1994.

[3] Hamid R. Hashemipour, Sumit Roy, and Alan J. Laub, "Decentralised structures for parrallel kalman filtering," *IEEE Transactions on Automatic Control*, vol. 33, no. 1, January 1988.

[4] Jean-François Chamberland and Venugopal V. Veeravalli, "Decentralized detection in sensor networks," IEEE Transactions on Signal Processing (to appear).

[5] Nigel J. Newton, "Observations preprocessing and quantization for nonlinear filters," *SIAM J. Control Optim.*, vol. 38, no. 2, pp. 482–502, 2000.

[6] Louis Shue, Subhrakanti Dey, Brian D. O. Anderson, and Franky Be Bruyne, "On state-estimation of a two-state hidden markov model with quantisation," *IEEE Transactions on Signal Processing*, vol. 49, no. 1, 2001.

[7] L. R. Rabiner, "A tutorial on hidden markov models and selected applications in speech recognition," *Proc. IEEE*, vol. 77, pp. 257–258, February 1989.

[8] X. Feng, K. Loparo, and Y. Fang, "Optimal state estimation for stochastic systems: An information theoretic approach," *IEEE Transactions on Automatic Control*, vol. 42, pp. 771–785, 1997.