

# On the Effect of Quantization on Performance at High Rates

Vijay Gupta, Amir F. Dana, Richard M Murray and Babak Hassibi

**Abstract**—We study the effect of quantization on the performance of a scalar dynamical system in the high rate regime. We evaluate the LQ cost for two commonly used quantizers: uniform and logarithmic and provide a lower bound on performance of any centroid-based quantizer based on entropy arguments. We also consider the case when the channel drops data packets stochastically.

## I. INTRODUCTION

In recent years analysis of linear systems in which the process and the controller communicate using a digital communication channel has gained increasing attention. This situation can model various situations in which the controller and the process are physically separated and communication between the two is assumed to be over a finite bandwidth channel, such as in networked or distributed control applications. Moreover, this area lies at the intersection of control theory and communications and as such can provide useful insights into both the areas.

As a result, this problem has been analyzed with increasing regularity since the seminal paper [6]. The problem of stabilization with finite communication bandwidth was considered by Wong and Brockett [24], [25]. Baillieul [1] also reported a tight bound on the data rate requirement for stabilizing a scalar system. Nair et al [19], [18] considered the stabilization of stochastic linear systems and Markov jump linear systems with finite data rates. Tatikonda [22] studied stabilization of finite-dimensional discrete-time noiseless linear processes and also presented results about the optimal LQG control of linear systems across noisy feedback links (see also [2]). Elia and Mitter [7] considered the question of optimal quantizer for stabilization. Various quantization and coding schemes for stabilization have been studied in the literature, (see, e.g., [20], [3], [16], [12], [13], [9]).

However, most of the work reported so far has focused on the effect of quantization on stability. It is worthwhile to also consider the question of performance of the system in the presence of quantization. The performance of a scalar statically quantized system with delays was considered in [23]. Lemmon and Ling [15] presented an upper bound for the quantization noise for the case when dynamic uniform quantization is done over a channel that drops packets. They defined the performance in terms of signal to quantization

ratio and presented some interesting trade-offs between the number of bits, locations of the system poles and the performance.

In this paper we study the effect of quantization on the LQR performance of the system. We consider a linear time-invariant scalar system with a control law in place and see how the performance degrades as less and less data is allowed to pass from the process to the controller. We come up with some interesting bounds for specific quantizers and some entropy-based general bounds on any centroid-based quantization and encoding scheme. We also consider extensions to dynamic quantization schemes and packet-dropping channels.

The paper is organized as follows. In the next section, we set up the problem and define some notation. We then study some specific static quantizers, dynamic quantizers and the extension to packet-dropping channels. We end with conclusions and outline some scope for future work.

## II. PROBLEM FORMULATION

Consider the situation described in Figure 1. We have a linear time-invariant process evolving in discrete time according to the relation

$$x_{k+1} = ax_k + u_k + w_k \quad (1)$$

with  $x_k \in \mathbf{R}$  as the process state,  $u_k$  as the control input and  $w_k$  as zero-mean white noise. The noise  $w_k$  is assumed to be bounded in an interval  $[-M, M]$  with variance  $\sigma^2$ . The initial condition  $x_0$  is assumed to be uniformly distributed in a symmetric interval  $[-u, u]$  and is assumed to be independent of  $w_k$ .

The process state is observed in a noiseless fashion by the encoder. The encoder denotes a mapping from the state  $x_k$  to a stream of bits  $b_k$ . The encoder has access to all the previous states  $\{x_j\}_{j=0}^k$  and the previous control signals  $\{u_j\}_{j=0}^{k-1}$  when it encodes  $x_k$ . We restrict our attention to encoders that merely perform the action of quantization and ignore the possibility of other source coding.

The bit-stream  $\{b_k\}$  is transmitted over a channel. We will chiefly be concerned with noiseless digital channels. After passing through the channel, the bits are received at the decoder. The decoder looks at the bit stream and outputs an estimate of the state  $\hat{x}_k$ . The decoder has access to all the previous bit streams  $\{b_j\}_{j=0}^k$ , the previous decoded estimates  $\{\hat{x}_j\}_{j=0}^{k-1}$  and the previous control signals  $\{u_j\}_{j=0}^{k-1}$  when it decodes  $\{b_k\}$ . We assume a linear control law of the form  $u_k = f\hat{x}_k$ . This control signal  $u_k$  is then used in the further evolution of the process described by (1). We assume that

Division of Engineering and Applied Science, California Institute of Technology, Pasadena, CA 91125, USA  
{vijay@cds., amirf@, murray@, hassibi@}caltech.edu  
Research supported in part by AFOSR grant F49620-01-1-0460 and by NSF grant CCR-0326554 for the first author and in part by the National Science Foundation under grant no. CCR-0133818, by David and Lucille Packard Foundation, and by Caltech's Lee Center for Advanced Networking for the second author.

there is no channel present between the controller and the process.

In the absence of any encoder, channel and decoder, we have  $\hat{x}_k = x_k$ . In general the two quantities would not be equal. For this process we consider the finite-time and infinite

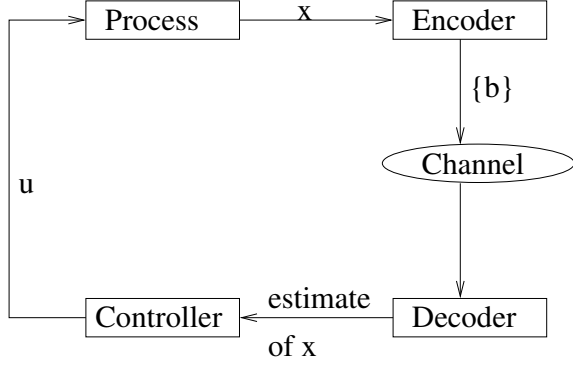


Fig. 1. Representation of the system considered.

time horizon LQ costs given by

$$J_K = E \left[ \sum_{k=0}^K x_k^2 q + u_k^2 r \right] \quad J_\infty = \lim_{K \rightarrow \infty} \frac{1}{K} J_K, \quad (2)$$

where, as usual,  $q$  is positive and  $r$  is non-negative. We assume that system has been sufficiently well-designed so that it remains stable and inside the range of operation of the quantizer at all times to avoid quantizer over-flow. The assumption of bounded noise is crucial in this. We wish to consider the effect of various quantizers on  $J_K$  and  $J_\infty$ .

We will denote the probability density function of a continuous random variable  $X$  by  $f_X(x)$  and its expectation by  $E[X]$ . The differential entropy of  $X$  is denoted by  $h(X)$  and defined according to (see, e.g., [5])

$$h(X) = - \int_{-\infty}^{\infty} f_X(x) \log f_X(x) dx,$$

where  $0 \log 0$  is interpreted as 0 and the log is taken to the base 2. A scalar quantizer  $Q$  of size  $N$  is a mapping from an interval on the real number line into a finite set  $\mathcal{C}$  containing  $N$  reproduction points called codewords. The interval is partitioned into  $N$  cells where the  $i$ -th cell, denoted by  $R_i$ , is an interval for which each point maps to the  $i$ -th codeword  $y_i$ . The quantizer is said to be mid-point based if the reconstruction level  $y_i$  is the mid-point of the cell  $R_i$ . It is said to be centroid-based if  $y_i$  is the centroid (conditional expectation of  $x$  in region  $R_i$ ) of region  $R_i$ . We will always assume that the boundaries of the cells are the mid-points of the quantization levels. We call a quantizer static or fixed if the mapping does not change with time, otherwise we call it dynamic. The rate of the quantizer is defined as  $r = \log_2(N)$ . We will be interested in the mean squared distortion of the quantizer defined as  $D = E[|X - \hat{X}|^2]$  when the scalar random variable  $X$  is quantized and reproduced as  $\hat{X}$ .

### III. ANALYSIS

We begin by considering the case when the quantizer is fixed. In general, the cost function is not easy to calculate analytically. This is because the quantization error depends on the probability density function of  $x_k$  which is not easy to calculate as time  $k$  evolves. To obtain a handle on the performance of different quantizers, we make the high rate approximation, which says that the rate of the quantizer is high (and hence the distortion is low). The results we obtain can thus be treated as approximations which become better as the rate of the code increases. At high rates, we can use the following result from source coding theory [11].

*Theorem 1:* Given a scalar quantizer with a mean square based distortion measure  $d(x, y) = \|x - y\|^2$ , the expected distortion of the random variable  $X$  being quantized can be bounded as follows

$$\bar{d} \geq \bar{d}_L = \frac{1}{12N^2} E[\lambda(X)^{-2}],$$

where  $\lambda(X)$  is the asymptotic quantizer density normalized to unit integral, obtained as we keep on increasing the number of quantization levels while  $N$  refers to the total number of quantization cells. Further, the lower bound becomes tighter as the rate of the code gets high.

*Uniform quantizer:* We first consider a mid-point based uniform quantizer, which is a very simple and commonly used quantizer. We first note that if the region to be quantized is  $[-t, t]$ , the asymptotic quantizer density is given by  $\lambda(x) = \frac{1}{2t}$ . Since  $N = \frac{2t}{\delta}$ , the distortion measure evaluates to  $\bar{d} = \frac{\delta^2}{12}$ . In addition we note the following result from [17] that is valid at high rates. For a mid-point based uniform quantizer, if we denote the variable being quantized as  $x_k$  and the quantization error by  $\delta_k$ , then  $E[x_k \delta_k] \ll E[\delta_k^2]$  and thus can be approximated to be zero<sup>1</sup>. Thus

$$J_K = (q + rf^2) E[x_0^2] \sum_{k=0}^K (a + f)^{2k} + rf^2 (K + 1) \frac{\delta^2}{12} + \frac{q + rf^2}{1 - (a + f)^2} \left( \frac{f^2 \delta^2}{12} + \sigma^2 \right) \sum_{k=0}^K (1 - (a + f)^{2k}).$$

For calculation of  $J_\infty$  we need to find conditions such that  $J_K/K$  does not diverge. To this end, we assume that  $(a + f)^2 < 1$  and that  $E[x_{k+1}^2] < E[x_k^2]$ . The first condition means  $f$  is stabilizing while the second condition places a limit on the size of the quantization cell. Assuming that there are  $N$  quantization cells, this condition implies that

$$N^2 \geq \frac{f^2 u^2}{3(1 - (a + f)^2) E[x_k^2] - 3\sigma^2}.$$

In particular for  $k = 0$ , this condition implies

$$N^2 \geq \frac{f^2 u^2}{(1 - (a + f)^2) u^2 - 3\sigma^2}.$$

<sup>1</sup>The result is true only under some technical conditions listed in [17] that, however, hold in our case.

Note that for the case when there is no noise and the control law  $f = -a$ , this reduces to the results derived in [22]. With these assumptions, the infinite-horizon cost is

$$J_\infty = rf^2 \frac{\delta^2}{12} + \frac{q + rf^2}{1 - (a + f)^2} \left( \frac{f^2 \delta^2}{12} + \sigma^2 \right).$$

*Logarithmic Quantizer:* We now calculate the cost for a mid-point based logarithmic quantizer that has been shown to be the most optimal quantizer for stabilization [7]. To apply Theorem 1 for a logarithmic quantizer that is operating over the union of the regions  $[-a, -\epsilon]$  and  $[\epsilon, a]$ , we note that the asymptotic quantizer density is given by

$$\lambda(x) = \frac{1}{|x|} \Big|_{\text{normalized to unit integral}} = \frac{1}{2|x| \ln(\frac{a}{\epsilon})}.$$

Now consider a logarithmic quantizer with ratio  $g$ . Thus the quantization cells are given for the positive axis by the intervals  $[\epsilon, g\epsilon]$ ,  $[g\epsilon, g^2\epsilon]$ ,  $\dots$ ,  $[g^{p-1}\epsilon, g^p\epsilon]$ , where  $p$  and  $N$  are related by  $2p = N$ . Since  $g^p\epsilon = a$ ,

$$\bar{d} = \frac{1}{3} \left( \frac{\ln(g)}{2} \right)^2 E[x_k^2] = \frac{(\ln g)^2}{12} E[x_k^2] \approx E[\Delta_k^2].$$

Using the fact that

$$-\sqrt{E[\Delta_k^2]E[x_k^2]} \leq E[\Delta_k x_k] \leq \sqrt{E[\Delta_k^2]E[x_k^2]},$$

we can obtain that

$$\begin{aligned} h_1 E[x_0^2] \frac{1 - g_1^{K+1}}{1 - g_1} + \frac{h_1 \sigma^2 (K - 1 + g_1^{K+1})}{1 - g_1} &\leq J_K \\ &\leq h_2 E[x_0^2] \frac{1 - g_2^{K+1}}{1 - g_2} + \frac{h_2 \sigma^2 (K - 1 + g_2^{K+1})}{1 - g_2}, \end{aligned}$$

where

$$\begin{aligned} g_1 &= (a + f)^2 + cf^2 - 2 |f(a + f)| \sqrt{c} \\ g_2 &= (a + f)^2 + cf^2 + 2 |f(a + f)| \sqrt{c} \\ h_1 &= q + rf^2 + rcf^2 - 2rf^2 \sqrt{c} \\ h_2 &= q + rf^2 + rcf^2 + 2rf^2 \sqrt{c} \end{aligned}$$

and  $c = \frac{(\ln g)^2}{12}$ . Thus a necessary condition for  $J_\infty$  to exist is  $g_1 \leq 1$  and a sufficient condition is  $g_2 \leq 1$ . Assuming these conditions exist, we obtain

$$\frac{h_1 \sigma^2}{1 - g_1} \leq J_\infty \leq \frac{h_2 \sigma^2}{1 - g_2}.$$

*Lower Bound for Centroid-based Quantizers:* It is well-known that the optimal quantizer minimizing the mean-square distortion error is a centroid-based quantizer [11]. Since the density function of  $x_k$  depends on the densities of all previous quantization errors, it is difficult to compute a priori and the optimal quantizer has to be obtained at every step through an iterative algorithm such as the Lloyd-Max algorithm [10], [11] or through a dynamic programming based algorithm [4]. We now consider such quantizers. To begin with, we note that for centroid-based quantizers (see, e.g., [17])  $E[\Delta_k \hat{x}_k] = 0$  for every time step  $k$ . Thus

$$E[x_{k+1}^2] = (a + f)^2 E[x_k^2] - (f^2 + 2af) E[\Delta_k^2] + \sigma^2.$$

Thus the cost can be evaluated as

$$\begin{aligned} J_K &= \sum_{k=0}^K \left[ -rf^2 E[\Delta_k^2] + (q + rf^2)(a + f)^{2k} E[x_0^2] \right. \\ &\quad \left. + \sigma^2 (q + rf^2) \sum_{j=0}^{k-1} (a + f)^{2j} - (f^2 + 2af) \times \right. \\ &\quad \left. (q + rf^2) \sum_{j=0}^{k-1} (a + f)^{2j} E[\Delta_{k-1-j}^2] \right]. \end{aligned}$$

The cost can easily be evaluated for specific quantizers such as uniform or logarithmic. Instead, we lower bound the cost function for any centroid-based quantizer using entropy arguments that do not require high-rate approximations. We note the following [5]

- Given  $n$  bits to describe a random variable  $X$  with differential entropy  $h(X)$ , the error can have differential entropy no less than  $h(X) - n$ .
- Given a random variable  $X$  with differential entropy  $h(X)$ , the lowest possible variance of  $X$  is  $\frac{1}{2\pi e} 2^{2h(X)}$ .
- *The Entropy-Power Inequality:* Given two independent random variables  $X$  and  $Y$  with differential entropy  $h(X)$  and  $h(Y)$  respectively,

$$2^{2h(X+Y)} \geq 2^{2h(X)} + 2^{2h(Y)}.$$

- Entropy of a random variable  $X$  is no less than the entropy of  $X$  given additional information about another random variable  $Y$ .

At time step  $k = 0$ , the entropy is simply  $h(x_0)$ , thus the entropy of  $\Delta_0$  is at least  $h(x_0) - n$ . At time step  $k = 1$ , we have

$$h(x_1) \geq h(x_1 | \hat{x}_0) = h(ax_0 + w_0 | \hat{x}_0).$$

Now  $x_0$  and  $w_0$  are independent (even given  $\hat{x}_0$ ). Denote the entropy of the noise by  $h(w)$ . Then

$$\begin{aligned} 2^{2h(x_1)} &\geq 2^{2h(ax_0 | \hat{x}_0)} + 2^{2h(w | \hat{x}_0)} \\ &= 2^{2\log(a) + 2h(x_0 | \hat{x}_0)} + 2^{2h(w)} \\ &\geq 2^{2\log(a)} 2^{2h(x_0) - 2n} + 2^{2h(w)}. \end{aligned}$$

Let  $c = a^2 2^{-2n}$ . Thus we obtain

$$\begin{aligned} h(x_1) &\geq \frac{1}{2} \log \left[ c 2^{2h(x_0)} + 2^{2h(w)} \right] \\ h(\Delta_1) &\geq \frac{1}{2} \log \left[ c 2^{2h(x_0)} + 2^{2h(w)} \right] - n. \end{aligned}$$

Similarly we may obtain

$$h(\Delta_k) \geq \frac{1}{2} \log \left[ c^k 2^{2h(x_0)} + \sum_{j=0}^{k-1} c^j 2^{2h(w)} \right] - n.$$

Finally the error variance at time step  $k$  is bounded by

$$E[\Delta_k^2] \geq \frac{1}{2\pi e} 2^{-2n} \left[ c^k 2^{2h(x_0)} + \sum_{j=0}^{k-1} c^j 2^{2h(w)} \right].$$

Thus we can evaluate the lower bound on cost function as

$$\begin{aligned} & \sum_{k=0}^K \left[ -rf^2 \frac{1}{2\pi e} 2^{-2n} \left( c^k 2^{2h(x_0)} + \sum_{j=0}^{k-1} c^j 2^{2h(w)} \right) \right. \\ & + (q + rf^2)(a + f)^{2k} E[x_0^2] + \sigma^2(q + rf^2) \sum_{j=0}^{k-1} (a + f)^{2j} \\ & \left. - (f^2 + 2af)(q + rf^2) \frac{1}{2\pi e} 2^{-2n} \sum_{j=0}^{k-1} (a + f)^{2j} \right. \\ & \left. \left( 2^{2h(x_0)} c^{k-j-1} + \sum_{i=0}^{k-j-2} 2^{2h(w)} c^i \right) \right]. \end{aligned}$$

Further, if we assume

$$(a + f)^2 \leq 1, \quad a^2 2^{-2n} \leq 1, \quad (3)$$

we obtain

$$\begin{aligned} J_\infty = & -rf^2 \frac{1}{2\pi e} 2^{-2n} 2^{2h(w)} \frac{1}{1-c} + \frac{\sigma^2(q + rf^2)}{1 - (a + f)^2} \\ & - (f^2 + 2af)(q + rf^2) \frac{1}{2\pi e} 2^{2(h(w)-n)} \frac{1}{1 - (a + f)^2} \frac{1}{1-c}. \end{aligned}$$

We do not yet have an analytic expression for the tightness of the bound. Note that the condition given in (3) is similar to the condition obtained for stability of a scalar unstable system in, e.g., [1]. Also, since we are interested only in quantizing the current state, we escape the complexity of having to define terms like average conditional entropy power as in [19].

So far we have assumed that the quantization is not followed by any noiseless coding. Moreover we have concentrated on the case of fixed rate quantization. Thus we defined the rate of the quantizer as  $\log(N)$ , where  $N$  is the number of quantization levels. If we assume that noiseless coding is permitted, it makes more sense to consider the entropy of the output vector as a measure of the rate. In such a case, we note the following result [11]

*Theorem 2:* The constrained entropy high rate quantizer bound is given by

$$\bar{d}_L \geq \frac{1}{12} e^{-2(H(q(X)) - h(X))},$$

where  $h(X)$  is the differential entropy of the random variable  $X$  while  $H(q(X))$  is the entropy of quantized variable  $q(X)$ . Furthermore, equality is achieved if and only if the asymptotic quantizer density  $\lambda(x)$  is a constant, that is, the quantizer reproduction vectors are uniformly distributed over some set having probability 1. Thus the bound is achieved by high rate lattice vector quantizers since they have a uniform density of quantization levels.

If we define rate as  $R = H(q(X))$  (which gives the average codeword length achievable using noiseless coding and hence the average rate), we obtain the above bound. For a fixed  $R$ , it can be proved that this distortion is lower than the one achieved for a given code rate. However, actually achieving this rate might require long codewords and hence might not

be practical in a real-time system. There also have been some works in the information theory literature, e.g. [21], that provide a way to bound the output entropy of a quantizer given the number of levels of the quantizer. Such works may provide an interesting way to achieve a trade-off between the two notions of rate that we have presented.

*Dynamic Quantization:* It is apparent that only the region corresponding to the uncertainty that the decoder has about  $x_k$  needs to be quantized and the information sent. We now consider this case of dynamic quantization in which the number of quantization levels  $N$  remains fixed; however the range over which quantization is being done varies with time. This is similar to schemes like prediction based encoding outlined in [12] and yields better performance, at the cost of added complexity due to a time-varying quantizer. Moreover it assumes some level of synchronization between the encoder and the decoder so that both agree on the specific quantizer to which the bits at time  $k$  pertain.

For simplicity, we consider only the infinite-time horizon cost function  $J_\infty$ . For the case of a uniform quantizer with  $N$  levels, the quantization step size at time  $k$  is given by

$$N\delta_k = l_k = a\delta_{k-1} + 2M.$$

The variance of the quantization error at time  $k$ ,  $E[\Delta_k^2]$  can be evaluated as before to be  $\frac{\delta_k^2}{12}$ . Thus

$$J_\infty = \frac{q + rf^2}{1 - (a + f)^2} \left( \sigma^2 + \frac{f^2 M^2}{3(N^2 - a^2)} \right) + \frac{rf^2 M^2}{3(N^2 - a^2)}.$$

The conditions for existence of  $J_\infty$  are

$$(a + f)^2 \leq 1, \quad \frac{a}{N} < 1.$$

Also note that this cost is equivalent to that of a static uniform quantizer with step size  $\delta = \frac{4M^2}{N^2 - a^2}$ . Since the cost function for a static uniform quantizer is an increasing function in the step size  $\delta$ , this gives us a relation between the parameters  $M$ ,  $N$  and  $a$  for determining which of the two quantizers, static or dynamic, is better.

*Stochastic Packet Drops:* So far we have assumed a perfect channel model, in that the bits  $\{b_k\}$  were transmitted to the decoder without fail. A more realistic channel model is one that suffers from stochastic data loss. In this paper, we model the data loss using the random packet loss model. At each time step, the channel can either be in a ‘good’ or a ‘bad’ state. In the good state, it transmits the bits  $\{b_k\}$  while in the bad state the packet containing the bits is dropped and no data is transmitted to the decoder. For simplicity we shall consider only the case when the channel transitions between these two states in an i.i.d. fashion although the results may readily be extended to the case when the transition occurs according to a Markov chain (the classical Gilbert-Elliott channel model [8]). Let the probability of packet drop at every time step be  $p$ . Note that the expectation in the cost function is now also over the probability of packet drops at each time step. For simplicity, we consider only  $J_\infty$  for the case of a midpoint-based uniform quantizer. In this case, the

state evolves according to the equation

$$x_{k+1} = \begin{cases} (a+f)x_k + f\Delta_k + w_k & \text{with prob } 1-p \\ ax_k + w_k & \text{with prob } p. \end{cases} \quad (4)$$

For a uniform quantizer,  $E[\Delta_k^2] = \frac{\delta^2}{12}$  while  $\Delta_k$  and  $w_k$  are independent of each other. Thus we can evaluate the steady-state covariance as

$$P_\infty = \frac{(1-p)f^2\frac{\delta^2}{12} + \sigma^2}{1 - (a+f)^2(1-p) - a^2p}.$$

Since the cost function is given by

$$J_\infty = (q + rf^2)P_\infty + rf^2\frac{\delta^2}{12},$$

it can be easily evaluated. For convergence, we have the additional condition  $(1-p)(a+f)^2 + pa^2 < 1$ .

*Example:* In this subsection, we consider a simple example to illustrate the above results. We consider the system parameter  $a = 2$ . The initial condition  $x_0$  is assumed to be uniformly distributed in the range  $[-20, 20]$  while the white noise  $w_k$  is assumed to be uniformly distributed in the range  $[-1, 1]$ . The cost function we consider is

$$J = \lim_{k \rightarrow \infty} E[x_k^2 + u_k^2].$$

For this cost function, the optimal control law without quantization turns out to be  $f = -1.618$ . We use this control law to consider the performance of various quantization schemes considered above. For the quantizers that operate on a fixed range, the minimum region to be quantized is  $[-20, 20]$ . We will assume that the control law does not allow the system to go outside this range, thus avoiding quantizer overflow.

Figures 2 and 3 show a comparison of our theoretical approximations with simulation results for uniform and logarithmic quantizers respectively. The simulation results refer to the cost in steady state averaged over 10000 runs for a system using the particular quantizer. The initial condition and the noise driving the system were chosen randomly for each run. It can be seen that the approximations are quite good, at least in this example. It can be seen that even for this simple system, logarithmic quantizer yields much better performance for the same number of bits. However, for convergence, the uniform quantizer requires 2 or more bits while for  $\epsilon = 0.1$ , the necessary and sufficient conditions for logarithmic quantizer require 3 and 4 bits respectively. Also, it may be noted that the plots provide merely a qualitative comparison since the expressions provided in the analysis are approximations. Figure 4 shows a comparison of the performance achieved by the dynamic uniform quantizer with the performance bound derived using entropy arguments. We see that the bound is reasonably tight in this example. Of course, static quantizers perform much worse than the dynamic quantizers, especially when a small number of bits are used. Figure 5 shows the performance of the system as a function of the packet loss probability across a channel that drops packets in an i.i.d. fashion. A uniform quantizer with 6

bits is used. The system becomes unstable at the theoretical value of  $p = 0.22$ .

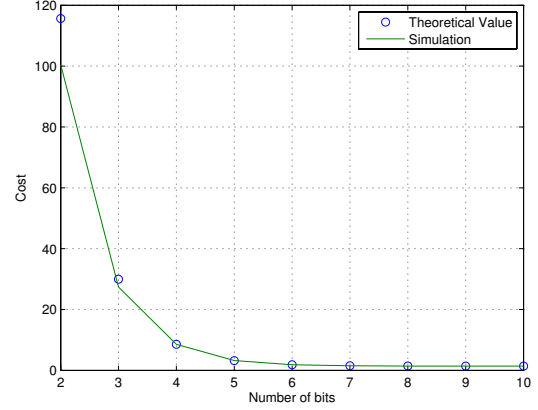


Fig. 2. Performance of approximations presented in the paper for uniform quantizer with simulation results.

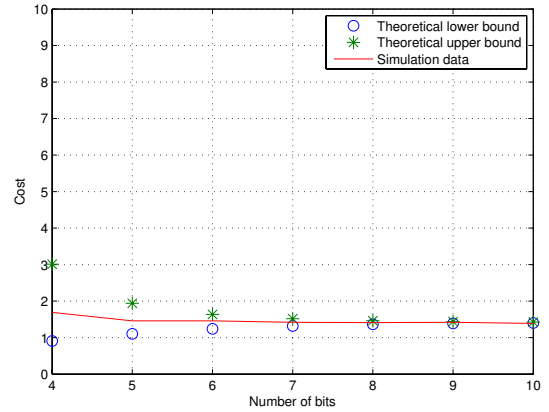


Fig. 3. Performance of approximations presented in the paper for logarithmic quantizer with simulation results.

#### IV. CONCLUSIONS AND FUTURE WORK

In this work, we looked at a scalar system in which the state is being quantized prior to its transmission to the controller. We looked at the problem of evaluating the performance of the controller in minimizing a quadratic cost for uniform and logarithmic quantizers. We saw that the problem is difficult to solve exactly in most cases; however through approximations like high-rate assumption, we were able to evaluate the performance. We also considered some simple extensions such as dynamic quantization and packet dropping channels.

There are many interesting directions in which this work can be extended. So far we have only considered the case of scalar processes. A more general case is when the process state  $x_k$  and measurement vector  $y_k$  are vectors. Studying quantization issues for such plants takes us into the realm of vector quantization theory, which is less well-developed

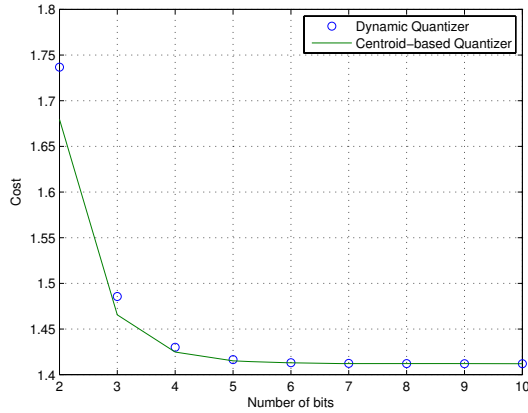


Fig. 4. Comparison of the dynamic quantizer performance with the lower bound derived in the paper.

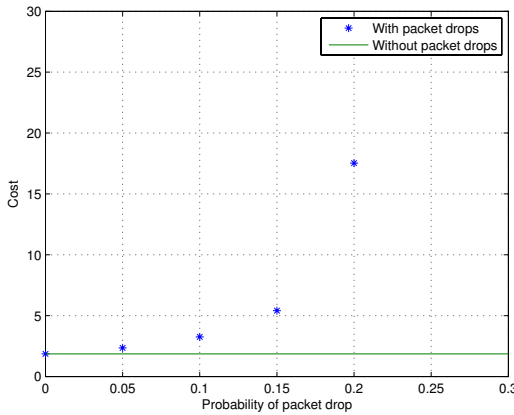


Fig. 5. Performance of the system across a packet dropping channel.

than its scalar counterpart and hence the extension is not trivial. The basic difficulty is that each component of the vector  $y_k$  carries information about other components and hence it is extremely wasteful to do scalar quantization on each component separately. If the system matrix  $A$  is diagonal (or diagonalizable) and the matrix  $C$  is invertible, and hence this dependence is not present, the results from scalar quantization that we derive above can be used on each component. We are currently working on extending our results to the case when these assumptions do not hold. Another important question is identification of the quantity that should be quantized and sent across the link. Some initial results are known in the area through the work of Ishwar et al. [14], but more research needs to be done.

## REFERENCES

- [1] J. Baillieul. Feedback designs for controlling device arrays with communication channel bandwidth constraints. In *ARO workshop on smart structures*, 1999.
- [2] V. S. Borkar, S. K. Mitter, and S. Tatikonda. Markov control problems with communication constraints. *Comm. in Information and Systems*, 1(1):16–33, 2001.

- [3] R. W. Brockett and D. Liberzon. Quantized feedback stabilization of linear systems. *IEEE Transactions on Automatic Control*, 45(7):1279–89, 2000.
- [4] J. D. Bruce. Optimum quantization. Technical Report 429, Research Laboratory of electronics, Massachusetts Institute of Technology, Cambridge, 1965.
- [5] T. M. Cover and J. A. Thomas. *Information Theory*. Wiley and Sons, New York, 1991.
- [6] D. F. Delchamps. Stabilizing a linear system with quantized state feedback. *IEEE Transactions on Automatic Control*, 35:916–924, 1990.
- [7] N. Elia and S. K. Mitter. Stabilization of linear systems with limited information. *IEEE Transactions on Automatic Control*, 46(9):1384–1400, 2001.
- [8] E. O. Elliott. Estimates of error rates for codes on burst-noise channels. *Bell Systems Technical Journal*, 42:1977–1997, September 1963.
- [9] F. Fagnani and S. Zampieri. Stability analysis and synthesis for scalar linear systems with a quantized feedback. *IEEE Transactions on Automatic Control*, 48(9):1569–1584, 2003.
- [10] A. Gersho and R. M. Gray. *Vector Quantization and Signal Compression*. Kluwer Academic Publishers, 1991.
- [11] R. M. Gray. *Source Coding Theory*. Kluwer Academic Publishers, 1990.
- [12] J. Hespanha, A. Ortega, and L. Vasudevan. Towards the control of linear systems with minimum bit-rate. In *Proc. of the 15th Int. Symp. Math. The. Netw. Sys.*, 2002.
- [13] H. Ishii and B. A. Francis. Quadratic stabilization of sampled-data systems with quantization. *Automatica*, 39:1793–1800, 2003.
- [14] P. Ishwar, R. Puri, K. Ramchandran, and S. S. Pradhan. On rate-constrained distributed estimation in unreliable sensor networks. *IEEE Journal on Selected Areas in Communications: Special issue on self-organizing distributed collaborative sensor networks*, 23(4):765–775, April 2005.
- [15] M. D. Lemmon and Q. Ling. Control system performance under dynamic quantization: The scalar case. In *Proceedings of the IEEE Conference on Decision and Control*, 2004.
- [16] D. Liberzon. On stabilization of linear systems with limited information. *IEEE Transactions on Automatic Control*, 48(2):304–307, 2003.
- [17] D. Marco and D. L. Neuhoff. The validity of the additive noise model for uniform scalar quantizers. *IEEE Transactions on Information Theory*, 51:1739–1755, May 2005.
- [18] G. N. Nair, S. Dey, and R. J. Evans. Infimum data rates for stabilising Markov jump linear systems. In *Proc. of the 42nd IEEE Conference on Decision and Control*, pages 1176–81, 2003.
- [19] G. N. Nair and R. J. Evans. Stabilizability of stochastic linear systems with finite feedback data rates. *SIAM Journal on Control and Optimization*, 43(2):413–436, July 2004.
- [20] I. R. Petersen and A. V. Savkin. Multi-rate stabilization of multivariable discrete-time linear systems via a limited capacity communication channel. In *Proc. of the 40th IEEE Conference on Decision and Control*, pages 304–309, 2001.
- [21] D. K. Sharma. Design of absolutely optimal quantizers for a wide class of distortion measures. *IEEE Transactions on Information Theory*, IT-24(6):693–702, November 1978.
- [22] S. Tatikonda. *Control under Communication Constraints*. PhD thesis, MIT, Cambridge, MA, 2000.
- [23] E. Verriest and M. Egerstedt. Control with delayed and limited information: A first look. In *Proceedings of the IEEE Conference on Decision and Control*, 2002.
- [24] W. S. Wong and R. W. Brockett. Systems with finite communication bandwidth-part I: State estimation problems. *IEEE Transactions on Automatic Control*, 42(9):1294–1298, 1997.
- [25] W. S. Wong and R. W. Brockett. Systems with finite communication bandwidth-part II: Stabilization with limited information feedback. *IEEE Transactions on Automatic Control*, 44(5), 1999.