2008 American Control Conference
Westin Seattle Hotel, Seattle, Washington, USA
June 11-13, 2008

Optimal LQG Coding and Control over Communication Channels: An Existence and an Infeasibility Result

Serdar Yüksel

Abstract— **This paper studies the problem of joint coding for and control of a plant over a discrete or continuous noisy memoryless channel. One main result the paper presents is that under certain assumptions, optimal coding and control policies exist and are stationary. It is also shown that for the remote control of an open-loop unstable system over a discrete noisy communication channel, if the system is driven by a Brownian disturbance, stationary memoryless coding schemes or innovation coding schemes lead to an unstable system. This result is true even when the capacity of the discrete channel is arbitrarily high, but finite. In particular, sensor and controller designs for control of systems driven by noise over communication networks when modeled as continuous alphabet erasure channels, lead to instability when are applied to realistic channels. Conditions ensuring stability leading to a recurrent Markov chain are presented. To achieve this, one needs to use coding policies with non-trivial memory constructions.**

I. Introduction

The use of discrete channels such as the Internet or bus lines (as in a Controller Area Network (CAN)) in control systems have become common place. Some salient examples include links in vehicle systems, large-scale printers and aerospace applications. The presence of such channels makes the traditional control approaches such as the principle of separation of estimation and control (due to the possible dual effect of control [1]), Kalman filtering (due to non-Gaussian disturbances and the dual effect of the control), linear control design (due to the non-classical information structure [2]) inapplicable or inefficient. There has been a period of scholar productivity in this area under different models and assumptions on the classes of systems considered. See [3] for some references.

This paper studies the problem of optimal control over a discrete noisy channel. The main contributions are an existence result, some sufficiency results, and a negative result.

In the following we describe the system model.

II. Model

We consider the discrete-time system

$$x_{t+1} = ax_t + bu_t + d_t \qquad (1)$$

where x_t is the state at time t, and $\{d_t\}$ is a sequence of zero-mean independent, identically distributed (i.i.d.) Gaussian random variables. The controller action at time t, u_t, is generated by partial access to the state of the system (Figure 1).

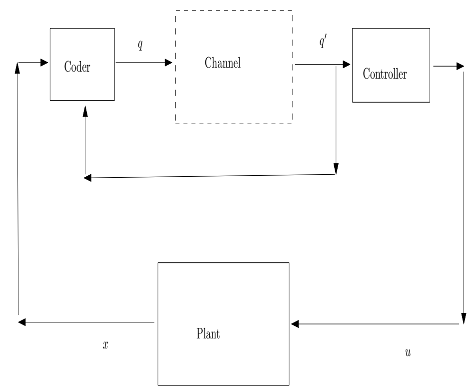

Fig. 1: Control over a noisy channel with causal feedback.

A. Problem Formulation

The problem that we are interested is the minimization of the infinite horizon average cost:

$$\limsup_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} E_{x_0}[x_t^2 + ru_t^2], \qquad (2)$$

where $E_{x_0}[]$ denotes the expectation over all sample paths with initial state given by x_0 and $r > 0$. To obtain certain specific results, we will also consider a quadratic cost functional with $r = 0$ in the development of the paper.

The optimization is over all admissible sensor (quantizer) and controller policies such that the sensor policies are measurable with respect to the sigma algebra generated by I_t^s:

$$I_t^s = \{I_{t-1}^s, x_t, I_t^c\},$$

with $I_0^s = \{x_0\}$ are are maps to $\mathcal{Q}$, the set of quantizers which will be described in the next section. The control policies are measurable functions of I_t^c:

$$I_t^c = \{I_{t-1}^c, q_t'\},$$

with $I_0^c = \{q_0'\}$.

B. Sensor Quantizer Classes for Discrete Channels

A channel coder maps the source symbols, state values, to corresponding channel inputs. This is done via quantizers in case the channel under consideration is discrete.

Definition 2.1: A *quantizer*, Q, for a scalar continuous variable is a Borel-measurable mapping from the real line to

Mathematics and Engineering, Department of Mathematics and Statistics, Queen's University, Kingston, ON K7L 3N6, Canada. E-mail:yuksel@mast.queensu.ca. Research supported by a Queen's University Research Initiation and the Principal's Advisory Research Committee (ARC) Grants.

a finite, characterized by corresponding bins $\{\mathcal{B}_i\}$ and their reconstruction levels $\{q^i\}$, such that $\forall i$, $Q(x) = q^i$ if and only if $x \in \mathcal{B}_i$. $\diamond$

We take $\mathcal{B}_i$ can be taken to be nonoverlapping semiopen intervals, $\mathcal{B}_i = [\delta_i, \delta_{i+1})$, with $\delta_i < \delta_{i+1}, i = 0, 1, \ldots, m$, where $m = |\mathcal{M}|$ is the cardinality of the channel input alphabet. We also have that $\delta_0 = -\infty, \delta_m = +\infty$. Here $\{\delta_i\}$ are termed "bin edges". Hence $\{\delta_i\} \in l_{\bar{m}}$; the space is an m-dimensional real vector space. We do not rule out $\delta_i = \delta_j$, in which case, all are represented as δ_i, effectively reducing the number of levels in the quantizer. We will also impose a compactness condition later when studying stable systems.

The quantizer outputs are transmitted via a noisy memoryless channel, hence the receiver has access to noisy versions of the quantizer/coder outputs for each time, which we denote by $\{q'^i\}$ generated according to a probability distribution for every fixed q^i. In this case, the quantizer design, in addition, needs to account for the errors in the transmission. Hence, the quantizer is in essence a *joint-source-channel* code.

In a dynamic, discrete-time setting, the construction of a quantizer at any time t could depend on the past quantizer values. In the following, we present five different classes of quantizers, successively being more restrictive.

1) Causal Quantizers: Let $\mathbb{R}$ be the input and the output space, and $\mathcal{Q}$ be the set of quantizers. Then, a causal quantizer policy Π_C at time t, to be denoted by f_t, is a mapping from $\mathbb{R}^{t+1} \times \mathcal{Q}^t$ to $\mathcal{Q}$, where $\mathbb{R}^s$, $s \in \mathbb{Z}$ is the s-product of $\mathbb{R}$. Such a quantizer is said to be *causal* (that is, it depends only on the past and present values of the input process and the channel outputs), in addition to being *dynamic*. Finally, we also introduce $g_t : \mathcal{Q}_0 \times \mathcal{Q}_1 \times \ldots \mathcal{Q}_t \rightarrow \mathbb{R}$ as the *decoder function*, which again has causal access to the past received values. The class of quantizers we have introduced above are deterministic causal quantizers, in the sense that for each fixed t, and given f_t and history $h_t := \{x_s, f_s \;\; s = 0, \ldots, t-1\}$ and x_t, the quantizer induced, $f_t(x_0^t, {q'_0}^{t-1})$, is a uniquely defined element of $\mathcal{Q}$. Let $\sigma(\mathcal{Q})$ be the σ-algebra of the set of quantizers. A more general class quantizers is the *randomized* ones, which assign a probability measure to selection of quantizers.

2) Markov Quantizers: A more restrictive class of quantizers are Markov quantizers, Π_{MR}. Let $\mathcal{P}$ be the set of conditional probability densities on $\mathbb{R}$):

$$P(x_t | f_0^{t-1}, {q'_0}^{t-1}).$$

A randomized Markov quantizer policy satisfies

$$P(.|\pi_t = \pi, t) = P(.|h_t, \pi_t = \pi), \quad a.s. \quad \forall h_t(w)$$

for each $\pi \in \mathcal{P}$, with $h_t(w)$ denoting the sample paths for the history process and that for each $\pi \in \mathcal{P}$, $q(.|\pi)$ is a probability measure on $\sigma(\mathcal{Q})$. For a deterministic Markov quantizer, $P(.|\pi_t, t)$ is a dirac measure.

3) Stationary Quantizers: A further restrictive class of quantizers are stationary quantizers Π_{SR}. Stationary quantizers can be both deterministic and randomized. Let $\mathcal{P}$ be the set of conditional probability densities on $\mathbb{R}$):

$$P(x_t | f_0^{t-1}, {q'_0}^{t-1}).$$

A randomized stationary quantizer policy satisfies

$$P(.|\pi_t = \pi) = P(.|h_t, \pi_t = \pi), \quad a.s. \quad \forall h_t(w)$$

for each $\pi \in \mathcal{P}$, with $h_t(w)$ denoting the sample paths for the history process and that for each $\pi \in \mathcal{P}$, $q(.|\pi)$ is a probability measure on $\sigma(\mathcal{P})$. For a deterministic stationary quantizer, $P(.|\pi)$ is a dirac measure.

4) Innovation Quantizers: An innovation quantizer policy Π_I is a time-invariant quantizer which has

$$e_t = x_t - aE[x_{t-1} | f_0^{t-1}, {q'_0}^{t-1}],$$

as its input argument. Hence, an innovation quantizer policy satisfies

$$P(.|e_t = x) = P(.|h_t, e_t = x), \quad a.s. \quad \forall h_t(w)$$

for each $x \in \mathbb{R}$.

5) Memoryless Quantizers: A further restrictive class of quantizers are memoryless quantizers $\mathcal{Q}_{MS}$. A randomized memoryless stationary quantizer policy satisfies

$$P(.|x_t = x) = P(.|h_t, x_t = x), \quad a.s. \quad \forall h_t(w)$$

for each $x \in \mathbb{R}$ and that for each $x \in \mathbb{R}$, $q(.|x)$ is a probability measure on $\mathcal{B}(R)$, and if, for every fixed $D \in \mathcal{B}(R)$, $q(D|.)$ is a well-defined function on $\mathbb{R}$.

C. Separation of Estimation and Control

It follows that the sensor (encoder) has access to the information available at the controller (decoder). The information structure thus falls within the class of nested information structures, for which it is known that, an optimal control problem admits a dynamic program, and there is no dual effect of control, as the control action is available at the sensor. The control is solely acting on the minimization of the infinite horizon cost, and the sensor is acting so facilitate the improvement on the estimation. The control action will be based on an intermediate partial observation: $P(x_t | {q'_0}^t)$ and will be solving a separate MDP problem given the statistics of the partial observations.

We review some notions from the theory of Markov Processes. A Markov chain with an invariant probability distribution is recurrent. A Markov channel is transient if the probability of return time of the state to a compact set being finite is less than one.

Definition 2.2: An open set $C \subset R$ is transient if

$$P\Bigg(\min(t > 0 : x_t \in C) < \infty | x_0 \in C \Bigg) < 1.$$

$\diamond$

Definition 2.3: A Markov chain is Lebesgue irreducible if for every open set $\mathcal{C} \subset \mathbb{R}$, $P(x_t \in \mathcal{C} | x_{t-1}) > 0$.

Under irreducibility, we define a Markov chain to be transient if there exists an an open set which is transient.

III. Instability Under Memoryless Policies

In a discrete memoryless channel setting, at any given time, the channel input has to have a finite cardinality. This observation leads to the following result:

Theorem 3.1: Suppose a discrete-time linear system as in (1), with $|a| > 1$, is being controlled over a discrete channel. A memoryless coding policy over a finite, but arbitrarily high, capacity leads to a transient Markov chain.

This theorem was essentially proved in [7] and [3], in the context of memoryless coding and control. The result extends to the current setting. Also see [6] for a similar discussion in noiseless channels.

This result immediately leads to the following corollary to Theorem 3.1:

Corollary 3.1: For the optimization problem (2), stationary memoryless policies are suboptimal.

IV. Performance of Innovation Coding Policies and Instability of Open-Loop Unstable Systems

Innovation coding is a widely popular coding scheme, and has variations, such as differential coding, predictive coding and $\Delta-$modulation, which has also recently been investigated in [33]. Furthermore, for the Gaussian setting, it is know that innovation coding is optimal, which was observed in [10]. It should also be noted that the celebrated Schalkwijk-Kailath [29] coding scheme also admits an innovation coding interpretation.

Control systems connected over Controller Area Networks use packets to communicate, and in certain applications these packets can be assumed to be losslessly transmitted. These packets usually have high numbers of bits. The following result shows that, with the further condition that the source has an invariant distribution and hence is *stable*, innovation coding can be very efficient at high-data rates. A modification of the analysis below can also be applied to communication networks with erasure information made available to the transmitter.

Theorem 4.1: For causal coding of the source given in (1) with stable dynamics $|a| < 1$, a lower bound on the average cost function $\limsup_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} x_t^2$, subject to a discrete noiseless channel with capacity R, is given by:

$$D(R) \geq D_{lower}(R) := a^2 \frac{\sigma^2}{(2^{2R} - a^2)} + E[d_t^2]$$

When, the channel capacity is high, then there exists a coding scheme such that the cost satisfies:

$$D(R) \leq D_{innovation}(R) := a^2 \frac{2\pi e \sigma^2}{(122^{2R} - 2\pi e a^2)} + E[d_t^2]$$

Furthermore, $D_{innovation}(R)$ is achieved via innovation coding.

However, for this bound to be valid, it is necessary that the source is stable. Unfortunately, these results do no extend to discrete noisy channels. We have a corollary to Theorem 3.1, the proof of which follow very similar steps, and hence is omitted.

Corollary 4.1: For causal coding of the source given in (1) with unstable dynamics $|a| \geq 1$, suppose a memoryless innovation coder is used. This leads to a transient Markov chain.

It should be noted that the arguments used above via Comparison theorem leading to the recurrence of the chain do not apply for the open-loop unstable sources. We will observe, in the following sections that, the results connected to the lower bound presented above applies for certain continuous alphabet channels, even when the source is open-loop unstable.

Upon observing these shortcomings, we now proceed to design the optimal coders.

V. Optimal Time-Varying Coders: Markov and Stationary Quantizers

We have two realistic assumptions that are used for the existence result.

Assumption A: The filter at the decoder has arbitrarily large, but finite memory: There exists $d \geq 1$ such that

$$\pi'_t(x) := P(x_t | f_{t-d}^{t-1}, q_{t-d}^{t-1}) = P(x_t | f_1^{t-1}, q_1^{t-1}).$$

$\diamond$

Assumption B: The set of quantizer bin edges satisfy $\sup_{\mathcal{Q}} |\delta_i| < M$, for some arbitrarily large but finite M. $\diamond$

The above two assumptions will be used in the existence result.

We observed that the main technical problem is the existence of an invariant distribution. Toward this end we provide the evolution of the conditional density of the state as seen by the decoder. Properties of conditional probability leads to the following expression for

$$\pi_n(x) := P(x_n = x | f_0^{n-1}, {q'}_0^{n-1})$$

$$\frac{\int \pi_{n-1}(x_{n-1}) P(.,.|.,.) P(x_n|x_{n-1}) dx_{n-1}}{\int\int \pi_{n-1}(x_{n-1}) P(.,.|.,.) P(x_n|x_{n-1}) dx_n dx_{n-1}}.$$

with $P(.,.|.,.)$ standing for $P(f_{n-1}, q'_{n-1} | \pi_{n-1}, x_{n-1})$, in the above.

Let $\mathcal{P}$ be the space of such probability distributions on $\mathbb{R}$ for the system considered, for $\pi_n(x), n \geq 1$, endowed with the topology of weak convergence [20] (we will see that for this space, uniform sup convergence may be used to develop certain properties). Then the conditional density and the quantization output process, $(\pi_n(x), f_n)$, form a joint Markov process in $\mathcal{P} \times \mathcal{Q}$.

Lemma 5.1: Under Assumptions A and B, the set of conditional density functions is uniformly tight, uniformly bounded and uniformly equi-continuous.

The above follows from the fact that the source is stable, the system noise has a uniformly continuous density (as a result of the effect of the Gaussian noise in the time update in filtering) and Assumption B.

Lemma 5.2: There is a countable set which is dense in the space of continuous and bounded functions on $\mathcal{P}$.

Proof: Pick a compact set $K_c = [-c, c]$. Let $\mathcal{P}_c$ be a compact support restriction of the elements in $\mathcal{P}$ to K_c. The restriction of all the density functions in this compact set will have bounded support, furthermore be totally bounded and uniformly equi-continuous. This is the Arzela-Ascoli characterization of compactness ([20], Section 7). It follows that there is a countable set that is dense in the set of continuous and bounded functions on $\mathcal{P}_c$, which we call $\mathcal{C}_b(\mathcal{P}_c)$. This can be shown by the construction of a countable approximation of any element $f(\pi)$ in $\mathcal{C}_b(\mathcal{P}_c)$ as a result of compactness. Now, let us increase the support set, by taking the truncation to be $c = c+1$ and the set to be $\mathcal{C}_b(\mathcal{P}_{c+1})$, and continue this iteration countably many times. Since the countable union of countable sets is countable, it follows that there is a countable set, which is dense in the space of continuous and bounded functions on $\mathcal{P}$, $C_b(\mathcal{P})$. $\diamond$

Theorem 5.1: For the system in (1), if $|a| < 1$, there exists an optimal stationary quantizer. Furthermore, the optimal stationary quantizer, is the optimal causal quantizer over all possible admissible policies.

We prove this result in the remainder of the section:

Proof:

Let $\mathcal{Q}_{SR}$ denote the set of stationary, randomized quantizer policies. Let $\mathcal{G}_\pi$ denote the set of ergodic occupation measures on the state, that is the set which satisfies the following for all continuous and bounded functions $f : \mathcal{P} \to R$:

$$\{v : \int f(\pi')v(d\pi') = \int (\int f(\pi)P^{\Pi_{SR}}(\pi|\pi')d\pi)v(d\pi')\},$$

where $P^{\Pi_{SR}}(\pi|\pi')$ stands for the transition kernel under quantizer policy Π_{SR} and for all $\Pi_{SR} \in \mathcal{Q}_{SR}$.

We now prove that this set is sequentially compact, and convex. We showed that $C_b(\mathcal{P})$ has a dense set which is countable. Now, a diagonalization argument can be used to show that the set of measures on $\mathcal{P}$, $P(\mathcal{P})$ is sequentially compact [20]: Every sequence $\{< v_n, f >\}$ has a converging subsequence, and by taking a further subsequence countably many times, and since there is a countable dense set to approximate every $f \in C_b(\mathcal{P})$ the desired result follows. Finally, we show that, the the cost function is continuous in v_n. This will imply the existence of a solution. Since the space of occupation measures is closed and connected, this can be achieved. But this is immediate from the definition of continuity: if v_n converges to v weakly, then $< v, c >$ also converges, with c being the second moment of the conditional estimation error, which is bounded and continuous on $\mathcal{P}$.

Finally, we show that the space of ergodic occupation measures is convex, so that any point is reachable from another point: By the arguments above, there is an optimal occupation measure. This measure can be achieved from any initial condition of states: Take a π^* which is visited infinitely often under a policy. Under the finite memory assumption, such a point exists, since the the visited states are atomic due to the finite memory assumption. One applies a quantizer such that this point is visited once, and then one applies the optimal quantizer policy thereafter.

$\diamond$

Let Γ denote the steady state distribution of $\phi_t = (f_t, \pi_t)$, and let $< v, z >$ denote the inner product between two functions v and z. One would then have the minimization of

$$< \Gamma, C > \tag{3}$$

over all admissible quantizers with a given number of levels, where $C(\phi)$ is the second moment of the estimation error applied to the conditional density.

By the infinite-dimensional linear programming approach, an optimal stationary policy lies on the extremal points of the convex set $\mathcal{G}_\Pi$:

Proposition 5.1: The optimal sensor and controller policies for the optimal control problem are deterministic.

VI. Existence of an Invariant Distribution

The following Lemma is a required technical result [3]:

Lemma 6.1: Let $\mathcal{P}$ be a set of uniformly continuous probability distributions on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ such that for some $M < \infty$, $\int P(x)x^2dx < M$, for all $P(\cdot) \in \mathcal{P}$ and $P(x)\log_2(P(x))$ is uniformly integrable over $\mathcal{P}$. Consider a sequence of random variables $\{X_t\}$ with densities $\{p_t(x_t)\}$ in $\mathcal{P}$. If $X_t \to X$ weakly, then the differential entropy satisfies $h(x_t) \to h(x)$.

We have that the conditional density process is recurrent, and hence $P(x_t|q_0'^t)$ does converge. Furthermore, this process lives in a uniformly integrable set: thus, the Lemma above can be invoked and there exists a limit for the conditional entropy sequence.

We have the following theorem:

Theorem 6.1: For the existence of an invariant density with a finite second moment, with

$$\lim_{t\to\infty} E[x_t^2] \le d,$$

the channel capacity should satisfy:

$$C \ge \frac{1}{2}\log(\frac{a^2 d}{d - \sigma^2}).$$

Clearly, this expression for the capacity presents a relationship between the achievable distortion and the information rate needed. Most of the time, however, the results utilizing entropy are unattainable. There are two immediate reasons for this: one due to the fact that codes achieving such a bound need to be codes of block length one, and the second one is due to the presence of the disturbance which does not let an invariant distribution exist when $|a| > 1$. We will observe that this bound is tight for Gaussian sources connected over Gaussian channels. As for the second issue, time-varying coding schemes, as presented in [9] and schemes as such [13] are necessary. In the following, we present a stochastic drift argument for the existence of a recurrent chain.

Theorem 6.2: Let

$$V(\pi) = E_\pi[x^2] = \int \pi(x)x^2dx,$$

where the expectation over the distribution $\pi(x)$. If a stationary quantizer policy satisfies, for some $\epsilon > 0, K < \infty$, and some set C weak* compact,

$$E[V(\pi_{t+1}|\pi_t, f_t] \leq (1-\epsilon)V(\pi_t) + 1_{\pi_t \in C}K$$

Then, the process is positive recurrent in the sense that,

$$P\bigg(\min\{t : \pi_t \in C | \pi_0 \in C\} < \infty\bigg) = 1$$

Proof:

Proof follows form stochastic drift characterization, and an extension of the comparison theorem due to Meyn and Tweedie [16].

It should be observed that, for $C_v := \{\pi : V(\pi) \leq v\}$, which is a weak* compact set (this can be proved following a similar discussion as in the proof of Theorem 5.1),

$$E[V(\pi_t|\pi_t, f_t] \leq (1-\epsilon)V(\pi_t) + 1_{\pi_t \in C}K$$

implies the weaker statement that

$$E[V(\pi_t)|\pi_t, f_t] \leq V(\pi_t) - c + 1_{\pi_t \in C_v}K'$$

for some constants c and K'. This follows from the fact that $C_v \subset C_u$ for $v < u$. Further, define $M_0 := V(\pi_0)$ and

$$M_k := V(\pi_k) + \sum_{i=0}^{k-1}(-c + K1_{\pi_i \in C})$$

Hence,

$$E[M_{(k+1)}|h_k] \leq M_k, \quad \forall k \geq 0$$

and thus, $\{M_k\}$ forms a supermartingale. Define a stopping time:

$$\tau^N = \min(N, \min\{i : V(\pi_i) + (1-\epsilon)V(\pi_i) + K1_{\pi_i} \geq N\})$$

This process is bounded in this interval. Hence, we have, by the optional sampling theorem:

$$E[M_{(\tau^N)}] \leq E[M_0]$$

Hence, we obtain

$$E[\sum_{i=0}^{\tau^N-1}]c \leq V(\pi_0) + KE[\sum_{i=0}^{\tau^N-1} 1_{\pi_i \in C}]$$

Hence, $cE[\tau^N - 1 + 1] \leq V(\pi_0) + K$, and by the Monotone convergence theorem,

$$c \lim_{N\to\infty} E[\tau^N] = cE[\tau] \leq V(\pi_0) + K$$

Hence the result follows. $\diamond$

One needs to come up with code-designs satisfying the drift condition in Theorem 6.2. The main challenge is to ensure that such an invariant distribution exists when the source is unstable. We observed that using memoryless policies lead to instability. One could use state-dependent stochastic drift arguments [3], [9] for the existence of a recurrent chain. The important observation to make is that, one needs to use time-varying quantization schemes; or state-dependent policies. It should be noted that, the time-varying coding schemes in [3], the zooming-out technique in [28], and schemes proposed in [13] can be interpreted from the viewpoint of Theorem 6.2.

VII. Continuous Alphabet Channels

The issues presented above with regard to transience of the chain do not arise in continuous alphabet channels. This is because any element in the source can be transmitted, hence the channel can be matched to the source. The problem with Gaussian channels was considered in [10], [22], [19], where it has been shown that innovation coders are optimal. For the Gaussian setup, a power constraint is associated to the encoder output, instead of a quantization bit constraint.

Hence, even when the open-loop source does not have a finite invariant density, the system can be stabilized via an innovation coder. We now show that the bound in (6.1) is tight for Gaussian policies. Our proof can be interpreted as an alternative derivation of the results presented in [10], [22] and [32], regarding the optimality of innovation coding of Gaussian first-order dynamic systems over Gaussian channels with feedback.

Lemma 7.1: For the optimization problem, with a stationary distortion d, if x_0, w_0 are Gaussian and the channel is a Gaussian channel with capacity C, the optimal stationary distortion and the capacity satisfy

$$C = \frac{1}{2}\log(\frac{a^2 D + \sigma^2}{D})$$

where $D = (d - \sigma^2)/a^2$

Proof: The bound in Theorem 6.1 provides a converse theorem, which is a lower bound for the Gaussian case as well. In the following we show that this bound is attainable via linear policies. Proof follows from a standard argument in joint source channel coding for Gaussian sources and channels [10]. Let x be a Gaussian random variable transmitted over a Gaussian channel with capacity C, and let $y = u + v$, with u being the channel input, v channel noise. The minimum estimation error covariance D_G of a Gaussian source transmitted over a Gaussian channel satisfies the following set of inequalities:

$$C = \sup_{p(u)} I(u;y) \geq I(x; E[x|y]) \geq \frac{1}{2}\log_2(1 + E[x^2]/D_G)$$

The above follow from the definitions of capacity, rate-distortion function $R_{D_G}(x)$, and the directed data processing inequality. See [10] for details. Furthermore, these all become tight and the minimum attainable distortion is given by $D_G = \frac{E[x^2]}{2^{2C}}$. Let the Gaussian channel have input power constraint P, and the Gaussian channel noise be $\{v_t\}$. Clearly, $C = \frac{1}{2}\log(1 + P/\sigma_v^2)$. This is achieved by a linear scaling of the input. In particular, the coder applies $z_t = \alpha(x_t - aE[x_{t-1}|I_{t-1}^c])$, with

$$\alpha = \frac{P}{E[(x_t - aE[x_{t-1}|I_{t-1}^c])^2]}.$$

The optimal decoder policy leads to

$$E[x_t|I_t^c] = aE[x_{t-1}|I_{t-1}^c]) + \frac{P}{P + \sigma_v^2}(z_t + v_t).$$

In this case, the estimation error satisfies the recursion

$$E[(x_t - E[x_t|I_t^c])^2] = \frac{E[a^2(x_{t-1} - E[x_{t-1}|I_{t-1}^c])^2] + \sigma_w^2}{1 + P/\sigma_v^2}.$$

Upon recognizing the capacity expression in the denominator: $C = \frac{1}{2}\log_2(1 + P/\sigma_v^2)$, the result follows. $\diamond$

Hence, the analysis of communication networks, driven by possibly unbounded noise, requires the treatment that we made earlier in this paper. This brings the question of using alternative, time-varying coding and control schemes; such as binning and variable length coding [3], [13].

VIII. Conclusion

In practice, all channels are discrete. Even Gaussian channels require modulation techniques and digitization: Despite the fact that capacity achieving distributions are Gaussian, the encoding schemes are capacity achieving infinite blocks generated via a random Gaussian codebook, and hence, in essence are discrete. We also observed the lack of robustness in the assumption of regarding high-rate packet networks as continuous alphabet channels; in practice, such a design leads to instability almost surely.

Capacity, entropy and Bode-integral related arguments have been presented in the literature as fundamental bounds, but these only have operational meanings for channels which require codebooks of length one to meet the rate-distortion and capacity theorems, as they do for the transmission of Gaussian sources over Gaussian channels, and as such they are very limited as they only are applicable for such sources. As had been observed by Walrand and Varaiya, and Sahai and Mitter, capacity is not an appropriate measure for control. Hence, there is a need for an operationally significant approach. We tried to present an alternative look at this problem within this perspective.

References

[1] E. Tse Y. Bar-Shalom, “Generalized certainty equivalence and dual effect in stochastic control”, *IEEE Trans. Automatic Control*, vol. 19, pp. 494 - 500, Oct. 1974.

[2] H. S. Witsenhausen, “A Counterexample in Stochastic Optimum Control”, *SIAM Journal on Control and Optimization* vol. 6, pp. 131-147, 1968.

[3] S. Yüksel and T. Başar, “Control over Noisy Forward and Reverse Channels,”*IEEE Trans. Automatic Control*, submitted, 2007.

[4] V. S. Borkar and S. K. Mitter, “LQG Control with Communication Constraints” in *Kailath Festschrift*, pp. 365-373, Kluwer Academic Publishers, Boston 1997.

[5] S. Yüksel and T. Başar, “Coding and control over discrete noisy forward and feedback channels,” in *Proc. IEEE Conference on Decision and Control/European Control Conference*, Sevilla, Spain, pp. 2517-2522, December 2005.

[6] G. N. Nair and R. J. Evans, “Stabilizability of stochastic linear systems with finite feedback data rates”, *SIAM Journal on Control and Optimization*, vol. 43, pp. 413 - 436, July 2004.

[7] S. Yüksel and T. Başar “Achievable Rates for Stability of LTI Systems over Noisy Forward and Feedback Channels”, *Proc. Conference on Information Sciences and Systems*, Baltimore, March 2005.

[8] A.S. Matveev and A. V. Savkin, “Shannon Zero Error Capacity and the Problem of Almost Sure Observability over Noisy Communication Channels”, in *Proc. IEEE CDC and ECC*, 2005, Seville, Spain, pp: 3219 - 3224.

[9] S. Yüksel and T. Başar, “Control of Multi-Dimensional Linear Systems over Noisy Channels” *Proc. IEEE Conference on Decision and Control*, San Diego, CA, 2006.

[10] R. Bansal and T. Başar, “Simultaneous design of measurement and control strategies for stochastic systems with feedback,” *Automatica*, vol. 25, pp. 679-694, September 1989.

[11] A. Manne, “Linear programming and sequential decision,” *Management Science*, 6, 1960.

[12] V. S. Borkar, “Convex Analytic Methods in Markov Decision Processes,” *Handbook of Markov Decision Processes: Methods and Applications*, Kluwer, Boston, MA.

[13] A. Sahai and S. Mitter, “The necessity and sufficiency of anytime capacity for stabilization of a linear system over a noisy communication link Part I: scalar systems,” *IEEE Trans. Inform. Theory*, vol. 52, pp. 3369-3395, Aug. 2006.

[14] N. C. Martins “Finite Gain lp Stabilization Requires Analog Control”, *Systems and Control Letters*, vol. 55, pp. 949-954, November 2006.

[15] J. C. Walrand and P. Varaiya, “Optimal causal coding-decoding problems,” *IEEE Trans. Information Theory*, vol. 19, pp. 814-820, Nov. 1983.

[16] S. P. Meyn and R. Tweedie, *Markov Chains and Stochastic Stability*, Springer Verlag, London (1993).

[17] S. Yüksel and T. Başar, “Optimal Signaling Policies for Decentralized Multi-Controller Stabilizability over Communication Channels”, *IEEE Trans. Automatic Control*, vol. 52, pp. 1969-1974, October 2007.

[18] S. Yüksel and T. Başar, “Communication Constraints for Decentralized Stabilizability with Time-Invariant Policies”, *IEEE Trans. Automatic Control*, vol. 52, pp. 1060-1066, June 2007.

[19] C. D. Charalambous and A. Farhadi, “A Mathematical Framework for Robust Control Over Uncertain Communication Channels,” in *Proc. IEEE Conference on Decision and Control/European Control Conference*, Sevilla, Spain, pp. 2530-2535, December 2005.

[20] P. Billingsley, *Convergence of Probability Measures*, New York, NY, John Wiley, 1968.

[21] F. Fagnani and S. Zampieri, “Stability analysis and synthesis for scalar linear systems with a quantized feedback,” *IEEE Trans. Automatic Control*, vol. 48, pp.1569-1584, Sept. 2003.

[22] S. Tatikonda and S. Mitter, “Control under communication constraints,” *IEEE Trans. Aut. Control*, vol. 49, pp. 1056-1068, July 2004.

[23] V. S. Borkar, S. K. Mitter, and S. Tatikonda “Optimal sequential vector quantization of Markov sources,” *SIAM J. Control and Optimization*, vol. 40, pp. 135-148, 2001.

[24] O. Hernandez-Lerma, J. Gonzales-Hernandez, and R. R. Lopez-Martinez, “Constrained average cost Markov control processes in Borel spaces,” *SIAM J. Control and Optimization*, vol. 42, pp. 442-468, 2003.

[25] G. N. Nair and R. J. Evans, “Exponential stabilisability of multidimensional linear systems with limited data rates,” *Automatica*, vol. 39, pp. 585-93, Apr. 2003

[26] A. Mahajan and Demosthenis Teneketzis “On jointly optimal encoding, decoding and memory update for noisy real-time communication systems,” *IEEE Transactions on Information Theory* submitted for publication.

[27] T. Linder and R. Zamir, “Causal coding of stationary sources and individual sequences with high resolution,”*IEEE Trans. Information Theory*, vol. 52, pp. 662-680, February 2006.

[28] D. Liberzon, “On stabilization of linear systems with limited information”, *IEEE Trans. on Automatic Control*, vol. 48, pp. 304-307, Feb 2003.

[29] J. P. M. Schalkwijk and T. Kailath, “A coding scheme for additive noise channels with feedbackPart I: No bandwidth constraint, *IEEE Trans. Inform. Theory*, vol. 12, pp. 172182, April 1966.

[30] N. C. Martins, M. A. Dahleh and N. Elia, “Feedback stabilization of uncertain systems in the presence of a direct link,” *IEEE Trans. Automatic Control*, vol. 51, pp. 438-447, March 2006.

[31] N. Elia, “When Bode meets Shannon: control-oriented feedback communication schemes,” *IEEE Trans. Automatic Control*, vol. 49, pp. 1477-1488, Sept. 2004.

[32] J.H. Braslavsky, R.H. Middleton and J.S. Freudenberg, “Feedback Stabilization over Signal-to-Noise Ratio Constrained Channels”, *IEEE Transactions on Automatic Control*, vol. 52, pp. 1391-1403, 2007.

[33] R.H. Middleton and G.C. Goodwin, “Improved finite word length characteristics in digital control using delta operators”, *IEEE Transactions on Automatic Control*, vol. 31, pp.1015-1021, Nov. 1986.