Optimal Design of Information Channels in Networked Control

Serdar Yüksel

Queen's University,
Department of Mathematics and Statistics

Stochastic Control

A controlled stochastic system is governed by the following state / measurement equations:

$$x_{t+1} = f(x_t, u_t, w_t), \tag{1}$$

$$y_t = g(x_t, v_t) \tag{2}$$

A control policy Π is a sequence of control functions $\{\gamma_0, \gamma_1, \cdots, \}$ each a causal function of the information vector

$$I_t = \{y_t; y_{[0,t-1]}, u_{[0,t-1]}\} \quad t \geq 1,$$

with control actions $u_t = \gamma_t(I_t)$.

Here, (2) defines a channel, a stochastic kernel.

Stochastic Control with Information Constraints

In stochastic control, typically a partial observation model/channel (parametrized above by $g(\cdot)$) is given and one looks for a control policy for optimization or stabilization.

In networked control systems, the observation channel itself and the information vector I_t are also design variables.

We can shape the observation / measurement channel, through encoding and decoding.

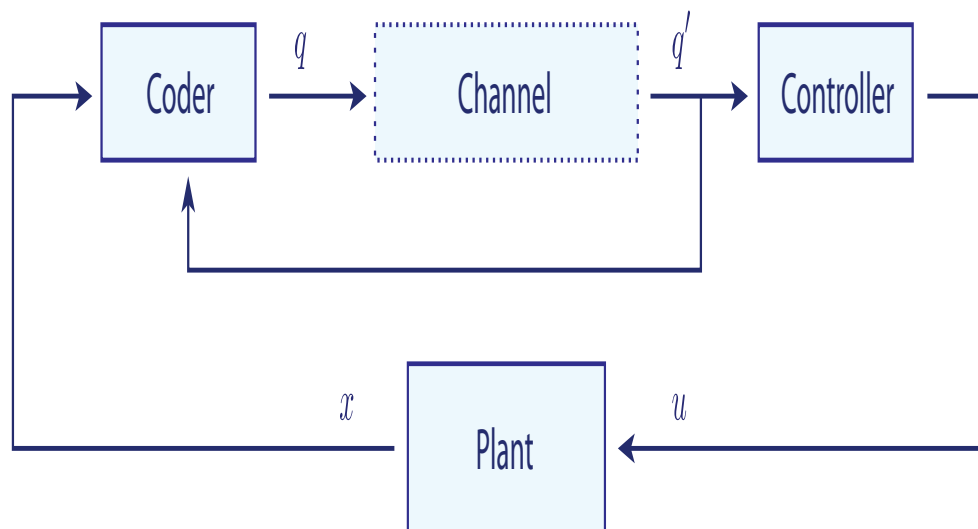# Stochastic Control with Information Constraints

Figure 1: Encoding shapes the conditional probability of the observation given the state.

Problem P1: Design of Information Channels for Stabilization

Given a system controlled over a channel, find the set of channels $\mathcal{Q}$ for which there exists a policy (both control and encoding), such that $\{x_t\}$ is stable.

Stochastic stability notions will be (i) ergodicity/asymptotic mean stationarity and (ii) existence of finite moments, to be specified later in this talk.

Problem P2: Design of Information Channels for Optimization

Given a controlled dynamical system, the goal is to minimize

$$J(P, Q) = E_P^{Q, \Pi} \left[\sum_{t=0}^{T-1} c(x_t, u_t) \right], \tag{3}$$

over the set of all admissible policies Π and channels in a family, $Q \in \mathcal{Q}$, and $c : \mathbb{X} \times \mathbb{U} \to \mathbb{R}_+$, a cost function.

Here $E_P^{Q, \Pi}$ denotes the expectation under Π and channel Q with initial prior P.

Problem P1: Design of Information Channels for Stabilization

We will consider a linear Gaussian unstable system model (results are applicable to higher-order systems)

$$x_{t+1} = ax_t + bu_t + d_t, \qquad t \geq 0 \tag{4}$$

It is assumed that $|a| \geq 1$ and $b \neq 0$.

This system is connected over a channel with a finite capacity to a controller.

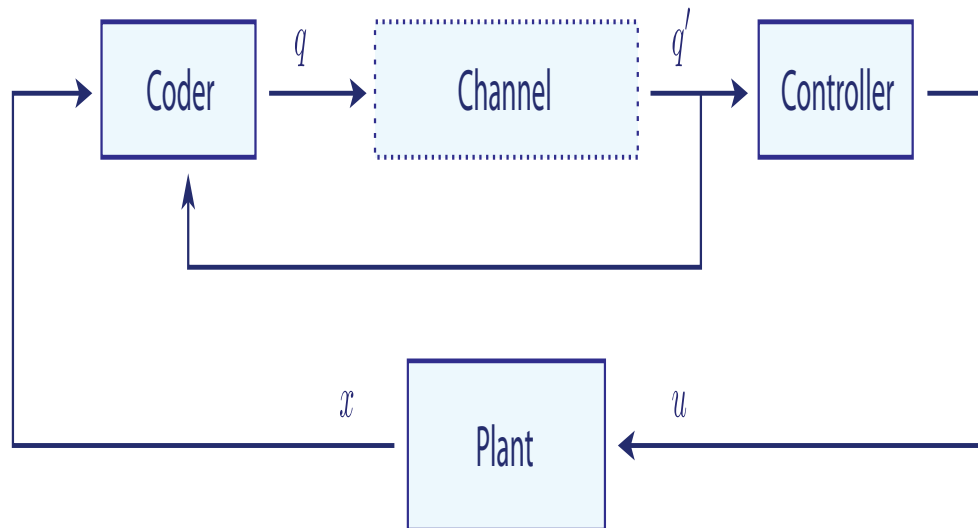# Causal Coding for Control

Figure 2: Control over a noisy channel

Information Channel

A discrete channel is a stochastic kernel such that for any $n \in \mathbb{N}$, an input sequence $q_{[0,n]}$ leads to an output $q'_{[0,n]}$ with probability $P(q'_{[0,n]} | q_{[0,n]})$.

The channel is memoryless, if (without feedback)

$$P(q'_{[0,n]} | q_{[0,n]}) = \prod_{k=0}^{n} P(q'_k | q_k).$$

Causal Coding for Control

The quantizer and the source coder policy is causal such that the channel input at time $t \geq 0$, q_t, is generated using the information:

$$I_t^s = \{x_{[0,t]}, q_{[0,t-1]}, q'_{[0,t-1]}\}$$

The quantizer outputs are transmitted through a channel, after being subjected to a channel encoder. The receiver has access to noisy versions of the quantizer/coder outputs for each time, which we denote by $q'_t \in \mathcal{M}'$.

The control policy at time t, also causal, only uses I_t^c, for $t \geq 0$:

$$I_t^c = \{q'_{[0,t]}\}$$

We will call such coding and control policies *admissible* policies.

Literature Review: Information Theory for Unstable Processes

Consider the following Gaussian AR process:

$$x_t = -\sum_{k=1}^{m} a_k x_{t-k} + w_k,$$

where $\{w_k\}$ is an independent and identical, zero-mean, Gaussian random sequence with variance $E[w_1^2] = \sigma^2$.

If the roots of:

$$H(z) = 1 + \sum_{k=1}^{m} a_k z^{-k}$$

are inside the unit circle, the process is (asymptotically) stationary.

Literature Review: Information Theory for Unstable Processes

The rate distortion function (distortion being the normalized Euclidean error) is given parametrically by the following (Kolmogorov'56) :

$$D_\theta = \frac{1}{2\pi} \int_{-\pi}^{\pi} \min(\theta, \frac{1}{g(w)}) dw,$$

$$R(D_\theta) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \max(1/2(\log \frac{1}{\theta g(w)}), 0) dw,$$

with

$$g(w) = \frac{1}{\sigma^2} |1 + \sum_{k=1} a_k e^{-ikw}|^2$$

Literature Review: Information Theory for Unstable Processes

If at least one root is on or outside the unit circle, $R(D_\theta)$ above should be replaced with (Gray (IT'70), Hashimoto-Arimoto (IT'80), Berger (IT'70)):

$$R(D_\theta) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \max\left(\frac{1}{2}\log(\frac{1}{\theta g(w)}), 0\right) dw + \sum_{k=1}^{m} \frac{1}{2} \max\left(0, \log(|\rho_k|^2)\right), \quad (5)$$

where $\{\rho_k\}$ are the roots of the polynomial.

Note that the encoding is non-causal.

Control Theory Literature and Causality Restrictions

Wong-Brockett (TAC'98), Nair-Evans (SICON'04), Tatikonda-Mitter (TAC'04) obtained that, in the mean-square sense, an average rate of information transmission needed for stabilizability is at least

$$\sum_{k=1}^{m} \frac{1}{2} \max \left(0, \log(|\rho_k|^2) \right)$$

Contrasting with the Gray/Hashimito-Arimoto result, this shows that the rate term is not due to the causality restriction, but due to the uncertainty inherent in the source (the differential entropy rate).

Causal Coding for Control: Presence of Unbounded System Noise with Noiseless Channels

Nair-Evans (SICON'04) considered a class of adaptive quantizer policies for such unstable linear systems driven by noise:

- with unbounded support set for its probability measure
- time-varying encoders
- controlled over noiseless channels, and obtained necessary and sufficient conditions for the boundedness of the following expression

$$\limsup_{t \to \infty} E[|x_t|^2] < \infty.$$

Gurt and Nair (Automatica'09) extended this result to erasure channels with variable rate coding.

Causal Coding for Control: Presence of Unbounded System Noise

With the lower bound attained, Y. (TAC'10) obtained the existence of a limit

$$\lim_{t \to \infty} E[|x_t|^2] < \infty,$$

- noise with unbounded support set for its probability measure
- fixed-rate encoders
- the process is stochastically stable in the sense that the joint process is a positive Harris recurrent Markov chain and the sample path ergodic theorem is applicable.

This was extended in Y.'09,Y.-Meyn (TAC'12) to erasure channels with similar ergodicity properties (Minero-Franceschetti-Dey-Nair'09's) result (variable-rate) shown to be sufficient in the same sense.

Causal Coding for Control: Noisy Channels

The particular notion of stochastic stability is critical in characterizing the conditions on the channel. Sahai-Mitter (IT'06) and Matveev-Savkin (SICON'07) considered almost sure stability and relation with zero-error capacity.

Sahai-Mitter (IT'06) also considered a characterization for reliability for controlling unstable processes, named, any-time capacity, defined for the finite moment criteria.

With a departure from the bounded noise assumption, Matveev (MCSS'08), considered a more general model of multi-dimensional systems driven by an unbounded noise process considering *stability in probability*: for large enough $p < 1$, there exists b s.t.

$$P(x_t \leq b) \geq p, \quad t \geq 0.$$

Martins-Dahleh (TAC'08), Sahai-Mitter (IT'06) and Matveev-Savkin (SICON'08) considered stability in probability also for bounded noise settings.

Causal Coding for Control: Noisy Channels

In today's talk (problem P1), the problem is to find, for the system

$$x_{t+1} = ax_t + bu_t + w_t,$$

the largest class of channels $\mathcal{Q}$, for which there exists a policy (both control and encoding), so that $\{x_t\}$ is stochastically stable:

When does an unstable linear system driven by unbounded noise, controlled over a channel (possibly with memory) is stochastically stabilizable in the following sense: Find $\{Q \in \mathcal{Q}\}$ for which there exist control and coding policies such that

- The ergodic theorem applies, the state process is asymptotically mean stationary.

- The state process has finite average second moment.

Stochastic Stability Notion: Asymptotic (Mean) Stationarity

Let $\mathbb{X} = \mathbb{R}^d$ and $\Sigma = \mathbb{X}^\infty$ denote the sequence space of all one-sided sequences drawn from $\mathbb{X}$. Thus, if $x \in \Sigma$ then $x = \{x_0, x_1, \dots\}$ with $x_i \in \mathbb{X}$.

Let $X_n : \Sigma \to \mathbb{X}$ denote the coordinate function such that $X_n(x) = x_n$.

Let T denote the shift operation on Σ, that is $X_n(Tx) = x_{n+1}$. $Tx = \{x_1, x_2, \dots\}$.

Definition .1. *A random process with measure P is $N-$stationary, (cyclo-stationary or periodically stationary with period N) if $P(T^N B) = P(B)$ for all $B \in \mathcal{B}(\Sigma)$. If $N = 1$, stationary.*

Definition .2. *A random process is $N-$ergodic if $A = T^N A$ implies that $P(A) \in \{0, 1\}$. If $N = 1$, it is ergodic.*

Stochastic Stability Notion: Asymptotic (Mean) Stationarity

Definition .3. *A process with a probability measure $(\Omega, \mathcal{F}, P)$ is asymptotically mean stationary (AMS) if there exists a probability measure $\bar{P}$*

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=0}^{n-1} P(T^{-k}F) = \bar{P}(F),$$

for all events F. Here $\bar{P}$ is the stationary mean of P.

This property is equivalent to the applicability of the ergodic theorem.

An N-stationary process is AMS.

Stochastic Stabilization over a DMC: Necessity and Sufficiency for AMS

Theorem .1. *[Y. (IT'12) + book chapter]*

(i) For stability over a DMC channel with any causal encoding and controller policy under the condition of the AMS property or that $\liminf_{t \to \infty} \frac{1}{t} h(x_t) \leq 0$, *the channel capacity must satisfy*

$$C \geq \log_2(|a|).$$

(ii) If

$$C > \log_2(|a|),$$

there exist coding and control policies such that the state process is AMS.

Also Applies for Channels with Memory

Let **Class A** be the set of channels which satisfy

a) the Markov chain condition:

$$q'_t \leftrightarrow q_t, q_{[0,t-1]}, q'_{[0,t-1]} \leftrightarrow x_{[0,t]},$$

for all $t \geq 0$ and

b) whose capacity with feedback is given by:

$$C = \lim_{T \to \infty} \max_{\{P(q_t|q_{[0,t-1]}, q'_{[0,t-1]}), \quad 0 \leq t \leq T-1\}} \frac{1}{T} I(q_{[0,T-1]} \to q'_{[0,T-1]}),$$

Theorem .1 is applicable except that for sufficiency, further conditions are needed, such as a channel which restarts itself (e.g. indecomposable Markov channels) and allows an exponential decay of error for rates less than capacity: see Tatikonda-Mitter (IT'09), Permuter-Weissman-Goldsmith (IT'09)).

Necessity: Sketch

The mutual information term satisfies

$$
\begin{aligned}
I(q_t'; q_{[0,t]} | q_{[0,t-1]}') \quad & = H(q_t' | q_{[0,t-1]}') - H(q_t' | q_{[0,t]}, q_{[0,t-1]}') \\
& = H(q_t' | q_{[0,t-1]}') - H(q_t' | q_{[0,t]}, x_t, q_{[0,t-1]}') \\
& \geq H(q_t' | q_{[0,t-1]}') - H(q_t' | x_t, q_{[0,t-1]}') \\
& = I(x_t; q_t' | q_{[0,t-1]}') \quad\quad\quad\quad\quad (6)
\end{aligned}
$$

Necessity: Sketch

$$
\begin{aligned}
C \; &\geq \; \limsup_{T \to \infty} (1/T) \left(\sum_{t=1}^{T-1} (x_t; q'_t | q'_{[0,t-1]})) + I(x_0; q_0) \right) \\[2mm]
&\geq \; \limsup_{T \to \infty} (1/T) \left(\sum_{t=1}^{T-1} \left(h(x_t | q'_{[0,t-1]}) - h(x_t | q'_{[0,t]}) \right) + I(x_0; q'_0) \right) \\[2mm]
&= \; \limsup_{T \to \infty} (1/T) \left(\left(\sum_{t=1}^{T-1} \log_2(|A|) \right) + h(x_0 | q'_0) - h(x_{T-1} | q'_{[0,T-1]}) + I(x_0; q'_0) \right) \\[2mm]
&= \; \log_2(|A|) - \liminf_{T \to \infty} \left((1/T) h(x_{T-1} | q'_{[0,T-1]}) \right)
\end{aligned}
$$

Additional steps, also involving an argument in Matveev'08 for the converse.

Sufficiency

Let n be a given block length. We will consider a class of uniform quantizers, defined by two parameters, with bin size $\Delta > 0$, and an even number $K(n) \geq 2$:

$$Q_{K(n)}^{\Delta}(x) = \begin{cases} (k - \frac{1}{2}(K(n) + 1))\Delta, & \text{if } x \in [(k - 1 - \frac{1}{2}K(n))\Delta, (k - \frac{1}{2}K(n))\Delta) \\ \mathcal{Z}, & \text{if } x \notin [-\frac{1}{2}K(n)\Delta, \frac{1}{2}K(n)\Delta), \end{cases}$$

where $\mathcal{Z}$ denotes the overflow symbol in the quantizer.
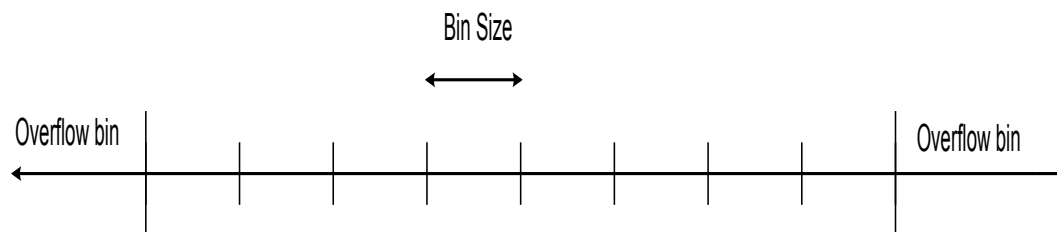
Bin Size

Overflow bin   Overflow bin

Figure 3: Quantizer.

Sufficiency

Adaptive Zooming Quantizer: An idea known since early 70's (Goodman-Gersho (TCOM'74)). Also used in control literature recently extensively.

Zoom-out when the state has escaped the quantizer.

Zoom-in when the state is inside the quantizer's granular region.

Our contribution here is on its stability analysis. To analyze stability, we consider the following.

Stochastic Stability: Markov Chains

Let $\{x_t, t \geq 0\}$ be a Markov chain with state space $(\mathbb{X}, \mathcal{B}(\mathbb{X}))$.

Definition .4. *For a Markov chain, a probability measure π is invariant on the Borel space $(\mathbb{X}, \mathcal{B}(\mathbb{X}))$ if*

$$\pi(D) = \int_{\mathbb{X}} P(x, D)\pi(dx), \quad D \in \mathcal{B}(\mathbb{X}).$$

Existence of a unique invariant probability measure (and thus positive Harris recurrence) lets the ergodic theorem hold: $\frac{1}{N}\sum_{t=0}^{N-1} f(x_t) \to \int \pi(dx)f(x)$, for all integrable f under π a.s..

Stochastic Stability: Markov Chains under Random-Time State-Dependent Drift

The following characterizes stabilization when control is applied at random times. Let τ_z, $z \geq 0$ with $\tau_0 = 0$ be such a sequence (of what is known as stopping times).

Theorem .2. *[Y. -Meyn (TAC'12)] Suppose that $\{x_t\}$ is a φ-irreducible Markov chain and $V : \mathbb{X} \to (0, \infty)$, $\delta : \mathbb{X} \to [1, \infty)$, $f : \mathbb{X} \to [1, \infty)$, a small set C, and a constant $b \in \mathbb{R}$, such that the following hold:*

$$E[V(x_{\tau_{z+1}}) \mid \mathcal{F}_{\tau_z}] \leq V(x_{\tau_z}) - \delta(x_{\tau_z}) + b1_{\{x_{\tau_z} \in C\}}$$

$$E\left[\sum_{k=\tau_z}^{\tau_{z+1}-1} f(x_k) \mid \mathcal{F}_{\tau_z} \right] \leq \delta(x_{\tau_z}), \qquad\qquad z \geq 0. \tag{7}$$

Then $\{x_t\}$ is positive Harris recurrent, and moreover $\pi(f) < \infty$, with π being the invariant distribution.

Stochastic Stability: Markov Chains

As a corollary:

Corollary .1. *[Y. - Meyn (TAC'12)] Suppose that $\{x_t\}$ is a φ-irreducible Markov chain. Suppose that there is a function $V : \mathbb{X} \to (0, \infty)$, a small set C, and a constant $b \in \mathbb{R}$, such that the following hold:*

$$E[V(x_{\tau_{z+1}}) \mid \mathcal{F}_{\tau_z}] \leq V(x_{\tau_z}) - 1 + b1_{\{x_{\tau_z} \in C\}}$$

$$\sup_{x \in \mathbb{X}, \ z \geq 0} E[\tau_{z+1} - \tau_z \mid \mathcal{F}_{\tau_z} = x] < \infty. \tag{8}$$

Then $\{x_t\}$ is positive Harris recurrent.

Sufficiency for AMS: Construction of the Stopping Times

Thus, when the decoder output is the overflow signal, then the quantizer is zoomed-out.

Zoom-in when state is within the granular region of the quantizer. Define $h_t := \frac{x_t}{\Delta_t 2^{R'-1}}$. We will say that the quantizer is perfectly zoomed when $|h_1| \leq 1$, and under-zoomed otherwise.

Define a sequence of stopping times (with n a block-length)

$$\tau_0 = 0, \quad \tau_{z+1} \;=\; \inf\{kn > \tau_z : |h_{kn}| \leq 1\}, \quad z, k \in \mathbb{Z}_+$$

Geometric Bound on the Stopping Time Distribution

For large $\Delta_{\tau z}$, roughly for $r > 1$ (exact expression is not included here):

$$P(\tau_{z+1} - \tau_z \geq kn | x_{\tau_z}, \Delta_{\tau_z}) \leq Mr^{-kn},$$

uniformly in $x_{\tau z}$.

By the geometric bound, $E[\tau_1] < \infty$, uniformly bounded for large Δ values. Hence, by the random-time drift criterion, we can show that, there exist $\psi > 0$, $|G| < \infty$ such that

$$E[\log(\Delta^2_{\tau_{z+1}}) | \Delta_{\tau z}, h_{\tau z}] \leq \log(\Delta^2_{\tau z}) - \psi + G1_{\{|\Delta_{\tau z}| \leq F\}} \tag{9}$$

By making the bin size process countable, the Markov chain becomes irreducible and the analysis can be completed.

Quadratic Stability - More Restrictive Conditions

Given a message set $\mathcal{M}(n) = \{1, 2, \ldots, K(n) + 1\}$, and a decoding function $\gamma : \mathcal{M}'^n \to \mathcal{M}(n)$ define:

- Type I-A: Error from a granular symbol to another granular symbol.

$$P^e_{g|g}(n) := \max_{c \in \mathcal{M}(n) \backslash \mathcal{Z}} P(\gamma(q'_{[0,n-1]}) \neq c, \gamma(q'_{[0,n-1]}) \neq \mathcal{Z}|c),$$

where conditioning on c means that the symbol c is transmitted.
- Type I-B: Error from a granular symbol to $\mathcal{Z}$:

$$P^e_{g|g}(n) := \max_{c \in \mathcal{M}(n) \backslash \mathcal{Z}} P(\gamma(q'_{[0,n-1]}) = \mathcal{Z}|c)$$

- Type II: Error from $\mathcal{Z}$ to a granular symbol:

$$P_e := P^e_{g|\mathcal{Z}}(n) = P(\gamma(q'_{[0,n-1]}) \neq \mathcal{Z}|\mathcal{Z})$$

Quadratic Stability

Theorem .3. *[Y. (IT'12)] A sufficient condition for quadratic stability over a DMC is that:*

$$\lim_{n\to\infty} \left(\frac{1}{n}\log(P^e_{\mathcal{Z}|g}(n)) + 2\log(|a|+\delta)\right) < 0,$$

$$\lim_{n\to\infty} \left(\kappa\frac{1}{n}\log(P^e_{g|\mathcal{Z}}(n)) + 2\log(|a|+\delta)\right) < 0,$$

$$\lim_{n\to\infty} \left(\kappa\frac{1}{n}\log(P^e_{g|g}(n)) + 2\log(|a|+\delta) + 2\kappa\log(\alpha)\right) < 0,$$

$$R'(n) > n\log_2(|a|/\alpha)$$

with arbitrarily small, positive $\eta > 0$ and $\kappa < \dfrac{1}{\log_{\frac{|a|+\delta}{|a|}}(\frac{|a|+\delta}{\alpha})}.$

A sufficient condition: The exponent under random coding satisfies $E(R) > \frac{2\log_2(|a|+\delta)}{\kappa}$ (Sahai-Mitter(IT'06) had $\kappa = 1$ with bounded noise).

Finite Moment Stability: Open Issues

The error exponent with feedback is typically improved with feedback, unlike capacity of DMCs.

However, a precise solution to the error exponent problem of fixed length block coding with noiseless feedback is not known.

Some partial results have been reported in the information theory literature: Dobrushin'62 (sphere packing bound is tight for a class of symmetric channels for rates above a critical number even with feedback), Csiszar-Körner, Haroutunian'77, Dyachkov'75, Nakiboĝlu et al'09.

Problem P2: Design of Information Channels for Optimization

Consider the optimization problem where the controller has access to channel outputs where $Q(dy|x)$ is the channel.

We consider first the single-stage case

$$
\begin{aligned}
J(P, Q) &= \inf_{\Pi} E_P^{Q,\Pi}\big[c(X_0, U_0)\big] \\
&= \inf_{\gamma \in \mathcal{G}} \int_{\mathbb{X} \times \mathbb{Y}} c(x, \gamma(y)) Q(dy|x) P(dx)
\end{aligned}
$$

in the channel Q, where $\mathcal{G}$ is the collection of all Borel measurable functions mapping $\mathbb{Y}$ into $\mathbb{U}$.

Problem P2: Design of Information Channels for Optimization

Let $\mathcal{P}(\mathbb{X})$ denote the family of probability measures on $\mathbb{X}$.

Let $\{\mu_n, \ n \in \mathbb{N}\}$ be a sequence in $\mathcal{P}(\mathbb{R}^N)$. Recall that $\{\mu_n\}$ is said to converge to $\mu \in \mathcal{P}(\mathbb{R}^N)$ *weakly* if

$$\int_{\mathbb{R}^N} c(x)\mu_n(dx) \to \int_{\mathbb{R}^N} c(x)\mu(dx)$$

for every continuous and bounded $c : \mathbb{R}^N \to \mathbb{R}$.

Definition .5. [Convergence of Channels]   (i) *A sequence of channels $\{Q_n\}$ converges to a channel Q weakly at input P if $PQ_n \to PQ$ weakly.*

(ii) *A sequence of channels $\{Q_n\}$ converges to a channel Q in* total variation *at input P if $PQ_n \to PQ$ in total variation, i.e., if $\|PQ_n - PQ\|_{TV} \to 0$.*

Continuity on the Space of Channels

Theorem .4. *[Y.-Linder (SICON'12)]*
We do not have continuity under weak convergence even for continuous cost functions.

Theorem .5. *[Y.-Linder (SICON'12)]*
If the cost function is measurable and bounded, the optimal cost $J(P, Q)$ is is continuous on the set of communication channels $\mathcal{Q}$ under under the topology of total variation.

This result will be useful to prove existence of optimal coding policies shortly.

Comparison of Information Channels

This problem is related to *comparison of experiments* as studied by D. Blackwell ('55) and Le Cam ('64). There is a partial order in the space of channels characterized by *simulation of outputs*.
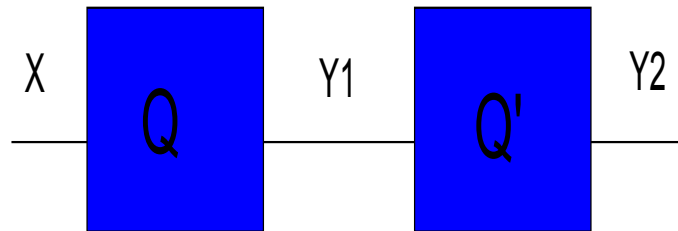

Figure 4: Blackwell'54: Q is more informative than the composite channel QQ'.

Converse holds under further technical conditions on the set of kernels.

Application: Quantizers as a class of channels

We can consider the problem of convergence and optimization of quantizers. We start with the definition of a quantizer.

Definition .6. *An M-cell vector quantizer, q, is a (Borel) measurable mapping from $\mathbb{X} = \mathbb{R}^n$ to the finite set $\{1, 2, \ldots, M\}$, characterized by a measurable partition $\{B_1, B_2, \ldots, B_M\}$ such that $B_i = \{x : q(x) = i\}$ for $i = 1, \ldots, M$. The B_i are called the cells (or bins) of q.*

A quantizer Q with cells $\{B_1, \ldots, B_M\}$, however, can also be characterized as a stochastic kernel Q on $\mathcal{B}(\mathbb{X} \times \{1, \ldots, M\})$ defined by

$$Q(i|x) = 1_{\{x \in B_i\}}, \quad i = 1, \ldots, M$$

so that $q(x) = \sum_{i=1}^{M} Q(i|x)$.

Existence of Optimal Quantizers: Convex Codecells

In the analysis, we will restrict the quantizers to have convex codecells.

By György-Linder (IT'03), there exist pairs of complementary closed half spaces $\{(H_{i,j}, H_{j,i}) : 1 \leq i, j \leq M, i \neq j\}$ such that for all $i = 1, \ldots, M$,

$$B_i \subset \bigcap_{j \neq i} H_{i,j}.$$

Since $\bar{B}_i := \bigcap_{j \neq i}$

If P admits a density, $P(\bar{B}_i \setminus B_i) = 0$ for all $i = 1, \ldots, M$.

One can obtain a (P–a.s) representation of Q by the $M(M-1)/2$ hyperplanes $h_{i,j} = H_{i,j} \cap H_{j,i}$.

Existence of Optimal Quantization Policies: Convex Codecells

Let $\mathcal{Q}_C(M)$ denote the set of quantizers with convex codecells. A sequence of quantizers converges if each of the coefficients defining hyperplanes in the quantizer converges pointwise.

Theorem .6. *[Y.-Linder (SICON'12)] The set $\mathcal{Q}_C(M)$ is compact under total variation at any input measure P that admits a density.*

Theorem .7. *[Y.-Linder (SICON'12)] Let P be absolutely continuous and suppose the goal is to find the best quantizer Q with M cells minimizing $J(P, Q) = \inf_\gamma E_P^{Q,\gamma} \int c(x, u)$, for measurable and bounded c. Then an optimal quantizer exists.*

Multi-Stage Case: Static Channels

A sequence of channels $\{Q_n\}$ converges to a channel Q *uniformly* in total variation if

$$\lim_{n\to\infty} \sup_{x\in\mathbb{X}} \big\|Q_n(\,\cdot\,|x) - Q(\,\cdot\,|x)\big\|_{TV} = 0.$$

The above is applicable to additive noise channels, where the additive noise admits a density.

Theorem .8. *Consider the multi-stage cost function with arbitrary $T \in \mathbb{N}$. If the cost function is bounded measurable, the optimization problem is continuous in the observation channel in the sense that if $\{Q_n\}$ is a sequence of channels converging to Q uniformly in total variation, then*

$$\lim_{n\to\infty} J(P, Q_n) = J(P, Q).$$

Multi-Stage Case: Dynamic Channels

Consider

$$x_{t+1} \quad = \quad f(x_t, u_t, w_t),$$

Suppose that the goal is the minimization,

$$\inf_{\Pi^{comp}} \inf_{\gamma} E_{\nu_0}^{\Pi^{comp}, \gamma} [\sum_{t=0}^{T-1} c(x_t, u_t)], \tag{10}$$

over all quantization policies Π^{comp}, and control policies γ with initial probability measure ν_0.
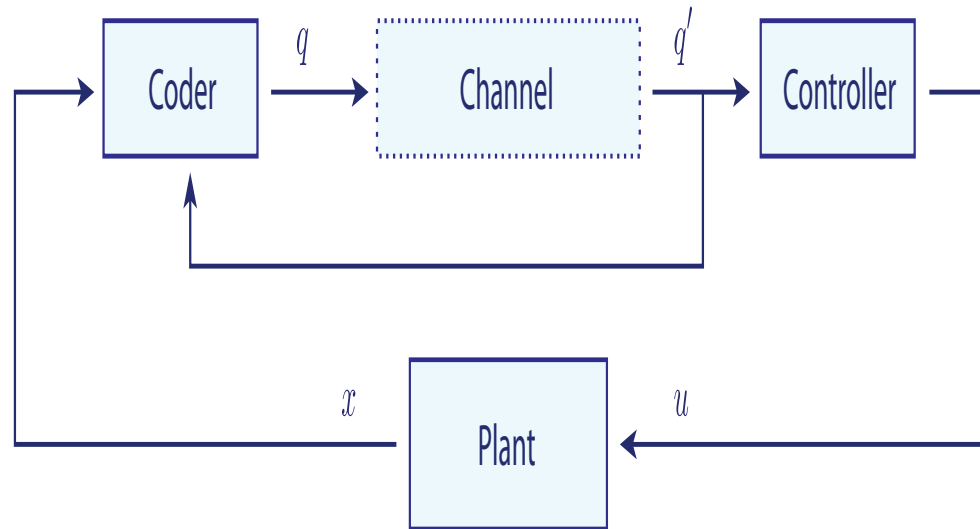
Figure 5: First consider a noiseless setting.

A Structural Result on Optimal Quantization Policies

Let $\mathcal{P}(\mathbb{X})$ denote the set of probability measures on $(\mathbb{X})$ under weak convergence and $\pi_t(\cdot) = P(x_t \in \cdot | q_{[0,t-1]}, u_{0,t-1})$.

The following is a minor extension of Walrand-Varaiya (IT'82), Teneketzis(IT'06) and Y. (IT'12).

Theorem .9. *Under the objective given in (10), any* causal composite quantization policy *can be replaced, without any loss in performance, by one which only uses the conditional probability measure* $\pi_t(\cdot) = P(x_t \in \cdot | q_{[0,t-1]})$, *the state* x_t, *and the time information* t, *at time* t.

This can be expressed as a quantization policy which only uses $\{\pi_t, t\}$ *to generate a quantizer, where the quantizer uses* x_t *to generate the quantization output* q_t.

Let Π_W denote this class of optimal policies.

MDP Formulation

Under Π_W:

$$\frac{\int \pi_{t-1}(dx_{t-1})P(q_{t-1}|\pi_{t-1}, x_{t-1})P(dx|x_{t-1}, u_{t-1})}{\int \int \pi_{t-1}(dx_{t-1})P(q_{t-1}|\pi_{t-1}, x_{t-1})P(dx|x_{t-1}, u_{t-1})}.$$

Here, $P(q_{t-1}|\pi_{t-1}, x_{t-1})$ is determined by the quantizer policy.

The sequence of conditional measures and quantizers $\{(\pi_t, Q_t)\}$ form a controlled Markov process in $\mathcal{P}(\mathbb{R}^n) \times \mathcal{Q}$, with cost to be optimized:

$$\inf_{\gamma} J_{\pi_0}(\Pi^{comp}, \gamma, T) = E_{\pi_0}^{\Pi^{comp}} \left[\frac{1}{T} \sum_{t=0}^{T-1} c(\pi_t, Q_t) \right],$$

where

$$c(\pi_t, Q_t) = \sum_{i=1}^{M} \inf_{u \in \mathbb{U}} \int_{Q_t^{-1}(i)} \pi_t(dx)c_0(x, u).$$

Existence

Assumptions:

(i) The evolution of the Markov source $\{x_t\}$ is given by

$$x_{t+1} = f(x_t) + w_t, \quad t \geq 0 \tag{11}$$

where $\{w_t\}$ is i.i.d. Gaussian.

(ii) Action space $\mathbb{U}$ is compact and $c_0 : \mathbb{R}^n \times \mathbb{U} \to \mathbb{R}_+$ is bounded and continuous.

(the compactness and boundedness conditions can be relaxed for quadratic cost functions).

Existence

Let Π_W^C be the set of coding policies in Π_W with quantizers having convex codecells.

Theorem .10. *[Y.-Linder (CDC'12)] For any $T \geq 1$, there exists a policy in Π_W^C such that*

$$\inf_{\Pi^{comp} \in \Pi_W^C} \inf_{\gamma} J_{\pi_0}(\Pi^{comp}, \gamma, T) \tag{12}$$

is achieved. Letting $J_T^T(\cdot) = 0$ and

$$J_0^T(\pi_0) := \min_{\Pi^{comp} \in \Pi_W^C, \gamma} J_{\pi_0}(\Pi^{comp}, \gamma, T),$$

the dynamic programming recursion holds.

$$T J_t^T(\pi_t) = \min_{Q \in \mathcal{Q}_c} \left(c(\pi_t, Q_t) + T E[J_{t+1}^T(\pi_{t+1}) | \pi_t, Q_t] \right)$$

Proof Sketch: Existence

Lemma .1. *For all $t \geq 0$, $\pi_t(dx)$ admits a probability density function and the sequence of density functions is uniformly continuous and is equi-continuous (due to the convolution with the Gaussian noise).*

Lemma .2. *(a) Let $\{\mu_n\}$ be a sequence of density functions on $\mathbb{R}^n$ which are uniformly equicontinuous and uniformly bounded and assume $\mu_n \to \mu$ weakly. Then $\mu_n \to \mu$ in total variation.*
(b) Let $\{Q_n\}$ be a sequence in $\mathcal{Q}_c$ such that $Q_n \to Q$ weakly at P for some $Q \in \mathcal{Q}_c$. If P admits a density, then $Q_n \to Q$ in total variation at P.

Hence, the existence theorem can be recalled since weak convergence implies total variation convergence under the setup.

Extensions

For controlled Markov sources, in the context of LQG systems, existence of optimal policies can be established. The predictive encoding structure suggested in Tatikonda-Sahai-Mitter (TAC'04) and Nair-Fagnani-Zampieri-Evans (Proc. IEEE'07) can be shown to be optimal together with an existence result.

The above results are also applicable to settings where the system is controlled over a noisy channel with noiseless feedback, almost identically.

If there is no feedback or there is noisy feedback, the analysis is very involved, and if there are no further restrictions; generally intractable as the horizon increases. Teneketzis (IT'06) and Mahajan-Teneketzis (JSAC'08), Mahajan-Teneketzis (SICON'09) investigated such settings.

Concluding Remarks and Two Directions

For stabilization, there is a total order on the set of channels: For asymptotic mean stationarity and ergodicity, Shannon capacity is the border for the converse and achievability for a large class of channels with memory and feedback.

For finite moments, the criteria we obtained are more stringent. Our bounds may not be tight for finite moments (we do not have a converse theorem); problems on fixed-length error exponents and unequal error exponents.

For optimization problems, there is a partial order: Blackwell ordering. Structural results and existence results can be obtained under conditions on the source process.

Possible direction: Empirical/Training and Approximation based design methods for optimal quantization to exploit the structural and existence results.

Possible direction: Topological issues on mismatch in the beliefs/priors in optimal teams. In the talk, we considered topologies and continuity on channels.