The Capacity of Unstable Dynamical Systems-Interaction of Control and Information Transmission

Charalambos D. Charalambous, Christos K. Kourtellaris, Sergey Loyka and Ioannis Tzortzis

***Abstract*— Feedback capacity is extended beyond classical communication channels, to stochastic dynamical systems, which may correspond to unstable control systems or unstable communication channels, subject to average cost constraints of total power $\kappa \in [0, \infty)$. It is shown that optimal conditional distributions or randomized strategies, have a dual role, to simultaneously control the output process and to encode information. The dual role is due to the interaction of control and information transmission; it states that encoders in communication channels operate as encoders-controllers, while controllers in control systems operate as controllers-encoders.**

The concepts are illustrated through the analysis of Gaussian control systems with randomized strategies, which are equivalent to Additive Gaussian Noise channels, *Stable or Unstable*, with arbitrary memory on past outputs, with an average constraint of quadratic form. It is shown that such unstable dynamical systems have *Control-Coding Capacity* which is operational, precisely as in Shannon's operational definition. However, the control-coding capacity is zero, unless the power κ allocated to the system, exceeds a threshold κ_{min}, where κ_{min} is the minimum cost of ensuring asymptotic stability and ergodicity. The excess power $\kappa - \kappa_{min}$ is turned into an achievable rate of information transmission over the dynamical system.

I. Introduction

This paper shows that Shannon's operational definition of capacity, extends to unstable dynamical systems with feedback, such as control systems, or communication channels. Further, it shows that such an extension is only possible due to an interaction between control and information transmission. The implications are striking.

In communication channels with feedback the encoder operates as an encoder-controller; the encoder part is responsible to encode information, while the controller part is responsible to control channel outputs, to achieve the control-coding communication rate. However, such a rate cannot be achieved, unless the overall power allocated to the communication system, denoted by $\kappa \in [0, \infty)$, is above a certain threshold, i.e., $\kappa \in (\kappa_{min}, \infty)$, where κ_{min} is the minimum cost required by the controller to control the communication system into asymptotic ergodicity. The excess cost $\kappa - \kappa_{min} > 0$ is the cost of communicating information at strictly positive rates.

In feedback control systems the controller operates as an controller-encoder; the controller part is responsible to controls the outputs of the control system, while the encoder part is responsible to encode information, to achieve the control-coding capacity of the control system, which acts as a communication. However, control systems do not have communication capabilities, unless randomized control strategies are used (i.e., conditional distributions), and the overall power is above the minimum cost of control κ_{min}.

The first objective of the paper is to relate the information definition of control-coding capacity of unstable dynamical systems to its corresponding operational definition, for general dynamical systems, which need not be stable, without any distinction whether these are control systems or communication channels (see Figure 1). For this reason, the terminology *control-coding capacity* of dynamical systems is used, instead of coding-capacity, although as far as the definition of operational meaning of achievable rate is concerned (see Definition 2.2), there is no fundamental difference.

The second objective of the paper is to illustrate the above concepts to dynamical systems, which are either

i) Additive Gaussian Noise (AGN) channels with memory, in which directed information is maximized over channel input distributions, subject to average cost constraints of quadratic form, and total power $\kappa \in [0, \infty)$, or

ii) Gaussian Feedback Control systems, in which randomized control strategies aim at minimizing the average of a quadratic cost, subject to an information theoretic rate constraint of total rate C.

Basically, a duality is established between i) and ii), from which all consequences of interaction of control and information transmission are obtained.

LQG-M. The channel or control system, with memory of order M, called Linear-Quadratic-Gaussian order M (LQG-M), is defined by

$$Y_i = \sum_{j=1}^{M} C_{i,i-j} Y_{i-j} + D_i A_i + V_i, \ Y_{-M}^{-1} = y_{-M}^{-1}, \tag{1}$$

$$\mathbf{P}_{V_i|V^{i-1},A^i,Y^{-1}} = \mathbf{P}_{V_i}(dv_i), \ V_i \sim N(0, K_{V_i}), \tag{2}$$

$$\mathcal{P}_{[0,n]}^{M}(\kappa) \triangleq \Big\{ P_i(da_i|a^{i-1}, y^{i-1}), i = 0, \ldots, n :$$
$$\frac{1}{n+1} \sum_{i=0}^{n} \mathbf{E}_\mu \Big\{ \langle A_i, R(i) A_i \rangle + \langle Y_{i-M}^{i-1}, Q_M(i) Y_{i-M}^{i-1} \rangle \Big\} \leq \kappa \Big\},$$
$$R(i) \in S_{++}^{q\times q}, \ Q_M(i) \in S_{+}^{Mp\times Mp}, \ i=0, \ldots, n.$$

Here $Y^n = Y_{-\infty}^n \equiv (Y^{-1}, Y_0^n) \triangleq \{Y_i : i = -\infty, \ldots, n\}$ is the controlled or channel output process, $A^n \triangleq \{A_i : i = 0, \ldots, n\}$ is the control or channel input process, $Y^{-1} \sim \mathbf{P}_{Y^{-1}} \equiv \mu(dy^{-1})$ is known to the encoder and decoder,

C. D. Charalambous, C. K. Kourtellaris and I. Tzortzis are with the Department of Electrical Engineering, University of Cyprus, Nicosia, Cyprus. E-mails: {chadcha,kourtellaris.christos,tzortzis.ioannis}@ucy.ac.cy.

S. Loyka is with the School of Electrical Engineering and Computer Science, University of Ottawa, Ottawa, ON, Canada. E-mail: sergey.loyka@ieee.org.

978-1-5090-4096-4/17/$31.00 ©2017 IEEE

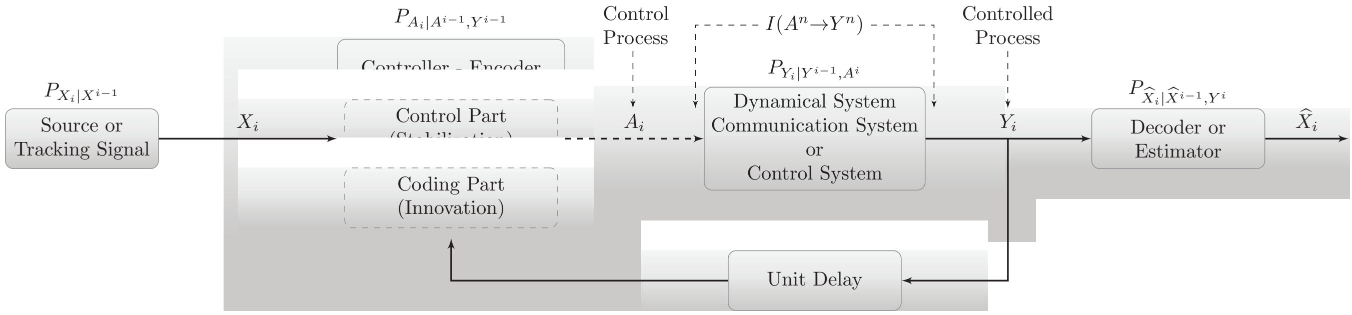

Fig. 1. *Shannon's communication block diagram and its analogy to Dynamical control systems.*

$Y_{-M}^{-1} \sim N(0, K_{Y_{-M}^{-1}})$ is the Gaussian initial condition of (1), $\big\{V_i \sim N(0, K_{V_i}) : i = 0, 1, \ldots, n\big\}$ is IID Gaussian, $\langle \cdot, \cdot \rangle$ denotes inner product of elements of linear spaces, $\mathbb{S}_{+}^{q\times q}$ denotes the set of symmetric positive semi-definite $q \times q$ matrices and $\mathbb{S}_{++}^{q\times q}$ the subset of positive definite matrices, $\kappa \in [0, \infty)$ is the total power, and $\mathbf{E}_\mu$ means the joint distribution depends on $\mu(\cdot)$.

First, we illustrate the claims, for the scalar example from [1] (see also [2] for more discussion), which corresponds to LQG-1, i.e., given by

$$Y_i = CY_{i-1} + DA_i + V_i, \quad Y_{-1} = y, \quad i = 0, \ldots, n, \tag{3}$$

$$\frac{1}{n+1}\mathbf{E}_\mu\Big\{\sum_{i=0}^{n-1}(R|A_i|^2 + Q|Y_{i-1}|^2) + R|A_n|^2 + M|Y_{n-1}|^2\Big\} \leq \kappa, \quad Y_{-1} \sim N(0, K_{Y_{-1}}) \tag{4}$$

where C can be stable or unstable, i.e., $|C| < 1$ or $|C| \geq 1$. By [1], the Shannon capacity is achieved by a Gaussian distribution of the form $\mathbf{P}_{A_i|Y_{i-1}}$, capacity is independent of y^{-1}, and the joint process $(A_i, B_i, V_i) : i = 0, \ldots, n$ is also Gaussian, with decomposition

$$A_i \stackrel{\triangle}{=} U_i + Z_i, \qquad i = 0, 1, \ldots, \tag{5}$$
$$U_i = \Gamma Y_{i-1}, \; Z_i \sim N(0, K_Z) \text{ IID}, \; Z_i \text{ independent of } V^i.$$

Here U_i is the control part, responsible to stabilize the LQG-1 model, i.e., (1), to ensure closed loop poles within the unit circle $|C + D\Gamma| < 1$, for an appropriate feedback gain Γ, and Z_i is the random part, responsible to encode information. The capacity is characterized by

$$C(\kappa) \stackrel{\triangle}{=} \lim_{n\longrightarrow\infty}\frac{1}{n+1}\max_{\substack{A_i=U_i+Z_i\\(4)\;holds}}\sum_{i=0}^{n} I(A_i; Y_i|Y_{i-1}) \tag{6}$$
$$= \sup_{K_Z \geq 0}\Big\{\frac{1}{2}\log\frac{|D^2K_Z + K_V|}{|K_V|} : RK_Z + P\big(D^2K_Z + K_V\big) = \kappa\Big\},$$
$$P = C^2P + Q - \big(CPD\big)^2\big(D^2P + R\big)^{-1}, \tag{7}$$
$$|C + D\Gamma^*| < 1, \quad \Gamma^* \stackrel{\triangle}{=} -\Big(D^2P + R\Big)^{-1}DPC \tag{8}$$

where (a) the optimal control part is $U_i = U_i^* = \Gamma^* Y_{i-1}^*$, $i = 0, \ldots$, and (b) the optimal random part $Z_i = Z_i^* \sim N(0, K_Z^*) : i = 0, \ldots$ is the solution of the generalized water-filling problem (6). Note that (7) is the well-known Riccati equation (with solutions, $P \geq 0$, $P < 0$) of LQG stochastic optimal control problem [3], and only stabilizing solutions are chosen, i.e., $P \geq 0$, such that $|C + D\Gamma^*| < 1$. Consider the following special cases.

No Cost on Output. Suppose $D = R = 1, Q = 0$.
(i) Then the cost constraint is independent of past channel output process, and the Riccati equation solutions are

$$P = \begin{cases} 0 & \text{if} \quad |C| < 1 \\ C^2 - 1 & \text{if} \quad |C| \geq 1. \end{cases} \tag{9}$$

The optimal strategy which achieves the capacity is given by

$$\Big(\Gamma^*, K_Z^*\Big) = \begin{cases} (0, \kappa), & \kappa \in [0, \infty) \text{ if } |C| < 1 \\ \Big(-\dfrac{C^2-1}{C}, \dfrac{\kappa + K_V(1-C^2)}{C^2}\Big), & \kappa \in [\kappa_{min}, \infty) \text{ if } |C| \geq 1 \\ \Big(-\dfrac{C^2-1}{C}, 0\Big), & \kappa \in [0, \kappa_{min}] \text{ if } |C| \geq 1 \end{cases}$$

where $\kappa_{min} \stackrel{\triangle}{=} (C^2 - 1)K_V$ is the part of the average power corresponding to $\{U_i^* : i = 0, \ldots,\}$, only. The above solution illustrates that if $Q = 0$ and the channel is stable, i.e., $|C| < 1$, then feedback does not increase capacity, and the capacity is that of a memoryless channel corresponding to $Y_i = DA_i + V_i, i = 0, \ldots, n$, $\frac{1}{n+1}\mathbf{E}\{\sum_{i=0}^{n}|A_i|^2\} \leq \kappa$. However, if the channel is unstable $|C| > 1$ then the minimum cost to control the channel into a stable regime is κ_{min}, and any $\kappa \in [0, \kappa_{min}]$ results in zero rate. For $C = 1$ there is a fundamental problem; the Riccati quadratic equation has solutions $P_1 = P_2 = 0$. This means $\kappa_{min} = 0$ but the LQG-1 model is unstable, and capacity is zero. This is due to the fact that $Q = 0$, implies the behavior of the output is not reflected in the constraint, and $U_i, i = 0, \ldots$, does not know it.
(ii) Let $C^{Stable}(\kappa)$ denote the capacity if the channel is stable, i.e., $|C| < 1$ and $C^{Unstable}(\kappa)$ denote the feedback capacity if the channel is unstable, i.e., $|C| > 1$. Then, there is a rate loss due to the instability given by

$$C^{Stable}(\kappa) - C^{Unstable}(\kappa) = \begin{cases} \frac{1}{2}\log\Big(1 + \frac{\kappa}{K_V}\Big), \kappa \leq \kappa_{min} \\ \log|C|, \quad \kappa \geq \kappa_{min}. \end{cases}$$

Cost on Output. Suppose $D = R = 1, Q > 0$. Then it can be verified by solving Riccati equation (7) that always

$\kappa_{min} > 0$, that is, there is cost of control which penalizes any deviations, in Mean-Square sense $|Y_i - r_{ref}|^2, i = 0, \ldots,$ from reference signal $r_{ref} = 0$ (this can be generalized to $r_{ref} \neq 0$ by generalizing (4)). Inclusion in the cost constraint, of such terms, is necessary in control application, because the output process is required to track specific signals. In communication applications, the need to control channel outputs becomes necessary, if the receiver is required to operate at certain operating point r_{ref}, because the receiver circuit can be driven into non-linear regimes away from r_{ref} or is sensitivity to excess power. Moreover, $Q \neq 0$ is sufficient to ensure unstable modes are reflected into the cost, to avoid solutions of Riccati equation $P_1 = P_2 = 0$, such as, when $C = 1$ discussed in (i).

The rest of the paper establishes the interaction of control and information transmission for general dynamical systems.

Cover and Pombra [4] characterized the feedback capacity of scalar-valued, Additive Gaussian Noise (AGN), with non-stationary, nonergodic noise. Several variants of the Cover and Pombra [4] AGN channel, such as, stationary ergodic noise, are investigated in [5]–[7]. However, the computation of capacity and capacity achieving distributiom for Cover's and Pombra's initial formulation [4], for finite n is not done.

Section II establishes analogies between stochastic optimal control systems and capacity of communication channels, for arbitrary models. Section III deals with LQG-M models.

II. The Control-Coding Capacity of General Stochastic Dynamical Systems

The analogy of stochastic control systems and communication channels, as shown in Figure 1 is, 1) $A^n \triangleq \{A_0, \ldots, A_n\}$ is the control or channel input process, and $A^n = a^n \in \times_{i=0}^n \mathbb{A}_i^n$ their actions. 2) $Y^n \equiv (Y^{-1}, Y_0^n) \triangleq \{\ldots, Y_{-1}, Y_0, \ldots, Y_n\}$ is the controlled or channel output process, taking values in $\times_{i=-\infty}^n \mathbb{Y}_i$. 3) $\{\mathbf{P}_{Y_i|Y^{i-1},A^i} : i = 0, \ldots, n\}$ is the control system or channel distribution. 4) $\mathcal{P}_{[0,n]} \triangleq \{\mathbf{P}_{A_i|A^{i-1},Y^{i-1}} \equiv P_i(da_i|a^{i-1}, y^{i-1}) : i = 0, \ldots, n\}$ is the set of randomized control strategies or channel input distributions (Y^{-1} is known to the encoder/decoder), and the admissible strategies are

$$\mathcal{P}_{[0,n]}(\kappa) \triangleq \Big\{P_i(da_i|a^{i-1}, y^{i-1}), i = 0, \ldots, n : \frac{1}{n+1}\mathbf{E}_\mu^P\Big(\ell_{0,n}(A^n, Y^n)\Big) \leq \kappa\Big\} \quad (10)$$

where $\ell_{0,n} : \mathbb{A}^n \times \mathbb{Y}^n \longmapsto (-\infty, \infty]$ is a measurable cost function, $\kappa \in [0, \infty)$ is the total cost or power, and notation $\mathbf{E}_\mu^P$ indicates the dependence of the joint distribution on $\mathcal{P}_{[0,n]}$ and initial distribution $\mathbf{P}_{Y^{-1}} \triangleq \mu(dy^{-1})$.
5) $\{X_i : i = 0, \ldots, n\}$ is the tracking signal or information process, taking values in $\times_{i=0}^n \mathbb{X}_i$.

Information Control-Coding Capacity. Define directed information from $\{A_0, A_1, \ldots, A_n\}$ to $\{Y_0, Y_1, \ldots, Y_n\}$ conditioned on Y^{-1} by [8]–[11]

$$I(A^n \rightarrow Y^n) \triangleq \mathbf{E}_\mu^P\Big\{\sum_{i=0}^n \log\Big(\frac{d\mathbf{P}_{Y_i|Y^{i-1},A^i}(\cdot|Y^{i-1}, A^i)}{d\mathbf{P}_{Y_i|Y^{i-1}}(\cdot|Y^{i-1}, A^{-1})}(Y_i)\Big)\Big\}. \quad (11)$$

A candidate for *Control-Coding Capacity* is the *Information Control-Coding Capacity* defined by

$$J_{A^\infty \rightarrow Y^\infty}(\kappa) \triangleq \liminf_{n\longrightarrow\infty} \frac{1}{n+1} J_{A^n\rightarrow Y^n}(P^*, \kappa),$$
$$J_{A^n\rightarrow Y^n}(P^*, \kappa) \triangleq \sup_{\mathcal{P}_{[0,n]}(\kappa)} I(A^n \rightarrow Y^n).$$

Under appropriate conditions, the direct and converse coding theorems [5], [11], [12], imply that $C(\kappa) \triangleq J_{A^\infty\rightarrow Y^\infty}(\kappa)$, is the *Control-Coding Capacity of the Dynamical System*. The operational meaning states that any control-coding rate R in bits/second below $C(\kappa)$, can be achieved[1].

Cost of Control and Communication. Next, we show, in general, $J_{A^n\rightarrow Y^n}(P^*, \kappa) = 0$, unless the power κ is above a critical value κ_{min}, which is precisely the minimum cost required to control $\{Y_i : i = 0, \ldots, n\}$. Let $\mathcal{P}_{[0,n]}^D$ denote the restriction of the randomized strategies $\mathcal{P}_{[0,n]}$ to the set of deterministic strategies defined by $\mathcal{P}_{[0,n]}^D \triangleq \Big\{a_0 = g_0(y^{-1}), \ldots, a_n = g_n(a^{n-1}, y^{n-1})\Big\}$. By [13], for any finite n, $C_{0,n}(\kappa) \triangleq J_{A^n\rightarrow Y^n}(P^*, \kappa)$ is a concave non-decreasing in $\kappa \in [0, \infty)$, and the inverse function of $C_{0,n}(\kappa)$ denoted by $\kappa_{0,n}(C)$ is a convex non-decreasing function of $C \in [0, \infty)$. This implies the following duality relation.

Dual Extremum Problem.

$$\begin{aligned}
&\kappa_{0,n}(C) \\
&\triangleq \inf_{\{P_i(da_i|a^{i-1},y^{i-1})\}_{i=0}^n : \frac{1}{n+1}I(A^n\rightarrow Y^n)\geq C} \mathbf{E}_\mu^P\{\ell_{0,n}(A^n, Y^n)\} \\
&\geq J_{0,n}^{SC}(P^*) \triangleq \inf_{\mathcal{P}_{[0,n]}} \mathbf{E}_\mu^P\Big\{\ell_{0,n}(A^n, Y^n)\Big\} \equiv \kappa_{0,n}(0) \\
&= \inf_{\mathcal{P}_{[0,n]}^D} \mathbf{E}_\mu^g\Big\{\ell_{0,n}(A^n, Y^n)\Big\} \equiv J_{0,n}^{SC}(g^*) \qquad (12)
\end{aligned}$$

where (12) follows from the well-known property of classical stochastic optimal control theory, that minimizing the pay-off $\mathbf{E}_\mu^P\{\ell_{0,n}(A^n, Y^n)\}$ over randomized control strategies $\mathcal{P}_{[0,n]}$ does not incur a better performance than minimizing it over deterministic strategies $\mathcal{P}_{[0,n]}^D$. Thus, at time n the minimum cost of control is $J_{0,n}^{SC}(P^*)$, and for information rate C, the cost of communication is given by

$$\kappa(C) - \kappa(0) \triangleq \lim_{n\longrightarrow\infty} \frac{1}{n+1}\kappa_{0,n}(C) - \lim_{n\longrightarrow\infty} \frac{1}{n+1}\kappa_{0,n}(0)$$

provided the limits exists and they are finite. We conclude that for non-zero transmission rate, the critical value is $\kappa_{min} = J_{0,n}^{SC}(P^*) \equiv \kappa_{0,n}(0)$, and this is precisely the minimum cost of control, using either randomized or deterministic strategies.

[1] In general, $R \equiv R_{y^{-1}}$ may depend on the initial data y^{-1}, unless ergodicity is shown.

Remark 2.1: There is a subtle issue regarding the control-coding capacity; for unstable control systems or channels, the rate $C(\kappa)$ may be zero, unless $\kappa \in (\kappa_{min}, \infty)$, is strictly positive, i.e., $\kappa_{min} > 0$ (see earlier example or [2]). This means, stationary ergodicity of the joint process $\{(A_i, Y_i) : i = \ldots, -1, 0, 1, \ldots\}$ is not sufficient to ensure a non-trivial value $C(\kappa) > 0$.

Definition 2.2: (Operational control-coding capacity of control systems) Consider a control system distribution $\big\{\mathbf{P}_{Y_i|Y^{i-1},A^i} : i = 0, \ldots, n\big\}$. A controller-encoder-decoder with power constraint consists of the following.
(a) A set of uniformly distributed messages $X^{(n)}$ with alphabet space $\mathcal{M}^{(n)} \triangleq \{1, \ldots, M^{(n)}\}$, known to both the encoder and decoder (the controller does not need to know these). The initial data Y^{-1} are known to the encoder-controller-decoder.
(b) A set of controller-encoder strategies $\mathcal{E}^S_{[0,n]} \triangleq \big\{a_0 = e_0(x^{(n)}, y^{-1}), \ldots, a_n = e_n(x^{(n)}, a^{n-1}, y^{n-1}), x^{(n)} \in \mathcal{M}^{(n)}\big\}$. The set of admissible controller-encoder strategies subject to power constraint κ is defined by

$$\mathcal{E}^S_{[0,n]}(\kappa) \triangleq \Big\{e_i(x^{(n)}, a^{i-1}, y^{i-1}), i = 0, \ldots, n :$$
$$\frac{1}{n+1}\mathbf{E}^e_\mu\Big(\sum_{i=0}^n \ell_{0,n}(A^n, Y^n)\Big) \leq \kappa\Big\} \subset \mathcal{E}^S_{[0,n]}, \quad \kappa \in [0, \infty).$$

(c) A decoder measurable mapping $d_n : \mathbb{Y}^n \longmapsto \mathcal{M}^{(n)}$, $\widehat{X}^{(n)} \triangleq d_n(Y^n)$ with average probability of decoding error

$$\frac{1}{M^{(n)}}\sum_{x^{(n)}} \mathbf{P}^e\{d_n(Y^n) \neq x^{(n)} | X^{(n)} = x^{(n)}, Y^{-1} = y^{-1}\} \leq \epsilon_n.$$

(d) The control-coding rate is $R^{(n)} \triangleq \frac{1}{n+1}\log M^{(n)}$. A control-coding rate $R > 0$ is said to be an achievable rate if $\lim_{n\longrightarrow\infty} \epsilon_n = 0$, and $\liminf_{n\longrightarrow\infty} \frac{1}{n+1}\log M^{(n)} \geq R$. The operational control-coding capacity of the control system is defined by $C(\kappa) \triangleq \sup\{R : R \text{ is achievable}\}$.

In general, $C(\kappa)$ depends on $Y^{-1} = y^{-1}$. Keeping in mind Remark 2.1, and, subject to technical assumptions, then standard coding theorems with some variations, i.e., [5], [10]–[12], can be applied to show $C(\kappa) \triangleq J_{A^\infty \to Y^\infty}(P^*, \kappa)$ is the control-coding capacity. For the LQG-M model achievability is shown in [2], using ergodic theory.

III. The Control-Coding Capacity of LQG-M

Consider the LQG-M model, and assume Y^{-1}_{-M} is Gaussian. From [2], we know that any candidate of the optimal randomized strategy which satisfies the cost constraint and maximizes $I(A^n \to Y^n)$, satisfies $P_i(da_i|a^{i-1}, y^{i-1}) = \pi^{M,g}(da_i|y^{i-1}_{i-M}), i = 0, \ldots, n$, it is Gaussian, with corresponding joint process which is also Gaussian, denoted by $\{(A_i, Y_i) \equiv (A^g_i, Y^g_i) : i = 0, \ldots, n\}$. Moreover,

$$A^g_i = g^M_i(Y^{g,i-1}_{i-M}) + Z_i \equiv \Gamma_M(i)Y^{g,i-1}_{i-M} + Z_i, \tag{13}$$

i) Z_i is independent of $(A^{g,i-1}, Y^{g,i-1})$, $i = 0, \ldots, n$,
ii) Z^i is independent of V^i, $i = 0, \ldots, n$,
iii) $\{Z_i \sim N(0, K_{Z_i}) : i = 0, 1, \ldots, n\}$ is an orthogonal innovations or independent Gaussian process.

The information control-coding capacity is [2]

$$J_{A^n \to Y^n}(\pi^{M,g,*}, \kappa) \triangleq \sup_{\mathcal{E}^M_{[0,n]}(\kappa)} H(Y^{g,n}) - H(V^n) \equiv C_{0,n}(\kappa), \tag{14}$$

$$H(Y^{g,n}) - H(V^n) = \frac{1}{2}\sum_{i=0}^n \log\frac{|D_i K_{Z_i} D_i^T + K_{V_i}|}{|K_{V_i}|}, \tag{15}$$

$$\mathcal{E}^M_{[0,n]}(\kappa) \triangleq \Big\{\big(\Gamma_M(i), K_{Z_i}\big), i = 0, \ldots, n : \mathbf{E}^{g^M}\Big[\sum_{i=0}^n \big(\langle A^g_i, R(i)A^g_i\rangle + \langle Y^{g,i-1}_{i-M}, Q_M(i)Y^{g,i-1}_{i-M}\rangle\big)\Big] \leq \kappa(n+1)\Big\},$$
$$Y^g_i = \Big(C_M(i) + D_i\Gamma_M(i)\Big)Y^{g,i-1}_{i-M} + D_iZ_i + V_i, \ A^g_i = (13).$$

Next, to aid the computation of the optimal strategy, we introduce additional variables as follows [14], $S^g_{i-1}(1) \triangleq Y^g_{i-2}$, $i = 0, \ldots, n$,

$$S^g_{i-1}(2) = Y^g_{i-3}, \ldots, S^g_{i-1}(M-1) \triangleq Y^g_{i-M},$$
$$S^g_{i-1} \triangleq [Y^g_{i-1}, S^g_{i-1}(1), S^g_{i-1}(2), \ldots, S^g_{i-1}(M-1)]^T.$$

Theorem 3.1: Consider the LQG-M. Define

$$A^g_i \triangleq U^g_i + Z_i, \quad U^g_i = g^M_i(S^g_{i-1}) \equiv \Gamma_M(i)S^g_{i-1}, \tag{16}$$

where $\{U^g_i : i = 0, \ldots, n\}$ is the deterministic part of the randomized strategy (control part) and $\{Z_i : i = 0, \ldots, n\}$ is the random part. Then we have the following.
(a) For $i = 0, \ldots, n$,

$$Y^g_i = C_M(i)S^g_{i-1} + D_iU^g_i + D_iZ_i + V_i, \tag{17}$$
$$S^g_i = C_M(i)S^g_{i-1} + D_M(i)U^g_i + D_M(i)Z_i + I_M(i)V_i, \tag{18}$$

where $\{C_M(i), Q_M(i), D_M(i), I_M(i) : i = 0, \ldots, n\}$ are appropriate matrices, and

$$J_{A^n \to Y^n}(\pi^{M,g,*}, \kappa) = \sup_{\big\{(g^M_i(\cdot), K_{Z_i}), i=0,\ldots,n\big\} \in \mathcal{E}^M_{[0,n]}(\kappa)} \sum_{i=0}^n H(Y^g_i|S^g_{i-1}) - H(V^n),$$
$$\sum_{i=0}^n H(Y^g_i|S^g_{i-1}) - H(V^n) = (15),$$
$$\mathcal{E}^M_{[0,n]}(\kappa) \triangleq \Big\{g^M_i : \mathbb{S}^{i-1}_{i-M} \longmapsto \mathbb{R}^q, \ u_i = g^M_i(s_{i-1}), \ K_{Z_i} \in S^{q\times q}_+,$$
$$\frac{1}{n+1}\mathbf{E}^{g^M}_\mu\Big\{\sum_{i=0}^n [\langle A^g_i, R(i)A^g_i\rangle + \langle S^g_{i-1}, Q_M(i)S^g_{i-1}\rangle]\Big\} \leq \kappa\Big\}.$$

(b) The optimal deterministic part of the randomized strategy, $\{g^{M,*}_i(\cdot) : i = 0, \ldots, n\}$ is given by the following equations.

$$g^{M,*}_i(s_{i-1}) = \Gamma^*_M(i)s_{i-1} = -\Big(D^T_M(i)P_M(i+1)D_M(i) + R(i)\Big)^{-1} D^T_M(i)P_M(i+1)C_M(i)\, s_{i-1}, \quad i = 0, \ldots, n, \tag{19}$$

$$g^{M,*}_n(s_{n-1}) = 0, \tag{20}$$

$$P_M(i) = C^T_M(i)P_M(i+1)C_M(i) + Q_M(i) - C^T_M(i)P_M(i+1)D_M(i)\Big(D^T_M(i)P_M(i+1)D_M(i) + R(i)\Big)^{-1} \Big(C^T_M(i)P_M(i+1)D_M(i)\Big)^T, \ P(n) = Q_M(n). \tag{21}$$

Moreover, $\{g_i^{M,*}(s_{i-1}) : i = 0, \dots, n\}$ given by (21) is the optimal solution of the following LQG control problem.

$$J_{0,n}(g^{M,*}) \equiv \kappa_{0,n}(0)$$
$$\stackrel{\triangle}{=} \inf_{g_i^{M,*}(\cdot): i=0,\dots,n} \mathbf{E}_\mu^{g^M}\Big\{\langle A_i, R(i)A_i\rangle + \langle Y_{i-M}^{i-1}, Q_M(i) Y_{i-M}^{i-1}\rangle\Big\}$$

where $\kappa_{0,n}(0)$ is the cost of control, corresponding to $K_{Z_i} = 0, i = 0, \dots, n$.
(c) The optimal random part of the strategy $\{K^*_{Z_i} : i = 0, \dots, n\}$ is the solution of the recursive equations

$$r_M(i) = r_M(i+1) + \sup_{K_{Z_i}\in S_+^{q\times q}} \Big\{\frac{1}{2}\log\frac{|D_i K_{Z_i} D_i^T + K_{V_i}|}{|K_{V_i}|}$$
$$-\lambda\, tr\Big(P_M(i+1)\Big[D_M(i)K_{Z_i}D_M^T(i) + I_M(i)K_{V_i}I_M^T(i)\Big]\Big)$$
$$-\lambda\, tr\Big(\ R(i)K_{Z_i}\Big)\Big\}, \quad i = 0, \dots, n-1, \tag{22}$$
$$r_M(n) = \sup_{K_{Z_n}\in S_+^{q\times q}} \Big\{\frac{1}{2}\log\frac{|D_{n,n}K_{Z_n}D_{n,n}^T + K_{V_n}|}{|K_{V_n}|}$$
$$-\lambda\, tr\Big(R(n)K_{Z_n}\Big) + \lambda(n+1)\kappa\Big\} \tag{23}$$

where $\{P_M(i) : i = 0, \dots, n\}$ satisfies (21).
(d) The information control-coding capacity is given by

$$J_{A^n\to Y^n}(\pi^{M,g,*}, \kappa)$$
$$= -\lambda\int_{\mathbb{Y}_{-M}^{-1}} \langle s_{-1}, P(0)s_{-1}\rangle \mathbf{P}_{S_{-1}}(ds_{-1}) + r_M(0)$$

where $\lambda \equiv \lambda_n(\kappa)$ is found from the cost constraint.

Proof: For the derivation see [2]. ■

Theorem 3.1 is analogous to the $n-$block capacity of Cover and Pombra [4], known to be diffucult to solve (see [7], page 58). It illustrates that the predictable part of the optimal randomized strategy given by (19), (20), controls the controlled process, precisely as in LQG stochastic optimal control theory [3], [14], while its non-predictable or random part is an innovations process with covariance $\{K^*_{Z_i} : i = 0, \dots, n\}$. In [2], using Theorem 3.1, the per unit time limit is analyzed and its operation is shown, via ergodic theory.

The example below demonstrates the water-filling properties of the random part of the optimal strategy.

Example 3.2: Consider the case $M = p = q = 1$. From (22) and (23) we obtain

$$C_{0,n}(\kappa) \stackrel{\triangle}{=} J_{A^n\to Y^n}(\pi^{1,g,*}, \kappa) = \frac{1}{2}\sum_{i=0}^{n}\log\frac{|D_i K^*_{Z_i} D_i^T + K_{V_i}|}{|K_{V_i}|}$$
$$= \frac{1}{2}\sum_{i=0}^{n-1}\Big\{\log\Big(\frac{D_i^2}{2\lambda\Big(P(i+1)D_i^2 + R_i\Big)K_{V_i}}\Big)\Big\}^+$$
$$+ \frac{1}{2}\Big\{\log\Big(\frac{D_i^2}{2\lambda R_n K_{V_n}}\Big)\Big\}^+ = \sum_{i=0}^{n} C_i(\kappa_i^*) \tag{24}$$

where $\{x\}^+ \stackrel{\triangle}{=} \max\{0, x\}$, $\lambda = \lambda_n(\kappa) \geq 0$ is the Lagrange multiplier chosen to satisfy constraint with equality given by

$$\sum_{i=0}^{n-1}\Big\{\Big\{\frac{1}{2\lambda} - \frac{\Big(P(i+1)D_i^2 + R_i\Big)K_{V_i}}{D_i^2}\Big\}^+ + P(i+1)K_{V_i}\Big\}$$
$$+ \Big\{\frac{1}{2\lambda} - \frac{R_n K_{V_n}}{D_n^2}\Big\}^+ + \mathbf{E}\Big(Y_{-1}\Big)^2 P(0) = \kappa(n+1). \tag{25}$$

Clearly, in general, for each i, $C_i(\kappa_i^*) > 0$ provided $\kappa_i^* \in (\kappa_{min,i}, \infty)$ and these critical values depend on whether $|C_{i,i-1}| \geq 1$ or $|C_{i,i-1}| < 1$, for $i = 0, \dots, n$.

IV. Conclusion

The control-coding capacity of dynamical systems is discussed, as a consequence of the universality of Shannon's capacity, which include unstable stochastic control systems or communication channels. For the LQG-M model, the interaction of control and information transmission is shown explicitly. Generalizations to models with past dependence on inputs $\{A_i : i = 0, \dots, n\}$ are found in [15].

References

[1] C. Kourtellaris and C. D. Charalambous, "Information structures of capacity achieving distributions for feedback channels with memory and transmission cost: Stochastic optimal control & variational equalities-part I," *IEEE Transactions on Information Theory*, 2015, submitted, November 2015. [Online]. Available: arxiv.org/abs/1512.04514

[2] C. D. Charalambous, C. Kourtellaris, and S. Loyka, "Capacity achieving distributions & information lossless randomized strategies for feedback channels with memory: The LQG theory of directed information: Part II," *IEEE Transactions on Information Theory*, submitted, April 2016. [Online]. Available: https://arxiv.org/abs/1604.01056

[3] P. R. Kumar and P. Varaiya, *Stochastic Systems: Estimation, Identification, and Adaptive Control*. Prentice Hall, 986.

[4] T. Cover and S. Pombra, "Gaussian feedback capacity," *IEEE Transactions on Information Theory*, vol. 35, no. 1, pp. 37–43, Jan. 1989.

[5] S. Ihara, *Information theory for Continuous Systems*. World Scientific, 1993.

[6] S. Yang, A. Kavcic, and S. Tatikonda, "On feedback capacity of power-constrained Gaussian noise channels with memory," *Information Theory, IEEE Transactions on*, vol. 53, no. 3, pp. 929–954, March 2007.

[7] Y.-H. Kim, "Feedback capacity of stationary Gaussian channels," *IEEE Transactions on Information Theory*, vol. 56, no. 1, pp. 57–85, 2010.

[8] H. Marko, "The bidirectional communication theory–A generalization of information theory," *IEEE Transactions on Communications*, vol. 21, no. 12, pp. 1345–1351, Dec. 1973.

[9] J. L. Massey, "Causality, feedback and directed information," in *International Symposium on Information Theory and its Applications (ISITA '90)*, Nov. 27-30 1990, pp. 303–305.

[10] G. Kramer, "Capacity results for the discrete memoryless network," *IEEE Transactions on Information Theory*, vol. 49, no. 1, pp. 4–21, Jan. 2003.

[11] S. Tatikonda and S. Mitter, "The capacity of channels with feedback," *IEEE Transactions on Information Theory*, vol. 55, no. 1, pp. 323–349, Jan. 2009.

[12] H. Permuter, T. Weissman, and A. Goldsmith, "Finite state channels with time-invariant deterministic feedback," *IEEE Transactions on Information Theory*, vol. 55, no. 2, pp. 644–662, Feb. 2009.

[13] C. D. Charalambous and P. A. Stavrou, "Directed information on abstract spaces: Properties and variational equalities," *IEEE Transactions on Information Theory*, vol. 62, no. 11, pp. 6019–6052, 2016.

[14] P. E. Caines, *Linear Stochastic Systems*, ser. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., New York, 1988.

[15] C. D. Charalambous, C. Kourtellaris, and I. Tzortzis, "Information structures of maximizing distributions of feedback capacity for general channel with memory and applications," *IEEE Transactions on Information Theory*, submitted, July 2016. [Online]. Available: https://arxiv.org/abs/1604.01063