2017 IEEE 56th Annual Conference on Decision and Control (CDC)
December 12-15, 2017, Melbourne, Australia

Signalling of Information in Networked Stochastic Dynamical Control Systems

Charalambos D. Charalambous, Christos Kourtellaris, and Ioannis Tzortzis

Abstract— **Signalling of information is made feasible from one controller to another controller, in a networked control system consisting of two interconnected control systems. In this application, controller 2 of control system 2 (CS-2) has access to feedback information from its output, while controller 1 of control system 1 (CS-1) does not have access to feedback information from its output.**

The methodology is based on computing the Control-Coding (CC) Capacity of CS-2, and then transforming the randomized control strategy which achieves the CC Capacity of CS-2, into a controller-encoder, which simultaneously controls the output of CS-2, and encodes the output of CS-1, then decodes it and applies it to the controller of CS-1, to minimize the pay-off of CS-1.

The paper demonstrates that randomized strategies of control systems simultaneously control and encode information, and signal information from one processor to another processor, through the control system, provided the rate at which information is signalled through the control system is below the CC Capacity of the control system.

I. Introduction

The operational definition of maximum rate of reliable information transfer or signalling through control systems stable or unstable is characterized recently in [1]. The supremum of all rates is given by the CC capacity of the control system, and is defined by the maximization of directed information from control processes to controlled processes over all randomized control strategies, which satisfy average cost constraints (see also [2], [3]). The operational definition and achievable CC rate are generalizations of Shannon's [4] operational definition and achievable coding rate of noisy communication channels [5] to stochastic dynamical control systems, with the encoder replaced by an controller-encoder (see Definitions 2.4, 2.5, Theorem 2.6, 2.8 in [6] or [3]). This means, in general, controllers in control systems can be replaced by controllers-encoders that simultaneously control and encode information, such that the encoded information can be reconstructed at the control system output. A direct consequence of CC theorems, is that, for any CC rate measured in bits/second below the CC Capacity of the control system, there exists an controller-encoder which stabilizes the control system and encodes information, and a decoder or estimator attached to control system outputs, that operate asymptotically with arbitrary small decoding error probability. Often, CC theorems follow from the ergodic theory of Markov decision [2], [3] or directed information stability [7].

The main objective of this paper is to apply the CC Capacity to a networked control system consisting of two interconnected control systems, each assigned one controller, as shown in Fig.I.1, to signal information from one controller to another controller. The controller of Control System (CS-2) has access to its output through feedback, while the controller of CS-1 does not have access to its output. To overcome the limitation of controller 1, then the CC Capacity of CS-2 is determined and the randomized control strategy which achieves it is found. Then the randomized control strategy of CS-2 is transformed into an controller-encoder, which simultaneously controls the output process $\{Y_i : i = 0, \dots, n\}$ and encodes the process $\{X_i : i = 0, \dots,\}$, and a decoder or estimator is designed $\{\widehat{X}_i : i = 0, \dots,\}$, with respect to a performance objective. Then the decoder output $\{\widehat{X}_i : i = 0, \dots,\}$ is made available to the controller of CS-1, that minimizes the pay-off of CS-1.

The main results state the following.

(a) If the CC Capacity of CS-2 is above the rate at which the output of CS-1 generates information, then the controlled process $\{X_i : i = 0, \dots,\}$ can be encoded into the control strategy of CS-1 and decoded with arbitrary small error probability, for large enough n. For this to hold, it it necessary that the total power allocated to CS-1 is above a critical level, which is the minimum cost of controlling CS-1. Any additional power is converted into information, which is communicated from its input to its output.

(b) Optimal control performance of the closed loop control interconnected system, and signalling of information can be achieved, with small error probability, by proper control-coding-decoding.

The dual role of control strategies, to control and encode information, although surprising, it is a direct generalization of Shannon's operational definition of capacity of noisy channels to unstable dynamical systems.

Past literature on capacity of noisy communication channels with feedback encoding, is mainly focussed on memoryless communication channels, additive Gaussian noise channels with memory, such as, [8], [9], and finite state channels with special structures, such as, [10]–[12]. Recent progress for general dynamical systems, that may represent stable or unstable control systems of communication channels with memory is found in [1], [7], [13].

Most material of this paper are extensively discussed in [3], where a detailed analysis is included.

II. Networked Control System with Signalling

In this section we introduce the networked control system, the definition of optimality for signalling of information, and we recall some technical results from [1].

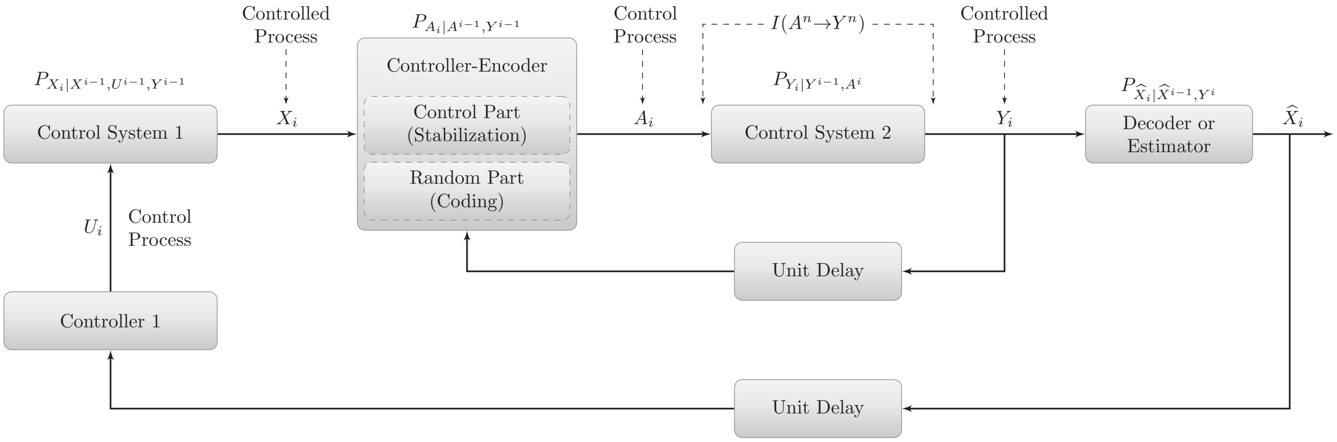

Fig. I.1. *Signalling of Information in Networked Control Systems via Control-Coding.*

A. Networked Control System

The networked control system consists of CS-1 and CS-2, as shown in Fig. I.1. The control process of CS-2 is $A^n \triangleq \{A_i : i = 0, 1, \ldots, n\}$ with values in $\mathbb{A}^n \triangleq \times_{i=0}^n \mathbb{A}_i$, while its controlled processes is $Y^n \triangleq \{Y_i : i = 0, 1, \ldots, n\}$, with values in $\mathbb{Y}^n \triangleq \times_{i=0}^n \mathbb{Y}_i$, with initial state $S \triangleq Y_{-1}$, with values in $\mathbb{S} \triangleq \mathbb{Y}_{-1}$. Similarly, the control process of CS-1 is $U^n \triangleq \{U_i : i = 0, 1, \ldots, n\}$, with values in $\mathbb{U}^n \triangleq \times_{i=0}^n \mathbb{U}_i$, while its controlled process is $X^n \triangleq \{X_i : i = 0, 1, \ldots, n\}$, with values in $\mathbb{X}^n \triangleq \times_{i=0}^n \mathbb{X}_i$. The networked system is described by the following elements.

(i) The Control System-1 (CS-1) is described by

$$\mathbf{P}_{X_i|X^{i-1},U^{i-1},A^{i-1},Y^{i-1},S} = \mathbf{P}_{X_i|X_{i-1},U_{i-1},Y_{i-1}} \equiv S_i(dx_i|x_{i-1}, u_{i-1}, y_{i-1}), \ i = 0, \ldots, n. \quad \text{(II.1)}$$

For $i = 0$, $\mathbf{P}_{X_0|X_{-1},U_{-1},Y_{-1}} = S_0(dx_0)$.

(ii) The Control System-2 (CS-2) is described by

$$\mathbf{P}_{Y_i|Y^{i-1},A^i,S,X^i,U^i} \equiv Q_i(dy_i|y_{i-1}, a_i), \quad i = 0, \ldots, n. \quad \text{(II.2)}$$

(iii) The Control Strategies of CS-1 are measurable maps

$$\mathcal{U}_{[0,n]} \triangleq \Big\{ g_i : \mathbb{U}^{i-1} \times \mathbb{Y}^{i-1} \times \mathbb{S} \longmapsto \mathbb{U}_i, u_0 = g_0(s), \ldots, u_i = g_i(u^{i-1}, y^{i-1}, s), \ i = 0, 1, \ldots, n \Big\} \quad \text{(II.3)}$$

That is, the controller of CS-1 does not have access to the signal $\{X_i : i = 0, \ldots, n\}$, but instead has access to the controlled process $\{Y_i : i = 0, \ldots, n\}$ and initial state $S = s$ of CS-2. However, since any control system has Control-Coding (CC) Capacity [1], our aim is to design a controller-encoder, which controls CS-2 and encodes the controlled process $\{X_i : i = 0, \ldots, n\}$, and then a decoder at the output of CS-2, which obtains an estimate $\{\widehat{X}_i : i = 0, \ldots, n\}$ of $\{X_i : i = 0, \ldots, n\}$, and then design a controller for CS-1, which uses this estimate to control the process $\{X_i : i = 0, \ldots, n\}$ of CS-1.

To this end, we introduce the CC-Capacity of CS-2 [1] and corresponding randomized control strategies that can be transformed into an contriller-encoder, which encodes the process $\{X_i : i = 0, \ldots, n\}$ and operates at the CC Capacity of CS-2.

(iv) The Randomized Control Strategies of (CS-2) are conditional distributions from the set

$$\mathcal{P}_{[0,n]} \triangleq \big\{ P_i(da_i|a^{i-1}, y^{i-1}, s) : i = 0, \ldots, n \big\}.$$

Admissible randomized control strategies belong to the power constraint set defined by

$$\mathcal{P}_{[0,n]}(\kappa) \triangleq \Big\{ P_i(da_i|a^{i-1}, y^{i-1}, s), i = 0, \ldots, n : \frac{1}{n+1} \mathbf{E}\Big(\gamma_{0,n}(A^n, Y^{n-1}) \Big) \leq \kappa \Big\} \subset \mathcal{P}_{[0,n]} \quad \text{(II.4)}$$

$$\gamma_{0,n}(A^n, Y^{n-1}) \triangleq \sum_{i=0}^n \gamma_i(a_i, y_{i-1}) \quad \text{(II.5)}$$

where $\gamma_{0,n}(\cdot,\cdot) : \mathbb{A}^n \times \mathbb{Y}^{n-1} \longmapsto (-\infty, \infty]$ is a measurable function and $\kappa \in [0, \infty]$ is the total power.

(v) The Pay-off or Performance Criterion of CS-2 is the directed information from $A^n \triangleq \{A_0, \ldots, A_n\}$ to $Y^n \triangleq \{Y_0, \ldots, Y_n\}$, conditioned on the initial data $S = s$, and defined by [14], [15]

$$I(A^n \rightarrow Y^n|s) \triangleq \mathbf{E}_s^P \Big\{ \sum_{i=0}^n \log \Big(\frac{dQ_i(\cdot|Y_{i-1}, A_i)}{d\mathbf{P}^P(\cdot|Y^{i-1}, S)}(Y_i) \Big) \Big\} \quad \text{(II.6)}$$

where for each i, $\mathbf{P}^P(dy_i|y^{i-1}, s) \equiv \mathbf{P}_{Y_i|Y^{i-1},S}$ is generated from $\big\{ Q_i(\cdot|\cdot), P_i(\cdot|\cdot) : i = 0, 1, \ldots, n \big\}$.

The performance objective for CS-2 is the information CC

Capacity of CS-2 defined by [1]

$$C(\kappa) = J_{A^\infty \rightarrow Y^\infty|s}(P^*, \kappa) \triangleq \lim_{n \longrightarrow \infty} \frac{1}{n+1} J_{A^n \rightarrow Y^n|s}(P^*, \kappa) \qquad \text{(II.7)}$$

$$C_{0,n}(\kappa) = J_{A^n \rightarrow Y^n|s}(P^*, \kappa) \triangleq \sup_{\mathcal{P}_{[0,n]}(\kappa)} I(A^n \rightarrow Y^n|s). \qquad \text{(II.8)}$$

$J_{A^n \rightarrow Y^n|s}(P^*, \kappa)$ is called the Finite-Time Horizon (FTH) Information CC Capacity of CS-2. Under certain conditions given in [6] (see also [3] for extensive discussion), then $C(\kappa)$ is an upper bound on the the supremum of all achievable CC rates, and any CC rate below $C(\kappa)$ is achievable (see Definitions 2.4, 2.5, Theorem 2.6, 2.8 in [6]).

(vi) The Pay-off or Performance Criterion of CS-1 is

$$J_{0,n}(g^*) \triangleq \inf_{\big\{g_i(\cdot):i=0,\ldots,n\big\} \in \mathcal{U}_{[0,n]}} \mathbf{E}\Big\{\ell_{0,n}(U^n, X^n)\Big\}, \qquad \text{(II.9)}$$

$$\ell_{0,n}(u^n, x^n) \triangleq \sum_{i=0}^{n} \ell_i(u_i, x_i) \qquad \text{(II.10)}$$

where $\ell_{0,n}(\cdot,\cdot)$ is a measurable function.

The algorithm concept based upon which signalling of information to CS-1 is shown, is the following.

(1) Compute the CC-Capacity of CS-2, and the optimal randomized control strategy $\big\{P_i^*(da_i|a^{i-1}, y^{i-1}, s) : i = 0, \ldots, n\big\} \in \mathcal{P}_{[0,n]}(\kappa)$, which achieves it.

(2) Transform the optimal randomized control strategy $\big\{P_i^*(da_i|a^{i-1}, y^{i-1}, s) : i = 0, \ldots, n\big\} \in \mathcal{P}_{[0,n]}(\kappa)$ into an controller-encoder, which controls the output process $\{Y_i : i = 0, \ldots, n\}$, encodes the output process $\{X_t : t = 0, \ldots, n\}$, and reconstructs or estimates the encoded process $\{X_t : t = 0, \ldots, n\}$, by a decoder or estimator to produce $\{\widehat{X}_t : t = 0, \ldots, n\}$.

(3) Apply the estimated process $\{\widehat{X}_t : t = 0, \ldots, n\}$ as input to the controller of CS-1 to control it, by minimizing the pay-off (II.9).

The set of admissible controller-encoder-decoder strategies for CS-2 is defined below.

Definition 2.1: (Admissible controller-encoder-decoders) **(a) Controller-Encoder Strategies of CS-2.** The controller-encoder strategies which control the controlled process $\{Y_i : i = 0, \ldots, n\}$ and encode the controlled process $\{X_i : i = 0, \ldots, n\}$ are measurable maps defined by

$$\mathcal{E}_{[0,n]}(\kappa) \triangleq \Big\{ e_i : \mathbb{X}^i \times \mathbb{A}^{i-1} \times \mathbb{Y}^{i-1} \times \mathbb{S} \longmapsto \mathbb{A}_i, \; a_i = e_i(x^i, a^{i-1}, y^{i-1}, s), \; i = 0, 1, \ldots, n : \frac{1}{n+1} \mathbf{E}_s\Big(\gamma_{0,n}(A^n, Y^{n-1})\Big) \leq \kappa \Big\}.$$

(b) Decoder Strategies of CS-2. The decoder strategies which reconstruct or estimate the process $\{X_i : i = 0, \ldots, n\}$ are square integrable sequences defined by

$$\mathcal{D}_{[0,n]} \triangleq \Big\{ d_i : \mathbb{Y}^i \times \mathbb{S} \longmapsto \widehat{\mathbb{X}}_i, \; \widehat{x}_i = d_i(y^i, s), i = 0, \ldots, n \Big\}.$$

Next, we state some preliminary results which we use in subsequent developments of the paper.

B. Information Structures of Randomized Control Strategies

By [1] the optimal randomized control strategies for the optimization problem (II.8) are Markov defined by

$$\overset{\circ}{\mathcal{P}}_{[0,n]}(\kappa) \triangleq \Big\{ \pi_i(da_i|y_{i-1}), \; i = 0, \ldots, n : \frac{1}{n+1} \mathbf{E}_s^\pi\Big(\sum_{i=0}^{n} \gamma_i(A_i, Y_{i-1})\Big) \leq \kappa \Big\} \subset \mathcal{P}_{[0,n]}(\kappa). \qquad \text{(II.11)}$$

Hence, optimization (II.8) reduces to the following Markov Decision (MD) problem with randomized control strategies.

$$C_{0,n}(\kappa) \triangleq J_{A^n \rightarrow Y^n|s}(\pi^*, \kappa) = \sup_{\overset{\circ}{\mathcal{P}}_{[0,n]}(\kappa)} \mathbf{E}_s^\pi\Big\{ \sum_{i=0}^{n} \log\Big(\frac{Q_i(\cdot|Y_{i-1}, A_i)}{\Pi_i^\pi(\cdot|Y_{i-1})}(Y_i)\Big)\Big\} \equiv \sup_{\overset{\circ}{\mathcal{P}}_{[0,n]}(\kappa)} \sum_{i=0}^{n} I^\pi(A_i; Y_i|Y_{i-1}) \qquad \text{(II.12)}$$

where the output process $\{Y_0, \ldots, Y_n\}$ is Markov, with corresponding transition probability distribution given by

$$\Pi_i^\pi(dy_i|y_{i-1}) = \int_{\mathbb{A}_i} Q_i(dy_i|y_{i-1}, a_i) \otimes \pi_i(da_i|y_{i-1}).$$

Thus, a candidate for the CC Capacity of CC-2 is

$$C(\kappa) \triangleq J_{A^\infty \rightarrow Y^\infty|s}(\pi^*, \kappa) = \lim_{n \longrightarrow \infty} \frac{1}{n+1} J_{A^n \rightarrow Y^n|s}(\pi^*, \kappa).$$

By [1], [16], the inverse function of $C_{0,n}(\kappa), \kappa \in (\kappa_{min}, \infty)$ denoted by $\kappa_{0,n}(C)$, exists and the following hold.

$$\kappa_{0,n}(C) \triangleq \inf_{\pi_i(da_i|y_{i-1}), i=0,\ldots,n: \frac{1}{n+1}\sum_{i=0}^{n} I^\pi(A_i;Y_i|Y_{i-1}) \geq C} \mathbf{E}_s^\pi\Big\{\gamma_{0,n}(A^n, Y^{n-1})\Big\}. \qquad \text{(II.13)}$$

$$\geq \inf_{\pi_i(da_i|y_{i-1}), i=0,\ldots,n} \mathbf{E}_s^\pi\Big\{\gamma_{0,n}(A^n, Y^{n-1})\Big\} \equiv \kappa_{0,n}(0). \qquad \text{(II.14)}$$

Note that $\kappa_{0,n}(C) - \kappa_{0,n}(0)$ is the cost of signalling $\{X_t : t = 0, \ldots, n\}$ to the output of CS-2.

C. Multi-Objective Optimality of Strategies

By the data processing inequalities [17], no controller-encoder strategy can operate at a higher information rate than $C_{0,n}(\kappa)$. Similarly, no decoder strategy no matter what post-processing is done by the decoder, can operate at an information rate higher than $C_{0,n}(\kappa)$. Thus, the fundamental problem of finding the controller-encoder-decoder strategy is reduced to the problem of determining an optimal controller-encoder-decoder with respect to an error criterion, which operates at $C_{0,n}(\kappa)$. In view of the above, we introduce the following definition of optimality.

Definition 2.2: (Optimal strategies)
A quadruple {controller-encoder, decoder, controller}, $(e^o(\cdot), d^o(\cdot), g^o(\cdot)) \in \mathcal{E}_{[0,n]}(\kappa) \times \mathcal{D}_{[0,n]} \times \mathcal{U}_{[0,n]}$ is optimal, if the following hold.
(i) For given $g(\cdot), d(\cdot)$ the strategy $e^o(\cdot, g(\cdot), d(\cdot)) \in \mathcal{E}_{[0,n]}(\kappa)$ operates at $J_{A^n \rightarrow Y^n|s}(\pi^*, \kappa)$, called information lossless.
(ii) For a given $g(\cdot)$ the decoder $d^o(\cdot) \in \mathcal{D}_{[0,n]}$ satisfies

$$\widehat{J}_{0,n}(g, d^o(\cdot, g), e^o(\cdot, g, d^o)) \triangleq \mathbf{E}_s^{g,e^o,d^o}\Big\{\sum_{i=0}^n \rho_i(X_i, \widehat{X}_i)\Big\}$$
$$\leq \widehat{J}_{0,n}(g, d(\cdot, g), e^o(\cdot, g, d)),\ \forall (d,g) \in \mathcal{D}_{[0,n]} \times \mathcal{U}_{[0,n]}$$

where $\rho_i : \mathbb{X}_i \times \widehat{\mathbb{X}}_i \longmapsto [0,\infty), (x,\widehat{x}) \longmapsto \rho_i(x,\widehat{x}), i = 0,\ldots,n$ is the error fidelity. The mean square error (MSE) fidelity is defined by $\rho_i(x,\widehat{x}) \triangleq |x-\widehat{x}|^2, i = 0,\ldots,n$.
(iii) The control strategy of the CS-1 $g^o(\cdot) \in \mathcal{U}_{[0,n]}$ satisfies

$$\begin{aligned} &J_{0,n}(g^o(\cdot), d^o(\cdot, g^o), e^o(\cdot, g^o, d^o)) \\ &\triangleq \mathbf{E}_s^{g^o,e^o,d^o}\Big\{\ell_{0,n}(X^n, U^n)\Big\} \\ &\leq J_{0,n}(g(\cdot), d^o(\cdot, g), e^o(\cdot, g, d^o)),\ \forall g(\cdot) \in \mathcal{U}_{[0,n]} \end{aligned} \tag{II.15}$$

Next, we wish to quantify the definition of an controller-encoder strategy to operate at $C_{0,n}(\kappa)$, i.e., to be information lossless.

Next, we give the characterization theorem of information lossless controller-encoder strategies.

Theorem 2.1: (Information structures of information lossless strategies)
For a given $g(\cdot), d(\cdot)$, let $e^o(\cdot, g(\cdot), d(\cdot)) \in \mathcal{E}_{[0,n]}(\kappa)$ be the strategy that maximizes the directed information from X^n to Y^n conditioned on $S = s$, defined by

$$\begin{aligned} I_{X^n \rightarrow Y^n|s}(e^*, \kappa) \triangleq \sup_{\mathcal{E}_{[0,n]}(\kappa)} \mathbf{E}_s^{e}\Big\{\sum_{i=0}^n \\ \log\Big(\frac{d\mathbf{P}(\cdot|Y_{i-1}, e_i(X^i, A^{i-1}, Y^{i-1}, S))}{\mathbf{P}^e(\cdot|Y^{i-1}, S)}(Y_i)\Big)\Big\}. \end{aligned} \tag{II.16}$$

Then the following hold.
(i) The optimal strategy in (II.16) occurs in the subset of Markov strategies in $\{X_i : i = 0,\ldots,n\}$, defined by

$$\begin{aligned} \overset{\circ}{\mathcal{E}}_{[0,n]}(\kappa) \triangleq \Big\{\mu_i{:}\mathbb{X}_i\times\mathbb{Y}^{i-1}\times\mathbb{S} \longmapsto \mathbb{A}_i,\ a_i = \mu_i(x_i, y^{i-1}, s), \\ i = 0,1,\ldots,n{:}\frac{1}{n+1}\mathbf{E}_s^{\mu}\Big(\gamma_{0,n}(A^n, Y^{n-1})\Big) \leq \kappa\Big\} \subset \mathcal{E}_{[0,n]}(\kappa) \end{aligned}$$

and moreover, $I_{X^n \rightarrow Y^n|s}(e^*, \kappa)$ reduces to the following (simplified) expression.

$$\begin{aligned} I_{X^n \rightarrow Y^n|s}(\mu^*, \kappa) \triangleq \sup_{\overset{\circ}{\mathcal{E}}_{[0,n]}(\kappa)} \mathbf{E}_s^{\mu}\Big\{\sum_{i=0}^n \\ \log\Big(\frac{d\mathbf{P}(\cdot|Y_{i-1}, \mu_i(X_i, Y^{i-1}, S))}{\mathbf{P}^{\mu}(\cdot|Y^{i-1}, S)}(Y_i)\Big)\Big\} \end{aligned} \tag{II.17}$$

$$\equiv \sup_{\overset{\circ}{\mathcal{E}}_{[0,n]}(\kappa)} \sum_{i=0}^n I^{\mu}(X_i; Y_i|Y^{i-1}, s) \tag{II.18}$$

$$\begin{aligned} \mathbf{P}^{\mu}(dy_i|y^{i-1}, s) = \int_{\mathbb{X}_i} \mathbf{P}(dy_i|y_{i-1}, \mu_i(x_i, y^{i-1}, s)) \\ \mathbf{P}^{\mu}(dx_i|y^{i-1}, s). \end{aligned} \tag{II.19}$$

(ii) An optimal information lossless controller-encoder strategy $\{\mu_i^*(\cdot,\cdot) : i = 0,\ldots,n\} \in \overset{\circ}{\mathcal{E}}_{[0,n]}(\kappa)$ satisfies the following identity.

$$I_{X^n \rightarrow Y^n|s}(\mu^*, \kappa) = J_{A^n \rightarrow Y^n|s}(\pi^*, \kappa). \tag{II.20}$$

Proof: See [3]. ■

The main point of this section is that, the information lossless characterization of optimal controller-encoder strategies relates the optimal randomized control strategy $\{\pi_i^*(da_i|y_{i-1}) : i = 0,\ldots,n\} \in \overset{\circ}{\mathcal{P}}_{[0,n]}(\kappa)$ corresponding to $J_{A^n \rightarrow Y^n|s}(\pi^*, \kappa)$, to the problem of synthesizing an optimal controller-encoder strategy $\{\mu_i^*(x_i, y^{i-1}, s) : i = 0,\ldots,n\} \in \overset{\circ}{\mathcal{E}}_{[0,n]}(\kappa)$, which encodes $\{X_t : t = 0,\ldots,n\}$ and achieves the control objectives of CS-2, while operating at $J_{A^n \rightarrow Y^n|s}(\pi^*, \kappa)$. This constructive methodology is illustrated in Section III.

III. Signalling in Gaussian Networks

CS-2 is a Gaussian Linear Decision Model (GL-DM-2) with quadratic cost function defined, for $i = 0,\ldots,n$, as follows.

$$Y_i = C_{i-1}\, Y_{i-1} + D_i\, A_i + V_i, \quad Y_{-1} = y_{-1} \equiv s, \tag{III.21}$$
$$\mathbf{P}_{V_i|V^{i-1}, A^i, Y_{-1}} = \mathbf{P}_{V_i}(dv_i), \quad V_i \sim N(0, K_{V_i}), \tag{III.22}$$
$$\gamma_i(a_i, y_{i-1}) \triangleq \langle a_i, R_i a_i\rangle + \langle y_{i-1}, Q_{i-1} y_{i-1}\rangle, \tag{III.23}$$
$$(C_{i-1}, D_i) \in \mathbb{R}^{p\times p} \times \mathbb{R}^{p\times q}, \quad (Q_{i-1}, R_i) \in \mathbb{S}_{+}^{p\times p} \times \mathbb{S}_{++}^{q\times q}. \tag{III.24}$$

The control system distribution is given by $Q_i(dy_i|y_{i-1}, a_i) \sim N(C_{i-1}y_{i-1} + D_i a_i, K_{V_i}), i = 0,\ldots,n$.
CS-1 is a Gaussian Linear Decision Model (GL-DM-1) with quadratic cost function described recursive over the horizon $i = 0,1,\ldots,n$, by

$$X_{i+1} = H_i Y_i + F_i X_i + B_i U_i + G_i W_i, \quad X_0 = x \in \mathbb{R}^q \tag{III.25}$$

$$\ell_i(x_i, u_i) \triangleq \langle u_i, \tilde{R}_i u_i\rangle + \langle x_i, \tilde{Q}_i x_i\rangle, \quad U_i = g_i(Y^{i-1}), \tag{III.26}$$

$$B_i \in \mathbb{R}^{q\times m}, \quad (\tilde{Q}_i, \tilde{R}_i) \in \mathbb{S}_{+}^{q\times q} \times \mathbb{S}_{++}^{m\times m} \tag{III.27}$$

where $\{W_i \sim N(0, K_{W_i}) : i = 0,\ldots,n-1\}$ are $\mathbb{W}_i = \mathbb{R}^k-$valued zero mean Gaussian processes independent of the Gaussian RV X_0, i.e., $\mathbf{P}_{X_0}(dx) \sim N(0, K_{X_0})$. By

(III.22), the noise process $\{W_i : i = 0, \dots, n-1\}$ is independent of the noises process $\{V_i : i = 0, 1, \dots, n\}$.

A. Information CC-Capacity: Hierarchical Optimality

Orthogonal Decomposition of Optimal Strategies. By [1], the optimal randomized control strategy is Gaussian, denoted by $\{\pi_i^g(da_i|y_{i-1}) : i = 0, \dots, n\}$, and the process $A_i = A_i^g$ is Gaussian, and realized by

$$A_i^g = e_i^g(Y_{i-1}^g, Z_i^g) = U_i^g + Z_i = \Gamma_i Y_{i-1}^g + Z_i^g, \quad U_i^g \triangleq \Gamma_i Y_{i-1}^g,$$
$$Y_i^g = \Big(C_{i-1} + D_i \Gamma_i\Big) Y_{i-1}^g + D_i Z_i^g + V_i, \quad Y_{-1}^g = y_{-1}$$

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

Hierarchical Decomposition and Separation Principle. (a) The optimal predictable part of the strategy $\{\Gamma_i^* : i = 0, \dots, n\}$, is given by

$$u_i^{g,*} = \overline{e}_i^{g,*}(y_{i-1}) = \Gamma_i^* y_{i-1}, \qquad i = 0, \dots, n, \tag{III.28}$$
$$\Gamma_i^* = -\Big(D_i^T P(i+1) D_i + R_i\Big)^{-1} D_i^T P(i+1) C_{i-1} \tag{III.29}$$

where $\Gamma_n^* = 0$ and $\Big\{P(i) : i = 0, \dots, n\Big\}$ is a solution of the Riccati difference matrix equation

$$\begin{aligned} P(i) &= C_{i-1}^T P(i+1) C_{i-1} + Q_{i-1} \\ &\quad - C_{i,i-1}^T P(i+1) D_i \Big(D_i^T P(i+1) D_i + R_i\Big)^{-1} \\ &\quad \Big(C_{i-1}^T P(i+1) D_i\Big)^T, \quad P(n) = Q_{n-1}. \end{aligned} \tag{III.30}$$

(b) The optimal randomized part of the strategy $\Big\{K_{Z_i}^* : i = 0, \dots, n\Big\}$ is obtained from the solution of the following water-filling problem.

$$\begin{aligned} r(i) =& r(i+1) + \sup_{K_{Z_i} \succeq 0} \Big\{ \frac{1}{2} \log \frac{|D_i K_{Z_i} D_i^T + K_{V_i}|}{|K_{V_i}|} \\ &- \lambda\, tr\Big(P(i+1)\Big[D_i K_{Z_i} D_i^T + K_{V_i}\Big]\Big) \\ &- \lambda\, tr\Big(R_i K_{Z_i}\Big)\Big\}, \quad i = n-1, \dots, 0, \end{aligned} \tag{III.31}$$

$$\begin{aligned} r(n) =& \sup_{K_{Z_n} \succeq 0} \Big\{ \frac{1}{2} \log \frac{|D_n K_{Z_n} D_n^T + K_{V_n}|}{|K_{V_n}|} \\ &- \lambda\, tr\Big(R_n K_{Z_n}\Big) + \lambda(n+1)\kappa \Big\} \end{aligned} \tag{III.32}$$

where $\lambda \equiv \lambda_n(\kappa) \geq 0$ is the Lagrange multiplier, which is found from the average constraint. Moreover,

$$J_{A^n \rightarrow Y^n|s}(\pi^{g,*}, \kappa) = -\lambda\, \langle y_{-1}, P(0) y_{-1} \rangle + r(0). \tag{III.33}$$

B. Signalling of Information: Controller-Encoder-Decoder Strategies

In this section, we state the main theorem that gives the optimal quadruple of strategies {controller-encoder, decoder, controller}, according to Definition 2.2.

Theorem 3.1: (optimal quadruple of strategies {controller-encoder, decoder, controller}) Consider the CS-1, i.e., $\{X_i : i = 0, 1, \dots, n\}$, which is to be encoded and transmitted over the CS-2 defined by (III.21)-(III.24). Let $\Big\{\big(\Gamma_i^*, K_{Z_i}^*\big) : i = 0, \dots, n\Big\}$ be the optimal strategy given by (III.29) and (III.31), (III.32) with corresponding optimal Gaussian conditional distribution $\{\pi_i^{g,*}(da_i|y_{i-1}) : i = 0, \dots, \}$ and joint process $\{(A_i^*, Y_i^*) : i = 0, \dots, n\}$, which achieves $J_{A^n \rightarrow Y^n|s}(\pi^{g,*}, \kappa)$.
Define the filter estimates[1] and conditional covariances by

$$\widehat{X}_{i|i-1} \triangleq \mathbf{E}_s\Big\{X_i \Big| Y^{*,i-1}\Big\}, \quad \widehat{X}_{i|i} \triangleq \mathbf{E}_s\Big\{X_i \Big| Y^{*,i}\Big\},$$
$$\Sigma_{i|i-1} \triangleq \mathbf{E}_s\Big\{\Big(X_i - \widehat{X}_{i|i-1}\Big)\Big(X_i - \widehat{X}_{i|i-1}\Big)^T \Big| Y^{*,i-1}\Big\},$$
$$\Sigma_{i|i} \triangleq \mathbf{E}_s\Big\{\Big(X_i - \widehat{X}_{i|i}\Big)\Big(X_i - \widehat{X}_{i|i}\Big)^T \Big| Y^{*,i}\Big\}, \ i = 0, \dots, n.$$

Then the encoder strategy[2] and corresponding controlled process, which operates at $J_{A^n \rightarrow Y^n|s}(\pi^{g,*}, \kappa)$, is given by

$$A_i^* = \mu_i^*(X_i, Y^{*,i-1}) = \Gamma_i^* Y_{i-1}^* + \Theta_i^* \Big\{X_i - \widehat{X}_{i|i-1}\Big\},$$
$$\Theta_i^* = K_{Z_i}^{*,\frac{1}{2}} \Sigma_{i|i-1}^{-\frac{1}{2}}, \quad \Theta_i^* \succeq 0,$$
$$Y_i^* = \Big(C_{i-1} + D_i \Gamma_i^*\Big) Y_{i-1}^* + D_i\, \Theta_i^* \Big\{X_i - \widehat{X}_{i|i-1}\Big\} + V_i,$$

for $i = 0, \dots, n$. Moreover, the following hold.
(a) Filter Estimates. The innovations process defined by $\Big\{\nu_i^* \triangleq Y_i^* - \mathbf{E}\{Y_i^* \Big| Y^{*,i-1}\Big\} : i = 0, \dots, n\Big\}$ satisfies

$$\nu_i^* = Y_i^* - \Big(C_{i-1} + D_i \Gamma_i^*\Big) Y_{i-1}^* = D_i\, \Theta_i^* \Big\{X_i - \widehat{X}_{i|i-1}\Big\} + V_i,$$
$$\mathbf{E}_s\Big\{\nu_i^* \Big| Y^{*,i-1}\Big\} = \mathbf{E}_s\Big\{\nu_i^*\Big\} = 0,$$
$$\mathbf{E}_s\Big\{\nu_i^* (\nu_i^*)^T \Big| Y^{*,i-1}\Big\} = D_i K_{Z_i}^* D_i^T + K_{V_i} = \mathbf{E}_s\Big\{\nu_i^* (\nu_i^*)^T\Big\}$$

and the sequence of RVs, $\Big\{\nu_i^* : i = 0, \dots, n\Big\}$, is uncorrelated. The optimal filter estimates satisfy the following recursions.

$$\begin{aligned} \widehat{X}_{i+1|i} &= H_i Y_i^* + F_i \widehat{X}_{i|i-1} + B_i g_i(Y^{*,i-1}) \\ &\quad + \Psi_{i|i-1} \nu_i^*, \quad \widehat{X}_{0|-1} = \text{Given}, \end{aligned} \tag{III.34}$$

$$\begin{aligned} \Sigma_{i+|i} &= F_i \Sigma_{i|i-1} F_i^T + G_i K_{W_i} G_i^T - F_i \Sigma_{i|i-1} \Big(D_i \Theta_i^*\Big)^T \\ &\quad \Big[D_i K_{Z_i}^* D_i^T + K_{V_i}\Big]^{-1} \Big(D_i \Theta_i^*\Big) \Sigma_{i|i-1} F_i^T, \end{aligned} \tag{III.35}$$

$$\Sigma_{0|-1} = \mathbf{E}_s\Big\{\Big(X_0 - \widehat{X}_{0|-1}\Big)\Big(X_0 - \widehat{X}_{0|-1}\Big)^T\Big\} \tag{III.36}$$

$$\Sigma_{i|i} = \Sigma_{i|i-1} - \overline{\Psi}_{i|i-1} \Big(D_i \Theta_i^*\Big) \Sigma_{i|i-1} \tag{III.37}$$

where the filter gains are difined by

$$\Psi_{i|i-1} \triangleq F_i \overline{\Psi}_{i|i-1},$$
$$\overline{\Psi}_{i|i-1} \triangleq \Sigma_{i|i-1} \Big(D_i \Theta_i^*\Big)^T \Big[D_i K_{Z_i}^* D_i^T + K_{V_i}\Big]^{-1} \tag{III.38}$$

[1]For simplicity the superscript on $\mathbf{E}_s$ is omitted.
[2]For any square matrix M with real entries $M^{\frac{1}{2}}$ is its square root.

and the controlled process $\left\{Y_i^*: i=0,\ldots,n\right\}$ is given by

$$Y_i^* = \Big(C_{i,i-1}+D_i\Gamma_i^*\Big)Y_{i-1}^*+\nu_i^*, \quad i=0,1,\ldots. \tag{III.39}$$

(b) Information Lossless Controller-Encoder Operating at $J_{A^n\rightarrow Y^n|s}(\pi^{g,*},\kappa)$. The controller-encoder $\{\mu_i^*(\cdot,\cdot): i=0,\ldots,\}$ is information lossless, that is,

$$\begin{aligned} I_{X^n\rightarrow Y^n|s}(\mu^*,\kappa) =& \sum_{i=0}^{n}\Big\{H(\nu_i^*)-H(V_i)\Big\} \\ =& J_{A^n\rightarrow Y^n|s}(\pi^{g,*},\kappa). \end{aligned} \tag{III.40}$$

(c) The optimal decoder is the Kalman Filter $\widehat{X}_{i|i}, i=0,\ldots,n$.
(d) The optimization problem of CS-1 is given by

$$\begin{aligned} J_{0,n}(g^*) \triangleq& \inf_{\left\{g_i(\cdot):i=0,\ldots,n\right\}\in\mathcal{U}_{[0,n]}} \mathbf{E}_s\Big\{\sum_{i=0}^{n}\Big(\langle U_i,\tilde{R}_iU_i\rangle \\ &+\langle X_i,\tilde{Q}_iX_i\rangle\Big)\Big\}, \quad U_i=g_i(Y^{*,i-1}) \end{aligned} \tag{III.41}$$

$$\begin{aligned} =& \inf_{\left\{g_i(\cdot):i=0,\ldots,n\right\}\in\mathcal{U}_{[0,n]}} \mathbf{E}_s\Big\{\sum_{i=0}^{n}\Big(\langle U_i,\tilde{R}_iU_i\rangle \\ &+\langle \widehat{X}_{i|i-1},\tilde{Q}_i\widehat{X}_{i|i-1}\rangle\Big)\Big\}+\sum_{i=0}^{n}Tr\Big(\tilde{Q}_i\Sigma_{i|i-1}\Big). \end{aligned} \tag{III.42}$$

subject to the constraint

$$\begin{aligned} \widehat{X}_{i+1|i} =& H_iY_i^*+F_i\widehat{X}_{i|i-1}+B_ig_i(Y^{*,i-1}) \\ &+\Psi_{i|i-1}\nu_i^*, \quad \widehat{X}_{0|-1}=\text{Given}, \end{aligned} \tag{III.43}$$

$$Y_i^* = \Big(C_{i,i-1}+D_i\Gamma_i^*\Big)Y_{i-1}^*+\nu_i^*, \quad i=0,1,\ldots. \tag{III.44}$$

Moreover, the optimal strategy of CS-1 is linear in $\overline{X}_i \triangleq (Y_{i-1}^*,\widehat{X}_{i|i-1}), i=0,\ldots,n$, i.e., $U_i^*=g_i^*(\overline{X}_i)=K_i\overline{X}_i, i=0,\ldots,n$, where $K_i, i=0,\ldots,n$ satisfied a control Riccati equation.

Proof: (a), (b) This is shown in [1]. (c), (d) follow from [3], [6]. Specifically, (III.42) is obtained by reconditioning. Substituting (III.44) into (III.43) we obtain

$$\begin{aligned} \widehat{X}_{i+1|i} =& H_i\Big(C_{i,i-1}+D_i\Gamma_i^*\Big)Y_{i-1}^*+F_i\widehat{X}_{i|i-1} \\ &+B_ig_i(Y^{*,i-1})+\Big(\Psi_{i|i-1}+I\Big)\nu_i^*, \ \widehat{X}_{0|-1}=\text{Given}. \end{aligned} \tag{III.45}$$

Since the state variable $\overline{X}_i \triangleq (Y_{i-1}^*,\widehat{X}_{i|i-1}), i=0,\ldots,n$ is Markov, i.e., $\mathbf{P}_{\overline{X}_{i+1}|\overline{X}^i,U^i}=\mathbf{P}_{\overline{X}_{i+1}|\overline{X}_i,U_i}, i=0,\ldots,n$, then the optimal strategy of CS-1 is linear in $\overline{X}_i$ (by LQG stochastic optimal control theory). ■

It is straight forward to invoke the above solution to determine the CC capacity of the control system, given by the per unit limit $C(\kappa)\triangleq\lim_{n\longrightarrow\infty}\frac{1}{n+1}J_{A^n\rightarrow Y^n|s}(\pi^*,\kappa)$, to analyze the asymptotic limit of Theorem 3.1, and to understand the implications of unstable control systems on the control-coding capacity. Such analysis is found in [1], [3], [6].

IV. Conclusions

A hierarchical constructive procedure is developed to synthesize *optimal controllers-encoders-decoders*, to signal information from the controller of one control system to the controllers of another control system. The method is based on CC Capacity of stochastic systems.

References

[1] Charalambous, C. Kourtellaris, and I. Tzortzis, "Information transfer of control strategies: Dualities of stochastic optimal control theory and feedback capacity of information theory," *IEEE Transactions on Automatic Control*, submitted, May 2016.

[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] C. D. Charalambous, C. Kourtellaris, and I. Tzortzis, "Ergodic control-coding capacity of stochastic controls systems: Information signalling and hierarchical optimality of gaussian systems," *IEEE Transactions on Automatic Control*, Submitted, February 2017.

[4] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech.*, vol. 27, pp. 379–423, July 1948.

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

[6] C. K. Charalambous, C. D. and I. Tzortzis, "Hierarchical decomposition of optimal control and information strategies in control-coding capacity of stochastic systems," in *Proceedings of 56th IEEE Conference on Decision and Control*, Melbourne, Australia, December 12-15 2017.

[7] P. A. Stavrou, C. D. Charalambous, and C. Kourtellaris, "Sequential necessary and sufficient conditions for capacity achieving distributions of channels with memory and feedback," *IEEE Transactions on Information Theory*, Accepted, April 2017.

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

[9] Y.-H. Kim, "A coding theorem for a class of stationary channels with feedback," *IEEE Transactions on Information Theory*, vol. 54, no. 4, pp. 1488–1499, 2008.

[10] S. Yang, A. Kavcic, and S. Tatikonda, "Feedback capacity of finite-state machine channels," *Information Theory, IEEE Transactions on*, vol. 51, no. 3, pp. 799–810, March 2005.

[11] H. Permuter, P. Cuff, B. Van Roy, and T. Weissman, "Capacity of the trapdoor channel with feedback," *IEEE Transactions on Information Theory*, 2010.

[12] O. Elishco and H. Permuter, "Capacity and coding of the ising channel with feedback," *IEEE Transactions on Information Theory*, vol. 60, no. 9, pp. 3138–5149, June 2014.

[13] C. D. Charalambous and C. Kourtellaris, "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

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

[15] 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.

[16] 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.

[17] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. John Wiley & Sons, Inc., Hoboken, New Jersey, 2006.

[18] C. D. Charalambous, C. Kourtellaris, and I. Tzortzis, "Information transfer in stochastic optimal control with randomized strategies and directed information criterion," in *Proceedings of 55th IEEE Conference on Decision and Control*, Ddecember 12-16 2016, to appear.