2018 IEEE Conference on Decision and Control (CDC)
Miami Beach, FL, USA, Dec. 17-19, 2018

Control-Coding Capacity of Decision Models Driven by Correlated Noise and Gaussian Application Examples

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

***Abstract*— We characterize the $n-$finite time feedback information (FTFI) control-coding capacity of decision models (DMs) driven by correlated noise. Under information stability the per unit limit, called control-coding (CC) capacity of the DM is operational, and analogous to Shannon's coding capacity of noisy communication channels, with the encoder replaced by a controller-encoder.**

We also analyze application examples of recursive linear DMs driven by correlated Gaussian noise, subject to an average cost constraint of quadratic form, called linear-quadratic Gaussian DMs (LQG-DMs). In one of the main theorems we show that the optimal randomized control strategies that achieve the $n-$FTFI CC capacity of the LQG-DMs, consist of multiple parts, that include control, estimation, and information transmission/signalling strategies, and that these strategies are determined using decentralized optimization techniques.

I. Introduction

Our main goal in this paper is to further develop Shannon's operational definition of coding capacity, in different scientific communities, which are motivated by feedback control system applications [1]. Hence, our underlying assumptions differ from those often imposed in the *information theory* literature of feedback capacity of noisy communication channels. Specifically, in control theory and its applications, the dominant mathematical models are unstable dynamical systems, and the role of feedback control inputs is to control output signals, and to achieve optimal performance.

Our main objectives are:

(1) to determine the characterization of $n-$finite time feedback information (FTFI) control-coding (CC) capacity of general nonlinear recursive decision models (DMs), driven by correlated noise, subject to average cost constraints.

Under the technical conditions of information stability then the per unit limit of the $n-$FTFI CC capacity, called CC capacity of the DM, is operational. This means, for any message set with rate in bits/second below the CC capacity, then randomized control strategies can be transformed into controllers-encoders that simultaneously control outputs, and signal messages to the decoder that reconstructs them with arbitrary small asymptotic error probability.

(2) to briefly analyze recursive linear DMs with past dependence on both inputs and output, and driven by correlated Gaussian noise, subject to an average cost of quadratic form, called LQG-DMs.

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

Previous work on the characterization of CC capacity of DMs with past dependence only on outputs is found in [2], while the analysis of linear DMs, stable or unstable, driven by independent Gaussian noise, is found in [3]. Specifically, in [2] (see Theorem 4.1, Theorem 5.2) it is shown that the optimal randomized strategy that achieves the CC capacity of the DM consists of a control part that controls the outputs, and an innovations part that is responsible to encode messages. We note that [1]–[3] demonstrate that Shannon's coding capacity extends naturally to unstable DMs, such as, stochastic control systems and unstable communication channels with memory. In [3], it is also shown that for unstable DMs, the operational definition of achievable rate is a variant of Shannon's *coding rate*, with the encoder replaced by a controller-encoder, called *control-coding* (CC) rate. Further, [4] demonstrates that optimal randomized strategies can be transformed into controller-encoders that simultaneously control outputs, encode Gaussian messages, and signal the messages to the decoder that reconstructs them with asymptotic arbitrary small error.

A. Mathematical Problem and Main Results

The optimization problems of CC capacity of the DMs considered in this paper are generalizations of stochastic optimal control problems with partial information, contrary to the DMs in [2], [3], which are generalizations of stochastic optimal control problems with complete information.

Definition 1.1: (Nonlinear decision model)
The DM is denoted by DM$\{M, L, T\}$, and defined by

$$B_i = h_i(B^{i-1}_{i-M}, A^i_{i-L}, V_i), \quad i = 0, \ldots, n, \tag{1}$$

$$\overline{S} \triangleq (B^{-1}_{-M}, A^{-1}_{-L}) = (b^{-1}_{-M}, a^{-1}_{-L}) \ \text{ is the initial state,} \tag{2}$$

$$\frac{1}{n+1}\mathbf{E}\Big\{\sum_{i=0}^{n} \gamma_i(V^i_{i-T}, A^i_{i-L}, B^i_{i-M})\Big\} \leq \kappa \in [0, \infty), \tag{3}$$

$$\mathbf{P}_{V_i|V^{i-1},A^i,B^{i-1}}(dv_i|v^{i-1}, a^i, b^{i-1}) = \mathbf{P}_{V_i|V^{i-1}_{i-T}}(dv_i|v^{i-1}_{i-T}), \quad i = 0, \ldots, n. \tag{4}$$

The realizations of the RVs are $A_i = a_i \in \mathbb{A}_i, B_i = b_i \in \mathbb{B}_i, V_i = v_i \in \mathbb{V}_i, \forall i \in \{\ldots, -1, 0, 1, \ldots, n\}$, where the spaces are finite-dimensional. We use the notation $z^m_k \triangleq \{z_k, z_{k+1}, \ldots, z_m\} \in \mathbb{Z}^m_k \triangleq \times^m_{j=k}\mathbb{Z}^m_k$, and convention $z^m \triangleq \{\ldots, z_1, z_0, z_1, \ldots, z_m\}$. We assume the following.
(a.i) $h_i : \mathbb{B}^{i-1}_{i-M} \times \mathbb{A}^i_{i-L} \times \mathbb{V}_i \longmapsto \mathbb{B}_i, \gamma_i : \mathbb{V}^i_{i-T} \times \mathbb{A}^i_{i-L} \times \mathbb{B}^i_{i-M} \longmapsto \mathbb{A}_i$ are measurable functions, $i = 0, \ldots, n$.
(a.ii) The inverse of the map $v_i \in \mathbb{V}_i \longmapsto h(b^{i-1}_{i-M_1}, a^i_{i-L_1}, v_i)$ exists and it is measurable.

DM$\{M, L, T\}$ includes nonlinear autoregressive models that can be expressed in state space form [5], and may correspond to unstable control systems or communication channels. Next, we introduce the application example we discuss in the paper.

Definition 1.2: (Linear-Quadratic Gaussian DM)
The Linear-Quadratic Gaussian DM, denoted by LQG-DM$\{M=1, L=1, T=1\}$, is defined by

$$B_i = C_{i,i-1}B_{i-1} + D_{i,i}A_i + D_{i,i-1}A_{i-1} + V_i, \tag{5}$$
$$V_i = F_{i-1}V_{i-1} + W_i, \ \{W_i : i = 0, \ldots, n\} \text{ ind. seq.}, \tag{6}$$
$$W_i \sim N(0, K_{W_i}) \ \text{ independent Gaussian, indep. of } \overline{S}, \tag{7}$$
$$\overline{S} \stackrel{\triangle}{=} (B_{-1}, A_{i-1}, V_{-1}) = (b_{-1}, a_{-1}, v_{-1}), \tag{8}$$
$$\overline{\gamma}_i(a_i, b_{i-1}) \stackrel{\triangle}{=} \langle a_i, R_{i,i}a_i\rangle + \langle b_{i-1}, Q_{i,i-1}b_{i-1}\rangle, \tag{9}$$
$$C_{i,i-1} \in \mathbb{R}^{p\times p}, \quad (D_{i,i}, D_{i,i-1}) \in \mathbb{R}^{p\times q} \times \mathbb{R}^{p\times q}, \tag{10}$$
$$R_{i,i} \in \mathbb{S}_{++}^{q\times q}, \quad Q_{i,i-1} \in \mathbb{S}_{+}^{p\times p}, \quad i = 0, \ldots, n \tag{11}$$

where $\langle \cdot, \cdot \rangle$ denotes inner product, $\mathbb{S}_{+}^{p\times p}$ (resp. $\mathbb{S}_{++}^{p\times p}$) the set of $p \times p$ symmetric positive semidefinite (resp. positive definite) matrices, and $N(\mu, \Sigma)$ means Gaussian distribution with mean μ and covariance Σ.

Main Results. Now, we describe the main results.

(a) Theorem 2.5 gives the converse CC theorem, i.e., that any achievable CC rate is bounded above by an extremum problem of an information measure, maximized over all randomized strategies $\mathbf{P}_{A_i|A^{i-1},B^{i-1},V^{i-1}}, i = 0, \ldots, n$ that satisfy the average power constraints, and conditions for the direct CC theorem.

(b) Theorem 2.6 gives the information structures of the maximizing distributions of the extremum problem in (a).

(c) Section III shows that, for the LQG-DM$\{M = 1, L = 1, T = 1\}$, the optimal input process A^n of the $n-$FTFI capacity is generated by multiple linear strategies, given by

$$A_i = e_i(B_0^{i-1}, A_{i-1}, V_{i-1}, Z_i, \overline{s}), \quad i = 0, \ldots, n \tag{12}$$
$$= U_i + \Lambda_{i,i-1}^1 A_{i-1} + \Lambda_{i,i-1}^2 V_{i-1} + Z_i \tag{13}$$
$$\equiv U_i + \Lambda_{i,i-1}S_{i-1} + Z_i, \tag{14}$$
$$U_i = \overline{e}_i(B_0^{i-1}, \overline{s}) = \Gamma_{i,i-1}^1 B_{i-1} + \Gamma_{i,i-1}^2 \widehat{S}_{i-1|i-1}, \tag{15}$$
$$\widehat{S}_{i-1|i-1} \stackrel{\triangle}{=} \mathbf{E}_{\overline{s}}\left\{S_{i-1}\middle|B_0^{i-1}\right\}, \tag{16}$$
$$i)\ Z_i \text{ indep. of } \left\{\overline{S}, B_0^{i-1}, A_0^{i-1}, V_0^{i-1}\right\}, \ i = 0, \ldots, n, \tag{17}$$
$$ii)\ Z_0^i \ \text{ indep. of } \ V_0^i, \quad i = 0, \ldots, n, \tag{18}$$
$$iii)\ \left\{Z_i \sim N(0, K_{Z_i}) : i = 0, \ldots, n\right\} \text{ independent}, \tag{19}$$
$$\Lambda_{i,i-1} \stackrel{\triangle}{=} \left[\ \Lambda_{i,i-1}^1 \quad \Lambda_{i,i-1}^2\ \right], \quad S_i \stackrel{\triangle}{=} \begin{bmatrix} A_i \\ V_i \end{bmatrix} \tag{20}$$

where $S_i, i = 0, \ldots, n$ is the state process, that is available to controller-encoder, but not the decoder.
The characterization of $n-$FTFI CC capacity is given by

$$C_{0,n}(\kappa) = \sup_{e_i(\cdot):i=0,\ldots,n:} \left\{H(B_0^n|\overline{s}) - H(V_0^n|v_{-1})\right\}, \tag{21}$$
$$K_{Z_i} \succeq 0, \ i = 0, \ldots, n, \ \mathbf{E}_{\overline{s}}\Big\{\sum_{i=0}^n \overline{\gamma}_i(A_i, B_{i-1})\Big\} \leq \kappa \tag{22}$$

where $H(X|\overline{s})$ is the entropy of RV X. We note that $e_i = (\overline{e}_i(\cdot), \Lambda_{i,i-1}^1, \Lambda_{i,i-1}^2, K_{Z_i})$, with $\{\Lambda_{i,i-1}^1, \Lambda_{i,i-1}^2\}$ deterministic matrices, is the strategy of the input process, that controls the output process B_i, and unless this ensures information stability or asymptotic stationarity and ergodicity of the output process, then the CC capacity does not exist, as shown in [2], [3], for $D_{i,i-1} = 0, F_{i-1} = 0, i = 0, \ldots, n$.

Special Case. If $C_{i,i-1} = 0, D_{i,i-1} = 0, Q_{i,i-1} = 0$, and the noise is described by a unit memory autoregressive AR(1) model, i.e., $T = 1$, then (13) degenerates to

$$A_i = \Lambda_{i,i-1}^2 \left(V_{i-1} - \mathbf{E}_{\overline{s}}\left\{V_{i-1}\middle|B_0^{i-1}\right\}\right) + Z_i. \tag{23}$$

i.e., it consists of estimation and innovation parts.

B. Related Literature in Information Theory

Our main results (a)-(c), extend the most general framework found in *information theory* literature, specifically the Cover and Pombra [6] characterization of the $n-$FTFI capacity of the scalar-valued, additive Gaussian noise (AGN) channel, defined by

$$B_i = A_i + V_i, \qquad i = 0, \ldots, n, \tag{24}$$
$$\frac{1}{n+1}\mathbf{E}\Big\{\sum_{i=0}^n |A_i|^2\Big\} \leq \kappa, \quad V_0^n \sim N(0, K_{V_0^n}) \tag{25}$$

where the distribution of the Gaussian noise is $\mathbf{P}_{V_0^n}$, i.e., not of finite memory. The computation of the $n-$FTFI capacity, remained to this date an open problem. Special cases of the scalar AGN channel (24), (25), when the noise is finite memory and stable, are analyzed in [7], [8], using frequency domain methods.

II. Characterization of $n-$FTFI CC Capacity

The main results of this section are, the converse CC Theorem 2.5, that identifies an upper bound on any achievable CC rate, and Theorem 2.6 that establishes a tighter bound, by determining the information structures of input distributions.

We start with the basic notation. A sequence measurable spaces is denoted by $\{(\mathbb{X}_i, \mathcal{B}(\mathbb{X}_i)) : i \in \mathbb{Z}\}$ while their product measurable space is denoted by $(\mathbb{X}^{\mathbb{Z}}, \mathcal{B}(\mathbb{X}^{\mathbb{Z}}))$. The probability distribution of a RV $X : (\Omega, \mathcal{F}) \longmapsto (\mathbb{X}, \mathcal{B}(\mathbb{X}))$ is denoted[1] by $\mathbb{P}\{X(\omega) \in \cdot\} = \mathbf{P}(\cdot) \equiv \mathbf{P}_X(\cdot)$. The conditional distribution of RV Y given $X = x$ is denoted by $\mathbf{P}_{Y|X}(dy|X = x) \equiv \mathbf{P}_{Y|X}(dy|x)$.

A. Control-Coding Capacity

By Definition 1.1, (a.ii), then

$$\mathbb{P}\{B_i \in db\Big|B^{i-1} = b^{i-1}, A^i = a^i\}$$
$$= Q_i(db|b_{i-M}^{i-1}, a_{i-L}^i, v_{i-T}^{i-1}), \quad i = 1, \ldots, n \tag{26}$$

and for $i = 0$, $Q_0(db|a_0, \overline{s})$, $\overline{s} \stackrel{\triangle}{=} (a_{-L}^{-1}, b_{-M}^{-1}, v_{-T}^{-1})$.

Next, we introduce the operational definition of CC capacity of the DM that explains the term "control-coding".

[1] The subscript on X is often omitted.

Definition 2.1: (CC Capacity of DM$\{M, L, T\}$)

For DM$\{M, L, T\}$, the operational definition for reliable communication or information signalling, and control performance, with noiseless feedback is a sequence of controller-encoder and decoder strategies, denoted by $\{(n, \mathcal{M}^{(n)}, \epsilon_n, \overline{s}, \kappa) : n = 0, 1, \ldots\}$ that consists of the following items.

(a) A set of uniformly distributed messages $X^{(n)}$ with alphabet space $\mathcal{M}^{(n)} \triangleq \{1, \ldots, M^{(n)}\}$, known to both the controller-encoder and decoder.

(b) A set of controller-encoder strategies mapping messages, the state $\overline{s}$, and noiseless feedback into inputs, defined by[2]

$$\mathcal{E}_{[0,n]}(\kappa) \triangleq \Big\{g_i : \mathcal{M}^{(n)} \times \mathbb{S} \times \mathbb{A}_0^{i-1} \times \mathbb{B}_0^{i-1} \longmapsto \mathbb{A}_i,$$
$$a_0 = g_0(w, \overline{s}), \ldots, a_n = g_n(w, \overline{s}, a_0^{n-1}, b_0^{n-1}), \quad w \in \mathcal{M}_n :$$
$$\frac{1}{n+1}\mathbf{E}_{\overline{s}}^{g}\Big\{\sum_{i=0}^{n} \gamma_i(V_{i-T}^i, A_{i-L}^i, B_{i-M}^i) \leq \kappa\Big\} \tag{27}$$

i.e, satisfying a cost constraint. The information structure of the controller-encoder is $\mathcal{I}_i^e \triangleq \{W, \overline{S}, A_0^{i-1}, B_0^{i-1}\}, A_i = g_i(\mathcal{I}_i^e)$.

(c) A decoder measurable mapping $d_{0,n}(\overline{s}, \cdot) : \mathbb{B}_0^n \longmapsto \mathcal{M}^{(n)}$ with average probability of decoding error

$$P_{error}^{(n)}(\overline{s}) \triangleq \frac{1}{M^{(n)}} \sum_{w \in \mathcal{M}^{(n)}} \mathbf{P}^g\Big\{d_{0,n}(\overline{S}, B_0^n) \neq w | W = w,$$
$$\overline{S} = \overline{s}\Big\} \equiv \mathbf{P}_{\overline{s}}^g\Big\{d_{0,n}(\overline{S}, B_0^n) \neq W\Big\} \leq \epsilon_n, \ \epsilon_n \in [0, 1]. \tag{28}$$

The CC rate is defined by $r_n \triangleq \frac{1}{n+1}\log M^{(n)}$.

A rate R is called an *achievable rate*, if there exists a controller-encoder and decoder sequence satisfying $\lim_{n \longrightarrow \infty} \epsilon_n = 0$ and $\liminf_{n \longrightarrow \infty} \frac{1}{n+1}\log M^{(n)} \geq R$. The operational definition of the *CC capacity* of the DM$\{M, L, T\}$ is $C(\kappa) \triangleq \sup\{R : R \text{ is achievable}\}$.

In general, $C(\kappa) \equiv C(\kappa, \overline{s})$ depends on initial state $\overline{S} = \overline{s}$.

B. Converse and Direct CC Theorems for DM$\{M, L, T\}$

We shall identify an upper bound on $C(\kappa)$, via the converse CC Theorem 2.6, using two lemmas.

Lemma 2.2: (Mutual information)

Consider the DM$\{M, L, T\}$. Given a strategy $g(\cdot) \in \mathcal{E}_{[0,n]}(\kappa)$, define the mutual information between messages $W \in \mathcal{M}^{(n)}$ to B_0^n conditioned on $\overline{S} = \overline{s}$ by

$$I^g(W; B_0^n|\overline{s}) \triangleq \mathbf{E}_{\overline{s}}^g\Big\{\log\Big(\frac{d\mathbf{P}_{B_0^n|W,\overline{S}}^g(\cdot|\overline{S}, W)}{\mathbf{P}_{B_0^n|\overline{S}}^g(\cdot|\overline{S})}(B_0^n)\Big)\Big\} \in [0, \infty].$$

Further, define $\overline{\mathcal{I}}_i^{\overline{e}} \triangleq \{W, \overline{S}, A_0^{i-1}, V_0^{i-1}, B_0^{i-1}\}, i = 0, \ldots, n$, and the set of controller-encoder strategies

$$\overline{\mathcal{E}}_{[0,n]}(\kappa) \triangleq \Big\{A_i = \overline{g}_i(\overline{\mathcal{I}}_i^{\overline{e}}), i = 0, \ldots, n :$$
$$\frac{1}{n+1}\mathbf{E}_{\overline{s}}^{\overline{g}}\Big(\sum_{i=0}^{n} \gamma_i(V_{i-T}^i, A_{i-L}^i, B_{i-M}^i) \leq \kappa\Big)\Big\}. \tag{29}$$

[2]The subscript/superscript on expectation operator $\mathbf{E}_{\overline{s}}^g$ indicates that the corresponding distribution $\mathbf{P} = \mathbf{P}_{\overline{s}}^g$ depends the encoding strategy g, and that $\overline{S} = \overline{s}$ is fixed.

Then the following hold.

$$\overline{\mathcal{E}}_{[0,n]}(\kappa) = \mathcal{E}_{[0,n]}(\kappa), \quad \text{and} \tag{30}$$
$$I^g(W; B_0^n|\overline{s}) =$$
$$\mathbf{E}_{\overline{s}}^{\overline{g}}\Big\{\sum_{i=0}^{n}\log\Big(\frac{dQ_i(\cdot|B_{i-M}^{i-1}, \{\overline{g}_j(\overline{\mathcal{I}}_j^{\overline{e}})\}_{j=i-L}^{i}, V_{i-T}^{i-1})}{\mathbf{P}^{\overline{g}}(\cdot|B_0^{i-1}, \overline{S})}(B_i)\Big)\Big\}$$
$$\equiv \sum_{i=0}^{n} I^{\overline{g}}(A_{i-L}^i, V_{i-T}^{i-1}; B_0|B_0^{i-1}, \overline{s}). \tag{31}$$

Proof: Follows from the chain rule of mutual information, and the invertibility condition of Definition 1.1, (a.ii). ■

$n-$**Finite Time Feedback Information CC Capacity.** Next, we derive the analog of Lemma 2.2, for directed information $I(A_0^n \rightarrow B_0^n|\overline{s})$, for the larger class of randomized strategies

$$\mathcal{P}_{[0,n]}(\kappa) \triangleq \Big\{P_i(da_i|a_0^{i-1}, b_0^{i-1}, \overline{s}), i = 0, \ldots, n :$$
$$\frac{1}{n+1}\mathbf{E}_{\overline{s}}^{P}\Big(\sum_{i=0}^{n} \gamma_i(V_{i-T}^i, A_{i-L}^i, B_{i-M}^i) \leq \kappa\Big)\Big\}. \tag{32}$$

Lemma 2.3: (Directed information)

Consider the DM$\{M, L, T\}$ and randomized strategies from $\mathcal{P}_{[0,n]}(\kappa)$. Given a $P(\cdot|\cdot) \in \mathcal{P}_{[0,n]}(\kappa)$, define the directed information from A_0^n to B_0^n conditioned on $\overline{S} = \overline{s}$ by

$$I^P(A_0^n \rightarrow B_0^n|\overline{s})$$
$$\triangleq \mathbf{E}_{\overline{s}}^P\Big\{\sum_{i=0}^{n}\log\Big(\frac{d\mathbf{P}_{B_i|B_0^{i-1}, A_0^i, \overline{S}}(\cdot|B_0^i, A_0^i, \overline{S})}{\mathbf{P}_{B_i|B_0^{i-1}, \overline{S}}^P(\cdot|B_0^i, \overline{S})}(B_i)\Big)\Big\}.$$

Further, define the randomized strategies

$$\overline{\mathcal{P}}_{[0,n]}(\kappa) \triangleq \Big\{\overline{P}_i(da_i|a_0^{i-1}, v_0^{i-1}, b_0^{i-1}, \overline{s}), i = 0, \ldots, n :$$
$$\frac{1}{n+1}\mathbf{E}_{\overline{s}}^{\overline{P}}\Big(\sum_{i=0}^{n} \gamma_i(V_{i-T}^i, A_{i-L}^i, B_{i-M}^i) \leq \kappa\Big)\Big\} \tag{33}$$

and directed information from (A_0^n, V_0^{n-1}) to B_0^n for $\overline{S} = \overline{s}$:

$$I^{\overline{P}}(A_0^n, V_0^{n-1} \rightarrow B_0^n|\overline{s}) \triangleq \sum_{i=0}^{n} I^{\overline{P}}(A_{i-L}^i, V_{i-T}^{i-1}; B_0|B_0^{i-1}, \overline{s}).$$

Then the following equality holds.

$$\overline{\mathcal{P}}_{[0,n]}(\kappa) = \mathcal{P}_{[0,n]}(\kappa), \quad \text{and} \tag{34}$$
$$I^P(A_0^n \rightarrow B_0^n|\overline{s}) = I^{\overline{P}}(A_0^n \rightarrow B_0^n|\overline{s}) \tag{35}$$
$$= \mathbf{E}_{\overline{s}}^{\overline{P}}\Big\{\sum_{i=0}^{n}\log\Big(\frac{dQ_i(\cdot|B_{i-M}^{i-1}, A_{i-L}^i, V_{i-T}^{i-1})}{\mathbf{P}^{\overline{P}}(\cdot|B_0^{i-1}, \overline{S})}(B_i)\Big)\Big\} \tag{36}$$
$$= I^{\overline{P}}(A_0^n, V_0^{n-1} \rightarrow B_0^n|\overline{s}). \tag{37}$$

Proof: The derivation is similar to Lemma 2.2. ■

In view of the above lemma, we define the $n-$FTFI CC capacity that will play a role in the identification of $C(\kappa)$.

Definition 2.4: ($n-$FTFI CC Capacity)

The $n-$FTFI CC capacity for DM$\{M, L, T\}$ is defined by

$$C_{A_0^n, V_0^{n-1} \rightarrow B_0^n|\overline{s}}(\kappa)$$
$$\triangleq \sup_{\overline{\mathcal{P}}_{[0,n]}(\kappa)} \sum_{i=0}^{n} I^{\overline{P}}(A_{i-L}^i, V_{i-T}^{i-1}; B_i|B_0^{i-1}, \overline{s}). \tag{38}$$

Theorem 2.5: (Converse and direct CC theorem) Consider the DM$\{M, L, T\}$.
(a) Converse CC Theorem. If there exists a sequence of controller-encoder-decoder strategies $\{(n, \mathcal{M}^{(n)}, \epsilon_n, \overline{s}, \kappa) : n = 0, 1, \dots\}$ given in Definition 2.1, then[3]

$$R \leq \liminf_{n\longrightarrow\infty} \frac{1}{n+1} \log M^{(n)} \tag{39}$$

$$\leq \liminf_{n\longrightarrow\infty} \sup_{\overline{\mathcal{E}}_{[0,n]}(\kappa)} \frac{1}{n+1} \sum_{i=0}^{n} I^{\overline{g}}(A_{i-L}^{i}, V_{i-T}^{i-1}; B_i | B_0^{i-1}, \overline{s})$$

$$\leq \liminf_{n\longrightarrow\infty} \frac{1}{n+1} C_{A_0^n, V_0^{n-1} \rightarrow B_0^n | \overline{s}}(\kappa) \tag{40}$$

provided the following conditions hold.
(i) The supremum of $\sum_{i=0}^{n} I^{\overline{P}}(A_{i-L}^{i}, V_{i-T}^{i-1}; B_i | B_0^{i-1}, \overline{s})$ over $\overline{\mathcal{P}}_{[0,n]}(\kappa)$ in (40) for any finite n in achieved in the set (i.e., the maximizing distribution exists).
(ii) The $\liminf_{n\longrightarrow\infty}$ in (40) is finite.
(b) Direct CC Theorem. Under the assumptions of Theorem 5.2 in [2] (with slight variation) then the CC capacity is $C(\kappa) = \liminf_{n\longrightarrow\infty} \frac{1}{n+1} C_{A_0^n, V_0^{n-1} \rightarrow B_0^n | \overline{s}}(\kappa)$.

Proof: (a) Suppose R is achievable, and hence $\lim_{n\longrightarrow\infty} \epsilon_n = 0$, $\liminf_{n\longrightarrow\infty} \frac{1}{n+1} \log M^{(n)} \geq R$. Then, for each n, since $W = w \in \mathcal{M}^{(n)}$ is uniformly distributed, by invoking Fano's inequality [9], the following hold.

$$\begin{aligned} \log M^{(n)} &= H(W|\overline{s}) = H^g(W|B_0^n, \overline{s}) \\ &+ I^g(W; B_0^n|\overline{s}), \quad \forall g(\cdot,\cdot) \in \mathcal{E}_{[0,n]}(\kappa) \\ &\leq h(\varepsilon_n) + \varepsilon_n \log M_n + I^g(W; B_0^n|\overline{s}) \end{aligned} \tag{41}$$

where $h(z) \triangleq -z \log z - (1-z) \log(1-z), z \in [0, 1]$. By Lemma 2.2 and Lemma 2.3, and standard arguments we obtain the claims. (b) This follows from Theorem 5.2 in [2]. ■

C. Information Structures of $n-$FTFI CC Capacity

Now, we give the information structures of optimal distributions for the $n-$FTFI CC capacity of Definition 2.4.

Theorem 2.6: (Information structures of DM$\{M, L, T\}$) Consider the DM$\{M, L, T\}$. Then we have the following.
(a) The optimal distribution of problem $C_{A_0^n, V_0^{n-1} \rightarrow B_0^n | \overline{s}}(\kappa)$ defined by (38) satisfies conditional independence

$$\overline{P}_i(da_i | a_0^{i-1}, v_0^{i-1}, b_0^{i-1}, \overline{s}) = \pi_i(da_i | a_{i-L}^{i-1}, v_{i-T}^{i-1}, b_{-M}^{i-1})$$

and the charactrization of the $n-$FTFI CC capacity is

$$C_{A_0^n, V_0^{n-1} \rightarrow B_0^n | \overline{s}}^{M,L,T}(\kappa) \triangleq \sup_{\overline{\mathcal{P}}_{[0,n]}^{M,L,T}(\kappa)} \mathbf{E}_{\overline{s}}^{\pi} \Big\{ \sum_{i=0}^{n} \log \Big(\frac{dQ_i(\cdot | B_{i-M}^{i-1}, A_{i-L}^{i}, V_{i-T}^{i-1})}{\mathbf{P}^{\pi}(\cdot | B_0^{i-1}, \overline{S})}(B_i) \Big) \Big\} \tag{42}$$

$$\overline{\mathcal{P}}_{[0,n]}^{M,L,T}(\kappa) \triangleq \Big\{ \pi_i(da_i | a_{i-L}^{i-1}, v_{i-T}^{i-1}, b_{-M}^{i-1}), i = 0, \dots, n : \frac{1}{n+1} \mathbf{E}_{\overline{s}}^{\pi} \Big(\sum_{i=0}^{n} \gamma_i(V_{i-T}^{i}, A_{i-L}^{i}, B_{i-M}^{i}) \Big) \leq \kappa \Big\} \tag{43}$$

[3]The superscript notation $I^{\overline{g}}(\cdot;\cdot|\cdot)$ indicates that the distributions depend on encoding strategies from the set $\overline{\mathcal{E}}_{[0,n]}(\kappa)$.

where $\overline{s} = (b_{-M}^{-1}, a_{-L}^{-1}, v_{-T}^{-1})$ is the initial state.

Proof: Follows by applying the tools from [2]. ■

By Theorem 2.6, for each i the controller-encoder observes the sequence $\{A_{i-L}^{i-1}, V_{i-T}^{i-1}, B_{i-M}^{i-1}\}$ and the initial state $\overline{s}$, while the decoder observes B_{i-M}^{i-1} and $\overline{s}$. Hence, $\{A_{i-L}^{i-1}, V_{i-T}^{i-1}\}$ needs to be estimated at the decoder using the á posteriori distribution $\mathbf{P}_{A_{i-L}^{i-1}, V_{i-T}^{i-1} | B_0^{i-1}, \overline{s}}$.

III. LQG-DM Driven by Correlated Noise

Consider the LQG-DM$\{M = 1, L = 1, T = 1\}$ defined by (5)-(11). By Theorem 2.6, the optimal input distribution is $\pi_i(da_i | a_{i-1}, v_{i-1}, b_{-1}^{i-1}), i = 0, \dots, n$. Moreover, the characterization of the $n-$FTFI CC capacity reduces to

$$C_{A_0^n, V_0^{n-1} \rightarrow B_0^n | \overline{s}}^{1,1,1}(\kappa) = \sup_{\overline{\mathcal{P}}_{[0,n]}^{1,1,1}(\kappa)} \sum_{i=0}^{n} I(A_{i-1}^{i}, V_{i-1}; B_i | B_0^{i-1}, \overline{s})$$

$$= \sup_{\overline{\mathcal{P}}_{[0,n]}^{1,1,1}(\kappa)} H(B_0^n | \overline{s}) - H(V_0^n | v_{-1}) \tag{44}$$

with $\gamma = \overline{\gamma}$. The material of this section are generalizations of those found in [3] for DMs with $D_{i,i-1} = 0, i = 0, \dots, n$ and $V_i, i = 0, \dots, n$ an independent Gaussian process.

(a) Gaussian Maximizing Distribution and Orthogonal Decomposition. By the maximum entropy principle of Gaussian distributions, then the supremum in (44) is confined to the orthogonal realization of a Gaussian process A^n, given by (12)-(20), and the power constraint is,

$$\overline{\mathcal{P}}_{[0,n]}^{1,1,1}(\kappa) \triangleq \Big\{ \Big(\overline{e}(\cdot), \Lambda_{i,i-1}, K_{Z_i} \succeq 0 \Big), i = 0, \dots, n : \frac{1}{n+1} \mathbf{E}_{\overline{s}}^{e} \Big(\sum_{i=0}^{n} \overline{\gamma}_i(A_i, B_{i-1}) \Big) \leq \kappa \Big\}. \tag{45}$$

where $\{(B_i, S_i) : i = 0, \dots, n\}$ satisfy the recursions

$$S_i = H_{i,i-1} S_{i-1} + G_{i,i-1} U_i + M_{i,i-1}^1 Z_i + M_{i,i-1}^2 W_i, \quad S_{-1} = s_{-1}, \quad i = 0, \dots, n, \tag{46}$$

$$B_i = C_{i,i-1} B_{i-1} + D_{i,i} U_i + \overline{\Lambda}_{i,i-1} S_{i-1} + D_{i,i} Z_i + W_i, \tag{47}$$

$$H_{i,i-1} \triangleq \begin{bmatrix} \Lambda_{i,i-1}^1 & \Lambda_{i,i-1}^2 \\ 0 & F_{i-1} \end{bmatrix}, \; G_{i,i-1} \triangleq \begin{bmatrix} I \\ 0 \end{bmatrix},$$

$$M_{i,i-1}^1 \triangleq \begin{bmatrix} I \\ 0 \end{bmatrix}, M_{i,i-1}^2 \triangleq \begin{bmatrix} 0 \\ I \end{bmatrix},$$

$$\overline{\Lambda}_{i,i-1} \triangleq D_{i,i} \Lambda_{i,i-1} + \overline{D}_{i,i-1}.$$

(c) Characterization of FTFI Capacity. By (44) to compute $H(B_0^n | \overline{s}) = \sum_{i=0}^{n} H(B_i | B_0^{i-1}, \overline{s})$, we use

$$\widehat{B}_{i|i-1} \triangleq \mathbf{E}_{\overline{s}}^{e} \Big\{ B_i \Big| B_0^{i-1} \Big\}, \quad \widehat{S}_{i|i} \triangleq \mathbf{E}_{\overline{s}}^{e} \Big\{ S_i \Big| B_0^{i} \Big\},$$

$$\widehat{S}_{i|i-1} \triangleq \mathbf{E}_{\overline{s}}^{e} \Big\{ S_i \Big| B_0^{i-1} \Big\},$$

$$K_{B_i | B_0^{i-1}} \triangleq \mathbf{E}_{\overline{s}}^{e} \Big\{ \Big(B_i - \widehat{B}_{i|i-1} \Big) \Big(B_i - \widehat{B}_{i|i-1} \Big)^T \Big| B_0^{i-1} \Big\},$$

$$P_{i|i} = \mathbf{E}_{\overline{s}}^{e} \Big\{ \Big(S_i - \widehat{S}_{i|i} \Big) \Big(S_i - \widehat{S}_{i|i} \Big)^T \Big| B_0^{i} \Big\}, \quad i = 0, \dots, n.$$

By (46), (47) the Kalman-filter recursions [5] are:

$$\begin{aligned}\widehat{S}_{i|i} &= H_{i,i-1}\widehat{S}_{i-1|i-1} + G_{i,i-1}U_i \\ &\quad + \Delta_{i|i-1}\nu_i^e, \qquad \widehat{S}_{-1|-1} = \overline{s}, \quad i=0,\ldots,n, \quad (48)\\ \widehat{B}_{i|i-1} &= C_{i,i-1}B_{i-1} + D_{i,i}U_i + \overline{\Lambda}_{i,i-1}\widehat{S}_{i-1|i-1}, \quad (49)\\ \nu_i^e &\triangleq B_i - \widehat{B}_{i|i-1} = \overline{\Lambda}_{i,i-1}\Big(S_{i-1} - \widehat{S}_{i-1|i-1}\Big)\\ &\quad + D_{i,i}Z_i + W_i, \quad (50)\\ K_{B_i|B_0^{i-1}} &= \overline{\Lambda}_{i,i-1}P_{i-1|i-1}\overline{\Lambda}_{i,i-1}^T + D_{i,i}K_{Z_i}D_{i,i}^T + K_{W_i}\end{aligned}$$

and $P_{i|i}$ solves the matrix Riccati difference equation (RDE)

$$\begin{aligned}&P_{i|i} = H_{i,i-1}P_{i-1|i-1}H_{i,i-1}^T + M_{i,i-1}^1 K_{Z_i} M_{i,i-1}^{1,T}\\ &+ M_{i,i-1}^2 K_{W_i} M_{i,i-1}^{2,T} - \Big(M_{i,i-1}^1 K_{Z_i} D_{i,i}^T + M_{i,i-1}^2 K_{W_i}\\ &+ H_{i,i-1}P_{i-1|i-1}\overline{\Lambda}_{i,i-1}^T\Big)\Phi_{i|i-1}\\ &.\Big(M_{i,i-1}^1 K_{Z_i} D_{i,i}^T + M_{i,i-1}^2 K_{W_i} + H_{i,i-1}P_{i-1|i-1}\overline{\Lambda}_{i,i-1}^T\Big)^T,\\ &\Phi_{i|i-1} \triangleq \Big[D_{i,i}K_{Z_i}D_{i,i}^T + K_{W_i} + \overline{\Lambda}_{i,i-1}P_{i-1|i-1}\overline{\Lambda}_{i,i-1}^T\Big]^{-1},\\ &\Delta_{i|i-1} \triangleq \Big(M_{i,i-1}^1 K_{Z_i} D_{i,i}^T + M_{i,i-1}^2 K_{W_i}\\ &+ H_{i,i-1}P_{i-1|i-1}\overline{\Lambda}_{i,i-1}^T\Big)\Phi_{i|i-1}. \quad (51)\end{aligned}$$

The innovations process, ν_i^e, of the controlled process B_0^n, in (48) is an orthogonal Gaussian process, and independent of the strategy $\overline{e}(\cdot)$ (by (13), (48)), and

$$\nu_i^e \triangleq B_i - \widehat{B}_{i|i-1} = \nu_i^e\Big|_{\overline{e}_i=0} \equiv \nu_i^0, \quad (52)$$

$$\nu_i^0 \sim N(0, K_{\nu_i^0}), \quad K_{\nu_i^0} = K_{B_i|B_0^{i-1}}, \quad i=0,\ldots,n \quad (53)$$

$$\mathbf{P}_{\overline{s}}^e(B_i \leq b_i|B_0^{i-1}) \sim N(\widehat{B}_{i|i-1}, K_{\nu^0}), i=0,\ldots,n. \quad (54)$$

From the above and (44) we obtain, for $i=0,\ldots,n$,

$$I(A_{i-1}^i, V_{i-1}; B_i|B_0^{i-1}, \overline{s}) = \log\frac{|K_{B_i|B_0^{i-1}}|}{|K_{W_i}|}. \quad (55)$$

The average cost is given by

$$\begin{aligned}&\frac{1}{n+1}\sum_{i=0}^n \mathbf{E}_{\overline{s}}^e\Big\{\overline{\gamma}_i(A_i, B_{i-1})\Big\} = \frac{1}{n+1}\sum_{i=0}^n \mathbf{E}_{\overline{s}}^e\Big\{\langle U_i, R_{i,i}U_i\rangle\\ &+ 2\langle\Lambda_{i,i-1}\widehat{S}_{i-1|i-1}, R_{i,i}U_i\rangle\\ &+ \langle\Lambda_{i,i-1}\widehat{S}_{i-1|i-1}, R_{i,i}\Lambda_{i,i-1}\widehat{S}_{i-1|i-1}\rangle + tr\Big(K_{Z_i}R_{i,i}\Big)\\ &+ tr\Big(\Lambda_{i,i-1}^T R_i\Lambda_{i,i-1}P_{i-1|i-1}\Big) + \langle B_{i-1}, Q_{i,i-1}B_{i-1}\rangle\Big\}\\ &\equiv \frac{1}{n+1}\sum_{i=0}^n \mathbf{E}_{\overline{s}}^e\Big\{\hat{\gamma}_i(U_i, \widehat{S}_{i-1|i-1}, B_{i-1}, \Lambda_{i,i-1}, K_{Z_i})\Big\}.\end{aligned}$$

Hence, the characterization of FTFI CC capacity is given by

$$\begin{aligned}&C_{A_0^n\rightarrow B_0^n|\overline{s}}^{1,1,1}(\kappa) = \frac{1}{2}\sup_{\overline{\mathcal{P}}_{[0,n]}^{1,1,1}(\kappa)}\Big\{ \quad (56)\\ &\sum_{i=0}^n \log\frac{|\overline{\Lambda}_{i,i-1}P_{i-1|i-1}\overline{\Lambda}_{i,i-1}^T + D_{i,i}K_{Z_i}D_{i,i}^T + K_{W_i}|}{|K_{W_i}|}\Big\}\\ &\overline{\mathcal{P}}_{[0,n]}^{1,1,1}(\kappa) \triangleq \Big\{e_i(\cdot) \triangleq \big(\overline{e}_i(\cdot,\cdot), \Lambda_{i,i-1}, K_{Z_i}\big), i=0,\ldots,n:\\ &\frac{1}{n+1}\sum_{i=0}^n \mathbf{E}_{\overline{s}}^e\Big(\overline{\gamma}_i(A_i, B_{i-1})\Big) \leq \kappa\Big\}. \quad (57)\end{aligned}$$

For problem (56), we prove a decentralized separation principle between the computation of the strategies $\{\overline{e}_i(\cdot) : i = 0,\ldots,n\}$ and $\{(\Lambda_{i,i-1}, K_{Z_i}) : i=0,\ldots,n\}$: it states that problem (56) is equivalent to a *decentralized optimization problem, with multiple decision makers (DMs), each having access to different information* [10].

Theorem 3.1: (Decentralized separation principle)
Consider the LQG-DM$\{M=1, L=1, T=1\}$ of Definition 1.2. Then the following hold.
(a) *The Cost-Rate*, i.e., dual of $C_{0,n}(\kappa) \equiv C_{A_0^n\rightarrow B_0^n|\overline{s}}^{1,1,1}(\kappa)$, is

$$\kappa_{0,n}(C) \triangleq \inf_{\big(\overline{e}_i(\cdot), \Lambda_{i,i-1}, K_{Z_i}\succeq 0\big), i=0,\ldots,n} \mathbf{E}_{\overline{s}}^e\Big\{\sum_{i=0}^n \overline{\gamma}_i(A_i, B_{i-1})\Big\}$$

$$\text{such that} \quad \frac{1}{2}\sum_{i=0}^n \log\frac{|K_{B_i|B_0^{i-1}}|}{|K_{W_i}|} \geq (n+1)C. \quad (58)$$

(b) *Decentralized Separation Principle.* Let $\{e^*(\cdot) \equiv (\overline{e}_i^*(\cdot), \Lambda_{i,i-1}^*, K_{Z_i}^*) : i=0,\ldots,n\}$ denote then optimal strategy of $C_{0,n}(\kappa) \equiv C_{A_0^n\rightarrow B_0^n|\overline{s}}(\kappa)$, defined by (56).
Then the following decentralized separation principle holds.
(i) For a fixed $\{(\Lambda_{i,i-1}, K_{Z_i}) : i=0,\ldots,n\}$, determine the optimal strategy $\{\overline{e}_i^*(\cdot, \Lambda, K_Z) : i=0,\ldots,n\}$ from the solution of the stochastic optimization problem

$$J_{0,n}^{SC}(\Lambda, K_Z, \overline{e}^*) \triangleq \inf_{\overline{e}_i(\cdot): i=0,\ldots,n} \mathbf{E}_{\overline{s}}^e\Big\{\sum_{i=0}^n \overline{\gamma}_i(A_i, B_{i-1})\Big\}. \quad (59)$$

(ii) Determine the optimal strategy $\{(\Lambda_{i,i-1}^*, K_{Z_i}^*) : i = 0,\ldots,n\}$ from (56), when $\{\overline{e}_i(\cdot) = \overline{e}_i^*(\cdot) : i=0,\ldots,n\}$.
(c) *Optimal Decentralized Strategies.* Define an augmented state variable as follows.

$$\overline{B}_{i-1} \triangleq \begin{bmatrix} B_{i-1} \\ \widehat{S}_{i-1|i-1}\end{bmatrix}, \; i=1,\ldots,n. \quad (60)$$

$\{\overline{e}_i(\cdot, \Lambda, K_Z) : i=0,\ldots,n\}$ of (59) is of the form

$$\begin{aligned}\overline{e}_i(B_0^{i-1}, \Lambda, K_Z, \overline{s}) &= \Gamma_{i,i-1}^1 B_{i-1} + \Gamma_{i,i-1}^2\widehat{S}_{i-1|i-1}, \quad (61)\\ &= \overline{\Gamma}_{i,i-1}\overline{B}_{i-1}, \; i=0,\ldots,n. \quad (62)\end{aligned}$$

where the components of $\{\overline{B}_i : i=0,\ldots,n\}$ satisfy (48), (53), and the augmented system is

$$\overline{B}_i = \overline{F}_{i,i-1}\overline{B}_{i-1} + \overline{E}_{i,i-1}U_i + \overline{G}_{i,i-1}\nu_i^e, \quad (63)$$

$$\overline{F}_{i,i-1} \triangleq \begin{bmatrix} C_{i,i-1} & \overline{\Lambda}_{i,i-1}\\ 0 & H_{i,i-1}\end{bmatrix}, \; \overline{E}_{i,i-1} \triangleq \begin{bmatrix} D_{i,i}\\ G_{i,i-1}\end{bmatrix} \quad (64)$$

$$\overline{G}_{i,i-1} \triangleq \begin{bmatrix} I \\ \Delta_{i|i-1}\end{bmatrix}. \quad (65)$$

The average cost is of quadratic form, given by

$$\mathbf{E}_{\overline{s}}^{e}\Big\{\sum_{i=0}^{n}\overline{\gamma}_i(A_i,B_{i-1})\Big\} \equiv \mathbf{E}_{\overline{s}}^{e}\Big\{\sum_{i=0}^{n}\hat{\gamma}_i(U_i,\overline{B}_{i-1},\Lambda_{i,i-1},K_{Z_i})\Big\}$$

$$\triangleq \mathbf{E}_{\overline{b}_{-1}}^{e}\Big\{\sum_{i=0}^{n}\Big(\begin{bmatrix}\overline{B}_{i-1}\\ U_i\end{bmatrix}^T\begin{bmatrix}\overline{M}_{i,i-1} & \overline{L}_{i,i-1}\\ \overline{L}_{i,i-1}^T & \overline{N}_{i,i-1}\end{bmatrix}\begin{bmatrix}\overline{B}_{i-1}\\ U_i\end{bmatrix}$$

$$+ tr\big(K_{Z_i}R_{i,i}\big) + tr\big(\Lambda_{i,i-1}^T R_{i,i}\Lambda_{i,i-1}P_{i-1|i-1}\big)\Big)\Big\}, \quad (66)$$

$$\overline{M}_{i,i-1} \triangleq \begin{bmatrix} Q_{i,i-1} & 0\\ 0 & \Lambda_{i,i-1}^T R_{i,i}\Lambda_{i,i-1}\end{bmatrix}, \quad (67)$$

$$\overline{L}_{i,i-1} \triangleq \begin{bmatrix} 0\\ \Lambda_{i,i-1}^T R_{i,i}\end{bmatrix}, \quad \overline{N}_{i,i-1} \triangleq R_{i,i}. \quad (68)$$

Moreover, the following hold.
(i) For a fixed $\{(\Lambda_{i,i-1},K_{Z_i}) : i=0,\ldots,n\}$, the optimal strategy $\{U_i^* = \overline{e}_i^*(B_0^{i-1},\Lambda,K_Z,\overline{s}) \equiv \overline{e}_i^*(\overline{B}_{i-1},\Lambda,K_Z,\overline{s}) : i=0,\ldots,n\}$ is given by the following equations.

$$\overline{e}_i^*(\overline{b}_{i-1},\Lambda,K_Z,\overline{s}) = \overline{\Gamma}_{i,i-1}^*\overline{b}_{i-1} \quad (69)$$

$$\overline{\Gamma}_{i,i-1}^* = -\Big(\overline{N}_{i,i-1}+\overline{E}_{i,i-1}^T\Sigma(i+1)\overline{E}_{i,i-1}\Big)^{-1}\Big(\overline{L}_{i,i-1}^T + \overline{E}_{i,i-1}^T\Sigma(i+1)\overline{F}_{i,i-1}\Big),\ i=0,\ldots,n-1, \quad (70)$$

$$\overline{\Gamma}_{n,n-1}^* = -\overline{N}_{n,n-1}^{-1}\overline{L}_{n,n-1}^T \quad (71)$$

where the symmetric positive semi-definite matrix $\{\Sigma(i) : i=0,\ldots,n\}$ satisfies the matrix difference Riccati equation

$$\Sigma(i)=\overline{F}_{i,i-1}^T\Sigma(i+1)\overline{F}_{i,i-1}-\Big(\overline{F}_{i,i-1}^T\Sigma(i+1)\overline{E}_{i,i-1}+\overline{L}_{i,i-1}\Big) .\Big(\overline{N}_{i,i-1}+\overline{E}_{i,i-1}^T\Sigma(i+1)\overline{E}_{i,i-1}\Big)^{-1}\Big(\overline{F}_{i,i-1}^T\Sigma(i+1)\overline{E}_{i,i-1} +\overline{L}_{i,i-1}\Big)^T+\overline{M}_{i,i-1}^T, \quad i=0,\ldots,n-1, \quad (72)$$

$$\Sigma(n) = diag\{\overline{M}_{n,n-1},0\} \quad (73)$$

and the optimal pay-off is given by

$$J^{SC}(\overline{e}^*(\cdot),\Lambda,K_Z) = \langle\overline{b}_{0|-1},\Sigma(0)\overline{b}_{0|-1}\rangle + \sum_{j=0}^{n}\Big\{tr\big(K_{Z_j}R_{j,j}\big) + tr\big(\Lambda_{j,j-1}^T R_{j,j}\Lambda_{j,j-1}P_{j-1|j-1}\big)\Big\} + \sum_{j=0}^{n-1}tr\Big(K_{\overline{B}_j|\overline{B}_{j-1}}\overline{G}_{j,j-1}^T\Sigma(j+1)\overline{G}_{j,j-1}\Big). \quad (74)$$

(ii) The optimal strategy $\{(\Lambda_{i,i-1}^*,K_{Z_i}^*) : i=0,\ldots,n\}$ is determined from (56), for fixed $\overline{e}(\cdot)=\overline{e}^*(\cdot)$.

Proof: The derivation follows from the structure of the strategies. ■

Remark 3.2: (Decentralized structure of optimal channel input process)

(1) Theorem 3.1 states that a decentralized separation principle holds, and (13) is decomposed into the strategies

$$A_i = \Gamma_{i,i-1}^1 B_{i-1} + \Gamma_{i,i-1}^2\widehat{S}_{i-1|i-1} + \Lambda_{i,i-1}^1 A_{i-1} + \Lambda_{i,i-1}^2 V_{i-1} + Z_i. \quad (75)$$

This separation principle is more general than the corresponding separation of estimation and control, of LQG stochastic optimal control theory [11].

(2) The direct control-coding theorem can be derived, as in [2], Theorem 4.1, using the ergodic theory of Markov decision problems.

Conjecture 3.3: For time-invariant (TI) LQG-DMs, and strategies in (75) restricted to TI, $(\Gamma^1,\Gamma^2,\Lambda^1,\Lambda^2,K_Z),\forall i$, and stable, in the sense that

$$\lim_{n\longleftrightarrow\infty} P_{n|n} = P, \quad \lim_{n\longleftrightarrow\infty}\overline{\Sigma}(n) = \overline{\Sigma} \quad (76)$$

exist and satisfy matrix Riccati algebraic equations, with stabilizing solutions, then the CC capacity is the per unit time limit of (56), evaluated at $\overline{e}(\cdot)=\overline{e}^*$, i.e.,

$$C(\kappa) = \lim_{n\longleftrightarrow\infty}\frac{1}{n+1}C_{A_0^n\rightarrow B_0^n|\overline{s}}^{1,1,1}(\kappa) \quad (77)$$

The conditions of the conjecture are sufficient to ensure the analog of Theorem 4.1 in [2] holds.

IV. Conclusions

The results of the he paper extend Shannon's operational definition of coding capacity to control-coding capacity for general unstable DMs, and provide answers to several open problems in Shannon's information theory, as applied to communication systems.

The analysis of the $n-$FTFI control-coding capacity of the LQG-DM given in Theorem 3.1, illustrates the decentralized separation principle of optimal strategies. Further, optimal strategies include control, estimation, and information transmission/signalling strategies, that interact in a specific order.

References

[1] C. D. Charalambous, C. Kourtellaris, I. Tzortzis, and Loyka, "The capacity of unstable dynamical systems-interaction of control and information transmission," in *IEEE International Symposium on Information Theory Proceedings (ISIT)*, 25-30, June 2017, pp. 2663–2667.
[2] 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," *IEEE Transactions on Information Theory*, vol. 64, no. 7, pp. 4962–4992, July 2018.
[3] C. D. Charalambous, C. Kourtellaris, and S. Loyka, "Capacity achieving distributions and separation principle for feedback channels with memory: The LQG theory of directed information," *IEEE Transactions on Information Theory*, vol. 64, no. 9, pp. 6384–6418, September 2018.
[4] 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*, vol. 62, no. 10, pp. 5010–5025, March 2017.
[5] P. E. Caines, *Linear Stochastic Systems*, ser. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., New York, 1988.
[6] T. Cover and S. Pombra, "Gaussian feedback capacity," *IEEE Transactions on Information Theory*, vol. 35, no. 1, pp. 37–43, Jan. 1989.
[7] Y.-H. Kim, "Feedback capacity of stationary Gaussian channels," *IEEE Transactions on Information Theory*, vol. 56, no. 1, pp. 57–85, 2010.
[8] 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.
[9] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. John Wiley & Sons, Inc., Hoboken, New Jersey, 2006.
[10] C. D. Charalambous and N. U. Ahmed, "Centralized versus decentralized optimization of distributed stochastic differential decision systems with different information structures-Part I: General theory," *IEEE Transactions on Automatic Control*, vol. 62, no. 3, pp. 1194–1209, 2017.
[11] P. R. Kumar and P. Varaiya, *Stochastic Systems: Estimation, Identification, and Adaptive Control*. Prentice Hall, 986.