ETH zürich

ETH Library

The Decentralized Structures of Capacity Achieving Distributions of Channels with Memory and Feedback

Conference Paper

Author(s):
Charalambous, Charalambos D.; Kourtellaris, Christos K.; Tzortzis, Ioannis; Loyka, Sergey

Publication date:
2018-02

Permanent link:
https://doi.org/10.3929/ethz-b-000245067

Rights / license:
In Copyright - Non-Commercial Use Permitted

This page was generated automatically upon download from the ETH Zurich Research Collection.
For more information, please consult the Terms of use.

International Zurich Seminar on Information and Communication (IZS), February 21 – 23, 2018

The Decentralized Structures of Capacity Achieving Distributions of Channels with Memory and Feedback

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

Abstract— **We consider extremum problems of feedback capacity for models with memory, subject to average cost constraints. We show the optimal input process that maximizes directed information consists of two parts, one responsible to control the output process, and one responsible to transmit new information that interact. Unlike [1], the decentralized structure of the optimal input process is demonstrated for Gaussian models with memory on past inputs and outputs. A semi-separation principle is shown that states, the optimal input process is generated from multiple strategies of a decentralized optimization problem, of control and information transmission. Further, it is shown that the derivation of directed information stability is semi-separable, in the sense that it separates into a statement about the ergodic properties of the stochastic optimal control problem with partial information, and a statement related to an information transmission problem.**

I. Introduction

Recently, it is shown that Shannon's coding capacity extends to unstable dynamic systems, irrespectively of whether these are communication channels or control systems [2] (see also [1], [3] for extensive analysis). Shannon's coding capacity is called *control-coding capacity* to emphasize the interaction of control and information transmission parts of the optimal input process, that achieves capacity.

MIMO G-RM. This paper utilizes some of the results found in the above references, to investigate Multiple-Input Multiple-Output (MIMO) Gaussian Recursive Models (G-RMs), with input process $A^n \stackrel{\triangle}{=} \{A_0, A_1, \ldots, A_n\}$ and output process $Y^n \stackrel{\triangle}{=} \{Y_0, Y_1, \ldots, Y_n\}$, described by

$$Y_i = C^{i-1}\, Y^{i-1} + D_{i,i}\, A_i + D_{i,i-1}A_{i-1} + V_i, \qquad (1)$$

$$S \stackrel{\triangle}{=} (Y^{-1}, A_{-1}) = (y^{-1}, a_{-1}) \equiv s,$$

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

$$(Y^{-1}, A_{-1}) \sim N(0, K_{Y^{-1},A_{-1}}),\ K_{Y^{-1},A_{-1}} \succ 0, \qquad (3)$$

$$\frac{1}{n+1}\mathbf{E}\Big\{\sum_{i=0}^{n}\langle A_i, R_i A_i\rangle + \langle Y_{i-1}, Q_{i,i-1}Y_{i-1}\rangle\Big\} \leq \kappa, \qquad (4)$$

$$(D_{i,i}, D_{i,i-1}) \in \mathbb{R}^{p\times q} \times \mathbb{R}^{p\times q}, \qquad (5)$$

$$R_i \in \mathbb{S}_{++}^{q\times q},\quad Q_{i,i-1} \in \mathbb{S}_{+}^{p\times p},\quad i = 0, \ldots, n. \qquad (6)$$

Here S is the initial data, $V_i \sim N(0, K_{V_i}), i = 0, 1, \ldots, n$ denotes zero mean Gaussian process, $\langle \cdot, \cdot \rangle$ 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, $\mathbb{S}_{++}^{q\times q}$ its subset of positive definite matrices, and κ is the power. The initial state $S = s$ is known to the encoder and the decoder.

Main Results. For the extremum problem of maximizing directed information from A^n to B^n given the initial state $S = s$, denoted by $I(A^n \rightarrow B^n|s)$, over conditional distributions $\mathbf{P}_{A_i|A^{i-1},B^{i-1},S}, i = 0, \ldots, n$, that satisfy the average cost constraint, it is shown that a *semi-separation principle* holds with the following consequences.

(a) Part of the optimal input process A^n is characterized by the solution of a stochastic optimal control problem with partial information,

(b) the rest is characterized by the solution of an information transmission problem that interacts with that of the stochastic control part, and

(c) their computation is directly related to the notion of Person-by-Person (PbP) optimality, and team or global optimality in problems of optimal control and games, where two or more strategies do not share the same information, and aim at optimizing a single pay-off.

(d) The derivation of directed information stability is semi-separable, into a statement related to the ergodic properties of the stochastic optimal control problem, and a statement related to an information transmission problem, that interact in a specific order.

The semi-separation principle and its consequences (a)-(d) are attributed to the property that a Gaussian input process $\{A_i = A_i^g : i = 0, \ldots, n\}$ with corresponding Gaussian ouput process $\{Y_i = Y_i^g : i = 0, \ldots, n\}$, maximizes directed information $I(A^n \rightarrow Y^n|s)$ (subject to the average cost constraint), and that such an optimal process is given by the following orthogonal decomposition.

$$A_i^g = \overline{e}_i(Y^{g,i-1}, A_{i-1}^g, Z_i^g),\ i = 0, \ldots, n,\ S = s, \qquad (7)$$

$$= U_i^g + \Lambda_{i,i-1}A_{i-1}^g + Z_i^g,\ \ U_i^g \stackrel{\triangle}{=} \Gamma^{i-1}Y^{g,i-1}, \qquad (8)$$

$$\equiv e_i(Y^{g,i-1}) + \Lambda_{i,i-1}A_{i-1}^g + Z_i^g \qquad (9)$$

where

$$e_i(y^{i-1}) \stackrel{\triangle}{=} \Gamma^{i-1}y^{i-1} \ \text{ is the control strategy,} \qquad (10)$$

$$Z_i^g \ \text{ is independent of } \Big(A^{g,i-1}, Y^{g,i-1}\Big),$$

$$Z^{g,i} \ \text{ is independent of } \ V^i,\ \ i = 0, \ldots, n, \qquad (11)$$

$$Z_i^g \sim N(0, K_{Z_i}) : i = 0, 1, \ldots, n \ \text{ is an independent Gaussian process} \qquad (12)$$

for some deterministic matrices $\{(\Gamma^{i-1}, \Lambda_{i,i-1}) : i = 0, \ldots, n\}$ of appropriate dimensions.

Indeed, the following properties hold.

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.

(P1) The optimal strategies $(e_i^*(\cdot), \Lambda_{i,i-1}^*, K_{Z_i}^*) : i = 0, \ldots, n\}$ are characterized by the solution of a decentralized optimization problem, where $e_i^*(\cdot), i = 0, \ldots, n$ is the solution of a stochastic optimal control problem, for a fixed $(\Lambda_{i,i-1}, K_{Z_i}) : i = 0, \ldots, n\}$, while the optimal $(\Lambda_{i,i-1}^*, K_{Z_i}^*) : i = 0, \ldots, n\}$ is the solution of an information transmission problem, with $e_i(\cdot) = e_i^*(\cdot), i = 0, \ldots, n$.

(P2) The following holds.

$$\text{If } K_{Z_i} = 0, i = 0, \ldots, n \text{ then } I(A^{g,n} \rightarrow Y^{g,n}|s) = 0.$$

(P2) is expected and easily verified, because the initial state $S = s$ is known to the encoder.

(P1) is an application of problems of optimal control and games, where two or more strategies do not share the same information, and aim at optimizing a single pay-off [4].

Special Cases of MIMO G-RM. Before we illustrate that the MIMO G-RM is fundamentally different from past investigations by other authors, we should mention that the MIMO G-RM is an infinite impulse response (IIR) model, and includes the following degenerate cases.

(1) Finite Impulse Response Model. If $C^{i-1} = 0, i = 0, \ldots, n$ then the MIMO G-RM reduces to a finite impulse response (FIR) model.

(2) No Dependence on Past Channel Inputs. If $D_{i,i-1} = 0, i = 0, \ldots, n$ then the MIMO G-RM reduces to the IIR model investigated in [1], [3].

A. Literature on Gaussian Channels with Memory & Feedback

For scalar-valued, Additive Gaussian Noise (AGN) channels with nonstationary and nonergodic noise, described by $Y_i = A_i + V_i$, $\frac{1}{n+1}\mathbf{E}\Big\{\sum_{i=0}^n |A_i|^2\Big\} \leq \kappa$, $\mathbf{P}_{V_i|V^{i-1},A^i} = \mathbf{P}_{V_i|V^{i-1}}, i = 0, \ldots, n$, $V^n \sim N(0, K_{V^n})$, then the feedback capacity is characterized by Cover and Pombra [5], via

$$C_{0,n}^{CP}(\kappa) \triangleq \frac{1}{2n} \max_{(\Gamma^n, K_{Z^n})} \log \frac{|(\Gamma^n + I)K_{V^n}(\Gamma^n + I)^T + K_{Z^n}|}{|K_{V^n}|} \tag{13}$$

$$\text{subject to } \frac{1}{n+1} tr\Big(\Gamma^n K_{V^n}(\Gamma^n)^T + K_{Z^n}\Big) \leq \kappa \tag{14}$$

where Z^n is a Gaussian process $N(0, K_{Z^n})$, orthogonal to V^n, and Γ^n is lower diagonal time-varying matrix with deterministic entries. Note that although, Z^n is called an "innovations process" in [5], this is not an orthogonal process. Note also that if $K_{Z^n} = 0$, since Γ^n is lower diagonal, then $C_{0,n}^{CP}(\kappa) = 0$, as expected. The closed form solution to (13) remains to this date an open problem.

The per unit time limit $C^{CP}(\kappa) \triangleq \lim_{n \longrightarrow \infty} \frac{1}{n+1} C_{0,n}^{CP}(\kappa)$, for the special case of stationary ergodic noise with finite memory, described by a power spectral density $S_V(\omega) = |H(e^{j\omega})|^2$, where the filter $H(\cdot)$ is rational with stable poles and marginally stable zeros, is analyzed in [6] and in [7]. Theorem 7 and Corollary 7.1 in [7] state that capacity is achieved, when the innovations part of the input processes is zero (i.e., eqn(125) in [7] with $e_t = 0, t = 0, \ldots,$).

We should mention that Theorem 3.1 (of our paper) cannot be obtained from [6]–[8], and that the methods applied in [6], [7] are not applicable. Our results are based on a semi-separation principle and its consequences (a)-(d).

II. Feedback Capacity and Decentralized Strategies

In this section we introduce a general channel or control model (CM), and we recall the decentralized structure of the input process, and its control and communication aspects.

Consider a CM model with input process $A^n \triangleq \{A_i : i = 0, 1, \ldots, n\}$, taking values in arbitrary alphabet spaces $\mathbb{A}^n \triangleq \times_{i=0}^n \mathbb{A}_i$, an output process $Y^n \triangleq \{Y_i : i = 0, 1, \ldots, n\}$ taking values in arbitrary alphabet spaces, $\mathbb{Y}^n \triangleq \times_{i=0}^n \mathbb{Y}_i$. The initial data is $S \triangleq (A^{-1}, Y^{-1}) = s \in \mathbb{S} \triangleq \mathbb{A}^{-1} \times \mathbb{Y}^{-1}$.

The channel or control model (CM) is a sequence of conditional distributions

$$\mathbf{P}_{Y_i|Y^{i-1},A^i,S} \equiv Q_i(dy_i|y^{i-1}, a^i, s), \quad i = 0, \ldots, n. \tag{15}$$

The conditional distributions of the input process are chosen from the set

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

The above definition means, the encoder (or controller-encoder to be precise) knows the initial data $s = (y^{-1}, a^{-1})$, and applies noiseless feedback. The conditional distributions of the input process are subject to a cost constraint[1]

$$\mathcal{P}_{[0,n]}(\kappa) \triangleq \Big\{P_i(da_i|a^{i-1}, y^{i-1}, s), i = 0, \ldots, n : \tag{16}$$

$$\frac{1}{n+1}\mathbf{E}_s^P\Big(\ell_{0,n}(A^n, Y^n)\Big) \leq \kappa\Big\} \subset \mathcal{P}_{[0,n]} \tag{17}$$

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

The pay-off 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 [9], [10]

$$I(A^n \rightarrow Y^n|s) \triangleq \sum_{i=0}^n I(A^i; Y_i|Y^{i-1}, s)$$

To connect directed information to the feedback capacity of the CM we introduce the following assumption [10].

Assumption 2.1: (i) If the information process to be encoded is $\{X_i : i = 0, \ldots, k\}$, then the following holds.

$$\mathbf{P}_{Y_i|Y^{i-1},A^i,S,X^k} = \mathbf{P}_{Y_i|Y^{i-1},A^i,S}, \forall k, i = 0, \ldots, n \tag{18}$$

(ii) The initial data $S = s$ is known to the encoder and decoder.

The finite-time horizon (FTH) information capacity (under Assumptions 2.1) is defined by

$$J_{A^n \rightarrow Y^n|s}(P^*, \kappa) \triangleq \sup_{\mathcal{P}_{[0,n]}(\kappa)} I(A^n \rightarrow Y^n|s). \tag{19}$$

[1]The notation $\mathbf{E}_s^P$ indicates the dependence of the joint distribution on elements of $\mathcal{P}_{[0,n]}$ and the initial state $S = s$.

Throughout we assume existence of a maximizing distribution (such conditions are extracted from [11]).

The information capacity is defined by

$$C(\kappa) \triangleq \lim_{n\longrightarrow\infty} \frac{1}{n+1} J_{A^n\rightarrow Y^n|s}(P^*,\kappa) \tag{20}$$

provided the limit exists and it is finite.

Coding Theorems. Recall [12], Appendix A (code definition and achievable rate). By the converse coding theorem [13], a tight upper bound on any achievable rate is $C(\kappa)$. Moreover, if the optimal joint process $\{(A_i, Y_i) : i = 0,\ldots,n\}$ is either asymptotically stationary and ergodic [14], [15], or it induces information stability of the directed information density (see [12], Appendix A), then any code rate below $C(\kappa)$ is achievable. In general, the rate may depend on the initial data $S = s$, i.e., $C(\kappa) = C_s(\kappa)$.

Dualities of Capacity and Stochastic Optimal Control. Let $\mathcal{P}^D_{[0,n]}$ denote the restriction of randomized strategies $\mathcal{P}_{[0,n]}$ to the set of deterministic strategies

$$\mathcal{P}^D_{[0,n]} \triangleq \Big\{a_0 = g_0(s),\ldots,a_n = g_n(s,a^{n-1},y^{n-1})\Big\}. \tag{21}$$

By [11], for any finite n, it can be shown that $C_{0,n}(\kappa) \triangleq J_{A^n\rightarrow Y^n|s}(P^*,\kappa), \kappa \in (\kappa_{min},\infty) \subset [0,\infty)$ is a concave strictly increasing in $\kappa \in (\kappa_{min},\infty)$, and the inverse function of $C_{0,n}(\kappa)$ denoted by $\kappa_{0,n}(C)$ is a convex non-decreasing in $C \in [0,\infty)$. This implies the following duality.

Dual Extremum Problem.

$$\kappa_{0,n}(C) \triangleq \inf_{\frac{1}{n+1}I(A^n\rightarrow Y^n|s)\geq C} \mathbf{E}^P_s\Big\{\ell_{0,n}(A^n,Y^n)\Big\} \tag{22}$$

$$\geq J^{SC}_{0,n}(P^*) \triangleq \inf_{\mathcal{P}_{[0,n]}} \mathbf{E}^P_s\Big\{\ell_{0,n}(A^n,Y^n)\Big\} \equiv \kappa_{0,n}(0) \tag{23}$$

$$= \inf_{\mathcal{P}^D_{[0,n]}} \mathbf{E}^g_s\Big\{\ell_{0,n}(A^n,Y^n)\Big\} \equiv J^{SC}_{0,n}(g^*). \tag{24}$$

Here (24) follows from classical stochastic optimal control theory, which states that minimizing $\mathbf{E}^P_s\{\ell_{0,n}(A^n,Y^n)\}$ over $\mathcal{P}_{[0,n]}$ does not incur a better performance than maximizing it over $\mathcal{P}^D_{[0,n]}$ [16]. The minimum cost of control is $J^{SC}_{0,n}(P^*)$, and for $C \geq 0$, the cost of communication is

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

provided the limits exists and they are finite. Hence, for rate $C > 0$, it is necessary that the total cost of the communication system exceeds the critical value is $\kappa_{min}(n+1) = J^{SC}_{0,n}(P^*) \equiv \kappa_{0,n}(0) = J^{SC}_{0,n}(g^*)$. This is precisely the minimum cost of control, when no communication occurs, i.e., $\kappa(C) \geq \kappa_{min}$, so power is allocated to the control process. For examples of the threshold effect see [1], [3].

Suppose the randomized strategies $\mathcal{P}_{[0,n]}$ are restricted to deterministic strategies, $\mathcal{P}^D_{[0,n]}$, then by recursive substitution, $g_j(s,y^{j-1},a^{j-1}) \equiv \overline{g}_j(s,y^{j-1})$, we have $\mathbf{P}^P(dy_i|y^{i-1},s)\Big|_{P\in\mathcal{P}^D_{[0,n]}} = Q_i(dy_i|y^{i-1},\{\overline{g}_0(s),\ldots,\overline{g}_i(s,y^{j-1})\}^i_{j=0},s)$. Hence,

$$J_{A^n\rightarrow Y^n}(P^*,\kappa)\Big|_{\Big\{P^*_i(\cdot|\cdot):i=0,\ldots,n\Big\}\in\mathcal{P}^D_{[0,n]}} = 0. \tag{25}$$

By (22), then $\kappa_{0,n}(C)\Big|_{\mathcal{P}_{[0,n]}=\mathcal{P}^D_{[0,n]}} = \kappa_{0,n}(0)$, and any optimal input process consists of a control process, which controls the output process, and a process which is responsible for information transmission.

III. Gaussian Recursive Model

Consider the G-RM (1)-(6), with $S = (Y^{-1}, A_{-1})$ known to encoder/decoder. By [17], the optimal distribution of the input is of the form $P_0(da_0|s), P_i(da_i|a_{i-1},y^{i-1},s), i = 1,\ldots,n$. The directed information from $A^n \triangleq \{A_0,\ldots,A_n\}$ to $Y^n \triangleq \{Y_0,\ldots,Y_n\}$ conditioned on $S = s$ is

$$I(A^n \rightarrow Y^n|s) = \sum_{i=0}^{n}\Big\{H(Y_i|Y^{i-1},s) - H(V_i)\Big\}. \tag{26}$$

Let $\{(A^g_i, Y^g_i, Z^g_i) : i = 0,\ldots,n\}$ denote a jointly Gaussian process, given $S = s$. By the maximum entropy property of Gaussian distributions it follows that the process given by (7)-(12), and satisfies the average constraint is optimal. Now, we prepare to compute directed information using (7)-(12). We need the following definitions[2].

$$\widehat{Y}_{i|i-1} \triangleq \mathbf{E}_s\Big\{Y^g_i\Big|Y^{g,i-1}\Big\}, \quad \widehat{A}_{i|i} \triangleq \mathbf{E}_s\Big\{A^g_i\Big|Y^{g,i}\Big\},$$

$$K_{Y_i|Y^{i-1}} \triangleq \mathbf{E}_s\Big\{\Big(Y^g_i - \widehat{Y}_{i|i-1}\Big)\Big(Y^g_i - \widehat{Y}_{i|i-1}\Big)^T\Big|Y^{g,i-1}\Big\}$$

$$P_{i|i} = \mathbf{E}_s\Big(A^g_i - \widehat{A}_{i|i}\Big)\Big(A^g_i - \widehat{A}_{i|i}\Big)^T, \quad i = 0,\ldots,n.$$

From [18], and using the independent properties of the noise process, i.e., (2), (8)-(12) then

$$\widehat{A}_{i|i} = \Lambda_{i,i-1}\widehat{A}_{i-1|i-1} + U^g_i + \Delta_{i|i-1}\Big(Y^g_i - \widehat{Y}_{i|i-1}\Big), \tag{27}$$

$$\widehat{Y}_{i|i-1} = C^{i-1}Y^{g,i-1} + D_{i,i}U^g_i + \overline{\Lambda}_{i,i-1}\widehat{A}_{i-1|i-1}, \tag{28}$$

$$K_{Y_i|Y^{i-1}} = \overline{\Lambda}_{i,i-1}P_{i-1|i-1}\overline{\Lambda}^T_{i,i-1} + D_{i,i}K_{Z_i}D^T_{i,i} \tag{29}$$

$$+ K_{V_i}, i = 0,\ldots,n, \ \widehat{Y}_{0|-1} = \mathbf{E}_s\{Y^g_0\}, \widehat{A}_{-1|-1} = \mathbf{E}_s\{A^g_{-1}\}$$

where

$$\overline{\Lambda}_{i,i-1} \triangleq D_{i,i}\Lambda_{i,i-1} + D_{i,i-1}, \quad i = 0,\ldots,n,$$

$$P_{i|i} = \Lambda_{i,i-1}P_{i-1|i-1}\Lambda^T_{i,i-1} + K_{Z_i} - \Big(K_{Z_i}D^T_{i,i} + \Lambda_{i,i-1}P_{i-1|i-1}\overline{\Lambda}^T_{i,i-1}\Big)$$

$$\Phi_{i|i-1}\Big(K_{Z_i}D^T_{i,i} + \Lambda_{i,i-1}P_{i-1|i-1}\overline{\Lambda}^T_{i,i-1}\Big)^T,$$

$$\Phi_{i|i-1} \triangleq \Big[D_{i,i}K_{Z_i}D^T_{i,i} + K_{V_i} + \overline{\Lambda}_{i,i-1}P_{i-1|i-1}\overline{\Lambda}^T_{i,i-1}\Big]^{-1},$$

$$\Delta_{i|i-1} \triangleq \Big(K_{Z_i}D^T_{i,i} + \Lambda_{i,i-1}P_{i-1|i-1}\overline{\Lambda}^T_{i,i-1}\Big)\Phi_{i|i-1}$$

The innovations process denoted by $\{\nu^{\overline{e}} : i = 0,\ldots,n\}$ is an orthogonal process, independent of $\{e_i(\cdot) : i = 0,\ldots,n\}$, and satisfies the following identities.

$$\nu^{\overline{e}}_i \triangleq Y^g_i - \widehat{Y}_{i|i-1} = \overline{\Lambda}_{i,i-1}\Big(A^g_{i-1} - \widehat{A}_{i-1|i-1}\Big) + D_{i,i}Z^g_i + V_i$$

$$= \nu^{\overline{e}}_i\Big|_{e=0} \equiv \nu^0_i, \ \nu^0_i \sim N(0, K_{Y_i|Y^{i-1}}), \ i = 0,\ldots,n \tag{30}$$

[2] $\mathbf{E}_s$ means conditional expectations are for fixed $S = s$.

where $\{\nu_i^0 : i = 0,\ldots,n\}$ indicates that the innovations process is independent of the strategy $\{e_i(\cdot) : i = 0,\ldots,n\}$. Then we obtain

$$I(A^{g,n} \rightarrow Y^{g,n}|s) = \frac{1}{2}\sum_{i=0}^{n} \log \frac{|K_{Y_i|Y^{i-1}}|}{|K_{V_i}|}. \quad (31)$$

Next, we give the decentralized semi-separation principle.

Theorem 3.1: (Decentralized semi-separation of control & information transmission) Consider the G-RM (1)-(6) with $S = (Y^{-1}, A_{-1}) = s$, fixed, and for simplicity assume $C^{i-1}Y^{i-1}$ in (1) is replaced by *unit memory* $C_{i,i-1}Y_{i-1}$. Then the following hold.

(a) *Equivalent Extremum Problem.* The process given by (7)-(12) is optimal, and the following hold.

$$\begin{aligned} &Y_i^g = C_{i,i-1}Y_{i-1}^g + \overline{\Lambda}_{i,i-1}A_{i-1}^g + D_{i,i}U_i^g + D_{i,i}Z_i^g \\ &\quad +V_i,\ i = 0,\ldots,n, \quad S \triangleq (Y_{-1}, A_{-1}) = s. \quad (32) \\ &\mathbf{E}_s^{\overline{e}}\Big\{\gamma_i(A_i^g, Y_{i-1}^g)\Big\} \\ &= \mathbf{E}_s^{\overline{e}}\Big\{\langle U_i^g, R_iU_i^g\rangle + 2\langle \Lambda_{i,i-1}\widehat{A}_{i-1|i-1}, R_iU_i^g\rangle \\ &+\langle \Lambda_{i,i-1}\widehat{A}_{i-1|i-1}, R_i\Lambda_{i,i-1}\widehat{A}_{i-1|i-1}\rangle + tr\Big(K_{Z_i}R_i\Big) \\ &+tr\Big(\Lambda_{i,i-1}^T R_i\Lambda_{i,i-1}P_{i-1|i-1}\Big) + \langle Y_{i-1}^g, Q_iY_{i-1}^g\rangle\Big\}. \quad (33) \end{aligned}$$

The FTH information capacity for fixed $S = s$ is given by

$$J_{A^n\rightarrow Y^n|s}(\overline{e}^*, \kappa, s) = \sup_{\overline{\mathcal{P}}^1_{[0,n]}(\kappa)} \frac{1}{2}\sum_{i=0}^{n} \log \frac{|K_{Y_i|Y^{i-1}}|}{|K_{V_i}|} \quad (34)$$

$$\begin{aligned} &\overline{\mathcal{P}}_{[0,n]}(\kappa) \triangleq \Big\{\overline{e}_i(\cdot) \triangleq \big(e_i(\cdot), \Lambda_{i,i-1}, K_{Z_i}\big), i = 0,\ldots,n : \\ &\frac{1}{n+1}\sum_{i=0}^{n} \mathbf{E}_s^{\overline{e}}\Big(\gamma_i(A_i^g, Y^{g,i-1})\Big) \le \kappa\Big\}. \quad (35) \end{aligned}$$

(b) *Decentralized Separation of Controller and Encoder Strategies.* The optimal strategy denoted by $\{\overline{e}^*(\cdot) \equiv (e_i^*(\cdot), \Lambda_{i,i-1}^*, K_{Z_i}^*) : i = 0,\ldots,n\}$ is the solution of the dual optimization problem

$$\kappa_{0,n}(C,s) \triangleq \inf_{\big(e_i(\cdot),\Lambda_{i,i-1},K_{Z_i}\big), i=0,\ldots,n: \frac{1}{2}\sum_{i=0}^{n} \log \frac{|K_{Y_i|Y^{i-1}}|}{|K_{V_i}|} \ge (n+1)C} \mathbf{E}_s^{\overline{e}}\Big\{\sum_{i=0}^{n} \gamma_i(A_i^g, Y_{i-1}^g)\Big\}. \quad (36)$$

Moreover, the following decentralized separation holds.

(i) The optimal strategy $\{e_i^*(\cdot) : i = 0,\ldots,n\}$ is the solution of the stochastic optimal control problem with partial information given by

$$\inf_{e_i(\cdot):i=0,\ldots,n} \mathbf{E}_s^{\overline{e}}\Big\{\sum_{i=0}^{n} \gamma_i(A_i^g, Y_{i-1}^g)\Big\} \quad (37)$$

for a fixed $\{\Lambda_{i,i-1}, K_{Z_i} : i = 0,\ldots,n\}$.

(ii) The optimal strategy $\{\Lambda_{i,i-1}^*, K_{Z_i}^* : i = 0,\ldots,n\}$ is the solution of (36) for $\{e_i(\cdot) = e_i^*(\cdot) : i = 0,\ldots,n\}$.

(c) *Optimal Strategies.* Any candidate of the control strategy $\{e_i(Y^{g,i-1}) : i = 0,\ldots,n\}$ is of the form

$$\begin{aligned} &e_i(Y^{g,i-1}) \triangleq \Gamma^1_{i,i-1}Y_{i-1}^g + \Gamma^2_{i,i-1}\widehat{A}_{i-1|i-1}, \quad (38) \\ &\equiv \overline{\Gamma}_{i,i-1}\overline{Y}_{i-1}^g, \quad \overline{Y}_{i-1}^g \triangleq \begin{bmatrix} Y_{i-1}^g \\ \widehat{A}_{i-1|i-1} \end{bmatrix}, \ i = 0,\ldots,n. \end{aligned}$$

Define the augmented system

$$\begin{aligned} &\overline{Y}_i^g = \overline{F}_{i,i-1}\overline{Y}_{i-1}^g + \overline{B}_{i,i-1}U_i^g + \overline{G}_{i,i-1}\nu_i^{\overline{e}}, \quad (39) \\ &\overline{F}_{i,i-1} \triangleq \begin{bmatrix} C_{i,i-1} & \overline{\Lambda}_{i,i-1} \\ 0 & \Lambda_{i,i-1} \end{bmatrix}, \overline{B}_{i,i-1} \triangleq \begin{bmatrix} D_{i,i} \\ I \end{bmatrix}, \\ &\overline{G}_{i,i-1} \triangleq \begin{bmatrix} I \\ \Delta_{i|i-1} \end{bmatrix}, \quad i = 0,\ldots,n \end{aligned}$$

and average cost

$$\begin{aligned} &\mathbf{E}_s^{\overline{e}}\Big\{\sum_{i=0}^{n} \gamma_i(A_i^g, Y_{i-1}^g)\Big\} \equiv \mathbf{E}_s^{\overline{e}}\Big\{\sum_{i=0}^{n} \overline{\gamma}_i(U_i^g, \overline{Y}_{i-1}^g)\Big\} \\ &\triangleq \mathbf{E}_s^{\overline{e}}\Big\{\sum_{i=0}^{n}\Big(\begin{bmatrix} \overline{Y}_{i-1}^g \\ U_i^g \end{bmatrix}^T \begin{bmatrix} \overline{M}_{i,i-1} & \overline{L}_{i,i-1} \\ \overline{L}_{i,i-1}^T & \overline{N}_{i,i-1} \end{bmatrix} \begin{bmatrix} \overline{Y}_{i-1}^g \\ U_i^g \end{bmatrix} \\ &\quad +tr\big(K_{Z_i}R_i\big) + tr\big(\Lambda_{i,i-1}^T R_i\Lambda_{i,i-1}P_{i-1|i-1}\big)\Big)\Big\}, \\ &\overline{M}_{i,i-1} \triangleq \begin{bmatrix} Q_{i,i-1} & 0 \\ 0 & \Lambda_{i,i-1}^T R_i\Lambda_{i,i-1} \end{bmatrix}, \\ &\overline{L}_{i,i-1} \triangleq \begin{bmatrix} 0 \\ \Lambda_{i,i-1}^T R_i \end{bmatrix}, \quad \overline{N}_{i,i-1} \triangleq R_i. \end{aligned}$$

Then the following hold.

(1) For a fixed $\{\Lambda_{i,i-1}, K_{Z_i} : i = 0,\ldots,n\}$ the optimal strategy $\{U_i^{g,*} = e_i^*(\overline{Y}^{g,i-1}) : i = 0,\ldots,n\}$ is the solution of the stochastic optimal control problem

$$J_{0,n}(e^*(\cdot), \Lambda, K_Z, \kappa, s) \triangleq \inf_{e_i(\cdot):i=0,\ldots,n} \mathbf{E}_s^{\overline{e}}\Big\{\sum_{i=0}^{n} \overline{\gamma}_i(U_i^g, \overline{Y}_{i-1}^g)\Big\}$$

where $\{\overline{Y}_i^g : i = 0,\ldots,n\}$ satisfy recursion (39). Moreover, the optimal strategy $\{U_i^{g,*} = e_i^*(\overline{Y}^{g,i-1}) : i = 0,\ldots,n\}$ is given by the following equations.

$$\begin{aligned} &e_i^*(\overline{y}^{i-1}) = \overline{\Gamma}_{i,i-1}\overline{y}_{i-1}, \quad (40) \\ &\overline{\Gamma}_{i,i-1} = -\Big(\overline{N}_{i,i-1} + \overline{B}_{i,i-1}^T\Sigma(i+1)\overline{B}_{i,i-1}\Big)^{-1} \\ &.\Big(\overline{L}_{i,i-1}^T + \overline{B}_{i,i-1}^T\Sigma(i+1)\overline{F}_{i,i-1}\Big), \quad i = 0,\ldots,n-1 \quad (41) \end{aligned}$$

$e_n^*(\overline{y}^{n-1}) = -\overline{N}_{n,n-1}^{-1}\overline{L}_{n,n-1}^T\overline{y}_{n-1}$, where the symmetric positive semidefinite matrix $\{\Sigma(i) : i = 0,\ldots,n\}$ satisfies a matrix difference Riccati equation, for $i = 0,\ldots,n-1$,

$$\begin{aligned} &\Sigma(i) = \overline{F}_{i,i-1}^T\Sigma(i+1)\overline{F}_{i,i-1} - (\overline{F}_{i,i-1}^T\Sigma(i+1)\overline{B}_{i,i-1} + \overline{L}_{i,i-1}) \\ &.\Big(\overline{N}_{i,i-1} + \overline{B}_{i,i-1}^T\Sigma(i+1)\overline{B}_{i,i-1}\Big)^{-1}\Big(\overline{B}_{i,i-1}^T\Sigma_{i,i-1}\overline{F}_{i,i-1} \\ &+\overline{L}_{i,i-1}^T\Big) + \overline{M}_{i,i-1}, \quad \Sigma(n) = diag\{Q_{n,n-1}, 0\} \end{aligned}$$

and the optimal pay-off is given by

$$J_{0,n}(e^*(\cdot),\Lambda,K_Z,\kappa,s) = \sum_{j=0}^{n}\Big\{tr\big(K_{Z_j}R_j\big)$$

$$+tr\big(\Lambda^T_{j,j-1}R_j\Lambda_{j,j-1}P_{j-1|j-1}\big)\Big\} + \sum_{j=0}^{n-1} tr\Big(K_{Y_j|Y^{j-1}}\overline{G}^T_{j,j-1}$$

$$.\Sigma(j+1)\overline{G}_{j,j-1}\Big) + \mathbf{E}\langle\overline{Y}_{-1|-1},\Sigma(0)\overline{Y}_{-1|-1}\rangle$$

(2) The optimal strategies $\{(\Lambda^*_{i,i-1}, K^*_{Z_i}) : i = 0,\ldots,n\}$ are the solutions of the optimization problem

$$\kappa_{0,n}(C,s) \stackrel{\triangle}{=} \inf_{\big(\Lambda_{i,i-1},K_{Z_i}\big),i=0,\ldots,n:\frac{1}{2}\sum_{i=0}^{n}\log\frac{|K_{Y_i|Y^{i-1}}|}{|K_{V_i}|}\geq(n+1)C} \Big\{ J_{0,n}(e^*(\cdot),\Lambda,K_Z,\kappa)\Big\}.$$

Proof: (a) This follows from (4) and (9). (33) is obtained using the reconditioning property of expectation. (b) (36) follows from the dual relation (22). (i), (ii) follow from the observation that the constraint in (36) depends only on $\{\Lambda, K_Z\}$ and not on $\{e_i(\cdot) : i = 0,\ldots,n\}$. (c), (i). (38) follows from (27), (30), because $\{Y_i, \widehat{A}_{i|i} : i = 0,\ldots,n\}$ is a sufficient statistics for the control process. The rest of the equations follows directly from the solution of partially observable stochastic optimal control problems [19]. ∎

Theorem 3.1, (1) and (2) are Person-by-Person Optimality statements of $\{e_i(\cdot) : i = 0,\ldots,\}$ and $\{\Lambda_{i,i-1}, K_{Z_i} : i = 0,\ldots,n\}$.

Theorem 3.1, (c) states that the optimal input process consists of 4 strategies, follows.

$$A^g_i = \Gamma^1_{i,i-1}Y^g_{i-1} + \Gamma^2_{i,i-1}\widehat{A}_{i-1|i-1} + \Lambda_{i,i-1}A^g_{i-1} + Z^g_i. \quad (42)$$

Remark 3.2: By Theorem 3.1, if $C_{i,i-1} = 0, Q_{i,i-1} = 0, i = 0,\ldots,n$ then $e^*_i(\overline{y}^{i-1}) = -\Lambda_{i,i-1}\widehat{A}_{i-1|i-1}, i = 0,\ldots,n$, and hence

$$A^g_i = \Lambda_{i,i-1}\Big(A^g_{i-1} - \widehat{A}_{i-1|i-1}\Big) + Z^g_i,\ i = 0,\ldots,n. \quad (43)$$

That is, $\widehat{A}_{i-1|i-1}, i = 0,\ldots,n$ is a sufficient statistic for the strategy $e_i(Y^{g,i-1}), i = 0,\ldots,n$, as expected.

Next, we discuss item Section I, (d).

Theorem 3.3: (Decentralized coding theorem)
Consider the G-RM of Theorem 3.1.
(a) If $D_{i,i-1} = 0, i = 0,\ldots,n$, then [2], Theorem IV.1 holds that states, directed information stability holds and separates into (i) a statement related to the ergodic properties of a stochastic optimal control problem with complete information, and (ii) a statement related to an information transmission problem.
(b) For the general G-RMs of Theorem 3.1 with $D_{i,i-1} \neq 0, i = 0,\ldots,n$, then (a) holds as in [2], Theorem IV.1, with some variations.

Proof: (b) This is done similar to [2], Theorem IV.1. ∎

IV. Conclusions

The decentralized features of extremum problems of capacity of models with memory and feedback are illustrated. For Gaussian recursive models with past dependence on inputs and outputs it is illustrated that a semi-separation principle holds, that makes calculations and the derivation of directed information stability simpler.

References

[1] 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,” *IEEE Transactions on Information Theory*, submitted in April 2016.

[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*, Accepted in November 2017, submitted in November 2015.

[3] 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.

[4] 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.

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

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

[7] 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.

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

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

[10] 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.

[11] 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.

[12] 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 in May 2017.

[13] J. Chen and T. Berger, “The capacity of finite-state Markov channels with feedback,” *IEEE Transactions on Information Theory*, vol. 51, no. 3, pp. 780–798, March 2005.

[14] 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.

[15] 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.

[16] I. I. Gihman and A. V. Skorohod, *Controlled Stochastic Processes*. Springer-Verlag, 1979.

[17] 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

[18] R. S. Liptser and A. N. Shiryaev, *Statistics of Random Processes: I. General Theory*, 2nd ed. Springer-Verlag, Berlin, Heidelberg, New York, 2001.

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