Stochastic Control over Finite Capacity Channels: Causality and Feedback

Charalambos D. Charalambous, Christos K. Kourtellaris and Photios Stavrou

Abstract—Optimal communication/control analysis and design of dynamical controlled systems, when there are finite capacity communication constraints, often involve information and control theoretic analysis. This paper employs information theoretic concepts which are subject to causality, and thus applicable to control/communication system analysis and design in which encoders, decoders and controllers are causal. The basic mathematical concepts are constructive: they are based on a modified version of Shannon’s self-mutual information, known as directed information, which describes how much information a random process conveys to another random process, to account for the causality of the probabilistic channel connecting these random processes, when feedback is used. Following this construction data processing inequalities are derived while the solution to the casual information rate distortion is obtained in which the optimal reconstruction kernel is causally dependent on the source and control. Further, in view of the complexity of the causal rate distortion function, a lower bound on the causal rate distortion function is derived which is easy to apply in many control/communication systems. Finally, using the information transmission theorem, the connection between stabilizability and observability of the control system is related to the causal rate distortion function and its lower bound.

I. INTRODUCTION

Over the past few years there has been a renewed interest in developing a mathematical framework consisting of control and information theoretic concepts for controlling dynamical systems over finite capacity communication channels or finite feedback data rates [1], [2], [3], [4], [5], [6], [7], [8], [9]. These type of control/communication systems are often represented by a control system whose output is the input to a communication channel, while the output of the communication channel is the input to the controller. There is, however, one fundamental difference when designing real-time feedback control or estimation strategies, which is causality. Causality is an inherent property of any real-time control and/or communication system or data compression system, especially when dealing with feedback communication systems with memory. Clearly, for the control/communication system, since control is applied in real-time, then the encoder, channel, decoder, and controller should be causal transformations on their corresponding input signals.

In this paper, some fundamental notions associated with Shannon’s theory of reliable transmission, are presented in the context of the control/communication system. These are the following.

1) Expressions for feedforward and feedback information with side information associated with the causal forward and feedback channel. The feedforward information is the directed information described in [13], while the information channel capacity is its maximization over all encoders.

2) An expression for causal rate distortion function with side information in terms of a causal reproduction kernel. An optimal reproduction kernel which involves a causal operation on the source data and control.

3) A tight lower bound on the causal rate distortion function, which depends on the conditional entropy rate of the source output given the decoder output the distortion measure and control.

4) General notions of observability and stabilizability and necessary conditions for the existence of an encoder, decoder, and controller, using a causal information transmission theorem and the causal rate distortion lower bound.

The material presented in this paper compliment those found in [13], while they shed light into some of the existing definitions and notions. Specifically, it is shown the directed information which is introduced in [13] and further explored in the control/communication context in [14], [4], is a consequence of Shannon’s self-mutual information restriction to causal channels. Specifically, mutual information for causal channels is defined via the Shannon [10] self-mutual information by restricting the Radon-Nykodym Derivative (RND) associated with self-mutual information to a non-anticipative (or causal) channel. Then by taking expectation of the restricted self-mutual information it is shown that the resulting mutual information is precisely the expression of directed information [13], [14], [4]. The later is viewed as the appropriate measure to define the capacity of channels with feedback. Subsequently, the information rate distortion function [10], [15], [16] is defined using the restricted self-mutual information and the optimal (minimizing) reproduction kernel is derived. The information transmission theorem is derived using the data processing inequalities and the connection between stabilizability and observability.
of the control system is related to the causal rate distortion function and its lower bound. Finally it is employed to establish necessary conditions for controlling systems over finite capacity channels.

II. PROBLEM FORMULATION

This paper is concerned with the control/communication system of Fig. II.1 defined on a complete probability space \((\Omega, \mathcal{F}(\Omega), P)\) with filtration \(\{\mathcal{F}_t\}_{t \geq 0}\), subject to causal feedback, encoding, decoding and control. Here, \(Y_t \in \mathcal{Y}, Z_t \in \mathcal{Z}, \tilde{Z}_t \in \tilde{\mathcal{Z}}, Y_t \in \mathcal{Y}, U_t \in \mathcal{U}\) are Random Variables (R.V.'s) denoting the source output, channel input, channel output, reproduction of the source output, and control input, respectively, at time \(t \in \mathbb{N}_+ = \{0, 1, 2, \ldots\}\). It is assumed that \(\mathcal{Y}, \mathcal{Z}, \tilde{\mathcal{Z}}, \mathcal{Y}, \mathcal{U}\) are complete separable metric spaces, and \(\{\mathcal{Y}_t, \mathcal{F}(\mathcal{Y}_t)\}, \{\mathcal{Z}_t, \mathcal{F}(\mathcal{Z}_t)\}, \{\tilde{\mathcal{Z}}_t, \mathcal{F}(\tilde{\mathcal{Z}}_t)\}\) and \(\{\mathcal{U}_t, \mathcal{F}(\mathcal{U}_t)\}\) are measurable spaces (e.g., \(\mathcal{F}(\mathcal{Y}_t)\)) being an algebra of subsets of the set \(\mathcal{Y}\) generated by closed sets. For \(T \in \mathbb{N}_+\), sequences of the R.V.'s with length \(T\) are identified with the product measurable spaces \((\mathcal{Y}_0, \mathcal{Z}_0, \tilde{\mathcal{Z}}_0, \mathcal{Y}_T, \mathcal{Z}_T, \tilde{\mathcal{Z}}_T)\) \(\triangleq \bigotimes_{k=0}^{T} \mathcal{Y}_k, \bigotimes_{k=0}^{T} \mathcal{Z}_k, \bigotimes_{k=0}^{T} \tilde{\mathcal{Z}}_k\). Throughout the paper, sequences of R.V.'s are denoted by \(Y_0, Y_1, \ldots, Y_T\) \(\in \mathcal{Y}_0, \mathcal{Y}_T\), where the probability measure induced by \(Y^T\) is denoted by \(\mathcal{M}_1(Y_0, T), T \in \mathbb{N}_+\). \(\log()\) denotes logarithm with base 2 and the subscript \(e\) is used to denote the natural logarithm.

Next, we introduce the definition of conditional independence which is vital in applications of information theory to channel capacity and rate distortion.

**Conditional Independence:** Conditionally independent R.V.'s are denoted by \((X, Y) \perp \perp Z\) or equivalently \(Y \rightarrow Z \rightarrow X\) form a Markov chain.

Encoders, decoders and controllers are defined via conditional distributions, hence the notion of a stochastic kernel is introduced.

**Stochastic Kernel:** Given the measurable spaces \((A, \mathcal{A}), (\hat{A}, \hat{\mathcal{A}})\), a stochastic Kernels is a mapping \(P : A \times \hat{A} \rightarrow [0, 1]\) satisfying the following two properties: 1) For every \(x \in A\), the set function \(P(:, x)\) is a probability measure (possibly finitely additive) on \(\hat{A}\), and 2) for every \(F \in \hat{\mathcal{A}}\), the function \(P(F,:)\) is \(\mathcal{A}\)-measurable.

The set of such stochastic Kernels is denoted by \(Q(A, \hat{A})\).

This section discusses the generalization of a communication channel with feedback in which the source is a dynamical control system as shown in Fig. II.1. The importance difference is the presence of the control law \(U_t = \mu(t, \omega)\) (where \(\mu(\cdot, \cdot)\) is the controller). In general, the presence of the control in the loop implies that the encoder and the decoder are functions of the control sequence generated by the control Law.

In such control/communication applications an interesting question is whether the design of the control and communication system can be separated and whether the encoder and decoder can be independent of the control Law (e.g. no side information of the control sequence is required at the encoder and the decoder). For Linear-Quadratic Gaussian Control systems and additive Gaussian channels such separation is shown in [9], where it is also shown that the encoder and decoder are independent of the control sequence. The following encoders, decoders and controllers will be discussed when considering specific control and communication systems.

**Information Source:** Given the control sequence for any \(t \in \mathbb{N}_+\) the information source is identified by the stochastic Kernel \(P(dY_t; y^{t-1}, u^t) \in Q(Y_t; \mathcal{Y}_{0,t-1} \times \mathcal{U}_{0,t})\)

**Encoder:** We define and discuss the following types of encoders:

Class A) This encoder at any time \(t \in \mathbb{N}_+\) is modeled by a stochastic kernel \(P(dZ_t; z^{t-1}, y^t, \tilde{z}^{t-1}) \in Q(Z_t; \mathcal{Z}_{0,t-1} \times \mathcal{Y}_{0,t} \times \tilde{\mathcal{Z}}_{0,t-1})\). Hence, the decoder assumes knowledge of the channel output and source output.

Class B) This encoder at any time \(t \in \mathbb{N}_+\) is modeled by a stochastic kernel \(P(dz_t; z^{t-1}, y^t, z^{t-1}, u^{t-1}) \in Q(Z_t; \mathcal{Z}_{0,t-1} \times \mathcal{Y}_{0,t} \times \mathcal{Z}_{0,t-1} \times \mathcal{U}_{0,t-1})\). Hence, this encoder has side information in the sense that it requires knowledge of the control sequence, in addition to the source output and the channel output.

**Communication Channel with Memory and Feedback:** A communication channel with memory and feedback is defined via the family of stochastic kernels \(\{d\tilde{Z}_t; \tilde{z}^{t-1}, z^t, u^t\} \in Q(\tilde{Z}_t; \mathcal{Z}_{0,t-1} \times \mathcal{Y}_{0,t} \times \mathcal{Z}_{0,t})\) where \(\tilde{Z}^t = \tilde{z}^t\) is the specific realization of the channel input, \(\tilde{Z}^{t-1} = \tilde{z}^{t-1}\) is a specific realization of the previous channel outputs and \(U^t = u^t\) is a specific realization of the previous controller outputs.

**Decoder:** We define and discuss the following types of decoders:

Class A) The decoder at any time \(t \in \mathbb{N}_+\) is modeled by a stochastic kernel \(P(d\tilde{Y}_t; \tilde{y}^{t-1}, \tilde{z}^t) \in Q(\tilde{Y}_t; \tilde{\mathcal{Y}}_{0,t-1} \times \mathcal{Z}_{0,t})\). Hence, the decoder assumes knowledge of the channel output and decoder output but it does not assume knowledge of the
control sequence.

Class B) The decoder at any time $t \in \mathbb{N}_+$ is modeled by a stochastic kernel $P(dU_t; \tilde{y}^{t-1}) \in \mathcal{Q}(U_t; \tilde{Y}_{0,t-1} \times U_{0,t-1})$. Hence, the decoder assumes knowledge of the channel output, previous decoder outputs, and control sequence.

**Controller:** We define and discuss the following types of controllers.

Class A) The controller at any time $t \in \mathbb{N}_+$ is modeled by a stochastic kernel $P(dU_t; y^{t-1}) \in \mathcal{Q}(U_t; \tilde{Y}_{0,t-1})$. In this case, the control law assumes knowledge of the decoder output and previous control sequence.

Class B) The controller at any time $t \in \mathbb{N}_+$ is modeled by a stochastic kernel $P(dU_t; y^{t-1}, \tilde{u}^{t-1}) \in \mathcal{Q}(U_t; \tilde{Y}_{0,t-1} \times U_{0,t-1})$. Hence, the control law assumes knowledge of the decoder output and previous control sequence.

Next, we define the notion of observability associated with the source output.

**Definition 2.1:** (Observability in Probability and r-Mean).

Consider a general control/communication system.

This system is called observable in probability if for a given $\delta > 0$, there exist a control sequence and an encoder and a decoder such that $\lim_{t \to \infty} \delta \sum_{i=0}^{T-1} dP(Y_i, \tilde{Y}_i) \leq D_v$, where

$$
\rho(Y, \tilde{Y}) \triangleq \begin{cases} 
1 & \text{if } ||Y - \tilde{Y}|| > \delta \\
0 & \text{if } ||Y - \tilde{Y}|| \leq \delta
\end{cases}
$$

in which $||.||$ is the norm on the product space $Y \times \tilde{Y}$, and $D_v \in [0, 1]$.

The system is called observable in $r$-mean if there exist a control sequence and an encoder and a decoder such that $\lim_{t \to \infty} \delta \sum_{i=0}^{T-1} dP(Y_i, \tilde{Y}_i) \leq D_v$, for some finite $D_v \geq 0$, where $\rho(Y, \tilde{Y}) = ||Y - \tilde{Y}||^r$, $r > 0$, and finite $D_v \geq 0$.

Next, we define the notion of stabilizability associated with any controlled source.

**Definition 2.2:** (Stabilizability in Probability and r-Mean).

Consider a general control/communication system with output $\{H_t; t \in \mathbb{N}_+\}$. The system output $\{H_t; t \in \mathbb{N}_+\}$ is called stabilizable in probability if for a given $\delta > 0$, there exist a controller, encoder, and decoder such that $\lim_{T \to \infty} \delta \sum_{i=0}^{T-1} dP(H_i, 0) \leq D_v$, where $\rho(., .)$ is defined by (II.1), and $D_v \in [0, 1]$.

This system is called stabilizable in $r$-mean if there exist a controller, encoder, and decoder such that $\lim_{T \to \infty} \delta \sum_{i=0}^{T-1} dP(H_i, 0) \leq D_v$, where $\rho(H, 0) = ||H - 0||^r$, $r > 0$, and finite $D_v \geq 0$.

Note that the reconstruction of the source output at the output of the decoder is part of the information used as input to the controller to stabilize control signals.

**III. INFORMATION THEORY FOR CONTROLLED SOURCES AND CHANNELS WITH FEEDBACK**

In this section, we derived basic information theory results [10] for causal feedback channels, with memory.

**A. Mutual Information: Feedforward and Feedback Rates**

First, we recall the classical definition of mutual information in order to constructively introduced feedforward and feedback information rates (see [9]).

Given a communication channel (with memory and feedback), the self-information and feedback mutual information between $Z^N$ and $\tilde{Z}^N$, defined via the channel $\{P(dZ^N; \tilde{Z}^N) \in \mathcal{Q}(\tilde{Z}_{0,N}; Z_{0,N}) ; N \in \mathbb{N}_+\}$ is defined by

$$
i(Z^N; \tilde{Z}^N) = \sum_{i=0}^{N} \log \frac{P(\tilde{Z}_i; \tilde{Z}_i^{t-1}, Z^N)}{P(\tilde{Z}_i; Z^N)} + \sum_{i=0}^{N} \log \frac{P(Z_i; Z_i^{t-1})}{P(Z_i; \tilde{Z}_t^{t-1})}
$$

and by taking expectation yields

$$
I(Z^N; \tilde{Z}^N) = I(Z^N \rightarrow \tilde{Z}^N) + I(Z^N \leftarrow \tilde{Z}^N)
$$

where

$$
I(Z^N \rightarrow \tilde{Z}^N) = \sum_{i=0}^{N} I(Z^N_i; \tilde{Z}_i^{t-1}) = \sum_{i=0}^{N} \int \log \frac{P(d\tilde{Z}_i; Z_i^{t-1})}{P(d\tilde{Z}_i; Z_i^{t-1})} P(d\tilde{Z}_i; Z_i^{t-1})
$$

and

$$
I(Z^N \leftarrow \tilde{Z}^N) = \sum_{i=0}^{N} I(\tilde{Z}_i^{t-1}; Z_i^{t-1}) = \sum_{i=0}^{N} \int \log \frac{P(d\tilde{Z}_i; \tilde{Z}_i^{t-1})}{P(d\tilde{Z}_i; \tilde{Z}_i^{t-1})} P(d\tilde{Z}_i; \tilde{Z}_i^{t-1})
$$

Notice that the term $I(Z^N \rightarrow \tilde{Z}^N)$ is the directed information from $Z^N$ to $\tilde{Z}^N$ and the mutual information between $Z^N$ and $\tilde{Z}^N$.

$$
(P(d\tilde{Z}_i; \tilde{Z}_i^{t-1}, Z_i^{t-1}) \in \mathcal{Q}(\tilde{Z}_i; \tilde{Z}_{i,t-1}^{t-1} ; Z_{i,t-1} ; t \in \mathbb{N}_+))
$$

denoted by $Z^N \rightarrow \tilde{Z}^N$, with $Z^N$ being the channel stochastic kernel.
hence it represents the feedforward information rate. It can be obtained from the self-mutual information $i(N; Z N)$ via the restriction of the RND $P(d Z N; Z N)$ as follows:

$$i(Z N; Z N)|_R \equiv \frac{P(d Z N; Z N)}{P(d Z N)} \bigg|_R = \pi N = 0 \frac{P(d Z N; Z N)}{P(Z N; Z N)},$$

where $P(d Z N; Z N)$ is the RND of $Z N_i$ to account for the direction $Z N \rightarrow Z N$. The definition of $I(Z N \rightarrow Z N)$ is given in [13]. On the other hand, $I(Z N \leftarrow Z N)$ is the directed information from $Z N$ to $Z N$ associated with the channel stochastic kernel $\{P(d Z i; z^{i-1}, z^i)\} \in Q(Z_i; Z_{0,i-1} \times Z_{0,i-1})$; $t \in N_+$. Hence, it represents the feedback information.

In the remaining of the paper we shall denote the channel $\{P(d Z i; z^{i-1}, z^i)\} \in Q(Z_i; Z_{0,i-1} \times Z_{0,i-1})$ by $Z N \rightarrow Z N$ and the channel $\{P(d Z i; z^{i-1}, z^i)\} \in Q(Z_i; Z_{0,i-1} \times Z_{0,i-1})$ by $Z N \leftarrow Z N$.

A channel with feedback always satisfies $I(Z N \rightarrow Z N) \geq 0$. On the other hand, if a channel has no feedback then by definition of feedback encoding we have $I(Z N \rightarrow Z N) = 0$, and $I(Z N; Z N) = I(Z N \rightarrow Z N)$. Hence, for a memoryless channel we have $I(Z N; Z N) \geq I(Z N \rightarrow Z N) = \sum N = 0 \log \frac{P(d Z i, d z_i)}{P(d z_i, d z_i)}$, with equality if only if $Z$, and $Z_i$ are conditional independent given $i$, for all $i \in \{0, \ldots, N\}$.

Finally, it is noted that the directed information from $Z N$ to $(Z N, U N)$ satisfies the identity:

$$I(Z N \rightarrow Z N, U N) = \sum N = 0^N 0 \sum \log \frac{P(d Z i, U i; Z i)}{P(d Z i, U i)} + I(Z N \rightarrow U N || Z N) - I(Z N \leftarrow Z N),$$

where

$$I(Z N \rightarrow Z N, U N) = \sum N = 0^N I(Z i; Z \leftarrow Z i, U i)$$

and

$$I(Z N \rightarrow Z N, U N) = \sum N = 0^N 0 \sum I(Z i; U i; Z i),$$

**B. Data Processing Inequalities for Control-Communication**

In this section, we shall introduce the data processing inequalities for causal systems. Suppose the receiver has access to the channel output and control sequence. The mutual information between $Y N$ and $(Z N, U N)$ is

$$I(Y N; Z N, U N) = I(Y N; Z N) + I(Y N; U N),$$

for all $i \in \{0, \ldots, N\}$. Similarly, if the decoder has access to the control sequence the mutual information between $Y N$ and $(Z N, U N)$ is

$$I(Y N; Z N, U N) = \sum N = 0^N \sum (H(Y i; Z i, U i) - H(Y i| Z i, U i, Z N))$$

Therefore we have the following theorem.

**Theorem 3.1:** (Data processing inequalities) Consider the controlled communication system of figure II.1 satisfying the Markov chain assumptions

$$Y N \rightarrow (Z i-1, U i-1, Z i) \rightarrow (Z i, U i), \forall i \in N_+$$

$$\hat{Y} N \rightarrow (Z i, U N, Y i-1) \rightarrow Y i, \forall i \in N_+$$

Then

$$I(Y N; \hat{Y} N, U N) \leq I(Y N; \hat{Y} N, U N) \leq I(Z N \rightarrow \hat{Z} N, U N)$$

Assuming an encoder and decoder of either Class A or Class B, then

$$I(Y N; \hat{Y} N, U N) = I(Y N; \hat{Y} N, U N) \leq I(Z N \rightarrow \hat{Z} N, U N)$$

**Proof:** Follows from above discussion.

**C. Information Capacity for Channels with Memory and Feedback**

For general controlled sources and feedback channels with memory shown in (II.1) we use the data processing inequalities to the expressions for information capacity.

**Definition 3.2:** (Information Capacity for Channels with Memory and Feedback) Consider a communication channel (with memory and feedback) $(U N, Z N) \rightarrow \hat{Z} N$ defined via \{\[P(d Z i; z^{i-1}, z^i, u^i)\] \in Q(Z i; Z_{0,i-1} \times Z_{0,i-1} \times U_{0,i}); i \in N_+\}. Let $D^N C \subset \{Q(Z i; Z_{0,i-1} \times Z_{0,i-1} \times U_{0,i})\}^N$ denote the channel input constraints. The information capacity for the time horizon $N$ is defined by

$$C_N \triangleq \sup_{\{P(d z_{i-1}; z^i, u^i)\} \in D^N C} I(Z N \rightarrow \hat{Z} N||U N),$$

while the information capacity in bits per time step is defined by

$$C \triangleq \lim_{N \rightarrow \infty} \frac{1}{N + 1} C_N.$$

In general, the Shannon information capacity definition is not equal to the operational capacity definition [16].
D. Rate Distortion for Sources with Memory and Feedback

Next, we define the rate distortion function for controlled sources with memory and decoders with feedback. Specifically, we state the definition of the information rate distortion, when the RND associated with self-mutual information is restricted to a non-anticipative source reproduction kernel. Moreover, we state the solution of the rate distortion function and the optimal reproduction kernel which is causally dependent on the decoding sequence, source sequence and control sequence.

Definition 3.3: (Causal Information Rate Distortion Function for sources with memory and feedback) Consider the data compression channel \((U^T, Y^T) \rightarrow \tilde{Y}^T\) defined via \(\{P(d\tilde{Y}_i; \tilde{y}_i^{-1}, y_i, u_i) \in Q(\tilde{Y}_i; \tilde{Y}_{0,i-1} \times Y_{0,i} \times U_{0,i}); \ i \in N_+\}\). Let \(D^R_{DC} = \left\{ P(d\tilde{Y}_i; \tilde{y}_i^{-1}, y_i, u_i) \in Q(\tilde{Y}_i; \tilde{Y}_{0,i-1} \times Y_{0,i} \times U_{0,i}) \right\}_{i=0}^{T-1} / \frac{1}{T^2} \sum_{i=0}^{T-1} E P(d\tilde{y}_i, d\tilde{y}_i^*) P_i(y_i, \tilde{y}_i) \leq D_o \} \) denote the distortion constraint, where \(D_o \geq 0\) is the distortion level and \(\{P_i(\cdot; t) \in N_+\}\) is a sequence of distortion measures (which are continuous in the second argument and non-negative functions).

The causal information rate distortion function for the time horizon \(T\) is defined by

\[
R_T(D_o) \triangleq \inf_{\{P(d\tilde{Y}_i; \tilde{y}_i^{-1}, y_i, u_i) \in Q(\tilde{Y}_i; \tilde{Y}_{0,i-1} \times Y_{0,i} \times U_{0,i}) \}} I(Y^T \rightarrow \tilde{Y}^T | U^T)
\]

and subsequently, the information rate sequential distortion function in each time step is

\[
R(D_o) \triangleq \lim_{T \to \infty} \frac{1}{T+1} R_T(D_o)
\]

provided the limit exists.

Next, we give the optimum non-anticipative (or causal) source reproduction kernel and the expression of the rate distortion function.

Theorem 3.4: (Causal Information Rate Distortion Function for Sources with Memory and Feedback) The information rate distortion function is given by

\[
R^D_T(D_o) = \sum_{t=0}^{T} \int_{Y_{0,t} \times Y_{0,t-1} \times U_{0,t}} \log \left\{ \int_{\tilde{Y}_t} e^{p_t(y_t, \tilde{y}_t)} P^*(d\tilde{y}_t; \tilde{y}_t^{-1}, y_t, u_t) \times P(du_t; u_t^{-1}, \tilde{y}_t^{-1}, y_t^{-1}, u_t^{-1}) \times P(dy_t; y_t^{-1}, u_t^{-1}, \tilde{y}_t^{-1}, y_t^{-1}) \right\} \times P(dy_t; y_t^{-1}, u_t^{-1}, \tilde{y}_t^{-1}, y_t^{-1}) \times P(du_t; u_t^{-1}, \tilde{y}_t^{-1}, y_t^{-1}, u_t^{-1})
\]

where \(s \leq 0\) is found from the constraint. Moreover, the non-anticipative stochastic Kernel which achieves the infimum of the rate distortion function is given by

\[
P^*(d\tilde{Y}_t; \tilde{y}_t^{-1}, y_t, u_t) = \frac{e^{s\rho_t(y_t, \tilde{y}_t)}}{\int_{\tilde{Y}_t} e^{s\rho_t(y_t, \tilde{y}_t)}} P(d\tilde{Y}_t; \tilde{y}_t^{-1}, y_t, u_t)
\]

Proof: The derivation is omitted due to space limitation.

Theorem 3.4 states that the infimizing reproduction kernel \(P^*(d\tilde{Y}_t; \tilde{y}_t^{-1}, y_t, u_t)\) is a causal mapping \(\tilde{Y}_{0,i-1} \times Y_{0,i} \times U_{0,i} \rightarrow \tilde{Y}_i\).

E. Information Transmission Theorem

Next, we present a necessary condition for end to end transmission up to distortion level \(D_o \geq 0\). This is the converse of information transmission theorem.

Theorem 3.5: (Information Transmission Theorem) A necessary condition for reproducing a sequence of source messages \(Y^T\) up to distortion level \(D_o\) by \(\tilde{Y}^T\) at the output of the decoder using a sequence of the channel inputs and outputs with length \(N (T \leq N)\) is

\[
C_N \geq R^D_T(D_o).
\]

Proof: Follows from the data processing inequalities.

F. Tight Lower Bound for Causal Rate Distortion Function

Clearly, the rate distortion function is very difficult to compute. Next, we present a lower bound which is practical in terms of providing a tight necessary condition for observability (reconstruction) and stabilizability.

Lemma 3.6: (Lower Bound for Causal Rate Distortion Function) Let \(Y^T, Y_i \in \mathbb{R}^d, 0 \leq t \leq T\) be a sequence with length \(T\) produced by the source, the kernel satisfies \(P(Y_i \in d\tilde{y}_i; \tilde{y}_i^{-1}, u_i) = f(y_i; \tilde{y}_i^{-1}, u_i) dy_i, t \in N_+\). Consider the following form of distortion measure \(\rho_t(y_t, \tilde{y}_t) = \rho(y_t, \tilde{y}_t) = \rho(y_t - \tilde{y}_t) : \mathbb{R}^d \rightarrow [0, \infty)\) is continuous. Then, i) A lower bound for \(\frac{1}{T+1} R^D_T(D_o)\) is given by

\[
\frac{1}{T+1} R^D_T(D_o) \geq \frac{1}{T+1} \sum_{i=0}^{T} H_S(Y_i|\tilde{Y}^{i-1}, U^i)
\]

and

\[
\max_{h \in G_D} H_S(h),
\]

where \(G_D\) is defined by \(G_D \triangleq \{ h : \mathbb{R}^d \rightarrow [0, \infty): \int_{\mathbb{R}^d} \rho(h)(\xi) d\xi = 1, \int_{\mathbb{R}^d} e^{s\rho(h)(\xi)} d\xi < \infty \text{ for all } s < 0, \text{ then } h^*(\xi) \in G_D \text{ that maximizes } H_S(h) \text{ is }
\]

\[
h^*(\xi) = \frac{e^{s\rho(h)}}{\int_{\mathbb{R}^d} e^{s\rho(h)} d\xi} \int_{\mathbb{R}^d} \rho(h) h^*(\xi) d\xi = D_o.
\]

Subsequently, when \(R^D(T, D_o)\) and \(H_S(Y) \triangleq \lim_{T \to \infty} \frac{1}{T+1} \sum_{i=0}^{T} H_S(Y_i|\tilde{Y}^{i-1}, U^i)\) exist, the lower bound denoted by \(R^D_S(D_o)\) is given by

\[
R^D(D_o) \geq H_S(Y) - \max_{h \in G_D} H_S(h) \triangleq R^D_S(D_o).
\]
ii) Suppose the difference distortion measure $p(.)$ satisfies the conditions a, b, d of ([17], pp. 2029), $\int_0^t e^{|p(U)|}d\xi < \infty$ for all $s < 0$, $\mathcal{H}_S(Y_t) > -\infty$ and there exists an $y^* \in \mathbb{R}^d$ such that $E[p(y - y^*)] < \infty$, $\forall y \in \mathbb{R}^d$.

Then, in the limit as $D_v \to 0$, the lower bound is asymptotically exact. That is, for the case when $R^D(D_v)$ and $\mathcal{H}_S(Y_t)$ exist, $\lim_{D_v \to 0} \left[ R^D(D_v) - \left( \mathcal{H}_S(Y_t) - \mathcal{H}(s) \right) \right] = 0$.

Proof: This is a generalization of [17].

Remark 3.7: For distortion measure $p(y, \tilde{y}) = ||y - \tilde{y}||$, $\max_{h \in D} \mathcal{H}_S(h) = \log e^\frac{r}{d} - \log (\frac{d}{d\pi^d}) \frac{d}{2}$ bits per source message, where $V_d$ is the volume of the unit sphere and $\Gamma$ is the gamma function [17].

IV. NECESSARY CONDITIONS FOR RECONSTRUCTION AND STABILITY OF GENERAL SYSTEMS

For the general control/communication system of Fig. II.1, the main theorem which connects reliable communication (e.g., reconstruction) and stability for general systems is given next.

Theorem 4.1: Consider the system of Fig. II.1 under the Markov chain assumption in which $Y^T, Y_t \in \mathbb{R}^d$ is the observed process to be reconstructed at the output of the decoder. Assume the Shannon entropy rate $\mathcal{H}_S(Y_t)$ exists and it is finite.

For the reconstruction of $Y^T$ in probability, a necessary condition on the channel capacity is

$$C \geq \mathcal{H}_S(Y_t) - \frac{1}{2} \log((2\pi e)^d \det \Gamma_g) \geq R^D(D_v), \quad (IV.12)$$

where $\mathcal{H}_S(Y_t) = \lim_{T \to \infty} \frac{1}{T} \sum_{t=0}^T \mathcal{H}(Y_t|Y^{t-1}, U^t)$ and $\Gamma_g$ is the covariance matrix of the Gaussian distribution $h^* = N(0, \Gamma_g)$, $\xi \in \mathbb{R}^d$ which satisfies

$$\int_{||\xi||^2} \xi^\top \xi ^{d/2} d\xi = D_v.$$ Moreover, a necessary condition on the channel capacity for the reconstruction of $Y^T, Y_t \in \mathbb{R}^d$ in the $r$-mean is

$$C \geq \mathcal{H}_S(Y_t) - \log e^\frac{r}{d} + \log \left( \frac{r}{d \mathcal{V}_d \Gamma_g^{1/2}} \right) \geq R^D(D_v), \quad (IV.13)$$

where $\Gamma(.)$ is the gamma function and $\mathcal{V}_d$ is the volume of the unit sphere (e.g., $V_d = \text{Vol}(S_d)$; $S_d = \{ \xi \in \mathbb{R}^d : ||\xi|| \leq 1 \}$).

Furthermore, for the case when the observed process, $Y^T, Y_t \in \mathbb{R}^d$, and the signal to be controlled, $H^T, H_t \in \mathbb{R}^d$, are related by $Y_t = H_t + Y_t$, (e.g., $\rho(H_t, 0) = \rho(Y_t - Y_t, 0)$), then (IV.12) and (IV.13) are also necessary conditions for the stability of the sequence $H^T, H_t \in \mathbb{R}^d$ in probability and $r$-mean.

Proof: Follows from Lemma 3.6 and Theorem 3.5

V. CONCLUSION AND FUTURE WORKS

Future work will investigate i) the existence of a separation theorem between the design of the communication and control subsystems, ii) the tightness of the lower bound, iii) necessary and sufficient conditions for existence of an encoder, decoder, controller for control/communication systems. Some of the answers will be addressed via generalizations of the direct and converse Shannon theorems.

REFERENCES


