Information Nonanticipative Rate Distortion Function and its Applications

Photios A. Stavrou, Christos K. Kourtellaris, and Charalampos D. Charalambous

Abstract

This paper investigates applications of nonanticipative Rate Distortion Function (RDF) in zero-delay Joint Source-Channel Coding (JSCC) based on excess distortion probability, in bounding the Optimal Performance Theoretically Attainable (OPTA) by noncausal and causal codes, and in computing the Rate Loss (RL) of zero-delay and causal codes with respect to noncausal codes. These applications are described using two running examples, the Binary Symmetric Markov Source with parameter $p$, (BSMS($p$)) and the multidimensional partially observed Gaussian-Markov source.

For the BSMS($p$), the solution of the nonanticipative RDF is derived, the RL of causal codes with respect to noncausal codes is computed, and an uncoded noisy coding theorem based on excess distortion probability is shown. For the multidimensional Gaussian-Markov source, the solution to the nonanticipative RDF is derived, its operational meaning using JSCC via a noisy coding theorem is shown by providing the optimal encoding-decoding scheme over a vector Gaussian channel, and the RL of causal and zero-delay codes with respect to noncausal codes is computed.

Further, to facilitate the applications of information nonanticipative RDF for abstract alphabet sources, existence of the optimal reproduction conditional distribution is showing using the topology of weak convergence of probability measures, the optimal stationary solution to the information nonanticipative RDF is provided, and several properties of the solution are derived including a new characterization which leads to a lower bound, analogous to the Shannon lower bound of the classical RDF.

The information nonanticipative RDF is shown to be equivalent to the Gorbunov and Pinsker $\epsilon$-entropy, which corresponds to the classical RDF with an additional causality or nonanticipative condition imposed on the optimal reproduction conditional distribution.

This work was financially supported by a medium size University of Cyprus grant entitled "DIMITRIS" and by QNRF, a member of Qatar Foundation, under the project NPRP 6-784-2-329. Part of the material in this paper was accepted for publication in IEEE International Symposium on Information Theory, 2014 [1].

The authors are with the Department of Electrical and Computer Engineering (ECE), University of Cyprus, 75 Kallipoleos Avenue, P.O. Box 20537, Nicosia, 1678, Cyprus, e-mail: {stavrou.fotios,kourtellaris.christos,chadcha@ucy.ac.cy}
Index Terms

Nonanticipative RDF, sources with memory, joint source-channel coding, binary symmetric Markov source, multidimensional Gaussian-Markov source, bounds.

CONTENTS

I Introduction
   I-A Motivation for Nonanticipative RDF ...................................... 6
   I-B Summary of Main Results and Related Literature ..................... 9

II Information Nonanticipative RDF on Abstract Spaces ............... 14
   II-A Equivalent Causal Conditioning Distributions ....................... 15
   II-B Information Nonanticipative RDF ...................................... 18

III Existence of Information Nonanticipative RDF and Relations to Nonanticipatory $\epsilon$-Entropy ............................. 20
   III-A Finite Horizon ............................................................... 21
   III-B Equivalence of Nonanticipative RDF and Nonanticipatory $\epsilon$-Entropy . 23
   III-C Infinite Horizon ............................................................ 25

IV Optimal Reproduction of Nonanticipative RDF, Properties, and Examples ......................... 28
   IV-A Stationary optimal reproduction distribution ....................... 29
   IV-B Properties of Nonanticipative RDF .................................. 31
   IV-C Examples of Nonanticipative RDF .................................... 33

V Coding Theorems and Bounds .................................................. 42
   V-A Noisy Coding Theorem: Nonanticipative Transmission ............ 42
   V-B Noiseless Coding Theorem ............................................... 50
   V-C Bounds on OPTA by Causal Codes ..................................... 51

VI Conclusion ................................................................. 55

Appendix A: Proof of Theorem III.2 ........................................... 56
I. Introduction

In lossy compression source coding with fidelity [2], [3], the sequence of real-valued symbols $X^\infty \triangleq \{X_0, X_1, \ldots\}$, $X_j \in \mathcal{X}$, $\forall j \geq 0$, generated by a source with distribution $P_{X^\infty}$, is transformed by the encoder into a sequence of symbols, the compressed representation $Z^\infty \triangleq \{Z_0, Z_1, \ldots\}$ (taking values in a finite alphabet set), which is then transmitted over a noiseless channel. The decoder at the channel output upon observing the compressed representation symbols produces the reproduction sequence $Y^\infty \triangleq \{Y_0, Y_1, \ldots\}$, $Y_j \in \mathcal{Y}$, $\forall j \geq 0$. Such a compression system is called causal [4] if the reproduction symbol $Y_n$ of the source symbol $X_n$, depends on the present and past source symbols $\{X_0, \ldots, X_n\}$ but not on the future source symbols $\{X_{n+1}, X_{n+2}, \ldots\}$. The cascade of the encoder-decoder is often called the reproduction coder, and it is a family of measurable functions such that $Y_n \triangleq f_n(X_0, \ldots, X_n)$, while the compressed representation itself may be noncausal and have variable rate [4]. Consequently, the decoder can generate the reproductions with arbitrary delay. The Optimal Performance Theoretically Attainable (OPTA) by noncausal codes is described via the Rate Distortion Function (RDF) [2], [3], while that of causal codes is described via the entropy rate of the reproduction sequence [4].

Zero-delay source coding is a sub-class of causal coding, with the additional constraint that the compressed representation symbol $Z_n$, depends on the past and present source symbols $X^n \triangleq \{X_0, X_1, \ldots, X_n\}$, while the reproduction at the decoder $Y_n$ of the present source symbol $X_n$, depends only on the compressed representation $Z^n \triangleq \{Z_0, Z_1, \ldots, Z_n\}$. Thus, a zero-delay coding system consists of a family of encoding-decoding measurable functions such that $Z_i = h_i(\{X_j : j = 0, 1, \ldots, i\})$ and $Y_i = f_i(\{Z_j : j = 0, 1, \ldots, i\})$, $\forall i \geq 0$ [5]–[9].

Joint Source-Channel Coding (JSCC) based on nonanticipative processing (i.e., the encoder-channel-decoder process at each time instant symbols causally), is perhaps, the most efficient zero-delay coding system, in the sense of optimal performance of matching the source characteristics to the channel characteristics, coded or uncoded [10]. Two such fascinating examples are a) the Independent Identically Distributed (IID) binary source with Hamming distortion transmitted uncoded over a symmetric memoryless channel (the distortion and channel parameter are made equal), and b) the IID Gaussian source with average squared-error distortion, transmitted over an Additive White Gaussian Noise (AWGN) channel, with the encoder and decoder scaling their
inputs. These examples demonstrate the simplicity of the JSCC system, in operating optimally
with zero-delay, that is, encoding information Symbol-by-Symbol (ShS), in complexity, when
this is compared to the asymptotic performance of optimally separating the encoder/decoder to
the source and channel encoders/decoders which may cause long processing delays.

In general, very little is known about JSCC based on nonanticipative processing and the OPTA
by causal and zero-delay codes. Often, bounds are introduced to quantify the Rate Loss (RL)
due to causality and zero-delay of the coding systems compared to that of the noncausal coding
systems. Such bounds are elaborated recently in [11], for Gaussian sources with square-error
distortion function. In many delay sensitive applications of lossy compression, limited end-to-
end decoding delay is often desirable, while for real-time systems, such as, communication for
control over finite rate channels [12]–[16], and in general, for systems involving feedback [17],
causal and more importantly zero-delay coding is preferable to noncausal coding.

In this paper, we consider information nonanticipative RDF [18], [19], a variant of the classical
information RDF, and we describe its applications in JSCC based on nonanticipative processing.
We show achievability by deriving a noisy coding theorem using coded and uncoded transmission,
with respect to the excess distortion probability (between the source and its reproduction).
Moreover, we use it to compute the Rate Loss (RL) or gap between the OPTA by causal [4]
and zero-delay codes with respect to the OPTA by noncausal codes, for both finite alphabet and
continuous alphabet valued sources. We also show equivalence of the information nonanticipative
RDF and its rate to Gorbunov and Pinsker nonanticipatory $\epsilon-$entropy and message generation
rates [20]–[22], which corresponds to Shannon information RDF with an additional causality
constraint imposed on the optimal reproduction distribution. Throughout the paper we illustrate
the achievability of JSCC to two running examples, the Binary Symmetric Markov Source
BSMS($p$) (with parameter $p$) and the multidimensional partially observed Gaussian-Markov
stationary source.

Further, to facilitate the application of the information nonanticipative RDF as explained above,
we establish the following additional results. Show existence of the optimal reproduction distri-
bution of the information nonanticipative RDF and its rate, give the expression of the optimal
reproduction for stationary processes and characterize several of its properties, and use it to
compute the nonanticipative RDF for the two running examples, the BSMS($p$) and the multidi-
mensional partially observed Gaussian-Markov stationary source.

The information nonanticipative RDF is defined as follows. Consider a source distribution \( P_{X^n}(dx^n) \), a causal sequence of reproduction distributions \( \{ P_{Y_i|Y_{i-1},X^i}(dy_i|y^{i-1},x^i) : i = 0, 1, \ldots, n \} \), a measurable distortion function \( d_{0,n}(x^n, y^n) : \mathcal{X}_{0,n} \times \mathcal{Y}_{0,n} \rightarrow [0, \infty] \), and an average fidelity set

\[
Q_{0,n}(D) \triangleq \left\{ P_{Y^n|X^n}(dy^n|x^n) \triangleq \otimes_{i=0}^n P_{Y_i|Y_{i-1},X^i}(dy_i|y^{i-1},x^i) : \right. \\
\left. \frac{1}{n+1} \int_{\mathcal{X}_{0,n} \times \mathcal{Y}_{0,n}} d_{0,n}(x^n, y^n)(P_{Y^n|X^n} \otimes P_{X^n})(dx^n, dy^n) \leq D \right\} .
\]

The information nonanticipative RDF is defined by

\[
R_{0,n}^{na}(D) \triangleq \inf_{P_{Y^n|X^n} \in Q_{0,n}(D)} \int_{\mathcal{X}_{0,n} \times \mathcal{Y}_{0,n}} \log \left( \frac{P_{Y^n|X^n}(dy^n|x^n)}{P_{Y^n}(dy^n)} \right) (P_{Y^n|X^n} \otimes P_{X^n})(dx^n, dy^n) = \inf_{P_{Y^n|X^n} \in Q_{0,n}(D)} \mathbb{I}_{X^n \rightarrow Y^n}(P_{X^n}, P_{Y^n|X^n}).
\]

Here, \( \mathbb{I}_{X^n \rightarrow Y^n}(\cdot, \cdot) \) is used to denote the functional dependence of \( R_{0,n}^{na}(D) \) on the two distributions \( \{ P_{X^n}, P_{Y^n|X^n} \} \). The information nonanticipative RDF rate is defined by

\[
R^{na}(D) \triangleq \lim_{n \rightarrow \infty} \frac{1}{n+1} R_{0,n}^{na}(D)
\]

provided the limit exists. Unlike classical RDF, the optimal reproduction distribution of the information nonanticipative RDF is a sequence of conditional distributions \( \{ P_{Y_i|Y_{i-1},X^i}(dy_i|y^{i-1},x^i) : i = 0, 1, \ldots, n \} \), hence at each time \( i \), it is causal with respect to the past and present source symbols and past reproduction symbols \( \{ X^i, Y^{i-1} \} \), \( i = 0, 1, \ldots, n \).

A. Motivation for Nonanticipative RDF

Next, we discuss some limitations of the classical information RDF with respect to its computation, and its applications to JSCC based on nonanticipative and SbS transmission, which

\( \odot \) denotes convolution of distributions.
motivated our interest in the information nonanticipative RDF \((I.3)\).

Recall that the classical information RDF with respect to the fidelity set of reproduction conditional distributions defined by

\[
R(D) \triangleq \lim_{n \to \infty} \frac{1}{n+1} R_{0,n}(D), \quad R_{0,n}(D) \triangleq \inf_{P_{Y^n|X^n}(x^n) \in \mathcal{Q}_{0,n}(D)} I(X^n; Y^n). \tag{I.5}
\]

where

\[
\mathcal{Q}_{0,n}(D) \triangleq \left\{ P_{Y^n|X^n}(dy^n|x^n) : \frac{1}{n+1} \int_{X^n \times Y^n} d_{0,n}(x^n, y^n)(P_{Y^n|X^n} \otimes P_{X^n})(dx^n, dy^n) \leq D \right\} \tag{I.6}
\]

Under general conditions \([2], [3], [23]\), it is already known that if the infimum over \(\mathcal{Q}_{0,n}(D)\) exists, then the limit \(R(D) = \lim_{n \to \infty} \frac{1}{n+1} R_{0,n}(D)\) exists, and \(R(D)\) is the OPTA by noncausal codes \([2]\). Moreover, it is also known that the optimal conditional distribution achieving the infimum in \((I.5)\) is given by the implicit expression

\[
P_{Y^n|X^n}^*(dy^n|x^n) = e^{s d_{0,n}(x^n, y^n)} P_{Y^n}(dy^n) \quad \int_{Y^n} e^{s d_{0,n}(x^n, y^n)} P_{Y^n}(dy^n), \quad s \leq 0 \tag{I.7}
\]

where \(s \in (-\infty, 0]\) is the Lagrange multiplier associated with the fidelity constraint \(\mathcal{Q}_{0,n}(D)\).

Although, establishing a noiseless coding theorem giving an operational meaning to \(R(D)\) as the OPTA by noncausal codes is by now standard, when applied to sources with memory, \(R(D)\) has certain limitations.

The first limitation of the classical information RDF is the computational complexity of obtaining the exact expression of \(R_{0,n}(D)\) and \(P_{Y^n|X^n}^*(dy^n|x^n)\), for finite \(n\), and that of \(R(D) \triangleq \lim_{n \to \infty} \frac{1}{n+1} R_{0,n}(D)\), even for stationary sources with memory. In fact, the exact expression of \(R(D)\) is only known for a small class of sources, which are either memoryless or Gaussian. For example, the exact characterization of the OPTA by noncausal codes of the BSMS\((p)\) with single letter Hamming distortion function is currently unknown; more precisely, the value of \(R(D)\) is only known for the distortion region \(0 \leq D \leq D_c, D_c = \frac{1}{2} \left(1 - \sqrt{1 - (\frac{q}{p})^2}\right)\), \(p = 1 - q, q \leq \frac{1}{2}\), and only bounds are available for \(D\) beyond \(D_c\) \([24]–[26]\).

The second limitation of the classical information RDF is the noncausal or anticipative nature of the optimal reproduction distribution \((I.7)\), which implies that for any \(n\), the reproduction at time \(i \leq n\) of \(x_i \in \mathcal{X}_i\) by \(y_i \in \mathcal{Y}_i\) depends on the past, present and future source symbols (i.e., its is noncausal), even for single letter distortion functions \(d_{0,n} \triangleq \sum_{i=0}^{n} \rho(x_i, y_i)\). The noncausality
of the optimal reproduction distribution (I.7) follows directly by applying Bayes’ rule applied in (I.7), which yields

\[ P_{Y^n|X^n}(dy^n|x^n) = \bigotimes_{i=0}^{n} P_{Y_i|Y_{i-1},X^n}(dy_i|y_{i-1}, x^n) \]  \hspace{1cm} (I.8)

\[ \neq \bigotimes_{i=0}^{n} P_{Y_i|Y_{i-1},X_i}(dy_i|y_{i-1}, x_i). \]  \hspace{1cm} (I.9)

Therefore, probabilistically, the optimal reproduction distribution (I.7) of the classical information RDF cannot be decomposed into a convolution of causal conditional distribution (i.e., it is anticipative). The anticipation of the optimal reproduction distribution (I.7) (or failure of (I.9) to hold) implies that, in general, the classical information RDF cannot be used in JSCC using nonanticipative processing or uncoded transmission, processing each symbol causally, also called probabilistic matching of the source to the channel [27], unless the source is memoryless, such as, the binary memoryless source transmitted over a binary symmetric channel [10], [28], [29]. Indeed, a necessary condition for JSCC using nonanticipative processing and probabilistic matching of the source to the channel is the realization of the optimal reproduction distribution by an encoder-channel-decoder which process symbols causally, as shown in Fig. I.1. For such realization to be feasible it is necessary that (I.8) equals (I.9), or equivalently, the following causality constraint expressed in terms of Markov chains (MC) should hold.

\[ X_{i+1}^{\infty} \leftrightarrow (X^i, Y^{i-1}) \leftrightarrow Y_i, \; i = 0, 1, \ldots \]  \hspace{1cm} (I.10)

Among all the classes of sources the only subclass for which the optimal reproduction distribution (I.7) of the classical information RDF is known to be nonanticipative (i.e., causal with respect
to future source symbols) is the independent source \( \{X_n: n = 0, 1, \ldots\} \) with single letter distortion. In this case, \( P_{Y | X^n}(dy^n|x^n) = \otimes_{i=0}^nP_{Y_i|X_i}(dy_i|x_i) - \text{a.a.} \) \( x^n \in \mathcal{X}_{0,n} \) and hence \( P_{Y_i|X_i}(dy_i|x_i) \) satisfies the necessary condition for JSCC or probabilistic matching of the source to the channel (memoryless) via nonanticipative processing. The realization of the optimal reproduction distribution as shown in Fig. I.1 is necessary for JSCC using nonanticipative transmission, and it is fundamental in the two examples mentioned earlier (see also [10], [27]), e.g., the binary IID source with a Hamming distortion, and the IID Gaussian source with mean-square distortion.

B. Summary of Main Results and Related Literature

Next, we present a summary of the contributions in this paper and their relations to existing literature.

**Information Nonanticipative RDF.** In Section II we formulate the information nonanticipative RDF on abstract spaces (Polish spaces), so that our analysis hold for both finite and continuous alphabets (because we present examples for both), and then we invoke certain result from [30], to define the information measure of the information nonanticipative RDF, \( \mathbb{I}_{X^n \rightarrow Y^n}(P_{X^n}, \frac{P_{Y^n|X^n}}{P_{Y^n|X^n}}) \), which is a special case of directed information [31], and the information nonanticipative RDF (I.3), and its rate (I.4). Since \( Q^{C1}_{0,n}(D) \subseteq Q_{0,n}(D) \) we immediately deduce the bounds

\[
R^{na}(D) \geq R(D), \quad R^*_{0,n}(D) \geq R_{0,n}(D), \quad \forall n \geq 0.
\]  

(I.11)

Note that these bounds hold with equality for memoryless sources with single letter distortion, because the optimal reproduction distribution of \( R_{0,n}(D) \) satisfies \( P_{Y^n|X^n}(dy^n|x^n) = \otimes_{i=0}^nP_{Y_i|X_i}(dy_i|x_i) \). This raises question whether \( R^{na}(D) \) is a tight bound on the OPTA by noncausal codes. In Section IV-C we derive the exact expression of \( R^{na}(D) \) for the BSMS\((p)\), and then we compare the bound \( R^{na}(D) \geq R(D) \) with the exact expression of \( R(D) \) which is only known for \( 0 \leq D \leq D_c \), and the upper bound on \( R(D) \) given in [25], illustrating the tightness of bound (I.11). Further, since we can show via the converse coding theorem that \( R^{na}(D) \) is a lower bound on the OPTA by causal codes [4], then we also use \( R^{na}(D) \) to evaluate the RL of causal codes with respect to noncausal codes.

**Existence of Information Nonanticipative RDF.** In Section III we first establish existence of the optimal nonanticipative distribution achieving the infimum in \( R^{na}_{0,n}(D) \) by invoking the
topology of weak convergence of probability measures and Prohorov’s theorems [32], [33]. This material generalizes in a nontrivial manner, the existence theorem derived in [23] for the classical RDF.

Second, we show that the information nonanticipative RDF, $R^\alpha_{0,n}(D)$, and its rate $R^\alpha(D)$, are equivalent to Gorbunov and Pinsker nonanticipatory $\epsilon$-entropy and message generation rate [20], denoted by $R^\alpha_{0,n}(D)$ and $R^\epsilon(D)$, respectively, and defined by

$$R^\epsilon(D) \triangleq \lim_{n \to \infty} \frac{1}{n+1} R^\epsilon_{0,n}(D) \equiv R^\alpha_{0,n}(D),$$

$$R^\epsilon_{0,n}(D) \triangleq \inf_{P_{X^n|Y^n} \in \mathcal{Q}_{0,n}(D)} \frac{1}{n+1} I(X^n;Y^n), \quad \forall n \geq 0. \quad (I.12)$$

We establish this equivalence by showing that the following Markov Chains (MCs) are equivalent.

$$X^{n+1}_i \leftrightarrow X^i \leftrightarrow Y^i \iff X^{n+1}_i \leftrightarrow (X^i, Y^{i-1}) \leftrightarrow Y^i, \quad i = 0, \ldots, n-1, \ n \geq 0. \quad (I.14)$$

The fact that the Left Hand Side (LHS) MC in (I.14) implies the Right Hand Side (RHS) is already known; the equivalence of the two statements is new and has not been documented elsewhere. Armed with the existence result, we further show that for general stationary sources the limit exists, and the limit and infimum operations can be interchanged, that is,

$$R^\alpha(D) = \lim_{n \to \infty} \inf_{P_{Y^n|X^n} \in \mathcal{Q}_{0,n}(D)} \frac{1}{n+1} \mathbb{H}_{X^n \to Y^n}(P_{X^n}, \overrightarrow{P_{Y^n|X^n}})$$

$$= \inf_{P_{Y^n|X^n} \in \mathcal{Q}_{0,n}(D)} \lim_{n \to \infty} \frac{1}{n+1} \mathbb{H}_{X^n \to Y^n}(P_{X^n}, \overrightarrow{P_{Y^n|X^n}}) \equiv R^\alpha_{0,n}(D) < \infty. \quad (I.15)$$

Moreover, using the existence of solution of the nonanticipative RDF derived in this paper, we consider consistent stationary sources (as defined by Gorbunov and Pinsker in [20]), and we invoke certain results which are derived therein and in [30] to establish that the optimal reproduction distribution for $R^\alpha_{0,n}(D)$ is realizable by jointly stationary source-reproduction pair $\{(X_i, Y_i) : i = 0, 1, \ldots\}$. Therefore, we give general sufficient conditions for existence of solution of the information nonanticipative RDF, its limit, and its stationary behaviour.

**Closed Form Expression of Information Nonanticipative RDF (Stationary Solution), Properties, and Examples.** Section [IV] consists of three subsections. In Section [IV-A] we consider stationary source-reproduction pairs and single letter distortion functions $d_{0,n} \triangleq \sum_{i=0}^n \rho(x_i, y_i)$, and we recall the expression of the optimal nonanticipative reproduction distribution for $R^\alpha_{0,n}(D)$.
[18] Section IV] given by
\[
\overrightarrow{P}_{Y^n|X^n}(dy^n|x^n) = \bigotimes_{i=0}^{n} P_{Y_i|Y_{i-1},X_i}(dy_i|y_{i-1},x_i) = \bigotimes_{i=0}^{n} e^{s_\rho(x_i,y_i)} P_{Y_i|Y_{i-1}}^*(dy_i|y_{i-1})
\]
and the closed-form expression for \( R_{0,n}^a(D) \).

In subsection [IV-B] we derive several properties of the nonanticipative RDF, which are the analogues to the properties of classical RDF [2], although their derivation and presentation is more involved. Some of these properties are employed in [34] to construct an algorithm for computing \( R_{m,n}^a(D) \), which is analogous to Blahut-Arimoto Algorithm (BAA) for single letter classical RDF, \( R(D) \) [35]. One of the properties is analogous to the Shannon lower bound of the classical RDF [2].

In subsection [IV-C] we invoke expression (I.16) together with some properties of the solution (presented in Section [IV-B]) to compute the information theoretic nonanticipative RDF in closed form for the following two running examples.

**BSMS(\( p \)) with single letter Hamming distortion.** We consider the BSMS(\( p \)) with single letter Hamming distortion we show that
\[
R_{m,n}^a(D) = \begin{cases} 
H(m) - H(D) & \text{if } D \leq \frac{1}{2}, m = 1 - p - D + 2pD \\
0 & \text{otherwise}
\end{cases}
\]

Note that, for \( p = \frac{1}{2} \), then \( R_{m,n}^a(D) \) reduces to the classical RDF, \( R(D) \), of IID memoryless source, as expected.

**Multidimensional Gaussian-Markov source with square-error distortion and JSCC.** We consider the multidimensional partially observed Gauss-Markov source (system), described in state space form by
\[
\begin{cases}
Z_{t+1} = AZ_t + BW_t, \ Z_0 = z, \ t = 0,1,\ldots \\
X_t = CZ_t + NV_t, \ t = 0,1,\ldots
\end{cases}
\] (I.17)

where \( Z_t \in \mathbb{R}^m \) is the state (unobserved) process driven by a multidimensional Gaussian noise process \( \{W_t: \ t = 0,1,\ldots\} \), and \( X_t \in \mathbb{R}^p \) is the observed source process corrupted by a multidimensional additive Gaussian noise process \( \{V_t: \ t = 0,1,\ldots\} \). In this application, the objective is to compress the data collected by the observation device generating, \( \{X_t: \ t = 0,1,\ldots,n\} \) is ergodic although
the $A$ matrix of $\{Z_t: t = 0, 1, \ldots, n\}$ may have unstable eigenvalues) we use the expression of reproduction distribution (I.16) and some of its properties to show that for a square error distortion function, the nonanticipative RDF conjectured in [18] is given by

$$R^{na}(D) = \frac{1}{2} \sum_{i=0}^{p} \log \left( \frac{\lambda_{\infty},i}{\delta_{\infty},i} \right) = \frac{1}{2} \log \left| \frac{\Lambda_{\infty}}{\Delta_{\infty}} \right|$$  \hspace{1cm} (I.18)

where

$$\Lambda_{\infty} = \text{diag}\{\lambda_{\infty,1}, \ldots, \lambda_{\infty,p}\}, \Delta_{\infty} = \text{diag}\{\delta_{\infty,1}, \ldots, \delta_{\infty,p}\}$$

are the steady state eigenvalues of the covariance of the error process, defined by $\Lambda_{\infty} \triangleq \lim_{n \to \infty} \Lambda_t$, $\Lambda_t \triangleq \mathbb{E}\left\{ \left( X_t - \mathbb{E}(X_t|Y^{t-1}) \right) \left( X_t - \mathbb{E}(X_t|Y^{t-1}) \right)^{tr} \right\}$, and

$$\delta_{\infty,i} \triangleq \begin{cases} \xi_{\infty} & \text{if } \xi_{\infty} \leq \lambda_{\infty,i} \\ \lambda_{\infty,i} & \text{if } \xi_{\infty} > \lambda_{\infty,i} \end{cases}, \quad i = 1, \ldots, p, \quad \sum_{i=1}^{p} \delta_{\infty,i} = D.$$ 

Further, we show that (I.18) is achievable using JSCC over a linear encoder and decoder, operating at zero-delay, over an additive Gaussian noise vector channel.

In Remark [V.11], we specialize (I.17) to the scalar Gaussian process (i.e., $X_t = Z_t$, $A = \alpha$, $B = 1$, $N = 0$, $W_t \sim \mathcal{N}(0, \sigma_w^2)$), and from (I.18) we obtain the well known expression

$$R^{na}(D) = \frac{1}{2} \ln(\alpha^2 + \frac{\sigma_w^2}{D})$$  \hspace{1cm} (I.19)
operational meaning to $R_{0,n}^a(D)$.

In Section V-C, we use $R_{n}^a(D)$ to derive bounds for the OPTA by causal codes [4], and the OPTA of noncausal codes. This is done by recalling Neuhoff and Gilbert [4] definition of OPTA by causal codes, denoted by $r_{c,+}(D)$, and the bounds

$$r_{c,+}(D) \geq R_{n}^a(D) \geq R(D). \quad (I.20)$$

Consequently, by analyzing and computing the exact expression of the nonanticipative RDF, $R_{n}^a(D)$, and its reproduction distribution, we are able to obtain bounds on the OPTA by noncausal codes, given by $R(D)$, and the OPTA by causal codes, given by $r_{c,+}(D)$, and evaluate the RL due to causality even for sources with memory. This part compliments previous work by Linder and Zamir [36], who showed using the results in [4], that at high resolution (small distortion), an optimal causal code for stationary source with finite differential entropy and square error distortion, consists of uniform quantizers followed by a sequence of entropy coders, and that the RL due to causality is given by the so-called space-filling loss of quantizers, which is at most

$$\frac{1}{2} \ln \left( \frac{2\pi e}{12} \right) \approx 0.254 \text{ bits/sample.}$$

Moreover, for arbitrary Gaussian stationary sources with memory and square-error distortion, we compute $R_{n}^a(D)$ explicitly, and then we use (I.20) to evaluate the RL due to causality. Our analysis on vector Gaussian stationary sources compliment Gorbunov and Pinsker computation in [21], [22], where it is shown (for scalar Gaussian sources), using power spectral densities that $R^\varepsilon(D)$ tends to Shannon’s RDF of noncausal codes, as $D \rightarrow 0$.

Along this direction, our results compliment recent contributions obtained in [11], on the gap between the OPTA by causal codes, nonanticipatory $\epsilon$-entropy, $R^\varepsilon(D)$, and classical RDF, $R(D)$, for stationary Gaussian sources with square-error distortion. Specifically, in [11] it is shown that for zero-mean Gaussian sources with square-error distortion (and bounded differential entropy rate), the OPTA by causal codes exceeds $R^\varepsilon(D)$ by less than approximately 0.254 bits/sample. The analysis in [11] includes a closed-form expression for $R^\varepsilon(D)$ when the source is first-order Markov with square-error distortion, while for general Gaussian sources it is shown that $r_{c,+}(D) \leq R^\varepsilon(D) + \frac{1}{2} \log_2(2\pi e)$ bits/sample. In addition, to analyze the gap between causal and noncausal codes for arbitrary Gaussian stationary sources, in [11] another nonanticipative information measure, namely, $R^\varepsilon(D)$, is introduced, which is then used to bound $R^\varepsilon(D)$ by $R^\varepsilon(D) \leq R^\varepsilon(D)$, in order to evaluate the gap $R^\varepsilon(D) - R(D) \leq R^\varepsilon(D) - R(D)$, via $R^\varepsilon(D)$. However, in [11] no closed-form expression for $R^\varepsilon(D)$ or $R^\varepsilon(D)$ is provided (except for first-
order Gauss-Markov source). Our results complement the existing bounds in [11], in the sense that we provide the optimal expression of the reproduction distribution of $R^\epsilon(D)$, which can be used to evaluate the gap between causal and noncausal codes for arbitrary Gaussian sources (partially observed).

In addition, we also show that for BSMS($p$) the RL due to causality, i.e., $R^{na}(D) - R(D)$, does not exceed $H(m) - H(p)$, for the region $0 \leq D \leq D_c$, $m = 1 - p - D + 2pD$, while for the region $D_c \leq D \leq \frac{1}{2}$, we compare the upper bound given in [25] with the upper bound of $R^{na}(D)$. Our simulations and properties of $R^{na}(D)$ illustrate that an upper bound on $R(D)$ based on $R^{na}(D)$ is more reliable compared to that in [25].

For multidimensional partially observable Gaussian stationary source we utilize the bounds $R^{na}(D) - R(D) \leq r^{c,+}(D) - R(D)$ and we deduce that the RL due to zero-delay codes does not exceed $\frac{1}{2} \log \frac{|\Lambda_\infty|}{|\Delta_\infty|} - R(D)$, where $R(D)$ can be computed using power spectral densities [2].

The paper is structured as follows. In Section II, we define the information nonanticipative RDF, while in Section III, we show existence of optimal solution to the information nonanticipative RDF and relate this measure to nonanticipatory $\epsilon$-entropy. In Section IV, we provide the optimal stationary reproduction distribution of nonanticipative RDF and we derive several of its properties. Here, we also present the two running examples, the BSMS($p$) with Hamming distortion and the multidimensional partially observed Gaussian-Markov source, to illustrate the various applications of information nonanticipative RDF in the evaluation of RDF for sources with memory and in JSCC. In Section V, we give an operational meaning to nonanticipative RDF and rate, by deriving a noisy coding theorem based on nonanticipative coded and uncoded transmission, for sources with memory, with respect to the excess distortion probability, and we then apply this to the BSMS($p$). In this section, we also derive bounds on the OPTA by causal codes and we evaluate these bounds for the two running examples of Section V.

II. INFORMATION NONANTICIPATIVE RDF ON ABSTRACT SPACES

In this section, we first introduce the definition of information nonanticipative RDF, for general source/reproduction alphabets modelled by Polish spaces (complete separable metric spaces). Past work utilizing Polish spaces and Prohorov’s theorem under the topology of weak convergence for single letter capacity without feedback and classical RDF analysis is found in [23], [37]. Our construction of various probability distributions is based on the methodology presented in [30],
where we deal with arbitrary \((n + 1)\)-fold convolution measures. In this section, we also state some properties of information nonanticipative RDF which follow directly from [30].

A. Equivalent Causal Conditioning Distributions

Let \(\mathbb{N} \triangleq \{0, 1, 2, \ldots\\} \) and \(\mathbb{N}^n \triangleq \{0, 1, 2, \ldots, n\}\). Introduce the source spaces \(\{(\mathcal{X}_n, \mathcal{B}(\mathcal{X}_n)) : n \in \mathbb{N}\}\) and the reproduction spaces \(\{(\mathcal{Y}_n, \mathcal{B}(\mathcal{Y}_n)) : n \in \mathbb{N}\}\) where \(\mathcal{X}_n, \mathcal{Y}_n, n \in \mathbb{N}\) are Polish spaces, and \(\mathcal{B}(\mathcal{X}_n)\) and \(\mathcal{B}(\mathcal{Y}_n)\) are Borel \(\sigma\)-algebras of subsets of \(\mathcal{X}_n\) and \(\mathcal{Y}_n\) respectively. Points in \(\mathcal{X}^n \triangleq \times_{n \in \mathbb{N}} \mathcal{X}_n\), \(\mathcal{Y}^n \triangleq \times_{n \in \mathbb{N}} \mathcal{Y}_n\) are denoted by \(x \triangleq \{x_0, x_1, \ldots\} \in \mathcal{X}^n\), \(y \triangleq \{y_0, y_1, \ldots\} \in \mathcal{Y}^n\), respectively, while their restrictions to finite coordinates for any \(n \in \mathbb{N}\) are denoted by \(x^n \triangleq \{x_0, x_1, \ldots, x_n\} \in \mathcal{X}_{0,n}\), \(y^n \triangleq \{y_0, y_1, \ldots, y_n\} \in \mathcal{Y}_{0,n}\).

Let \(\mathcal{B}(\mathcal{X}^n) \triangleq \circledast_{i \in \mathbb{N}} \mathcal{B}(\mathcal{X}_i)\) denote the \(\sigma\)-algebra on \(\mathcal{X}^n\) generated by cylinder sets \(\{x = (x_0, x_1, \ldots) \in \mathcal{X}^n : x_0 \in A_0, x_1 \in A_1, \ldots, x_n \in A_n\}, A_i \in \mathcal{B}(\mathcal{X}_i), 0 \leq i \leq n, n \geq 1\), and similarly for \(\mathcal{B}(\mathcal{Y}^n) \triangleq \circledast_{i \in \mathbb{N}} \mathcal{B}(\mathcal{Y}_i)\). Thus, \(\mathcal{B}(\mathcal{X}_{0,n})\) and \(\mathcal{B}(\mathcal{Y}_{0,n})\) denote the \(\sigma\)-algebras of cylinder sets in \(\mathcal{X}^n\) and \(\mathcal{Y}^n\), respectively, with bases over \(A_i \in \mathcal{B}(\mathcal{X}_i)\) and \(B_i \in \mathcal{B}(\mathcal{Y}_i), 0 \leq i \leq n\), respectively.

Source Distribution. The source is specified by the collection of functions \(\{p_n(dx_n; x^{n-1}) : n \in \mathbb{N}\}\), which satisfies the following conditions.

i) For \(n \in \mathbb{N}\), \(p_n(\cdot; x^{n-1})\) is a probability measure on \(\mathcal{B}(\mathcal{X}_n)\);

ii) For \(n \in \mathbb{N}\), \(A_n \in \mathcal{B}(\mathcal{X}_n), p_n(A_n; x^{n-1})\) is \(\bigcircledast_{i=0}^{n-1} \mathcal{B}(\mathcal{X}_i)\)–measurable in \(x^{n-1} \in \mathcal{X}_{0,n-1}\).

Thus, for each \(n \in \mathbb{N}\), \(p_n(dx_n; x^{n-1})\) is a stochastic kernel on \((\mathcal{X}_n, \mathcal{B}(\mathcal{X}_n))\) given \((\mathcal{X}_{0,n-1}, \mathcal{B}(\mathcal{X}_{0,n-1}))\), denoted by \(Q_n(\mathcal{X}_n; \mathcal{X}_{0,n-1})\), and hence for \(B \in \mathcal{B}(\mathcal{X}_{0,n})\) a cylinder set of the form \(B \triangleq \{x \in \mathcal{X}^n : x_0 \in B_0, x_1 \in B_1, \ldots, x_n \in B_n\}\), \(B_i \in \mathcal{B}(\mathcal{X}_i), 0 \leq i \leq n\), we can define the family of measures \(P(\cdot)\) on \(\mathcal{B}(\mathcal{X}^n)\) via the \((n + 1)\)-fold convolution

\[
P(B) \triangleq \int_{B_0} p_0(dx_0) \ldots \int_{B_n} p_n(dx_n; x^{n-1}) \equiv \mu_{0,n}(B_{0,n}), \quad B_{0,n} = x_{i=0}^n B_i. \tag{II.1}
\]

Here, we use the notation \(\mu_{0,n}(\cdot)\) to denote the restriction of the measure \(P(\cdot)\) on the cylinder set \(B \in \mathcal{B}(\mathcal{X}_{0,n})\), for \(n \in \mathbb{N}\).

Reproduction Distribution. The reproduction channel is specified by a collection of functions \(\{q_n(dy_n; y^{n-1}, x^n) : n \in \mathbb{N}\}\) which satisfies the following conditions.

iii) For \(n \in \mathbb{N}\), \(q_n(\cdot; y^{n-1}, x^n)\) is a probability measure on \(\mathcal{B}(\mathcal{Y}_n)\);

iv) For \(n \in \mathbb{N}\), \(B_n \in \mathcal{B}(\mathcal{Y}_n), q_n(B_n; y^{n-1}, x^n)\) is \(\bigcircledast_{i=0}^{n-1} \mathcal{B}(\mathcal{Y}_i) \mathcal{B}(\mathcal{X}_i) \mathcal{B}(\mathcal{X}_n)\)–measurable.
function of $x^n \in X_{0,n}$ and $y^{n-1} \in Y_{0,n-1}$.

By our notation, for each $n \in \mathbb{N}$, $q_n(dy_n; y^{n-1}, x^n) \in Q_n(Y_n; Y_{0,n-1} \times X_{0,n})$ is a stochastic kernel, hence, a version of regular conditional distribution $P_{Y_n|Y^{n-1}, X^n}(dy_n|Y^{n-1} = y^{n-1}, X^n = x^n)$.

Given a cylinder set $C \triangleq \{y \in Y^N : y_0 \in C_0, y_1 \in C_1, \ldots, y_n \in C_n\}$, $C_i \in B(Y_i)$, $0 \leq i \leq n$, we can define a family of measures on $B(Y^N)$ by the $(n+1)$-fold convolution

$$Q(C|x) \triangleq \int_{C_0} q_0(dy_0; x_0) \ldots \int_{C_n} q_n(dy_n; y^{n-1}, x^n) \equiv \overrightarrow{Q}_{0,n}(C_{0,n}|x^n), \quad C_{0,n} = \times_{i=0}^{n} C_i. \quad \text{(II.2)}$$

Consequently, the RHS of (II.2) defines a consistent family of finite-dimensional distributions, and hence, there exists a unique measure on $(Y^N, B(Y^N))$ from which the collection of distributions $\{q_n(dy_n; y^{n-1}, x^n) : n \in \mathbb{N}\}$ is obtained by conditioning on appropriate events (this will be done shortly).

Moreover, the family of measures $Q(\cdot|x)$ on $(Y^N, B(Y^N))$ defined by (II.2) satisfies the following consistency condition.

**C1:** If $D \in B(Y_{0,n})$, then $Q(D|x)$ is $B(X_{0,n})$-measurable function of $x \in X^N$.

Since $\{Y_n : n \in \mathbb{N}\}$ are Polish spaces, it follows that for any family of measures $Q(\cdot|x)$ on $(Y^N, B(Y^N))$ satisfying consistency condition **C1**, we can construct a collection of functions $\{q_n(dy_n; y^{n-1}, x^n) : n \in \mathbb{N}\}$ satisfying conditions **iii** and **iv** which are connected with $Q(\cdot|x)$ via relation (II.2) [30]. Thus, we have two equivalent definitions of causal conditioning distribution of the reproduction channel. The first definition is described by a family of measures on $Q(\cdot|x)$ on $(Y^N, B(Y^N))$ via (II.2). The second equivalent definition is described by a family of measures $Q(\cdot|x)$ on $(Y^N, B(Y^N))$ satisfying the consistency condition **C1**.

Next, we construct the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and the sequences of RV’s $\{(X_i, Y_i) : i \in \mathbb{N}\}$ on it. Given the basic measures $P(\cdot)$ on $X^N$ defined by (II.1) and $Q(\cdot|x)$ on $Y^N$ defined by (II.2) satisfying consistency condition **C1**, we can construct a sequence of RV’s $\{(X_i, Y_i) : i \in \mathbb{N}\}$ or the collections of conditional distributions $\{p_n(dx_n; x^{n-1}) : n \in \mathbb{N}\}$ and $\{q_n(dy_n; y^{n-1}, x^n) : n \in \mathbb{N}\}$ and a probability space as follows.

Let $A^{(n)} = \{x : x_n \in A\}$, $A \in B(X_n)$, and let $P(A^{(n)}|B(X_{0,n-1}))$ denote the conditional probability of $A^{(n)}$ with respect to $B(X_{0,n-1})$ calculated on the probability space $(X^N, B(X^N), P(\cdot))$. 

May 8, 2014 DRAFT
Then
\[ P\{X_n \in A | X^{n-1} = x^{n-1}\} = P\{\{x : x_n \in A\} | B(\mathcal{X}^{0,n-1})\} = p_n(A_n; x^{n-1}) - a.a. x^{n-1} \in \mathcal{X}^{0,n-1}. \]

(II.3)

Hence, \( p_n(\cdot; \cdot) \in \mathcal{Q}(\mathcal{X}^n; \mathcal{X}^{0,n-1}) \) (e.g., satisfying i), ii) are determined from \( P(\cdot) \).

Let \( B^{(n)} = \{y : y_n \in B\}, B \in \mathcal{B}(\mathcal{Y}^n) \), and let \( Q(B^{(n)}|x) | \mathcal{B}(\mathcal{Y}^{0,n-1}) \) denote the conditional probability of \( B^{(n)} \) with respect to \( \mathcal{B}(\mathcal{Y}^{0,n-1}) \) calculated on the probability space \( (\mathcal{Y}^n, \mathcal{B}(\mathcal{Y}^n), Q(\cdot|x)) \).

Then
\[ P\{Y_n \in B | Y^{n-1} = y^{n-1}, X^n = x^n\} = Q(\{y : y_n \in B\} | \mathcal{B}(\mathcal{Y}^{0,n-1})) = q_n(B_n; y^{n-1}, x^n) - a.a. (x^{n-1}, y^{n-1}) \in \mathcal{X}^{0,n} \times \mathcal{Y}^{0,n-1}. \]

(II.4)

Hence, \( q_n(\cdot; \cdot; \cdot) \in \mathcal{Q}(\mathcal{Y}^n; \mathcal{X}^{0,n-1} \times \mathcal{X}^{0,n}) \) are determined from \( Q(\cdot; \cdot) \). Note that (II.4) states that the family of measures \( Q(\cdot|x) \) on \( (\mathcal{Y}^n, \mathcal{B}(\mathcal{Y}^n)) \) uniquely defines the collection \( \{q_n(dy_n; y^{n-1}, x^n) : n \in \mathbb{N}\} \) of conditional distributions satisfying iii), iv).

Moreover, the joint distribution of RV’s \( \{(X_n, Y_n) : n \in \mathbb{N}\} \) on \( (\mathcal{X}^n \times \mathcal{Y}^n, \mathcal{B}(\mathcal{X}^n) \circ \mathcal{B}(\mathcal{Y}^n)) \) is defined by the convolution of the two measures as follows.
\[ P\{X_0 \in A_0, Y_0 \in B_0, \ldots, X_n \in A_n, Y_n \in B_n\} \]
\[ \triangleq \int_{A_0} p_0(dx_0) \int_{B_0} q_0(dy_0; x_0) \ldots \int_{A_n} p_n(x_n; x^{n-1}) \int_{B_n} q_n(dy_n; y^{n-1}, x^n). \]

Hence, for any \( P(\cdot) \) defined by (II.1) and \( Q(\cdot; \cdot) \) satisfying consistency condition C1 there exist a probability space \( (\Omega, \mathcal{F}, P) \equiv (\mathcal{X}^n \times \mathcal{Y}^n, \mathcal{B}(\mathcal{X}^n) \circ \mathcal{B}(\mathcal{Y}^n), P) \) and a sequence of RV’s \( \{(X_i, Y_i) : i \in \mathbb{N}\} \) defined on it, whose joint probability distribution is defined uniquely via \( P(\cdot) \) and \( Q(\cdot; \cdot) \). Finally, we note that Kolmogorov’s extension theorem [38] guarantees the construction of countable additive probability measures for both \( P(\cdot) \) and \( Q(\cdot|x) \).

In [30], we have utilized the equivalent definition of any family of reproduction conditional distributions satisfying consistency condition C1 to show that this family is convex almost surely (a.s.). First, we recall the definition of a.s.-convexity of regular conditional distributions given in [30] Appendix, Definition VI.17]. It is well-known that the set of regular conditional distributions \( P_{Y^n|X^n}(\cdot|x^n) \in \mathcal{M}_1(\mathcal{Y}_{0,n}) \) is a convex set. Define the subset of \( \mathcal{Q}(\mathcal{Y}^n; \mathcal{X}^n) \) consisting of all conditional distributions which satisfy consistency condition C1 by
\[ \mathcal{Q}^{C1}(\mathcal{Y}^n; \mathcal{X}^n) \triangleq \{Q(\cdot|x) \in \mathcal{M}_1(\mathcal{Y}^n) : Q(\cdot|x) \text{ satisfies consistency condition C1}\} \]
and denote $Q^{C1}(Y_{0,n}; X_{0,n})$ its projection to finite number of coordinates by

$$Q^{C1}(Y_{0,n}; X_{0,n}) \overset{\Delta}{=} \left\{ Q_{0,n}(\cdot | x^n) \in M_1(Y_{0,n}) : Q_{0,n}(\cdot | x^n) \text{ satisfies consistency condition } C1 \right\}.$$ 

The following theorem is derived in [30].

**Theorem II.1** [30] *(Convexity of the set $Q^{C1}(Y^N; X^N)$)*

Let $\{Y_n : n \in \mathbb{N}\}$ be a Polish spaces with $\{B(Y_n) : i \in \mathbb{N}\}$ the $\sigma$–algebras of Borel sets. Then $Q^{C1}(Y_{0,n}; X_{0,n})$ is a convex set.

Theorem [II.1] states that for any $\lambda \in [0, 1]$, and $\overrightarrow{Q}_{0,n}^1(\cdot | x^n)$, $\overrightarrow{Q}_{0,n}^2(\cdot | x^n)$ two probability measures on $(Y^N, B(Y^N))$ of the form (II.2), then $\lambda \overrightarrow{Q}_{0,n}^1(\cdot | x^n) + (1 - \lambda) \overrightarrow{Q}_{0,n}^2(\cdot | x^n)$ is also a probability measure on $(Y^N, B(Y^N))$ of the form (II.2).

**B. Information Nonanticipative RDF**

In this section, we formally define the information nonanticipative RDF.

First, we construct the various measures of interest. Given the source distribution $P(\cdot) \in M_1(X^N)$ and reproduction distribution $Q(\cdot | x) \in Q^{C1}(Y^N; X^N)$ define the following measures.

**P1:** The joint distribution on $X^N \times Y^N$ defined uniquely for $A_i \in B(X_i), B_i \in B(Y_i), \forall i \in \mathbb{N}$, by

$$(\mu_{0,n} \otimes \overrightarrow{Q}_{0,n})(\times_{i=0}^n (A_i \times B_i)) \overset{\Delta}{=} \int_{A_0} p_0(dx_0) \int_{B_0} q_0(dy_0; x_0) \cdots \int_{A_n} p_n(x_n; x^{n-1}) \int_{B_n} q_n(dy_n; y^{n-1}, x^n)

\equiv P\left\{ X_0 \in A_0, Y_0 \in B_0, \ldots, X_n \in A_n, Y_n \in B_n \right\}. \quad \text{(II.5)}$$

**P2:** The marginal distributions on $Y^N$ defined uniquely for $B_i \in B(Y_i), \forall i \in \mathbb{N}$, by $^2$

$$\nu_{0,n}(\times_{i=0}^n B_i) \overset{\Delta}{=} (\mu_{0,n} \otimes \overrightarrow{Q}_{0,n})(\times_{i=0}^n (X_i \times B_i))

\equiv P\left\{ X_0 \in X_0, Y_0 \in B_0, \ldots, X_n \in X_n, Y_n \in B_n \right\}. \quad \text{(II.6)}$$

**P3:** The product measure $\overrightarrow{\Pi}_{0,n} : B(X_{0,n}) \otimes B(Y_{0,n}) \longrightarrow [0, 1]$ defined uniquely for $A_i \in B(X_i), B_i \in B(Y_i), \forall i \in \mathbb{N}$, by

$$(\times_{i=0}^n (A_i \times B_i)) \overset{\Delta}{=} (\mu_{0,n} \times \nu_{0,n})(\times_{i=0}^n (A_i \times B_i))

\equiv \int_{A_0} p_0(dx_0) \int_{B_0} \nu_0(dy_0) \cdots \int_{A_n} p_n(x_n; x^{n-1}) \int_{B_n} \nu_n(dy_n; y^{n-1}).$$

$^2$The proper definition is $\nu_{0,n}^{P \otimes Q}(\cdot)$ to indicate the dependence on $P \otimes Q$; however, for simplicity we use $\nu_{0,n}(\cdot) \equiv \nu_{0,n}^{P \otimes Q}(\cdot)$. 

May 8, 2014
The information theoretic measure associated with nonanticipative RDF is a special case of directed information\cite{[7]}, defined via relative entropy $\mathbb{D}(\cdot||\cdot)$ as follows.

$$
I_{\mu,0,n}(X^n \rightarrow Y^n) \triangleq \mathbb{D}(\mu_{0,n} \otimes Q_{0,n}||\tilde{Q}_{0,n})
$$

(I.7)

$$
= \int_{X_0,n \times Y_0,n} \log \left( \frac{Q_{0,n}(dy^n|x^n)}{\nu_{0,n}(dy^n)} \right) (\mu_{0,n} \otimes \tilde{Q}_{0,n})(dx^n, dy^n)
$$

(I.8)

$$
\equiv \mathbb{I}_{X^n \rightarrow Y^n}(\mu_{0,n}, \tilde{Q}_{0,n}).
$$

(I.9)

The RHS of (I.8) is obtained by using the chain rule of relative entropy \cite{[32]} and the following observations, $\mu_{0,n} \otimes \tilde{Q}_{0,n} \ll \nu_{0,n}$ if and only if $\tilde{Q}_{0,n}(\cdot|x^n) \ll \nu_{0,n}(\cdot)$ for $\mu_{0,n}$—almost all $x^n \in X_0,n$. In (I.9) we use the notation $\mathbb{I}_{X^n \rightarrow Y^n}(\cdot, \cdot)$ to indicate the functional dependence of $I_{\mu,0,n}(X^n \rightarrow Y^n)$ on $\{\mu_{0,n}, \tilde{Q}_{0,n}\}$.

The following convexity result is derived in \cite{[30]}, and we recall it for subsequent use.

**Theorem II.2.** \cite{[30]} (Convexity of information nonanticipative RDF)

Let $\{X_n : n \in \mathbb{N}\}$ and $\{Y_n : n \in \mathbb{N}\}$ be Polish spaces.

Then $\mathbb{I}_{X^n \rightarrow Y^n}(\mu_{0,n}, \tilde{Q}_{0,n})$ is a convex functional of $\tilde{Q}_{0,n}(\cdot|x^n) \in \mathcal{Q}^{C1}(Y_0,n; X_0,n)$ for a fixed $\mu_{0,n}(\cdot) \in \mathcal{M}_1(X_0,n)$, and a concave function of $\mu_{0,n}(\cdot) \in \mathcal{M}_1(X_0,n)$ for a fixed $\tilde{Q}_{0,n}(\cdot|x^n) \in \mathcal{Q}^{C1}(Y_0,n; X_0,n)$.

We are now ready to introduce the definition of information nonanticipative RDF. To this end, for each $n \in \mathbb{N}$, let $d_{0,n} : X_0,n \times Y_0,n \mapsto [0, \infty]$ be a measurable distortion function. The fidelity of reproduction of $y^n \in Y_0,n$ by $x^n \in X_0,n$ is defined by the set of conditional distributions

$$
\mathcal{Q}_{0,n}^{C1}(D) \triangleq \left\{ Q_{0,n}(\cdot|x^n) \in \mathcal{Q}^{C1}(Y_0,n; X_0,n) : \frac{1}{n+1} \int_{X_0,n \times Y_0,n} d_{0,n}(x^n, y^n)(\mu_{0,n} \otimes Q_{0,n})(dx^n, dy^n) \leq D \right\}
$$

(I.10)

$$
\equiv \left\{ \tilde{Q}_{0,n}(\cdot|x^n) \in \mathcal{M}_1(Y_0,n) : \frac{1}{n+1} \int_{X_0,n \times Y_0,n} d_{0,n}(x^n, y^n)(\mu_{0,n} \otimes \tilde{Q}_{0,n})(dx^n, dy^n) \leq D \right\}
$$

\footnote{Directed information corresponds to $\{p_n(dx; x^{n-1}) : n \in \mathbb{N}\}$ and $\mu_{0,n}(\cdot)$ replaced by $\{p_n(x_n; x^{n-1}, y^{n-1}) : n \in \mathbb{N}\}$ and $\tilde{F}_{0,n}(dx^n|y^{n-1}) \triangleq \otimes_{i=0}^n p_i(dx, x^{i-1}, y^{i-1})$, respectively, in the construction of measures P1-P3, and (I.7).}
for some \( D \geq 0 \). Denote by \( Q^{C1}_{0,\infty}(D) \) the corresponding set in (II.10) when the fidelity is replaces by \( \lim_{n \rightarrow \infty} \frac{1}{n+1} \int_{X_0,n \times Y_0,n} d_{0,n}(x^n, y^n)(\mu_{0,n} \otimes \vec{Q}_{0,n})(dx^n, dy^n) \leq D \).

The information nonanticipative RDF is defined as follows.

**Definition II.3. (Information nonanticipative RDF)**

Consider the fidelity of reproduction \( Q^{C1}_{0,n}(D) \) given by (II.10).

1. The information nonanticipative RDF is defined by
   \[
   R^{na}_{0,n}(D) \triangleq \inf_{Q_{0,n}(\cdot|x^n) \in Q^{C1}_{0,n}(D)} \|X^n \rightarrow Y^n(\mu_{0,n}, \vec{Q}_{0,n}) \]  
   provided the infimum over \( Q^{C1}_{0,n}(D) \) in (II.11) exists; if not we set \( R^{na}_{0,n}(D) = +\infty \).

2. The information nonanticipative RDF rate is defined by
   \[
   R^{na}(D) = \lim_{n \rightarrow \infty} \frac{1}{n+1} R^{na}_{0,n}(D) 
   \]  
   provided the limit on the RHS of (II.12) exists; if the infimum over \( Q^{C1}_{0,n}(D) \) in (II.11) does not exist then we set \( R^{na}(D) = +\infty \).

In addition, define
\[
\vec{R}^{na}(D) \triangleq \inf_{Q_{0,\infty}^{\infty}(\cdot|x^n) \in Q^{C1}_{0,\infty}(D)} \lim_{n \rightarrow \infty} \frac{1}{n+1} \|X^n \rightarrow Y^n(\mu_{0,n}, \vec{Q}_{0,n}) \geq R^{na}(D). \]  

Since, in general, \( \vec{R}^{na}(D) \geq R^{na}(D) \), then, \( R^{na}(D) \) is more natural than \( \vec{R}^{na}(D) \). By analogy with the definition of classical RDF, one may assume that \( \{(X_i, Y_i) : i = 0, 1, \ldots\} \) is jointly stationary and ergodic process or \( \frac{1}{n+1} \log \left( \frac{Q_{0,n}(\cdot|x^n)}{\nu_{0,n}(\cdot)} \right)(y^n) \) is information stable. However, we do not know á priori whether the joint process \( \{(X_i, Y_i) : i = 0, 1, \ldots\} \) is stationary. We will show that under general conditions the inequality in (II.13) holds with equality (i.e., the infimum and the limit are interchangeable), the infimum in (II.11) exists, and the limit in (II.12) also exists.

### III. Existence of Information Nonanticipative RDF and Relations to Nonanticipatory \( \epsilon \)-Entropy

This section consists of three subsections. In the first subsection, we show existence of an optimal reproduction distribution \( \vec{Q}_{0,n}(\cdot|x^n) \in Q^{C1}_{0,n}(D) \), which achieves the infimum of \( R^{na}_{0,n}(D) \), thus establishing finiteness of \( R^{na}_{0,n}(D) \) for some finite \( n \in \mathbb{N} \), under very general conditions.
In the second subsection, we show equivalence of the information nonanticipative RDF (see Definition II.3) to Gorbunov and Pinsker [20] nonanticipatory \( \epsilon \)-entropy. In the third subsection, we consider consistent stationary sources as defined by Gorbunov and Pinsker [20], to establish equality of the limiting expressions in (I.15), finiteness of \( R_{0,\alpha}^{(\epsilon)}(D) \), and that for stationary sources the infimum over \( \overrightarrow{Q}_{0,n}(\cdot|x^n) \in Q_{0,n}^{C_1}(D) \) is achieved and it is realizable by stationary source-reproduction pairs \( \{(X_n, Y_n) : n \in \mathbb{N}\} \).

Since our source and reproduction alphabets are general Polish spaces, to address the question of existence of solution to the information nonanticipative RDF we shall invoke the topology of weak convergence of probability measures and Prohorov’s theorems. Let \( Z \) be a Polish space and \( BC(Z) \) the set of bounded continuous real-valued functions \( h : Z \mapsto \mathbb{R} \) endowed with the uniform norm \( ||h||_{BC(Z)} \triangleq \sup_{z \in Z} |h(z)| \). We say that a sequence of probability measures \( \{P^\alpha(\cdot) : \alpha = 1, 2, \ldots\} \subset M_1(Z) \) converges weakly to \( P^o(\cdot) \in M_1(Z) \) if

\[
\lim_{\alpha \to \infty} \int_Z f(z) P^\alpha(dz) = \int_Z f(z) P^o(dz) \iff P^\alpha \overset{w}{\rightharpoonup} P^o, \forall f \in BC(Z).
\]

A. Finite Horizon

First, we investigate the finite time or horizon case and then, in a subsequent section we proceed with the discussion of the infinite horizon case.

Our main assumptions, which are natural generalizations of those imposed in [23], to establish existence of solution for the single letter classical information RDF are the following.

**Assumption III.1.** For all \( n \in \mathbb{N} \),

(A1) \( \mathcal{Y}_{0,n} \) is a compact Polish space and \( \mathcal{X}_{0,n} \) is a Polish space;

(A2) For all \( h(\cdot) \in BC(\mathcal{Y}_n) \), the function mapping

\[
(x^n, y^{n-1}) \in \mathcal{X}_{0,n} \times \mathcal{Y}_{0,n-1} \mapsto \int_{\mathcal{Y}_n} h(y) q_n(dy; y^{n-1}, x^n) \in \mathbb{R}
\]

is continuous jointly in the variables \( (x^n, y^{n-1}) \in \mathcal{X}_{0,n} \times \mathcal{Y}_{0,n-1} \);

(A3) The distortion function \( d_{0,n}(x^n, \cdot) : \mathcal{X}_{0,n} \times \mathcal{Y}_{0,n} \mapsto [0, \infty] \) is Borel measurable relative to \( \mathcal{B}(\mathcal{X}_{0,n}) \circ \mathcal{B}(\mathcal{Y}_{0,n}) \) and continuous on \( y^n \in \mathcal{Y}_{0,n} \);

(A4) The distortion level \( D \) is such that there exist sequence \( (x^n, y^n) \in \mathcal{X}_{0,n} \times \mathcal{Y}_{0,n} \) satisfying \( d_{0,n}(x^n, y^n) < D \).
Assumption III.1 (A1) is also required for the existence of the single letter classical information RDF. Assumption III.1 (A2) is a weak continuity of the mapping defined by (III.1); this is preferable over strong continuity of \( q_n(A; y^{n-1}, x^n) \) as a function of \( (y^{n-1}, x^n) \in Y_{0,n-1} \times X_{0,n} \) for every Borel set \( A \in \mathcal{B}(Y_n) \) in order not to exclude reproduction distribution described by delta measures. Assumption III.1 (A3) is preferable over the bounded distortion function in order to allow unbounded distortion measures, such as square-error distortions for continuous alphabet sources. Assumption III.1 (A4) ensures that the set \( Q_{C1}^0(D) \) is non-empty, since it implies existence of some \( \epsilon > 0 \) such that the set \( Q_{C1}^0(D(1-\epsilon)) \) is non-empty.

The proof of existence and finiteness of \( R_{n,n}^m(D) \) (see Definition II.3) for finite \( n \in \mathbb{N} \), is based on Weierstrass’ theorem, therefore we need to establish lower semicontinuity of the functional \( I_{X_n \rightarrow Y_n}(\mu_{0,n}, \tilde{Q}_{0,n}) \) with respect to \( \tilde{Q}_{0,n}(\cdot|x^n) \in Q_{C1}^0(D) \) for a fixed \( \mu_{0,n}(\cdot) \in M_1(X_{0,n}) \), and compactness of the set \( Q_{C1}^0(D) \), using the topology of weak convergence of probability measures.

First, we show compactness of the constraint set \( Q_{C1}^0(D) \).

**Theorem III.2. (Compactness of fidelity set)**

Under Assumption III.1 the following hold.

1. The set of probability measures \( Q_{C1}^0(Y_{0,n}; X_{0,n}) \) is closed and tight (i.e., compact);
2. the set \( Q_{C1}^0(D) \) is a closed subset of \( Q_{C1}^0(Y_{0,n}; X_{0,n}) \) (i.e., compact).

*Proof:* See Appendix A.

Next, we establish lower semicontinuity of the functional \( I_{X_n \rightarrow Y_n}(\mu_{0,n}, \tilde{Q}_{0,n}) \) with respect to \( \tilde{Q}_{0,n}(\cdot|x^n) \in Q_{C1}^0(Y_{0,n}; X_{0,n}) \) for a fixed \( \mu_{0,n}(\cdot) \in M_1(X_{0,n}) \).

**Lemma III.3. (Lower semicontinuity of \( I_{X_n \rightarrow Y_n}(\mu_{0,n}, \tilde{Q}_{0,n}) \))**

Let Assumption III.1 (A1), (A2), hold. Then, the functional \( I_{X_n \rightarrow Y_n}(\mu_{0,n}, \tilde{Q}_{0,n}) \) is lower semicontinuous on \( \tilde{Q}_{0,n}(\cdot|x^n) \in Q_{C1}^0(Y_{0,n}; X_{0,n}) \) for a fixed \( \mu_{0,n}(\cdot) \in M_1(X_{0,n}) \).

*Proof:* See Appendix B.

Utilizing Theorem III.2 and Lemma III.3 next, we show existence of a solution of the information nonanticipative RDF.

**Theorem III.4. (Existence of information nonanticipative RDF)**
Under Assumption III.1 the infimum over \( \hat{Q}_{0,n}(\cdot|x^n) \in \mathcal{Q}^{C1}_{0,n}(D) \) of
\[
R_{0,n}^{na}(D) \triangleq \inf_{\hat{Q}_{0,n}(\cdot|x^n) \in \mathcal{Q}^{C1}_{0,n}(D)} \frac{1}{n+1} \mathbb{I}_{X^n \rightarrow Y^n}(\mu_{0,n}, \hat{Q}_{0,n})
\] (III.2)
is achieved by some \( \hat{Q}^*_{0,n}(\cdot|x^n) \in \mathcal{Q}^{C1}_{0,n}(D) \).

Proof: By Lemma III.3, the functional \( \mathbb{I}_{X^n \rightarrow Y^n}(\mu_{0,n}, \hat{Q}_{0,n}) \) is lower semicontinuous with respect to \( \hat{Q}_{0,n}(\cdot|x^n) \in \mathcal{Q}^{C1}_{0,n}(D) \) for fixed \( \mu_{0,n} \in \mathcal{M}_1(X_{0,n}) \), and by Theorem III.2, \( \mathcal{Q}^{C1}_{0,n}(D) \) is compact. Therefore, by invoking Weierstrass’ theorem [39] we deduce that the infimum in (III.2) is achieved by some \( \hat{Q}^*_{0,n}(\cdot|x^n) \in \mathcal{Q}^{C1}_{0,n}(D) \). This completes the derivation.

Therefore, we conclude by stating that under the general conditions of Assumption III.1 the information nonanticipative RDF is finite for any finite \( n \in \mathbb{N} \), that is, there exists a \( K \in (0, \infty) \) such that \( R_{0,n}^{na}(D) < K \), for any finite \( n \in \mathbb{N} \).

B. Equivalence of Nonanticipative RDF and Nonanticipatory \( \epsilon \)-Entropy

In this section, we recall Gorbunov-Pinsker’s definition of nonanticipatory \( \epsilon \)-entropy [20], we show equivalence of certain statements regarding conditional independence, and we use them to show equivalence of the information nonanticipative RDF (II.11) (respectively information nonanticipative RDF rate (II.12)), and Gorbunov and Pinsker’s definition of nonanticipatory \( \epsilon \)-entropy (respectively message generation rates).

For a given a source \( P_{X^n} \in \mathcal{M}_1(X_{0,n}) \) and a reproduction \( P_{Y^n|X^n} \in \mathcal{Q}_{0,n}(D) \subset \mathcal{Q}(Y_{0,n}; X_{0,n}) \), Gorbunov and Pinsker restricted the fidelity set of classical RDF, \( \mathcal{Q}_{0,n}(D) \), to those reproduction distributions which satisfy the following MC.

\[
X_{n+1}^\infty \leftrightarrow X^n \leftrightarrow Y^n \iff P_{Y^n|X^n}(dy^n|x^\infty) = P_{Y^n|X^n}(dy^n|x^n) \quad \text{a.a.} \ x^\infty \in X^\mathbb{N}, n = 0, 1, \ldots.
\] (III.3)

Then they introduced the nonanticipatory \( \epsilon \)-entropy and nonanticipatory message generation rate as follows.

**Definition III.5.** [20]**(Nonanticipatory \( \epsilon \)-entropy and message generation rate)**

Consider the fidelity constraint set \( \mathcal{Q}_{0,n}(D) \) defined by (I.6).

The nonanticipatory \( \epsilon \)-entropy is defined by
\[
R_{0,n}^{\epsilon}(D) \triangleq \inf_{P_{Y^n|X^n}(\cdot|x^n) \in \mathcal{Q}_{0,n}(D): X_{i+1} \leftrightarrow X^i \leftrightarrow Y^i, i=0,1,\ldots,n-1} I(X^n; Y^n)
\] (III.4)
provided the infimum in (III.4) over $Q_{0,n}(D)$ and $X_{i+1}^n \leftrightarrow X^i \leftrightarrow Y^i$, $i = 0, 1, \ldots, n - 1$, exists; if not then we set $R_{0,n}^\epsilon(D) = +\infty$.

The nonanticipatory message generation rate of the source is defined by

\[ R^\epsilon(D) \triangleq \lim_{n \to \infty} \frac{1}{n + 1} R_{0,n}^\epsilon(D) \]  

(III.5)

provided the limit in the RHS of (III.5) exists; if the infimum in (III.4) does not exist we set $R^\epsilon(D) = +\infty$.

In addition, define

\[ \tilde{R}^\epsilon(D) \triangleq \inf_{P_{Y^n|X^n}(\cdot|x^n) \in Q_{0,n}(D)} \lim_{n \to \infty} \frac{1}{n + 1} I(X^n;Y^n) \geq R^\epsilon(D). \]  

(III.6)

The MC constraint (III.3) is a probabilistic version of the deterministic causal reproduction coder in [4], defined as the cascade of an encoder-decoder (ED) as follows.

**Definition III.6.** [4] (Causal reproduction coder)

A reproduction coder $f_i : \mathcal{X}_{0,n} \mapsto \mathcal{Y}_i$, $\forall i = 0, 1, \ldots, n$, is called causal if the mapping $x^n \mapsto f_i(x^n)$ is measurable $\forall i \in \mathbb{N}^n$ and

\[ f_i(x^n) = f_i(\hat{x}^n) \quad \text{whenever} \quad x^i = \hat{x}^i, \quad \forall n \geq i, \quad n \in \mathbb{N}. \]

Thus, a source code is called causal if the reproduction coder is causal. Since the class of randomized reproduction coders embeds deterministic coders, then probabilistically, a reproduction coder is causal if and only if the following MC holds $X_{i+1}^\infty \leftrightarrow X^i \leftrightarrow Y_i$, $\forall i \in \mathbb{N}$. Therefore, nonanticipatory $\epsilon$-entropy, $R_{0,n}^\epsilon(D)$, imposes a probabilistic causality constraint on the optimal reproduction distribution.

Gorbunov and Pinsker [21], [22] proceeded further to compute $R^\epsilon(D) \triangleq \lim_{n \to \infty} \frac{1}{n + 1} R_{0,n}^\epsilon(D)$, whenever the limit exists, for the class of scalar Gaussian stationary ergodic sources by working on the frequency domain using power spectral densities. Further, in [21], [22] it is also shown that in the limit, as $D \to 0$, the nonanticipatory message generation rate $R^\epsilon(D)$ of Gaussian stationary sources converges to the classical information RDF. Recently, in [11] several bounds for the OPTA by causal and noncausal codes are derived for Gaussian sources, with quadratic distortion function, utilizing an upper bound to the nonanticipatory $\epsilon$-entropy.

Now, we are ready to establish the connection between nonanticipatory $\epsilon$-entropy (III.4) (e.g.,
Lemma III.7. (Equivalent nonanticipative statements)

The following statements are equivalent.

**MC1:** \( P_{Y^n|X^n}(dy^n|x^n) = P_{Y^n|X^n}(dy^n|x^n) = \otimes_{i=0}^{n} P_{Y_{i+1},X_{i+1}}(dy_{i+1},x_{i+1}), \quad \forall n \in \mathbb{N}; \)

**MC2:** \( Y_i \leftrightarrow (X^i, Y^{i-1}) \leftrightarrow (X_{i+1}, X_{i+2}, \ldots, X_n) \) forms a MC, for each \( i = 0, 1, \ldots, n - 1 \), \( \forall n \in \mathbb{N}; \)

**MC3:** \( Y_i \leftrightarrow X^i \leftrightarrow X_{i+1} \) forms a MC, for each \( i = 0, 1, \ldots, n - 1 \), \( \forall n \in \mathbb{N}; \)

**MC4:** \( X_{n+1}^i \leftrightarrow X^i \leftrightarrow Y^i \) forms a MC, for each \( i = 0, 1, \ldots, n - 1 \), \( \forall n \in \mathbb{N}. \)

**Proof:** See Appendix C.

The equivalence of **MC1,** **MC2,** **MC3** is easily shown, and so is the fact that **MC4** implies any of **MC1,** **MC2,** **MC3.** What is new in Lemma III.7 is the equivalence of **MC4** with any of **MC1,** **MC2,** **MC3.** We note that **MC3** of Lemma III.7 is precisely Granger’s definition of temporal causality [40], which is used in econometrics to unravel complex relations between macroeconomic variables from time series observation. It is also applied in bioengineering [40], [41], and more recently in neuroimaging to infer that \( \{Y_n : n \in \mathbb{N}\} \) does not cause \( \{X_n : n \in \mathbb{N}\}. \)

In the next theorem, we utilize Lemma III.7 and more specifically, the fact that **MC4** is equivalent to **MC2** and **MC1,** to show that the extremum of the nonanticipatory \( \epsilon \)-entropy (III.4), \( R_{\epsilon,0,n}(D) \), is equivalent to the extremum of nonanticipative RDF given by (II.11), \( R_{\text{na},0,n}(D) \).

Theorem III.8. (Equivalence of \( R_{\text{na},0,n}(D) \) and \( R_{\epsilon,0,n}(D) \))

**Definition II.3** and **Definition III.5** are equivalent, i.e., \( R_{\text{na},0,n}(D) = R_{\epsilon,0,n}(D), \quad \forall n \in \mathbb{N}. \)

**Proof:** The derivation follows directly from Lemma III.7.

C. Infinite Horizon

In this section, we first invoke Theorem III.4 to investigate existence of the information nonanticipative RDF rate, and the validity of interchanging the limit and infimum operations in (I.15).
One may also consider the two-sided definition of nonanticipative RDF by replacing \( R_{0,n}^{na}(D) \) by \( R_{n_1,n_2}^{na}(D) \), \( n_2 > n_1 \), in which case, the rate is defined by \( \lim_{n_2-n_1 \to \infty} \frac{1}{n_2-n_1+1} R_{n_1,n_2}^{na}(D) \), provided the limit exists. However, the rate is defined if and only if the following limit is defined for some \( n_1 \): \( \lim_{n_2 \to \infty} \frac{1}{n_2-n_1} R_{n_1,n_2}^{na}(D) \). Hence, without loss of generality, we let \( n_1 = 0 \).

First, we prove the following inequality.

**Lemma III.9.**

Suppose that \( \mathcal{X}_{0,n} \) and \( \mathcal{Y}_{0,n} \) are Polish spaces. Then

\[
R^{na}(D) \leq \overrightarrow{R}^{na}(D) \triangleq \inf_{Q_{0,\infty}(\cdot | x^\infty) \in Q_{0,\infty}^{C1}(D)} \lim_{n \to \infty} \frac{1}{n+1} \mathbb{I}_{X^n \to Y^n}(\mu_{0,n}, \overrightarrow{Q}_{0,n}). \tag{III.7}
\]

**Proof:** If the infimum in (III.2) does not exist there is nothing to prove. Hence, suppose this infimum exists. By definition we have

\[
R_{0,n}^{na}(D) \leq \mathbb{I}_{X^n \to Y^n}(\mu_{0,n}, \overrightarrow{Q}_{0,n}), \quad \forall \overrightarrow{Q}_{0,n}(\cdot | x^n) \in Q_{0,n}^{C1}(D)
\]

and hence, by taking the limit on both sides we obtain

\[
\lim_{n \to \infty} \frac{1}{n+1} R_{0,n}^{na}(D) \leq \lim_{n \to \infty} \frac{1}{n+1} \mathbb{I}_{X^n \to Y^n}(\mu_{0,n}, \overrightarrow{Q}_{0,n}), \quad \forall \overrightarrow{Q}_{0,\infty}(\cdot | x^\infty) \in Q_{0,\infty}^{C1}(D).
\]

Taking the infimum over \( \overrightarrow{Q}_{0,\infty}(\cdot | x^\infty) \in Q_{0,\infty}^{C1}(D) \) we obtain (III.7). This shows that \( R^{na}(D) \leq \overrightarrow{R}^{na}(D) \).

Since by Theorem III.8 \( R_{0,n}^{na}(D) = R_{0,n}^e(D) \), \( \forall n \in \mathbb{N} \), and since we have shown existence of solutions to the information nonanticipative RDF (e.g., Theorem III.4), all technical results derived in [20] Theorems 1-4] are directly applicable to \( R_{0,n}^{na}(D) \) and its rate, \( R^{na}(D) \), without assuming finiteness of \( R_{0,n}^{na}(D) \) for some \( n \), as in [20]. Next, we summarize these results in order to show that the limit \( \lim_{n \to \infty} \frac{1}{n+1} R_{0,n}^{na}(D) \) is finite by first introducing some definitions from [20].

Consider the following class of sources [20]. Let \( Z \triangleq \{Z_i : i = 0, 1, \ldots \} \in \mathcal{A} \subseteq \mathcal{X}^\infty \times \mathcal{Y}^\infty \) denote the sets of points of the random process \( \{Z_i : i = 0, 1, \ldots \} \) and define \( \mathcal{A}_s \subseteq \mathcal{X}^\infty \times \mathcal{Y}^\infty \) as the set of points \( Z_s \triangleq \{Z_{i-s} : i = 0, 1, \ldots \} \) representing shifts of points in \( Z \), for \( s = 0, 1, \ldots \). In addition, assume a real-valued nonnegative function \( \alpha(\cdot) : [s_1, s_2] \to [0, \infty) \) such that \( \sum_{t=s_1}^{s_2} \alpha_t = 1 \), \( t \in (0, \infty) \). Define the joint distribution \( P_{X^n,Y^n} \) on \( \mathcal{X}^\infty \times \mathcal{Y}^\infty \) by

\[
P_{X^n,Y^n}(\mathcal{A}) = (P_{Y^n|X^n} \otimes P_{X^n})(\mathcal{A}) \triangleq \sum_{s=s_1}^{s_2} \alpha_s (P_{Y^n|X^n} \otimes P_{X^n})(\mathcal{A}_s), \forall n = 0, 1, \ldots
\]
Definition III.10. ([20] Specified, consistent, and stationary sources)
The source \( \{ X_n : n \in \mathbb{N} \} \) is called

(1) “specified” if

(a) \( P_{Y^k | X^k} \in Q_{0,k}(D) \) and \( P_{Y_{k+1}^n | X_{k+1}^n} \in Q_{k+1,n}(D) \), \( Y_{k+1}^n \overset{\Delta}{=} \{ Y_{k+1}, \ldots, Y_n \} \)

implies

(b) the joint conditional distribution of the concatenated RVs satisfies

\[ P_{Y^k, Y_{k+1}^n | X^k, X_{k+1}^n} \in Q_{0,n}(D), \forall k = 0, 1, \ldots, n - 1; \]

(2) “consistent” if (b) implies (a);

(3) “stationary” if the random process \( \{ X_n : n = 0, 1, \ldots \} \) is stationary and for any \( k = 1, 2, \ldots \), the sets \( Q_{0,n}(D) \) and \( Q_{k,n+k}(D) \) are copies of the same set.

The fidelity set is called

(4) “shift invariant” if given the source \( P_{X^n} (\cdot) \), then for any \( \alpha (\cdot) \) defined above the following holds.

\[ P_{Y^n | X^n} \in Q_{0,n}(D) \implies P_{\tilde{Y}^n | X^n} \in Q_{0,n}(D), \forall n = 0, 1, \ldots \]  \hspace{1cm} (III.8)

Note that in general, Definition [III.10] (2) does not imply (1). This point is further elaborate in [20]. Notice that condition (III.8) holds for the Letter-by-Letter fidelity \( \mathbb{E} \{ \rho(X_i, Y_i) \} \leq D_i \), \( D_i > 0, i = 0, 1, \ldots, n \) (see [43]).

The following theorem is a direct consequence of Lemma III.9, Theorem III.4, and Gorbunov-Pinsker [20, Theorem 2].

Theorem III.11. (Limits)

Suppose Assumption [III.7] holds and the source is stationary.

Then the following hold.

\[ R^{na}(D) = \lim_{n \to \infty} \inf_{Q_{0,n}(\cdot|x^n) \in \mathcal{Q}^0_n(D)} \frac{1}{n + 1} \mathbb{I}_{X^n \to Y^n} (\mu_{0,n}, \overrightarrow{Q}_{0,n}) < \infty \]  \hspace{1cm} (III.9)

e.g., the limit exists and it is finite. Moreover,

\[ R^{na}(D) = \overrightarrow{R}^{na}(D) \equiv \inf_{Q_{0,\infty}(\cdot|x^\infty) \in \mathcal{Q}^1_\infty(D)} \lim_{n \to \infty} \frac{1}{n + 1} \mathbb{I}_{X^n \to Y^n} (\mu_{0,n}, \overrightarrow{Q}_{0,n}). \]  \hspace{1cm} (III.10)

Proof: The derivation utilizes Theorem III.4 and the subadditivity of \( R^{na}_{0,n}(D) \), that is, \( R^{na}_{0,n}(D) \leq R^{na}_{0,k}(D) + R^{na}_{k+1,n}(D) \), \( 0 < k < n \). Since by Theorem III.8 we have \( R^{na}_{0,n}(D) = \)
$R_{0,n}(D)$, and by \cite[Lemma 1]{[20]}, $R_{0,n}(D)$ is subadditive, then $R_{0,n}(D)$ is also subadditive. Moreover, under the conditions of Theorem III.4 we know that $R_{0,n}(D)$ is finite for any finite $n \in \mathbb{N}$. Utilizing this and the subadditivity of $R_{0,n}(D)$ we deduce that the limit in (III.9) exist and it is finite. (III.10) follows from \cite[Theorem 2]{[20]}, the stationarity of the source, and Theorem III.8 which states that $R_{0,n}(D) = R_{0,n}(D)$.

Often, in the derivation of classical RDF rate it is assumed that the process $\{(X_n, Y_n) : n \in \mathbb{N}\}$ is jointly stationary. This is not very natural because one does not know à priori whether the reproduction process is stationary. The next theorem utilizes Theorem III.4 and \cite{[20]} to show that the infimum of the information nonanticipative RDF rate is achieved by a stationary reproduction distribution, which makes the process $\{(X_n, Y_n) : n \in \mathbb{N}\}$ jointly stationary.

**Theorem III.12.** (Stationarity of reproduction distribution)
Suppose Assumption III.1 holds, the source $\{X_n : n \in \mathbb{N}\}$ is stationary and consistent, and the fidelity set is shift invariant. Then, the infimum in (III.10) is achieved by some $\tilde{Q}_{0,n}^*(\cdot|x^n) \in Q_{0,n}^{C1}(D)$ such that the source-reproduction pair $\{(X_n, Y_n) : n \in \mathbb{N}\}$ is jointly stationary.

**Proof:** By Assumption III.1 the statement of Theorem III.4 holds. Moreover, by Theorem III.8 we have $R_{0,n}^{na}(D) = R_{0,n}^e(D)$. By invoking \cite[Theorem 4]{[20]} we establish the claim of stationarity of the joint process.

Therefore, under Assumption III.1 by showing existence of solution to the information nonanticipative RDF (e.g., Theorem III.4), we have strengthened the results described in \cite[Theorems 1-4]{[20]} because we have removed the assumption that $R_{0,n}^e(D)$ is finite for some finite $n \in \mathbb{N}$.

**IV. Optimal Reproduction of Nonanticipative RDF, Properties, and Examples**

In this section, we recall the closed form expression of the optimal reproduction conditional distribution of the information nonanticipative RDF, $R_{0,n}^{na}(D)$, under the assumption that the reproduction distributions $\{q_n(\cdot;y^{n-1},x^n) : n \in \mathbb{N}\}$ are stationary derived in \cite[Section IV]{[18]}. Further, we derive important properties of information nonanticipative RDF, and we present two running examples, the BSMS($p$) and the multidimensional partially observed Gaussian-Markov stationary source.
Conditions for stationarity to hold are given in Theorem [III.12]. The main assumption we impose is the following.

**Assumption IV.1.**

The $(n + 1)$-fold convolution of causal conditional distribution $\overrightarrow{Q}_{0,n}^*(|x^n) = \otimes_{i=0}^{n} q_i(\cdot; y^{i-1}, x^i)$ which achieves the infimum of $R_{0,n}^{\text{con}}(D)$, is a convolution of stationary conditional distributions.

**A. Stationary optimal reproduction distribution**

Theorem [III.12] gives general conditions for Assumption [IV.1] to hold. Clearly, (II.11) is a constrained problem, which is convex due to the convexity of the fidelity set, and the convexity of $\mathbb{I}_{X^n \to Y^n}(\mu_{0,n}, \cdot)$, as a functional of $\overrightarrow{Q}_{0,n}^*(|x^n) \in Q^{C^1}(Y_{0,n}; X_{0,n})$, (see Theorem [II.2]). Therefore, we apply duality theory [39] to convert the constrained problem into an unconstrained problem using Lagrange multipliers, and then we verify the equivalence of the constrained and unconstrained problems. This procedure is done in [18, Theorem IV.3] hence, it is omitted. Assumption [IV.1] facilitates the computation of the Gateaux differential of $\mathbb{I}_{X^n \to Y^n}(\mu_{0,n}, \overrightarrow{Q}_{0,n})$ at the optimal $\overrightarrow{Q}_{0,n}^*(|x^n)$ in only one direction $(\overrightarrow{Q}_{0,n}^*(|x^n) - \overrightarrow{Q}_{0,n}^*(|x^n))$ (due to stationarity of $q_i(\cdot; y^{i-1}, x^i)$), rather than varying along each direction of $\{q_i(\cdot; y^{i-1}, x^i) : i = 0, 1, \ldots, n\}$, according to $q_i^*(\cdot; y^{i-1}, x^i) = q_i^*(\cdot; y^{i-1}, x^i) + \epsilon(q_i(\cdot; y^{i-1}, x^i) - q_i^*(\cdot; y^{i-1}, x^i))$, $i = 0, 1, \ldots, n$, and then compute the Gateaux differential of $\mathbb{I}_{X^n \to Y^n}(\mu_{0,n}, \overrightarrow{Q}_{0,n}) \equiv \mathbb{I}_{X^n \to Y^n}(\mu_{0,n}, q_i : i = 0, 1, \ldots, n)$ as a functional of $q_i(\cdot; y^{i-1}, x^i)$, $i = 0, 1, \ldots, n$ in each direction. Clearly, the nonstationary case is much more involved.

Next, we state the main theorem.

**Theorem IV.2. (Optimal stationary reproduction distribution)**

Suppose Assumption [III.1] and Assumption [IV.1] hold, and $d_{0,n} = \sum_{i=0}^{n} \rho_i(x^i, y^i)$.

The following hold.

1. The infimum is attained at $\overrightarrow{Q}_{0,n}^*(|x^n) \in Q^{C^1}_{0,n}(D)$ given by $^4$

   $$\overrightarrow{Q}_{0,n}^*(x^n_{i=0}B_i|x^n) = \int_{B_0} q_0^*(dy_0; x_0) \int_{B_1} q_1^*(dy_1; y_0, x_1) \ldots \int_{B_n} q_n^*(dy_n; y^{n-1}, x^n)$$

   (IV.1)

---

$^4$Due to stationarity assumption $\nu_i^*(\cdot; y^{i-1}) = \nu^*(\cdot; y^{i-1})$ and $q_i^*(\cdot; y^{i-1}, x^i) = q^*(\cdot; y^{i-1}, x^i)$.  

May 8, 2014 DRAFT
where

\[ q^*_i(dy_i; y_{i-1}, x^i) = \frac{e^{s\rho_i(x^i, y^i)}\nu^*_i(dy_i; y_{i-1})}{\int_{\mathcal{Y}_i} e^{s\rho_i(x^i, y^i)}\nu^*_i(dy_i; y_{i-1})}, \quad i = 0, 1, \ldots, n, \quad s \leq 0 \]  \quad (IV.2)

and \( \nu^*_i(\cdot; y_{i-1}) \in \mathcal{Q}(\mathcal{Y}_i; \mathcal{Y}_{0,i-1}), \quad i = 0, 1, \ldots, n. \)

(2) The information nonanticipative RDF is given by

\[
R^{na}_{0,n}(D) = sD(n + 1) - \sum_{i=0}^{n} \int_{\mathcal{X}_{0,i} \times \mathcal{Y}_{0,n-1}} \log \left( \int_{\mathcal{Y}_i} e^{s\rho_i(x^i, y^i)}\nu^*_i(dy_i; y_{i-1}) \right) \\
\times \mathcal{Q}^*_0(dy_{i-1}|x_{i-1}) \otimes \mu_{0,i}(dx^i). \quad (IV.3)
\]

Moreover, if \( R^{na}_{0,n}(D) > 0 \) then \( s < 0, \) and

\[
\frac{1}{n+1} \sum_{i=0}^{n} \int_{\mathcal{X}_{0,i} \times \mathcal{Y}_{0,i}} \rho_i(x^i, y^i) \mathcal{Q}^*_0(dy^i|x^i) \otimes \mu_{0,i}(dx^i) = D. \quad (IV.4)
\]

Proof: The proof is described in [18, Theorem IV.4].

For \( i \leq n, \) let \( B^{(i)} = \{ y^n : y_i \in B \}, \) \( B \in \mathcal{B}(\mathcal{Y}_i), \) and let \( \mathcal{Q}^*_0(\cdot|B(\mathcal{Y}_{0,i-1})) \) denote the conditional probability of \( B^{(i)} \) with respect to \( \mathcal{B}(\mathcal{Y}_{0,i-1}) \) calculated on \( (Y_{0,n}, B(Y_{0,n}), \mathcal{Q}^*_0(\cdot|x^n)) \) then

\[
\mathcal{Q}^*_0(\{y^n : y_i \in B\}|x^i|B(\mathcal{Y}_{0,i-1})) = q^*(dy_i; y_{i-1}, x^i)
\]

\[
= \frac{e^{s\rho_i(x^i, y^i)}\nu^*_i(dy_i; y_{i-1})}{\int_{\mathcal{Y}_i} e^{s\rho_i(x^i, y^i)}\nu^*_i(dy_i; y_{i-1})}, \quad a.a. (x^i, y_{i-1}) \in \mathcal{X}_{0,i} \times \mathcal{Y}_{0,i-1}. \quad (IV.5)
\]

The RHS term of \( (IV.5) \) determines, for each \( i = 0, 1, \ldots, \), the dependence of the reproduction distribution \( q^*(\cdot; y_{i-1}, x^i) \) on the past reproduction \( y_{i-1} \) and the past and present source symbols \( x^i. \) Below we list a few observations regarding the structure of \( (IV.5). \)

Remark IV.3. (Properties of the stationary optimal reproduction distribution)

(1) If \( \{X_n : n \in \mathbb{N}\} \) is Gaussian stationary and \( \rho_i(x^i, y^i) = ||x_i - y_i||^2, \) a quadratic function of \( (x^n, y^n) \) then for each \( (y_{i-1}, x^i) \in \mathcal{Y}_{0,i-1} \times \mathcal{X}_{0,i}, q^*(\cdot; y_{i-1}, x^i) \) is Gaussian.

This follows from the fact that the exponent in the RHS of \( (IV.5) \) is quadratic in \( (x_i, y_i) \in \mathcal{X}_i \times \mathcal{Y}_i, \) and thus by assuming \( \nu^*_i(\cdot|y_{i-1}) \) is conditionally Gaussian then the RHS of \( (IV.5) \) will be of exponential quadratic form in \( (x_i, y^i). \) Hence, this RHS can be matched by a conditional Gaussian distribution for \( q^*(\cdot; y_{i-1}, x^i). \) The procedure is standard and involves completion of squares.
If the distortion function is $\rho_i(x^i, y^i) = \rho_i(x_i, y_i)$ then

$$q^*(\cdot; y^{i-1}, x^i) = q^*(\cdot; y^{i-1}, x_i) - a.a. (x^i, y^{i-1}) \in \mathcal{X}_{0,i} \times \mathcal{Y}_{0,i-1}, \ i = 0, 1, \ldots$$

That is, the reconstruction kernel is Markov in $\{X_n : n \in \mathbb{N}\}$. However, even if we further restrict the distortion function to single letter $\rho_i(x_i, y_i)$, we cannot deduce how far into the past $q^*(\cdot; y^{i-1}, x_i)$ will depend on the reproduction symbols $y^{i-1}$. If the distortion function is of the form $\rho_i(x_i, x_{i-1}, y_i)$ then

$$q^*(\cdot; y^{i-1}, x^i) = q^*(\cdot; y^{i-1}, x_i, x_{i-1}) - a.a. (x^i, y^{i-1}) \in \mathcal{X}_{0,i} \times \mathcal{Y}_{0,i-1}, \ i = 0, 1, \ldots$$

Despite the above observations, for specific stationary sources, it turns out that the closed form expression for the optimal reproduction distribution, and that of the nonanticipative RDF are relatively easy to compute.

B. Properties of Nonanticipative RDF

In this section, we derive some properties of information nonanticipative RDF, $R^{na}(D)$, under stationarity of the joint process $\{(X_i, Y_i) : i = 0, 1, \ldots\}$. These are analogous of the well-known properties of the single letter classical RDF, $R(D)$, given in [2].

In the next Lemma we state the convexity and monotonicity property of $R^{na}(D)$ with respect to a distortion level $D$.

**Lemma IV.4. (Convexity and monotonicity of $R^{na}(D)$)**

$R^{na}(D)$ is a convex, nonincreasing function of $D \in [0, \infty)$.

**Proof:** By Theorem II.1 the set $\mathcal{Q}^{C1}_{0,n}(D)$ is convex, and by Theorem II.2, $\mathbb{I}_{X^n \to Y^n} (\mu_{0,n}, \tilde{Q}_{0,n})$ is a convex functional of $\tilde{Q}_{0,n} (|x^n) \in \mathcal{Q}^{C1}(\mathcal{Y}_{0,n}; X_{0,n})$ for a fixed $\mu_{0,n}(dx^n) \in \mathcal{M}_1(\mathcal{X}_{0,n})$. Hence, the result follows.

In the next lemma we provide the exact expression of $D_{\text{max}}$.

**Lemma IV.5. ($D_{\text{max}}$)**

$R^{na}_{0,n}(D) > 0$ for all $D < D_{\text{max}}$ and $R^{na}_{0,n}(D) = 0$ for all $D \geq D_{\text{max}}$, where

$$D_{\text{max}} = \Delta \min_{y^i \in \mathcal{Y}_{0,n}} \frac{1}{n+1} \sum_{i=0}^{n} \int_{\mathcal{X}_{0,i}} \rho_i(x^i, y^i) \mu_{0,i}(dx^i)$$

provided the minimum exists.
Proof: The derivation is similar to the one for the classical RDF, hence it is omitted.

Next, we describe a geometrical interpretation of the slope of information nonanticipative RDF.

Lemma IV.6. (Property of the slope of nonanticipative RDF)

Assume \( \rho(x^i, y^i) e^{\xi \rho(x^i, y^i)} \in L^1(\nu_i(dy_i; y^{i-1})) - a.a. \) \( (x^i, y^{i-1}) \) \( e^{\xi \rho(x^i, y^i)} \in L^1(\nu_i(dy_i; y^{i-1})) - a.a. \ (x^i, y^{i-1}), \ \forall i \in \mathbb{N}^n, \) for some \( \xi \in \mathbb{R}, \) then the Lagrange multiplier \( s \) in Theorem IV.2 is always non-positive.

Moreover,

\[
\frac{d}{dD} \frac{1}{n+1} R_{0,n}^{na}(D) = s, \quad s \leq 0. \tag{IV.6}
\]

Proof: The derivation is similar to the one for the classical RDF, hence it is omitted.

In the next lemma we give an alternative equivalent characterization of the optimal reproduction conditional distribution.

Lemma IV.7. (Equivalent characterization of optimal reproduction)

The solution to the minimization problem of nonanticipative RDF is such that

\[
\int_{X_0,i} e^{s \rho_i(x^i, y^i)} \lambda_i(x^i, y^{i-1}) P_{0,i}^*(dx^i|y^{i-1}) = 1, \quad \nu_i^*(dy_i; y^{i-1}) - a.s., \ \forall i \in \mathbb{N}^n \tag{IV.7}
\]

where

\[
\lambda_i(x^i, y^{i-1}) = \left( \int_{Y_i} e^{s \rho_i(x^i, y^i)} \nu_i^*(dy_i; y^{i-1}) \right)^{-1}, \ \forall i \in \mathbb{N}^n \tag{IV.8}
\]

and \( P_{0,i}^*(dx^i|y^{i-1}) \in Q(X_0,i; Y_{0,i-1}). \)

Proof: See Appendix D.

Note that in Lemma IV.7 we prove the solution to the optimization problem described in Theorem IV.2 on a set of \( \nu_i^* \)-measure 1. It can be shown, by utilizing measure theoretic arguments, that a necessary and sufficient condition for existence of a solution in Theorem IV.2 is the condition

\[
\int_{X_0,i} e^{s \rho_i(x^i, y^i)} \lambda_i(x^i, y^{i-1}) P_{0,i}^*(dx^i|y^{i-1}) \leq 1, \ \forall y^i \in Y_{0,i}, \ i \in \mathbb{N}^n. \tag{IV.9}
\]

Finally, by utilizing the previous results, in the next theorem, we present an alternative definition of the solution of the information nonanticipative RDF, as a maximization over a certain class of functions. By using this property, we can derive a lower bound on \( R_{0,n}^{na}(D) \), which is analogous.
to Shannon’s lower bound. In fact, we use this bound to derive the information nonanticipative RDF of the multidimensional Gaussian-Markov process given by (1.17).

**Theorem IV.8.** (Alternative characterization of solution of the information nonanticipative RDF)

An alternative expression of the information nonanticipative RDF, $R_{0,n}^{na}(D)$, is

$$R_{0,n}^{na}(D) = \sup_{s \leq 0} \lambda \in \Psi_s \left\{ sD(n + 1) + \sum_{i=0}^{n} \int_{X_0,i \times Y_0,i-1} \log \left( \lambda_i(x^i, y^{i-1}) \right) P_{0,i-1}(dx^{i-1}, dy^{i-1}) \otimes p_i(dx^i; x^{i-1}) \right\}$$  \hspace{1cm} (IV.10)

where

$$\Psi_s \triangleq \left\{ \lambda \triangleq \{ \lambda_i(x^i, y^{i-1}) \geq 0 : i = 0, 1, \ldots, n \} : \int_{X_0,i} e^{\rho_i(x^i,y^i)} \lambda_i(x^i, y^{i-1}) P_{0,i}(dx^i|y^{i-1}) \leq 1, \forall y^i \in Y_{0,i}, i = 0, 1, \ldots, n \right\}.$$  \hspace{1cm} (IV.11)

Moreover, for each $s \leq 0$, a necessary and sufficient condition for $\lambda$ to achieve the supremum in (IV.10) is the existence of $\nu^*_i(\cdot; \cdot)$ related to $\lambda_i(\cdot; \cdot)$ via (IV.8) such that (IV.7) holds with equality a.a. $y_i \in Y_i$, where $\nu_i(dy_i; y^{i-1}) > 0$, $i = 0, 1, \ldots, n$.

**Proof:** See Appendix E

C. Examples of Nonanticipative RDF

In this section, we compute the optimal reproduction distribution of the information nonanticipative RDF and rate $R_{0,n}^{na}(D)$ for (i) a finite alphabet source with memory, the BSMS($p$), and (ii) a continuous alphabet source with memory, the multidimensional partially observed Gaussian-Markov stationary system described in state space form.

**Binary Symmetric Markov Source (BSMS)(p):**

Consider a BSMS($p$), with stationary transition probabilities $\{ P_{X_i|X_{i-1}}(x_i|x_{i-1}) : (x_i, x_{i-1}) \in \{0, 1\} \times \{0, 1\} \}$ given by $P_{X_i|X_{i-1}}(0|0) = P_{X_i|X_{i-1}}(1|1) = 1 - p$, $P_{X_i|X_{i-1}}(1|0) = P_{X_i|X_{i-1}}(0|1) = p$, $i \in 0, 1, \ldots$. We consider a single letter Hamming distortion criterion, $\rho(x, y) = 0$ if $x = y$ and $\rho(x, y) = 1$ if $x \neq y$. The main results are given in the next theorem.
Theorem IV.9. The nonanticipative RDF rate, $R_{n^a}(D)$, for the BSMS($p$) with single letter Hamming distortion function is given by

$$R_{n^a}(D) = \begin{cases} H(m) - H(D) & \text{if } D \leq \frac{1}{2}, \ m = 1 - p - D + 2pD \\ 0 & \text{otherwise} \end{cases}$$

the optimal (stationary) reproduction distribution is given by

$$P_{Y_i|Y_{i-1},X_i}(y_i|y_{i-1},x_i) = \begin{bmatrix} \alpha & 1 - \beta & \beta & 1 - \alpha \\ 1 - \alpha & \beta & 1 - \beta & \alpha \end{bmatrix} \quad (IV.12)$$

where

$$\alpha = \frac{(1-p)(1-D)}{1-p-D+2pD}, \ \beta = \frac{p(1-D)}{p+D-2pD}. \quad (IV.13)$$

and $\{Y_i: \ i = 0, 1 \ldots \}$ is a first-order Markov with the same transition probability as that of the source.

Proof: The steady state distribution of the source is given by $P_X(X_i = 0) = P_X(0) = P_X(X_i = 1) = P_X(1) = 0.5$. By Theorem [IV.9] (see Remark [IV.3]), the stationary reproduction distribution is Markov with respect to the source, and it is given by

$$P_{Y_i|Y_{i-1},X_i}(y_i|y_{i-1},x_i) = \frac{e^{s\rho(x,y)}}{\sum_{y_i \in \{0,1\}} e^{s\rho(x,y)}} P_{Y_i|Y_{i-1}}(y_i|y_{i-1}), \ i = 0, 1, \ldots \quad (IV.14)$$

We calculate $P_{Y_i|Y_{i-1}}(dy_i|y_{i-1})$ by reconditioning it on $X_i$ and then substitute it into the RHS of (IV.14), to deduce that the optimal stationary nonanticipative reproduction distribution depends only on $Y_{i-1}$, and it is conditional independent of $Y_{i-2}$, that is, $P_{Y_i|Y_{i-1},X_i}(y_i|y_{i-1},x_i) = P_{Y_i|Y_{i-1},X_i}(y_i|y_{i-1},x_i) = a.a. (y_{i-1},x_i)$, and it is given by (IV.12). Moreover, by using (IV.12), we obtain the transition probability of the optimal reproduction $\{Y_i: \ i = 0, 1, \ldots \}$, given by $P_{Y_i|Y_{i-1}}(0|0) = P_{Y_i|Y_{i-1}}(1|1) = 1 - p$ and $P_{Y_i|Y_{i-1}}(1|0) = P_{Y_i|Y_{i-1}}(0|1) = p$. Notice that $\{Y_i: \ i = 0, 1, \ldots \}$ is a first-order Markov with the same transition probability as that of the source $\{X_i: \ i = 0, 1, \ldots \}$.

The Lagrange multiplier “$s$” is found from fidelity constraint as follows.

$$E\{\rho(x_i,y_i)\} = \frac{e^s}{1+e^s} = D \quad \Rightarrow e^s = \frac{D}{1-D}, \ D \leq 0.5 \quad (IV.15)$$
Finally, the information nonanticipative RDF is computed using the expression

\[ R^{na}(D) = \sum_{x_i, y_i, y_i-1} P^*_X(x_i, y_i, y_i-1) \log \left( \frac{P^*_Y(y_i | y_i-1, x_i)}{P^*_Y(Y_i | Y_i-1, X_i(X_i, Y_i-1, x_i))} \right). \]  

(IV.16)

The graph of \( R^{na}(D) \) is illustrated in Fig. IV.2, which shows that, as \( p \) tends to \( \frac{1}{2} \), and the source becomes less correlated the rate increases.

Note that for \( p = \frac{1}{2} \), then BSMS(\( \frac{1}{2} \)) is the IID Bernoulli source, and \( R^{na}(D) = 1 - H(D) \equiv R(D), \; D < \frac{1}{2} \), as expected.

Fig. IV.2. \( R^{na}(D) \) for different values of parameter \( p \).

In the next section we use \( R^{na}(D) \) to compute the RL of causal codes, and to compare \( R^{na}(D) \) to the upper bound given in [25].

**Multidimensional Partially Observed Gaussian Process and JSCC:**

Here, we consider a multidimensional partially observed Gaussian-Markov process and we compute the closed form expression of the information nonanticipative RDF, \( R^{na}(D) \).

Consider the following discrete-time multidimensional partially observed linear Gauss-Markov system described by

\[
\begin{align*}
Z_{t+1} &= AZ_t + BW_t, \; Z_0 = z, \; t \in \mathbb{N} \\
X_t &= CZ_t + NV_t, \; t \in \mathbb{N}
\end{align*}
\]  

(IV.17)
where \( Z_t \in \mathbb{R}^m \) is the state (unobserved) process and \( X_t \in \mathbb{R}^p \) is the information source, obtained from noisy measurements of \( CZ_t \). The model in (IV.17) is often encountered in applications where the process \( \{Z_t : t \in \mathbb{N}\} \) is not directly observed; instead, what is directly observed is the process \( \{X_t : t \in \mathbb{N}\} \) which is a noisy version of it. This is a realistic model for any sensor which collects information for the underlying process \( CZ_t \), since the sensor is a measurement device often subject to additive Gaussian noise. Hence, in this application the objective is to compress the sensor data, which is the only observable information. Next, we introduce certain assumptions which are sufficient for existence of the limit, \( R_{na}^{0,n}(D) \triangleq \lim_{n \to \infty} \frac{1}{n+1} R_{na}^{0,n}(D) \).

(G1) \((C,A)\) is detectable and \((A, \sqrt{BB^T})\) is stabilizable, \((N \neq 0)\) \(^{[44]}\);  
(G2) The state and observation noise \( \{(W_t, V_t) : t \in \mathbb{N}^n\} \) are Gaussian IID vectors \( W_t \in \mathbb{R}^k \), \( V_t \in \mathbb{R}^d \), mutually independent with parameters \( N(0, I_{k \times k}) \) and \( N(0, I_{d \times d}) \), independent of the Gaussian RV \( Z_0 \), with parameters \( N(\bar{z}_0, \bar{\Sigma}_0) \).  
(G3) The distortion function is single letter defined by \( d_{0,n}(x^n, y^n) \triangleq \sum_{t=0}^{n} ||x_t - y_t||_2^2 \).

For the fully observed scalar case corresponding to \( X_t = Z_t \in \mathbb{R} \), the reconstruction of \( \{Z_t : t \in \mathbb{N}^n\} \) and its realization over a scalar additive white Gaussian noise (AWGN) channel is discussed in \(^{[12]}\), while the partially observed (IV.17) for the scalar case \( X_t \in \mathbb{R} \), is discussed in \(^{[14]}\) via indirect methods. However, as pointed out in \(^{[11]}\), the computation of the nonanticipative RDF for the vector Gaussian process is unsolved. Here, we show that the conjecture stated in \(^{[18]}\) is indeed true, by invoking an upper bound and Theorem (IV.8) which facilitates the computation of a lower bound similar to Shannon’s lower bound, which is achievable. To this end, we provide a closed form expression to the nonanticipative RDF for the vector Gaussian process.
According to Theorem IV.2, the optimal stationary reproduction distribution is given by

\[ P_{Y_t|Y_{t-1},X_t}^*(dy_t|y^{t-1},x^t) = \frac{e^{s||y_t-x_t||_2^2} P_{Y_t|Y_{t-1}}^*(dy_t|y^{t-1})}{\int_{Y_t} e^{s||y_t-x_t||_2^2} P_{Y_t|Y_{t-1}}^*(dy_t|y^{t-1})}, \quad s \leq 0 \]

\[ \equiv P_{Y_t|Y_{t-1},X_t}^*(dy_t|y^{t-1},x_t) - a.a. (y^{t-1}, x_t). \]  

(IV.18)

Hence, from (IV.18), it follows that the optimal reproduction is Markov with respect to the process \( \{X_t : t \in \mathbb{N}\} \). Moreover, since the exponential term \( ||y_t - x_t||_2^2 \) in the RHS of (IV.18) is quadratic in \( (x_t, y_t) \), and \( \{Z_t : i \in \mathbb{N}\} \) is Gaussian then \( \{(Z_t, X_t) : t \in \mathbb{N}\} \) are jointly Gaussian, and it follows that a Gaussian distribution \( P_{Y_t|Y_{t-1},X_t}(\cdot|y^{t-1},x_t) \) (for a fixed realization of \( (y^{t-1}, x_t) \)), and Gaussian distribution \( P_{Y_t|Y_{t-1}}(\cdot|y^{t-1}) \) can match the left and right side of (IV.18). Therefore, at any time \( t \in \mathbb{N} \), the output \( Y_t \) of the optimal reconstruction channel depends on \( X_t \) and the previous outputs \( Y^{t-1} \), and its conditional distribution is Gaussian. Hence, the channel connecting \( \{X_t : t \in \mathbb{N}\} \) to \( \{Y_t : t \in \mathbb{N}\} \) has the general form

\[ Y_t = \bar{A} X_t + B Y^{t-1} + V^c_t, \quad t \in \mathbb{N} \]  

(IV.19)

where \( \bar{A} \in \mathbb{R}^{p \times p}, \bar{B} \in \mathbb{R}^{p \times t_p}, \) and \( \{V^c_t : t \in \mathbb{N}\} \) is an independent sequence of Gaussian vectors \( N(0; Q) \).

Introduce the error estimate \( \{K_t : t \in \mathbb{N}\} \), and its covariance \( \{\Lambda_t : t \in \mathbb{N}\} \), defined by

\[ K_t \triangleq X_t - \hat{X}_{t|t-1}, \quad \hat{X}_{t|t-1} \triangleq \mathbb{E}\left\{ X_t | \sigma\{Y^{t-1}\} \right\}, \quad \Lambda_t \triangleq \mathbb{E}\{K_t K_t^T\}, \quad t \in \mathbb{N} \]  

(IV.20)

where \( \sigma\{Y^{t-1}\} \) is the \( \sigma \)-algebra generated by the sequence \( \{Y^{t-1}\} \). The covariance is diagonalized by introducing a unitary transformation \( \{E_t : t \in \mathbb{N}\} \) such that

\[ E_t \Lambda_t E_t^T = diag\{\lambda_{t,1}, \ldots, \lambda_{t,p}\}, \quad \Gamma_t \triangleq E_t K_t, \quad t \in \mathbb{N}. \]  

(IV.21)

Note that although \( \{\Gamma_t : t \in \mathbb{N}\} \) has independent Gaussian components, each component is correlated. Analogously, introduce to the process \( \{\hat{K}_t : t \in \mathbb{N}\} \) defined by

\[ \hat{K}_t \triangleq Y_t - \hat{X}_{t|t-1}, \quad \hat{\Gamma}_t = E_t \hat{K}_t, \quad t \in \mathbb{N}. \]  

(IV.22)

We shall compute the information nonanticipative RDF by considering the realization shown in Fig. IV.4, where \( \{V^t_c : t = 0, 1, \ldots\} \) is Gaussian \( N(0; Q) \), and \( \{A_t, B_t : t = 0, 1, \ldots\} \) are to be determined. Note that the square error fidelity criterion \( d_{0,n}(\cdot, \cdot) \) is not affected by the preprocessing and post processing of \( \{(X_t, Y_t) : t \in \mathbb{N}\} \), since \( d_{0,n}(X^n, Y^n) = d_{0,n}(K^n, \hat{K}^n) = \)
Next, we state the main results.

**Theorem IV.10.** \((R_{na}(D)\) of multidimensional partially observed Gaussian source)  
Under Assumptions (G1)-(G3), the information nonanticipative RDF rate for the multidimensional partially observed Gaussian source (IV.17) is given by  
\[
R_{na}(D) = \frac{1}{2} \sum_{i=1}^{p} \log \left( \frac{\lambda_{\infty,i}}{\delta_{\infty,i}} \right) \]  
where \(\text{diag}\{\lambda_{\infty,1}, \ldots, \lambda_{\infty,p}\} = \lim_{t \to \infty} E_t \Lambda_t E_t^{tr} = E_\infty \Lambda_\infty E_\infty^{tr},\)  
\[
\Lambda_\infty = \lim_{t \to \infty} \mathbb{E} \left\{ \left( C(Z_t - \mathbb{E}\{Z_t|Y^{t-1}\}) + NV_i \right) \left( C(Z_t - \mathbb{E}\{Z_t|Y^{t-1}\}) + NV_i \right)^{tr} \right\} = C \lim_{t \to \infty} \Sigma_t C^{tr} + NN^{tr} = C \Sigma_\infty C^{tr} + NN^{tr} \]  
\[
\delta_{\infty,i} \triangleq \begin{cases} \xi_\infty & \text{if } \xi_\infty \leq \lambda_{\infty,i} \\
\lambda_{\infty,i} & \text{if } \xi_\infty > \lambda_{\infty,i} \end{cases}, \quad i = 2, \ldots, p \]
and $\xi_\infty$ is chosen such that $\sum_{i=1}^p \delta_{\infty,i} = D$. Moreover, $\Sigma_\infty$ is the steady state covariance of the error $Z_t - \mathbb{E}\{Z_t|Y_t-1\} \sim N(0, \Sigma_\infty)$, $\hat{Z}_{t|t-1} \triangleq \mathbb{E}\{Z_t|Y_{t-1}\}$, of the Kalman filter given by

$$\hat{Z}_{t+1|t} = A\hat{Z}_{t|t-1} + A\Sigma_\infty(E_{t|t}^{tr}H_{t|t}E_{t|t}C)^{tr}M_\infty^{-1}(Y_t - C\hat{Z}_{t|t-1}), \quad \hat{Z}_0 = \mathbb{E}\{Z_0|Y^{-1}\}, \quad Z_0 - \hat{Z}_0 \sim N(0, \Sigma_\infty).$$

(IV.27)

$$\Sigma_\infty = A\Sigma_\infty A^{tr} - A\Sigma_\infty(E_{t|t}^{tr}H_{t|t}E_{t|t}C)^{tr}M_\infty^{-1}(E_{t|t}^{tr}H_{t|t}E_{t|t}C)\Sigma_\infty A^{tr} + BB_\infty^{tr}$$

(IV.28)

$$M_\infty = E_{t|t}^{tr}H_{t|t}E_{t|t}C\Sigma_\infty(E_{t|t}^{tr}H_{t|t}E_{t|t}C)^{tr} + E_{t|t}^{tr}H_{t|t}E_{t|t}NN^{tr}(E_{t|t}^{tr}H_{t|t}E_{t|t})^{tr} + E_{t|t}^{tr}B_{t|t}^{tr}B_{t|t}^{tr}E_{t|t}$$

(IV.29)

and

$$H_\infty = \lim_{t \to \infty} H_t, \quad H_t \triangleq diag\{\eta_{t,1}, \ldots, \eta_{t,p}\}, \quad \eta_{t,i} = 1 - \frac{\delta_{t,i}}{\lambda_{t,i}}, \quad i = 1, \ldots, p, \quad t \in \mathbb{N}$$

(IV.30)

$$B_\infty = \lim_{t \to \infty} B_t = \sqrt{H_\infty\Delta_\infty Q^{-1}}, \quad B_t \triangleq \sqrt{H_t\Delta_t Q^{-1}}, \quad t \in \mathbb{N}$$

(IV.31)

$$\Delta_\infty = \lim_{t \to \infty} \Delta_t, \quad \Delta_t = diag\{\delta_{t,1}, \ldots, \delta_{t,p}\}, \quad t \in \mathbb{N}.$$  

(IV.32)

Proof: See Appendix F. □

In the next remark, we confirm that Theorem [IV.10] gives, as a special case, the value of $R^{na}(D)$ for scalar Gaussian stationary source found in [11, Theorem 3].

**Remark IV.11.** Consider the special case of first-order (scalar) Gaussian-Markov source [11, Theorem 3]

$$X_{t+1} = \alpha X_t + \sigma_w W_t, \quad W_t \sim N(0, 1).$$

This corresponds to the dynamical system [IV.17] with $m = p = 1$, $C = 1$, $N = 0$, $A = \alpha$, $B = \sigma_w$, i.e., $\sigma_w W_t \sim N(0, \sigma_w^2)$, hence $X_t = Z_t$. Clearly, $\Lambda_\infty = \Sigma_\infty$, $\Delta_\infty = D$, where $H_\infty = 1 - \frac{D}{\Sigma_\infty}$ and $E_\infty = 1$.

Using (IV.29), we have

$$M_\infty = \Sigma_\infty \Sigma_\infty^2 + H_\infty D = H_\infty(\Sigma_\infty H_\infty + D) = H_\infty(\Sigma_\infty(1 - \frac{D}{\Sigma_\infty}) + D) = \Sigma_\infty H_\infty.$$  

(IV.33)
Also, by (IV.28), we get
\[\Sigma_{\infty} = \alpha^2\Sigma_{\infty} - \alpha^2\Sigma_{\infty}^2H_{\infty}^2M^{-1} + \sigma_w^2 \equiv \alpha^2\Sigma_{\infty} - \alpha^2\Sigma_{\infty}^2H_{\infty}^2H_{\infty}^{-1}\Sigma_{\infty}^{-1} + \sigma_w^2 = \alpha^2\Sigma_{\infty} - \alpha^2\Sigma_{\infty}H_{\infty} + \sigma_w^2 = \alpha^2\Sigma_{\infty} - \alpha^2\Sigma_{\infty}(1 - \frac{D}{\Sigma_{\infty}}) + \sigma_w^2 = \alpha^2D + \sigma_w^2 \quad (IV.34)\]
where (a) follows from (IV.33). Finally, by substituting (IV.34) in the expression of the nonanticipative RDF (IV.24) we obtain
\[R^{na}(D) = \frac{1}{2} \log \frac{|A_{\infty}|}{|\Delta_{\infty}|} = \frac{1}{2} \log \frac{\Sigma_{\infty}}{D} = \frac{1}{2} \log \left(\frac{\alpha^2D + \sigma_w^2}{D}\right) = \frac{1}{2} \log \left(\frac{\alpha^2 + \sigma_w^2}{D}\right). \quad (IV.35)\]
which is the expression derived in [11, Theorem 3]. Hence, Theorem [IV.10] generalizes previous work to multidimensional (vector) Gaussian-Markov stationary process.

In the following lemma, we show that \(\{\tilde{K}_t : t \in \mathbb{N}\}\) is the innovation process of \(\{Y_t : t \in \mathbb{N}\}\), and hence the two processes generate the same \(\sigma\)-algebras (they contain the same information).

**Lemma IV.12. (Equivalence of \(\sigma\)-algebras)**

The following hold.
\[\mathcal{F}^Y_{0,t} \triangleq \sigma\{Y_s : s = 0, 1, \ldots, t\} = \mathcal{F}^\tilde{K}_{0,t} \triangleq \sigma\{\tilde{K}_s : s = 0, 1, \ldots, t\}, \forall t \in \mathbb{N}.\]
that is, \(\mathcal{F}^Y_{0,t} \subseteq \mathcal{F}^\tilde{K}_{0,t}\) and \(\mathcal{F}^\tilde{K}_{0,t} \subseteq \mathcal{F}^Y_{0,t}\), \(\forall t \in \mathbb{N}\).

**Proof:** Since \(\tilde{K}_s = Y_s - \mathbb{E}\{X_s|Y^{s-1}\}, 0 \leq s \leq t\), then \(\mathcal{F}^\tilde{K}_{0,t} \subseteq \mathcal{F}^Y_{0,t}, \forall t \in \mathbb{N}\). Hence, we need to show that \(\mathcal{F}^Y_{0,t} \subseteq \mathcal{F}^\tilde{K}_{0,t}, \forall t \in \mathbb{N}\). The innovation process of \(\{Y_t : t \in \mathbb{N}\}\) is by definition (see Fig. [IV.4], (IV.22). (F.1))
\[
I_t = Y_t - \mathbb{E}\{Y_t|Y^{t-1}\} = E^{H}_\infty H_{\infty}E_{\infty}\left(X_t - \mathbb{E}\{X_t|Y^{t-1}\}\right) + E^{V}_\infty B_{\infty}V_{t} + \mathbb{E}\{X_t|Y^{t-1}\} - \mathbb{E}\{X_t|Y^{t-1}\} = E^{H}_\infty H_{\infty}E_{\infty}\left(X_t - \mathbb{E}\{X_t|Y^{t-1}\}\right) + E^{V}_\infty B_{\infty}V_{t} = \tilde{K}_t. \quad (IV.36)
\]
Since the innovation process \(\{I_s : 0 \leq s \leq t\}\) and the optimal reproduction process \(\{Y_s : 0 \leq s \leq t\}\) generates the same \(\sigma\)-algebras, then \(\mathcal{F}^I_{0,t} \subseteq \mathcal{F}^Y_{0,t}, \mathcal{F}^Y_{0,t} \subseteq \mathcal{F}^I_{0,t}\), i.e., \(\mathcal{F}^Y_{0,t} = \mathcal{F}^I_{0,t}\), and hence, by (IV.36) we also obtain \(\mathcal{F}^Y_{0,t} \subseteq \mathcal{F}^\tilde{K}_{0,t}\). This completes the proof. \(\Box\)

We now observe the following consequence of Lemma [IV.12]
Remark IV.13. By Lemma IV.12 all conditional expectations with respect to the process \( \{Y_t : t = 0,1,\ldots\} \) can be replaced by conditional expectations with respect to the independent process \( \{\tilde{K}_t : t = 0,1,\ldots\} \). Hence, the process \( \{K_t : t = 0,1,\ldots\} \) can be written as \( K_t = X_t - \mathbb{E}\{X_t|\sigma\{Y^{t-1}\}\} = X_t - \mathbb{E}\{X_t|\sigma\{\tilde{K}^{t-1}\}\} \), and its reconstruction is given by

\[
\tilde{K}_t = E_{\infty}^{tr} H_\infty E_\infty \left(X_t - \mathbb{E}\{X_t|\tilde{K}^{t-1}\}\right) + E_{\infty}^{tr} B_\infty V^c_t = E_{\infty}^{tr} H_\infty K_t + E_{\infty}^{tr} B_\infty V^c_t, \quad t = 0,1,\ldots
\]

Moreover, by Lemma IV.12 \( K_t \) and \( \tilde{K}_t \) are independent of \( Y_0,\ldots,Y_{t-1} \), and \( \tilde{K}_0,\ldots,\tilde{K}_{t-1} \), \( t = 0,1,\ldots \). This property is analogous to the JSCC of a scalar RV over a scalar additive Gaussian noise channel with feedback \([45\, \text{Theorem 5.6.1}]\).

Joint Source-Channel Coding. We are now ready to show that Fig. IV.4 is indeed the realization of the optimal reproduction distribution, over an additive Gaussian vector channel with feedback, and that it corresponds to JSCC using SbS transmission. That is, for a given \( D \in [D_{\text{min}}, D_{\text{max}}] \), there exists a power \( P \) such that \( R^a(D) = C(P) \), where \( C(P) \) is the capacity of the additive Gaussian vector channel with total power constraint and the end-to-end average distortion is satisfied. Consider a vector channel \( B_t = A_t + V^c_t \), \( t \in \mathbb{N} \), where \( V^c_t \) is Gaussian zero mean, \( Q \triangleq \text{Cov}(V^c_t) = \text{diag}\{q_1,q_2,\ldots,q_p\} \), and \( A_t \in \mathbb{R}^p \). By the memoryless nature of the channel we know that directed information \([31]\) is equal to mutual information and that \( I(A^n \rightarrow B^n) = I(A^n; B^n) \geq I_{X^n 
rightarrow Y^n}(P_{X^n}, \hat{P}_{Y^n|X^n}) \). Recall that the capacity of a parallel memoryless Gaussian channel with feedback subject to a power constraint \( \frac{1}{n+1} \mathbb{E}\{\sum_{t=0}^n \|A_t\|_{\mathbb{R}^p} \leq P \} \), is given by

\[
C(P) = \lim_{n \rightarrow \infty} \frac{1}{2} \frac{1}{n+1} \sum_{t=0}^n \sum_{i=1}^p \log(1 + \frac{\mathbb{E}(A_i|^2)}{(2q_i)}) \equiv \frac{1}{2} \sum_{i=1}^p \log(1 + \frac{P^i}{q_i}) \equiv C(P^1,\ldots,P^p),
\]

\[
\sum_{i=1}^p P^i = P, \quad P^i = \lim_{t \rightarrow \infty} \mathbb{E}(A_i^t)^2.
\]

In view of Fig. IV.4 the conditional distribution of nonanticipative RDF is realized via an encoder-channel decoder. Moreover, the channel consists of parallel additive Gaussian noise channels with feedback given by

\[
B^i_t = A^i_\infty(X_t,B_t^{t-1}) + V^c_{t,i}, \quad t \in \mathbb{N}, \quad i = 1,\ldots,p.
\]

where \( A_\infty(X_t,B_t^{t-1}) = A_\infty E_\infty K_t \) and \( A_\infty \triangleq \sqrt{Q \Delta^{-1}_\infty H_\infty} \). Then for a given \( D > 0 \), we can let
\( P = D \), and \( \frac{P_i}{q_i} = \frac{\lambda_{\infty,i}}{\delta_{\infty,i}} - 1 \), to obtain

\[
R_{na}(D) = \lim_{n \to \infty} \inf_{P_{Y|X} \in \mathcal{P}_{0,1}(D)} \frac{1}{n+1} \mathbb{I}_{X^n \to Y^n}(P_{X^n}, P_{Y^n|X^n})
\]

\[
= \lim_{n \to \infty} \frac{1}{2n+1} \sum_{i=0}^{n} \sum_{i=1}^{p} \log \left( \frac{\lambda_{t,i}}{\delta_{t,i}} \right) = \frac{1}{2} \sum_{i=1}^{p} \log \left( \frac{\lambda_{\infty,i}}{\delta_{\infty,i}} \right) = \frac{1}{2} \log \left| \Lambda_{\infty} \right| = C\left(P_{1}^{\infty}, \ldots, P_{p}^{\infty} \right) \bigg|_{\frac{P_i}{q_i} = \frac{\lambda_{\infty,i}}{\delta_{\infty,i}} - 1, i=1, \ldots, p}.
\]

(IV.37)

Thus, for a given distortion level \( D > 0 \), the realization shown in Fig. IV.4 is optimal in the sense that the end-to-end nonanticipative RDF, \( R_{na}(D) \) is achieved (with the prescribed average distortion), the encoder achieves the capacity, and (IV.37) is satisfied. Thus, \( R_{na}(D) \) is achievable over the noisy channel.

In the next section, we also discuss achievability with respect to excess distortion probability.

V. Coding Theorems and Bounds

In section, we describe a noisy (and noiseless) coding theorem based on nonanticipative transmission, for sources with memory, with respect to the excess distortion probability, establishing an operational meaning for \( R_{na}(D) \). Subsequently, we apply it to the BSMS(\( p \)), to establish an operational meaning to its rate \( R_{na}(D) \). Moreover, we use \( R_{na}(D) \) to derive bounds for the OPTA by causal codes [4], and for the OPTA of noncausal codes.

A. Noisy Coding Theorem: Nonanticipative Transmission

The noisy coding theorem is similar to the one illustrated in Fig. IV.4 for Gaussian sources. We focus on JSCC based on nonanticipative code, that is, the encoder and decoder at each time instant \( i \) process samples causally, with memory on past symbols, and without anticipation with respect to symbols occurring at times \( j > i \). For this scheme, we show that coded and uncoded nonanticipative transmission of sources with memory has an operational meaning, in the sense that excess distortion probability can be made arbitrarily small, based only on the properties of the information nonanticipative RDF. Fig. V.5 describes the block diagram of JSCC based on nonanticipative transmission. We assume that the cost of transmitting symbols over the channel is a measurable function

\[
c_{0,n} : \mathcal{A}_{0,n} \times \mathcal{B}_{0,n-1} \to [0, \infty), \quad c_{0,n}(a^n, b^{n-1}) \triangleq \sum_{i=0}^{n} \gamma_i(a^i, b^{i-1}).
\]

(V.1)

May 8, 2014 DRAFT
We use the following definition of a nonanticipative code.

**Definition V.1. (Nonanticipative code)**

An \((n, d, \epsilon, P)\) nonanticipative code is a tuple

\[
\left( \mathcal{X}_0, n, \mathcal{A}_0, n, \mathcal{B}_0, n, \mathcal{Y}_0, n, P_{X^n}, \overrightarrow{P}_{A^n|B^{n-1}, X^n}, \overrightarrow{P}_{B^n|A^n}, \overrightarrow{P}_{Y^n|B^n}, d_{0,n}, c_{0,n} \right)
\]

where \(\overrightarrow{P}_{A^n|B^{n-1}, X^n} \sim \{ P_{A_i|A_{i-1}, B_{i-1}, X_i} (\cdot|\cdot, \cdot) : i \in \mathbb{N}^n \} \), \(\overrightarrow{P}_{Y^n|B^n} \sim \{ P_{Y_i|Y_{i-1}, B_i} (\cdot|\cdot, \cdot) : i \in \mathbb{N}^n \} \) is the code, \(\overrightarrow{P}_{B^n|A^n} \sim \{ P_{B_i|B_{i-1}, A_i} (\cdot|\cdot, \cdot) : i \in \mathbb{N}^n \} \) is the channel, with excess distortion probability

\[
P \left\{ d_{0,n}(X^n, Y^n) > (n+1)d \right\} \leq \epsilon, \ \epsilon \in (0, 1), \ d > 0
\]

and transmission cost

\[
\frac{1}{n+1} \mathbb{E} \left\{ c_{0,n}(A^n, B^{n-1}) \right\} \leq P, \ P > 0
\]

where \(P\) and \(\mathbb{E}\) are taken with respect to the joint distribution induced by source-encoder-channel-decoder \(P_{X^n, A^n, B^n, Y^n}(dx^n, da^n, db^n, dy^n)\).

An uncoded nonanticipative code, denoted by \((n, d, \epsilon, P)\), is a subset of an \((n, d, \epsilon, P)\) nonanticipative code in which an encoder and decoder are identity maps, \(P_{A_i|A_{i-1}, B_{i-1}, X_i} (da_i|a^{i-1}, b^{i-1}, x^i) = \delta_{X_i}(da_i), \ P_{Y_i|Y_{i-1}, B_i} (dy_i|y^{i-1}, b^i) = \delta_{B_i}(dy_i)\), that is, \(A_i = X_i, Y_i = B_i, i = 0, 1, \ldots, n\).

Next, we define the minimum excess distortion as follows.
Definition V.2. (Minimum excess distortion)
The minimum excess distortion achievable by a nonanticipative code \((n,d,\epsilon,P)\) is defined by

\[
D^o(n,\epsilon,P) \triangleq \inf \left\{ d : \exists (n,d,\epsilon,P) \text{ nonanticipative code} \right\}. \tag{V.2}
\]

For uncoded nonanticipative code \((V.2)\) is replaced by

\[
\bar{D}^o(n,\epsilon,P) \triangleq \inf \left\{ d : \exists (n,d,\epsilon) \text{ nonanticipative code} \right\}. \tag{V.3}
\]

Note that in our definition of nonanticipative code \((n,d,\epsilon,P)\) we have assumed indirectly that the finite time information capacity is defined by

\[
C_{0,n}(P) \triangleq \sup_{\{P_{A_i|A_{i-1},B_{i-1}}(a_i|a^{i-1},b^{i-1}) : i=0,1,\ldots,n\} \in \mathcal{P}_{0,n}(P)} I(A^n \rightarrow B^n) \tag{V.4}
\]

where the average power constraint is

\[
\mathcal{P}_{0,n}(P) \triangleq \left\{ \{P_{A_i|A_{i-1},B_{i-1}}(a_i|a^{i-1},b^{i-1}) : i = 0,1,\ldots,n\} : \frac{1}{n+1} \mathbb{E}\{c_{0,n}(a^n,b^{n-1})\} \leq P \right\} \tag{V.5}
\]

and \(I(A^n \rightarrow B^n)\) is the directed information from \(A^n\) to \(B^n\) defined by

\[
I(A^n \rightarrow B^n) \triangleq \sum_{i=0}^{n} I(A^i;B_i|B^{i-1}). \tag{V.6}
\]

The information channel capacity is given by

\[
C(P) = \lim_{n \to \infty} \frac{1}{n+1} C_{0,n}(P). \tag{V.7}
\]

Thus, we assumed that the supremum \((V.4)\) is finite and the limit exists.

Since we consider nonanticipative transmission, a necessary condition for JSCC is the probabilistic realization of the optimal nonanticipative reproduction distribution of the information nonanticipative RDF by an encoder-channel-decoder processing symbols causally. Hence, we introduce the following definition of realization.

Definition V.3. (Realization)

Given a source \(\{P_{X_i|X_{i-1}}(dx_i|x^{i-1}) : i \in \mathbb{N}^n\}\), a channel \(\{P_{B_i|B_{i-1},A_i}(db_i|b^{i-1},a^i) : i \in \mathbb{N}^n\}\) is a realization of the optimal reproduction distribution \(\{P^*_{Y_i|Y_{i-1},X_i}(dy_i|y^{i-1},x^i) : i \in \mathbb{N}^n\}\), if there
exists a pre-channel encoder \( \{ P_{A_i|A^{i-1},B^{i-1},X^i} : i \in \mathbb{N} \} \) and a post-channel decoder \( \{ P_{Y_i|Y^{i-1},B_i} : i \in \mathbb{N} \} \) such that

\[
\overrightarrow{P}_{Y^n|X^n}(dy^n|x^n) = \otimes_{i=0}^nP_{Y_i|Y^{i-1},X^i}(dy_i|y^{i-1},x^i) = \otimes_{i=0}^nP_{Y_i|Y^{i-1},X^i}(dy_i|y^{i-1},x^i) \quad (V.8)
\]

where the joint distribution from which the RHS of \((V.8)\) is obtained is precisely

\[
P_{X^n,A^n,B^n,Y^n}(dx^n,da^n,db^n,dy^n) = \otimes_{i=0}^nP_{Y_i|Y^{i-1},B_i}(dy_i|y^{i-1},b^i) \otimes P_{B_i|B^{i-1},A_i}(db_i|b^{i-1},a^i)
\]

\[
\otimes P_{A_i|A^{i-1},B^{i-1},X^i}(da_i|a^{i-1},b^{i-1},x^i) \otimes P_{X_i|X^{i-1}}(dx_i|x^{i-1}).
\]

Moreover, we say that \( R_{n,a}(D) \) is realizable if in addition the realization operates with average distortion \( D \) and

\[
\lim_{n \to \infty} \frac{1}{n+1} \mathbb{I}_{X^n \to Y^n}(P_{X^n}, \overrightarrow{P}_{Y^n|X^n}) = R_{n,a}(D) \triangleq \lim_{n \to \infty} \frac{1}{n+1} R_{0,n}^{n,a}(D).
\]

Using the above definition of realization we now prove achievability of the nonanticipative code for sources with memory.

**Theorem V.4.** (Achievability of nonanticipative code)

**Part A.** (Coded transmission)

Suppose the following conditions hold.

1. \( R_{0,n}^{n,a}(D) \) has a solution and the optimal reproduction distribution is stationary.
2. \( C_{0,n}(P) \) has a solution and the maximizing distribution is stationary.
3. The optimal stationary reproduction distribution \( \overrightarrow{P}_{Y^n|X^n}(dy^n|x^n) \) given by Theorem IV.2 is realizable, and \( R_{n,a}(D) = \lim_{n \to \infty} \frac{1}{n+1} R_{0,n}^{n,a}(D) \) is also realizable.
4. For a given \( D \in [D_{\min}, D_{\max}] \) there exists a \( P \) such that \( R_{n,a}(D) = C(P) \).

If

\[
\mathbb{P}\left\{ d_{0,n}(X^n,Y^n) > (n+1)d \right\} \leq \epsilon, \quad d > D
\]

(V.9)

where \( \mathbb{P} \) is taken with respect to \( P_{Y^n,X^n}(dy^n, dx^n) = \overrightarrow{P}_{Y^n|X^n}(dy^n|x^n) \otimes P_{X^n}(dx^n) \), then there exists an \( (n,d,\epsilon,P) \) nonanticipative code.

**Part B.** (Uncoded transmission)

Suppose the following conditions hold.

1. **Condition Part A.** (1) holds.
2. The encoder and the decoder are identity maps (uncoded), and the channel \( P_{B_i|B^{i-1},A_i} \) corresponds to \( P_{Y_i|Y^{i-1},X^i} \) (i.e., \( A_i = X_i, Y_i = B_i \)), \( i = 0, 1, \ldots, n \).
(3) For a given $D \in [D_{\text{min}}, D_{\text{max}}]$, the expression $\lim_{n \to \infty} \frac{1}{n+1} I(A^n \to B^n)$ corresponding to the optimal reproduction distribution of $R_{na}(D)$ is finite.

If

$$\mathbb{P}\left\{ d_{0,n}(X^n, Y^n) > (n+1)d \right\} \leq \epsilon, \ d > D \quad \text{(V.10)}$$

where $\mathbb{P}$ is taken with respect to $P_{Y^n, X^n}(dy^n, dx^n) = \hat{P}^*_{Y^n|X^n}(dy^n|x^n) \otimes P_X(dx^n)$, then there exists an uncoded $(n, d, \epsilon)$ nonanticipative code.

**Proof:** Part A. If conditions (1)-(3) hold then the optimal reproduction distribution is stationary, it is realizable, and this realization achieves $R_{na}(D)$. By (4) $R_{na}(D) = C(P)$. If (V.9) is satisfied then an nonanticipative code exists. Part B. This is a special case of Part A.; by the data processing and condition Part B. (3), we know that $R_{na}(D) \leq \lim_{n \to \infty} \frac{1}{n+1} I(A^n \to B^n) < \infty$. Hence, if (V.10) holds, there exists an uncoded $(n, d, \epsilon)$ nonanticipative code. □

The method described in Theorem [V.4 Part A.], ensures JSCC so that the channel operates at the supremum of all achievable rates, and hence $R(D)$ is the minimum rate of reproducing source messages at the decoder, i.e., $R_{na}(D) = C(P)$. This noisy coding theorem is the one applied in the Gaussian example, with respect to the average distortion instead of the excess distortion probability. The method described in Theorem [V.4 Part B.], is simpler; find the optimal reproduction distribution of $R_{na}(D)$, then use this distribution as the channel and ensure that (V.10) holds, which implies achievability of the uncoded nonanticipative code. The only disadvantage is the loss of resources, because in general the channel corresponding to the optimal reproduction distribution of $R_{na}(D)$ will have higher capacity than the value of $R_{na}(D)$. Below, we apply Theorem [V.4 Part B.], to the BSMS($p$).

**Excess Distortion Probability of BSMS($p$).** Consider the BSMS($p$) and its nonanticipative rate $R_{na}(D)$ computed in Theorem [IV.9]. For a given $D > 0$, using uncoded transmission as in Theorem [V.4 Part B.], the exact calculation of the excess distortion probability $\mathbb{P}\left\{ d_{0,n-1}(X^{n-1}, Y^{n-1}) > nd \right\} \leq \epsilon, \ \epsilon \in (0, 1), \ d > D$, is not as straightforward as it in the case of the IID Bernoulli source [28]. However, if we can show that the joint process $\{(X_i, Y_i) : i = 0, 1, \ldots\}$ induced by the optimal reproduction distribution and the BSMS($p$) with alphabet $\Sigma \triangleq \{(x, y) : x \in \{0, 1\}, y \in \{0, 1\}\}$ is Markov, then we can find an upper bound for the excess distortion probability, for
finite but large enough $n$ so the bound is non-trivial. Express the distortion function as follows.

$$z_i \triangleq (x_i, y_i), \quad S_n \triangleq \sum_{i=0}^{n-1} f(z_i), \quad f(z) \triangleq \rho(x, y) = x \oplus y, \quad i = 0, 1, \ldots, n-1$$

and introduce the mapping $\phi : \Sigma \mapsto \Sigma \triangleq \{1, 2, 3, 4\}, (x, y) \in \Sigma$ such that under $\phi$, $(0, 0) \mapsto 1, (0, 1) \mapsto 2, (1, 0) \mapsto 3, (1, 1) \mapsto 4$.

**Theorem V.5.** The joint process $\{Z_i \triangleq (X_i, Y_i) : i = 0, 1, \ldots\}$ generated by the optimal reproduction distribution $P^{*}_{Y_i|Y_{i-1}, X_i}(y_i|y_{i-1}, x_i)$ and the BSMS(p), $P_{X_i|X_{i-1}}(x_i|x_{i-1})$ is Markov, that is, $P^{*}_{Z_i|Z_{i-1}}(z_i|z_{i-1}) = P^{*}_{Z_i|Z_{i-1}}(z_i|z_{i-1})$, $i = 0, 1, \ldots$, and its transition probability matrix denoted by $\Pi = \{\pi(i, j) \equiv P^{*}_{Z_i|Z_{i-1}}(i|j) : (i, j) \in \Sigma \times \Sigma\}$ is given by

$$\Pi = \begin{bmatrix}
\alpha(1-p) & \beta(1-p) & \alpha p & \beta p \\
(1-\alpha)(1-p) & (1-\beta)(1-p) & (1-\alpha)p & p(1-\beta) \\
(1-\beta)p & (1-\alpha)p & (1-\beta)(1-p) & (1-\alpha)(1-p) \\
\beta p & \alpha p & \beta(1-p) & \alpha(1-p)
\end{bmatrix} \quad (V.11)$$

where each column $\{\pi(\cdot, j) : j \in \{1, 2, 3, 4\}\}$ is a probability vector, where $\alpha$ and $\beta$ are defined in \[IV.13\].

**Proof:** Recall the solution of the BSMS(p) in Theorem \[IV.9\] Then for $i = 1, 2, \ldots$, we have

$$P^{*}_{X_i,Y_i|X_{i-1},Y_{i-1}}(x_i, y_i|x_{i-1}, y_{i-1}) = P^{*}_{Y_i|Y_{i-1}, X_i}(y_i|y_{i-1}, x_i)P_{X_i|X_{i-1}}(x_i|x_{i-1})$$

$$= P^{*}_{Y_i|Y_{i-1}, X_i}(y_i|y_{i-1}, x_i)P_{X_i|X_{i-1}}(x_i|x_{i-1}),$$

$$= P^{*}_{X_i,Y_i|X_{i-1},Y_{i-1}}(x_i, y_i|x_{i-1}, y_{i-1}).$$

This shows that the joint process is Markov. By simple algebra, we obtain (V.11). 

Since the process $\{Z_i : i = 0, 1, \ldots\}$ is Markov we can apply Hoeffding’s inequality \[46\], which bounds the probability of a function of the Markov chain to obtain an upper bound for the excess distortion probability as follows.

$$\mathbb{P}^*(\frac{S_n - \mathbb{E}[S_n]}{n} \geq \gamma) \leq \exp\left(-\frac{\lambda^2(n\gamma - 2\|f\|m/\lambda^2)^2}{2n\|f\|^2m^2}\right), \quad n > 2\|f\|m/(\lambda\gamma) \quad (V.12)$$
Fig. V.6. Hoeffding bound for excess distortion probability for \( p = 0.3, D = 0.1 \) and \( \gamma = 0.1 \) \((d = D + \gamma)\).
This bound is illustrated in Fig. V.7 and performs much better than Hoeffding’s bound, as illustrated in the comparison shown in Fig. V.8. Clearly, 

Next, we use results from large deviations theory to compute the error exponent of the excess distortion probability, as \( n \to \infty \). Let \( \mathbb{P}^{\pi} \) denote the joint probability associated with \( \Pi \), having an initial state \( Z_0 = \sigma \in \Sigma \) given by 

\[
\mathbb{P}^{\pi}(Z_1 = z_1, \ldots, Z_{n-1} = z_{n-1}) = \pi(z_1, \sigma) \otimes_{i=2}^{n-1} \pi(z_i, z_{i-1})
\]

and denote the expectation with respect to \( \mathbb{P}^{\pi} \) by \( \mathbb{E}^{\pi}_{\sigma}(\cdot) \). It can be verified that \( \Pi \) is irreducible, therefore we can describe the error exponent of the excess distortion probability by the Perron-Frobenius eigenvalue of a certain non-negative matrix \cite{49}.

Let \( \lambda \in \mathbb{R} \) and define the non-negative matrix \( \Pi_{\lambda} \) with eigenvalues 

\[
\pi_{\lambda}(j, i) = \pi(j, i) \exp\{\lambda f(j)\}, \quad i, j \in \Sigma.
\]

Let \( \rho(\Pi_{\lambda}) \) denote the Perron-Frobenius eigenvalue of \( \Pi_{\lambda} \). Then by \cite{49} Theorem 3.1.2, we have the following. For any set \( \Gamma \in \mathbb{R} \) and initial state \( \sigma \in \Sigma \), 

\[
-\inf_{\theta \in \Gamma} I(\theta) \leq \lim \inf_{n \to \infty} \frac{1}{n} \log \mathbb{P}^{\pi}_{\sigma}\left(\frac{S_n}{n} \in \Gamma\right) \leq \lim \sup_{n \to \infty} \frac{1}{n} \log \mathbb{P}^{\pi}_{\sigma}\left(\frac{S_n}{n} \in \Gamma\right) \leq -\inf_{\theta \in \Gamma} I(\theta)
\]
Fig. V.8. Comparison of Hoeffding bound and R-FSMC bound for excess distortion probability for \( p = 0.3, D = 0.1 \) and \( \gamma = 0.1 \) (\( d = D + \gamma \)). A logarithmic scale is applied for time index \( n \).

where \( \Gamma^o \) is the interior of \( \Gamma \) and \( \bar{\Gamma} \) is the closure of \( \Gamma \), and \( I : \mathbb{R} \rightarrow [0, \infty] \) is a convex rate function defined by

\[
I(\theta) \triangleq \sup_{\lambda \in \mathbb{R}} \left\{ \lambda \theta - \log \rho(\Pi_\lambda) \right\}.
\]  

(V.14)

As \( n \) becomes very large (i.e., \( n \rightarrow \infty \)) then

\[
\mathbb{P}_\sigma\left( \frac{S_n}{n} \geq d \right) \sim \exp\left(-n \inf_{\theta \in [d, \infty]} I(\theta)\right), \quad d \geq D
\]  

(V.15)

and the exponential decay of this probability is obtained by minimizing the rate function over \( \theta \in [d, \infty] \). Since for the evaluation of (V.15), \( I(\theta) \) is convex, non-decreasing function of \( \theta \in [d, \infty] \), then

\[
\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}_\sigma\left( \frac{S_n}{n} \geq d \right) = -I(d).
\]

The graph of \( I(\theta) \), for \( \theta \geq D \) is shown in Fig. V.9 and indicates the rate of exponential decay of the excess distortion probability as a function of \( \theta \geq D \).

B. Noiseless Coding Theorem

It is straight forward to verify by invoking Lemma III.7 and Theorem III.8 that the the coding theorem derived in [7] Chapter 5] for two dimensional sources \( X^{n,s} \triangleq \{X_{i,j} : i = 0, \ldots, n, j = \ldots, s \} \)
0, . . . , s} ∈ ×_{i=0}^{n} ×_{j=0}^{s} X_{i,j}, where i represents time index and j represents spatial index, gives an operational meaning to $R_{0,n}^{a}(D)$ (even for finite n, when the process is IID in the spatial component). For completeness, we describe this connection in Appendix G.

C. Bounds on OPTA by Causal Codes

In this section, we illustrate another application of the nonanticipative RDF by first recalling Neuhoff and Gilbert [4] OPTA by causal codes denoted by $r^{c,+}(D)$, and then showing that $r^{c,+}(D)$ is bounded below by the nonanticipative information RDF.

Consider a causal source code and define the average fidelity by

$$d^{+}(x, y) \triangleq \limsup_{k \to \infty} \frac{1}{n+1} \mathbb{E}\left\{d_{0,n}(x^{n}, y^{n})\right\}, \quad d_{0,n}(x^{n}, y^{n}) \triangleq \sum_{i=0}^{n} \rho(x_{i}, y_{i}).$$

Let $l_{n}(x^{\infty})$ denote the total number of bits received at the decoder at the time it reproduces the output sequence $\{Y_{n} : n \in \mathbb{N}\}$, when the source is $\{X_{n} : n \in \mathbb{N}\}$. In [4] the average rate of the encoder-decoder pairs using causal reproduction coders is measured by

$$\limsup_{n \to \infty} \frac{1}{n+1} \mathbb{E}\left\{l_{n}(X^{\infty})\right\}.$$ 

Moreover, the operational causal RDF for a source $\{X_{n} : n = 0, 1, \ldots\}$ is defined as follows.

Fig. V.9. Rate function (V.14) for $p = 0.3$ and $D = 0.1$. 
Definition V.6. (OPTA by causal codes)

The OPTA by causal codes subject to fidelity is defined by

\[ r_{c,+}^c(D) \triangleq \inf_{y_i = f_i(x^i), \forall i \in \mathbb{N}, d^c(x,y) \leq D} \limsup_{n \to \infty} \frac{1}{n+1} \mathbb{E} \left\{ l_n(X^n) \right\}. \quad (V.16) \]

Causal codes based on [4] are analyzed and further generalized in [50] for stationary ergodic sources, under a variety of side information available at the encoder and decoder. Although expression (V.16) is very attractive, its computation for general sources is very difficult.

Next, we show that the OPTA by causal codes is bounded below by the expression of information nonanticipative RDF rate. Consider the joint distribution defined by \( P_{X^n}(d.x^n) \), and a causal conditioning reproduction distribution \( \overrightarrow{R}_{Y^n|X^n} \), so that the randomized coders are consistent with Definition III.6. Then, by data processing inequality we have the following bounds.

\[ \mathbb{E} \left\{ l_n(X^n) \right\} \geq H(Y^n) \geq \sum_{i=0}^{n} \left\{ H(Y_i|Y^{i-1}) - H(Y_i|Y^{i-1}, X^i) \right\} \overset{(b)}{=} \ll_{X^n \to Y^n}(P_{X^n}, \overrightarrow{R}_{Y^n|X^n}) \]

where (b) follows from the fact that the joint distribution is defined by \( P_{X^n}(d.x^n) \) and the conditional reproduction distribution \( \overrightarrow{R}_{Y^n|X^n} \). Therefore, given a distortion function \( d_{0,n}(x^n, y^n) \) and a distortion level \( D \geq 0 \), for any finite time \( n \in \mathbb{N} \), by using (V.17), and taking the infimum over the reproduction codes as in (V.16), over randomized reproduction distribution \( \overrightarrow{R}_{Y^n|X^n} \in Q_{0,n}^C(D) \) we have the following bounds \( \forall n \geq 0 \).

\[ r_{0,n}^c(D) \triangleq \inf_{y_i = f_i(x^i), \forall i \in \mathbb{N}, d^c(x,y) \leq D} \frac{1}{n+1} \mathbb{E} \left\{ l_n(X^n) \right\} \geq \frac{1}{n+1} R_{0,n}^{na}(D) \geq \frac{1}{n+1} R_{0,n}(D). \]

In the previous bounds we can first take \( \limsup_{n \to \infty} \) and then the infimum giving

\[ r_{c,+}^c(D) \geq R_{0,n,+}^{na}(D) \triangleq \inf_{Q_{0,n,+}^C(D)} \limsup_{n \to \infty} \frac{1}{n+1} R_{0,n}^{na}(D) \geq R_+^+(D) \overset{\triangle}{=} \inf_{Q_{0,n,+}^C(D)} \limsup_{n \to \infty} \frac{1}{n+1} R(D). \]

Therefore, the information nonanticipative RDF, \( R_{na}(D) \), and rate \( R_{na,+}(D) \), is a lower bounds on \( r_{c,+}(D) \), the OPTA by causal codes, and an upper bounds to the classical RDF and rate \( R^+(D) \), the OPTA by noncausal codes. These bounds are investigated recently in [11] for quadratic fidelity and Gaussian stationary sources.

Evaluation of bounds for BSMS(p).
The classical RDF for the BSMS($p$) is only known for the distortion region $0 \leq D \leq D_c$ [24], and is given by

$$R(D) = H(p) - H(D) \text{ if } D \leq D_c = \frac{1}{2} \left(1 - \sqrt{1 - \left(\frac{p}{q}\right)^2}\right), \quad p \leq 0.5.$$  \hspace{1cm} (V.18)

For the remainder of the distortion region $D > D_c$ only upper and lower bounds on $R(D)$ are known, and these are derived in [25]. In fact, it is shown by Gray in [24] that (V.18) is also a lower bound for $R(D)$ in the region $D_c \leq D \leq \frac{1}{2}$. Our expression of the nonanticipative RDF provides an upper bound on the classical RDF for all possible values of $D$, $0 \leq D \leq 0.5$. Next, we compare the exact value of $R(D)$ given by (V.18) (i.e. valid for $D \leq D_c$), the lower bound (given by the same expression for all values of $D_c \leq D \leq \frac{1}{2}$), and the upper bound derived by Berger in [25, equation 46], which hold for $0 \leq D \leq \frac{1}{2}$. We illustrate that the upper bound in [25] is not as tight as the one based on $R^{na}(D)$. For the BSMS($p$), we also compute the RL of causal codes with respect to noncausal codes, by using the fact that this RL is at most $R^{na}(D) - R(D), \forall D \leq D_c$ bits/sample.

Figure [V.10] shows the exact value of $R(D)$ for $0 \leq D \leq D_c$, its lower bound (based on (V.18)), the upper bound derived in [25], Shannon’s lower bound, and the upper bound based on $R^{na}(D)$. We observe that for $p = 0.25$, the upper bound based on $R^{na}(D)$ does slightly
Fig. V.11. Comparison of the functional behaviour of $R^{na}(D)$ and $R(D)$ for BSMS($p$) with $p = 0.12$.

better than Berger’s upper bound. Moreover, since $R^{na}(D)$ is nonincreasing and convex as a function of $D$, and nonincreasing for all values of $p \in [0, 0.5]$ (these are easily shown), then the upper bound based on $R^{na}(D)$ is convex, when compared to Berger’s upper bound which is not necessarily convex and nonincreasing (as illustrated by the blue graph in Figure [V.11]).

Finally, we use the bound $R(D) \leq R^{na}(D) \leq r_{c,+}(D)$ to deduce that the RL of causal codes for the BSMS($p$) cannot exceed

$$RL = R^{na}(D) - R(D) \leq \begin{cases} H(m) - H(p) & \text{if } 0 \leq D \leq p \\ H(m) - H(D) & \text{if } D_c < D \leq 0.5. \end{cases}$$

This bound on the RL is illustrated in Figure [V.12] which demonstrates the fluctuation of the RL for $p \in [0, 0.5]$. It is interesting to see that the maximum value of the RL is 0.2144 and corresponds to $(p = 0.1012, D = 0.1012)$. This bound is exact for $D \leq D_c \leq p$. For high resolution ($D \rightarrow 0$), the classical RDF and the nonanticipative RDF are equivalent and equal to $H(p)$.

Bounds on multidimensional partially observed Gaussian source.

For the multidimensional partially observable Gaussian-Markov stationary source given in our
example, the RL of causal codes with respect to noncausal codes is at most $R^{na}(D) - R(D)$ bits/sample, where $R^{na}(D)$ is given in Theorem [IV.10] while the expression for $R(D)$ is found in [2]. On the other hand, $R^{na}(D) - R(D)$ is the RL of zero-delay codes with respect to noncausal codes. To facilitate the computation of RL of zero delay codes with respect to noncausal codes in frequency domain, one can derive the equivalent expression for $R^{na}(D)$ in frequency domain using the solution given in Theorem [IV.10]. For the scalar Gaussian stationary process such an expression is given in [21], [22].

VI. CONCLUSION

A variant of the classical RDF of the OPTA by noncausal codes, called information nonanticipative RDF, which imposes a nonanticipative or causality constraint on the optimal reproduction conditional distribution is investigated in the context of its applications in JSCC using nonanticipative transmission schemes with respect to the excess distortion probability, and in evaluating the RL of the OPTA by zero-delay and causal codes with respect to noncausal codes. These applications are employed to two working examples, the BSMS($p$) with Hamming distance distortion, and the multidimensional partially observed Gaussian-Markov source. It is our belief that the results derived in this paper provide a crucial step towards the complete investigation.

![Comparison of the Rate Loss (RL) for $p \in [0, 0.5]$.](image)
of two fundamental problems in information theory, the evaluation of nonanticipative RDF in systems where sources with memory are considered, and in nonanticipative JSCC.

APPENDIX A

PROOF OF THEOREM III.2

(1) The tightness of the set of probability measures $Q^{C\ref{thm:III.2}}(\mathcal{Y}_{0,n};\mathcal{X}_{0,n})$ is shown in [30, Theorem III.5] hence, it is omitted. By Prohorov’s theorem [32] the compactness of $Q^{C\ref{thm:III.2}}(\mathcal{Y}_{0,n};\mathcal{X}_{0,n})$ will follow if we show it is closed, that is, given \( \{\overrightarrow{Q}_{0,n}^{\alpha}(\cdot|x^n) : \alpha = 1, 2, \ldots \} \subset Q^{C\ref{thm:III.2}}(\mathcal{Y}_{0,n};\mathcal{X}_{0,n}) \) with $\overrightarrow{Q}_{0,n}^{\alpha}(\cdot|x^n) \xrightarrow{w} \overrightarrow{Q}_{0,n}^{0}(\cdot|x^n)$ then $\overrightarrow{Q}_{0,n}^{0}(\cdot|x^n) \in Q^{C\ref{thm:III.2}}(\mathcal{Y}_{0,n};\mathcal{X}_{0,n})$. Since the family of measures $\overrightarrow{Q}_{0,n}^{0}(\cdot|x^n) \in Q^{C\ref{thm:III.2}}(\mathcal{Y}_{0,n};\mathcal{X}_{0,n})$ and $\{q_{i}(\cdot;y^{i-1},x^{i}) : i = 0, 1, \ldots, n \}$, are tight, and $q_{i}(\cdot;y^{i-1},x^{i})$ are probability measures on $\mathcal{M}_{1}(\mathcal{Y}_{i})$, $i = 0, 1, \ldots, n$, then, for $\overrightarrow{Q}_{0,n}^{\alpha}(\cdot|x^n) \in Q^{C\ref{thm:III.2}}(\mathcal{Y}_{0,n};\mathcal{X}_{0,n})$ there is a collection of probability measures $\{q_{i}^{\alpha}(\cdot;y^{i-1},x^{i}) : i = 0, 1, \ldots, n \}$ such that

\[
q_{i}^{\alpha}(\cdot;y^{i-1},x^{i}) \xrightarrow{w} q_{i}^{0}(\cdot;y^{i-1},x^{i}), \ i = 0, 1, \ldots, n.
\]

Hence, to show closedness of $Q^{C\ref{thm:III.2}}(\mathcal{Y}_{0,n};\mathcal{X}_{0,n})$ it suffices to show that

\[
\otimes_{i=0}^{n} q_{i}^{\alpha}(\cdot;y^{i-1},x^{i}) \xrightarrow{w} \otimes_{i=0}^{n} q_{i}^{0}(\cdot;y^{i-1},x^{i})
\]

whenever $q_{i}^{\alpha}(\cdot;y^{i-1},x^{i}) \xrightarrow{w} q_{i}^{0}(\cdot;y^{i-1},x^{i}), \ i = 0, 1, \ldots, n$. This will be shown by induction. Consider $n = 0$. For any $h_{0}(\cdot) \in BC(\mathcal{Y}_{0})$, by the definition of weak convergence

\[
\lim_{\alpha \rightarrow \infty} \int_{\mathcal{Y}_{0}} h_{0}(y_{0})q_{0}^{\alpha}(dy_{0};x_{0}) = \int_{\mathcal{Y}_{0}} h_{0}(y_{0})q_{0}^{0}(dy_{0};x_{0}).
\]

Consider $n = 1$. For any $h(y_{0}, y_{1}) = h_{0}(y_{0})h_{1}(y_{1}), h_{0}(\cdot) \in BC(\mathcal{Y}_{0}), h_{1}(\cdot) \in BC(\mathcal{Y}_{1})$ we need to show that for any $\epsilon > 0$ there exists a $N \in \mathbb{N}$ such that for $\alpha \geq N$ we have

\[
\left| \int_{\mathcal{Y}_{0}} h_{0}(y_{0})q_{0}^{\alpha}(dy_{0};x_{0}) \int_{\mathcal{Y}_{1}} h_{1}(y_{1})q_{1}^{\alpha}(dy_{1};y_{0},x_{1}) - \int_{\mathcal{Y}_{0}} h_{0}(y_{0})q_{0}^{0}(dy_{0};x_{0}) \int_{\mathcal{Y}_{1}} h_{1}(y_{1})q_{1}^{0}(dy_{1};y_{0},x_{1}) \right| \leq \epsilon.
\]

(A.1)
Write the left hand side (LHS) of (A.1) as follows.
\[
\int_{Y_0 \times Y_1} h_0(y_0) h_1(y_1) q_1^0(dy_1; y_0, x^1) q_0^0(dy_0; x_0) - \int_{Y_0 \times Y_1} h_0(y_0) h_1(y_1) q_1^0(dy_1; y_0, x^1) q_0^0(dy_0; x_0)
\]
\[
\leq \int_{Y_0 \times Y_1} h_0(y_0) h_1(y_1) q_1^0(dy_1; y_0, x^1) q_0^0(dy_0; x_0) - \int_{Y_0 \times Y_1} h_0(y_0) h_1(y_1) q_1^0(dy_1; y_0, x^1) q_0^0(dy_0; x_0)
\]
\[
+ \int_{Y_0 \times Y_1} h_0(y_0) h_1(y_1) q_1^0(dy_1; y_0, x^1) q_0^0(dy_0; x_0) - \int_{Y_0 \times Y_1} h_0(y_0) h_1(y_1) q_1^0(dy_1; y_0, x^1) q_0^0(dy_0; x_0).
\]
(A.2)

Let \( \epsilon > 0 \) be given. By the continuity of the function mapping \((y_0, x^1) \in Y_0 \times X_{0,1} \mapsto \int_{Y_1} h(y_1) q_1(dy_1; y_0, x^1) \) and the weak convergence \( q_1^0(\cdot; y^0, x^1) \xrightarrow{w} q_1^0(\cdot; y^0, x^1) \), for each \((y_0, x^1) \in Y_0 \times X_{0,1} \), there exists a \( N \in \mathbb{N} \) such that for all \( \alpha \geq N \)
\[
\left| \int_{Y_0} h_0(y_0) \left( \int_{Y_1} h_1(y_1) q_1^0(dy_1; y_0, x^1) \right) (q_0^0(dy_0; x_0) - q_0^0(dy_0; x_0)) \right| \leq \epsilon
\]
and
\[
\left| \int_{Y_1} h_1(y_1) q_1^0(dy_1; y_0, x^1) - \int_{Y_1} h_1(y_1) q_1^0(dy_1; y_0, x^1) \right| \leq \epsilon.
\]

Hence, by (A.2) we have
\[
\left| \int_{Y_0 \times Y_1} h_0(y_0) h_1(y_1) q_1^0(dy_1; y_0, x^1) q_0^0(dy_0; x_0) - \int_{Y_0 \times Y_1} h_0(y_0) h_1(y_1) q_1^0(dy_1; y_0, x^1) q_0^0(dy_0; x_0) \right|
\leq \epsilon + \sup_{y_0 \in Y_0} |h_0(y_0)| \epsilon, \ \forall \alpha \geq N.
\]

Since \( \epsilon > 0 \) is arbitrary then (A.1) holds. Suppose that for \( n = k \) and for any \( h_i(\cdot) \in BC(Y_i), i = 0, 1, \ldots, k, \ \forall \ \epsilon > 0 \), there exists a \( N \in \mathbb{N} \) such that for any \( \alpha \geq N \)
\[
\left| \int_{Y_{0,k}} \otimes_{i=0}^k h_i(y_i) q_i^0(dy_i; y^{i-1}, x^i) - \int_{Y_{0,k}} \otimes_{i=0}^k h_i(y_i) q_i^0(dy_i; y^{i-1}, x^i) \right| \leq \epsilon.
\]
(A.3)

We show that (A.3) holds for \( n = k + 1 \), e.g.,
\[
\otimes_{i=0}^{k+1} q_i^0(\cdot; y^{i-1}, x^i) \xrightarrow{w} \otimes_{i=0}^{k+1} q_i^0(\cdot; y^{i-1}, x^i)
\]
whenever \( q_i^0(\cdot; y^{i-1}, x^i) \xrightarrow{w} q_i^0(\cdot; y^{i-1}, x^i), i = 0, 1, \ldots, k + 1 \), and \( \otimes_{i=0}^k q_i^0(\cdot; y^{i-1}, x^i) \xrightarrow{w} \otimes_{i=0}^k q_i^0(\cdot; y^{i-1}, x^i) \).
Consider $n = k + 1$. For any $h_i(\cdot) \in BC(\mathcal{Y}_i)$, $i = 0, 1, \ldots, k + 1$, then
\[
C^\alpha \triangleq \left| \int_{\mathcal{Y}_{0,k+1}} \otimes_{i=0}^{k+1} h_i(y_i) q_i^\alpha(dy_i; y_i, x_i^i) - \int_{\mathcal{Y}_{0,k+1}} \otimes_{i=0}^{k+1} h_i(y_i) q_i^0(dy_i; y_i, x_i^i) \right|
\]
\[
= \left| \int_{\mathcal{Y}_{0,k}} \left( \int_{\mathcal{Y}_{k+1}} h_{k+1}(y_{k+1}) q_{k+1}^\alpha(dy_{k+1}; y_{k}, x_{k+1}^k) \right) \otimes_{i=0}^{k} h_i(y_i) q_i^\alpha(dy_i; y_i, x_i^i) 
- \int_{\mathcal{Y}_{0,k}} \left( \int_{\mathcal{Y}_{k+1}} h_{k+1}(y_{k+1}) q_{k+1}^0(dy_{k+1}; y_{k}, x_{k+1}^k) \right) \otimes_{i=0}^{k} h_i(y_i) q_i^0(dy_i; y_i, x_i^i) \right|
\]
\[
\leq \left| \int_{\mathcal{Y}_{0,k}} \left( \int_{\mathcal{Y}_{k+1}} h_{k+1}(y_{k+1}) q_{k+1}^\alpha(dy_{k+1}; y_{k}, x_{k+1}^k) \right) \otimes_{i=0}^{k} h_i(y_i) q_i^\alpha(dy_i; y_i, x_i^i) 
- \int_{\mathcal{Y}_{0,k}} \left( \int_{\mathcal{Y}_{k+1}} h_{k+1}(y_{k+1}) q_{k+1}^0(dy_{k+1}; y_{k}, x_{k+1}^k) \right) \otimes_{i=0}^{k} h_i(y_i) q_i^0(dy_i; y_i, x_i^i) \right|
\]
\[
+ \left| \int_{\mathcal{Y}_{0,k}} \left( \int_{\mathcal{Y}_{k+1}} h_{k+1}(y_{k+1}) q_{k+1}^0(dy_{k+1}; y_{k}, x_{k+1}^k) \right) \otimes_{i=0}^{k} h_i(y_i) q_i^\alpha(dy_i; y_i, x_i^i) 
- \int_{\mathcal{Y}_{0,k}} \left( \int_{\mathcal{Y}_{k+1}} h_{k+1}(y_{k+1}) q_{k+1}^0(dy_{k+1}; y_{k}, x_{k+1}^k) \right) \otimes_{i=0}^{k} h_i(y_i) q_i^0(dy_i; y_i, x_i^i) \right|.
\]

(A.4)

Let $\epsilon > 0$ be given. By the continuity of the function mapping $(y^{i-1}, x^i) \in \mathcal{Y}_{i-1} \times \mathcal{X}_{0,i} \mapsto \int_{\mathcal{Y}_i} h_i(y_i) q_i(dy_i; y_i, x^i)$ and the weak convergence $q_i^\alpha(\cdot; y_i, x^i) \xrightarrow{w} q_i^0(\cdot; y_i, x^i)$, for each $(y^{i-1}, x^i) \in \mathcal{Y}_{i-1} \times \mathcal{X}_{0,i}$, where $i = 0, 1, \ldots, k + 1$, there exists an $N \in \mathbb{N}$ such that for all $\alpha \geq N$

\[
C^\alpha \leq \epsilon + \sup_{y^k \in \mathcal{Y}_{0,k}} |h_0(y_0) \ldots h_k(y_k)| \epsilon, \ \forall \alpha \geq N.
\]

Since $\epsilon > 0$ is arbitrary the derivation is complete.

(2) Here, we will show that for $\{\overrightarrow{Q}^\alpha_{0,n}(\cdot | x^n) : \alpha = 1, 2, \ldots\} \in \mathcal{Q}^{C_1}(D)$ such that $\overrightarrow{Q}^\alpha_{0,n}(\cdot | x^n) \xrightarrow{w} \overrightarrow{Q}^0_{0,n}(\cdot | x^n)$, then $\overrightarrow{Q}^0_{0,n}(\cdot | x^n) \in \mathcal{Q}^{C_1}(D)$. Let $\{\overleftarrow{Q}^\alpha_{0,n}(\cdot | x^n) : \alpha = 1, 2, \ldots\} \in \mathcal{Q}^{C_1}(D) \subset \mathcal{Q}^{C_1}(\mathcal{Y}_{0,n}; \mathcal{X}_{0,n})$. Since $\mathcal{Q}^{C_1}(\mathcal{Y}_{0,n}; \mathcal{X}_{0,n})$ is closed and tight, and hence compact, there exists a subsequence $\{\overrightarrow{Q}^\alpha_{0,n}(\cdot | x^n) : i = 1, 2, \ldots\} \in \mathcal{Q}^{C_1}(\mathcal{Y}_{0,n}; \mathcal{X}_{0,n})$ and a measure $\overrightarrow{Q}^0_{0,n}(\cdot | x^n) \in \mathcal{Q}^{C_1}(\mathcal{Y}_{0,n}; \mathcal{X}_{0,n})$ such that $\overrightarrow{Q}^\alpha_{0,n}(\cdot | x^n) \xrightarrow{w} \overrightarrow{Q}^0_{0,n}(\cdot | x^n)$ for each $x^n \in \mathcal{X}_{0,n}$. Recall that $d_{0,n} : \mathcal{X}_{0,n} \times \mathcal{Y}_{0,n} \mapsto [0, \infty]$ is a Borel measurable, non-negative, and continuous function on $y^n \in \mathcal{Y}_{0,n}$. Consider the sequence $\{d_{0,n}^{(k)} \triangleq d_{0,n} \wedge k : k \in \mathbb{N}\}$ which is bounded, and continuous function in
the second argument \( y^n \in \mathcal{Y}_{0,n} \). By Lebesgue’s monotone convergence theorem, we have

\[
\lim_{i \to \infty} \int_{X_{0,n} \times Y_{0,n}} d_{0,n}(x^n, y^n)(\mu_{0,n} \otimes \overrightarrow{Q}_{0,n}^{\alpha_i})(dx^n, dy^n) \\
= \lim_{i \to \infty} \lim_{k \to \infty} \int_{X_{0,n} \times Y_{0,n}} d_{0,n}(x^n, y^n)(\mu_{0,n} \otimes \overrightarrow{Q}_{0,n}^{\alpha_i})(dx^n, dy^n) \\
= \lim_{i \to \infty} \sup_{k \in \mathbb{N}} \int_{X_{0,n} \times Y_{0,n}} d_{0,n}(x^n, y^n)(\mu_{0,n} \otimes \overrightarrow{Q}_{0,n}^{\alpha_i})(dx^n, dy^n). \tag{A.5}
\]

Given \( \epsilon > 0 \), there exists \( k_0 \in \mathbb{N} \) such that

\[
\sup_{k \in \mathbb{N}} \int_{X_{0,n} \times Y_{0,n}} d_{0,n}(x^n, y^n)(\mu_{0,n} \otimes \overrightarrow{Q}_{0,n}^{\alpha_i})(dx^n, dy^n) - \epsilon \\
\leq \int_{X_{0,n} \times Y_{0,n}} d_{0,n}(x^n, y^n)(\mu_{0,n} \otimes \overrightarrow{Q}_{0,n}^{\alpha_i})(dx^n, dy^n). \tag{A.6}
\]

Substituting (A.6) in (A.5) and using the fact that \( d_{0,n}^{(k_0)} \) is bounded, and continuous in \( y^n \in \mathcal{Y}_{0,n} \), we obtain the following

\[
\lim_{i \to \infty} \int_{X_{0,n} \times Y_{0,n}} d_{0,n}(x^n, y^n)(\mu_{0,n} \otimes \overrightarrow{Q}_{0,n}^{\alpha_i})(dx^n, dy^n) \\
\leq \lim_{i \to \infty} \int_{X_{0,n} \times Y_{0,n}} d_{0,n}^{(k_0)}(x^n, y^n)(\mu_{0,n} \otimes \overrightarrow{Q}_{0,n}^{\alpha_i})(dx^n, dy^n) + \epsilon \\
= \int_{X_{0,n} \times Y_{0,n}} d_{0,n}^{(k_0)}(x^n, y^n)(\mu_{0,n} \otimes \overrightarrow{Q}_{0,n}^{\alpha_i})(dx^n, dy^n) + \epsilon \\
\leq \int_{X_{0,n} \times Y_{0,n}} d_{0,n}(x^n, y^n)(\mu_{0,n} \otimes \overrightarrow{Q}_{0,n}^{\alpha_i})(dx^n, dy^n) + \epsilon. \tag{A.7}
\]

On the other hand, for every \( k \in \mathbb{N} \), we have

\[
\lim_{i \to \infty} \int_{X_{0,n} \times Y_{0,n}} d_{0,n}(x^n, y^n)(\mu_{0,n} \otimes \overrightarrow{Q}_{0,n}^{\alpha_i})(dx^n, dy^n) \\
\geq \lim_{i \to \infty} \int_{X_{0,n} \times Y_{0,n}} d_{0,n}^{(k)}(x^n, y^n)(\mu_{0,n} \otimes \overrightarrow{Q}_{0,n}^{\alpha_i})(dx^n, dy^n) \\
= \int_{X_{0,n} \times Y_{0,n}} d_{0,n}^{(k)}(x^n, y^n)(\mu_{0,n} \otimes \overrightarrow{Q}_{0,n}^{\alpha_i})(dx^n, dy^n).
\]

Letting \( k \to \infty \) we get

\[
\lim_{i \to \infty} \int_{X_{0,n} \times Y_{0,n}} d_{0,n}(x^n, y^n)(\mu_{0,n} \otimes \overrightarrow{Q}_{0,n}^{\alpha_i})(dx^n, dy^n) \\
\geq \int_{X_{0,n} \times Y_{0,n}} d_{0,n}(x^n, y^n)(\mu_{0,n} \otimes \overrightarrow{Q}_{0,n}^{\alpha_i})(dx^n, dy^n). \tag{A.8}
\]

May 8, 2014 DRAFT
By combining (A.7) and (A.8) and letting $\epsilon \to 0$ we get the following equality.

$$
\lim_{i \to \infty} \int_{X_0, n \times Y_0, n} d_{0, n}(x^n, y^n)(\mu_{0, n} \otimes \widetilde{Q}_{0, n}^\alpha)(dx^n, dy^n) = \int_{X_0, n \times Y_0, n} d_{0, n}(x^n, y^n)(\mu_{0, n} \otimes \widetilde{Q}_{0, n}^0)(dx^n, dy^n).
$$

Next, we show that $\widetilde{Q}_{0, n}^0 (\cdot|x^n) \in \mathcal{Q}^{C_1}(D)$. By non-negativity of $d_{0, n}$, (A.8) and Fatou’s lemma we have the following result

$$
\int_{X_0, n \times Y_0, n} d_{0, n}(x^n, y^n)(P_{0, n} \otimes \widetilde{Q}_{0, n}^0)(dx^n, dy^n) = \int_{X_0, n \times Y_0, n} \lim_{i \to \infty} d_{0, n}(x^n, y^n)(\mu_{0, n} \otimes \widetilde{Q}_{0, n}^\alpha_i)(dx^n, dy^n)
$$

$$
= \int_{X_0, n \times Y_0, n} \liminf_{i \to \infty} d_{0, n}(x^n, y^n)(\mu_{0, n} \otimes \widetilde{Q}_{0, n}^\alpha_i)(dx^n, dy^n)
$$

$$
\leq \liminf_{i \to \infty} \int_{X_0, n \times Y_0, n} d_{0, n}(x^n, y^n)(\mu_{0, n} \otimes \widetilde{Q}_{0, n}^\alpha_i)(dx^n, dy^n) \leq D
$$

where the last inequality follows from the fact that $\widetilde{Q}_{0, n}^\alpha_i (\cdot|x^n) \in \mathcal{Q}^{C_1}(D)$. Hence, $\mathcal{Q}^{C_1}_{0, n}(D)$ is weakly closed. Since a weakly closed subset of a weakly compact set is weakly compact, then $\mathcal{Q}^{C_1}_{0, n}(D)$ is compact. This completes the derivation.

\[\square\]

**Appendix B**

**Proof of Lemma III.3**

We need to show that for any sequence $\{ \widetilde{Q}_{0, n}^\alpha (\cdot|x^n) : \alpha = 1, 2, \ldots \} \subset \mathcal{Q}^{C_1}(\mathcal{Y}_{0, n}; \mathcal{X}_{0, n})$ such that $\widetilde{Q}_{0, n}^\alpha (\cdot|x^n) \overset{w}{\to} \widetilde{Q}_{0, n}^0 (\cdot|x^n)$ then

$$
\mathbb{I}_{X^n \to Y^n}(\mu_{0, n}, \widetilde{Q}_{0, n}^0) \leq \liminf_{\alpha \to \infty} \mathbb{I}_{X^n \to Y^n}(\mu_{0, n}, \widetilde{Q}_{0, n}^\alpha).
$$

By Assumption III.1 (A1), (A2), we know from Theorem III.2 that $\widetilde{Q}_{0, n}^\alpha (\cdot|x^n) \overset{w}{\to} \widetilde{Q}_{0, n}^0 (\cdot|x^n) \in \mathcal{Q}^{C_1}(\mathcal{Y}_{0, n}; \mathcal{X}_{0, n})$. Utilizing this, define the sequence $P_{0, n}^\alpha (dx^n, dy^n) \equiv (\mu_{0, n} \otimes \widetilde{Q}_{0, n}^\alpha)(dx^n, dy^n) \in \mathcal{M}_1(\mathcal{X}_{0, n} \times \mathcal{Y}_{0, n})$. We show weak convergence $\overset{w}{\to} (\mu_{0, n} \otimes \widetilde{Q}_{0, n}^0)(dx^n, dy^n) \equiv P_{0, n}^0 (dx^n, dy^n) \in \mathcal{M}_1(\mathcal{X}_{0, n} \times \mathcal{Y}_{0, n})$ by considering integrals with respect to a test function $\phi_{0, n}(\cdot, \cdot) \in BC(\mathcal{X}_{0, n} \times \mathcal{Y}_{0, n})$ via

$$
\int_{\mathcal{X}_{0, n} \times \mathcal{Y}_{0, n}} \phi_{0, n}(x^n, y^n) P_{0, n}^\alpha (dx^n, dy^n) = \int_{\mathcal{X}_{0, n} \times \mathcal{Y}_{0, n}} \phi_{0, n}(x^n, y^n)(\mu_{0, n} \otimes \widetilde{Q}_{0, n}^\alpha)(dx^n, dy^n).
$$

May 8, 2014
Without loss of generality, we take \( \phi_{0,n}(x^n, y^n) = g_{0,n}(x^n)h_{0,n}(y^n), g_{0,n}(x^n) \in \mathcal{B}(\mathcal{X}_0,n), h_{0,n}(y^n) \in \mathcal{B}(\mathcal{Y}_0,n) \). Utilizing Lebesgue’s Dominated Convergence Theorem we have

\[
\lim_{\alpha \to \infty} \int_{\mathcal{Y}_0,n \times \mathcal{X}_0,n} h_{0,n}(y^n)g_{0,n}(x^n)\tilde{Q}^{\alpha}_{0,n}(dy^n|x^n) \otimes \mu_{0,n}(dx^n)
= \lim_{\alpha \to \infty} \int_{\mathcal{X}_0,n} \left( \int_{\mathcal{Y}_0,n} h_{0,n}(y^n)\tilde{Q}^{\alpha}_{0,n}(dy^n|x^n) \right) g_{0,n}(x^n)\mu_{0,n}(dx^n)
= \int_{\mathcal{Y}_0,n \times \mathcal{X}_0,n} h_{0,n}(y^n)g_{0,n}(x^n)\tilde{Q}^{0}_{0,n}(dy^n|x^n) \otimes \mu_{0,n}(dx^n).
\]

Hence, \( P^{\alpha}_{0,n}(dx^n, dy^n) \rightarrow^{w} P^{0}_{0,n}(dx^n, dy^n) \in \mathcal{M}_1(\mathcal{X}_0,n \times \mathcal{Y}_0,n) \). Since \( \mathcal{X}_0,n \times \mathcal{Y}_0,n \) are separable and the joint measure \( P^{\alpha}_{0,n}(dx^n, dy^n) \rightarrow^{w} P^{0}_{0,n}(dx^n, dy^n) \), then by the continuous mapping theorem \([33]\) it follows that the marginal measures (its projection to \( \mathcal{Y}_0,n \)) also converge weakly, \( \nu^{\alpha}_{0,n} \rightarrow^{w} \nu^{0}_{0,n} \in \mathcal{M}_1(\mathcal{Y}_0,n) \). Next, we define the product measure \( \overline{\Pi}^{\alpha}_{0,n} \triangleq \mu_{0,n} \times \nu^{\alpha}_{0,n} \in \mathcal{M}_1(\mathcal{X}_0,n \times \mathcal{Y}_0,n) \), where \( \{\nu^{\alpha}_{0,n} : \alpha = 1, 2, \ldots\} \) are the marginals of \( \{\tilde{Q}^{\alpha}_{0,n} \otimes \mu_{0,n} : \alpha = 1, 2, \ldots\} \). Then, by the continuous mapping theorem \([33]\) we also have

\[
\overline{\Pi}^{\alpha}_{0,n} = \mu_{0,n} \times \nu^{\alpha}_{0,n} \rightarrow^{w} \overline{\Pi}^{0}_{0,n} = \mu_{0,n} \times \nu^{0}_{0,n}.
\]

Recall the definition of directed information via relative entropy (see \((\text{II.7})-(\text{II.9})\)) given by

\[
\mathbb{D}(P_{0,n}||\overline{\Pi}^{\alpha}_{0,n}) = \mathbb{D}(\mu_{0,n} \otimes \tilde{Q}^{\alpha}_{0,n}||\mu_{0,n} \times \nu^{\alpha}_{0,n}) = \mathbb{I}_{X^n \rightarrow Y^n}(\mu_{0,n}, \tilde{Q}^{\alpha}_{0,n}). \quad (\text{B.1})
\]

Relative entropy is lower semicontinuous \([32]\) Lemma 1.4.3, hence

\[
\mathbb{D}(P^{0}_{0,n}||\overline{\Pi}^{0}_{0,n}) = \mathbb{D}(\mu_{0,n} \otimes \tilde{Q}^{0}_{0,n}||\overline{\Pi}^{0}_{0,n}) \leq \lim_{\alpha \to \infty} \inf \mathbb{D}(P^{\alpha}_{0,n}||\overline{\Pi}^{\alpha}_{0,n}). \quad (\text{B.2})
\]

By \((\text{B.1})\) it follows that \((\text{B.2})\) is also equivalent to

\[
\mathbb{I}_{X^n \rightarrow Y^n}(\mu_{0,n}, \tilde{Q}^{0}_{0,n}) \leq \lim_{\alpha \to \infty} \inf \mathbb{I}_{X^n \rightarrow Y^n}(\mu_{0,n}, \tilde{Q}^{\alpha}_{0,n}).
\]

Hence, directed information is lower semicontinuous as a functional of \( \tilde{Q}^{\alpha}_{0,n}(\cdot|x^n) \in \mathcal{Q}^{\mathcal{C}1}(\mathcal{Y}_0,n; \mathcal{X}_0,n) \) for a fixed \( \mu_{0,n}(\cdot) \in \mathcal{M}_1(\mathcal{X}_0,n) \). This completes the derivation. \(\square\)
APPENDIX C

PROOF OF LEMMA [III.7]

The equivalence of MC1, MC2 and MC3 is straightforward hence it is omitted. To this end, we show equivalence of MC4 to any of MC1, MC2 and MC3. We proceed with the derivation, by often assuming existence of densities, which are denoted by lower case letters \( \tilde{p}(\cdot | \cdot) \), to avoid lengthy measure theoretic arguments.

**MC4 \implies MC3**: Since for \( i = 0, \ldots, n - 1 \), by MC4 we have

\[
P_{X_{i+1}^{n}|X_1^i, Y_1^i}(dx_{i+1}^{n}|x^i, y^i) = P_{X_{i+1}^{n}|X_1^i}(dx_{i+1}^{n}|x^i)
\]

then by integrating over \( X_{i+2}^{n} \) both sides of the previous identity we obtain MC3.

**MC4 \iff MC3**: Since MC3 \iff MC2, we show that if \( X_{i+1}^n \leftrightarrow (X^i, Y^{i-1}) \leftrightarrow Y_i \) forms a MC for \( i = 0, 1, \ldots, n - 1 \), then \( X_{i+1}^n \leftrightarrow X^i \leftrightarrow Y^i \) forms a MC for \( i = 0, 1, \ldots, n - 1 \). We show this by induction. First, we show that \( (X_{i+1}, X_{i+2}) \leftrightarrow X^i \leftrightarrow Y^i \) forms a MC, or equivalently, \( \tilde{p}(x_{i+1}, x_{i+2}|x^i, y^i) = \tilde{p}(x_{i+1}, x_{i+2}|x^i) \). Since

\[
\tilde{p}(x_{i+1}, x_{i+2}|x^i, y^i) = \frac{\tilde{p}(x_i, x_{i+1}, x_{i+2}, y^i)}{\tilde{p}(x^i, y^i)} = \frac{\tilde{p}(y_i|y^{i-1}, x_{i+2})\tilde{p}(y^{i-1}, x_{i+2})}{\tilde{p}(x^i, y^i)}
\]

\[
\tilde{p}(y_i|y^{i-1}, x^i) \tilde{p}(x_{i+2}|x^{i+1}, y^{i-1}) \tilde{p}(x^{i+1}, y^{i-1})
\]

\[
= \frac{\tilde{p}(y_i|y^{i-1}, x^i) \tilde{p}(x_{i+2}|x^{i+1}, y^{i-1}) \tilde{p}(x^{i+1}, y^{i-1})}{\tilde{p}(y_i|y^{i-1}, x^i) \tilde{p}(x^i, y^{i-1})}
\]

\[
= \frac{\tilde{p}(x_{i+2}|x^{i+1}) \tilde{p}(x_{i+1}|x^i)}{\tilde{p}(x_{i+2}, x_{i+1}|x^i)}
\]

where \((a)\) is implied from MC2, while \((b)\), \((c)\) follows from MC3 \iff MC2. Hence, MC4 holds for \( n = i + 2 \).

Suppose \( X_{i+1}^k \leftrightarrow X^i \leftrightarrow Y^i \) forms a MC, for some \( i + 2 \leq k < n - 1 \). We show that it holds for
Moreover,

\[
\forall P \text{ is expressed as information nonanticipative RDF. Then, by Theorem IV.2 the optimal reproduction distribution (where } (d), (e) \text{ follow from MC3 } \iff \text{MC2. This completes the derivation.}\]

**APPENDIX D**

**PROOF OF THEOREM IV.7**

Let \( s \) be the the Lagrange multiplier which is part of the optimal solution which solves the information nonanticipative RDF. Then, by Theorem IV.2 the optimal reproduction distribution is expressed as

\[
q_i^*(F_i; y^{i-1}, x^i) = \int_{F_i} e^{s p_i(x^i, y^i)} \lambda_i(x^i, y^{i-1}) \nu_i^*(dy_i; y^{i-1}), \quad \forall F_i \in \mathcal{B}(\mathcal{Y}_i), \quad \forall i \in \mathbb{N}^n. \tag{D.1}
\]

By integrating (D.1) with respect to \( P_{0,i}^*(dx^i|y^{i-1}) \) we obtain the expression

\[
\int_{\mathcal{X}_{0,i}} q_i^*(F_i; y^{i-1}, x^i) \otimes P_{0,i}^*(dx^i|y^{i-1}) = P_{0,i}^*(\mathcal{X}_{0,i} \times F_i|y^{i-1})
\]

\[
= \int_{F_i \times \mathcal{X}_{0,i}} e^{s p_i(x^i, y^i)} \lambda_i(x^i, y^{i-1}) \nu_i^*(dy_i; y^{i-1}) \otimes P_{0,i}^*(dx^i|y^{i-1}), \quad \forall F_i \in \mathcal{B}(\mathcal{Y}_i), \quad i \in \mathbb{N}^n.
\]

Moreover, \( \forall F_i \in \mathcal{B}(\mathcal{Y}_i), \quad i \in \mathbb{N}^n, \)

\[
\nu_i^*(F_i; y^{i-1}) = \int_{F_i} \nu_i^*(dy_i; y^{i-1}) = P_{0,i}^*(\mathcal{X}_{0,i} \times F_i|y^{i-1})
\]

\[
= \int_{F_i \times \mathcal{X}_{0,i}} e^{s p_i(x^i, y^i)} \lambda_i(x^i, y^{i-1}) P_{0,i}^*(dx^i|y^{i-1}) \otimes \nu_i^*(dy_i; y^{i-1}). \tag{D.2}
\]

Utilizing (D.2) we finally obtain

\[
\int_{\mathcal{X}_{0,i}} e^{s p_i(x^i, y^i)} \lambda_i(x^i, y^{i-1}) P_{0,i}^*(dx^i|y^{i-1}) = 1, \quad \nu_i^* \text{ - a.s., } \forall i \in \mathbb{N}.
\]

This completes the derivation. \( \square \)
APPENDIX E

PROOF OF THEOREM IV.8

Let $s \leq 0$, $\lambda \in \Psi_s$ and $\overrightarrow{Q}_{0,n}(\cdot|x^n) \in \mathcal{Q}^{C_1}(D)$ be given. Then, using the fact that

$$
\frac{1}{n+1} \sum_{i=0}^{n} \int_{X_{x,i} \times Y_{0,i}} \rho_i(x^i, y^i)(\mu_{0,i} \otimes \overrightarrow{Q}_{0,i})(dx^i, dy^i) \leq D
$$

gives

$$
\mathbb{I}_{X^n \rightarrow Y^n}(\mu_{0,n}, \overrightarrow{Q}_{0,n}) - sD(n+1) - \sum_{i=0}^{n} \int_{X_{x,i} \times Y_{0,i}} \log \left( \frac{q_i(dy_1, y_i^1, 1, x^i)}{\nu_i(dy_1, y_i^1, 1, x^i)} \right)(\mu_{0,i} \otimes \overrightarrow{Q}_{0,i})(dx^i, dy^i)
$$

$$
- \int_{X_{x,0} \times Y_{0,0}} \log (\lambda_i(x^i, y_i^1))(\mu_{0,i} \otimes \overrightarrow{Q}_{0,i})(dx^i, dy^i)
$$

$$
= \sum_{i=0}^{n} \int_{X_{x,i} \times Y_{0,i}} \log \left( \frac{q_i(dy_1, y_i^1, 1, x^i)e^{-sp_i(x^i, y_i^1)}}{\nu_i(dy_1, y_i^1, 1, x^i)\lambda_i(x^i, y_i^1)} \right)(\mu_{0,i} \otimes \overrightarrow{Q}_{0,i})(dx^i, dy^i)
$$

$$
= \int_{X_{x,0} \times Y_{0,0}} \left\{ \int_{X_{x,i} \times Y_i} \log \left( \frac{q_i(dy_1, y_i^1, 1, x^i)e^{-sp_i(x^i, y_i^1)}}{\nu_i(dy_1, y_i^1, 1, x^i)\lambda_i(x^i, y_i^1)} \right) \right\} \otimes P_{0,i-1}(dx^{i-1}, dy^{i-1})
$$

\[\mathbb{I}_{X^n \rightarrow Y^n}(\mu_{0,n}, \overrightarrow{Q}_{0,n}) - sD(n+1) - \sum_{i=0}^{n} \int_{X_{x,i} \times Y_{0,i}} \log \left( \frac{q_i(dx_1, x_i)}{\nu_i(dx_1, x_i)} \right)(\mu_{0,i} \otimes \overrightarrow{Q}_{0,i})(dx^i, dy^i)
$$

$$
- \int_{X_{x,0} \times Y_{0,0}} \log (\lambda_i(x^i, y_i^1))(\mu_{0,i} \otimes \overrightarrow{Q}_{0,i})(dx^i, dy^i)
$$

$$
= \sum_{i=0}^{n} \int_{X_{x,i} \times Y_{0,i}} \log \left( \frac{q_i(dx_1, x_i)e^{-sp_i(x^i, y_i^1)}}{\nu_i(dx_1, x_i)\lambda_i(x^i, y_i^1)} \right)(\mu_{0,i} \otimes \overrightarrow{Q}_{0,i})(dx^i, dy^i)
$$

$$
= \int_{X_{x,0} \times Y_{0,0}} \left\{ \int_{X_{x,i} \times Y_i} \left( 1 - \frac{e^{sp_i(x^i, y_i^1)\nu_i(dx_1, x_i)\lambda_i(x^i, y_i^1)}}{q_i(dx_1, x_i)} \right) \right\} \otimes P_{0,i-1}(dx^{i-1}, dy^{i-1})
$$

$$
= \sum_{i=0}^{n} \left\{ 1 - \int_{X_{x,i} \times Y_{0,i-1}} \int_{X_{x,i} \times Y_i} e^{sp_i(x^i, y_i^1)\lambda_i(x^i, y_i^1)\nu_i(dx_1, x_i)} \right\} \otimes P_{0,i-1}(dx^{i-1}, dy^{i-1})
$$
\[
= \sum_{i=0}^{n} \left\{ 1 - \int_{\mathcal{Y}_i} \nu_i(dy_i; y_i^{i-1}) \int_{\mathcal{Y}_{0,i-1}} \nu_{0,i-1}(dy_i^{i-1}) \left( \int_{\mathcal{X}_{0,i}} e^{s\rho_i(x_i, y_i^{i-1})} \lambda_i(x_i, y_i^{i-1}) \otimes P_{0,i}(dx_i|y_i^{i-1}) \right) \right\}
\]
\[
\geq \sum_{i=0}^{n} \left( 1 - \int_{\mathcal{Y}_{0,i}} \nu_0(dy_i) \right) = 0 \quad (b)
\]
where (a) follows from the inequality \( \log x \geq 1 - \frac{1}{x}, \ x > 0 \), and (b) follows from (IV.11).

Hence, we obtain
\[
P_{0,n}^{na}(D) \geq \sup_{s \leq 0} \sup_{\lambda \in \Psi_s} \left\{ sD(n + 1) + \sum_{i=0}^{n} \int_{\mathcal{X}_{0,i} \times \mathcal{Y}_{0,i-1}} \log \left( \lambda_i(x_i, y_i^{i-1}) \right) P_{0,i-1}(dx_i^{i-1}, dy_i^{i-1}) \otimes p_i(dx_i; x_i^{i-1}) \right\}
\]
\[
= \sum_{i=0}^{n} \left\{ 1 - \int_{\mathcal{Y}_i} \nu_i(dy_i; y_i^{i-1}) \int_{\mathcal{Y}_{0,i-1}} \nu_{0,i-1}(dy_i^{i-1}) \left( \int_{\mathcal{X}_{0,i}} e^{s\rho_i(x_i, y_i^{i-1})} \lambda_i(x_i, y_i^{i-1}) \otimes P_{0,i}(dx_i|y_i^{i-1}) \right) \right\}
\]  
\[
\geq \sum_{i=0}^{n} \left( 1 - \int_{\mathcal{Y}_{0,i}} \nu_0(dy_i) \right) = 0 \quad (b)
\]
However, equality in (c) holds if
\[
\lambda_i(x_i, y_i^{i-1}) \triangleq \left( \int_{\mathcal{Y}_i} e^{s\rho_i(x_i, y_i^{i-1})} \nu_i^*(dy_i; y_i^{i-1}) \right)^{-1}, \ i = 0, 1, \ldots, n.
\]
This completes the derivation.

\[\Box\]

APPENDIX F

Proof of Theorem [IV.10]

The derivation is based on showing that \( R_{0,n}^{na}(D) \) is bounded above and below by the RHS of (IV.24). The lower bound is obtained by using Theorem IV.8 to derive a lower bound analogous to Shannon’s lower bound.

Define
\[
H_\infty = \lim_{t \to \infty} H_t, \ H_t \triangleq \text{diag}\{\eta_{t,1}, \ldots, \eta_{t,p}\}, \ \eta_{t,i} = 1 - \frac{\delta_{t,i}}{H_{t,i}}, \ i = 1, \ldots, p, \ t \in \mathbb{N}.
\]

Consider the additive noisy channel with feedback of the form
\[
\tilde{K}_t = E_t^{tr} H_t E_t \left( X_t - \mathbb{E}\{X_t|\sigma\{Y_t^{t-1}\}\} \right) + E_t^{tr} B_t V_t^c = E_t^{tr} H_t E_t K_t + E_t^{tr} B_t V_t^c, \ t \in \mathbb{N}^n \quad (F.1)
\]
where \( \{V_t^c : t \in \mathbb{N}\} \) is an independent Gaussian zero mean process with covariance \( \text{cov}(V_t^c) = Q = \text{diag}\{q_1, \ldots, q_p\} \), and \( \{B_t : t \in \mathbb{N}\} \) is to be determined.

Next, we show that by letting \( B_\infty = \lim_{t \to \infty} B_t \), where \( B_t = \sqrt{H_t \Delta_t Q^{-1}} \), and \( \Delta_t \triangleq \text{diag}\{\delta_{t,1}, \ldots, \delta_{t,p}\} \), then \( \Lambda_\infty = \lim_{t \to \infty} \Lambda_t = \lim_{t \to \infty} \mathbb{E}\{K_t K_t^{tr}\} \), and also \( \lim_{n \to \infty} \frac{1}{n+1} \mathbb{E}\{\sum_{t=0}^{n} \|X_t - Y_t\|^2_{\mathbb{R}^p}\} = \)
\[
\lim_{n \to \infty} \frac{1}{n+1} \mathbb{E}\left\{ \sum_{t=0}^{n} \|K_t - \tilde{K}_t\|_{\mathbb{F}_p}^2 \right\} = D. \text{ Clearly, by (IV.20), (IV.22), (F.1)}
\]

\[
\lim \mathbb{E}\left\{ (X_t - Y_t)^{tr}(X_t - Y_t) \right\} = \lim \text{Trace } \mathbb{E}\left\{ (K_t - \tilde{K}_t)(K_t - \tilde{K}_t)^{tr} \right\}
\]

\[
= \lim \text{Trace } \mathbb{E}\left\{ (K_t - E_t^{tr}H_tE_tK_t - E_t^{tr}B_tV_t^c)(K_t - E_t^{tr}H_tE_tK_t - E_t^{tr}B_tV_t^c)^{tr} \right\}
\]

\[
= \lim \text{Trace } \mathbb{E}\left\{ (I - E_t^{tr}H_tE_t)K_t - E_t^{tr}B_tV_t^c)((I - E_t^{tr}H_tE_t)K_t - E_t^{tr}B_tV_t^c)^{tr} \right\}
\]

\[
= \lim \text{Trace}\left\{ (I - E_t^{tr}H_tE_t)\Lambda_t(I - E_t^{tr}H_tE_t)^{tr} + E_t^{tr}B_tQ_tB_t^{tr}E_t \right\}
\]

\[
= \lim \text{Trace}\left\{ E_t^{tr}(I - H_t)\Lambda_{11,0} + \Lambda_{11,n} + \Lambda_{n,1} + \Lambda_{n,n} \right\} = D
\]

where \(a\) holds by setting \(B_\infty = B_t\) as in (IV.31).

Also, by (IV.23),

\[
\lim_{n \to \infty} \frac{1}{n+1} \mathbb{I}_{\mathbb{F}_{0,n}^{K_n},\tilde{K}_n}(D) \leq \lim_{n \to \infty} \frac{1}{n+1} \mathbb{I}_{X_n 
arrow Y^n}(P_{K_n}, \tilde{\mathbb{P}}_{\tilde{K}_n|K_n})
\]

\[
= \lim_{n \to \infty} \frac{1}{n+1} \sum_{t=0}^{n} I(K_t; \tilde{K}_t|\tilde{K}^{t-1})
\]

\[
= \lim_{n \to \infty} \frac{1}{n+1} \sum_{t=0}^{n} \left( H(\tilde{K}_t|\tilde{K}^{t-1}) - H(\tilde{K}_t|\tilde{K}^{t-1}, K_t) \right)
\]

\[
\overset{(b)}{\leq} \lim_{n \to \infty} \frac{1}{n+1} \sum_{t=0}^{n} \left( H(\tilde{K}_t) - H(\tilde{K}_t|\tilde{K}^{t-1}, K_t) \right)
\]

\[
\overset{(c)}{\leq} \lim_{n \to \infty} \frac{1}{n+1} \sum_{t=0}^{n} \left( H(\tilde{K}_t) - H(\tilde{K}_t|K_t) \right)
\]

\[
\overset{(d)}{=} \lim_{n \to \infty} \frac{1}{n+1} \sum_{t=0}^{n} \left( H(\tilde{K}_t) - H(E_t^{tr}B_tV_t^c) \right)
\]

(F.2)

where \(b\) follows from the fact that conditioning reduces entropy, \(c\) follows from the fact that \(\tilde{K}_t = E_t^{tr}H_tE_tK_t + E_t^{tr}B_tV_t^c\) is a memoryless Gaussian channel, and \(d\) follows from the orthogonality of \(K_t\) and \(V_t^c\), \(\forall t \in \mathbb{N}\). Next, we compute the entropy rates appearing in (F.2) from the covariances of the corresponding processes. The covariance of the Gaussian zero mean, noise
process \( \{E_{t}^{tr} B_{t} V_{t}^{c}, t \in \mathbb{N}\} \) is obtained as follows.

\[
\lim_{t \to \infty} E \left\{ (E_{t}^{tr} B_{t} V_{t}^{c}) (E_{t}^{tr} B_{t} V_{t}^{c})^{tr} \right\} = \lim_{t \to \infty} E \left\{ E_{t}^{tr} B_{t} V_{t}^{c} (E_{t}^{tr} B_{t} V_{t}^{c})^{tr} \right\} = \lim_{t \to \infty} E_{t}^{tr} B_{t} Q B_{t}^{tr} E_{t} = \lim_{t \to \infty} \left\{ E_{t}^{tr} \sqrt{H_{t} \Delta_{t} Q^{-1} Q \sqrt{H_{t} \Delta_{t} Q^{-1}}} E_{t} \right\} = \lim_{t \to \infty} E_{t}^{tr} H_{t} \Delta_{t} E_{t} = \lim_{t \to \infty} E_{t}^{tr} diag \{ \eta_{t,1}, \ldots, \eta_{t,p} \delta_{t,p} \} E_{t} = E_{\infty}^{tr} diag \{ \eta_{\infty,1} \delta_{\infty,1}, \ldots, \eta_{\infty,p} \delta_{\infty,p} \} E_{\infty}. \tag{F.3}
\]

The covariance of the process \( \{\tilde{K}_{t} : t \in \mathbb{N}\} \) is obtained as follows.

\[
\lim_{t \to \infty} E \left\{ \tilde{K}_{t} \tilde{K}_{t}^{tr} \right\} = \lim_{t \to \infty} E \left\{ (E_{t}^{tr} H_{t} E_{t} K_{t} + E_{t}^{tr} B_{t} V_{t}^{c}) (E_{t}^{tr} H_{t} E_{t} K_{t} + E_{t}^{tr} B_{t} V_{t}^{c})^{tr} \right\} = \lim_{t \to \infty} E \left\{ E_{t}^{tr} H_{t} E_{t} K_{t} K_{t}^{tr} E_{t}^{tr} H_{t} E_{t} + E_{t}^{tr} B_{t} V_{t}^{c} V_{t}^{c}^{tr} B_{t}^{tr} E_{t} \right\} = \lim_{t \to \infty} \left\{ E_{t}^{tr} H_{t} E_{t} \Delta_{t} E_{t}^{tr} H_{t} E_{t} + E_{t}^{tr} \sqrt{H_{t} \Delta_{t} Q^{-1} Q \sqrt{H_{t} \Delta_{t} Q^{-1}}} E_{t} \right\} = \lim_{t \to \infty} E_{t}^{tr} diag \{ \lambda_{t,1} - \delta_{t,1}, \ldots, \lambda_{t,p} - \delta_{t,p} \} E_{t}, \lambda_{t,i} - \delta_{t,i} \geq 0, \forall t
\]

\[
= E_{\infty}^{tr} diag \{ \lambda_{\infty,1} - \delta_{\infty,1}, \ldots, \lambda_{\infty,p} - \delta_{\infty,p} \} E_{\infty}. \tag{F.4}
\]

Using (F.4) we obtain the first term of (F.2) as follows

\[
\lim_{n \to \infty} \frac{1}{n + 1} \sum_{t=0}^{n} H(\tilde{K}_{t}) = \lim_{n \to \infty} \frac{1}{2n + 1} \sum_{t=0}^{n} \log \left\{ (2\pi e) diag \{ \lambda_{t,1} - \delta_{t,1}, \ldots, \lambda_{t,p} - \delta_{t,p} \} \right\} = \lim_{n \to \infty} \frac{1}{2n + 1} \sum_{t=0}^{n} \log \left\{ (2\pi e) \times_{i=1}^{p} (\lambda_{t,i} - \delta_{t,i})^{+} \right\} = \lim_{n \to \infty} \frac{1}{2n + 1} \sum_{t=0}^{n} \sum_{i=1}^{p} \log \left\{ (2\pi e) (\lambda_{t,i} - \delta_{t,i})^{+} \right\} = \frac{1}{2} \sum_{i=1}^{p} \log \left\{ (2\pi e) (\lambda_{\infty,i} - \delta_{\infty,i})^{+} \right\}. \tag{F.5}
\]
Also, by (F.3), we obtain the second term in (F.2) as follows.

\[
\lim_{n \to \infty} \frac{1}{n + 1} \sum_{t=0}^{n} H(E_t^c B, V_t^c) = \lim_{n \to \infty} \frac{1}{2n + 1} \sum_{t=0}^{n} \log \left\{ (2\pi e) \text{diag}\{\eta_{t,1}, \ldots, \eta_{t,p}\} \right\}
\]

\[
= \frac{1}{2} \sum_{t=0}^{n} \log \left\{ (2\pi e) \times p_{i=1}^{p} (\eta_{\infty,i} \delta_{\infty,i}) \right\} = \frac{1}{2} \sum_{i=1}^{p} \log \left\{ (2\pi e) (\eta_{\infty,i} \delta_{\infty,i}) \right\}.
\]  \hspace{1cm} (F.6)

By using (F.5) and (F.6) in (F.2) we have the following upper bound

\[
\lim_{n \to \infty} R_{0,n}^{\alpha,K^n}(D) \leq \frac{1}{2} \sum_{i=1}^{p} \log \left\{ (2\pi e) (\lambda_{\infty,i} - \delta_{\infty,i})^+ \right\} - \frac{1}{2} \sum_{i=1}^{p} \log \left\{ (2\pi e) (\eta_{\infty,i} \delta_{\infty,i}) \right\} - \frac{1}{2} \sum_{i=1}^{p} \log \frac{\lambda_{\infty,i}}{\delta_{\infty,i}}
\]

where \(\delta_{\infty,i} = \min\{\xi_{\infty,i}, \lambda_{\infty,i}\} \) and \(\sum_{i=1}^{p} \delta_{\infty,i} = D\).

**Lower Bound (Analogous to Shannon’s Lower Bound).** Next, we apply Theorem IV.8 to obtain a lower bound for the nonanticipative RDF \(R_{0,n}^{\alpha,K^n}(D) = \lim_{n \to \infty} \frac{1}{n+1} R_{0,n}^{\alpha,K^n}(D) = \lim_{n \to \infty} \frac{1}{n+1} R_{0,n}^{\alpha,K^n}(D)\).

Applying Theorem IV.8 to \(R_{0,n}^{\alpha,K^n}(D)\), the set \(\Psi_s\) is defined by

\[
\Psi_s \triangleq \left\{ \lambda_t(k_t, \tilde{k}^{t-1}) : t = 0, 1, \ldots, n \right\} : \lambda_t(k_t, \tilde{k}^{t-1}) \geq 0,
\]

\[
\int e^{s||K_t - \tilde{K}_t||_F^2} \lambda(k_t, \tilde{k}^{t-1}) \bar{p}(k_t|\tilde{k}^{t-1})dk_t \leq 1, \; t = 0, 1, \ldots, n \}
\]  \hspace{1cm} (F.7)

where \(\bar{p}(k_t|\tilde{k}^{t-1})\) denotes the conditional density of \(k_t\) given \(\tilde{k}^{t-1}\). Choose \(s \leq 0\) and take \(\lambda_t(k_t, \tilde{k}^{t-1}) \in \Psi_s\) such that

\[
\lambda_t(k_t, \tilde{k}^{t-1}) = \frac{\alpha_t}{\bar{p}(k_t|\tilde{k}^{t-1})}
\]  \hspace{1cm} (F.8)

for some \(\alpha_t\) not depending on \(k_t\). Substituting (F.8) into the reduced integral inequality in (F.7) we obtain

\[
\alpha_t \int e^{s||K_t - \tilde{K}_t||_F^2} dk_t \leq 1, \; t = 0, 1, \ldots, n.
\]

By changing the variable of integration we also obtain

\[
\alpha_t \int e^{s||z_t||_F^2} dz_t = \alpha_t \sqrt{\frac{(\pi s)^2}{4}} = \alpha_t \left(\frac{\pi}{s}\right)^p \leq 1, \; t = 0, 1, \ldots, n.
\]  \hspace{1cm} (F.9)
By setting $\alpha_t = \left(-\frac{s}{\pi}\right)^2$, $t = 0, 1, \ldots, n$, the inequality of (F.9) holds with equality. Then, by Theorem [IV.8] we have
\[
\lim_{n \to \infty} \frac{1}{n+1} R^{n}_{0,n} (D) \geq sD + \lim_{n \to \infty} \frac{1}{n+1} \sum_{t=0}^{n} \int_{K_t \times K_{0,t-1}} \bar{p}(k_t, \bar{k}^{t-1}) \log \left( \frac{\lambda_t(k_t, \bar{k}^{t-1})}{\bar{p}(k_t, \bar{k}^{t-1})} \right) dk_t d\bar{k}^{t-1}
\]
\[
\overset{(e)}{=} sD + \lim_{n \to \infty} \frac{1}{n+1} \sum_{t=0}^{n} \int_{K_t \times K_{0,t-1}} \bar{p}(k_t, \bar{k}^{t-1}) \log \left( \frac{\left(-\frac{s}{\pi}\right)^2}{\bar{p}(k_t, \bar{k}^{t-1})} \right) dk_t d\bar{k}^{t-1}
\]
\[
= sD + \lim_{n \to \infty} \frac{1}{n+1} \sum_{t=0}^{n} \log \left(-\frac{s}{\pi}\right)^2 + \lim_{n \to \infty} \frac{1}{n+1} \sum_{t=0}^{n} H(K_t|\bar{K}^{t-1})
\]
\[
\overset{(f)}{=} sD + \lim_{n \to \infty} \frac{1}{n+1} \sum_{t=0}^{n} \log \left(-\frac{s}{\pi}\right)^2 + \lim_{n \to \infty} \frac{1}{n+1} \sum_{t=0}^{n} H(K_t)
\]
\[(F.10)\]

where (e) follows from (F.8), and (f) follows from the orthogonality of $K_t$ and $\bar{K}^{t-1}$. Next, we need to find the Lagrangian multiplier “$s$” so that the lower bound (F.10) equals $\frac{1}{2} \sum_{i=1}^{p} \log \frac{\lambda_{\infty,i}}{\delta_{\infty,i}}$.

To this end, we need to ensure existence of some $s < 0$ such that the following identity holds.

\[
sD + \lim_{n \to \infty} \frac{1}{2n+1} \sum_{t=0}^{n} \sum_{i=1}^{p} \log \left(-\frac{s}{\pi}\right)^2 + \lim_{n \to \infty} \frac{1}{2n+1} \sum_{t=0}^{n} \log 2\pi e|\Lambda_t| = \lim_{n \to \infty} \frac{1}{2n+1} \sum_{t=0}^{n} \sum_{i=1}^{p} \log \lambda_{t,i} - \delta_{t,i}
\]
\[
\implies sD = \lim_{n \to \infty} \frac{1}{2n+1} \sum_{t=0}^{n} \sum_{i=1}^{p} \log \left(-\frac{s}{\pi}\right)^2 + \lim_{n \to \infty} \frac{1}{2n+1} \sum_{t=0}^{n} \sum_{i=1}^{p} \log \lambda_{t,i} - \delta_{t,i}
\]
\[
\implies \lim_{n \to \infty} \frac{1}{2n+1} \sum_{t=0}^{n} \sum_{i=1}^{p} \log e^{2\pi e|\Lambda_t|} = \lim_{n \to \infty} \frac{1}{2n+1} \sum_{t=0}^{n} \sum_{i=1}^{p} \log \lambda_{t,i} - \delta_{t,i}
\]
\[
= \frac{1}{2} \sum_{i=1}^{p} \log \frac{\lambda_{\infty,i}}{\delta_{\infty,i}} - \frac{1}{2} \sum_{i=1}^{p} \log \lambda_{t,i} - \delta_{t,i}
\]
\[
\implies \lim_{n \to \infty} \frac{1}{2n+1} \sum_{t=0}^{n} \sum_{i=1}^{p} \log e^{2\pi e|\Lambda_t|} = \lim_{n \to \infty} \frac{1}{2n+1} \sum_{t=0}^{n} \sum_{i=1}^{p} \log \lambda_{t,i} - \delta_{t,i}
\]
\[
\implies \lim_{n \to \infty} \frac{1}{2n+1} \sum_{t=0}^{n} \sum_{i=1}^{p} \log e^{2\pi e|\Lambda_t|} = \frac{1}{2} \sum_{i=1}^{p} \log \frac{1}{2\pi e \delta_{\infty,i}}
\]
\[
\implies \frac{1}{2} \sum_{i=1}^{p} \log e^{2\pi e|\Lambda_t|} = \frac{1}{2} \sum_{i=1}^{p} \log \frac{1}{2\pi e \delta_{\infty,i}} \implies e^{2\pi e|\Lambda_t|} = \left(\frac{1}{2\pi e \delta_{\infty,i}}\right)^{2} \implies s = -\frac{1}{2\delta_{\infty,i}}
\]

where $\delta_{\infty,i} = \{\xi_{\infty}, \lambda_{\infty,i}\}$. Now, if $\delta_{\infty,i} = \xi_{\infty}$ then $s = -\frac{1}{2\delta_{\infty,i}}$ and the nonanticipative RDF is bounded below be the following expression

\[
\lim_{n \to \infty} R^{n}_{0,n} (D) \geq \lim_{n \to \infty} \frac{1}{2n+1} \sum_{t=0}^{n} \sum_{i=1}^{p} \log \frac{\lambda_{t,i}}{\delta_{t,i}} = \frac{1}{2} \sum_{i=1}^{p} \log \frac{\lambda_{\infty,i}}{\delta_{\infty,i}}
\]
\[(F.11)\]
which is the desired lower bound with $\sum_{i=1}^{p} \delta_{\infty,i} = D$. In the case where $\delta_{\infty,i} = \lambda_{\infty,i}$, then no encoding is performed and there is no sense in proving a lower bound to $R_{0,n}^{na,K^n,K^n}(D)$. This completes the proof of (IV.24).

Next, we determine the expression of $\Lambda_{\infty}$. By definition, $\Lambda_{\infty} = \lim_{t \to \infty} \Lambda_t$, where $\Lambda_t = \text{cov}(X_t - \mathbb{E}\{X_t|\sigma\{Y_t^{-1}\}\})$. Since $X_t - \mathbb{E}\{X_t|\sigma\{Y_t^{-1}\}\} = CZ_t + NV_t - C\mathbb{E}\{Z_t|\sigma\{Y_t^{-1}\}\}$ then $\Lambda_t = C\sum_t C^{tr} + NN^{tr}$. Let $\hat{Z}_{t|t-1} = \mathbb{E}\{Z_t|\sigma\{Y_t^{-1}\}\}$. Clearly, $\Sigma_t \triangleq \mathbb{E}\{(Z_t - \mathbb{E}\{Z_t|Y_t^{-1}\})(Z_t - \mathbb{E}\{Z_t|Y_t^{-1}\})^t\}$. Moreover, $\Lambda_{\infty} = C\lim_{t \to \infty} \Sigma_t C^{tr} + NN^{tr}$. Therefore, to determine $\Sigma_{\infty} \triangleq \lim_{t \to \infty} \Sigma_t$, we need the equation of the error $e_t \triangleq Z_t - \hat{Z}_{t|t-1}$, hence the equation of the least-squares filter of $Z_t$ given all the previous outputs $Y_t^{-1}$, namely $\hat{Z}_{t|t-1}$. From Fig. IV.4, we deduce that $Y_t = \hat{K}_t + \hat{X}_{t|t-1}$, where $\{\hat{X}_{t|t-1} : t \in \mathbb{N}\}$ is obtained from the modified Kalman filter as follows. Recall that

$$Y_t = \hat{K}_t + \hat{X}_{t|t-1} = E^{tr}_t H_{\infty} E_{\infty} (X_t - \hat{X}_{t|t-1}) + E^{tr}_t B_{\infty} V^{c}_t + \hat{X}_{t|t-1}$$

$$= E^{tr}_t H_{\infty} E_{\infty} (CZ_t + NV_t - C\hat{Z}_{t|t-1}) + E^{tr}_t B_{\infty} V^{c}_t + C\hat{Z}_{t|t-1}$$

$$= E^{tr}_t H_{\infty} E_{\infty} CZ_t + \hat{Z}_{t|t-1} + \hat{Z}_{t|t-1} + E^{tr}_t H_{\infty} E_{\infty} NV_t + E^{tr}_t B_{\infty} V^{c}_t$$

where $\{V_t : t \in \mathbb{N}\}$ and $\{V^{c}_t : t \in \mathbb{N}\}$ are independent Gaussian vectors. Then $\hat{Z}_{t|t-1}$ is given by the modified Kalman filter \cite{44} (IV.27), (IV.28). Notice that the ergodic is Gaussian with initial condition $\hat{Z}_0 = \mathbb{E}\{Z_0|Y^{-1}\}$ and $\Sigma_0$ the covariance of $Z_0 - \hat{Z}_0$ which is Gaussian $N(0, \Sigma_{\infty})$.

\section*{APPENDIX G}

\subsection*{NOISELESS CODING THEOREM AND NONANTICIPATIVE RDF}

We introduce the following additional notation. For each $i \in \{0, 1, \ldots, n\}$, let $X_i^r \triangleq \times_{i=0}^{s} X_{i,j}$, for each $j \in \{0, 1, \ldots, s\}$, let $X^s_i \triangleq \times_{i=0}^{s} X_{i,j}$, hence $X^{n,s} \triangleq \times_{i=0}^{n} \times_{j=0}^{s} X_{i,j}$. Thus, for a fixed $i \in \{0, 1, \ldots, n\}$, $x_i^r \in X_{i}^r$ and for a fixed $j \in \{0, 1, \ldots, s\}$, $x_j^n \in X_{i}^n$. Similarly, denote its reproduction by $Y^{n,s} \triangleq \{Y_{i,j} : i = 0, \ldots, n, j = 0, \ldots, s\} \in \times_{i=0}^{n} \times_{j=0}^{s} Y_{i,j}$. Such two-dimensional sources are utilized recently in video coding applications \cite{43}, where the authors derived a coding theorem giving an operational meaning to $R_{0,n}^{na}(D)$. Here, we establish an operational meaning for $R_{0,n}^{na}(D)$ by invoking the coding theorem derived in \cite{7} which is based on the following definition of sequential quantizer.

\begin{definition} \textit{(Sequential quantizer)} \end{definition}

A sequential quantizer is a sequence of measurable functions $f_{i}^{n,s} = \{f_{i}^{s} : i = 0, 1, \ldots, n\}$
defined by \( f_i^s : \mathcal{X}_i \times \mathcal{Y}_i \rightarrow \mathcal{Y}_i \), \( y_i^s = f_i^s(x_i^s, y_{i-1}^s), \ i = 0, 1, \ldots, n \). The set of all such quantizers is denoted by \( \mathcal{F}_{n,s} \).

Thus, a sequential quantizer is nonanticipative with respect to its time index. The operational meaning of the sequential RDF is defined as follows.

**Definition G.2.** (Operational meaning of sequential RDF)

Let \( Q_{0,n,s}^{SRD,o}(D) \) denote the fidelity set

\[
Q_{0,n,s}^{SRD,o} \triangleq \left\{ f^{n,s} \in \mathcal{F}_{n,s} : \frac{1}{n+1} \sum_{i=0}^{n} \mathbb{E}_{P_{X_i,s,Y_i}} \{ \rho_s(X_i^s, Y_i^s) \} \leq D \right\} \quad (G.1)
\]

where \( \rho_s : \mathcal{X}_i \times \mathcal{Y}_i \rightarrow [0, \infty) : i = 0, 1, \ldots, n \) is measurable. The operational sequential RDF is defined by

\[
R_{0,n,s}^{SRD,o}(D) \triangleq \inf_{f^{n,s} \in Q_{0,n,s}^{SRD,o}(D)} \frac{1}{n+1} H(Y_0^s, \ldots, Y_n^s) \quad (G.2)
\]

provided the infimum exists and it is set to \(+\infty\) if it does not.

The operational sequential RDF rate is defined by

\[
R_{0,n}^{SRD,o}(D) \triangleq \lim_{s \to \infty} R_{0,n,s}^{SRD,o}(D),
\]

provided the limit exists and it is set to \(+\infty\) if the infimum in the RHS of \((G.2)\) does not exist.

The information sequential RDF for which a coding theorem is derived in \([7]\) is the following.

Given the two dimensional source \( P_{X^{n,s}}(dx^{n,s}) \), and a reproduction distribution \( P_{Y^{n,s}|X^{n,s}}(dy^{n,s}|x^{n,s}) \), the fidelity set is defined by

\[
Q_{0,n,s}^{SRD}(D) \triangleq \left\{ P_{Y^{n,s}|X^{n,s}}(dy^{n,s}|x^{n,s}) : \frac{1}{n+1} \sum_{i=0}^{n} \mathbb{E}_{P_{X_i,s,Y_i}} \{ \rho_s(X_i^s, Y_i^s) \} \leq D \right\}, \quad (G.3)
\]

The information sequential RDF is defined as follows.

**Definition G.3.** (Information sequential RDF)

Consider fidelity set \( Q_{0,n,s}^{SRD}(D) \) given by \((G.3)\).

The information sequential RDF is defined by

\[
R_{0,n,s}^{SRD}(D) = \inf_{P_{Y^{n,s}|X^{n,s}} \in Q_{0,n,s}^{SRD}(D)} \frac{1}{n+1} I(X^{n,s}; Y^{n,s}), \quad (G.4)
\]

provided the infimum exists and it is set to \(+\infty\) otherwise.
The first part of the following sequential source coding theorem is derived in [7].

**Theorem G.4.** (Sequential source coding theorem)

Suppose \( \{X_{i,j} : i = 0, 1, \ldots, n, j = 0, 1, \ldots, s\} \) are finite alphabets spaces, \( P_{X^n,s}(dx^n) = \otimes_{j=0}^{s} P(dx^j) \), and \( \{X^n_j : j = 0, 1, \ldots, s\} \) are identically distributed (with respect to the spatial index), and there exists an \( x_0 \) and \( D_{\text{max}} > 0 \) such that \( E_{P_{X_{i,j}}} \rho_s(X_{i,j}, x_0) < D_{\text{max}} \), for all \( i = 0, 1, \ldots, n, \ j = 0, 1, \ldots, s \).

**(1)** For any \( \epsilon > 0 \) and finite \( n \in \mathbb{N} \), there exists an integer \( s(\epsilon, n) \) such that for all \( s \geq s(\epsilon, n) \) we have

\[
R_{0,n,s}^{\text{SRD},\epsilon}(D + \epsilon) \leq R_{0,n}^{\text{SRD}}(D) + \epsilon
\]

where

\[
R_{0,n}^{\text{SRD}}(D) \triangleq \inf_{P_{Y^n,|X^n} : \frac{1}{n+1} \mathbb{E}_{P_{X^n,Y^n}} \left\{ \sum_{i=0}^{n} \rho(X_i, Y_i) \leq D \right\} I(X^n; Y^n). \tag{G.6}
\]

**(2)** The following hold.

\[
R_{0,n}^{\text{SRD}}(D) = R_{0,n}^{\text{na}}(D), \ \forall n \in \mathbb{N}
\]

and

\[
\lim_{n \to \infty} \frac{1}{n+1} R_{0,n}^{\text{SRD}}(D) = R_{0,n}^{\text{na}}(D) \triangleq \lim_{n \to \infty} \frac{1}{n+1} R_{0,n}^{\text{na}}(D). \tag{G.8}
\]

**Proof:** (1) The derivation of the coding theorem is found in [7, Chapter 5].

(2) Next, we show (G.7). Notice that \( R_{0,n}^{\text{SRD}}(D) \) given by (G.6) is precisely Gorbunov-Pinsker’s nonanticipatory \( \epsilon \)-entropy \( R_{0,n}^{\epsilon}(D) \). Therefore, by Theorem III.8 we have

\[
R_{0,n}^{\text{SRD}}(D) = R_{0,n}^{\epsilon}(D) = R_{0,n}^{\text{na}}(D), \ n \in \mathbb{N}. \tag{G.9}
\]

Dividing by \( (n + 1) \) and taking the limit as \( n \to \infty \) in both sides of (G.9) yields (G.8).

We conclude this section by stating that Theorem G.4 is derived for finite time index \( n \), and that no assumption is imposed regarding the process \( \{X_{i,j} : i = 0, 1, \ldots, n\} \) for fixed \( j \in \{0, 1, \ldots, s\} \) such as, stationarity, ergodicity, etc.
REFERENCES

[1] P. A. Stavrou, C. K. Kourtellaris, and C. D. Charalambous, “Applications of information nonanticipative rate distortion function,” in IEEE International Symposium on Information Theory (ISIT)(accepted), June 29-July 4 2014. [Online]. Available: [http://arxiv.org/abs/1401.5828v4](http://arxiv.org/abs/1401.5828v4)

[2] T. Berger, Rate Distortion Theory: A Mathematical Basis for Data Compression. Englewood Cliffs, NJ: Prentice-Hall, 1971.

[3] T. S. Han, Information-Spectrum Methods in Information Theory, 2nd ed. Springer-Verlag, Berlin, Heidelberg, New York, 2003.

[4] D. Neuhoff and R. Gilbert, “Causal source codes,” IEEE Transactions on Information Theory, vol. 28, no. 5, pp. 701–713, Sep. 1982.

[5] T. Ericson, “A result on delay-less information transmission,” in IEEE International Symposium on Information Theory (ISIT), 1979.

[6] N. Gaarder and D. Slepian, “On optimal finite-state digital transmission systems,” IEEE Transactions on Information Theory, vol. 28, no. 2, pp. 167–186, 1982.

[7] S. C. Tatikonda, “Control over communication constraints,” Ph.D. dissertation, Mass. Inst. of Tech. (M.I.T.), Cambridge, MA, 2000.

[8] T. Linder and G. Lagosi, “A zero-delay sequential scheme for lossy coding of individual sequences,” IEEE Transactions on Information Theory, vol. 47, no. 6, pp. 2533–2538, 2001.

[9] E. Akyol, K. Viswanatha, K. Rose, and T. A. Ramstad, “On zero delay source-channel coding,” submitted to IEEE Transactions on Information Theory, 2013. [Online]. Available: [http://arxiv.org/abs/1302.3660](http://arxiv.org/abs/1302.3660)

[10] M. Gastpar, B. Rimoldi, and M. Vetterli, “To code, or not to code: Lossy source-channel communication revisited,” IEEE Transactions on Information Theory, vol. 49, no. 5, pp. 1147–1158, May 2003.

[11] M. S. Derpich and J. Østergaard, “Improved upper bounds to the causal quadratic rate-distortion function for gaussian stationary sources,” IEEE Transactions on Information Theory, vol. 58, no. 5, pp. 3131–3152, May 2012.

[12] S. Tatikonda and S. Mitter, “Control over noisy channels,” IEEE Transactions on Automatic Control, vol. 49, no. 7, pp. 1196–1201, July 2004.

[13] G. N. Nair and R. J. Evans, “Stabilizability of Stochastic Linear Systems with Finite Feedback Data Rates,” SIAM Journal on Control and Optimization, vol. 43, no. 2, pp. 413–436, 2004.

[14] C. D. Charalambous and A. Farhadi, “LQG optimality and separation principle for general discrete time partially observed stochastic systems over finite capacity communication channels,” Automatica, vol. 44, no. 12, pp. 3181–3188, 2008.

[15] C. D. Charalambous, C. K. Kourtellaris, and P. Stavrou, “Stochastic control over finite capacity channels: Causality, feedback and uncertainty,” in Proceedings of the 48th IEEE Conference on Decision and Control, held jointly with the 28th Chinese Control Conference (CDC/CCC), 2009, pp. 5889–5894.

[16] A. Farhadi and C. D. Charalambous, “Stability and reliable data reconstruction of uncertain dynamic systems over finite capacity channels,” Automatica, vol. 46, no. 5, pp. 889–896, 2010.

[17] 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.

[18] C. D. Charalambous, P. A. Stavrou, and N. U. Ahmed, “Nonanticipative rate distortion function and relations to filtering theory,” IEEE Transactions on Automatic Control, vol. 59, no. 4, pp. 937–952, April 2014.
[19] C. D. Charalambous and P. A. Stavrou, “On the relation of nonanticipative rate distortion function and filtering theory,” in 12th Biannual European Control Conference (ECC), 17–19 June 2013, pp. 1627–1632.

[20] A. K. Gorbunov and M. S. Pinsker, “Nonanticipatory and prognostic epsilon entropies and message generation rates,” Problems of Information Transmission, vol. 9, no. 3, pp. 184–191, July-Sept. 1973, translation from Problemy Peredachi Informatsii, vol. 9, no. 3, pp. 12-21, July-September 1973.

[21] M. S. Pinsker and A. K. Gorbunov, “Epsilon-entropy with delay for small mean-square reproduction error,” Problems of Information Transmission, vol. 23, no. 2, pp. 91–95, 1987, translation from Problemy Peredachi Informatsii, vol. 23, no. 2, pp. 3-8, April-June 1987.

[22] A. K. Gorbunov and M. S. Pinsker, “Asymptotic behavior of nonanticipative epsilon-entropy for Gaussian processes,” Problems of Information Transmission, vol. 27, no. 4, pp. 361–365, 1991, translation from Problemy Peredachi Informatsii, vol. 27, no. 4, pp. 100-104, October-December 1991.

[23] I. Csiszár, “On an extremum problem of information theory,” Studia Scientiarum Mathematicarum Hungarica, vol. 9, pp. 57–71, 1974.

[24] R. Gray, “Information rates of stationary ergodic finite-alphabet sources,” IEEE Transactions on Information Theory, vol. 17, no. 5, pp. 516–523, 1971.

[25] T. Berger, “Explicit bounds to r(d) for a binary symmetric markov source,” IEEE Transactions on Information Theory, vol. 23, no. 1, pp. 52–59, 1977.

[26] S. Jalali and T. Weissman, “New bounds on the rate-distortion function of a binary markov source,” in IEEE International Symposium on Information Theory (ISIT), June 2007, pp. 571–575.

[27] T. Berger, “Living information theory,” IEEE Information Theory Society Newsletter, vol. 53, no. 1, pp. 6–19, Mar 2003.

[28] V. Kostina and S. Verdu, “Fixed-length lossy compression in the finite blocklength regime,” IEEE Transactions on Information Theory, vol. 58, no. 6, pp. 3309–3338, 2012.

[29] ———, “Lossy joint source-channel coding in the finite blocklength regime,” IEEE Transactions on Information Theory, vol. 59, no. 5, pp. 2545–2575, 2013.

[30] C. D. Charalambous and P. A. Stavrou, “Directed information on abstract spaces: Properties and variational equalities,” submitted to IEEE Transactions on Information Theory, 2013. [Online]. Available: http://arxiv.org/abs/1302.3971

[31] 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.

[32] P. Dupuis and R. S. Ellis, A Weak Convergence Approach to the Theory of Large Deviations. John Wiley & Sons, Inc., New York, 1997.

[33] P. Billingsley, Convergence of Probability Measures, 2nd ed., ser. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons Inc., 1999.

[34] P. A. Stavrou and C. D. Charalambous, “Variational equalities of directed information and applications,” in IEEE International Symposium on Information Theory (ISIT), 7-12 July 2013, pp. 2577–2581.

[35] R. Blahut, “Computation of channel capacity and rate-distortion functions,” IEEE Transactions on Information Theory, vol. 18, no. 4, pp. 460–473, July 1972.

[36] T. Linder and R. Zamir, “Causal coding of stationary sources and individual sequences with high resolution,” IEEE Transactions on Information Theory, vol. 52, no. 2, pp. 662–680, Feb. 2006.

[37] I. Csiszár, “Arbitrarily varying channels with general alphabets and states,” IEEE Transactions on Information Theory, vol. 38, no. 6, pp. 1725–1742, Nov. 1992.
[38] D. P. Bertsekas and S. E. Shreve, *Stochastic Optimal Control: The Discrete-Time Case*, ser. Optimization and Neural Computation Series. Athena Scientific, 2007.

[39] D. G. Luenberger, *Optimization by Vector Space Methods*. John Wiley & Sons, Inc., New York, 1969.

[40] V. Solo, “On causality and mutual information,” in *47th IEEE Conference on Decision and Control (CDC ’08)*, Dec. 2008, pp. 4939–4944.

[41] F. H. Lin, K. Hara, V. Solo, M. Vangel, J. W. Belliveau, S. T. Stufflebeam, and H. M. S., “Dynamic Granger-Geweke causality modeling with application to interictal spike propagation,” *Human Brain Mapping*, vol. 30, no. 6, pp. 1877–1886, June 2009.

[42] J. Pearl, *Causality: models, reasoning, and inference*, 2nd ed. New York, NY, USA: Cambridge University Press, 2009.

[43] N. Ma and P. Ishwar, “On delayed sequential coding of correlated sources,” *IEEE Transactions on Information Theory*, vol. 57, no. 6, pp. 3763–3782, 2011.

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

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

[46] P. W. Glynn and D. Ormoneit, “Hoeffding’s inequality for uniformly ergodic markov chains,” *Statistics & Probability Letters*, vol. 56, no. 2, pp. 143–146, 2002.

[47] C. A. León and F. Perron, “Optimal hoeffding bounds for discrete reversible markov chains,” *The Annals of Applied Probability*, vol. 14, no. 2, pp. 958–970, May 2004.

[48] F. P. Kelly, *Reversibility and Stochastic Networks*. New York, NY, USA: Cambridge University Press, 2011.

[49] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*, 2nd ed. Springer-Verlag Inc., New York, 1998.

[50] T. Weissman and N. Merhav, “On causal source codes with side information,” *IEEE Transactions on Information Theory*, vol. 51, no. 11, pp. 4003–4013, 2005.