

Nonanticipative Rate Distortion Function and Relations to Filtering Theory

Charalambos D. Charalambous, Photios A. Stavrou, *Student Member, IEEE*,
and Nasir U. Ahmed

Abstract

The relation between nonanticipative Rate Distortion Function (RDF) and filtering theory is discussed on abstract spaces. The relation is established by imposing a realizability constraint on the reconstruction conditional distribution of the classical RDF. Existence of the extremum solution of the nonanticipative RDF is shown using weak*-convergence on appropriate topology. The extremum reconstruction conditional distribution is derived in closed form, for the case of stationary processes. The realization of the reconstruction conditional distribution which achieves the infimum of the nonanticipative RDF is described. Finally, an example is presented to illustrate the concepts.

Index Terms

Nonanticipative Rate Distortion Function (RDF), filtering, realization, weak*-convergence, optimal reconstruction conditional distribution

I. INTRODUCTION

Shannon's information theory [2] for reliable communication evolved over the years without much emphasis on nonanticipation imposed on the communication sub-systems. In particular, the

This work was financially supported by a medium size University of Cyprus grant entitled "DIMITRIS" and by European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. INFOS-ICT-223844. Part of this work was presented in 20th International Symposium on Mathematical Theory of Networks and Systems (MTNS '12) [1].

Charalambos D. Charalambous and Photios A. Stavrou 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: chadcha@ucy.ac.cy, stavrou.fotios@ucy.ac.cy

Nasir U. Ahmed is with the School of Electrical Engineering and Computer Science, University of Ottawa, 161 Louis Pasteur, P.O. Box 450, Stn A, Ottawa, Ontario, Canada, K1N 6N5, e-mail: ahmed@site.uottawa.ca

classical rate distortion function (RDF) for source data compression or quantization deals with the characterization of the optimal reconstruction conditional distribution subject to a fidelity criterion [3], [4], without regard to nonanticipation.

On the other hand, filtering theory is developed by imposing real-time realizability on the estimators with respect to measurement data. Although, both reliable communication and filtering (state estimation for control) are concerned with reconstruction of processes, the main underlying assumptions characterizing them are different.

In this paper, the intersection of rate distortion function (RDF) and real-time realizable filtering theory is established by invoking a nonanticipative constraint on the reconstruction conditional distribution to be realizable via real-time operations, while the optimal nonanticipative reconstruction conditional distribution is derived. Consequently, the connection between nonanticipative RDF, its characterization via the optimal reconstruction conditional distribution, and real-time realizable filtering theory is established under very general conditions on the source (including Markov sources).

The fundamental advantage of the new filtering approach based on nonanticipative RDF, is the ability to ensure average or probabilistic estimation error constraints, which is non-trivial task if Bayesian filtering techniques are employed to formulate such constraints. The motivations includes nonanticipative data compression over noisy channels, such as control over networks, where the controlled system and controller may be connected via a noisy channel [5]–[10]. In such applications, filtering via nonanticipative RDF approximates sensor measurements by the reconstruction process taking values in a set of smaller cardinality, while the approximation is quantified by the distortion function. Given the recent interest in developing controller and estimator architectures processing quantized information and, in general, communication schemes for control applications, nonanticipative RDF is necessary for developing zero-delay or limited delay quantization schemes. Moreover, nonanticipative RDF is necessary for the realization of the compression channel by communication systems processing information causally.

The first relation between information theory and filtering via distortion rate function is discussed by R. S. Bucy in [11], by carrying out the computation of a realizable (nonanticipative) distortion rate function with square criteria for two samples of the Ornstein-Uhlenbeck Gaussian process. Related work on nonanticipative rate distortion theory is pursued by A. K. Gorbunov and M. S. Pinsker in [12], [13]. Specifically, [12] discussed nonanticipative RDF for general

stationary processes and establishes existence of the infinite horizon limit, while [13] computes a closed form expression for nonanticipative RDF (called ϵ -entropy) for stationary Gaussian processes using power spectral methods. Further elaborations on the similarities and differences between [11]–[13] and this paper will be discussed in subsequent parts of the paper. Moreover, over the years several papers appeared in the literature in which controller or estimator are designed based on information theoretic measures [14]–[16]. An earlier work designing filters via information theoretic measures is [17], while [18] analyzes mutual information for Gaussian processes.

In this paper, the connection between nonanticipative rate distortion theory and filtering theory is further examined, under a nonanticipative condition defined by the family of conditional distributions (reconstructions), for general distortion functions and random processes on abstract Polish spaces. The connection is established via optimization on the space of conditional distributions with average distortion constraint and almost sure (*a.s.*) constraints to account for the nonanticipative condition on the reconstruction conditional distribution. The main results are the following.

- (1) Existence of the nonanticipative RDF using the topology of weak*-convergence;
- (2) Closed form expression for reconstruction conditional distribution minimizing the nonanticipative RDF for stationary processes;
- (3) Realization procedure of the filter based on the nonanticipative RDF;
- (4) Example to demonstrate the realization of the filter.

It is important to point out that items (1)–(4) above are not addressed in the related papers [11]–[13]. Moreover, (2) together with (3) are important in reliable communication for filtering and control applications, because one can develop communication architectures which operate with zero-delay or limited delay, as opposed to the classical RDF which is anticipative.

Next, we give a high level discussion on Bayesian filtering theory and nonanticipative RDF, and we present some aspects of the problem pursued in this paper. Consider a discrete-time process $X^n \triangleq \{X_0, X_1, \dots, X_n\} \in \mathcal{X}_{0,n} \triangleq \times_{i=0}^n \mathcal{X}_i$, and its reconstruction $Y^n \triangleq \{Y_0, Y_1, \dots, Y_n\} \in \mathcal{Y}_{0,n} \triangleq \times_{i=0}^n \mathcal{Y}_i$, where \mathcal{X}_i and \mathcal{Y}_i are Polish spaces (complete separable metric spaces). The objective is to reconstruct X^n by Y^n via nonanticipative operations subject to a distortion or fidelity criterion. That is, for each $i = 0, 1, \dots$, the reconstruction Y_i of X_i should depend on past and present information $\{X_0, Y_0, X_1, Y_1, \dots, X_{i-1}, Y_{i-1}, X_i\}$. Once this mapping is found a

procedure is introduced to realize the filter of Y_i from auxiliary¹ measurements.

A. Bayesian Estimation Theory

In classical filtering [19], one is given a mathematical model that generates the process X^n , via its conditional distribution $\{P_{X_i|X^{i-1}}(dx_i|x^{i-1}) : i = 0, 1, \dots, n\}$ or via discrete-time recursive dynamics, a mathematical model that generates observed data obtained from sensors, say, Z^n , $\{P_{Z_i|Z^{i-1}, X^i}(dz_i|z^{i-1}, x^i) : i = 0, 1, \dots, n\}$, while Y^n are the causal estimates of some function of the process X^n based on the observed data Z^n . Note that for a memoryless channel that generates the observation sequence $\{Z_i : i = 0, 1, \dots, n\}$ then $P_{Z_i|Z^{i-1}, X^i}(dz_i|z^{i-1}, x^i) = P_{Z_i|X_i}(dz_i|x_i) - a.s., i = 0, 1, \dots, n$.

In Bayesian estimation one is interested in causal estimators of some function $\Phi : \mathcal{X}_n \mapsto \mathbb{R}$, $Y_n \triangleq \Phi(X_n)$ based on the observed data $Z^{n-1} \triangleq \{Z_0, Z_1, \dots, Z_{n-1}\}$. With respect to minimizing the least-squares error pay-off, the best estimate of $\Phi(X_i)$ given Z^{i-1} , denoted by $\hat{\Phi}(X_i)$, is given by the conditional mean

$$\hat{\Phi}(X_i) \triangleq \mathbb{E}\{\Phi(X_i)|Z^{i-1}\} = \int_{\mathcal{X}_i} \Phi(x) P_{X_i|Z^{i-1}}(dx|z^{i-1}), \quad i = 0, 1, \dots, n.$$

For non-linear problems, Bayesian filtering is often addressed via the conditional distribution $\{P_{X_i|Z^{i-1}}(dx_i|z^{i-1}) : i = 0, 1, \dots, n\}$ or its unnormalized versions which satisfy discrete-recursions [19], and forms a sufficient statistic for the filtering problem.

Consider the simplified example of the multi-dimensional Gaussian-Markov processes modeled by

$$\begin{cases} X_{k+1} = A_k X_k + B_k W_k, & X_0 \sim N(0; \Sigma_{x_0}), \quad k = 0, 1, \dots, n-1 \\ Z_k = C_k X_k + D_k V_k, & k = 0, 1, \dots, n \end{cases} \quad (\text{I.1})$$

where $\{A_k, B_k, C_k, D_k\}$ are time-varying matrices having appropriate dimensions, $W_k \sim N(0; \Sigma_{W_k})$ (Gaussian with mean zero and covariance Σ_{W_k}), $V_k \sim N(0; \Sigma_{V_k})$, $k = 0, 1, \dots, n$, while the processes $\{W_k : k = 0, 1, \dots, n-1\}$, $\{V_k : k = 0, 1, \dots, n\}$ are mutually independent, and independent of X_0 . The classical Kalman Filter [19] is a well-known example for which the optimal reconstruction $\hat{X}_i = \mathbb{E}[X_i|Z^{i-1}]$, $i = 0, 1, \dots, n$, is the conditional mean which minimizes the average least-squares estimation error. Thus, in classical filtering theory both models which

¹This point is explained in Subsection I-B.

generate the unobserved and observed processes, X^n and Z^n , respectively, are given á priori, and the estimator \hat{X}_i is a nonanticipative function of the past information Z^{i-1} , $i = 0, 1, \dots, n$. Fig. I.1 illustrates the cascade block diagram of the Bayesian filtering problem.

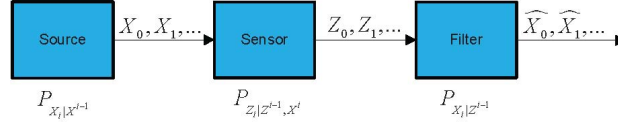


Fig. I.1. Block Diagram of Bayesian Filtering Problem

B. Nonanticipative Rate Distortion Theory and Estimation

In nonanticipative rate distortion theory one is given the process X^n , which induces the conditional distributions $\{P_{X_i|X^{i-1}}(dx_i|x^{i-1}) : i = 0, 1, \dots, n\}$ and determines the nonanticipative reconstruction conditional distribution $\{P_{Y_i|Y^{i-1}, X^i}(dy_i|y^{i-1}, x^i) : i = 0, 1, \dots, n\}$ which minimizes the mutual information between X^n and Y^n subject to a distortion or fidelity constraint, via a nonanticipative or realizability constraint. The filter or estimate $\{Y_i : i = 0, 1, \dots, n\}$ of $\{X_i : i = 0, 1, \dots, n\}$ is found by realizing the reconstruction distribution $\{P_{Y_i|Y^{i-1}, X^i}(dy_i|y^{i-1}, x^i) : i = 0, 1, \dots, n\}$ via a cascade of sub-systems as shown in Fig. I.2. The point to be

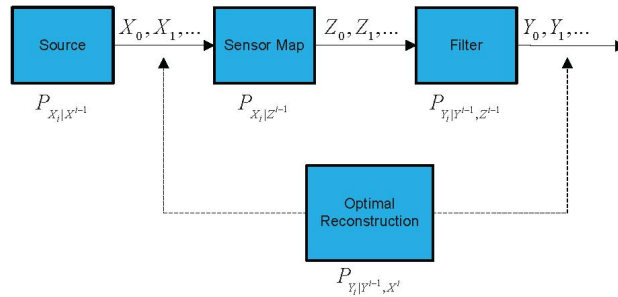


Fig. I.2. Block Diagram of Filtering via Nonanticipative Rate Distortion Function

made here is that the auxiliary random sequence $\{Z_0, Z_1, \dots\}$ which is the analogue of sensor measurements (in the above discussion of Bayesian estimation) is identified during the realization of the optimal reconstruction distribution $\{P_{Y_i|Y^{i-1}, X^i}(dy_i|y^{i-1}, x^i) : i = 0, 1, \dots, n\}$. Thus, in Bayesian estimation, the sensor map is given á priori, while in nonanticipative rate distortion

theory, this map is identified during the realization of the optimal reconstruction conditional distribution $\{P_{Y_i|Y^{i-1}, X^i}(dy_i|y^{i-1}, x^i) : i = 0, 1, \dots, n\}$, so that the end-to-end nonanticipative RDF from X^n to Y^n is achieved.

The precise problem formulation of nonanticipative RDF is defined by first introducing the distortion or fidelity constraint and mutual information. The distortion function [3] or fidelity constraint between x^n and its reconstruction y^n , is a measurable function defined by

$$d_{0,n} : \mathcal{X}_{0,n} \times \mathcal{Y}_{0,n} \mapsto [0, \infty], \quad d_{0,n}(x^n, y^n) \triangleq \sum_{i=0}^n \rho_{0,i}(x^i, y^i).$$

For single letter distortion $d_{0,n}(x^n, y^n) \equiv \sum_{i=0}^n \rho(x_i, y_i)$, and for single letter square error distortion $d_{0,n}(x^n, y^n) \equiv \sum_{i=0}^n \|x_i - y_i\|^2$ [3]. Moreover, for finite alphabet spaces \mathcal{X}_i and \mathcal{Y}_i , the distortion function can be defined in terms of the Hamming distance [4].

The mutual information between X^n and Y^n , for a given distribution $P_{X^n}(dx^n)$, and conditional distribution $P_{Y^n|X^n}(dy^n|x^n)$, is defined by² [3]

$$I(X^n; Y^n) \triangleq \int_{\mathcal{X}_{0,n} \times \mathcal{Y}_{0,n}} \log \left(\frac{P_{Y^n|X^n}(dy^n|x^n)}{P_{Y^n}(dy^n)} \right) P_{Y^n|X^n}(dy^n|x^n) \otimes P_{X^n}(dx^n). \quad (\text{I.2})$$

Next, introduce the nonanticipative constraint on the reconstruction distribution. To this end, define the $(n+1)$ -fold nonanticipative convolution measure

$$\vec{P}_{Y^n|X^n}(dy^n|x^n) \triangleq \otimes_{i=0}^n P_{Y_i|Y^{i-1}, X^i}(dy_i|y^{i-1}, x^i) - a.s. \quad (\text{I.3})$$

The set of nonanticipative reconstruction distributions is defined by

$$\vec{\mathcal{Q}}_{ad} \triangleq \left\{ P_{Y^n|X^n}(dy^n|x^n) : P_{Y^n|X^n}(dy^n|x^n) = \vec{P}_{Y^n|X^n}(dy^n|x^n) - a.s. \right\}. \quad (\text{I.4})$$

Note that without the nonanticipative constraint specified by $\vec{\mathcal{Q}}_{ad}$, the connection between filtering theory and rate distortion theory cannot be established, since in general by Bayes' rule $P_{Y^n|X^n}(dy^n|x^n) = \otimes_{i=0}^n P_{Y_i|Y^{i-1}, X^n}(dy_i|y^{i-1}, x^n) - a.s.$, and hence, for each $i = 0, 1, \dots, n$, the conditional distribution $P_{Y_i|Y^{i-1}, X^n}(\cdot|\cdot, \cdot)$ of Y_i will depend on future symbols $\{X_{i+1}, X_{i+2}, \dots, X_n\}$, in addition to the past and present symbols $\{Y^{i-1}, X^i\}$. However, by imposing the nonanticipative constraint (I.4), then at each time instant $i = 0, 1, \dots$, the reconstruction Y_i of X_i will depend on the past reconstructions $\{Y_0, \dots, Y_{i-1}\}$ and past and present symbols $\{X_0, \dots, X_i\}$. For filtering and control applications, the nonanticipative constraint is necessary to avoid anticipative

²The precise definition of a convolution of measures denoted by \otimes is given in Section II.

processing of information, while for quantization or compression applications it offers the possibility to realize the compression channel (optimal reconstruction distribution) via causal operations and achieve an end-to-end compression with zero-delay.

Nonanticipative Distortion Rate Function. The nonanticipative distortion rate function is defined by the minimization over $P_{Y^n|X^n}(dy^n|x^n)$ of the average distortion function subject to a constraint on the mutual information rate $I(X^n; Y^n) \leq R$ and the nonanticipative constraint (I.4) as follows.

$$D_{0,n}^{na}(R) \triangleq \inf_{P_{Y^n|X^n}(dy^n|x^n) \in \vec{\mathcal{Q}}_{ad}: I(X^n; Y^n) \leq R} \mathbb{E} \left\{ d_{0,n}(X^n, Y^n) \right\}. \quad (\text{I.5})$$

The classical distortion rate function does not impose the nonanticipative constraint $P_{Y^n|X^n}(dy^n|x^n) = \vec{P}_{Y^n|X^n}(dy^n|x^n) - a.s.$, hence the resulting optimal reconstruction distribution of symbol y_i will depend on (y^{i-1}, x^i) and on future symbols (x_{i+1}, \dots, x_n) . Thus, by solving (I.5) and then realizing the conditional distribution the optimal causal filter will be defined.

At this stage it is important to point out that the nonanticipative condition (I.4) is different from the realizability condition in [11], in which is assumed that Y_i is independent of $X_{j|i}^* \triangleq X_j - \mathbb{E}(X_j|X^i)$, $j = i+1, i+2, \dots$. Moreover, the nonanticipative condition (I.4) is implied by the nonanticipative condition found in [12], [13], defined by $X_{n+1}^\infty \leftrightarrow X^n \leftrightarrow Y^n$ forms a Markov chain for any $n = 0, 1, \dots$ (e.g., $P_{Y^n|X^n, X_{n+1}^\infty}(dy^n|x^n, x_{n+1}^\infty) = P_{Y^n|X^n}(dy^n|x^n)$, $n = 0, 1, \dots$). The claim here is that the nonanticipative condition (I.4) is more natural and applies to processes which are not necessarily Gaussian with square error distortion function, and subject to slight modification to controlled sources in which the control is a function of the reconstruction process (we shall discuss this point further).

Nonanticipative Rate Distortion Function. An equivalent problem to (I.5) is the nonanticipative RDF defined by

$$R_{0,n}^{na}(D) \triangleq \inf_{P_{Y^n|X^n}(dy^n|x^n) \in \vec{\mathcal{Q}}_{ad}: \mathbb{E} \left\{ d_{0,n}(X^n, Y^n) \right\} \leq D} I(X^n; Y^n). \quad (\text{I.6})$$

The two problems defined by (I.5) and (I.6) are equivalent in the sense that the solution of (I.5)

gives that of (I.6) and vice-versa [4]. Moreover, it can be shown that

$$\begin{aligned}
P_{Y^n|X^n}(dy^n|x^n) &= \vec{P}_{Y^n|X^n}(dy^n|x^n) - a.s. \iff \\
I(X^n; Y^n) &= \int_{\mathcal{X}_{0,n} \times \mathcal{Y}_{0,n}} \log \left(\frac{\vec{P}_{Y^n|X^n}(dy^n|x^n)}{P_{Y^n}(dy^n)} \right) \vec{P}_{Y^n|X^n}(dy^n|x^n) \otimes P_{X^n}(dx^n) \\
&\equiv \mathbb{I}(P_{X^n}, \vec{P}_{Y^n|X^n})
\end{aligned} \tag{I.7}$$

where the notation $\mathbb{I}(P_{X^n}, \vec{P}_{Y^n|X^n})$ is used to point out the functional dependence of $I(X^n; Y^n)$ on $\{P_{X^n}, \vec{P}_{Y^n|X^n}\}$. The nonanticipative distortion rate function and RDF can be generalized to controlled sources.

The paper is organized as follows. Section II discusses the problem formulation on abstract spaces. Section III establishes existence of optimal minimizing reconstruction distribution, and Section IV derives the stationary solution. Section V describes the real-time realization of nonanticipative RDF. Finally, Section VI demonstrates the filter realization via an example.

II. FORMULATION OF NONANTICIPATIVE RDF ON ABSTRACT SPACES

Throughout the paper we use the notation defined on Table I. The source and reconstruction alphabets, respectively, are sequences of Polish spaces $\{\mathcal{X}_t : t \in \mathbb{N}\}$ and $\{\mathcal{Y}_t : t \in \mathbb{N}\}$, associated with their corresponding measurable spaces $(\mathcal{X}_t, \mathcal{B}(\mathcal{X}_t))$ and $(\mathcal{Y}_t, \mathcal{B}(\mathcal{Y}_t))$, $t \in \mathbb{N}$. Sequences of alphabets are identified with the product spaces $(\mathcal{X}_{0,n}, \mathcal{B}(\mathcal{X}_{0,n})) \triangleq \times_{k=0}^n (\mathcal{X}_k, \mathcal{B}(\mathcal{X}_k))$, and $(\mathcal{Y}_{0,n}, \mathcal{B}(\mathcal{Y}_{0,n})) \triangleq \times_{k=0}^n (\mathcal{Y}_k, \mathcal{B}(\mathcal{Y}_k))$. The source and reconstruction are random processes denoted by $X^n \triangleq \{X_t : t \in \mathbb{N}^n\}$, $X : \{t\} \times \Omega \mapsto \mathcal{X}_t$, and by $Y^n \triangleq \{Y_t : t \in \mathbb{N}^n\}$, $Y : \{t\} \times \Omega \mapsto \mathcal{Y}_t$, respectively.

The reconstruction conditional distribution will be defined via stochastic kernels. Note that the random variable (RV) Z is called conditional independent of RV X given the RV Y if and only if $X \leftrightarrow Y \leftrightarrow Z$ forms a MC in both directions, equivalently $P_{X,Z|Y}(dx, dz|y) = P_{X|Y}(dx|y)P_{Z|Y}(dz|y) - a.s.$, or $P_{Z|Y,X}(dz|y, x) = P_{Z|Y}(dz|y) - a.s.$

Definition II.1. [20] Let $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$, $(\mathcal{Y}, \mathcal{B}(\mathcal{Y}))$ be measurable spaces in which \mathcal{Y} is a Polish Space.

A stochastic kernel on \mathcal{Y} given \mathcal{X} is a mapping $q : \mathcal{B}(\mathcal{Y}) \times \mathcal{X} \rightarrow [0, 1]$ satisfying the following two properties:

TABLE I
TABLE OF NOTATION

$\mathbb{N} \triangleq \{0, 1, \dots\}$	Set of nonnegative integers
$\mathbb{N}^n \triangleq \{0, 1, \dots, n\}$	Set of first $(n + 1)$ nonnegative integers
$\mathcal{X}_t, \mathcal{Y}_t, t \in \mathbb{N}$	Source and reconstruction alphabets
$\mathcal{B}(\mathcal{X}_t), \mathcal{B}(\mathcal{Y}_t)$	σ -algebra of Borel sets generated by $\mathcal{X}_t, \mathcal{Y}_t$
$\mathcal{X}_{0,n} \triangleq \times_{k=0}^n \mathcal{X}_k, \mathcal{Y}_{0,n} \triangleq \times_{k=0}^n \mathcal{Y}_k$	Cartesian product of source and reconstruction alphabets
$\mathcal{B}(\mathcal{X}_{0,n}) \triangleq \times_{k=0}^n \mathcal{B}(\mathcal{X}_k), \mathcal{B}(\mathcal{Y}_{0,n}) \triangleq \times_{k=0}^n \mathcal{B}(\mathcal{Y}_k)$	σ -algebra of Borel sets generated by $\mathcal{X}_{0,n}, \mathcal{Y}_{0,n}$
$x^n \triangleq \{x_0, \dots, x_n\}, y^n \triangleq \{y_0, \dots, y_n\}$	Sequence of source and reconstruction symbols
$\mathcal{M}_1(\mathcal{Z})$	Set of probability measures on a measurable space $(\mathcal{Z}, \mathcal{B}(\mathcal{Z}))$
$\mathcal{Q}(\mathcal{Y}; \mathcal{X})$	Set of stochastic kernels on $(\mathcal{Y}, \mathcal{B}(\mathcal{Y}))$ given $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$
$X \leftrightarrow Y \leftrightarrow Z \Leftrightarrow P_{Z X,Y}(dz x,y) = P_{Z Y}(dz y) - a.s.$	Markov Chain (MC) or conditional independence
$BC(\mathcal{X})$	Vector space of bounded continuous real-valued functions defined on a Polish space \mathcal{X}
$L_1(\mu, BC(\mathcal{X}))$	Set of all μ -integrable functions defined on \mathcal{X} with values in $BC(\mathcal{X})$
$\ \cdot\ _\mu$	Norm with respect to $L_1(\mu, BC(\mathcal{X}))$
\mathcal{X}^*	Topological dual of a Banach space \mathcal{X}
$M_{rba}(\mathcal{X})$	Space of finitely additive regular bounded signed measures on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$
$\Pi_{rba}(\mathcal{X}) \subset M_{rba}(\mathcal{X})$	Space of finitely additive regular bounded probability measures on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$

- (1) For every $x \in \mathcal{X}$, the set function $q(\cdot; x)$ is a probability measure (possibly finitely additive) on $\mathcal{B}(\mathcal{Y})$;
- (2) For every $F \in \mathcal{B}(\mathcal{Y})$, the function $q(F; \cdot)$ is $\mathcal{B}(\mathcal{X})$ -measurable.

Stochastic kernels can be used to define anticipative and nonanticipative convolution of reconstruction kernels and associated classical and nonanticipative RDF.

Definition II.2. Given measurable spaces $(\mathcal{X}_{0,n}, \mathcal{B}(\mathcal{X}_{0,n}))$, $(\mathcal{Y}_{0,n}, \mathcal{B}(\mathcal{Y}_{0,n}))$, and their product spaces, data compression channels are classified as follows.

- 1) An Anticipative Data Compression Channel is a stochastic kernel $q_{0,n}(dy^n; x^n) \in \mathcal{Q}(\mathcal{Y}_{0,n}; \mathcal{X}_{0,n})$. Such a kernel admits a factorization into a convolution of a sequence of anticipative

stochastic kernels as follows

$$q_{0,n}(dy^n; x^n) = \otimes_{i=0}^n q_i(dy_i; y^{i-1}, x^n) - a.s. \quad (\text{II.8})$$

where $q_i(dy_i; y^{i-1}, x^n) \in \mathcal{Q}(\mathcal{Y}_i; \mathcal{Y}_{0,i-1} \times \mathcal{X}_{0,n}), i = 0, \dots, n, n \in \mathbb{N}$.

- 2) A Nonanticipative Convolution Data Compression Channel is a convolution of a sequence of nonanticipative stochastic kernels defined by

$$\vec{q}_{0,n}(dy^n; x^n) \triangleq \otimes_{i=0}^n q_i(dy_i; y^{i-1}, x^i) - a.s. \quad (\text{II.9})$$

where $q_i(dy_i; y^{i-1}, x^i) \in \mathcal{Q}(\mathcal{Y}_i; \mathcal{Y}_{0,i-1} \times \mathcal{X}_{0,i}), i = 0, \dots, n, n \in \mathbb{N}$.

- 3) A Restricted Nonanticipative Data Compression Channel is a stochastic kernel $q_{0,n}(dy^n; x^n) \in \mathcal{Q}(\mathcal{Y}_{0,n}; \mathcal{X}_{0,n})$ which is a convolution of a sequence of nonanticipative stochastic kernels obtained by imposing the almost sure (a.s.) constraint defined by

$$q_{0,n}(dy^n; x^n) = \otimes_{i=0}^n q_i(dy_i; y^{i-1}, x^i) - a.s. \quad (\text{II.10})$$

where $q_i \in \mathcal{Q}(\mathcal{Y}_i; \mathcal{Y}_{0,i-1} \times \mathcal{X}_{0,i}), i = 0, \dots, n, n \in \mathbb{N}$.

As stated earlier, the classical RDF is concerned with optimizing (I.2) with respect to anticipative stochastic kernels (II.8). In this paper we address problem (I.5) or (I.6), i.e., when the conditional distribution (stochastic kernel) is restricted nonanticipative, and we discuss generalizations based on (II.9). That is, for a given distribution $P_{X^n}(dx^n)$, nonanticipative RDF imposes the a.s.-constraint (II.10) on the reconstruction conditional distribution, and hence on the joint distribution $P_{X^n, Y^n}(dx^n, dy^n)$ generated by them, unlike the classical RDF which does not impose such a constraint. However, when the source is independently distributed, i.e., $P_{X^n}(dx^n) = \otimes_{i=0}^n P_{X_i}(dx_i) - a.s.$, it is well known that the optimal reconstruction conditional distribution of the classical RDF has the property $P_{Y^n|X^n}^*(dy^n|x^n) = \otimes_{i=0}^n P_{Y_i|X_i}^*(dy_i|x_i) - a.s.$ It is also well known that for sources with memory (i.e., Markov sources) the optimal reconstruction conditional distribution of the classical RDF is anticipative, i.e., $P_{Y^n|X^n}^*(dy^n|x^n) = \otimes_{i=0}^n P_{Y_i|Y^{i-1}, X^n}^*(dy_i|y^{i-1}, x^n) - a.s.$ Therefore, to ensure a nonanticipative reconstruction conditional distribution one has to impose the constraint (II.10) to the classical RDF. On the other hand, a nonanticipative convolution data compression channel (II.9) does not impose any constraint on the joint distribution $P_{X^n, Y^n}(dx^n, dy^n)$. This point is further explained below by discussing generalizations of distortion rate function (I.5) and RDF (I.6).

Remark II.3. *The nonanticipative distortion rate function and the nonanticipative RDF can be generalized as follows. Given a sequence of conditional distributions $\{P_{X_i|X^{i-1}, Y^{i-1}}(dx_i|x^{i-1}, y^{i-1}) : i = 0, 1, \dots, n\}$ then (I.5) and (I.6) can be generalized to*

$$\vec{D}_{0,n}^{na}(R) \triangleq \inf_{\vec{P}_{Y^n|X^n}(dy^n|x^n): I(X^n \rightarrow Y^n) \leq R} E\{d_{0,n}(X^n, Y^n)\} \quad (\text{II.11})$$

$$\vec{R}_{0,n}^{na}(D) \triangleq \inf_{\vec{P}_{Y^n|X^n}(dy^n|x^n): E\{d_{0,n}(X^n, Y^n)\} \leq D} I(X^n \rightarrow Y^n) \quad (\text{II.12})$$

where $I(X^n \rightarrow Y^n)$ is the directed information measure from X^n to Y^n defined by

$$\begin{aligned} I(X^n \rightarrow Y^n) &\triangleq \int_{\mathcal{X}_{0,n} \times \mathcal{Y}_{0,n}} \log \left(\frac{\vec{P}_{Y^n|X^n}(dy^n|x^n)}{P_{Y^n}(dy^n)} \right) \vec{P}_{Y^n|X^n}(dy^n|x^n) \otimes \overleftarrow{P}_{X^n|Y^{n-1}}(dx^n|y^{n-1}) \\ &\equiv \mathbb{I}_{X^n \rightarrow Y^n}(\overleftarrow{P}_{X^n|Y^{n-1}}, \vec{P}_{Y^n|X^n}) \end{aligned}$$

where $\overleftarrow{P}_{X^n|Y^{n-1}}$ is defined by

$$\overleftarrow{P}_{X^n|Y^{n-1}}(dx^n|y^{n-1}) \triangleq \otimes_{i=0}^n P_{X_i|X^{i-1}, Y^{i-1}}(dx_i|x^{i-1}, y^{i-1}) - a.s.$$

Clearly, (II.11) and (II.12) do not assume $P_{X_i|X^{i-1}, Y^{i-1}}(dx_i|x^{i-1}, y^{i-1}) = P_{X_i|X^{i-1}}(dx_i|x^{i-1}) - a.s.$, and hence the process X^n can be affected by Y^n causally. It is easy to show that if (II.10) holds then $P_{X_i|X^{i-1}, Y^{i-1}}(dx_i|x^{i-1}, y^{i-1}) = P_{X_i|X^{i-1}}(dx_i|x^{i-1}) - a.s.$, $i = 0, 1, \dots, n$, also holds, and hence (II.11), (II.12) reduce to (I.5), (I.6). The generalizations (II.11), (II.12), covers conditionally Gaussian sources as a special case [21]. It also covers the case when the source is a controlled process, controlled over a finite rate channel based on the quantized or reconstruction process. These generalizations will be investigated elsewhere, since they will require new topological spaces on which existence of optimal solution to (II.11) and (II.12) can be shown.

A. Nonanticipative RDF

In this subsection the nonanticipative RDF is rigorously defined on abstract spaces. Given a source probability measure $\mu_{0,n} \in \mathcal{M}_1(\mathcal{X}_{0,n})$ (possibly finitely additive) and a reconstruction kernel $q_{0,n} \in \mathcal{Q}(\mathcal{Y}_{0,n}; \mathcal{X}_{0,n})$, one can define three probability measures as follows.

(P1): The joint measure $P_{0,n} \in \mathcal{M}_1(\mathcal{Y}_{0,n} \times \mathcal{X}_{0,n})$:

$$\begin{aligned} P_{0,n}(G_{0,n}) &\triangleq (\mu_{0,n} \otimes q_{0,n})(G_{0,n}), \quad G_{0,n} \in \mathcal{B}(\mathcal{X}_{0,n}) \times \mathcal{B}(\mathcal{Y}_{0,n}) \\ &= \int_{\mathcal{X}_{0,n}} q_{0,n}(G_{0,n}, x^n; x^n) \mu_{0,n}(dx^n) \end{aligned}$$

where G_{0,n,x^n} is the x^n -section of $G_{0,n}$ at point x^n defined by $G_{0,n,x^n} \triangleq \{y^n \in \mathcal{Y}_{0,n} : (x^n, y^n) \in G_{0,n}\}$ and \otimes denotes the convolution.

(P2): The marginal measure $\nu_{0,n} \in \mathcal{M}_1(\mathcal{Y}_{0,n})$:

$$\begin{aligned} \nu_{0,n}(F_{0,n}) &\triangleq P_{0,n}(\mathcal{X}_{0,n} \times F_{0,n}), \quad F_{0,n} \in \mathcal{B}(\mathcal{Y}_{0,n}) \\ &= \int_{\mathcal{X}_{0,n}} q_{0,n}((\mathcal{X}_{0,n} \times F_{0,n})_{x^n}; x^n) \mu_{0,n}(dx^n) = \int_{\mathcal{X}_{0,n}} q_{0,n}(F_{0,n}; x^n) \mu_{0,n}(dx^n). \end{aligned}$$

(P3): The product measure $\pi_{0,n} : \mathcal{B}(\mathcal{X}_{0,n}) \times \mathcal{B}(\mathcal{Y}_{0,n}) \mapsto [0, 1]$ of $\mu_{0,n} \in \mathcal{M}_1(\mathcal{X}_{0,n})$ and $\nu_{0,n} \in \mathcal{M}_1(\mathcal{Y}_{0,n})$ for $G_{0,n} \in \mathcal{B}(\mathcal{X}_{0,n}) \times \mathcal{B}(\mathcal{Y}_{0,n})$:

$$\pi_{0,n}(G_{0,n}) \triangleq (\mu_{0,n} \times \nu_{0,n})(G_{0,n}) = \int_{\mathcal{X}_{0,n}} \nu_{0,n}(G_{0,n,x^n}) \mu_{0,n}(dx^n).$$

The precise definition of mutual information between two sequences of Random Variables X^n and Y^n , denoted $I(X^n; Y^n)$ is defined via the Kullback-Leibler distance (or relative entropy) between the joint probability distribution of (X^n, Y^n) and the product of its marginal probability distributions of X^n and Y^n , using the Radon-Nikodym derivative as follows.

Definition II.4. [22] *Given a measurable space $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$, the relative entropy between two probability measures $P, Q \in \mathcal{M}_1(\mathcal{X})$ is defined by*

$$\mathbb{D}(P||Q) \triangleq \begin{cases} \int_{\mathcal{X}} \log \left(\frac{dP}{dQ} \right) dP = \int_{\mathcal{X}} \log \left(\frac{dP}{dQ} \right) \frac{dP}{dQ} dQ & \text{if } P \ll Q \\ +\infty & \text{otherwise} \end{cases}$$

where $\frac{dP}{dQ}$ denotes the Radon-Nikodym derivative (density) of P with respect to Q , and $P \ll Q$ denotes absolute continuity of Q with respect to P .

Hence, by the construction of probability measures (P1)-(P3), and the chain rule of relative entropy [20], the following equivalent definitions of mutual information are obtained.

$$I(X^n; Y^n) \triangleq \mathbb{D}(P_{0,n} || \pi_{0,n}) \tag{II.13}$$

$$= \int_{\mathcal{X}_{0,n} \times \mathcal{Y}_{0,n}} \log \left(\frac{d(\mu_{0,n} \otimes q_{0,n})}{d(\mu_{0,n} \times \nu_{0,n})} \right) d(\mu_{0,n} \otimes q_{0,n}) \tag{II.14}$$

$$= \int_{\mathcal{X}_{0,n} \times \mathcal{Y}_{0,n}} \log \left(\frac{q_{0,n}(dy^n; x^n)}{\nu_{0,n}(dy^n)} \right) q_{0,n}(dy^n; dx^n) \otimes \mu_{0,n}(dx^n) \tag{II.15}$$

$$\begin{aligned} &= \int_{\mathcal{X}_{0,n}} \mathbb{D}(q_{0,n}(\cdot; x^n) || \nu_{0,n}(\cdot)) \mu_{0,n}(dx^n) \\ &\equiv \mathbb{I}(\mu_{0,n}, q_{0,n}). \end{aligned} \tag{II.16}$$

Note that (II.16) states that mutual information is expressed as a functional of $\{\mu_{0,n}, q_{0,n}\}$ and it is denoted by $\mathbb{I}(\mu_{0,n}, q_{0,n})$. Note also that $\mu_{0,n} \otimes q_{0,n} \ll \mu_{0,n} \times \nu_{0,n}$ if and only if $q(\cdot; x^n) \ll \nu_{0,n}(\cdot)$, $\mu_{0,n}$ -a.s., which is used to establish that (II.14) is equivalent to (II.15). Necessary and sufficient conditions for existence of a Radon-Nikodym derivative for finitely additive measures can be found in [23].

Next, the classical RDF [3] is introduced, since the definition of nonanticipative RDF will be based on the classical definition by imposing the nonanticipative constraint (I.4) (or Definition II.2-3).

Definition II.5. [3](**Classical Rate Distortion Function**) *Let $d_{0,n} : \mathcal{X}_{0,n} \times \mathcal{Y}_{0,n} \rightarrow [0, \infty]$, be an $\mathcal{B}(\mathcal{X}_{0,n}) \times \mathcal{B}(\mathcal{Y}_{0,n})$ -measurable distortion function, and let $Q_{0,n}(D) \subset \mathcal{Q}(\mathcal{Y}_{0,n}; \mathcal{X}_{0,n})$ (assuming is nonempty) denotes the average distortion or fidelity constraint defined by*

$$Q_{0,n}(D) \triangleq \left\{ q_{0,n} \in \mathcal{Q}(\mathcal{Y}_{0,n}; \mathcal{X}_{0,n}) : \int_{\mathcal{X}_{0,n} \times \mathcal{Y}_{0,n}} d_{0,n}(x^n, y^n) q_{0,n}(dy^n; x^n) \otimes \mu_{0,n}(dx^n) \leq D \right\} \quad (\text{II.17})$$

for $D \geq 0$. The classical RDF associated with the anticipative kernel $q_{0,n} \in \mathcal{Q}(\mathcal{Y}_{0,n}; \mathcal{X}_{0,n})$ is defined by

$$R_{0,n}(D) \triangleq \inf_{q_{0,n} \in Q_{0,n}(D)} \mathbb{I}(\mu_{0,n}, q_{0,n}). \quad (\text{II.18})$$

Existence in (II.18) is shown by assuming $d_{0,n}(x^n, \cdot)$ is bounded continuous on $\mathcal{Y}_{0,n}$ while $\mathcal{Y}_{0,n}$ is compact, using weak-convergence of probability measures in [24], and for more general conditions $d_{0,n}(x^n, \cdot)$ which is only continuous on $\mathcal{Y}_{0,n}$ using weak*-convergence of measures on Polish spaces [25].

Unfortunately, for general sources and distortion function $d_{0,n}(x^n, y^n)$, the optimal reconstruction $q_{0,n}^*(dy^n; x^n) = \otimes_{i=0}^n q_i^*(dy_i; y^{i-1}, x^n)$ is anticipative, and hence the link to filtering theory cannot be established due to dependence of y_i on (y^{i-1}, x^i) and on future symbols (x_{i+1}, \dots, x_n) . This raises the question whether the classical RDF can be reformulated so that the optimal reconstruction kernel is nonanticipative. Before the definition of nonanticipative RDF we introduced a Lemma which gives insight into how classical and nonanticipative RDF are related.

The next lemma relates nonanticipative convolution reconstruction kernels and conditional independence.

Lemma II.6. *The following are equivalent for each $n \in \mathbb{N}$.*

- 1) $q_{0,n}(dy^n; x^n) = \vec{q}_{0,n}(dy^n; x^n)$ -a.s., (see Definition II.2-3).
- 2) For each $i = 0, 1, \dots, n-1$, $Y_i \leftrightarrow (X^i, Y^{i-1}) \leftrightarrow (X_{i+1}, X_{i+2}, \dots, X_n)$, forms a MC.
- 3) For each $i = 0, 1, \dots, n-1$, $Y^i \leftrightarrow X^i \leftrightarrow X_{i+1}$ forms a MC.

Moreover, $X_{i+1}^n \leftrightarrow X^i \leftrightarrow Y^i$, forms a MC for each $i = 0, 1, \dots, n-1$, implies any of the statements 1), 2), 3).

Proof. This is straight forward hence the derivation is omitted. \square

According to Lemma II.6–1), for a restricted nonanticipative stochastic kernel the mutual information becomes

$$\begin{aligned} I(X^n; Y^n) &= \int_{\mathcal{X}_{0,n} \times \mathcal{Y}_{0,n}} \log \left(\frac{\vec{q}_{0,n}(dy^n; x^n)}{\nu_{0,n}(dy^n)} \right) \vec{q}_{0,n}(dy^n; dx^n) \otimes \mu_{0,n}(dx^n) \\ &\equiv \mathbb{I}(\mu_{0,n}, \vec{q}_{0,n}) \end{aligned} \quad (\text{II.19})$$

where (II.19) states that $I(X^n; Y^n)$ is a functional of $\{\mu_{0,n}, \vec{q}_{0,n}\}$. Hence, nonanticipative RDF is defined by optimizing $\mathbb{I}(\mu_{0,n}, q_{0,n})$ over $q_{0,n} \in Q_{0,n}(D)$ subject to the realizability constraint $q_{0,n}(dy^n; x^n) = \vec{q}_{0,n}(dy^n; x^n)$ - a.s., which satisfies a distortion constraint.

Definition II.7. (Nonanticipative Rate Distortion Function) Suppose $d_{0,n}(x^n, y^n) \triangleq \sum_{i=0}^n \rho_{0,i}(x^i, y^i)$, where $\rho_{0,i} : \mathcal{X}_{0,i} \times \mathcal{Y}_{0,i} \rightarrow [0, \infty]$, is a sequence of $\mathcal{B}(\mathcal{X}_{0,i}) \times \mathcal{B}(\mathcal{Y}_{0,i})$ -measurable distortion functions, for $i = 0, 1, \dots, n$, and let $\vec{Q}_{0,n}(D)$ (assuming is nonempty) denotes the average distortion or fidelity constraint defined by

$$\vec{Q}_{0,n}(D) \triangleq Q_{0,n}(D) \cap \left\{ q_{0,n} \in \mathcal{Q}(\mathcal{Y}_{0,n}; \mathcal{X}_{0,n}) : q_{0,n}(dy^n; x^n) = \vec{q}_{0,n}(dy^n; x^n) - \text{a.s.} \right\} \quad (\text{II.20})$$

The nonanticipative RDF associated with the restricted nonanticipative stochastic kernel is defined by

$$R_{0,n}^{na}(D) \triangleq \inf_{q_{0,n} \in \vec{Q}_{0,n}(D)} \mathbb{I}(\mu_{0,n}, q_{0,n}). \quad (\text{II.21})$$

Thus, $R_{0,n}^{na}(D)$ is characterized by minimizing mutual information or equivalently $\mathbb{I}(\mu_{0,n}, q_{0,n})$ over the $Q_{0,n}(D)$ and the nonanticipative constraint (I.4). In the work of [12], nonanticipative RDF is called ϵ -entropy and nonanticipation is defined via $X_{i+1}^n \leftrightarrow X^i \leftrightarrow Y^i$, forms a MC for each $i = 0, 1, \dots, n-1$, which implies $q_{0,n}(dy^n; x^n) = \vec{q}_{0,n}(dy^n; x^n)$. Clearly, Gorbunov and

Pinsker [12] nonanticipative RDF which imposes the constraint $X_{i+1}^n \leftrightarrow X^i \leftrightarrow Y^i$ forms a MC for each $i = 0, 1, \dots, n-1$, implies $P_{X_{i+1}|X^i, Y^i}(dx_{i+1}|x^i, y^i) = P_{X_{i+1}|X^i}(dx_{i+1}|x^i) - a.s., i = 0, 1, \dots, n-1$, and hence, it does not allow the generalizations discussed in Remark II.3.

III. EXISTENCE OF OPTIMAL RECONSTRUCTION KERNEL

In this section, appropriate topologies and function spaces are introduced and existence of the minimizing nonanticipative product kernel in (II.21) is proved. The construction of spaces is based on [25].

A. Abstract Spaces

Let $BC(\mathcal{Y}_{0,n})$ denote the vector space of bounded continuous real valued functions defined on the Polish space $\mathcal{Y}_{0,n}$. Furnished with the sup norm topology, this is a Banach space. Denote by $L_1(\mu_{0,n}, BC(\mathcal{Y}_{0,n}))$ the space of all $\mu_{0,n}$ -integrable functions defined on $\mathcal{X}_{0,n}$ with values in $BC(\mathcal{Y}_{0,n})$, so that for each $\phi \in L_1(\mu_{0,n}, BC(\mathcal{Y}_{0,n}))$ its norm is defined by

$$\|\phi\|_{\mu_{0,n}} \triangleq \int_{\mathcal{X}_{0,n}} \|\phi(x^n, \cdot)\|_{BC(\mathcal{Y}_{0,n})} \mu_{0,n}(dx^n) < \infty.$$

The norm topology $\|\phi\|_{\mu_{0,n}}$ makes $L_1(\mu_{0,n}, BC(\mathcal{Y}_{0,n}))$ a Banach space. The topological dual of $BC(\mathcal{Y}_{0,n})$ denoted by $(BC(\mathcal{Y}_{0,n}))^*$ is isometrically isomorphic to the Banach space of finitely additive regular bounded signed measures on $\mathcal{Y}_{0,n}$ [26], denoted by $M_{rba}(\mathcal{Y}_{0,n})$. Let $\Pi_{rba}(\mathcal{Y}_{0,n}) \subset M_{rba}(\mathcal{Y}_{0,n})$ denote the set of regular bounded finitely additive probability measures on $\mathcal{Y}_{0,n}$. Clearly if $\mathcal{Y}_{0,n}$ is compact, then $(BC(\mathcal{Y}_{0,n}))^*$ will be isometrically isomorphic to the space of countably additive signed measures, as in [24]. It follows from the theory of “lifting” [27] that the dual of the space $L_1(\mu_{0,n}, BC(\mathcal{Y}_{0,n}))$ is $L_\infty^w(\mu_{0,n}, M_{rba}(\mathcal{Y}_{0,n}))$, denoting the space of all $M_{rba}(\mathcal{Y}_{0,n})$ valued functions $\{q\}$ which are weak*-measurable in the sense that for each $\phi \in BC(\mathcal{Y}_{0,n})$, $x^n \rightarrow q_{x^n}(\phi) \triangleq \int_{\mathcal{Y}_{0,n}} \phi(y^n) q(dy^n; x^n)$ is $\mu_{0,n}$ -measurable and $\mu_{0,n}$ -essentially bounded.

B. Weak*-Compactness and Existence

Next, we prepare to prove existence of solution to $R_{0,n}^{na}(D)$. Define an admissible set of stochastic kernels associated with classical rate distortion function by

$$Q_{ad} \triangleq L_\infty^w(\mu_{0,n}, \Pi_{rba}(\mathcal{Y}_{0,n})) \subset L_\infty^w(\mu_{0,n}, M_{rba}(\mathcal{Y}_{0,n})).$$

Clearly, Q_{ad} is a unit sphere in $L_\infty^w(\mu_{0,n}, M_{rba}(\mathcal{Y}_{0,n}))$. For each $\phi \in L_1(\mu_{0,n}, BC(\mathcal{Y}_{0,n}))$ we can define a linear functional on $L_\infty^w(\mu_{0,n}, M_{rba}(\mathcal{Y}_{0,n}))$ by

$$\ell_\phi(q_{0,n}) \triangleq \int_{\mathcal{X}_{0,n}} \left(\int_{\mathcal{Y}_{0,n}} \phi(x^n, y^n) q_{0,n}(dy^n; x^n) \right) \mu_{0,n}(dx^n).$$

This is a bounded, linear and weak*-continuous functional on $L_\infty^w(\mu_{0,n}, M_{rba}(\mathcal{Y}_{0,n}))$ as it is shown below.

$$\begin{aligned} |\ell_\phi(q_{0,n})| &= \left| \int_{\mathcal{X}_{0,n}} \left(\int_{\mathcal{Y}_{0,n}} \phi(x^n, y^n) q_{0,n}(dy^n; x^n) \right) \mu_{0,n}(dx^n) \right| \\ &\leq \int_{\mathcal{X}_{0,n}} \left| \left(\int_{\mathcal{Y}_{0,n}} \phi(x^n, y^n) q_{0,n}(dy^n; x^n) \right) \right| \mu_{0,n}(dx^n) \\ &\leq \int_{\mathcal{X}_{0,n}} \|\phi(x^n, \cdot)\|_{BC(\mathcal{Y}_{0,n})} \|q_{0,n}(\cdot; x^n)\|_{TV} \mu_{0,n}(dx^n) \\ &\leq \int_{\mathcal{X}_{0,n}} \|\phi(x^n, \cdot)\|_{BC(\mathcal{Y}_{0,n})} \mu_{0,n}(dx^n) \\ &= \|\phi\|_{L_1(\mu_{0,n}, BC(\mathcal{Y}_{0,n}))} < \infty. \end{aligned}$$

So given $\phi \in L_1(\mu_{0,n}, BC(\mathcal{Y}_{0,n}))$, there exists a $c_\phi < \infty$ such that $\|\ell_\phi\| < c_\phi$. Therefore, ℓ_ϕ is a bounded, linear functional on $L_\infty^w(\mu_{0,n}, \Pi_{rba}(\mathcal{Y}_{0,n}))$ and hence on $L_\infty^w(\mu_{0,n}, M_{rba}(\mathcal{Y}_{0,n}))$. Thus, it is continuous in the weak*-sense.

For $d_{0,n} : \mathcal{X}_{0,n} \times \mathcal{Y}_{0,n} \rightarrow [0, \infty)$ measurable and $d_{0,n} \in L_1(\mu_{0,n}, BC(\mathcal{Y}_{0,n}))$ the distortion constraint set of the classical RDF is given by

$$Q_{0,n}(D) \triangleq \{q \in Q_{ad} : \ell_{d_{0,n}}(q_{0,n}) \leq D\}.$$

The next result is shown in [25]; it utilizes the Alaoglu's theorem [26], which states that a closed and bounded subset of a weak*-compact set is weak*-compact. These will be used to establish existence of minimizer in \vec{Q}_{ad} for the nonanticipative RDF $R_{0,n}^{na}(D)$.

Lemma III.1. [25] For $d_{0,n} \in L_1(\mu_{0,n}, BC(\mathcal{Y}_{0,n}))$, the set $Q_{0,n}(D)$ is bounded and weak*-closed subset of Q_{ad} (hence weak*-compact).

Now we prepare to consider the problem stated in Definition II.7. First, we show weak*-compactness of \vec{Q}_{ad} defined as a subset of Q_{ad} as follows.

$$\vec{Q}_{ad} = \left\{ q_{0,n} \in Q_{ad} : q_{0,n}(dy^n; x^n) = \vec{q}_{0,n}(dy^n; x^n) - a.s. \right\}.$$

The average distortion function for the nonanticipative RDF is defined by

$$\begin{aligned}\vec{Q}_{0,n}(D) &\triangleq \left\{ q_{0,n} \in Q_{ad} : \ell_{d_{0,n}}(q_{0,n}) \triangleq \int_{\mathcal{X}_{0,n}} \left(\int_{\mathcal{Y}_{0,n}} d_{0,n}(x^n, y^n) q_{0,n}(dy^n; x^n) \right) \right. \\ &\quad \left. \otimes \mu_{0,n}(dx^n) \leq D \right\} \cap \vec{Q}_{ad} \\ &= \left\{ q_{0,n} \in \vec{Q}_{ad} : \ell_{d_{0,n}}(q_{0,n}) \triangleq \int_{\mathcal{X}_{0,n}} \left(\int_{\mathcal{Y}_{0,n}} d_{0,n}(x^n, y^n) q_{0,n}(dy^n; x^n) \right) \right. \\ &\quad \left. \otimes \mu_{0,n}(dx^n) \leq D \right\}, \quad D \geq 0.\end{aligned}$$

Since we are interested in proving existence of nonanticipative RDF of Definition II.7, we shall first show that \vec{Q}_{ad} is weak*-closed, and then utilize Lemma III.1 to establish weak*-compactness for \vec{Q}_{ad} and then weak*-compactness of $\vec{Q}_{0,n}(D)$.

Lemma III.2. *Let $\mathcal{X}_{0,n}$ and $\mathcal{Y}_{0,n}$ be Polish spaces and introduce the net $\{q_i^\alpha(dy_i; y^{i-1}, x^i)\}$, where $\alpha \in (\mathcal{D}, \succeq)$, and $q_i^\alpha \in \mathcal{Q}(\mathcal{Y}_i; \mathcal{Y}_{0,i-1}, \mathcal{X}_{0,i})$. Assume*

- (a) $q_i^\alpha(\cdot; y^{i-1}, x^i) \xrightarrow{w^*} q_i^0(\cdot; y^{i-1}, x^i)$ for $i = 1, \dots, n$;
- (b) for all $h_i(\cdot, \cdot) \in L_1(\mu_i, BC(\mathcal{Y}_i))$ the function

$$(x^i, y^{i-1}) \in \mathcal{X}_{0,i} \times \mathcal{Y}_{0,i-1} \mapsto \int_{\mathcal{X}_i} \int_{\mathcal{Y}_i} h_i(y) q_i(dy; y^{i-1}, x^i) \mu_i(dx_i; x^{i-1})$$

is $L_1(\mu_{0,i-1}, BC(\mathcal{Y}_{0,i-1}))$ for $i = 0, 1, \dots, n$;

- (c) for all $h_i(\cdot, \cdot) \in L_1(\mu_i, BC(\mathcal{Y}_i))$ and $\forall \epsilon > 0$ there exists $\alpha \succ \alpha_\epsilon$ such that

$$\begin{aligned}&\int_{\mathcal{X}_i} \sup_{y^{i-1} \in \mathcal{Y}_{0,i-1}} \left| \int_{\mathcal{Y}_i} h_i(x_i, y_i) q_i^\alpha(dy_i; y^{i-1}, x^i) \right. \\ &\quad \left. - \int_{\mathcal{Y}_i} h_i(x_i, y_i) q_i^0(dy_i; y^{i-1}, x^i) \right| \mu_i(dx_i; x^{i-1}) < \epsilon, \quad \forall x^{i-1} \in \mathcal{X}_{0,i-1}.\end{aligned}$$

Then the convolution of stochastic kernels converges in weak*-sense as follows.

$$\vec{q}_{0,n}^\alpha \xrightarrow{w^*} \vec{q}_{0,n}^0 \quad (\text{III.22})$$

e.g, the set \vec{Q}_{ad} is weak*-closed.

Proof. See Appendix. □

Next, we utilize the weak*-compactness of \vec{Q}_{ad} to show that $\vec{Q}_{0,n}(D)$ is also weak*-compact.

Remark III.3. *There are certain important cases in which $d_{0,n}$ may not be bounded. This is the case when $d_{0,n}$ is a metric of a linear metric space. The next theorem is crucial in showing the*

weak*-closedness property of $\vec{Q}_{0,n}(D)$ to those distortion functions $d_{0,n}$ which are not necessarily bounded, since they are measurable functions from the class $d_{0,n} \in L_1(\mu_{0,n}, BC(\mathcal{Y}_{0,n}))$.

Theorem III.4. *Let $\mathcal{X}_{0,n}, \mathcal{Y}_{0,n}$ be two Polish spaces and $d_{0,n} : \mathcal{X}_{0,n} \times \mathcal{Y}_{0,n} \mapsto [0, \infty]$, a measurable, nonnegative, extended real valued function, such that for a fixed $x^n \in \mathcal{X}_{0,n}$, $y^n \rightarrow d(x^n, \cdot)$ is continuous on $\mathcal{Y}_{0,n}$, for $\mu_{0,n}$ -almost all $x^n \in \mathcal{X}_{0,n}$ and suppose the conditions of Lemma III.2 hold. For any $D \in [0, \infty)$, the set $\vec{Q}_{0,n}(D)$ is a weak*-closed subset of \vec{Q}_{ad} and hence weak*-compact.*

Proof. Let $\{\vec{q}_{0,n}^\alpha\} \in \vec{Q}_{0,n}(D) \subset \vec{Q}_{ad}$ be a net. Since \vec{Q}_{ad} is weak*-compact, there exists a subnet of the net $\{\vec{q}_{0,n}^\alpha\}$, relabelled as the original net, and an element $\vec{q}_{0,n}^0 \in \vec{Q}_{ad}$ such that $\vec{q}_{0,n}^\alpha \xrightarrow{w^*} \vec{q}_{0,n}^0$ ³. We must show that $\vec{q}_{0,n}^0 \in \vec{Q}_{0,n}(D)$. Considering the sequence $\{d_{0,n}^k \triangleq d_{0,n} \wedge k, k \in N\}$, which are bounded, measurable functions (continuous in the second argument), it follows from the weak*-convergence of the sequence $\{\vec{q}_{0,n}^\alpha\}$ to $\vec{q}_{0,n}^0$ that

$$\begin{aligned} & \int_{\mathcal{X}_{0,n}} \left(\int_{\mathcal{Y}_{0,n}} d_{0,n}^k(x^n, y^n) \vec{q}_{0,n}^0(dy^n; x^n) \right) \mu_{0,n}(dx^n) \\ &= \lim_{\alpha} \int_{\mathcal{X}_{0,n}} \left(\int_{\mathcal{Y}_{0,n}} d_{0,n}^k(x^n, y^n) \vec{q}_{0,n}^\alpha(dy^n; x^n) \right) \mu_{0,n}(dx^n) \end{aligned} \quad (\text{III.23})$$

for each $k \in N$. Since $d_{0,n}$ is non-negative and $d_{0,n}^k \uparrow d_{0,n}$ as $k \rightarrow \infty$ and $\vec{q}_{0,n}^\alpha \in \vec{Q}_{0,n}(D)$, we have

$$\begin{aligned} & \int_{\mathcal{X}_{0,n}} \left(\int_{\mathcal{Y}_{0,n}} d_{0,n}^k(x^n, y^n) \vec{q}_{0,n}^0(dy^n; x^n) \right) \mu_{0,n}(dx^n) \\ &= \lim_{\alpha} \int_{\mathcal{X}_{0,n}} \left(\int_{\mathcal{Y}_{0,n}} d_{0,n}^k(x^n, y^n) \vec{q}_{0,n}^\alpha(dy^n; x^n) \right) \mu_{0,n}(dx^n) \\ &\leq \lim_{\alpha} \int_{\mathcal{X}_{0,n}} \left(\int_{\mathcal{Y}_{0,n}} d_{0,n}(x^n, y^n) \vec{q}_{0,n}^\alpha(dy^n; x^n) \right) \mu_{0,n}(dx^n) \leq D \end{aligned}$$

which is valid for all $k \in N$. Since $d_{0,n}^k \uparrow d_{0,n}$ and they are non-negative, it follows from Lebesgue's monotone convergence theorem and non-negativity of stochastic kernels that

$$\int_{\mathcal{X}_{0,n}} \left(\int_{\mathcal{Y}_{0,n}} d_{0,n}(x^n, y^n) \vec{q}_{0,n}^0(dy^n; x^n) \right) \mu_{0,n}(dx^n) \leq D.$$

³i.e. $\left| \int_{\mathcal{X}_{0,n}} \int_{\mathcal{Y}_{0,n}} \phi(x^n, y^n) \vec{q}_{0,n}^\alpha(dy^n; x^n) \otimes \mu_{0,n}(dx^n) - \int_{\mathcal{X}_{0,n}} \int_{\mathcal{Y}_{0,n}} \phi(x^n, y^n) \vec{q}_{0,n}^0(dy^n; x^n) \otimes \mu_{0,n}(dx^n) \right| \rightarrow 0$ for any $\phi \in L_1(\mu_{0,n}; BC(\mathcal{Y}_{0,n}))$.

This shows that the weak*-limit $\vec{q}_{0,n}^0 \in \vec{Q}_{0,n}(D)$ and hence we have proved that the set $\vec{Q}_{0,n}(D)$ is a weak*-closed subset of \vec{Q}_{ad} . By Alaoglu's theorem [26] being a weak*-closed subset of a weak*-compact set, it is weak*-compact. \square

Based on Theorem III.4 and lower semicontinuity of relative entropy, we show existence of the optimal reconstruction conditional distribution for nonanticipative RDF.

Theorem III.5. (Existence) *Under the conditions of Theorem III.4, $R_{0,n}^{na}(D)$ has a minimum.*

Proof. This follows from Theorem III.4 provided lower semicontinuity of $\mathbb{I}(\mu_{0,n}, \cdot)$ on \vec{Q}_{ad} is established. First we prove that $\vec{q}_{0,n} \rightarrow \mathbb{I}(\mu_{0,n}, \cdot)$ is weak*-lower semicontinuous. Let $\{\vec{q}_{0,n}^\alpha\}$ be a net from \vec{Q}_{ad} and suppose it is weak*-convergent to $\vec{q}_{0,n}^0$. Define the net $P_{0,n}^\alpha \in \Pi_{rba}(\mathcal{X}_{0,n} \times \mathcal{Y}_{0,n})$ given by the convolution product $P_{0,n}^\alpha \equiv \mu_{0,n}(dx^n) \otimes \vec{q}_{0,n}^\alpha(dy^n; x^n)$. Take any $\varphi(\cdot) \in BC(\mathcal{X}_{0,n} \times \mathcal{Y}_{0,n})$ and consider the expression

$$\int_{\mathcal{X}_{0,n} \times \mathcal{Y}_{0,n}} \varphi_{0,n}(x^n, y^n) P_{0,n}^\alpha(dx^n, dy^n) \equiv \int_{\mathcal{X}_{0,n} \times \mathcal{Y}_{0,n}} \varphi_{0,n}(x^n, y^n) \vec{q}_{0,n}^\alpha(dy^n; x^n) \otimes \mu_{0,n}(dx^n).$$

Since $\vec{q}_{0,n}^\alpha \xrightarrow{w^*} \vec{q}_{0,n}^0$ in $L_\infty^w(\mu_{0,n}, \Pi_{rba}(\mathcal{Y}_{0,n}))$, it is clear from the above expression that

$$P_{0,n}^\alpha \xrightarrow{w^*} P_{0,n}^0 \equiv \mu_{0,n} \otimes \vec{q}_{0,n}^0 \quad \text{in } \Pi_{rba}(\mathcal{X}_{0,n} \times \mathcal{Y}_{0,n}). \quad (\text{III.24})$$

Similarly one can easily verify that the net of the product measures $\{\pi_{0,n}^\alpha\}$ converges to the product measure $\pi_{0,n}^0$,

$$\pi_{0,n}^\alpha \equiv \nu_{0,n}^\alpha \times \mu_{0,n} \xrightarrow{w^*} \nu_{0,n}^0 \times \mu_{0,n} \equiv \pi_{0,n}^0$$

where $\{\nu_{0,n}^\alpha\}$ are the marginals of $\{P_{0,n}^\alpha\}$ on $\mathcal{Y}_{0,n}$ and $\nu_{0,n}^0$ is its weak*-limit. Now we use the lower semicontinuity property of relative entropy [20, Lemma 1.4.3, p. 36]. Following [20] it is verified that the same procedure holds true not only for countably additive measures but also for finitely additive ones. Using this fact we conclude that

$$\mathbb{D}(P_{0,n} || \pi_{0,n}) \leq \liminf_{\alpha \rightarrow \infty} \mathbb{D}(P_{0,n}^\alpha || \pi_{0,n}^\alpha).$$

By (II.13), this is equivalent to

$$\mathbb{I}(\mu_{0,n}, q_{0,n}) \leq \liminf_{\alpha \rightarrow \infty} \mathbb{I}(\mu_{0,n}, q_{0,n}^\alpha). \quad (\text{III.25})$$

This proves weak*-lower semicontinuity of $\mathbb{I}(\mu_{0,n}, \cdot)$ on \vec{Q}_{ad} . We have already observed in Theorem III.4 that the set $\vec{Q}_{0,n}(D)$ is weak*-compact, and we have just seen that $\mathbb{I}(\mu_{0,n}, \cdot)$ is

weak*-lower semicontinuous. Hence $\mathbb{I}(\mu_{0,n}, \cdot)$ attains its infimum on $\vec{\mathcal{Q}}_{0,n}(D)$. So there exists a $\vec{q}_{0,n}^* \in \vec{\mathcal{Q}}_{0,n}(D)$ such that $R_{0,n}^{na}(D) = \mathbb{I}(\mu_{0,n}, \vec{q}_{0,n}^*)$. \square

IV. NECESSARY CONDITIONS OF OPTIMALITY FOR NONANTICIPATIVE RDF

In this section the form of the optimal nonanticipative convolution reconstruction kernels is derived under a stationarity assumption. The method is based on calculus of variations on the space of measures [28].

Assumption IV.1. *The family of measures $\vec{q}_{0,n}(dy^n; x^n)$ defined in (II.9), is the convolution of stationary conditional distributions.*

Assumption IV.1 holds for stationary process $\{(X_i, Y_i) : i \in \mathbb{N}\}$ and single letter distortion $d_{0,n}(x^n, y^n) \equiv \sum_{i=0}^n \rho(x_i, y_i)$. It also holds for distortion defined by $\rho(T^i x^n, T^i y^n)$, for which $T^i x^n = \tilde{x}^n$ is the i^{th} shift operator on the input sequence x^n , where $\tilde{x}_k = x_{k+i}$ (similarly for $T^i y^n$), and $\sum_{i=0}^n \rho(T^i x^n, T^i y^n)$ depends only on the components of (x^n, y^n) [29]. Utilizing Assumption IV.1, which holds for stationary processes and a single letter distortion function, the Gateaux differential of $\mathbb{I}(\mu_{0,n}, \vec{q}_{0,n})$ is taken at $\vec{q}_{0,n}^*$ in the direction of $\vec{q}_{0,n} - \vec{q}_{0,n}^*$, via the definition $\vec{q}_{0,n}^\epsilon \triangleq \vec{q}_{0,n} + \epsilon(\vec{q}_{0,n} - \vec{q}_{0,n}^*)$, $\epsilon \in [0, 1]$, since under the stationarity assumption, the functionals $\{q_i(dy_i; y^{i-1}, x^i) \in \mathcal{Q}(\mathcal{Y}_i; \mathcal{Y}_{0,i-1} \times \mathcal{X}_{0,i}) : i = 0, 1, \dots, n\}$ are identical.

Theorem IV.2. *Suppose $\mathbb{I}_{\mu_{0,n}}(\vec{q}_{0,n}) \triangleq \mathbb{I}(\mu_{0,n}, \vec{q}_{0,n})$ is well defined for every $\vec{q}_{0,n} \in L_\infty^w(\mu_{0,n}, \Pi_{rba}(\mathcal{Y}_{0,n}))$ possibly taking values from the set $[0, \infty]$. Then $\vec{q}_{0,n} \rightarrow \mathbb{I}_{\mu_{0,n}}(\vec{q}_{0,n})$ is Gateaux differentiable at every point in $L_\infty^w(\mu_{0,n}, \Pi_{rba}(\mathcal{Y}_{0,n}))$, and the Gateaux derivative at the point $\vec{q}_{0,n}^*$ in the direction $\vec{q}_{0,n} - \vec{q}_{0,n}^*$ is given by*

$$\delta \mathbb{I}_{\mu_{0,n}}(\vec{q}_{0,n}^*, \vec{q}_{0,n} - \vec{q}_{0,n}^*) = \int_{\mathcal{X}_{0,n}} \int_{\mathcal{Y}_{0,n}} \log \left(\frac{\vec{q}_{0,n}^*(dy^n; x^n)}{\nu_{0,n}^*(dy^n)} \right) (\vec{q}_{0,n} - \vec{q}_{0,n}^*)(dy^n; x^n) \otimes \mu_{0,n}(dx^n)$$

where $\nu_{0,n}^* \in \mathcal{M}_1(\mathcal{Y}_{0,n})$ is the marginal measure corresponding to $\vec{q}_{0,n}^* \otimes \mu_{0,n} \in \mathcal{M}_1(\mathcal{Y}_{0,n} \times \mathcal{X}_{0,n})$.

Proof. The proof, although lengthy, it is similar to the one in [25], hence it is omitted. \square

The constrained problem defined by (II.21) can be reformulated using Lagrange multipliers. The equivalence of constrained and unconstrained problems is established in the following theorem.

Theorem IV.3. Suppose $d_{0,n}(x^n, y^n) \triangleq \sum_{i=0}^n \rho(T^i x^n, T^i y^n)$, where $d_{0,n} : \mathcal{X}_{0,n} \times \mathcal{Y}_{0,n} \rightarrow \bar{R}_0 \equiv [0, \infty]$ is continuous in the second argument and the set $\Gamma \equiv \{(x^n, y^n) \in \mathcal{X}_{0,n} \times \mathcal{Y}_{0,n} : d_{0,n}(x^n, y^n) < D\}$ is nonempty. Then the constrained problem as stated in Theorem III.5, is equivalent to an unconstrained problem stated below.

$$\begin{aligned} \inf_{\vec{q}_{0,n} \in \vec{Q}_{0,n}(D)} \mathbb{I}(\mu_{0,n}, \vec{q}_{0,n}) &= \max_{s \leq 0} \inf_{\vec{q}_{0,n}} \{ \mathbb{I}(\mu_{0,n}, \vec{q}_{0,n}) - sG(\vec{q}_{0,n}) \}, \quad G(\vec{q}_{0,n}) \triangleq \ell_{d_{0,n}}(\vec{q}_{0,n}) - D \\ &= \max_{s \leq 0} \inf_{\vec{q}_{0,n}} \left\{ \mathbb{I}(\mu_{0,n}, \vec{q}_{0,n}) - s \left(\int_{\mathcal{X}_{0,n}} \int_{\mathcal{Y}_{0,n}} d_{0,n}(x^n, y^n) \right. \right. \\ &\quad \left. \left. \vec{q}_{0,n}(dy^n; x^n) \otimes \mu_{0,n}(dx^n) - D \right) \right\} \end{aligned}$$

where $\vec{q}_{0,n} \equiv \vec{q}_{0,n}(dy^n; x^n) = \otimes_{i=0}^n q_i(dy_i; y^{i-1}, x^i)$ -a.s. Further the infimum occurs on the boundary of the set $\vec{Q}_{0,n}(D)$.

Proof. See Appendix. □

Utilizing Theorem IV.3, we can reformulate the constraint problem as an unconstrained problem, hence we have

$$R_{0,n}^{na}(D) = \sup_{s \leq 0} \inf_{\vec{q}_{0,n}} \left\{ \mathbb{I}(\mu_{0,n}, \vec{q}_{0,n}) - s(\ell_{d_{0,n}}(\vec{q}_{0,n}) - D) \right\}. \quad (\text{IV.26})$$

Note that $\vec{q}_{0,n} \in \mathcal{M}_1(\mathcal{Y}_{0,n})$ are probability measures on $\mathcal{Y}_{0,n}$ therefore, one should introduce another set of Lagrange multipliers.

Moreover, $\vec{q}_{0,n}(dy^n; x^n) = \otimes_{i=0}^n q_i(dy_i; y^{i-1}, x^i)$ is a consistent probability measure on $\mathcal{Y}_{0,n}$, therefore for each $k = 0, 1, \dots, n$, $\int_{\mathcal{Y}_{0,k}} \vec{q}_{0,k}(dy^k; x^k) = 1$. This constraint is expressed via

$$\begin{aligned} &\sum_{i=0}^n \int_{\mathcal{X}_{0,i} \times \mathcal{Y}_{0,i}} \lambda_i(x^i, y^{i-1}) \left(\vec{q}_{0,i}(dy^i; x^i) - 1 \right) \mu_{0,i}(dx^i) \\ &= \sum_{i=0}^n \int_{\mathcal{X}_{0,n} \times \mathcal{Y}_{0,n}} \lambda_i(x^i, y^{i-1}) \left(\vec{q}_{0,n}(dy^n; x^n) - 1 \right) \mu_{0,n}(dx^n) \end{aligned} \quad (\text{IV.27})$$

where $\{\lambda_i(\cdot, \cdot) : i = 0, 1, \dots, n\}$ are Lagrange multipliers.

Utilizing the additional constraint (IV.27) in (IV.26), then we derive the optimal reconstruction kernel for the nonanticipative RDF, $R_{0,n}^{na}(D)$. This is given in the following theorem.

Theorem IV.4. Suppose $d_{0,n}(x^n, y^n) = \sum_{i=0}^n \rho(T^i x^n, T^i y^n)$ and the conditions of Lemma III.2 and Theorem III.4 hold. Then the infimum in (IV.26) is attained at $\vec{q}_{0,n}^* \in L_\infty^w(\mu_{0,n}, \Pi_{rba}(\mathcal{Y}_{0,n}))$

given by⁴

$$\begin{aligned}\vec{q}_{0,n}^*(dy^n; x^n) &= \otimes_{i=0}^n q_i^*(dy_i; y^{i-1}, x^i) - a.s \\ &= \otimes_{i=0}^n \frac{e^{s\rho(T^i x^n, T^i y^n)} \nu_i^*(dy_i; y^{i-1})}{\int_{\mathcal{Y}_i} e^{s\rho(T^i x^n, T^i y^n)} \nu_i^*(dy_i; y^{i-1})}, \quad s \leq 0\end{aligned}\quad (\text{IV.28})$$

and $\nu_i^*(dy_i; y^{i-1}) \in \mathcal{Q}(\mathcal{Y}_i; \mathcal{Y}_{0,i-1})$. The nonanticipative RDF is given by

$$\begin{aligned}R_{0,n}^{na}(D) &= sD - \sum_{i=0}^n \int_{\mathcal{X}_{0,i} \times \mathcal{Y}_{0,i-1}} \log \left(\int_{\mathcal{Y}_i} e^{s\rho(T^i x^n, T^i y^n)} \nu_i^*(dy_i; y^{i-1}) \right) \\ &\quad \times \vec{q}_{0,i-1}^*(dy^{i-1}; x^{i-1}) \otimes \mu_{0,i}(dx^i)\end{aligned}\quad (\text{IV.29})$$

where “ s ” is the optimal value of (IV.26).

If $R_{0,n}^{na}(D) > 0$ then $s < 0$ and

$$\sum_{i=0}^n \int_{\mathcal{X}_{0,i}} \int_{\mathcal{Y}_{0,i}} \rho(T^i x^n, T^i y^n) \vec{q}_{0,i}^*(dy^i; x^i) \otimes \mu_{0,i}(dx^i) = D \quad (\text{IV.30})$$

and s is obtained from the equality condition (IV.30).

Proof. The fully unconstrained problem of (IV.26) is obtained by introducing another set of Lagrange multipliers $\{\lambda_i(\cdot, \cdot) : i = 0, 1, \dots, n\}$ as in (IV.27). Using the pair of Lagrange multipliers $\{s, \lambda \triangleq \{\lambda_i(\cdot, \cdot) : i = 0, 1, \dots, n\}\}$ introduce the extended pay-off functional

$$\begin{aligned}\mathbb{I}_D^{s,\lambda}(\mu_{0,n}, \vec{q}_{0,n}) &\triangleq \mathbb{I}(\mu_{0,n}, \vec{q}_{0,n}) - s \left(\ell_{d_{0,n}}(\vec{q}_{0,n}) - D \right) \\ &\quad + \sum_{i=0}^n \int_{\mathcal{X}_{0,n}} \int_{\mathcal{Y}_{0,n}} \lambda_i(x^i, y^{i-1}) \left(\vec{q}_{0,n}(dy^n; x^n) - 1 \right) \mu_{0,n}(dx^n).\end{aligned}$$

This is a fully unconstrained problem on the vector space $L_\infty^w(\mu_{0,n}, M_{rba}(\mathcal{Y}_{0,n}))$. Utilizing Theorem IV.2, the Gateaux derivative of $\mathbb{I}_D^{s,\lambda}$ on $L_\infty^w(\mu_{0,n}, M_{rba}(\mathcal{Y}_{0,n}))$ at any point $\vec{q}_{0,n}^*$ in the direction $\vec{q}_{0,n} - \vec{q}_{0,n}^*$ is given by

$$\begin{aligned}\delta \mathbb{I}_D^{s,\lambda}(\vec{q}_{0,n}^*; \vec{q}_{0,n} - \vec{q}_{0,n}^*) &= \int_{\mathcal{X}_{0,n} \times \mathcal{Y}_{0,n}} \log \left(\frac{\vec{q}_{0,n}^*(dy^n; x^n)}{\nu_{0,n}^*(dy^n)} \right) (\vec{q}_{0,n} - \vec{q}_{0,n}^*)(dy^n; x^n) \otimes \mu_{0,n}(dx^n) \\ &\quad - s \int_{\mathcal{X}_{0,n} \times \mathcal{Y}_{0,n}} d_{0,n}(x^n, y^n) (\vec{q}_{0,n} - \vec{q}_{0,n}^*)(dy^n; x^n) \otimes \mu_{0,n}(dx^n) \\ &\quad + \sum_{i=0}^n \int_{\mathcal{X}_{0,n} \times \mathcal{Y}_{0,n}} \lambda_i(x^i, y^{i-1}) (\vec{q}_{0,n} - \vec{q}_{0,n}^*)(dy^n; x^n) \otimes \mu_{0,n}(dx^n)\end{aligned}$$

⁴Due to stationarity assumption $\nu_i(\cdot; \cdot) = \nu(\cdot; \cdot)$ and $q_i^*(\cdot; \cdot, \cdot) = q^*(\cdot; \cdot, \cdot)$, $\forall i = 0, 1, \dots, n$.

$$= \int_{\mathcal{X}_{0,n} \times \mathcal{Y}_{0,n}} \log \left(e^{\sum_{i=0}^n (-s\rho(T^i x^n, T^i y^n) + \lambda_i(x^i, y^{i-1}))} \frac{\vec{q}_{0,n}^*(dy^n; x^n)}{\nu_{0,n}^*(dy^n)} \right) \\ (\vec{q}_{0,n} - \vec{q}_{0,n}^*)(dy^n; x^n) \otimes \mu_{0,n}(dx^n), \quad \forall \vec{q}_{0,n} \in L_\infty^w(\mu_{0,n}, M_{rba}(\mathcal{Y}_{0,n})).$$

Since $\mathbb{I}_D^{s,\lambda}(\mu_{0,n}, \vec{q}_{0,n})$ is convex in $\vec{q}_{0,n}$, it follows from the calculus of variations principle that a necessary and sufficient condition for $\vec{q}_{0,n}^*$ to be a minimizer is $\delta \mathbb{I}_D^{s,\lambda}(\vec{q}_{0,n}^*; \vec{q}_{0,n} - \vec{q}_{0,n}^*) = 0, \forall \vec{q}_{0,n} \in L_\infty^w(\mu_{0,n}, M_{rba}(\mathcal{Y}_{0,n}))$. Since the Gateaux derivative must be zero for all $\vec{q}_{0,n} \in L_\infty^w(\mu_{0,n}, M_{rba}(\mathcal{Y}_{0,n}))$ then

$$\frac{\vec{q}_{0,n}^*(dy^n; x^n)}{\nu_{0,n}^*(dy^n)} = e^{\sum_{i=0}^n (s\rho(T^i x^n, T^i y^n) - \lambda_i(x^i, y^{i-1}))} - a.s.$$

Equivalently,

$$\otimes_{i=0}^n \frac{q_i^*(dy_i; y^{i-1}, x^i)}{\nu_i^*(dy_i; y^{i-1})} = \otimes_{i=0}^n e^{(s\rho(T^i x^n, T^i y^n) - \lambda_i(x^i, y^{i-1}))} - a.s.$$

Since $\int_{\mathcal{Y}_i} q_i^*(dy_i; y^{i-1}, x^i) = 1$, then

$$\lambda_i(x^i, y^{i-1}) = \log \int_{\mathcal{Y}_i} e^{s\rho(T^i x^n, T^i y^n)} \nu_i^*(dy_i; y^{i-1}), \quad i = 0, 1, \dots, n.$$

Hence,

$$\begin{aligned} \vec{q}_{0,n}^*(dy^n; x^n) &= \otimes_{i=0}^n q_i^*(dy_i; y^{i-1}, x^i) - a.s \\ &= \otimes_{i=0}^n \frac{e^{s\rho(T^i x^n, T^i y^n)} \nu_i^*(dy_i; y^{i-1})}{\int_{\mathcal{Y}_i} e^{s\rho(T^i x^n, T^i y^n)} \nu_i^*(dy_i; y^{i-1})}. \end{aligned}$$

Since $s \leq 0$ and $\lambda_i \geq 0, i = 0, 1, \dots, n$ then $\vec{q}_{0,n}^* \in L_\infty^w(\mu_{0,n}, \Pi_{rba}(\mathcal{Y}_{0,n}))$. Substituting $\vec{q}_{0,n}^*$ into $\mathbb{I}_D^{s,\lambda}(\mu_{0,n}, \vec{q}_{0,n})$ gives (IV.29).

Note that for $s = 0$ then $R_{0,n}^{na}(D) = 0$ and $\vec{q}_{0,n}^*(dy^n; x^n) = \nu_{0,n}^*(dy^n), \mu_{0,n}$ -almost all $x^n \in \mathcal{X}_{0,n}$. This is trivial so we must have $s < 0$. From Theorem IV.3 the solution occurs on the boundary of $\vec{Q}_{0,n}(D)$ giving (IV.30) for $s < 0$. \square

Often it is interesting to identify conditions so that the optimal reconstruction is Markov with respect to $\{X_i : i = 0, 1, \dots, n\}$. The next remark discusses this case.

Remark IV.5. Note that if the distortion function satisfies $\rho(T^i x^n, T^i y^n) = \rho(x_i, T^i y^n)$ then according to Theorem IV.4 we have

$$q_i^*(dy_i; y^{i-1}, x^i) = q_i^*(dy_i; y^{i-1}, x_i) - a.s., \quad i \in \mathbb{N}^n \quad (\text{IV.31})$$

that is, the reconstruction kernel is Markov in X^n . However, even if $\rho(T^i x^n, T^i y^n) = \rho(x_i, y_i)$ (single letter) one cannot claim that the optimal reconstruction distribution is also Markov with respect to $\{Y_i : i = 0, 1, \dots, n\}$ because the right hand side (RHS) of (IV.28) does not satisfy $\nu_i(dy_i; y^{i-1}) = \nu_i(dy_i; y_{i-1})$.

The relation between nonanticipative RDF and filtering theory is developed for fixed source distribution. In the next remark we discuss extensions of the nonanticipative RDF for a class of sources and relations to robust filtering.

Remark IV.6. *Nonanticipative RDF can be generalized to a class of sources to address robustness of the filter. One such class is defined by a relative entropy constraint between the unknown or true distribution P_{X^n} with respect to the nominal distribution $P_{X^n}^0$ via*

$$\mathcal{M}_{P_{X^n}^0}(d) \triangleq \{P_{X^n} \in \mathcal{M}_1(\mathcal{X}_{0,n}) : \mathbb{D}(P_{X^n} || P_{X^n}^0) \leq d\}$$

where d is the radius of uncertainty. Such a model of uncertainty or class of distributions is often employed in filtering and control applications because it is related to robust filtering and control using minimax methods [30], [31].

Therefore, the nonanticipative RDF for the class of sources $\mathcal{M}_{P_{X^n}^0}(d)$ is now defined using minimax strategies by

$$R_{0,n}^{na,+}(D, d) = \inf_{\vec{P}_{Y^n|X^n} \in \vec{\mathcal{Q}}_{0,n}(D)} \sup_{P_{X^n} \in \mathcal{M}_{P_{X^n}^0}(d)} \mathbb{I}(P_{X^n}, \vec{P}_{Y^n|X^n}) \quad (\text{IV.32})$$

Through (IV.32) one can obtain relations to minimax filtering strategies via nonanticipative RDF. An example using this formulation for control of Gaussian state space systems over limited rate channels is found in [32]. The investigation of the classical RDF for such a relative entropy class of sources is discussed in [33], where it is also shown that the Von-Neumann minimax theorem holds and hence one can interchange infimum and supremum operations. The validity of the Von-Neumann minimax theorem for (IV.32), will imply that the optimal reconstruction distribution for the minimax nonanticipative RDF is (IV.28), and hence the remaining task is to perform the infimum operation over the relative entropy class of the solution to the nonanticipative RDF given by (IV.29). This is the simplest approach to relate nonanticipative RDF for a class of sources and minimax filtering techniques. Unlike minimax filtering techniques, the filtering obtained from (IV.32) will always satisfy the fidelity criterion which can be defined with respect to probability

of error or the average error.

However, it is not clear how one can apply sensitivity minimization to nonanticipative RDF filter, because only the source distribution is given, while the observation map and filter are obtained from the realization of the optimal reconstruction distribution (see Fig. I.2). This is contrary to sensitivity minimization approach, where the input-output maps are given and depend on design functions, such as, the controller or the filter [34], [35]. Nevertheless, when the source is a second order Gaussian process described by a Power Spectral Density (PSD) and the fidelity of reconstruction is the mean-square error, then it might be possible to apply robust filtering and control techniques to address uncertainty of the PSD similar to the computation of capacity of channels with memory [36].

V. REALIZATION OF NONANTICIPATIVE RDF

The realization of the nonanticipative RDF (optimal reconstruction kernel and nonanticipative RDF) is equivalent to identifying the sensor mapping (see Fig. I.2) which generates the auxiliary random process $\{Z_i : i = 0, 1, \dots, n\}$ so that the optimal reconstruction conditional distribution is matched from the output of the source to the output of the filter. This intermediate mapping consists of an encoder followed by a channel. Thus, the realization of the nonanticipative optimal reconstruction distribution consists of a communication channel, an encoder and a decoder such that the reconstruction from the sequence X^n to the sequence Y^n matches the nonanticipative rate distortion minimizing reconstruction kernel. Fig. V.3 illustrates a cascade of subsystems that realizes the nonanticipative RDF. For the single letter expression of classical RDF this is related to the so-called source-channel matching of information theory [37]. It is also described in [38] and [39] for control over finite capacity communication channels, since this technique allows one to design encoding/decoding schemes without encoding and decoding delays. The realization of the optimal reconstruction kernel is given below.

Definition V.1. Given a source $\{P_{X_i|X^{i-1}}(dx_i|x^{i-1}) : i = 0, \dots, n\}$, a channel $\{P_{B_i|B^{i-1}, A^i}(db_i|b^{i-1}, a^i) : i = 0, \dots, n\}$ is a realization of the optimal nonanticipative reconstruction kernel $\{q_i^*(dy_i; y^{i-1}, x^i) : i = 0, \dots, n\}$ if there exists a pre-channel encoder $\{P_{A_i|A^{i-1}, B^{i-1}, X^i}(da_i|a^{i-1}, b^{i-1}, x^i) : i = 0, \dots, n\}$ and a post-channel decoder $\{P_{Y_i|Y^{i-1}, B^i}(dy_i|y^{i-1}, b^i) : i = 0, \dots, n\}$

such that

$$\begin{aligned}\vec{q}_{0,n}^*(dy^n; x^n) &\triangleq \otimes_{i=0}^n q_i^*(dy_i; y^{i-1}, x^i) \\ &= \otimes_{i=0}^n P_{Y_i|Y^{i-1}, X^i}(dy_i|y^{i-1}, x^i) - a.s.\end{aligned}\tag{V.33}$$

where the joint distribution is

$$\begin{aligned}P_{X^n, A^n, B^n, Y^n}(dx^n, da^n, db^n, dy^n) \\ &= \otimes_{i=0}^n P_{Y_i|Y^{i-1}, B^i, A^i, X^i}(dy_i|y^{i-1}, b^i, a^i, x^i) \\ &\quad \otimes P_{B_i|B^{i-1}, A^i, X^i, Y^{i-1}}(db_i|b^{i-1}, a^i, x^i, y^{i-1}) \otimes P_{A_i|A^{i-1}, X^i, Y^{i-1}, B^{i-1}}(da_i|a^{i-1}, x^i, y^{i-1}, b^{i-1}) \\ &\quad \otimes P_{X_i|X^{i-1}, A^{i-1}, B^{i-1}, Y^{i-1}}(dx_i|x^{i-1}, a^{i-1}, b^{i-1}, y^{i-1}) - a.s., \\ &= \otimes_{i=0}^n P_{Y_i|Y^{i-1}, B^i}(dy_i|y^{i-1}, b^i) \otimes P_{B_i|B^{i-1}, A^i}(db_i|b^{i-1}, a^i) \\ &\quad \otimes P_{A_i|A^{i-1}, B^{i-1}, X^i}(da_i|a^{i-1}, b^{i-1}, x^i) \otimes P_{X_i|X^{i-1}}(dx_i|x^{i-1}) - a.s.\end{aligned}\tag{V.34}$$

The filter is given by $\{P_{X_i|B^{i-1}}(dx_i|b^{i-1}) : i = 0, \dots, n\}$.

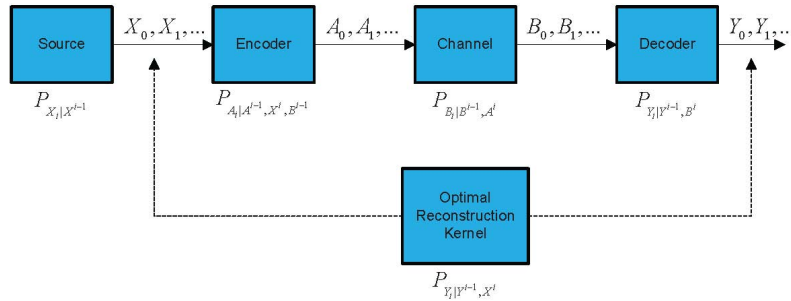


Fig. V.3. Block Diagram of Realizable Nonanticipative Rate Distortion Function

Thus, $\{B_i : i = 0, 1, \dots, n\}$ is the auxiliary random process which is obtained during the realization procedure in order to define the filter $\{P_{X_i|B^{i-1}}(dx_i|b^{i-1}) : i = 0, \dots, n\}$. Note that unlike Bayesian filtering in which the auxiliary process represents the observations which are given á priori, in nonanticipative RDF this is identified during the realization procedure. In the Definition V.1, the following MC assumptions are assumed.

- 1) $(X^i, A^i) \leftrightarrow (Y^{i-1}, B^i) \leftrightarrow Y_i;$
- 2) $(X^i, Y^{i-1}) \leftrightarrow (B^{i-1}, A^i) \leftrightarrow B_i;$

- 3) $Y^{i-1} \leftrightarrow (A^{i-1}, B^{i-1}, X^i) \leftrightarrow A_i;$
- 4) $(A^{i-1}, B^{i-1}, Y^{i-1}) \leftrightarrow X^{i-1} \leftrightarrow X_i.$

These conditional independent assumptions are natural since they correspond to data processing inequalities [4]. Thus, if $\{P_{B_i|B^{i-1}, A^i}(db_i|b^{i-1}, a^i) : i = 0, \dots, n\}$ is a realization of the nonanticipative RDF minimizing kernel $\{q_i^*(dy_i; y^{i-1}, x^i) : i = 0, \dots, n\}$ then the channel connecting the source, encoder, channel, decoder achieves the nonanticipative RDF, and the filter is obtained via $\{P_{X_i|B^{i-1}}(dx_i|b^{i-1}) : i = 0, \dots, n\}$. Moreover, the above MCs imply the following data processing inequality, $I(A^n \rightarrow B^n) \triangleq \sum_{i=0}^n I(A^i; B_i|B^{i-1}) \geq I(X^n; Y^n)$. The optimal realization (encoder-channel-decoder) is defined as the one for which the last inequality holds with equality.

VI. EXAMPLE

In this section, we present the filter for Gaussian Markov partially-observable processes by utilizing the realization procedure of Section V.

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

$$\begin{cases} X_{t+1} = AX_t + BW_t, & X_0 = X, \quad t \in \mathbb{N}^n \\ Y_t = CX_t + NV_t, & t \in \mathbb{N}^n \end{cases} \quad (\text{VI.35})$$

where $X_t \in \mathbb{R}^m$ is the state (unobserved) process of information source (plant), and $Y_t \in \mathbb{R}^p$

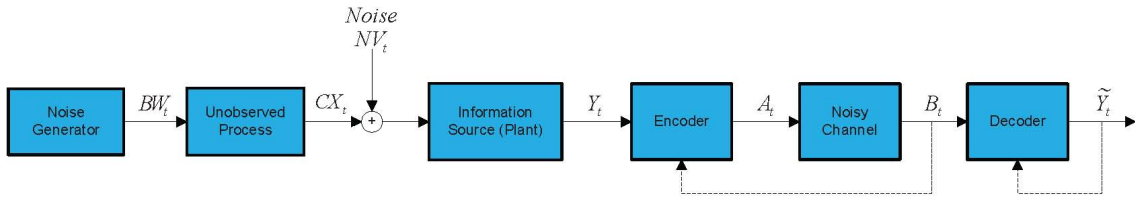


Fig. VI.4. Communication System

is the partially observed (measurement) process. The model in (VI.35) consists of a process $\{X_t : t \in \mathbb{N}^n\}$ which is not directly observed; instead what is directly observed is the process $\{Y_t : t \in \mathbb{N}^n\}$ which is a noisy version of $\{X_t : t \in \mathbb{N}^n\}$. This is a realistic model for any sensor which collects information for an underlying process, since the sensor is a measurement

device which is often subject to additive Gaussian noise. Hence, in this application the objective is to compress the sensor data. Since we only treat the stationary case, we assume that (C, A) is detectable and $(A, \sqrt{BB^{tr}})$ is stabilizable, $(N \neq 0)$ [18]. The state and observation noise $\{(W_t, V_t) : t \in \mathbb{N}^n\}$ are mutually independent, independent of the Gaussian RV X_0 , with parameters $N(\bar{x}_0, \bar{\Sigma}_0)$, where $W_t \in \mathbb{R}^k$ and $V_t \in \mathbb{R}^d$, are Gaussian IID processes with zero mean and identity covariances.

The realization will be done following Fig. VI.4. The goal is to reconstruct $\{Y_t : t \in \mathbb{N}^n\}$ by $\{\tilde{Y}_t : t \in \mathbb{N}^n\}$ causally. The distortion is single letter defined by

$$d_{0,n}(y^n, \tilde{y}^n) \triangleq \frac{1}{n+1} \sum_{t=0}^n \|y_t - \tilde{y}_t\|^2.$$

The objective is to compute

$$R_{0,n}^{na}(D) = \inf_{\vec{P}_{\tilde{Y}^n|Y^n} \in \vec{Q}_{0,n}(D)} \frac{1}{n+1} \mathbb{I}(P_{Y^n}, \vec{P}_{\tilde{Y}^n|Y^n}) \quad (\text{VI.36})$$

where $\vec{Q}_{0,n}(D) \triangleq \{\vec{P}_{\tilde{Y}^n|Y^n} : E\{d_{0,n}(Y^n, \tilde{Y}^n)\} \leq D\}$, and realize the reconstruction distribution. The reconstruction of $\{X_t : t \in \mathbb{N}^n\}$ when it is fully observed, i.e., when $Y_t = X_t$, is realized over a scalar additive white Gaussian noise (AWGN) channel in [9], while the partially observed scalar reconstruction of $\{Y_t : t \in \mathbb{N}^n\}$ is realized over a scalar AWGN channel in [38] via indirect methods (utilizing upper bounds which are achievable).

Here, we consider the vector process $Y_t \in \mathbb{R}^p$ and realize it over a vector AWGN channel. The methodology is based on the explicit formulae of optimal reconstruction of Theorem IV.4. According to Theorem IV.4, the optimal reconstruction is given by

$$\vec{P}_{\tilde{Y}^n|Y^n}^*(d\tilde{y}^n|y^n) = \otimes_{t=0}^n \frac{e^{s\|\tilde{y}_t - y_t\|^2} P_{\tilde{Y}_t|\tilde{Y}^{t-1}}(d\tilde{y}_t|\tilde{y}^{t-1})}{\int_{Y_t} e^{s\|\tilde{y}_t - y_t\|^2} P_{\tilde{Y}_t|\tilde{Y}^{t-1}}(d\tilde{y}_t|\tilde{y}^{t-1})}, \quad s \leq 0 \quad (\text{VI.37})$$

where each term in the RHS is identical because our results are derived based on the stationarity assumption. Hence, from (VI.37) it follows that $P_{\tilde{Y}_t|\tilde{Y}^{t-1}, Y^t} = P_{\tilde{Y}_t|\tilde{Y}^{t-1}, Y_t}(d\tilde{y}_t|\tilde{y}^{t-1}, y_t)$ —a.s., that is the reconstruction is Markov with respect to the process $\{Y_t : t \in \mathbb{N}^n\}$. Moreover, since the exponential term $\|\tilde{y}_t - y_t\|^2$ in the RHS of (VI.37) is quadratic in (y_t, \tilde{y}_t) , and $\{X_t : t \in \mathbb{N}^n\}$ is Gaussian then $\{(X_t, Y_t) : t \in \mathbb{N}^n\}$ is jointly Gaussian, and it follows that a Gaussian distribution $P_{\tilde{Y}_t|\tilde{Y}^{t-1}, Y_t}(\cdot|\tilde{y}^{t-1}, y_t)$ (for a fixed realization of (\tilde{y}^{t-1}, y_t)), and Gaussian distribution $P_{\tilde{Y}_t|\tilde{Y}^{t-1}}(\cdot|\tilde{y}^{t-1})$ can match the left and right side of (VI.37). Therefore, at time $t \in \mathbb{N}^n$, the output \tilde{Y}_t of the optimal reconstruction channel depends on Y_t and the previous channel outputs

\tilde{Y}^{t-1} , and its conditional distribution is Gaussian. Hence, the channel connecting $\{Y_t : t \in \mathbb{N}^n\}$ to $\{\tilde{Y}_t : t \in \mathbb{N}^n\}$ has the general form

$$\tilde{Y}_t = \bar{A}_t Y_t + \bar{B}_t \tilde{Y}^{t-1} + Z_t, \quad t \in \mathbb{N}^n \quad (\text{VI.38})$$

where $\bar{A}_t \in \mathbb{R}^{p \times p}$, $\bar{B}_t \in \mathbb{R}^{p \times tp}$, and $\{Z_t : t \in \mathbb{N}^n\}$ is an independent sequence of Gaussian vectors. Since we treat the stationary case, the finite horizon analysis below is only an intermediate state before we give the stationary solution.

The communication channel (VI.38) can be realized via a memoryless additive Gaussian noise channel with feedback [4] defined by

$$B_t = A_t + Z_t, \quad t \in \mathbb{N}^n \quad (\text{VI.39})$$

where the encoder, at time t , is a mapping $A_t = \Phi_t(Y_t, \tilde{Y}^{t-1})$ with power $P_t \triangleq \text{Trace} E\{A_t A_t^{tr}\}$, and the decoder at time $t \in \mathbb{N}^n$ receives B^t and computes the reconstruction $\tilde{Y}_t = \Psi_t(B^t, \tilde{Y}^{t-1})$. By Section V, in view of the MCs we have the data processing inequality $\mathbb{I}(P_{Y^n}, \vec{P}_{\tilde{Y}^n|Y^n}) \leq I(A^n \rightarrow B^n) = I(A^n; B^n)$, where the last equality holds because the channel is memoryless [4].

For the realization, the first step is the whitening of the source $\{Y_t : t \in \mathbb{N}^n\}$ by introducing the Gaussian innovation process $\{K_t : t \in \mathbb{N}^n\}$, defined by

$$K_t \triangleq Y_t - E\{Y_t | \sigma\{\tilde{Y}^{t-1}\}\}, \quad t \in \mathbb{N}^n \quad (\text{VI.40})$$

whose covariance is defined by

$$\Lambda_t \triangleq E\{K_t K_t^{tr}\}, \quad t \in \mathbb{N}^n. \quad (\text{VI.41})$$

The second step is the diagonalization of the covariance $\{\Lambda_t : t \in \mathbb{N}^n\}$ by introducing a unitary transformation $\{E_t : t \in \mathbb{N}^n\}$ such that

$$E_t \Lambda_t E_t^{tr} = \text{diag}\{\lambda_{t,1}, \dots, \lambda_{t,p}\}, \quad t \in \mathbb{N}^n. \quad (\text{VI.42})$$

Thus, $\Gamma_t \triangleq E_t K_t$, where $\{\Gamma_t : t \in \mathbb{N}^n\}$ has independent components for each $t \in \mathbb{N}^n$. In practise, the encoder consists of a pre-encoder which preprocesses the observations $\{Y_t : t \in \mathbb{N}^n\}$ by generating $\{K_t : t \in \mathbb{N}^n\}$ and then applies $\{E_t : t \in \mathbb{N}^n\}$ to it. At the decoder end, there is a pre-decoder which generates $\{\tilde{K}_t : t \in \mathbb{N}^n\}$ defined by

$$\tilde{K}_t \triangleq \tilde{Y}_t - E\{Y_t | \sigma\{\tilde{Y}^{t-1}\}\}, \quad t \in \mathbb{N}^n \quad (\text{VI.43})$$

on which the unitary transformation $\{E_t : t \in \mathbb{N}^n\}$ is applied to generate $\tilde{\Gamma}_t = E_t \tilde{K}_t$. Next, we calculate the RDF by taking advantage of the preprocessing at the encoder-decoder. Note that the fidelity criterion $d_{0,n}(\cdot, \cdot)$ is not affected by the preprocessing at the encoder-decoder since $d_{0,n}(Y^n, \tilde{Y}^n) = d_{0,n}(K^n, \tilde{K}^n) = \frac{1}{n+1} \sum_{t=0}^n \|\tilde{K}_t - K_t\|^2 = \frac{1}{n+1} \sum_{t=0}^n \|\tilde{\Gamma}_t - \Gamma_t\|^2$. Now, we show that

$$\begin{aligned} \mathbb{I}(P_{Y^n}, \vec{P}_{\tilde{Y}^n|Y^n}) &= \sum_{t=0}^n \left(H(\tilde{K}_t | \tilde{K}^{t-1}) - H(\tilde{K}_t | \tilde{K}^{t-1}, K_t) \right) \\ &= \sum_{t=0}^n \left(H(\tilde{\Gamma}_t | \tilde{\Gamma}^{t-1}) - H(\tilde{\Gamma}_t | \tilde{\Gamma}^{t-1}, \Gamma_t) \right). \end{aligned}$$

By (VI.37),

$$\vec{P}_{\tilde{Y}^n|Y^n}(d\tilde{y}^n|x^n) = \otimes_{t=0}^n P_{\tilde{Y}_t|\tilde{Y}^{t-1}, Y_t}(d\tilde{y}_t|\tilde{y}^{t-1}, y_t) - a.s.$$

Hence,

$$\mathbb{I}(P_{Y^n}, \vec{P}_{\tilde{Y}^n|Y^n}) = \sum_{t=0}^n \left(H(\tilde{Y}_t | \tilde{Y}^{t-1}) - H(\tilde{Y}_t | \tilde{Y}^{t-1}, Y_t) \right).$$

Since conditional entropy is translation invariant, utilizing (VI.43) gives

$$\begin{aligned} H(\tilde{Y}_t | \tilde{Y}^{t-1}) &= H(\tilde{K}_t | \tilde{Y}^{t-1}) \\ &= H(\tilde{K}_t | \tilde{Y}_{-1}, \tilde{Y}_0, \dots, \tilde{Y}_{t-1}) \\ &\stackrel{(a)}{=} H(\tilde{K}_t | \tilde{Y}_{-1}, \tilde{Y}_0, \dots, \tilde{Y}_{t-2}, \tilde{K}_{t-1} + E(\tilde{Y}_{t-1} | \sigma\{\tilde{Y}^{t-2}\})) \\ &= H(\tilde{K}_t | \tilde{Y}_{-1}, \tilde{Y}_0, \dots, \tilde{Y}_{t-2}, \tilde{K}_{t-1}) \\ &= H(\tilde{K}_t | \tilde{K}^{t-1}) \end{aligned} \tag{VI.44}$$

and repeated application of step (a) gives (VI.44). Similarly, $H(\tilde{Y}_t | \tilde{Y}^{t-1}, Y_t) = H(\tilde{K}_t | \tilde{Y}^{t-1}, K_t) = H(\tilde{K}_t | \tilde{K}^{t-1}, K_t)$. Hence,

$$\mathbb{I}(P_{Y^n}, \vec{P}_{\tilde{Y}^n|Y^n}) = \sum_{t=0}^n \left(H(\tilde{K}_t | \tilde{K}^{t-1}) - H(\tilde{K}_t | \tilde{K}^{t-1}, K_t) \right) \equiv \sum_{t=0}^n I(K_t; \tilde{K}_t | \tilde{K}^{t-1}).$$

Since the unitary transformation is non-singular then $\mathbb{I}(P_{Y^n}, \vec{P}_{\tilde{Y}^n|Y^n}) = \sum_{t=0}^n I(K_t; \tilde{K}_t | \tilde{K}^{t-1}) = \sum_{t=0}^n I(\Gamma_t; \tilde{\Gamma}_t | \tilde{\Gamma}^{t-1})$, $t \in \mathbb{N}^n$. Therefore, (VI.36) is equivalent to the following expression.

$$R_{0,n}^{na}(D) = R_{0,n}^{na, \Gamma^n, \tilde{\Gamma}^n}(D) \triangleq \inf_{\vec{P}_{\tilde{\Gamma}^n|\Gamma^n}: E\{d_{0,n}(\Gamma^n, \tilde{\Gamma}^n) \leq D\}} \frac{1}{n+1} \mathbb{I}(P_{\Gamma^n}, \vec{P}_{\tilde{\Gamma}^n|\Gamma^n}). \tag{VI.45}$$

By [3] (invoking an upper bound and Shannon's lower bound if necessary) the stationary solution of (VI.45) is given by

$$\lim_{n \rightarrow \infty} R_{0,n}^{na, \Gamma^n, \tilde{\Gamma}^n}(D) = \lim_{n \rightarrow \infty} \frac{1}{2} \frac{1}{n+1} \sum_{t=0}^n \sum_{i=1}^p \log \left(\frac{\lambda_{t,i}}{\delta_{t,i}} \right)$$

where

$$\delta_{t,i} \triangleq \begin{cases} \xi_t & \text{if } \xi_t \leq \lambda_{t,i} \\ \lambda_{t,i} & \text{if } \xi_t > \lambda_{t,i} \end{cases}, \quad t \in \mathbb{N}^n, \quad i = 2, \dots, p$$

and $\{\xi_t : t \in \mathbb{N}^n\}$ is chosen such that $\sum_{i=1}^p \delta_{t,i} = D$. Define $\eta_{t,i} \triangleq 1 - \frac{\delta_{t,i}}{\lambda_{t,i}}$, $i = 1, \dots, p$, $\Delta_t \triangleq \text{diag}\{\delta_{t,1}, \dots, \delta_{t,p}\}$, and $H_t \triangleq \text{diag}\{\eta_{t,1}, \dots, \eta_{t,p}\} \in \mathbb{R}^{p \times p}$.

As a result, the reconstruction conditional distribution is given by

$$P_{\tilde{\Gamma}^n | \Gamma^n}^*(d\tilde{\gamma}^n | \gamma^n) = \otimes_{t=0}^n P_{\Gamma_t | \tilde{\Gamma}_t}^*(d\tilde{\gamma}_t | \gamma_t) - a.s.$$

where $P_{\tilde{\Gamma}_t | \Gamma_t}^*(\cdot | \cdot) \sim N(H_t \Gamma_t, H_t \Delta_t)$.

Realization of Nonanticipative RDF Over Vector AWGN Channel. Consider a vector channel

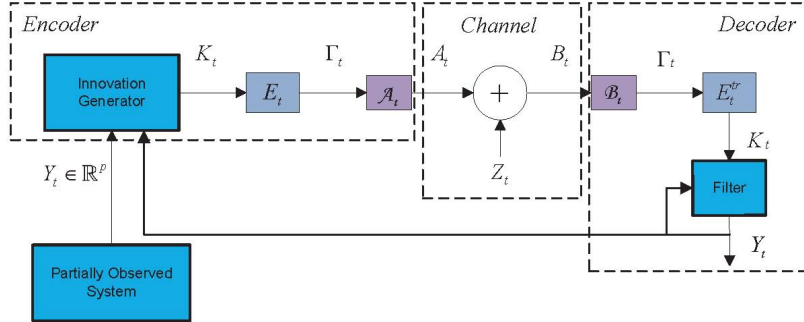


Fig. VI.5. Design of Discrete-Time Communication System

$B_t = A_t + Z_t$, $t \in \mathbb{N}^n$, where Z_t is Gaussian zero mean, $Q \triangleq \text{Cov}(Z_t) = \text{diag}\{q_1, q_2, \dots, q_p\}$, and $A_t \in \mathbb{R}^p$. By Section V, and the memoryless nature of the channel we know that $I(A^n \rightarrow B^n) = I(A^n; B^n) \geq \mathbb{I}(P_{Y^n}, \vec{P}_{\tilde{Y}^n | Y^n})$. Hence, we compress the source and transmit it to the decoder over the vector channel, so that the RDF is equal to the capacity of the channel, i.e., $\lim_{n \rightarrow \infty} R_{0,n}^{na}(D) = \lim_{n \rightarrow \infty} \frac{1}{n+1} I(A^n; B^n)$. That is, we match the source to the channel. Therefore, we need to design the operators $\{(\mathcal{A}_t, \mathcal{B}_t) : t \in \mathbb{N}^n\}$ so that the compressed signal $A_t = \mathcal{A}_t \Gamma_t$, is sent through an AWGN channel with feedback (shown in Fig. VI.5), after which

the received signal is decompressed by $\tilde{\Gamma}_t = \mathcal{B}_t B_t$ at the pre-decoder. By the knowledge of the channel output at the decoder, the mean square estimator \hat{X}_t is generated at the decoder since $\hat{X}_t \triangleq E\{X_t | \sigma\{\tilde{Y}^{t-1}\}\}$ (one may also use $\sigma\{B^{t-1}\}$ to find the filter of $\{X_t : t \in \mathbb{N}^n\}$).

The compression operator $\{\mathcal{A}_t : t \in \mathbb{N}^n\}$ is chosen so that $\lim_{n \rightarrow \infty} R_{0,n}^{na}(D) = \lim_{n \rightarrow \infty} \frac{1}{n+1} I(A^n; B^n)$. Recall that $B_t = A_t + Z_t$, $A_t = \mathcal{A}_t E_t K_t$, $Q \triangleq \text{Cov}(Z_t)$, $\text{Trace} E\{A_t A_t^{tr}\} = P_t$, $t = 0, 1, \dots, n$. Hence, we find $\{\mathcal{A}_t : t \in \mathbb{N}^n\}$ so that the following holds.

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{n+1} C_{0,n}(P_0, \dots, P_n) &\triangleq \lim_{n \rightarrow \infty} \frac{1}{n+1} I(A^n; B^n) = \lim_{n \rightarrow \infty} \frac{1}{2} \frac{1}{n+1} \sum_{t=0}^n \log |I + E\{A_t A_t^{tr}\} Q^{-1}| \\ &= \lim_{n \rightarrow \infty} \frac{1}{2} \frac{1}{n+1} \sum_{t=0}^n \log \frac{|\Lambda_t|}{|\Delta_t|} = \lim_{n \rightarrow \infty} R_{0,n}^{na}(D). \end{aligned}$$

From the previous equality we obtain

$$\mathcal{A}_t \triangleq \sqrt{Q \Delta_t^{-1} H_t}, \quad t \in \mathbb{N}^n.$$

The decompression operator $\{\mathcal{B}_t : t \in \mathbb{N}^n\}$ is chosen so that the desired distortion is achieved by the above realization. The decompressed channel output $\tilde{\Gamma}_t = \mathcal{B}_t B_t$ due to transmitting the compressed input $A_t = \mathcal{A}_t \Gamma_t$ is

$$\begin{aligned} \tilde{\Gamma}_t &= \mathcal{B}_t B_t = \mathcal{B}_t (A_t + Z_t) = \mathcal{B}_t (\mathcal{A}_t \Gamma_t + Z_t), \quad \Gamma_t = E_t K_t \\ &= H_t E_t K_t + \mathcal{B}_t Z_t, \quad t \in \mathbb{N}^n. \end{aligned} \tag{VI.46}$$

By pre-multiplying $\tilde{\Gamma}_t$ by E_t^{tr} we can construct

$$\tilde{K}_t = E_t^{tr} \tilde{\Gamma}_t = E_t^{tr} H_t E_t K_t + E_t^{tr} \mathcal{B}_t Z_t, \quad t \in \mathbb{N}^n.$$

The reconstruction of Y_t is given by the sum of \tilde{K}_t and $C \hat{X}_t$ as follows.

$$\begin{aligned} \tilde{Y}_t &= \Psi_t(B^t, \tilde{Y}^{t-1}) \\ &= \tilde{K}_t + C \hat{X}_t, \quad \hat{X}_t = E\{X_t | \sigma\{\tilde{Y}^{t-1}\}\} \end{aligned} \tag{VI.47}$$

$$= E_t^{tr} H_t E_t K_t + E_t^{tr} \mathcal{B}_t Z_t + C \hat{X}_t, \quad t \in \mathbb{N}^n. \tag{VI.48}$$

Next, we determine $\{\mathcal{B}_t : t \in \mathbb{N}^n\}$.

First, we notice that

$$E\{(Y_t - \tilde{Y}_t)^{tr} (Y_t - \tilde{Y}_t)\} = \text{Trace} \left(E\{(Y_t - \tilde{Y}_t)(Y_t - \tilde{Y}_t)^{tr}\} \right).$$

Then we can compute

$$\begin{aligned}
E\left\{(Y_t - \tilde{Y}_t)^{tr}(Y_t - \tilde{Y}_t)\right\} &= \text{Trace}E\left\{(K_t - \tilde{K}_t)(K_t - \tilde{K}_t)^{tr}\right\} \\
&= \text{Trace}E\left\{(K_t - E_t^{tr}\tilde{\Gamma}_t)(K_t - E_t^{tr}\tilde{\Gamma}_t)^{tr}\right\} \\
&= \text{Trace}E\left\{(K_t - E_t^{tr}H_tE_tK_t - E_t^{tr}\mathcal{B}_tZ_t)(K_t - E_t^{tr}H_tE_tK_t - E_t^{tr}\mathcal{B}_tZ_t)^{tr}\right\} \\
&= \text{Trace}E\left\{((I - E_t^{tr}H_tE_t)K_t - E_t^{tr}\mathcal{B}_tZ_t)((I - E_t^{tr}H_tE_t)K_t - E_t^{tr}\mathcal{B}_tZ_t)^{tr}\right\} \\
&= \text{Trace}\left\{(I - E_t^{tr}H_tE_t)\Lambda_t(I - E_t^{tr}H_tE_t)^{tr} + E_t^{tr}\mathcal{B}_tQ\mathcal{B}_t^{tr}E_t\right\} \\
&= \text{Trace}\left\{(I - E_t^{tr}H_tE_t)E_t^{tr}\text{diag}(\lambda_{t,1}, \dots, \lambda_{t,p})E_t(I - E_t^{tr}H_tE_t)^{tr} + E_t^{tr}\mathcal{B}_tQ\mathcal{B}_t^{tr}E_t\right\} \\
&= \text{Trace}\left\{E_t^{tr}\left((I - H_t)\text{diag}(\lambda_{t,1}, \dots, \lambda_{t,p})(I - H_t)^{tr} + (\mathcal{B}_tQ\mathcal{B}_t^{tr})\right)E_t\right\} \\
&\stackrel{(b)}{=} \text{Trace}\left\{\text{diag}(\delta_{t,1}, \dots, \delta_{t,p})\right\} = D
\end{aligned}$$

where (b) holds if we set

$$\mathcal{B}_t \triangleq \sqrt{H_t\Delta_tQ^{-1}}, \quad t \in \mathbb{N}^n.$$

This shows that the realization of Fig. VI.5 achieves end-to-end average distortion equal to D .

Decoder. The decoder is $\tilde{Y}_t = \tilde{K}_t + C\hat{X}_t$, where $\{\hat{X}_t : t \in \mathbb{N}^n\}$ is obtained from the modified Kalman filter as follows. Recall that

$$\begin{aligned}
\tilde{Y}_t &= \tilde{K}_t + C\hat{X}_t \\
&= E_t^{tr}H_tE_t(Y_t - C\hat{X}_t) + E_t^{tr}\mathcal{B}_tZ_t + C\hat{X}_t \\
&= E_t^{tr}H_tE_t(CX_t + NV_t - C\hat{X}_t) + E_t^{tr}\mathcal{B}_tZ_t + C\hat{X}_t \\
&= E_t^{tr}H_tE_t(CX_t - C\hat{X}_t) + C\hat{X}_t + E_t^{tr}H_tE_tNV_t + E_t^{tr}\mathcal{B}_tZ_t \tag{VI.49}
\end{aligned}$$

where $\{V_t : t \in \mathbb{N}^n\}$ and $\{Z_t : t \in \mathbb{N}^n\}$ are independent Gaussian vectors. Then $\hat{X}_t = E\{X_t | \sigma\{\tilde{Y}^{t-1}\}\}$ is given by the modified Kalman filter

$$\hat{X}_{t+1} = A\hat{X}_t + A\Sigma_t(E_t^{tr}H_tE_tC)^{tr}M_t^{-1}(\tilde{Y}_t - C\hat{X}_t), \quad \hat{X}_0 = \bar{x}_0 \tag{VI.50}$$

$$\Sigma_{t+1} = A\Sigma_tA^{tr} - A\Sigma_t(E_t^{tr}H_tE_tC)^{tr}M_t^{-1}(E_t^{tr}H_tE_tC)\Sigma_tA + BB_t^{tr}, \quad \Sigma_0 = \bar{\Sigma}_0 \tag{VI.51}$$

where

$$M_t = E_t^{tr}H_tE_tC\Sigma_t(E_t^{tr}H_tE_tC)^{tr} + E_t^{tr}H_tE_tNN^{tr}(E_t^{tr}H_tE_t)^{tr} + E_t^{tr}\mathcal{B}_t\mathcal{B}_t^{tr}E_t.$$

Stationary Solution: Infinite Horizon. Now, we are ready to give the complete solution to the stationary nonanticipative RDF and its realization. As $t \rightarrow \infty$, under the assumption that the linear Gauss-Markov system is stabilizable and detectable, we have

$$\Sigma_\infty = A\Sigma_\infty A^{tr} - A\Sigma_\infty (E_\infty^{tr} H_\infty E_\infty C)^{tr} M_\infty^{-1} (E_\infty^{tr} H_\infty E_\infty C) \Sigma_\infty A + B B_\infty^{tr}$$

where

$$M_\infty = E_\infty^{tr} H_\infty E_\infty C \Sigma_\infty (E_\infty^{tr} H_\infty E_\infty C)^{tr} + E_\infty^{tr} H_\infty E_\infty N N^{tr} (E_\infty^{tr} H_\infty E_\infty)^{tr} + E_\infty^{tr} \mathcal{B}_\infty \mathcal{B}_\infty^{tr} E_\infty$$

and E_∞ is the unitary matrix that diagonalizes Λ_∞ given by

$$E_\infty \Lambda_\infty E_\infty^{tr} = \text{diag}(\lambda_{\infty,1}, \dots, \lambda_{\infty,p})$$

and

$$\delta_{\infty,i} \triangleq \begin{cases} \xi_\infty & \text{if } \xi_\infty \leq \lambda_{\infty,i} \\ \lambda_{\infty,i} & \text{if } \xi_\infty > \lambda_{\infty,i} \end{cases}, \quad i = 1, \dots, p$$

satisfying $\sum_{i=1}^p \delta_{\infty,i} = D$.

Define

$$\Delta_\infty = \text{diag}(\delta_{\infty,1}, \dots, \delta_{\infty,p}), \quad H_\infty = \text{diag}(\eta_{\infty,1}, \dots, \eta_{\infty,p})$$

where $\eta_{\infty,i} = 1 - \frac{\delta_{\infty,i}}{\lambda_{\infty,i}}$. The nonanticipative RDF can be computed as follows.

$$\begin{aligned} R^{na}(D) &= \lim_{n \rightarrow \infty} \inf_{P_{\tilde{Y}^n|Y^n}(d\tilde{y}^n|y^n) \in \vec{\mathcal{Q}}_{0,n}(D)} \frac{1}{n+1} \mathbb{I}(P_{Y^n}, \vec{P}_{\tilde{Y}^n|Y^n}) \\ &= \lim_{n \rightarrow \infty} \left(\frac{1}{2} \frac{1}{n+1} \sum_{t=0}^n \sum_{i=1}^p \log \left(\frac{\lambda_{t,i}}{\delta_{t,i}} \right) \right) \\ &= \frac{1}{2} \sum_{i=1}^p \log \left(\frac{\lambda_{\infty,i}}{\delta_{\infty,i}} \right) = \frac{1}{2} \log \frac{|\Lambda_\infty|}{|\Delta_\infty|} = \frac{1}{2} \lim_{n \rightarrow \infty} C_{0,n}(P_0, \dots, P_n) \stackrel{(c)}{=} C(P) \end{aligned} \quad (\text{VI.52})$$

where (c) comes from the fact that the power constraint satisfies $\lim_{t \rightarrow \infty} \text{Trace} E\{A_t A_t^{tr}\} = \lim_{t \rightarrow \infty} P_t = P$. Thus, for a given distortion level D , $C(P) = R^{na}(D)$ is the minimum capacity under which there exists a realizable filter for the data reconstruction of $\{Y_t : t \in \mathbb{N}\}$ by $\{\tilde{Y}_t : t \in \mathbb{N}\}$ ensuring an average distortion equal to D . Note that for $\frac{D}{p} < \min_i \lambda_{\infty,i}$ then $R^{na}(D) = \frac{1}{2} \log \frac{|\Lambda_\infty|}{(\frac{D}{p})^p}$, e.g., $\delta_{\infty,i} = \frac{D}{p}$. Hence, from (VI.52) we have $D = p \left(|\Lambda_\infty| e^{-2R^{na}} \right)^{\frac{1}{p}}$. As a result, we have the direct relation between the reconstruction error D and the rate R^{na} . Finally, the filter is the steady state version of (VI.50), (VI.51) with initial condition $\hat{X}_0 = E\{X_0|Y^{-1}\}$ and Σ_0 the covariance of $X_0 - \hat{X}$ which is Gaussian $N(0, \Sigma_\infty)$.

VII. CONCLUSION

This paper investigates nonanticipative RDF on abstract spaces. Existence of the optimal reconstruction conditional distribution is shown, while closed form expression is derived for the stationary case. The relation between filtering theory and nonanticipative rate distortion theory is discussed via a realization procedure. Finally, an example is presented which illustrates the realization of the nonanticipative RDF.

APPENDIX

A. Proof of Lemma III.2

To show closedness of \vec{Q}_{ad} as a subset of Q_{ad} it suffices to show that

$$\otimes_{i=0}^n q_i^\alpha(\cdot; y^{i-1}, x^i) \xrightarrow{w^*} \otimes_{i=0}^n q_i^0(\cdot; y^{i-1}, x^i).$$

This will be shown by induction. Consider $n = 0$. For any $h_0(x_0, y_0) \in L_1(\mu_0, BC(\mathcal{Y}_0))$, by definition of weak*-convergence it follows from (a) that

$$\lim_{\alpha \rightarrow \infty} \int_{\mathcal{X}_0 \times \mathcal{Y}_0} h_0(x_0, y_0) q_0^\alpha(dy_0; x_0) \mu_0(dx_0) = \int_{\mathcal{X}_0 \times \mathcal{Y}_0} h_0(x_0, y_0) q_0^0(dy_0; x_0) \mu_0(dx_0).$$

Consider $n = 1$. For $\tilde{h}_0(\cdot, \cdot) \in L_1(\mu_0, BC(\mathcal{Y}_0))$, $\tilde{h}_1(\cdot, \cdot) \in L_1(\mu_1, BC(\mathcal{Y}_1))$ We need to show that

$$\begin{aligned} & \lim_{\alpha \rightarrow \infty} \left| \int_{\mathcal{X}_0} \left(\int_{\mathcal{Y}_0} h_0(x_0, y_0) \left(\int_{\mathcal{X}_1} \left(\int_{\mathcal{Y}_1} h_1(x_1, y_1) q_1^\alpha(dy_1; y_0, x^1) \right) \mu_1(dx_1; x_0) \right) q_0^\alpha(dy_0; x_0) \right) \mu_0(dx_0) \right. \\ & \left. - \int_{\mathcal{X}_0} \left(\int_{\mathcal{Y}_0} h_0(x_0, y_0) \left(\int_{\mathcal{X}_1} \left(\int_{\mathcal{Y}_1} h_1(x_1, y_1) q_1^0(dy_1; y_0, x^1) \right) \mu_1(dx_1; x_0) \right) q_0^0(dy_0; x_0) \right) \mu_0(dx_0) \right| = 0. \end{aligned}$$

The latter equation is written as follows.

$$\begin{aligned} & \left| \int_{\mathcal{X}_0} \left(\int_{\mathcal{Y}_0} h_0(x_0, y_0) \left(\int_{\mathcal{X}_1} \left(\int_{\mathcal{Y}_1} h_1(x_1, y_1) q_1^\alpha(dy_1; y_0, x^1) \right) \mu_1(dx_1; x_0) \right) q_0^\alpha(dy_0; x_0) \right) \mu_0(dx_0) \right. \\ & \left. - \int_{\mathcal{X}_0} \left(\int_{\mathcal{Y}_0} h_0(x_0, y_0) \left(\int_{\mathcal{X}_1} \left(\int_{\mathcal{Y}_1} h_1(x_1, y_1) q_1^0(dy_1; y_0, x^1) \right) \mu_1(dx_1; x_0) \right) q_0^0(dy_0; x_0) \right) \mu_0(dx_0) \right| \end{aligned}$$

$$\begin{aligned}
&\leq \left| \int_{\mathcal{X}_0} \left(\int_{\mathcal{Y}_0} h_0(x_0, y_0) \underbrace{\left(\int_{\mathcal{X}_1} \left(\int_{\mathcal{Y}_1} h_1(x_1, y_1) q_1^0(dy_1; y_0, x^1) \right) \mu_1(dx_1; x_0) \right)}_{\tilde{h}_1(x_0, y_0)} q_0^\alpha(dy_0; x_0) \right) \mu_0(dx_0) \right. \\
&\quad \left. - \int_{\mathcal{X}_0} \left(\int_{\mathcal{Y}_0} h_0(x_0, y_0) \left(\int_{\mathcal{X}_1} \left(\int_{\mathcal{Y}_1} h_1(x_1, y_1) q_1^0(dy_1; y_0, x^1) \right) \mu_1(dx_1; x_0) \right) q_0^0(dy_0; x_0) \right) \mu_0(dx_0) \right| \\
&\quad + \left| \int_{\mathcal{X}_0} \left(\int_{\mathcal{Y}_0} h_0(x_0, y_0) \left(\int_{\mathcal{X}_1} \left(\int_{\mathcal{Y}_1} h_1(x_1, y_1) q_1^\alpha(dy_1; y_0, x^1) \right) \mu_1(dx_1; x_0) \right) q_0^\alpha(dy_0; x_0) \right) \mu_0(dx_0) \right. \\
&\quad \left. - \int_{\mathcal{X}_0} \left(\int_{\mathcal{Y}_0} h_0(x_0, y_0) \left(\int_{\mathcal{X}_1} \left(\int_{\mathcal{Y}_1} h_1(x_1, y_1) q_1^0(dy_1; y_0, x^1) \right) \mu_1(dx_1; x_0) \right) q_0^\alpha(dy_0; x_0) \right) \mu_0(dx_0) \right|.
\end{aligned}$$

We need to show that both RHS terms go to zero as $a \rightarrow \infty$. Let $\epsilon > 0$ be given. Then, there exists an $\alpha_\epsilon \in \mathcal{D}$ such that for all $\alpha \succeq \alpha_\epsilon$ the first RHS term can be written as

$$\begin{aligned}
&\left| \int_{\mathcal{X}_0} \left(\int_{\mathcal{Y}_0} h_0(x_0, y_0) \tilde{h}_1(x_0, y_0) \left(q_0^\alpha(dy_0; x_0) - q_0^0(dy_0; x_0) \right) \right) \mu_0(dx_0) \right| \\
&= \left| \int_{\mathcal{X}_0} \left(\int_{\mathcal{Y}_0} h_0(x_0, y_0) \tilde{h}_1(x_0, y_0) \left(q_0^\alpha(dy_0; x_0) - q_0^0(dy_0; x_0) \right) \right) \mu_0(dx_0) \right| \\
&\leq \int_{\mathcal{X}_0} \left| \int_{\mathcal{Y}_0} h_0(x_0, y_0) \tilde{h}_1(x_0, y_0) \left(q_0^\alpha(dy_0; x_0) - q_0^0(dy_0; x_0) \right) \right| \mu_0(dx_0) \\
&\leq \epsilon, \quad \forall \epsilon > 0 \text{ and } \forall \alpha \succ \alpha_\epsilon
\end{aligned}$$

where the last inequality follows from condition (b), e.g., $\tilde{h}_0(\cdot, \cdot) \in L_1(\mu_0, BC(\mathcal{Y}_0))$.

The second RHS term can be written as

$$\begin{aligned}
&\left| \int_{\mathcal{X}_0} \left(\int_{\mathcal{Y}_0} h_0(x_0, y_0) \underbrace{\left(\int_{\mathcal{X}_1} \left(\int_{\mathcal{Y}_1} h_1(x_1, y_1) \left(q_1^\alpha(dy_1; y_0, x^1) - q_1^0(dy_1; y_0, x^1) \right) \right) \mu_1(dx_1; x_0) \right)}_{\tilde{h}_1^\alpha(x_0, y_0)} \right. \right. \\
&\quad \left. \left. \otimes q_0^\alpha(dy_0; x_0) \right) \mu_0(dx_0) \right| = \int_{\mathcal{X}_0} \int_{\mathcal{Y}_0} h_0(x_0, y_0) \tilde{h}_1^\alpha(x_0, y_0) q_0^\alpha(dy_0; x_0) \otimes \mu_0(dx_0). \quad (\text{A.53})
\end{aligned}$$

By condition (c) for $i = 1$, and $\forall \epsilon > 0$ and $\alpha \succ \alpha_\epsilon$ we have

$$\sup_{y_0 \in \mathcal{Y}_0} \int_{\mathcal{X}_1} \left| \int_{\mathcal{Y}_1} h_1(x_1, y_1) q_1^\alpha(dy_1; y_0, x^1) - \int_{\mathcal{Y}_1} h_1(x_1, y_1) q_1^0(dy_1; y_0, x^1) \right| \mu_1(dx_1; x_0) \leq \epsilon, \quad \forall x_0 \in \mathcal{X}_0.$$

Utilizing the last inequality into (A.53) yields that in the limit as $\alpha \rightarrow \infty$, then (A.53) goes to zero.

Next, suppose that for $n = k$ and for all $\epsilon > 0$ there exists $\alpha_\epsilon \in \mathcal{D}$ such that for any $\alpha \succeq \alpha_\epsilon$

$$\begin{aligned} & \left| \int_{\mathcal{X}_0} \left(\int_{\mathcal{Y}_0} h_0(x_0, y_0) \dots \left(\int_{\mathcal{X}_k} \left(\int_{\mathcal{Y}_k} h_k(x_k, y_k) q_k^\alpha(dy_k; y^{k-1}, x^k) \right) \mu_k(dx_k; x^{k-1}) \right) \right. \right. \\ & \quad \left. \dots q_0^\alpha(dy_0; x_0) \right) \mu_0(dx_0) \\ & - \int_{\mathcal{X}_0} \left(\int_{\mathcal{Y}_0} h_0(x_0, y_0) \dots \left(\int_{\mathcal{X}_k} \left(\int_{\mathcal{Y}_k} h_k(x_k, y_k) q_k^0(dy_k; y^{k-1}, x^k) \right) \mu_k(dx_k; x^{k-1}) \right) \right. \\ & \quad \left. \dots q_0^0(dy_0; x_0) \right) \mu_0(dx_0) \Big| \leq \epsilon. \end{aligned}$$

To conclude the derivation we need to show that for $n = k + 1$

$$\otimes_{i=0}^{k+1} q_i^\alpha(\cdot; y^{i-1}, x^i) \xrightarrow{w^*} \otimes_{i=0}^{k+1} q_i^0(\cdot; y^{i-1}, x^i).$$

Consider $n = k + 1$. We need to show that for all $\epsilon > 0$ there exists $\alpha_\epsilon \in \mathcal{D}$ such that for any

$\alpha \succeq \alpha_\epsilon$

$$\begin{aligned} & \left| \int_{\mathcal{X}_0} \left(\int_{\mathcal{Y}_0} h_0(x_0, y_0) \dots \left(\int_{\mathcal{X}_{k+1}} \left(\int_{\mathcal{Y}_{k+1}} h_{k+1}(x_{k+1}, y_{k+1}) q_{k+1}^\alpha(dy_{k+1}; y^k, x^{k+1}) \right) \mu_{k+1}(dx_{k+1}; x^k) \right) \right. \right. \\ & \quad \left. \dots q_0^\alpha(dy_0; x_0) \right) \mu_0(dx_0) \\ & - \int_{\mathcal{X}_0} \left(\int_{\mathcal{Y}_0} h_0(x_0, y_0) \dots \left(\int_{\mathcal{X}_{k+1}} \left(\int_{\mathcal{Y}_{k+1}} h_{k+1}(x_{k+1}, y_{k+1}) q_{k+1}^0(dy_{k+1}; y^k, x^{k+1}) \right) \mu_{k+1}(dx_{k+1}; x^k) \right) \right. \\ & \quad \left. \dots q_0^0(dy_0; x_0) \right) \mu_0(dx_0) \Big| \leq \epsilon. \end{aligned}$$

Since,

$$\begin{aligned} & \left| \int_{\mathcal{X}_0} \left(\int_{\mathcal{Y}_0} h_0(x_0, y_0) \dots \left(\int_{\mathcal{X}_{k+1}} \left(\int_{\mathcal{Y}_{k+1}} h_{k+1}(x_{k+1}, y_{k+1}) q_{k+1}^\alpha(dy_{k+1}; y^k, x^{k+1}) \right) \mu_{k+1}(dx_{k+1}; x^k) \right) \right. \right. \\ & \quad \left. \dots q_0^\alpha(dy_0; x_0) \right) \mu_0(dx_0) \\ & - \int_{\mathcal{X}_0} \left(\int_{\mathcal{Y}_0} h_0(x_0, y_0) \dots \left(\int_{\mathcal{X}_{k+1}} \left(\int_{\mathcal{Y}_{k+1}} h_{k+1}(x_{k+1}, y_{k+1}) q_{k+1}^0(dy_{k+1}; y^k, x^{k+1}) \right) \mu_{k+1}(dx_{k+1}; x^k) \right) \right. \\ & \quad \left. \dots q_0^0(dy_0; x_0) \right) \mu_0(dx_0) \Big| \end{aligned}$$

$$\begin{aligned}
&\leq \left| \int_{\mathcal{X}_{0,k}} \int_{\mathcal{Y}_{0,k}} \otimes_{i=0}^k h_i(x_i, y_i) \left(\int_{\mathcal{X}_{k+1}} \int_{\mathcal{Y}_{k+1}} h_{k+1}(x_{k+1}, y_{k+1}) \left(q_{k+1}^\alpha(dy_{k+1}; y^k, x^{k+1}) \right. \right. \right. \\
&\quad \left. \left. \left. - q_{k+1}^0(dy_{k+1}; y^k, x^{k+1}) \right) \mu_{k+1}(dx_{k+1}; x^k) \right) \otimes_{i=0}^k q_i^\alpha(dy_i; y^{i-1}, x^i) \otimes \mu_i(dx_i; x^{i-1}) \right| \\
&+ \left| \int_{\mathcal{X}_{0,k}} \int_{\mathcal{Y}_{0,k}} \otimes_{i=0}^k h_i(x_i, y_i) \underbrace{\left(\int_{\mathcal{X}_{k+1}} \int_{\mathcal{Y}_{k+1}} h_{k+1}(x_{k+1}, y_{k+1}) q_{k+1}^0(dy_{k+1}; y^k, x^{k+1}) \mu_{k+1}(dx_{k+1}; x^k) \right)}_{\tilde{h}_{k+1}(x^k, y^k)} \right. \\
&\quad \left. \otimes_{i=0}^k \left(q_i^\alpha(dy_i; y^{i-1}, x^i) - q_i^0(dy_i; y^{i-1}, x^i) \right) \otimes \mu_i(dx_i; x^{i-1}) \right|.
\end{aligned}$$

By condition (c) the following inequality holds, $\forall x^k \in \mathcal{X}_{0,k}$,

$$\begin{aligned}
&\sup_{y^k \in \mathcal{Y}_{0,k}} \int_{\mathcal{X}_{k+1}} \left| \int_{\mathcal{Y}_{k+1}} h_{k+1}(x_{k+1}, y_{k+1}) \left(q_{k+1}^\alpha(dy_{k+1}; y^k, x^{k+1}) - \right. \right. \\
&\quad \left. \left. q_{k+1}^0(dy_{k+1}; y^k, x^{k+1}) \right) \mu_{k+1}(dx_{k+1}; x^k) \right| \\
&\leq \epsilon, \quad \forall \epsilon > 0 \text{ and } \forall \alpha \succ \alpha_\epsilon.
\end{aligned}$$

Also, by condition (b), $\tilde{h}_{k+1} \in L_1(\mu_{0,k}, BC(\mathcal{Y}_{0,k}))$. Utilizing the previous observations and the induction hypothesis $\otimes_{i=0}^k q_i^\alpha(\cdot; y^{i-1}, x^i) \xrightarrow{w^*} \otimes_{i=0}^k q_i^0(\cdot; y^{i-1}, x^i)$ in the two inequalities above, then in the limit as $\alpha \rightarrow \infty$, the terms in the inequality go to zero.

As a result, \vec{Q}_{ad} is a weak*-closed set. Being a weak*-closed subset of the weak*-compact set Q_{ad} , \vec{Q}_{ad} is also weak*-compact.

B. Proof of Theorem IV.3

The proof is based on Lagrange Duality theorem [28, Theorem 1, p. 224]. We choose $X \triangleq L_\infty^w(\mu_{0,n}, M_{rba}(\mathcal{Y}_{0,n}))$ which is clearly a vector space. For the set Ω the natural choice is the set $\Omega = \vec{Q}_{ad} \equiv L_\infty^w(\mu_{0,n}, \Pi_{rba}(\mathcal{Y}_{0,n})) \subseteq X$. Define

$$\begin{aligned}
G(\vec{q}_{0,n}) &\triangleq \ell_{d_{0,n}}(\vec{q}_{0,n}) - D, \quad \vec{q}_{0,n} \in L_\infty^w(\mu_{0,n}, M_{rba}(\mathcal{Y}_{0,n})) \\
&\triangleq \int_{\mathcal{X}_{0,n}} \left(\int_{\mathcal{Y}_{0,n}} d_{0,n}(x^n, y^n) \vec{q}(dy^n; x^n) \right) \mu_{0,n}(dx^n) - D.
\end{aligned}$$

It is clear that $G(\cdot)$ is a convex mapping from $L_\infty^w(\mu_{0,n}, M_{rba}(\mathcal{Y}_{0,n}))$ into the real line with the natural ordering $(\mathbb{R}, \preceq) \triangleq \mathbb{Z}$. Also recall that $\vec{q}_{0,n} \rightarrow \mathbb{I}(\mu_{0,n}; \vec{q}_{0,n})$ is convex and well defined on Ω and that, by Theorem III.5, $\inf_{\vec{q}_{0,n} \in \vec{Q}_{0,n}(D)} \mathbb{I}(\mu_{0,n}; \vec{q}_{0,n})$ exists and is finite. Thus,

according to the Lagrange duality theorem referred to above, it suffices to show that there exists a $\vec{q}_{0,n}^1 \in \Omega$ such that

$$G(\vec{q}_{0,n}^1) = \int_{\mathcal{X}_{0,n}} \left\{ \int_{\mathcal{Y}_{0,n}} d_{0,n}(x^n, y^n) \vec{q}_{0,n}^1(dy^n; x^n) \right\} \mu_{0,n}(dx^n) - D < 0.$$

Introduce the sets $A_1 \triangleq \{x^n \in \mathcal{X}_{0,n} : \Gamma_{x^n} \neq \emptyset\}$ and $A_0 \triangleq \mathcal{X}_{0,n} \setminus A_1$, with Γ_{x^n} denoting the x^n -section of Γ . Define the measure valued function $\vec{q}_{0,n}^1$ as follows

$$\vec{q}_{0,n}^1(\Gamma_{x^n}; x^n) = 0, \quad \forall x \in A_0; \quad \vec{q}_{0,n}^1(\mathcal{Y}_{0,n}; x^n) = 1, \quad \forall x^n \in \mathcal{X}_{0,n}$$

$$0 \leq \vec{q}_{0,n}^1(B; x^n) \leq 1, B \subset \Gamma_{x^n}, \quad \vec{q}_{0,n}^1(\Gamma_{x^n}; x^n) = 1, \quad \forall x^n \in A_1$$

where $B \in \mathcal{B}(\mathcal{Y}_{0,n})$. Since by hypothesis $\Gamma \neq \emptyset$ we have $\mu_{0,n}(A_1) > 0$ and thus the kernel $\vec{q}_{0,n}^1$ is well defined and it belongs to $L_\infty^w(\mu_{0,n}, \Pi_{rba}(\mathcal{Y}_{0,n}))$. Using this kernel in the expression for $\ell_{d_{0,n}}(\vec{q}_{0,n})$, one can easily verify that $\ell_{d_{0,n}}(\vec{q}_{0,n}^1) < D$ and hence $G(\vec{q}_{0,n}^1) < 0$. Then, by the Lagrange Duality theory, we arrive at the conclusion of the theorem as stated. Also it follows from the same duality theory that if the infimum is achieved by some $\vec{q}_{0,n}^* \in L_\infty^w(\mu_{0,n}, \Pi_{rba}(\mathcal{Y}_{0,n}))$, then

$$s \left(\int_{\mathcal{X}_{0,n}} \int_{\mathcal{Y}_{0,n}} d_{0,n}(x^n, y^n) \vec{q}_{0,n}^*(dy^n; x^n) \otimes \mu_{0,n}(dx^n) - D \right) = 0. \quad (\text{A.54})$$

In other words, for non-zero $s \in (-\infty, 0]$, solution occurs on the boundary. This completes the proof.

ACKNOWLEDGEMENT

The authors wish to thank the associate editor and anonymous reviewers for their valuable comments which improved significantly the presentation of this paper.

REFERENCES

- [1] P. A. Stavrou and C. D. Charalambous, "Causal rate distortion function and relations to filtering theory," in *20th International Symposium on Mathematical Theory of Networks and Systems (MTNS)*, July 9-13 2012, pp. 1–4 (Extended Abstract). [Online]. Available: <http://www.mtns2012.conference.net.au>
- [2] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech.*, vol. 27, pp. 379–423, July 1948.
- [3] T. Berger, *Rate Distortion Theory: A Mathematical Basis for Data Compression*. Englewood Cliffs, NJ: Prentice-Hall, 1971.
- [4] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. John Wiley & Sons, Inc., Hoboken, New Jersey, 2006.

- [5] S. Yuksel, “A tutorial on quantizer design for networked control systems: stabilization and optimization,” *International Journal of Applied and Computational Mathematics*, vol. 10, no. 3, pp. 365–401, 2011.
- [6] V. Gupta, A. F. Dana, J. P. Hespanha, R. M. Murray, and B. Hassibi, “Data transmission over networks for estimation and control,” *IEEE Transactions on Automatic Control*, vol. 54, no. 8, pp. 1807–1819, Aug. 2009.
- [7] J. S. Freudenberg and R. H. Middleton, “Feedback control performance over a noisy communication channel,” in *IEEE Information Theory Workshop (ITW)*, May 2008, pp. 232–236.
- [8] V. S. Borkar, S. Mitter, and S. Tatikonda, “Markov control problems under communication constraints,” *Communications in Information and Systems*, vol. 1, no. 2, pp. 15–32, Jan. 2001.
- [9] S. Tatikonda and S. Mitter, “Control over noisy channels,” *IEEE Transactions on Automatic Control*, vol. 49, no. 7, pp. 1196–1201, July 2004.
- [10] S. Yuksel and S. P. Meyn, “Random-time, state-dependent stochastic drift for Markov chains and application to stochastic stabilization over erasure channels,” *to appear in IEEE Transactions on Automatic Control*. [Online]. Available: <http://arxiv.org/abs/1104.0862v3>
- [11] R. S. Bucy, “Distortion rate theory and filtering,” *IEEE Transactions on Information Theory*, vol. 28, no. 2, pp. 336–340, Mar. 1982.
- [12] 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.
- [13] —, “Asymptotic behavior of nonanticipative epsilon-entropy for Gaussian processes,” *Problems of Information Transmission*, vol. 27, no. 4, pp. 361–365, 1991.
- [14] J. I. Galdos and D. E. Gustafson, “Information and distortion in reduced-order filter design,” *IEEE Transactions on Information Theory*, vol. 23, no. 2, pp. 183–194, Mar. 1977.
- [15] X. Feng, A. Loparo, and Y. Fang, “Optimal state estimation for stochastic systems: An information theoretic approach,” *IEEE Transactions on Automatic Control*, vol. 42, no. 6, pp. 771–785, June 1997.
- [16] L. Guo, L. Yin, H. Wang, and T. Chai, “Entropy optimization filtering for fault isolation of nonlinear non-Gaussian stochastic systems,” *IEEE Transactions on Automatic Control*, vol. 54, no. 4, pp. 804–810, April 2009.
- [17] D. Teneketzis, “Communication in decentralized control,” Ph.D. dissertation, Mass. Inst. of Tech. (M.I.T.), Cambridge, MA, 1979.
- [18] P. E. Caines, *Linear Stochastic Systems*, ser. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., New York, 1988.
- [19] R. J. Elliott, L. Aggoun, and J. B. Moore, *Hidden Markov Models: Estimation and Control*. Springer-Verlag, Berlin, Heidelberg, New York, 1995.
- [20] P. Dupuis and R. S. Ellis, *A Weak Convergence Approach to the Theory of Large Deviations*. John Wiley & Sons, Inc., New York, 1997.
- [21] R. S. Liptser and A. N. Shiryaev, *Statistics of Random Processes: II. Applications*, 2nd ed. Springer-Verlag, Berlin, Heidelberg, New York, 2001.
- [22] S. Ihara, *Information theory - for Continuous Systems*. World Scientific, 1993.
- [23] H. B. Maynard, “A Radon-Nikodým theorem for finitely additive bounded measures,” *Pacific Journal of Mathematics*, vol. 83, no. 2, pp. 401–413, 1979.
- [24] I. Csiszár, “On an extremum problem of information theory,” *Studia Scientiarum Mathematicarum Hungarica*, vol. 9, pp. 57–71, 1974.

- [25] F. Rezaei, N. U. Ahmed, and C. D. Charalambous, "Rate distortion theory for general sources with potential application to image processing," *International Journal of Applied Mathematical Sciences*, vol. 3, no. 2, pp. 141–165, 2006.
- [26] N. Dunford and J. T. Schwartz, *Linear Operators Part I: General Theory*. John Wiley & Sons, Inc., Hoboken, New Jersey, 1988.
- [27] A. Ionescu Tulcea and C. Ionescu Tulcea, *Topics in the Theory of Lifting*. Springer-Verlag, Berlin, Heidelberg, New York, 1969.
- [28] D. G. Luenberger, *Optimization by Vector Space Methods*. John Wiley & Sons, Inc., New York, 1969.
- [29] R. M. Gray, *Entropy and Information Theory*, 2nd ed. Springer Sciences and Business Media, New York, 2010.
- [30] L. Xie, V. A. Ugrinovskii, and I. R. Petersen, "A duality relationship for regular conditional relative entropy," in *16th International Federation of Automatic Control (IFAC) World Congress*, vol. 16, July 4-8 2005, pt. 1.
- [31] C. D. Charalambous and F. Rezaei, "Stochastic uncertain systems subject to relative entropy constraints: Induced norms and monotonicity properties of minimax games," *IEEE Transactions on Automatic Control*, vol. 52, no. 4, pp. 647–663, April 2007.
- [32] 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.
- [33] F. Rezaei, C. D. Charalambous, and P. A. Stavrou, "Rate distortion function for a class of relative entropy sources," in *19th International Symposium on Mathematical Theory of Networks and Systems (MTNS)*, July 5-9 2010, pp. 1853–1858.
- [34] B. Hassibi, A. H. Sayed, and T. Kailath, *Indefinite-Quadratic Estimation and Control: A Unified Approach to H^2 and H^∞ Theories*. SIAM studies in applied and numerical mathematics, 1999.
- [35] T. Zhou, "Sensitivity penalization based robust state estimation for uncertain linear systems," *IEEE Transactions on Automatic Control*, vol. 55, no. 4, pp. 1018–1024, 2010.
- [36] S. Z. Denic, C. D. Charalambous, and S. M. Djouadi, "Information theoretic bounds for compound MIMO Gaussian channels," *IEEE Transactions on Information Theory*, vol. 55, no. 4, pp. 1603–1617, April 2009.
- [37] 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.
- [38] 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.
- [39] S. C. Tatikonda, "Control over communication constraints," Ph.D. dissertation, Mass. Inst. of Tech. (M.I.T.), Cambridge, MA, 2000.