
This is an electronic reprint of the original article.
This reprint may differ from the original in pagination and typographic detail.

Charalambous, Charalambos D.; Kourtellaris, Christos; Charalambous, Themistoklis; Van Schuppen, Jan H.

Generalizations of Nonanticipative Rate Distortion Function to Multivariate Nonstationary Gaussian Autoregressive Processes

Published in:

Proceedings of the 58th IEEE Conference on Decision and Control, CDC 2019

DOI:

[10.1109/CDC40024.2019.9029859](https://doi.org/10.1109/CDC40024.2019.9029859)

Published: 01/12/2019

Document Version

Peer-reviewed accepted author manuscript, also known as Final accepted manuscript or Post-print

Please cite the original version:

Charalambous, C. D., Kourtellaris, C., Charalambous, T., & Van Schuppen, J. H. (2019). Generalizations of Nonanticipative Rate Distortion Function to Multivariate Nonstationary Gaussian Autoregressive Processes. In *Proceedings of the 58th IEEE Conference on Decision and Control, CDC 2019* (pp. 8190-8195). Article 9029859 (Proceedings of the IEEE Conference on Decision and Control; Vol. 2019-December). IEEE.
<https://doi.org/10.1109/CDC40024.2019.9029859>

This material is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the repository collections is not permitted, except that material may be duplicated by you for your research use or educational purposes in electronic or print form. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone who is not an authorised user.

© 2019 IEEE. This is the author's version of an article that has been published by IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

Generalizations of Nonanticipative Rate Distortion Function to Multivariate Nonstationary Gaussian Autoregressive Processes

Charalambos D. Charalambous, Christos Kourtellaris, Themistoklis Charalambous,
and Jan H. van Schuppen

Abstract—The characterizations of nonanticipative rate distortion function (NRDF) on a finite horizon are generalized to nonstationary multivariate Gaussian order L autoregressive, AR(L), source processes, with respect to mean square error (MSE) distortion functions. It is shown that the optimal reproduction distributions are induced by a reproduction process, which is a linear function of the state of the source, its best mean-square error estimate, and a Gaussian random process.

I. INTRODUCTION

Motivated by applications of communication systems, in which encoders and decoders are required to process information with minimum coding and decoding delay, respectively, and in some cases, in real-time, such as control system applications, Gorbunov and Pinsker [1], [2] introduced the nonanticipatory epsilon entropy of the source subject to either point-wise distortion or average distortion.

The nonanticipatory epsilon entropy of Gauss-Markov processes subject to a point-wise distortion is analyzed extensively in the literature, under various names, such as *sequential, nonanticipative RDF* (see, for example, [3]–[5]). In [3]–[5] various applications are identified, that include control of linear Gaussian control systems over memoryless communication channels with finite transmission rates [3], bounds on the optimal performance theoretically attainable by noncausal and causal codes [4], filtering subject to a fidelity [5], and joint source and channel coding and decoding design that operate in real-time [6]. In view of the difficulty to characterize finite-time NRDF and to compute its value, recently semidefinite programming is proposed to compute numerically its value for multivariate Gauss-Markov sources subject to a point-wise distortion [7]. The characterization of the NRDF for the multivariate Gauss-Markov process with average distortion is recently derived in [8], and includes the optimal realization coefficients. It should be mentioned that the identification of the optimal realization coefficients was unknown since the work of Gorbunov and Pinsker [2]. Hence, [8] completed the characterization of [2, Theorem 5], and gave a dynamic reverse-waterfilling, to find the optimal realization coefficients that turns out to be related to the solution of a certain difference Riccati matrix equation.

C. D. Charalambous and C. Kourtellaris are with the Department of Electrical and Computer Engineering, University of Cyprus, Nicosia, Cyprus. Emails: {chadcha,kourtellaris.christos}@ucy.ac.cy.

T. Charalambous is with the Department of Electrical Engineering and Automation, School of Electrical Engineering, Aalto University, Espoo, Finland. Email: themistoklis.charalambous@aalto.fi.

J. H. van Schuppen is affiliated with the company Van Schuppen Control Research, Amsterdam, Netherlands. Email: jan.h.van.schuppen@xs4all.nl.

Despite the literature on the analysis of nonanticipative epsilon entropy of Gauss-Markov sources (e.g., [3]–[5], [7]), an analysis of the characterization of NRDF which parallels the work found in [2] for multivariate Gaussian autoregressive AR(L) process with point-wise and average distortion functions is missing. The present paper aims to close this gap. Our main results state that the characterization of the NRDF is fundamentally different from that of Gauss-Markov sources.

II. NOTATION

$\mathbb{R} \triangleq (-\infty, \infty)$, $\mathbb{Z} \triangleq \{\dots, -1, 0, 1, \dots\}$, $\mathbb{Z}_0 \triangleq \{0, 1, 2, \dots\}$, $\mathbb{N} \triangleq \{1, 2, \dots\}$, $\mathbb{N}^n \triangleq \{1, \dots, n\}$, $n \in \mathbb{N}$. For any matrix $A \in \mathbb{R}^{p \times m}$, $(p, m) \in \mathbb{N} \times \mathbb{N}$, we denote its transpose by A^T , and for $m = p$, we denote its trace by $\text{tr}(A)$, and by $\text{diag}\{A\}$, the matrix with diagonal entries A_{ii} , $i = 1, \dots, p$, and zero elsewhere. $S_+^{p \times p}$ denotes the set of symmetric positive semidefinite matrices $A \in \mathbb{R}^{p \times p}$, and $S_{++}^{p \times p}$ its subset of positive definite matrices. The statement $A \succeq A'$ (resp. $A \succ A'$) means that $A - A'$ is symmetric positive semidefinite (resp. positive definite). For $x \in \mathbb{R}$, then $\{x\}^+ \triangleq \max\{1, x\}$.

$\{(\mathbb{X}_n, \mathcal{B}(\mathbb{X}_n)) : n \in \mathbb{Z}\}$ denotes a sequence of measurable spaces, where \mathbb{X}_n are confined to complete separable metric spaces or Polish space, and $\mathcal{B}(\mathbb{X}_n)$ the Borel σ -algebras of subsets of \mathbb{X}_n . Points in the product space $\mathbb{X}^{\mathbb{Z}} \triangleq \times_{n \in \mathbb{Z}} \mathbb{X}_n$ are denoted by $x_{-\infty}^{\infty} \triangleq (\dots, x_{-1}, x_0, x_1, \dots) \in \mathbb{X}^{\mathbb{Z}}$, and their restrictions to finite coordinates for any $(m, n) \in \mathbb{Z} \times \mathbb{Z}$ by $x_m^n \triangleq (x_m, \dots, x_n) \in \mathbb{X}_m^n$, $n \geq m$. Hence, $\mathcal{B}(\mathbb{X}^{\mathbb{Z}}) \triangleq \otimes_{t \in \mathbb{Z}} \mathcal{B}(\mathbb{X}_t)$ denotes the σ -algebra on $\mathbb{X}^{\mathbb{Z}}$, and similarly $\mathcal{B}(\mathbb{X}_m^n)$.

Given a random variable (RV) $X : (\Omega, \mathcal{F}) \mapsto (\mathbb{X}, \mathcal{B}(\mathbb{X}))$, we denote by $\mathbf{P}_X(dx) \equiv \mathbf{P}(dx)$ the distribution induced by X on $(\mathbb{X}, \mathcal{B}(\mathbb{X}))$, and by $\mathcal{M}(\mathbb{X})$ the set of probability distributions on \mathbb{X} . Given another RV, $Y : (\Omega, \mathcal{F}) \mapsto (\mathbb{Y}, \mathcal{B}(\mathbb{Y}))$ we denote by $\mathbf{P}_{Y|X}(dy|X = x) \equiv \mathbf{P}(dy|x)$ the conditional distribution of RV Y for a fixed $X = x$.

III. INFORMATION STRUCTURES OF NRDF

This section presents the mathematical formulation and the preliminary Theorem 1, which states: if the source distribution is of L -th order memory, $\mathbf{P}_{X_t|X^{t-1}, Y^{t-1}} = \mathbf{P}_{X_t|X_t^{t-L}, L \in \{1, 2, \dots\}}$, then the optimal reproduction distribution of the NRDF, $R_{0,n}^{na}(D)$, is $\mathbf{P}_{Y_t|Y^{t-1}, X_t^{t-L}, t = 0, \dots, n}$.

¹The subscript notation is often omitted when it is clear from the arguments of the distribution.

Definition 1: (Conditional independence)

Consider three RVs $X : \Omega \rightarrow \mathbb{X}$, $Y : \Omega \rightarrow \mathbb{Y}$, and $Z : \Omega \rightarrow \mathbb{Z}$ defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We say that RVs (X, Y) are conditionally independent given RV Z if $\mathbf{P}_{X|Y,Z} = \mathbf{P}_{X|Z}$ -a.a. $(y, z) \in \mathbb{Y} \times \mathbb{Z}$. •

Definition 2: (Source and reproduction distributions)

Let $x^n \triangleq \{x_0, x_1, \dots, x_n\} \in \mathbb{X}^n \triangleq \times_{i=0}^n \mathbb{X}_i$ denote a sequence generated by the source and let $y^n \triangleq \{y_0, y_1, \dots, y_n\} \in \mathbb{Y}^n \triangleq \times_{i=0}^n \mathbb{Y}_i$ denote its reproduction sequence.

(a) The source generates sequences from the set of distributions that satisfy a conditional independence condition

$$\mathcal{S}_{[0,n]} \triangleq \left\{ \mathbf{P}_{X^n} = \mu(dx_0) \otimes_{t=1}^n S_t(dx_t|x^{t-1}) : \right. \\ \left. \mathbf{P}_{X_t|X^{t-1}, Y^{t-1}} = S_t(dx_t|x^{t-1}) - a.a., \quad t = 1, \dots, n \right\}.$$

where $(x^{t-1}, y^{t-1}) \in \mathbb{X}^{t-1} \times \mathbb{Y}^{t-1}$ and $\mu(dx_0)$ is the initial distribution.

(b) The reproduction sequences are generated from the set of distributions

$$\mathcal{Q}_{[0,n]} \triangleq \left\{ \mathbf{P}_{Y_t|Y^{t-1}, X^t} = Q_t(dy_t|y^{t-1}, x^t), t = 0, \dots, n \right\}, \quad (\text{III.1})$$

where $Q_0(dy_0|y^{-1}, x^0) = Q_0(dy_0|x_0)$. If initial states $x^{-1} \triangleq (\dots, x_{-1}) \in \mathbb{X}^{-1} \triangleq \mathbb{X}_{-\infty}^{-1}$ and $y^{-1} \triangleq (\dots, y_{-1}) \in \mathbb{Y}^{-1} \triangleq \mathbb{Y}_{-\infty}^{-1}$ are available, then (x^n, y^n) are replaced by $((x^{-1}, x^n), (y^{-1}, y^n))$, and $\mathbf{P}_{X^{-1}, Y^{-1}} = \mathbf{P}_{X^{-1}|Y^{-1}} \otimes \nu(dy^{-1})$, where $\nu(dy^{-1})$, is the initial distribution of RV Y^{-1} . Moreover, when such initial states are available, then $\mathcal{S}_{[0,n]}$ and $\mathcal{Q}_{[0,n]}$ are appropriately defined. •

For each $t = 0, 1, \dots$, we introduce the space \mathbb{G}^t of admissible source and reproduction histories up to time t :

$$\mathbb{G}^t \triangleq \mathbb{X}_0 \times \mathbb{Y}_0 \times \dots \times \mathbb{X}_t \times \mathbb{Y}_t, \quad t = 0, 1, \dots \quad (\text{III.2})$$

A typical element of \mathbb{G}^t is $(x^t, y^t) \triangleq (x_0, y_0, \dots, x_t, y_t)$. Given any elements of $\mathcal{S}_{[0,n]}$, $\mathcal{Q}_{[0,n]}$, and an initial distribution $\mathbf{P}_{X_0}(dx_0) \equiv \mu(dx_0)$, by Ionescu-Tulcea theorem, there exists a unique probability measure \mathbf{P}_μ^P on $(\mathbb{G}^\infty, \mathcal{B}(\mathbb{G}^\infty))$, with $\mathbf{P}_\mu^Q(\mathbb{G}^\infty) = 1$, and carrying the sequence of RVs $\{(X_t, Y_t) : t = 0, 1, \dots\}$, defined by

$$\mathbf{P}_\mu^Q(dx_0, dy_0, dx_1, \dots, dy_{n-1}, dx_n, dy_n) = \mu(dx_0) \\ \otimes Q_0(dy_0|x_0) \otimes S_1(dx_1|x_0) \otimes \dots \\ \otimes Q_{n-1}(dy_{n-1}|y^{n-2}, x^{n-1}) \otimes S_n(dx_n|x^{n-1}) \\ \otimes Q_n(dy_n|y^{n-1}, x^n), \quad (\text{III.3})$$

The conditional distribution of Y_t given Y^{t-1} is

$$\mathbf{P}_t^Q(dy_t|y^{t-1}) \triangleq \int_{\mathbb{X}^t} Q_t(dy_t|y^{t-1}, x^t) \otimes S_t(dx_t|x^{t-1}) \\ \otimes \mathbf{P}_t^Q(dx^{t-1}|y^{t-1}), \quad t = 1, \dots, n, \quad (\text{III.4})$$

$$\mathbf{P}_0^Q(dy_0) \triangleq \int_{\mathbb{X}_0} Q_0(dy_0|x_0) \otimes \mu(dx_0). \quad (\text{III.5})$$

For problems with initial states Y^{-1} and $X_{-\infty}^0$, the above distributions should be modified.

Definition 3: (Nonanticipative RDF)

Consider the source and reproduction distributions of Definition 2. The information measure is

$$\mathbf{E}_\mu^Q \left\{ \sum_{t=0}^n \log \left(\frac{Q_t(\cdot|Y^{t-1}, X^t)}{\mathbf{P}_t^Q(\cdot|Y^{t-1})}(Y_t) \right) \right\} \\ = \mathbf{E}_\mu^Q \left\{ \log \left(\frac{Q_0(\cdot|X_0)}{\mathbf{P}_0^Q(\cdot)}(Y_0) \right) \right\} \\ + \mathbf{E}_\mu^Q \left\{ \sum_{t=1}^n \log \left(\frac{Q_t(\cdot|Y^{t-1}, X^t)}{\mathbf{P}_t^Q(\cdot|Y^{t-1})}(Y_t) \right) \right\} \in [0, \infty) \\ \equiv I(X_0; Y_0) + \sum_{t=1}^n I(X^t; Y_t|Y^{t-1}),$$

where $I(X_t; Y_t|Y^{t-1})$ denotes the conditional mutual information between X^t and Y_t conditioned on Y^{t-1} . The NRDF of the source subject to a total distortion is defined by

$$R_{0,n}^{na}(D) \triangleq \inf_{\mathcal{Q}_{[0,n]}(D)} \mathbf{E}_\mu^Q \left\{ \sum_{t=0}^n \log \left(\frac{Q_t(\cdot|Y^{t-1}, X^t)}{\mathbf{P}_t^Q(\cdot|Y^{t-1})}(Y_t) \right) \right\}, \quad (\text{III.6})$$

$$\mathcal{Q}_{[0,n]}(D) \triangleq \left\{ Q_t(dy_t|y^{t-1}, x^t), t = 0, \dots, n : \right. \\ \left. \frac{1}{n+1} \mathbf{E}_\mu^Q \left\{ d_{0,n}(X^n, Y^n) \right\} \leq D \right\} \quad (\text{III.7})$$

where $d_{0,n}(x^n, y^n) \triangleq \sum_{t=0}^n \rho_t(x_t, y_t)$ and $d_{0,n} : \mathbb{X}^n \times \mathbb{Y}^n \rightarrow [0, \infty)$ is a measurable function. If the set $\mathcal{Q}_{[0,n]}(D)$ is empty, then we define $R_{0,n}^{na}(D) = \infty$. •

It should be mentioned that existence of the infimum is shown in [9], under appropriate conditions.

Theorem 1: (Information structures of optimal distributions)

Consider the NRDF of Definition 3.

(a) If the source distribution is Markov, that is, $S_t(dx_t|x_{t-1}), t = 1, \dots, n, \mu_0(dx_0)$ and single-letter distortion is used, $d_{0,n}(x^n, y^n) = \sum_{t=0}^n \rho_t(x_t, y_t)$, then

$$R_{0,n}^{na}(D) = \inf_{\mathcal{Q}_{[0,n]}^M(D)} \mathbf{E}_\mu^{Q^M} \left\{ \sum_{t=0}^n \log \left(\frac{Q_t^M(\cdot|Y^{t-1}, X_t)}{\mathbf{P}_t^{Q^M}(\cdot|Y^{t-1})}(Y_t) \right) \right\}, \quad (\text{III.8})$$

$$\mathcal{Q}_{[0,n]}^M \triangleq \left\{ Q_t(dy_t|y^{t-1}, x_t), t = 0, \dots, n : \right. \\ \left. \frac{1}{n+1} \mathbf{E}_\mu^{Q^M} \left\{ \sum_{t=0}^n \rho_t(x_t, y_t) \right\} \leq D \right\}, \quad (\text{III.9})$$

where the joint and conditional distributions in (III.8) are given by

$$\mathbf{P}^{Q^M}(dx^t, dy^t) \triangleq \mu_0(dx_0) \otimes Q_0(dy_0|x_0) \\ \otimes_{i=1}^t \left(Q_i^M(dy_i|y^{i-1}, x_i) \otimes S_t(dx_i|x_{i-1}) \right), \quad t = 1, \dots, n, \\ \mathbf{P}_t^{Q^M}(dy_t|y^{t-1}) \triangleq \int_{\mathbb{X}^t} Q_t^M(dy_t|y^{t-1}, x_t) \otimes \mathbf{P}_t^{Q^M}(dx_t|y^{t-1}), \\ \mathbf{P}_0^{Q^M}(dx_0, dy_0) \triangleq Q_0(dy_0|x_0) \otimes \mu_0(dx_0), \\ \mathbf{P}_0^{Q^M}(dy_0) \triangleq \int_{\mathbb{X}_0} Q_0^M(dy_0|x_0) \otimes \mu_0(dx_0).$$

(b) If in part (a) the source distribution is replaced by $S_t(dx_t|x_{t-L}^{t-1}), t = L, \dots, n, \mu_t(dx_0, dx_1, \dots, dx_t), t =$

$0, \dots, L-1, L \in \{1, 2, \dots\}$, then the optimal reproduction distribution is of the form

$$Q_t^M(dy_t|y^{t-1}, x_{t-L+1}^t), t = L, \dots, n, \quad (\text{III.10})$$

$$Q_0^M(dy_0|x_0), Q_1^M(dy_1|y_0, x_0, x_1), \dots, \quad (\text{III.11})$$

$$Q_{L-1}^M(dy_{L-1}|y_0, \dots, y_{L-2}, x_0, \dots, x_{L-1}). \quad (\text{III.12})$$

Proof: Due to [8], see also [10]. \blacksquare

IV. THE NRDF OF GAUSSIAN AR(L) PROCESSES SUBJECT TO MSE FIDELITY

We introduce the definitions of time-varying multivariate Gaussian AR(L) processes, for which we derive the main results of this section.

Definition 4: (Multivariate Gaussian AR(L) processes) Consider a tuple of stochastic processes (X^n, Y^n) each of which is \mathbb{R}^p valued defined on some $(\Omega, \mathcal{F}, \mathbb{P})$.

(a) The distribution induced by the process X^n is said to be of memory order L , if it is a subclass of Definition 2.(a), and satisfies

$$\mathbf{P}_{X_t|X^{t-1}, Y^{t-1}} = S_t(dx_t|x_{t-L}^{t-1}) - a.a.(x^{t-1}, y^{t-1}), \quad (\text{IV.1})$$

for $t = 0, \dots, n$, where $X_0 \sim \mu(dx_0)$, and x_{-L+1}^{-1} is assumed to generate the trivial information, i.e., $\sigma\{X_{-L+1}^{-1}\} = \{\Omega, \emptyset\}$; otherwise, the initial distribution of X_{-L+1}^0 is $\mu(dx_{-L+1}^0)$.

(b) The process X^t of part (a), is called Gaussian, of memory order L , if $S_t(dx_t|x_{t-L}^{t-1})$, $t = 0, \dots, n$ are Gaussian, and

$$\mathbf{E}\{X_t|X^{t-1}\} \text{ is linear in } X_{t-L}^{t-1}, \quad t = 0, \dots, n, \quad (\text{IV.2a})$$

$$\text{cov}(X_t, X_t|X^{t-1}) \text{ is nonrandom, } \quad t = 0, \dots, n. \quad (\text{IV.2b})$$

(c) The process X^t of part (b), with $L = 1$, is called a Gauss-Markov process, if its state-space representation is

$$X_t = A_{t-1}X_{t-1} + B_{t-1}W_t, \quad X_0 = x_0, \quad (\text{IV.3})$$

for $t = 1, \dots, n$, where

(i) $A_t \in \mathbb{R}^{p \times p}, B_t \in \mathbb{R}^{p \times q}, t = 1, \dots, n-1$ are non-random matrices;

(ii) $\{W_t : t = 1, \dots, n\}$ is an \mathbb{R}^q -valued sequence of independent Gaussian distributed RVs, $N(0, K_{W_t}), K_{W_t} \in S_+^{q \times q}$;

(iii) $X_0 \in \mathbb{R}^p$ is Gaussian $N(0, K_{X_0})$, independent of W^n .

(d) The process X^t of part (b) is called Gaussian AR(L), if its representation is

$$X_t = \sum_{k=1}^L A_{t,k}X_{t-k} + W_t, \quad t = 0, 1, \dots, \quad L \in \mathbb{N}, \quad (\text{IV.4})$$

$$\sigma\{X_{-L}^{-1}\} = \sigma\{\Omega, \emptyset\}, \quad \text{or } S_0 \triangleq X_{-L}^{-1} = s_0, \quad (\text{IV.5})$$

where $\{W_t : t = 0, \dots, n\}$ is a sequence of independent Gaussian distributed RVs (i.e., $N(0, K_{W_t})$), independent of the RV S_0 , and $A_{t,k} \in \mathbb{R}^{p \times p}$. \bullet

A. The NRDF of Multivariate Gauss-Markov Processes with Average Distortion

The main results of this section are for AR(1), which are included herein to compare the results for the AR(L) source. Specifically, Theorem 2, which identifies sufficient conditions, for a Markov Gaussian joint distribution $\mathbf{P}_{X^n, Y^n}(dx^n, dy^n)$ to achieve the minimum in the definition of $R_{0,n}^{na}(D)$, and the weak realization of the joint processes (X^n, Y^n) . Theorem 3, then characterizes the NRDF, and gives the construction of the corresponding joint distribution $\mathbf{P}_{X^n, Y^n}(dx^n, dy^n)$, and the parametric realization of the joint process (X^n, Y^n) that achieves the characterization. Since some of the statements of Theorem 2 and Theorem 3 are derived also in [8], we omit the proofs herein due to space limitations.

We shall need the following definitions from mean-square estimation theory. The filter estimates $\hat{X}_{t|t} \triangleq \mathbf{E}\{X_t|Y^t\}$, $\hat{X}_{t|t-1} \triangleq \mathbf{E}\{X_t|Y^{t-1}\}$, for $t = 0, \dots, n$, where $\hat{X}_{0|-1} \triangleq \mathbf{E}\{X_0\} = 0$, and error covariances

$$\Sigma_t \triangleq \mathbf{E}\left\{\left(X_t - \hat{X}_{t|t}\right)\left(X_t - \hat{X}_{t|t}\right)^T\right\}, \quad t = 0, \dots, n,$$

$$\Sigma_t^- \triangleq \mathbf{E}\left\{\left(X_t - \hat{X}_{t|t-1}\right)\left(X_t - \hat{X}_{t|t-1}\right)^T\right\}, \quad t = 1, \dots, n.$$

Theorem 2: (Preliminary characterization of $R_{0,n}^{na}(D)$ for a Gauss-Markov processes) Consider the Gauss-Markov process X^n of Definition 4.(b), and the distortion function $d_{0,n}(x^n, y^n) \triangleq \frac{1}{n+1} \sum_{t=0}^n \|X_t - Y_t\|^2$. Assume $R_{0,n}^{na}(D) \in [0, \infty)$ for $D \in [0, D_{max}] \subseteq [0, \infty)$. For any distribution $\mathbf{P}_\mu(dx^n, dy^n)$ induced by the joint process (X^n, Y^n) , the following hold.

(a) Given any arbitrary joint distribution of the joint process (X^n, Y^n) that achieves the minimum of the NRDF $R_{0,n}^{na}(D)$, then there exists a jointly Gaussian distribution defined by

$$\mathbf{P}_\mu^Q(dx^n, dy^n) = \mu_0(dx_0) \otimes Q_0(dy_0|x_0) \otimes_{t=1}^n \left(Q_t(dy_t|y^{t-1}, x_t) \otimes S_t(dx_t|x_{t-1})\right) \quad (\text{IV.6})$$

and induced by the process X^n and the reproduction process

$$Y_t = H_t X_t + g_t(Y^{t-1}) + V_t, \quad t = 0, \dots, n \quad (\text{IV.7a})$$

$$= \begin{cases} H_t X_t + (I - H_t) A_{t-1} \hat{X}_{t-1|t-1} + V_t, & t = 1, \dots, n \\ H_t X_t + V_t, & t = 0 \end{cases} \quad (\text{IV.7b})$$

such that

$$H_t, t = 0, \dots, n \text{ are nonrandom,} \quad (\text{IV.8a})$$

$$g_t(\cdot), t = 1, \dots, n \text{ is a measurable function,} \quad (\text{IV.8b})$$

$$V_t \sim N(0, K_{V_t}), K_{V_t} = K_{V_t}^T \geq 0, t = 0, \dots, n, \quad (\text{IV.8c})$$

$$\forall t = 0, \dots, n, V_t \text{ is independent of } X_0 \text{ and } W_s, \quad (\text{IV.8d})$$

$$s = 0, 1, \dots, t$$

Moreover, the reproduction distribution is parametrized by $(H_t, K_{V_t}), t = 0, \dots, n$, and satisfies

$$Q_t(dy_t|y^{t-1}, x_t) = \mathbf{P}_t(dy_t|y^{t-1}, \hat{x}_{t-1|t-1}, x_t) \quad (\text{IV.9a})$$

$$\equiv Q_t^1(dy_t|\hat{x}_{t-1|t-1}, x_t), \quad t = 1, \dots, n,$$

$$Q_0(dy_0|x_0) \equiv Q_0^1(dy_0|x_0), \quad (\text{IV.9b})$$

while the pay-off satisfies

$$I(X_0; Y_0) + \sum_{t=1}^n I(X^t; Y_t|Y^{t-1})$$

$$= I(X_0; Y_0) + \sum_{t=1}^n I(X_t; Y_t|Y^{t-1}, \hat{X}_{t-1|t-1}), \quad (\text{IV.10a})$$

where

$$\mathbf{P}_t^{Q^1}(dy_t|y^{t-1}) = \int Q_t^1(dy_t|\hat{x}_{t-1|t-1}, x_t) \otimes \mathbf{P}^{Q^1}(dx_t|y^{t-1}),$$

$$t = 1, \dots, n,$$

$$\mathbf{P}_0^{Q^1}(dy_0) = \int Q_0^1(dy_0|x_0) \otimes \mu(dx_0),$$

and the average distortion is given by

$$\mathbf{E}_\mu^{Q^1} \left\{ \sum_{t=0}^n \|X_t - Y_t\|^2 \right\} \leq (n+1)D. \quad (\text{IV.12})$$

(b) For any joint distribution $\mathbf{P}_\mu^{Q^1}(dx^n, dy^n)$ of part (a), then the following inequality holds.

$$I(X_0; Y_0) + \sum_{t=1}^n I(X_t; Y_t|Y^{t-1}, \hat{X}_{t-1|t-1})$$

$$\geq I(X_0; \hat{X}_{0|0}) + \sum_{t=1}^n I(X_t; \hat{X}_{t|t}|Y^{t-1}, \hat{X}_{t-1|t-1}). \quad (\text{IV.13})$$

(c) Consider the statement of part (a). If there exists $(H_t, K_{V_t}) \in \mathbb{R}^{p \times p} \times \mathcal{S}_+^{p \times p}, t = 0, \dots, n$, such that $\hat{X}_{t|t} = Y_t - a.s., t = 0, \dots, n$ then the inequality (IV.13) holds with equality, and the characterization of NRDF is given by

$$R_{0,n}^{na}(D) \triangleq \inf_{Q_{0,n}^1(D)} \left\{ I(X_0; Y_0) + \sum_{t=1}^n I(X_t; Y_t|Y_{t-1}) \right\}$$

$$= \inf_{Q_{0,n}^1(D)} \mathbf{E}_\mu^{Q^1} \left\{ \log \left(\frac{Q_0^1(\cdot|X_0)}{\mathbf{P}_0^{Q^1}(\cdot)}(Y_0) \right) \right.$$

$$\left. + \sum_{t=1}^n \log \left(\frac{Q_t^1(\cdot|Y_{t-1}, X_t)}{\mathbf{P}_t^{Q^1}(\cdot|Y_{t-1})}(Y_t) \right) \right\}, \quad (\text{IV.14})$$

where

$$Q_{0,n}^1(D) \triangleq \left\{ Q_t^1(dy_t|y_{t-1}, x_t), t = 0, \dots, n : \right.$$

$$\left. \frac{1}{n+1} \mathbf{E}_\mu^{Q^1} \left\{ \sum_{t=0}^n \|X_t - Y_t\|^2 \right\} \leq D \right\} \quad (\text{IV.15})$$

and the joint distribution of (X^n, Y^n) is Markov, and it is induced by the representation

$$X_t = A_{t-1}X_{t-1} + B_{t-1}W_t, \quad X_0 = x_0, \quad t = 1, \dots, n,$$

$$Y_t = H_tX_t + (I - H_t)A_{t-1}Y_{t-1} + V_t, \quad t = 1, \dots, n,$$

$$Y_0 = H_0X_0 + V_0.$$

Moreover, the joint distribution of the process (X^n, Y^n) is Gaussian, defined by

$$\mathbf{P}_\mu^{G^1}(dx^n, dy^n) = \mu(dx_0) \otimes Q_0^1(dy_0|x_0)$$

$$\otimes_{t=1}^n \left(Q_t^1(dy_t|y_{t-1}, x_t) \otimes S_t(dx_t|x_{t-1}) \right). \quad (\text{IV.16})$$

In the next theorem, we address Theorem 2.(c), i.e., we identify sufficient conditions such that there exists $(H_t, K_{V_t}) \in \mathbb{R}^{p \times p} \times \mathcal{S}_+^{p \times p}, t = 0, \dots, n$, with the property $\hat{X}_{t|t} = Y_t - a.s., t = 0, \dots, n$, and we give their precise expressions, thus completing the characterization of $R_{0,n}^{na}(D)$.

Theorem 3: (Characterization of $R_{0,n}^{na}(D)$ for a Gauss-Markov processes) Consider the statement Theorem 2.(c). Then, the following hold.

(a) The representation of Y^n , with $(H_t, K_{V_t}), t = 0, \dots, n$, defined below, satisfies $\hat{X}_{t|t} = Y_t - a.s., t = 0, \dots, n$.

$$Y_t = H_tX_t + (I - H_t)A_{t-1}Y_{t-1} + V_t, \quad t = 1, \dots, n,$$

$$= H_tA_{t-1}(X_{t-1} - Y_{t-1}) + A_{t-1}Y_{t-1} + H_tB_{t-1}W_t + V_t,$$

$$Y_0 = H_0X_0 + V_0,$$

where $(H_t, K_{V_t}), t = 0, \dots, n$ are given by

$$H_t \triangleq I - \Sigma_t(\Sigma_t^-)^{-1}, \quad (\text{IV.17a})$$

$$K_{V_t} \triangleq \Sigma_t H_t^T = \Sigma_t - \Sigma_t(\Sigma_t^-)^{-1}\Sigma_t \succeq 0, \quad (\text{IV.17b})$$

$$\Sigma_t^- \triangleq A_{t-1}\Sigma_{t-1}A_{t-1}^T + B_{t-1}K_{W_t}B_{t-1}^T, \quad \Sigma_0^- = K_{X_0}.$$

(b) The representation of Y^n of part (a) induces a distribution $Q^1(dy_t|y_{t-1}, x_t), t = 0, \dots, n$, which achieves the characterization of NRDF $R_{0,n}^{na}(D)$ given by (IV.14)-(IV.15).

(c) The characterization of the NRDF is equivalent to the following optimization problem.

$$R_{0,n}^{na}(D) = \inf_{Q_{0,n}^1(D)} \left\{ I(X_0; Y_0) + \sum_{t=1}^n I(X_t; Y_t|Y_{t-1}) \right\}$$

$$= \inf_{Q_{[0,n]}^1(D)} \left\{ \frac{1}{2} \log \left\{ \frac{|\Sigma_{X_0}|}{|\Sigma_0|} \right\}^+ \right. \quad (\text{IV.18})$$

$$\left. + \frac{1}{2} \sum_{t=1}^n \log \left\{ \frac{|A_{t-1}\Sigma_{t-1}A_{t-1}^T + B_{t-1}K_{W_t}B_{t-1}^T|}{|\Sigma_t|} \right\}^+ \right\}$$

where the constraint set is characterized by

$$\overset{\circ}{Q}_{[0,n]}^1(D) \triangleq \{ \Sigma_t \in \mathcal{S}_+^{p \times p}, t = 0, \dots, n :$$

$$\Sigma_t \preceq A_{t-1}\Sigma_{t-1}A_{t-1}^T + B_{t-1}K_{W_t}B_{t-1}^T,$$

$$\Sigma_0 \preceq K_{X_0}, t = 1, \dots, n, \quad \frac{1}{n+1} \sum_{t=0}^n \text{tr}(\Sigma_t) \leq D \}.$$

Note that Theorem 3.(c), that is, $R_{0,n}^{na}(D)$ given by (IV.18), is the generalization of Gorbunov and Pinsker [1, Example 1] to multivariate sources with total distortion function.

B. The Nonanticipative RDF of Multivariate Gaussian Processes with Arbitrary Memory with average Distortion

Now, we generalize the results of Section IV-A to AR(L) models, i.e., to time-varying multivariate Gaussian processes X^n . We consider a slight variation of Section IV-A, when

$(x_{-L+1}^{-1}, y^{-1}) \in \mathbb{X}_{-L+1}^{-1} \times \mathbb{Y}_{-\infty}^{-1}$ are also available. We define a variant of the NRDF (III.6) by

$$\begin{aligned} & R_{0,n}^{na}(D) \\ & \triangleq \inf_{\mathcal{Q}_{0,n}(D)} \mathbf{E}^Q \left\{ \sum_{t=0}^n \log \left(\frac{Q_t(\cdot | Y_{-\infty}^{t-1}, X_{-L+1}^t)(Y_t)}{\mathbf{P}_t^Q(\cdot | Y_{-\infty}^{t-1})} \right) \right\} \\ & \quad \text{(IV.19)} \\ & = \inf_{\mathcal{Q}_{0,n}(D)} \left\{ I(X_{-L+1}^0; Y_0 | Y_{-\infty}^{-1}) + \sum_{t=1}^n I(X_{-L+1}^t; Y_t | Y_{-\infty}^{t-1}) \right\}, \\ & \quad \text{(IV.20)} \end{aligned}$$

where $X_{-L+1}^t \triangleq (X_{-L+1}, \dots, X_{-1}, X_0, X_1, \dots, X_t)$, $Y_{-\infty}^t \triangleq (\dots, Y_{-1}, Y_0, Y_1, \dots, Y_t)$. The distribution of the initial data is $\mathbf{P}_{X_{-L+1}^{-1}, Y_{-\infty}^{-1}} = \mathbf{P}_{X_{-L+1}^{-1} | Y_{-\infty}^{-1}} \otimes \nu(dy_{-\infty}^{-1})$.

We show the following structural property. The optimal reproduction distribution of the NRDF $R_{0,n}^{na}(D)$ of a Gaussian process X^n , with memory of order L , and MSE distortion $d_{0,n}(x^n, y^n) = \sum_{t=0}^n \|x_t - y_t\|^2$, is of the form

$$Q_t(dy_t | Y_{-\infty}^{t-1}, X_{-L+1}^t) = Q_t^L(dy_t | \mathbf{E}\{S_{t-1} | Y_{-\infty}^{t-1}\}, S_t).$$

where S_t is the state of the source process. That is, at each time $t = 0, \dots, n$ the conditional distribution depends only on the estimate of the state $\hat{S}_{t-1|t-1} = \mathbf{E}\{S_{t-1} | Y_{-\infty}^{t-1}\}$ and S_t . For any finite integer $L \in \{1, \dots\}$, we derive the main theorem, by first introducing, a state space representation for

$$\begin{aligned} S_t & \triangleq \text{vector} \left(S_t^1, S_t^2, \dots, S_t^L \right) \\ & = \text{vector} \left(X_t, \dots, X_{t-L+1} \right), \quad t = 0, \dots, n. \end{aligned} \quad \text{(IV.21)}$$

as follows:

$$S_t = A_{t-1}S_{t-1} + B_{t-1}W_t, \quad S_0 = s_0, \quad t = 1, \dots, n, \quad \text{(IV.22)}$$

$$X_t = C_t S_t, \quad \text{(IV.23)}$$

$$A_t \triangleq \begin{pmatrix} A_{t,1} & A_{t,2} & \dots & A_{t,L} \\ I & 0 & \dots & 0 \\ 0 & I & \dots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 \dots & I & 0 \end{pmatrix}, \quad B_t \triangleq \begin{pmatrix} I \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \text{(IV.24)}$$

$$C_t \triangleq (I \ 0 \ \dots \ 0). \quad \text{(IV.25)}$$

for some non-random matrices (A_t, B_t, C_t) , where $W_t \sim N(0, K_{W_t})$ is an independent Gaussian sequence, independent of $S_0 \sim N(0, K_{S_0})$.

Next, we state the main theorem.

Theorem 4: (Characterization of $R_{0,n}^{na}(D)$ for Gaussian processes with memory of order L , and MSE distortion)

Consider the time-varying multivariate Gaussian process X^n , AR(L), and MSE distortion, of Definition 4.(b), and Gaussian distribution $\mu(dx_{-L+1}^0)$. Assume the infimum in $R_{0,n}^{na}(D)$ defined by (IV.19) exists, i.e., $R_{0,n}^{na}(D) \in [0, \infty)$ for some $D < D_{max} \subseteq [0, \infty)$. The following hold.

(a) For any finite integer $L \in \{1, \dots\}$, then (IV.21)-(IV.25)

is a representation of the process X_{-L+1}^t . Moreover, the process S^t is Markov, i.e.,

$$\mathbf{P}_t(ds_t | s^{t-1}) = \mathbf{P}_t(ds_t | s_{t-1}), \quad t = 1, \dots, n. \quad \text{(IV.26)}$$

(b) The joint distribution that achieves the infimum of the nonanticipative RDF defined by (IV.19) is jointly Gaussian given by

$$\begin{aligned} \mathbf{P}_\mu^Q(dx_{-L+1}^n, dy_{-\infty}^n) & = \mathbf{P}_\mu(ds_0, dy_{-\infty}^{-1}) \otimes Q_0(dy_0 | s_0, y_{-\infty}^{-1}) \\ & \quad \otimes_{t=1}^n \left(Q_t(dy_t | y_{-\infty}^{t-1}, s_t) \otimes S_t(dx_t | s_{t-1}) \right) \end{aligned} \quad \text{(IV.27)}$$

and induced by process S^n and reproduction process $Y_{-\infty}^n$

$$Y_t = H_t S_t + g_t^*(Y_{-\infty}^{t-1}) + V_t, \quad t = 0, \dots, n \quad \text{(IV.28)}$$

$$\stackrel{(a)}{=} H_t S_t + (C_t - H_t) A_{t-1} \hat{S}_{t-1|t-1} + V_t \quad \text{(IV.29)}$$

$$\stackrel{(b)}{=} H_t A_{t-1} (S_{t-1} - \hat{S}_{t-1|t-1}) + C_t A_{t-1} \hat{S}_{t-1|t-1} + H_t B_{t-1} W_t + V_t, \quad \text{(IV.30)}$$

where (a) is due to

$$g_t^*(Y_{-\infty}^{t-1}) = (C_t - H_t) A_{t-1} \hat{S}_{t-1|t-1}, \quad \text{(IV.31)}$$

(b) is due to the substitution of S_t , i.e., (IV.22), and

$$\text{i) } H_t, t = 0, \dots, n \text{ are non-random,} \quad \text{(IV.32)}$$

$$\text{ii) } V_t \sim N(0, K_{V_t}), K_{V_t} = K_{V_t}^\top \succeq 0, t = 0, \dots, n, \quad \text{(IV.33)}$$

$$\text{iii) } V_t \text{ is independent of } W_t, t = 0, \dots, n \text{ and } S_0. \quad \text{(IV.34)}$$

Further, the reproduction distribution (parametrized by $(H_t, K_{V_t}), t = 0, \dots, n$), satisfies

$$\begin{aligned} Q_t(dy_t | y_{-\infty}^{t-1}, s_t) & = \mathbf{P}_t(dy_t | y_{-\infty}^{t-1}, \hat{s}_{t-1|t-1}, s_t) \\ & \equiv Q_t^L(dy_t | \hat{s}_{t-1|t-1}, s_t), \quad t = 0, \dots, n, \end{aligned} \quad \text{(IV.35)}$$

and the pay-off in (IV.19) is expressed as

$$\mathbf{E}^Q \left\{ \sum_{t=0}^n \log \left(\frac{Q_t(\cdot | Y_{-\infty}^{t-1}, X_{-L+1}^t)(Y_t)}{\mathbf{P}_t^Q(\cdot | Y_{-\infty}^{t-1})} \right) \right\} \quad \text{(IV.36)}$$

$$= \mathbf{E}^{Q^L} \left\{ \sum_{t=0}^n \log \left(\frac{Q_t^L(\cdot | \hat{S}_{t-1|t-1}, S_t)}{\mathbf{P}_t^{Q^L}(\cdot | Y_{-\infty}^{t-1})} \right) \right\} \quad \text{(IV.37)}$$

$$= \sum_{t=0}^n I(S_t; Y_t | Y_{-\infty}^{t-1}), \quad \text{(IV.38)}$$

where the conditional distribution of Y_t given $Y_{-\infty}^{t-1}$ is

$$\begin{aligned} \mathbf{P}_t^{Q^L}(dy_t | y_{-\infty}^{t-1}) & = \int Q_t(dy_t | \hat{s}_{t-1|t-1}, s_t) \\ & \quad \otimes \mathbf{P}^{Q^L}(ds_t | y_{-\infty}^{t-1}), \quad t = 0, \dots, n. \end{aligned} \quad \text{(IV.39)}$$

Moreover, the following inequality holds

$$\begin{aligned} \sum_{t=0}^n I(S_t; Y_t | Y_{-\infty}^{t-1}) & = \sum_{t=0}^n I(S_t; Y_t | Y_{-\infty}^{t-1}, \hat{S}_{t-1|t-1}) \\ & \geq \sum_{t=0}^n I(S_t; \hat{X}_{t|t} | Y_{-\infty}^{t-1}, \hat{S}_{t-1|t-1}), \end{aligned} \quad \text{(IV.40)}$$

and it is achieved if $\widehat{X}_{t|t} = Y_t - a.s.$

(c) In part (b) the information measure $\sum_{t=0}^n I(S_t; Y_t | Y_{-\infty}^{t-1})$, i.e., (IV.37) is given by

$$\begin{aligned} & \sum_{t=0}^n I(S_t; Y_t | Y_{-\infty}^{t-1}) \\ &= \frac{1}{2} \sum_{t=0}^n \log \left\{ \frac{|\Sigma_t^Y|}{|H_t B_{t-1} K_{W_t} (H_t B_{t-1})^T + K_{V_t}|} \right\}^+ \\ &= \frac{1}{2} \sum_{t=0}^n \log \left\{ \frac{|\Sigma_t^-|}{|\Sigma_t|} \right\}^+, \end{aligned}$$

where

$$\Sigma_t^Y = \text{cov}(Y_t, Y_t | Y_{-\infty}^{t-1}) = H_t A_{t-1} \Sigma_{t-1} (H_t A_{t-1})^T + H_t B_{t-1} K_{W_t} (H_t B_{t-1})^T + K_{V_t}, \quad (\text{IV.41})$$

$$\Sigma_t \triangleq \mathbf{E} \left\{ (S_t - \widehat{S}_{t|t}) (S_t - \widehat{S}_{t|t})^T \right\}, \quad (\text{IV.42})$$

$$\Sigma_t^- \triangleq \mathbf{E} \left\{ (S_t - \widehat{S}_{t|t-1}) (S_t - \widehat{S}_{t|t-1})^T \right\}, \quad (\text{IV.43})$$

and Σ_t satisfies the Kalman-filter Riccati equation for estimating S_t from $Y_{-\infty}^t$, and similarly Σ_t^- .

The average distortion constraint is

$$\mathbf{E} \left\{ d_{0,n}(X^n, Y^n) \right\} = \mathbf{E} \left\{ \sum_{t=0}^n \|C_t S_t - Y_t\|^2 \right\} \leq D(n+1) \quad (\text{IV.44})$$

(d) The characterization of NRDF is given by the following optimization problem.

$$R_{0,n}^{na}(D) = \inf_{(H_t, K_{V_t}), t=0, \dots, n: (\text{IV.44}) \text{ holds}} \sum_{t=0}^n I(S_t; Y_t | Y_{-\infty}^{t-1}).$$

and the relation between (H_t, K_{V_t}) is found from the condition $\widehat{X}_{t|t} = Y_t - a.s.$, which ensure the lower bound (IV.40) is achieved.

Proof: The derivation is based on the techniques of Theorem 2 and Theorem 3.

(a) The process X^n is not Markov; however, $\{S_t : t = 0, \dots, n\}$ defined by (IV.22) is Markov, as easily shown by an application of Bayes' theorem. The state-space representation of $\{S_t : t = 0, \dots, n\}$ is one way to represent X^n .

(b) Note that by a slight variation of Theorem 1, to account for the initial data $(X_{-L+1}^0, Y_{-\infty}^{-1})$, the minimization over $Q_t(dy_t | y_{-\infty}^{t-1}, x_{-L+1}^t, t = 0, \dots, n$ in (IV.19) is of the form $Q_t^L(dy_t | y_{-\infty}^{t-1}, x_{-L+1}^t, t = 0, \dots, n$. Further, since the RVs $(X_{-L+1}^0, Y_{-\infty}^{-1})$ are jointly Gaussian, the minimization in (IV.19) occurs in the set of jointly Gaussian distributions defined by (IV.27). The rest of the statements (IV.28)-(IV.39) are shown by following Theorem 2, (a).

(c) This is simply an evaluation of the information measure and average distortion using part (b).

(d) This follows from part (c) and (IV.40). \blacksquare

In the next remark we state some observations regarding Theorem 4.

Remark 1: Discussion on Theorem 4

(a) There is a clear and fundamental difference between

Theorem 4, which treats AR(L) sources, and analogous results for AR(1).

(b) Whether the optimization of Theorem 4 can be further simplified, is not a subject of analysis in this paper.

V. CONCLUSIONS

We generalized the NRDF to nonstationary, multivariate Gaussian process of memory order L , with MSE distortion. Characterizations of the NRDF and corresponding optimal reproduction distributions, and their realizations are obtained, and shown to depend on the state of the source and its mean-square error estimate.

ACKNOWLEDGEMENTS

This work was supported in parts by the European Regional Development Fund and the Republic of Cyprus through the Research Promotion Foundation (Project: EXCELLENCE/1216/0365). The work of T. Charalambous was supported by the Academy of Finland under Grant 317726.

REFERENCES

- [1] 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, 1973.
- [2] A. K. Gorbunov and M. S. Pinsker, "Prognostic epsilon entropy of a Gaussian message and a Gaussian source," *Problems of Information Transmission*, vol. 10, no. 2, pp. 93–109, 1974.
- [3] S. Tatikonda, A. Sahai, and S. Mitter, "Stochastic linear control over a communication channel," *IEEE Transactions on Automatic Control*, vol. 49, pp. 1549–1561, 2004.
- [4] M. S. Derpich and J. Østergaard, "Improved upper bounds to the causal quadratic rate-distortion function for Gaussian stationary sources," *IEEE Transactions on Information Theory*, vol. 58, pp. 3131–3152, May 2012.
- [5] C. D. Charalambous, P. A. Stavrou, and N. U. Ahmed, "Nonanticipative rate distortion function and relations to filtering theory," *IEEE Transactions on Automatic Control*, vol. 59, pp. 937–952, April 2014.
- [6] C. K. Kourtellis, C. D. Charalambous, and J. J. Boutros, "Nonanticipative transmission for sources and channels with memory," in *IEEE International Symposium on Information Theory (ISIT)*, pp. 521–525, June 2015.
- [7] T. Tanaka, K. K. Kim, P. A. Parrilo, and S. K. Mitter, "Semidefinite programming approach to Gaussian sequential rate-distortion trade-offs," *IEEE Transactions on Automatic Control*, vol. 62, pp. 1896–1910, April 2017.
- [8] P. A. Stavrou, T. Charalambous, C. D. Charalambous, and S. Loyka, "Optimal estimation via nonanticipative rate distortion function for time-varying Gauss-Markov processes," *SIAM Journal on Control and Optimization (SICON)*, vol. 56, no. 5, pp. 3731–3765, 2018.
- [9] C. D. Charalambous and P. A. Stavrou, "Directed information on abstract spaces: Properties and variational equalities," *IEEE Transactions on Information Theory*, vol. 62, pp. 6019–6052, Nov 2016.
- [10] C. D. Charalambous and P. A. Stavrou, "Optimization of directed information and relations to filtering theory," in *European Control Conference (ECC)*, pp. 1385–1390, June 2014.