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

Structural Properties of Nonanticipatory Epsilon Entropy of Multivariate Gaussian Sources

Published in:
Proceedings of the IEEE International Symposium on Information Theory, ISIT 2020

DOI:
10.1109/ISIT44484.2020.9174319

Published: 01/01/2020

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

Please cite the original version:
Charalambous, C. D., Charalambous, T., Kourtellaris, C., & Van Schuppen, J. H. (2020). Structural Properties of Nonanticipatory Epsilon Entropy of Multivariate Gaussian Sources. In *Proceedings of the IEEE International Symposium on Information Theory, ISIT 2020* (pp. 2867-2872). Article 9174319 (IEEE International Symposium on Information Theory). IEEE. https://doi.org/10.1109/ISIT44484.2020.9174319

Structural Properties of Nonanticipatory Epsilon Entropy of Multivariate Gaussian Sources

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

Abstract— **The complete characterization of the Gorbunov and Pinsker [1], [2] nonanticipatory epsilon entropy of multivariate Gauss-Markov sources with square-error fidelity is derived, which remained an open problem since 1974. Specifically, it is shown that the optimal matrices of the stochastic realization of the optimal test channel or reproduction distribution, admit spectral representations with respect to the same unitary matrices, and that the optimal reproduction process is generated, subject to pre-processing and post-processing by memoryless parallel additive Gaussian noise channels. The derivations and analyses are new and bring out several properties of such optimization problems over the space of conditional distributions and their realizations.**

I. Introduction

Motivated by a wide range of applications in which information is required to be transferred in real time or with low latency, Gorbunov and Pinsker [1], [2] introduced the nonanticipatory epsilon entropy and message generation rates of sources with respect to a fidelity criterion. This is a variant of the rate distortion function (RDF) and its rate of lossy compression [3]–[5], i.e., the optimal performance theoretically attainable (OPTA) by noncausal codes, with an additional causality constraint imposed on the optimal reproduction distribution (i.e., the test channel distribution.

Despite the broad range of applications of nonanticipatory epsilon entropy, often called nonanticipative rate distortion function (NRDF), in causal and zero-delay codes, and control over limited rate channels (e.g., [6]–[13]), with focus on Gauss-Markov sources, its characterization for multivariate Gaussian sources is currently not known. In particular, the only complete characterizations of nonanticipatory epsilon entropy known today, are for scalar Gaussian Markov sources with pointwise square-error fidelity due to Gorbunov and Pinsker [2, Example 2], and the scalar Gaussian Markov source with total square-error fidelity presented in [10]. For multivariate Gauss-Markov sources with pointwise square-error fidelity a partial characterization of nonanticipatory epsilon entropy is given by Gorbunov and Pinsker in [2, Theorem 5]. Specifically, the properties of the matrices that define the realization, and induce the optimal test channel distribution are not identified in [2, Theorem 5], and remained to this date unknown. In view of the difficulties to characterize and compute the nonanticipatory epsilon entropy for multivariate Gauss-Markov sources subject to a point-wise fidelity, numerical algorithms are developed based on semidefinite programming [9]. However, the technical question whether the numerical algorithm produces a valid realization of the optimal test channel distribution is not apparent. Another attempt reported recently in [11] for the multivariate Gauss-Markov process with average fidelity, led to characterization of the nonanticipatory epsilon entropy, i.e., the NRDF, through Kuhn-Tucker conditions, via the solution of a certain difference algebraic Riccati equation. However, due to the lack of knowledge of fundamental structural properties of the realization that induces the optimal test channel distribution, a suboptimal water-filling solution is proposed, that corresponds to an upper bound on the NRDF.

This paper provides for the first time, the complete characterization of the problem posed and partly solved by Gorbunov and Pinsker [2], for Gauss-Markov sources subject to a point-wise square error fidelity. What is shown in this paper is the structural property that states: the matrices that define the realization of the reproduction process that induces the optimal test channel distribution admit a spectral representations with respect to the same unitary matrices. It then follows the optimal reproduction process is realized, subject to a pre-processing and post-processing, by memoryless parallel additive Gaussian noise channels. This property further implies the *rate loss of causal and zero delay codes with respect to noncausal codes* derived in [7] can be computed explicitly, and the optimal realization can be transformed into an encoder-decoder pair based on subtractive dither with uniform scalar quantization (SDUSQ) and lattice codes, in complete analogy to the RDF of scalar Gaussian random variables [14]–[16].

II. Notation and Preliminaries

Notation. $\mathbb{R} \triangleq (-\infty, \infty)$, $\mathbb{Z} \triangleq \{\ldots, -1, 0, 1, \ldots\}$, $\mathbb{Z}_0 \triangleq \{0, 1, 2, \ldots\}$, $\mathbb{N} \triangleq \{1, 2, \ldots\}$, $\mathbb{N}^n \triangleq \{1, \ldots, n\}$, $n \in \mathbb{N}$. For any matrix $A \in \mathbb{R}^{p\times m}, (p, m) \in \mathbb{N} \times \mathbb{N}$, we denote its transpose by $A^{\intercal}$, and for $m = p$, we denote its trace by $\mathrm{tr}(A)$, and the matrix with diagonal entries $A_{ii},\ i = 1, \ldots, p$ and zero elsewhere by $\mathrm{diag}\{A\}$. $\mathcal{S}_{+}^{p\times p}$ denotes the set of symmetric positive semidefinite matrices $A \in \mathbb{R}^{p\times p}$, and $\mathcal{S}_{++}^{p\times p}$ 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 $x \in \mathbb{R}$, we

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.

define $\{x\}^+ \triangleq \max\{1, x\}$.

Definition 1 (Multivariate Gauss-Markov Source). *A multivariate Gauss-Markov process is an $\mathbb{R}^p$-valued random process $X_t : \Omega \to \mathbb{R}^p, t = 0, \ldots, n$ defined by the recursion*

$$X_t = A_{t-1}X_{t-1} + B_{t-1}W_t, \; X_0 = x_0, \; t = 1, \ldots, n \quad (1)$$

where (i) $A_t \in \mathbb{R}^{p\times p}, B_t \in \mathbb{R}^{p\times q}$ *are non-random matrices;* (ii) $\{W_t : t = 1, \ldots, n\}$ *is an $\mathbb{R}^q$-valued independent Gaussian process, $N(0, K_{W_t})$, $K_{W_t} \succeq 0$, independent of X_0;* (iii) $X_0 \in \mathbb{R}^p$ *is Gaussian $N(0, K_{X_0}), K_{X_0} \succeq 0$.*

We evaluate the reproduction $Y^n \triangleq \{Y_0, Y_1, \ldots, Y_n\}$, $Y_t : \Omega \to \mathbb{Y}_t \triangleq \mathbb{R}^p$, of $X^n \triangleq \{X_0, X_1, \ldots, X_n\}$, with respect to the mean-square error of the total square-error fidelity, $d_{0,n}(\cdot,\cdot)$, defined by $\frac{1}{n+1}\mathbf{E}\Big\{d_{0,n}(X^n, Y^n)\Big\} \leq D$, $d_{0,n}(X^n, Y^n) \triangleq \sum_{t=0}^n ||X_t - Y_t||^2_{\mathbb{R}^p}$, $D \in [0, \infty)$. We also consider the special case of pointwise mean square-error, $\mathbf{E}\Big\{d_t(X_t, Y_t)\Big\} \leq D_t$, $D_t \in [0, \infty)$, $d_t(X_t, Y_t) \triangleq ||X_t - Y_t||^2_{\mathbb{R}^p}$, $t = 0, \ldots, n$.

The next well-known proposition of conditionally Gaussian RVs is extensively used in our derivations.

Proposition 1. *Consider a pair of* RVs $X = (X_1, \ldots, X_k)^T : \Omega \to \mathbb{R}^k$ *and* $Y = (Y_1, \ldots, Y_l)^T : \Omega \to \mathbb{R}^l$, *defined on some probability distribution $\Big(\Omega, \mathcal{F}, \mathbb{P}\Big)$. Let $\mathcal{G} \subseteq \mathcal{F}$ be a sub$-\sigma-$algebra. Assume the conditional distribution of (X, Y) conditioned on $\mathcal{G}$, i.e., $\mathbf{P}(dx, dy|\mathcal{G})$ is $\mathbb{P}-$*a.s. *(almost surely) Gaussian, with conditional means*

$$\mu_{X|\mathcal{G}} \triangleq \mathbf{E}\left\{X\Big|\mathcal{G}\right\}, \quad \mu_{Y|\mathcal{G}} \triangleq \mathbf{E}\left\{Y\Big|\mathcal{G}\right\} \quad (2)$$

*and conditional covariances $Q_{X,X|\mathcal{G}} \triangleq \mathrm{cov}\Big(X, X\Big|\mathcal{G}\Big)$, $Q_{X,Y|\mathcal{G}} \triangleq \mathrm{cov}\Big(X, Y\Big|\mathcal{G}\Big)$, $Q_{Y,Y|\mathcal{G}} \triangleq \mathrm{cov}\Big(Y, Y\Big|\mathcal{G}\Big)$. Then, the vectors of conditional expectations $\mu_{X|Y,\mathcal{G}} \triangleq \mathbf{E}\left\{X\Big|Y, \mathcal{G}\right\}$ and matrices of conditional covariances $Q_{X,X|Y,\mathcal{G}} \triangleq \mathrm{cov}\left(X, X\Big|Y, \mathcal{G}\right)$ are given, $\mathbb{P}-$*a.s., *by*[1]*:*

$$\mu_{X|Y,\mathcal{G}} = \mu_{X|\mathcal{G}} + Q_{X,Y|\mathcal{G}}Q^{-1}_{Y,Y|\mathcal{G}}\Big(Y - \mu_{Y|\mathcal{G}}\Big), \quad (3)$$

$$Q_{X,X|Y,\mathcal{G}} \triangleq Q_{X,X|\mathcal{G}} - Q_{X,Y|\mathcal{G}}Q^{-1}_{Y,Y|\mathcal{G}}Q^T_{X,Y|\mathcal{G}}. \quad (4)$$

If $\mathcal{G}$ is the trivial information, i.e., $\mathcal{G} = \{\Omega, \emptyset\}$, then $\mathcal{G}$ is removed from the above expressions, and (3), (4) degenerate to the well-known conditional mean.

From Proposition 1 then follows Theorem 1 (below), which we apply to identify a realization of the test channel.

Theorem 1 (Equivalent statements). *Consider the statement of Proposition 1. Any two of the following three conditions imply the third:*

$$\textit{Condition 1:} \;\; Q_{X,Y|\mathcal{G}}Q^{-1}_{Y,Y|\mathcal{G}} = I. \quad (5)$$

$$\textit{Condition 2:} \;\; \mu_{X|\mathcal{G}} = \mu_{Y|\mathcal{G}}. \quad (6)$$

$$\textit{Condition 3:} \;\; \mu_{X|Y,\mathcal{G}} = Y - \text{a.s.}. \quad (7)$$

[1]If the inverse $Q^{-1}_{X,Y|\mathcal{G}}$ does not exists then it is replaced by the pseudo inverse $Q^{\dagger}_{X,Y|\mathcal{G}}$.

In subsequent sections, we make use of the following definitions of conditional means and error covariances: $\widehat{X}_{t|t} \triangleq \mathbf{E}_\mu\{X_t\Big|Y^t\}$, $\widehat{X}_{t|t-1} \triangleq \mathbf{E}_\mu\left\{X_t\Big|Y^{t-1}\right\}$, $\Sigma_t \triangleq \mathbf{E}_\mu\left\{\Big(X_t - \widehat{X}_{t|t}\Big)\Big(X_t - \widehat{X}_{t|t}\Big)^T\right\}$, $\Sigma_t^- \triangleq \mathbf{E}_\mu\left\{\Big(X_t - \widehat{X}_{t|t-1}\Big)\Big(X_t - \widehat{X}_{t|t-1}\Big)^T\right\}$, with initial conditions $\widehat{X}_{0|-1} \triangleq \mathbf{E}_\mu\{X_0\} = 0$ and $\Sigma_0^- \triangleq \mathbf{E}_\mu\left\{\Big(X_0 - \widehat{X}_{0|-1}\Big)\Big(X_0 - \widehat{X}_{0|-1}\Big)^T\right\}$.

III. The Nonanticipative RDF of Multivariate Gauss-Markov Sources Subject to MSE Fidelity

A. Characterization of $R^{na}_{0,n}(D)$ of Multivariate Gauss-Markov Sources with Total MSE Distortion

Theorem 2 ($R^{na}_{0,n}(D)$ of Gauss-Markov processes with total MSE distortion [11]). *Consider the Gauss-Markov process $X^n \triangleq \{X_0, \ldots, X_n\}$ of Definition 1 and a total MSE distortion function $d_{0,n}(x^n, y^n) \triangleq \sum_{t=0}^n ||X_t - Y_t||^2$. Assume $R^{na}_{0,n}(D) \in [0, \infty)$ for $D \in [0, D_{\max}) \subseteq [0, \infty)$. Then, the following hold.*
(a) The distribution of the joint process (X^n, Y^n) that achieves the minimum of the nonanticipative RDF $R^{na}_{0,n}(D)$ is jointly Gaussian distribution and it is induced by the process X^n (of Definition 1) and the reproduction process

$$Y_t = \begin{cases} H_0X_0 + V_0, & t = 0, \\ H_tX_t + \Big(I - H_t\Big)A_{t-1}\widehat{X}_{t-1|t-1} + V_t, & \textit{otherwise}, \end{cases}$$

where $H_t, t = 0, \ldots, n$ are nonrandom, $V_t \sim N(0, K_{V_t}), K_{V_t} = K^T_{V_t} \succeq 0, t = 0, \ldots, n$, is independent and Gaussian, $V_t, t = 0, \ldots, n$, is independent of $W_t, t = 1, \ldots, n$, and X_0, and

$$\Sigma_t = \Sigma_t^- - \Sigma_t^- H_t^T\left[H_t\Sigma_t^- H_t^T + K_{V_t}\right]^{-1}\left(\Sigma_t^- H_t^T\right)^T, \quad (8)$$

for $t = 0, \ldots, n$ and $\Sigma_0^- = K_{X_0}$, and

$$\Sigma_t^- = A_{t-1}\Sigma_{t-1}A^T_{t-1} + B_{t-1}K_{W_t}B^T_{t-1}, \quad (9)$$

for $t = 1, \ldots, n$. Moreover, $\widehat{X}_{t|t}$ satisfies the Kalman-filter recursion

$$\begin{aligned}\widehat{X}_{t|t} &= A_{t-1}\widehat{X}_{t-1|t-1} + \Sigma_t^- H_t^T\Big(H_t\Sigma_t^- H_t^T + K_{V_t}\Big)^{-1} \quad (10)\\ &.\Big(Y_t - H_tA_{t-1}\widehat{X}_{t-1|t-1} - \Big(I - H_t\Big)A_{t-1}\widehat{X}_{t-1|t-1}\Big),\\ &\equiv f_t(\widehat{X}_{t-1|t-1}, Y_t), \quad t = 1, \ldots, n, \;\; \widehat{X}_{0|0} = \textit{given}.\end{aligned}$$

(b) The nonanticipative RDF is given by

$$R^{na}_{0,n}(D) = \inf_{\mathcal{Q}^{H,K_V}_{[0,n]}(D)} \frac{1}{2}\sum_{i=0}^n \log\left\{\frac{|\Sigma_t^-|}{|\Sigma_t|}\right\}^+ \quad (11)$$

$$= \inf_{\mathcal{Q}^{H,K_V}_{[0,n]}(D)} \frac{1}{2}\sum_{i=0}^n \log\left\{\frac{|H_t\Sigma_t^- H_t^T + K_{V_t}|}{|K_{V_t}|}\right\}^+, \quad (12)$$

where the average distortion constraint is

$$\mathcal{Q}^{H,K_V}_{[0,n]}(D) \triangleq \Big\{H_t \in \mathbb{R}^{p\times p},\ K_{V_t} \in \mathcal{S}_+^{p\times p}, t=0,\ldots,n:$$
$$\mathbf{E}^Q_\mu\Big\{\sum_{t=0}^n ||X_t - Y_t||^2_{\mathbb{R}^p}\Big\} = \tag{13}$$
$$\sum_{t=0}^n \text{tr}\Big((I-H_t)\Sigma_t^-(I-H_t)^T + K_{V_t}\Big) \le (n+1)D\Big\}.$$

Theorem 3 (below) is obtained by invoking Hadamard's inequality to identify a fundamental *structural property* of the realization coefficients $\{H_t, K_{V_t}, \Sigma_t, \Sigma_t^-\}$ of Theorem 2.

Theorem 3 (Structural properties of realization of Theorem 2)**.** *Consider the characterization of nonanticipative RDF given in Theorem 2. Then, the following hold.*
(a) For any element of $\mathcal{Q}^{H,K_V}_{[0,n]}(D)$, i.e, $(\Sigma_t^-, \Sigma_t, H_t, K_{V_t})$, then

$$\frac{1}{2}\sum_{t=0}^n \log\frac{|\Sigma_t^-|}{|\Sigma_t|} = -\frac{1}{2}\sum_{t=0}^n \log\Big|I - \Lambda_t\Big[\Lambda_t + U_t^T K_{V_t} U_t\Big]^{-1}\Big|$$
$$\ge -\frac{1}{2}\sum_{t=0}^n \log\prod_{i=1}^p \Big(1 - \lambda_{t,i}\Big[\lambda_{t,i} + \Big(U_t^T K_{V_t} U_t\Big)_{ii}\Big]^{-1}\Big), \tag{14}$$

where Λ_t is the diagonal matrix of the spectral representation of $H_t\Sigma_t^- H_t^T \succeq 0$, that is, $H_t\Sigma_t^- H_t^T = U_t\Lambda_t U_t^T$, $\Lambda_t = \text{diag}\{\lambda_{t,1},\ldots,\lambda_{t,p}\}$, $U_tU_t^T = I$, $U_t^TU_t = I$, $\lambda_{t,1} \ge \lambda_{t,2} \ge \ldots \ge \lambda_{t,p}$, $t=0,\ldots,n$.
(b) The inequality in (14) holds with equality if and only if the spectral representation of $K_{V_t} \succeq 0$ is expressed w.r.t. the same unitary matrix of $H_t\Sigma_t^- H_t^T \succeq 0$, for $t=0,\ldots,n$, i.e.,

$$K_{V_t} = U_t D_{K_{V_t}} U_t^T, \quad D_{K_{V_t}} = \text{diag}\{\sigma_{t,1},\ldots,\sigma_{t,p}\},$$
$$\sigma_{t,1} \ge \sigma_{t,2} \ge \ldots \ge \sigma_{t,p}, \quad t=0,\ldots,n. \tag{15}$$

Theorem 4 (below) states the following: If there exists matrix-valued parameters $(H_t, K_{V_t}) \in \mathbb{R}^{p\times p}\times\mathcal{S}_+^{p\times p}, t=0,\ldots,n$, in the set $\mathcal{Q}^{H,K_V}_{[0,n]}(D)$, such that $\widehat{X}_{t|t} = Y_t - a.s., t=0,\ldots,n$, then the reproduction distribution satisfies $\mathbf{P}_{Y_t|Y^{t-1},X_t} = \mathbf{P}_{Y_t|Y_{t-1},X_t}$, and hence the joint process (X^n, Y^n) that achieves $R^{na}_{0,n}(D)$ is Markov.

Theorem 4 ($R^{na}_{0,n}(D)$ of Gauss-Markov processes with total MSE distortion)**.** *Consider the statement of Theorem 2.(a). Then, if there exists $(H_t, K_{V_t}) \in \mathbb{R}^{p\times p}\times\mathcal{S}_+^{p\times p}, t=0,\ldots,n$, such that $\widehat{X}_{t|t} = Y_t - a.s., t=0,\ldots,n$ then the joint distribution of (X^n, Y^n) is Markov, and it is induced by the representation*

$$X_t = A_{t-1}X_{t-1} + B_{t-1}W_t,\ X_0 = x,\ t=1,\ldots,n,$$
$$Y_t = H_tX_t + \Big(I-H_t\Big)A_{t-1}Y_{t-1} + V_t,\ t=1,\ldots,n,$$
$$Y_0 = H_0X_0 + V_0.$$

and $\mathbf{P}_{Y_t|Y^{t-1},X^t} = Q^1(dy_t|x_t,y_{t-1})$.

Theorem 5 (below) establishes existence of the matrices $(H_t, K_{V_t}) \in \mathbb{R}^{p\times p}\times\mathcal{S}_+^{p\times p}, t=0,\ldots,n$, such that $\widehat{X}_{t|t} = Y_t - a.s., t=0,\ldots,n$ holds, by constructing these matrices.

Theorem 5 (Preliminary characterization of $R^{na}_{0,n}(D)$ of Gauss-Markov processes with total MSE distortion)**.** *The following hold.*
(a) Conditions (C1) and (C2) stated below are sufficient for the equalities $\widehat{X}_{t|t} = Y_t - a.s, t=0,\ldots,n$ to hold.

$$\text{cov}\Big(X_t, Y_t|Y^{t-1}\Big)\Big\{\text{cov}\Big(Y_t, Y_t|Y^{t-1}\Big)\Big\}^{-1} = I, \tag{C1}$$
$$\mathbf{E}^{Q^1}_\mu\Big\{X_t\Big|Y^{t-1}\Big\} = \mathbf{E}^{Q^1}_\mu\Big\{Y_t\Big|Y^{t-1}\Big\},\ t=0,\ldots,n. \tag{C2}$$

Moreover, the representation of the reproduction process Y^n, with $(H_t, K_{V_t}), t=0,\ldots,n$, defined below, satisfies $\widehat{X}_{t|t} = Y_t -$ a.s., $t=0,\ldots,n$.

$$Y_t = H_tX_t + (I-H_t)A_{t-1}Y_{t-1} + V_t,\ t=1,\ldots,n, \tag{16}$$
$$Y_0 = H_0X_0 + V_0, \tag{17}$$

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

$$H_t \triangleq I - \Sigma_t(\Sigma_t^-)^{-1} \in \mathbb{R}^{p\times p}, \tag{18a}$$
$$K_{V_t} \triangleq \Sigma_t H_t^T = \Sigma_t - \Sigma_t(\Sigma_t^-)^{-1}\Sigma_t \succeq 0, \tag{18b}$$

for $t=0,\ldots,n$, and where

$$\Sigma_t^- \triangleq A_{t-1}\Sigma_{t-1}A_{t-1}^T + B_{t-1}K_{W_t}B_{t-1}^T, \tag{19}$$

for $t=1,\ldots,n$ and $\Sigma_0^- = K_{X_0}$.
(c) The characterization of the nonanticipative RDF is equivalent to the following optimization problem:

$$R^{na}_{0,n}(D) = \inf_{\overset{\circ}{\mathcal{Q}}{}^1_{[0,n]}(D)} \Big\{\frac{1}{2}\log\Big\{\frac{|\Sigma_{X_0}|}{|\Sigma_0|}\Big\}^+ \tag{20}$$
$$+\frac{1}{2}\sum_{t=1}^n \log\Big\{\frac{|A_{t-1}\Sigma_{t-1}A_{t-1}^T + B_{t-1}K_{W_t}B_{t-1}^T|}{|\Sigma_t|}\Big\}^+\Big\}.$$

where the constraint set $\overset{\circ}{\mathcal{Q}}{}^1_{[0,n]}(D)$ is characterized by

$$\overset{\circ}{\mathcal{Q}}{}^1_{[0,n]}(D) \triangleq \Big\{\Sigma_t \in \mathcal{S}_+^{p\times p},\ t=0,\ldots,n:$$
$$\Sigma_t \preceq A_{t-1}\Sigma_{t-1}A_{t-1}^T + B_{t-1}K_{W_t}B_{t-1}^T,$$
$$\Sigma_0 \preceq K_{X_0},\ t=1,\ldots,n,\ \frac{1}{n+1}\sum_{t=0}^n \text{tr}(\Sigma_t) \le D\Big\}. \tag{21}$$

Corollary 1 (Equivalent characterization of $R^{na}_{0,n}(D)$ of Gauss-Markov processes with MSE distortion)**.** *Consider the Gauss-Markov processes with total MSE distortion of Theorem 5. Define $Z_t \triangleq X_t - Y_t$, $\tilde{X}_t \triangleq X_t - A_{t-1}Y_{t-1}$, $\tilde{Y}_t \triangleq Y_t - A_{t-1}Y_{t-1}$, $\tilde{X}_0 = X_0$, $\tilde{Y}_0 = Y_0$, for $t=1,\ldots,n$, where $\mathbf{E}^{Q^1}_\mu\{X_t\Big|Y^{t-1}\} = \mathbf{E}^{Q^1}_\mu\{Y_t\Big|Y^{t-1}\} = A_{t-1}Y_{t-1}$.*
(a) The representation of Theorem 5.(a) is equivalently expressed as

$$X_t = Y_t + Z_t, \quad t=1,\ldots,n, \quad X_0 = x_0, \tag{22}$$
$$Z_t = \Big(I-H_t\Big)A_{t-1}Z_{t-1} + \Big(I-H_t\Big)B_{t-1}W_t - V_t, \tag{23}$$

where $Z_0 = z_0 = x_0 - y_0$. Also,

$$Y_t = H_tA_{t-1}Z_{t-1} + A_{t-1}Y_{t-1} + H_tB_{t-1}W_t + V_t, \tag{24}$$
$$Y_0 = H_0X_0 + V_0, \tag{25}$$

where W_t is independent of Z_{t-1} for $t=0,\dots,n$ and (18a)-(18b) *hold.*
(b) Processes $(\tilde{X}^n, \tilde{Y}^n, Z^n)$ satisfy the following recursions.

$$\tilde{X}_t = A_{t-1}Z_{t-1} + B_{t-1}W_t, \ \tilde{X}_0 = \tilde{x}_0, \ t=1,\dots,n, \quad (26)$$
$$\tilde{Y}_t = H_t A_{t-1}Z_{t-1} + H_t B_{t-1}W_t + V_t, \ t=1,\dots,n, \quad (27)$$
$$= H_t\tilde{X}_t + V_t, \quad (28)$$
$$\tilde{Y}_0 = H_0\tilde{X}_0 + V_0, \quad (29)$$

where Z_t satisfies (23)*, W_t is independent of Z_{t-1}, and $(H_t, K_{V_t}), t=0,\dots,n$, are given by* (18)*. Further, the pay-off and average MSE of the characterization of nonanticipative RDF are equivalently expressed as*

$$I(X_0;Y_0) + \sum_{t=1}^{n} I(X_t;Y_t|Y_{t-1}) = I(\tilde{X}_0;\tilde{Y}_0) + \sum_{t=1}^{n} I(\tilde{X}_t;\tilde{Y}_t)$$
$$= \frac{1}{2}\log\Big\{\frac{|K_{X_0}|}{|\Sigma_0|}\Big\}^+ \quad (30)$$
$$+ \frac{1}{2}\sum_{t=1}^{n}\log\Big\{\frac{|A_{t-1}\Sigma_{t-1}A_{t-1}^{T} + B_{t-1}K_{W_t}B_{t-1}^{T}|}{|\Sigma_t|}\Big\}^+,$$

and

$$\mathbf{E}_{\mu}^{Q^1}\Big\{\sum_{t=0}^{n}||X_t - Y_t||^2\Big\} = \mathbf{E}_{\mu}^{Q^1}\Big\{\sum_{t=0}^{n}||Z_t||^2\Big\}$$
$$= \mathbf{E}_{\mu}^{Q^1}\Big\{\sum_{t=0}^{n}||\tilde{X}_t - \tilde{Y}_t||^2\Big\} = \sum_{t=0}^{n}\mathrm{tr}\Big(\Sigma_t\Big), \quad (31)$$

where $A_{t-1}\Sigma_{t-1}A_{t-1}^{T} + B_{t-1}K_{W_t}B_{t-1}^{T} \succeq \Sigma_t, t=1,\dots,n, K_{X_0} \succeq \Sigma_0$.

Remark 1 (On Corollary 1)**.** *The translated realization in* (28) *corresponds to the memoryless realization of Fig. 1 (left), while the pay-off and average distortions* (30) *and* (31)*, remain invariant with respect to the translation.*

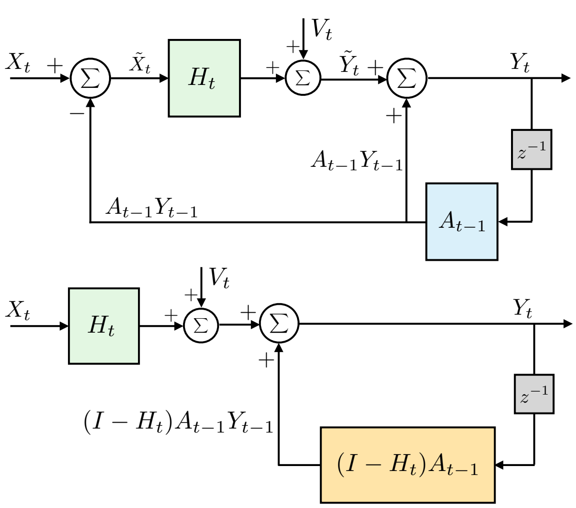

Fig. 1: Equivalent Realizations-Block diagrams of two realizations for multivariate Gaussian sources with feedback (top) and without feedback (bottom). The realization with feedback is coming from (26)-(28), while the one without feedback from (16).

The next lemma is preliminary to a subsequent main theorem, on the *structural properties* of the realization matrices $\Sigma_t^-, \Sigma_t, H_t, K_{V_t}, t=0,\dots,n$, of Theorem 5. It identifies sufficient conditions such that the lower bound of Theorem 3 holds (i.e., Hadamard's inequality holds with equality), which then implies $H_t = H_t^{T}$, hence $H_t \in \mathcal{S}_+^{p\times p}$, i.e., it is symmetric and nonnegative definite, and that $\Sigma_t^-, \Sigma_t, H_t, K_{V_t}$ have spectral decomposition w.r.t. the same unitary matrix $U_tU_t^{T} = I$, for $t=0,\dots,n$.

Lemma 1 (Preliminary structural properties of realization of Theorem 5)**.** *Consider the optimal reproduction distribution, and its realization given in Theorem 5. Suppose $\Sigma_t^- \in \mathcal{S}_+^{p\times p}$ defined by* (19) *and $\Sigma_t \in \mathcal{S}_+^{p\times p}$ commute, that is,*

$$\Sigma_t^-\Sigma_t = \Sigma_t\Sigma_t^-, \quad t=0,\dots,n, \quad (32)$$

where $\Sigma_t^- \triangleq A_{t-1}\Sigma_{t-1}A_{t-1}^{T} + B_{t-1}K_{W_t}B_{t-1}^{T}, t = 1,\dots,n, \Sigma_0^- = K_{X_0}$. Then, $H_t \triangleq I - \Sigma_t(\Sigma_t^-)^{-1} = H_t^{T} \in \mathcal{S}_+^{p\times p}$, $K_{V_t} = \Sigma_t H_t^{T} = \Sigma_t H_t = K_{V_t}^{T} \in \mathcal{S}_+^{p\times p}$, $t=0,\dots,n$, that is, H_t is also a symmetric positive semi-definite matrix, and moreover $\{\Sigma_t, \Sigma_t^-, H_t, K_{V_t}\}$ have the same spectral decomposition w.r.t. the same unitary matrix $U_tU_t^{T} = I, U_t^{T}U_t = I$, for $t=0,\dots,n$. Moreover, (32) *is a sufficient condition for* (15) *to hold, and for the inequality in* (14) *to hold with equality.*

With the aid of Lemma 1, we identify sufficient conditions for Hadamard's inequality of Theorem 3 to hold with equality, and to establish the fundamental *structural properties* of the optimal realization coefficients $\{H_t, K_{V_t}, \Sigma_t, \Sigma_t^-\}$, as stated in the next theorem. Thus, we complete characterization of the nonanticipative RDF $R_{0,n}^{na}(D)$.

Theorem 6 (Complete characterization of $R_{0,n}(D)$)**.** *Consider the Gauss-Markov process $X^n \triangleq \{X_0,\dots,X_n\}$ of Definition 1.(b), and a total square-error distortion function $d_{0,n}(x^n,y^n) \triangleq \sum_{t=0}^{n}||X_t - Y_t||^2$. Assume $R_{0,n}^{na}(D) \in [0,\infty)$ for $D \in [0, D_{\max}) \subseteq [0,\infty)$. Then, the following hold.*
(a) The RDF $R_{0,n}^{na}(D)$ is completely characterized by

$$R_{0,n}^{na}(D) = \inf_{\Delta_0,\dots,\Delta_n:\, \sum_{t=0}^{n}\mathrm{tr}\left(\Delta_t\right)\leq D} \frac{1}{2}\sum_{t=0}^{n}\log\frac{|\Lambda_t|}{|\Delta_t|}, \quad (33)$$

where

$$\Sigma_t^- = A_{t-1}\Sigma_{t-1}A_{t-1}^{T} + B_{t-1}K_{W_t}B_{t-1}^{T}$$
$$= U_t\Lambda_tU_t^{T}, \quad \Lambda_t = \mathrm{diag}\{\lambda_{t,1},\dots,\lambda_{t,p}\} \quad (34)$$
$$\Sigma_t = U_t\Delta_tU_t^{T}, \quad \Delta_t = \mathrm{diag}\{\delta_{t,1},\dots,\delta_{t,p}\}. \quad (35)$$

(b) Let $(H_t, K_{V_t}, \Delta_t^*), t=0,\dots,n$ denote the optimal values of $R_{0,n}^{na}(D)$ of part (a). Then, the realization $Y^n = Y^{*,n}$ given below achieves $R_{0,n}^{na}(D)$.

$$Y_t^* = H_t^*X_t + (I - H_t^*)A_{t-1}Y_{t-1}^* + V_t^*, \ t=1,\dots,n,$$
$$Y_0^* = H_0^*X_0 + V_0^*,$$

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

$$H_t^* \triangleq I - \Sigma_t^*(\Sigma_t^{*,-})^{-1} = H^{*,T}, \quad (36)$$
$$K_{V_t^*} \triangleq \Sigma_t^*H_t^{*,T} \succeq 0, \quad t=0,\dots,n, \quad (37)$$
$$\Sigma_t^{*,-} \triangleq A_{t-1}\Sigma_{t-1}^*A_{t-1}^{T} + B_{t-1}K_{W_t}B_{t-1}^{T}, \quad (38)$$

for $t = 1, \ldots, n$, $\Sigma_0^{*,-} = K_{X_0}$ *and where the following the property holds:* $\{\Sigma_t^*, \Sigma_t^{*,-}, H_t^*, K_{V_t^*}\}$ *have the same spectral decomposition w.r.t. the same unitary matrix* $U_t^* U_t^{*,^T} = I, U_t^{*,^T} U_t^* = I$.
(c) The optimal $Y^{*,n}$ *that achieves* $R_{0,n}(D)$ *is realized by parallel channels. For the case for which there is feedback, the realization is shown in Fig. 2.*

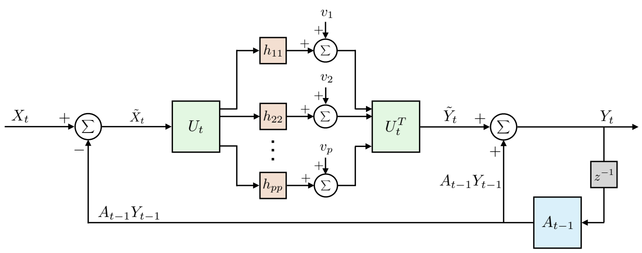

Fig. 2: Feedback realization of the optimal reproduction process by parallel channels for multivariate Gaussian sources.

B. Characterization of $R_{0,n}(D_0, \ldots, D_n)$ of Multivariate Gauss-Markov Sources with Pointwise MSE Distortion

From Theorem 6 it also follows the complete characterization of $R_{0,n}^{na}(D_0, D_1, \ldots, D_n)$ as given in the next corollary.

Corollary 2 (Complete characterization of $R_{0,n}^{na}(\{D_t\}_{t=0}^n)$)**.** *Consider the Gauss-Markov process* $X^n \triangleq \{X_0, \ldots, X_n\}$ *of Definition 1 and a pointwise square-error distortion function* $d_t(x_t, y_t) \triangleq ||X_t - Y_t||^2, t = 0, \ldots, n$. *Assume* $R_{0,n}^{na}(D_0, \ldots, D_n) \in [0, \infty)$ *for* $D_t \in [0, D_{t,\max}) \subseteq [0, \infty), t = 0, \ldots, n$. *Then, the RDF* $R_{0,n}^{na}(\{D_t\}_{t=0}^n)$ *is given by the solution of the optimization problem*

$$R_{0,n}^{na}(\{D_t\}_{t=0}^n) = \inf_{\{\Delta_t\}_{t=0}^n \colon \operatorname{tr}\left(\Delta_t\right) \le D_t, t=0,\ldots,n} \frac{1}{2} \sum_{t=0}^{n} \log \frac{|\Lambda_t|}{|\Delta_t|},$$

where (Λ_t, Δ_t) *are the diagonal elements defined by (34) and (35). The optimal realization of* $Y^{*,n}$ *of the RDF* $R_{0,n}^{na}(D_0, \ldots, D_n)$ *is (36)-(38).*

Remark 2 (Upper bound solution of $R_{0,n}^{na}(D_0, \ldots, D_n)$ of Gauss-Markov sources using water-filling)**.** *Consider the complete characterization of* $R_{0,n}^{na}(D_0, D_1, \ldots, D_n)$ *of Corollary 2, of the Gauss-Markov processes with pointwise distortion function. Then, a simple upper bound is determined as follows.*

$$R_0(D_0) = \frac{1}{2} \log \left\{ \frac{|K_{X_0}|}{|\Sigma_0^*|} \right\}^+ = \frac{1}{2} \sum_{j=1}^{p} \log \left\{ \frac{\lambda_{0,j}^*}{\theta_0} \right\}^+, \quad (39)$$

$$R_t(D_t) = \frac{1}{2} \log \left\{ \frac{|\Sigma_t^{-,*}|}{|\Sigma_t^*|} \right\}^+ = \frac{1}{2} \sum_{j=1}^{p} \log \left\{ \frac{\lambda_{t,j}^*}{\theta_t} \right\}^+, \quad (40)$$

for $t = 1, \ldots, n$, *where the constants* $\theta_t > 0, t = 0, \ldots, n$ *are uniquely determined by*

$$\sum_{j=1}^{p} \min\left\{\lambda_{0,j}^*, \theta_0\right\} = D_0, \quad (41)$$

$$\sum_{j=1}^{p} \min\left\{\lambda_{t,j}^*, \theta_t\right\} = D_t, \quad t = 1, \ldots, n. \quad (42)$$

IV. Numerical Example

We provide a numerical example in which we compute an upper bound of the nonanticipative RDF of a multivariate Gaussian process, based on sequential water-filling over the spatial dimensions of a multi-variate Gauss-Markov source, as given in Remark 2. Specifically, we choose $p = 3$, i.e., a 3-dimensional source X_t and a time-horizon $n = 3$. Matrices A_t, B_t, K_{W_t}, and K_{X_0} are randomly chosen. The solution is shown in Fig. 3.

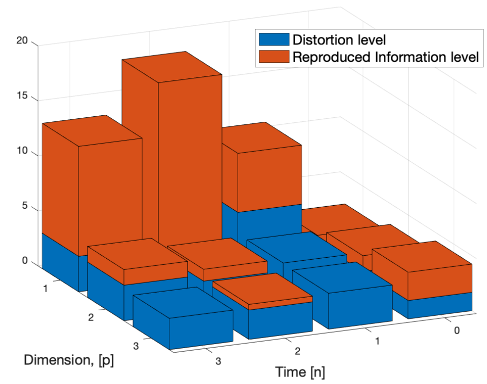

Fig. 3: Water-filling subject to pointwise distortion in time-domain for $n = 3$ time units.

In Figure 4, we plot the nonanticipative RDF of our proposed upper bound and of the solution given by [9] via Semidefinite programming (SDP) (which does not include a realization of the test channel distribution).

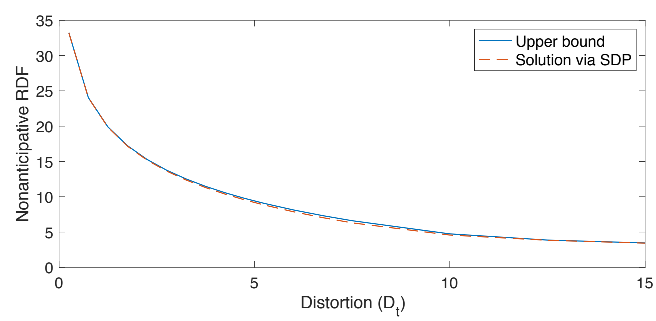

Fig. 4: Nonanticipative RDF for different values of pointwise distortion for the upper bound proposed and for the optimal solution given via SDP [9].

V. Conclusions

Structural properties of the optimal matrices that define two realizations of reproduction processes that induce the optimal test channel distribution are derived. Additionally, an upper bound on $R_{0,n}^{na}(D_0, \ldots, D_n)$ for Gauss-Markov sources using space-time water-filling is proposed. It is shown via simulations that the upper bound is tight.

Acknowledgements

This work was co-funded by the European Regional Development Fund and the Republic of Cyprus through the Research and Innovation 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] R. T. Gallager, *Information Theory and Reliable Communication*. John Wiley & Sons, Inc., New York, 1968.

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

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

[6] 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.

[7] 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.

[8] 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.

[9] T. Tanaka, K. K. K. Kim, P. A. Parrilo, and S. K. Mitter, "Semidefinite programming approach to Gaussian sequential rate-distortion tradeoffs," *IEEE Transactions on Automatic Control*, vol. 62, pp. 1896–1910, April 2017.

[10] P. A. Stavrou, T. Charalambous, and C. D. Charalambous, "Finite-time nonanticipative rate distortion function for time-varying scalar-valued Gauss-Markov sources," *IEEE Control Systems Letters*, vol. 2, pp. 175–180, Jan. 2018.

[11] 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.

[12] P. A. Stavrou, T. Charalambous, C. D. Charalambous, S. Loyka, and M. Skoglund, "Asymptotic Reverse-Waterfilling Characterization of Nonanticipative Rate Distortion Function of Vector-Valued Gauss-Markov Processes with MSE Distortion," in *2018 IEEE Conference on Decision and Control (CDC)*, pp. 14–20, Dec. 2018.

[13] P. A. Stavrou, T. Tanaka, and S. Tatikonda, "The Time-Invariant Multidimensional Gaussian Sequential Rate-Distortion Problem Revisited," *IEEE Transactions on Automatic Control*, pp. 1–1, 2019 (Early access).

[14] R. Zamir and M. Feder, "On universal quantization by randomized uniform/lattice quantizers," *IEEE Transactions on Information Theory*, vol. 38, pp. 428–436, March 1992.

[15] R. Zamir and M. Feder, "On lattice quantization noise," *IEEE Transactions on Information Theory*, vol. 42, pp. 1152–1159, July 1996.

[16] R. Zamir, *Lattice Coding for Signals and Netwroks*. Cambridge, U.K.: Cambridge Univ. Press, 2014.