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

**Complete Characterization of Gorbunov and Pinsker Nonanticipatory Epsilon Entropy of Multivariate Gaussian Sources**

*Published in:*
IEEE Transactions on Information Theory

*DOI:*
10.1109/TIT.2021.3124888

*Published: 01/03/2022*

*Document Version*
Publisher's PDF, also known as Version of record

*Published under the following license:*
CC BY

*Please cite the original version:*
Charalambous, C. D., Charalambous, T., Kourtellaris, C., & van Schuppen, J. H. (2022). Complete Characterization of Gorbunov and Pinsker Nonanticipatory Epsilon Entropy of Multivariate Gaussian Sources: Structural Properties. *IEEE Transactions on Information Theory, 68*(3), 1440-1464. https://doi.org/10.1109/TIT.2021.3124888
Complete Characterization of Gorbunov and Pinsker Nonanticipatory Epsilon Entropy of Multivariate Gaussian Sources: Structural Properties

Charalambos D. Charalambous, Themistoklis Charalambous, Senior Member, IEEE, Christos Kourtellaris, and Jan H. van Schuppen, Life Member, IEEE

Abstract—This paper derives the optimal test channel distribution and the complete characterization of the classical Gorbunov and Pinsker (1973), Gorbunov and Pinsker (1974) nonanticipatory epsilon entropy of multivariate Gaussian Markov sources with square-error fidelity, which remained an open problem since 1974. The paper also formulates a state dependent nonanticipatory epsilon entropy, in which past reproductions are available to the decoder but not to the encoder, the test channel is specified with respect to an auxiliary (state) random process, and the reproduction process is a causal function of past reproduction and the auxiliary random process. This variation is analogous to the Wyner and Ziv (1976) and Wyner (1978) rate distortion function (RDF), of memoryless sources. It is shown that the operational rate of zero-delay codes, with past reproductions available to the decoder but not to the encoder is bounded below by the state dependent nonanticipatory epsilon entropy rate. For the case of multivariate Gaussian Markov sources with square-error fidelity, the optimal test channel distribution and the complete characterization of the state dependent nonanticipatory epsilon entropy are derived, and also shown that the two nonanticipatory epsilon entropies coincide. The derivations are new; they are based on structural properties of the stochastic realizations of the reproduction process that induce the optimal test channel distributions. They are derived using, achievable lower bounds on information theoretic measures, properties of mean-square estimation theory, Hadamard’s inequality, and canonical correlation coefficients of a tuple of multivariate jointly Gaussian random processes. Applications of the nonanticipatory epsilon entropy and its state dependent variation are discussed to the areas of control of unstable Gaussian systems over limited memory channels, design of causal estimators for Gaussian Markov sources with a fidelity criterion, computation of the rate loss of causal and zero-delay codes of Gaussian Markov sources with respect to non-causal codes.

Index Terms—Nonanticipatory epsilon entropy, multivariate Gaussian processes, square-error fidelity criterion.

Gorbunov and Pinsker [2], [3] introduced the nonanticipatory epsilon entropy,1 and message generation rates of sources with respect to a fidelity criterion. These are variants of the classical rate distortion function2 (RDF) and its rate—the information theoretic definition of infimum of all achievable rates of non-causal codes subject to a fidelity [6]–[8], often called, the optimal performance theoretically attainable (OPTA) by noncausal codes. Over the last 20 years the nonanticipatory epsilon entropy of Gaussian Markov processes with symbol-by-symbol and block symbol square error distortion criteria is investigated extensively in the literature, under the names, sequential and nonanticipative RDF (see [9]–[13]), and applied in the following three main applications areas:

1) Quantification of the rate loss of the OPTA by causal codes [14] and zero-delay codes [15]–[18], with respect to noncausal codes, for Gaussian Markov sources with letter-by-letter square error fidelity [10]. For scalar sources, it is shown in [10] that the OPTA by causal codes, when feedback is used from the decoder to the encoder, exceeds Gorbunov’s and Pinsker’s message generation rates by less than $\frac{1}{2}\ln\left(\frac{2\pi e}{4}\right) \simeq 0.254$ bits/sample, while that of zero-delay codes exceeds Gorbunov’s and Pinsker’s message generation rates by less than $\frac{1}{2}\ln\left(\frac{2\pi e}{4}\right) + 1 \simeq 1.254$ bits/sample. Further, causal and zero-delay codes, based on subtractive dither with uniform scalar quantization (SDUSQ) [19], [20] and lattice codes [21] are constructed in [22] by making use of the test channel. In this application area, the main problem is to determine tight bounds on causal and zero-delay codes, and to construct

1The terminology epsilon entropy is often used in the Russian literature for Shannon’s Rate Distortion Function defined on abstract spaces, with average distortion less than or equal to $\epsilon^2$.

2An additional causality constraint is imposed on the optimal reproduction distribution of the source, known as the test channel distribution.
causal and zero-delay codes which operate with respect to such rate loss requirements.

2) Control of linear Gaussian Markov control systems over finite rate, noiseless or noisy, communications channels [9], [23], [24]. In this application area, the main problem is to determine necessary and sufficient conditions to stabilize unstable control systems and to design controllers, encoders and decoders to control the control system using a noisy or noiseless channel with finite rate constraints.

3) Synthesis of recursive, causal filters of Gaussian Markov processes subject to letter-by-letter and block letter square error fidelity criteria [11], [13], [25]. In this application, further clarified in Fig. 2, the main problem is to compute the optimal test channel distribution of Gorbunov’s and Pinsker’s nonanticipatory epsilon entropy and then realize it by designing an optimal encoder, noisy channel, and decoder, such that the Gaussian Markov process is reconstructed at the output of the decoder or estimator, with the pre-specified end-to-end average distortion. An earlier formulation of the causal estimation problem is put forward by Bucy [26], where use is made of mutual information subject to a causality constraint on the estimator.

In Bayesian estimation [27], [28] of Fig. 1, one is given a model that generates the unobserved process \( X^n = \{ X_t : t = 0, \ldots, n \} \), and a model that generates observed data \( O^n = \{ O_0, O_1, \ldots, O_n \} \), i.e., the sensor model. At each time \( t \), an estimate of the unobserved process \( X_t \), that minimizes the posterior expected value of a loss function, denoted by \( \hat{X}_t = g_t(O^{t-1}) \) for some nonanticipative measurable function \( g_t(\cdot) \), is constructed causally, from the observed data \( O^{t-1} \), for \( t = 0, \ldots, n \). Bayesian filtering theory presupposes that both models, which generate \( (X^n, O^n) \), are given \( \text{a priori} \), while at each time \( t \), the estimator is \( \hat{X}_t = g_t(O^{t-1}) \) is often computed recursively, like the Kalman filter of mean-square estimation.

In information-based estimation, one is given a model that generates the unobserved process \( X^n \) and a fidelity criterion of reconstructing \( X^n \) by \( Y^n \), and the objective is to determine the optimal nonanticipative test channel conditional distribution \( \{ P_{Y_t|Y^{t-1}, X_t} : t = 0, \ldots, n \} \) of the Gorbunov and Pinsker nonanticipatory epsilon entropy, denoted hereinafter by \( P^n_{\text{GP}}(D) \), and to realize this distribution by an \{encoder, channel, decoder\} so that the end-to-end average distortion criterion is met. Thus, in information-based estimation, the observation model is constructed by the cascade of the \{encoder, channel\}, that generates \( O^n = B^n \) according to Fig. 2 if there is feedback, and \( O^n = S^n \) according to Fig. 3 if there is no feedback and the filter is the decoder, which satisfies the end-to-end average distortion.

Despite the broad range of applications of nonanticipatory epsilon entropy, such as, 1)-3), two fundamental questions are not answered to this date.

Question 1. The first question is concerned with the mathematical problem of the complete characterization of Gorbunov’s and Pinsker’s nonanticipatory epsilon entropy for multivariate Gaussian sources with letter-by-letter and block letter distortion criteria. That is, the specification of the optimal test channel distribution, and its properties, that achieves the nonanticipatory epsilon entropy, and gives rise to a specific optimization problem over a set of parameters to be optimized, i.e., to ensure such reproduction process exist.

Application areas 1)-3) above are currently limited by the fact that, for multivariate Gaussian Markov sources with letter-by-letter or block letter square error distortion criteria, neither Gorbunov’s and Pinsker’s nonanticipatory epsilon entropy and message generation rate, nor the optimal test channel distributions that achieve these are known. For example, in [22], suboptimal rate losses are computed and suboptimal codes are constructed, and similarly in [29].

Question 2. The second question is concerned with the practical aspects of application areas of Gorbunov’s and Pinsker’s nonanticipatory epsilon entropy; more specifically, whether it is possible to achieve nonanticipatory epsilon entropy using a state dependent variation, that uses a state dependency in the mathematical problem of the complete characterization of Gorbunov’s and Pinsker’s nonanticipatory epsilon entropy for multivariate Gaussian sources with letter-by-letter and block letter distortion criteria, denoted by \( P^n_{\text{GP}}(D) \), and to realize this distribution by an \{encoder, channel, decoder\} so that the end-to-end average distortion criterion is met. Thus, in information-based estimation, the observation model is constructed by the cascade of the \{encoder, channel\}, that generates \( O^n = B^n \) according to Fig. 2 if there is feedback, and \( O^n = S^n \) according to Fig. 3 if there is no feedback and the filter is the decoder, which satisfies the end-to-end average distortion.

3This distinction is well-known for memoryless sources, due to Wyner and Ziv [4] and Wyner [5].
use of past reproduction symbols as side information. This requirement imposes a heavy burden on design specifications due to the communication of past reproductions to the encoder, compared to encoders that do not use past reproductions as side information. For example, in application area 3), the synthesis of recursive, causal filters can be implemented using:

(2.1) the feedback realization of Fig. 2 that uses past reproductions, as side information, to generate both the outputs of the sensor map and the decoder/estimator, or

(2.2) the nonfeedback realization of Fig. 3 that uses past reproductions, as side information, to generate the outputs of the decoder/estimator, and does not use past reproductions to generate the outputs of the sensor map.

The fundamental question however, is whether the nonfeedback synthesis of Fig. 3 achieves the value of the feedback synthesis of Fig. 2.

Prior literature did not address Questions 1 and 2, and this limits the full use of nonanticipatory epsilon entropy and message generation rates in the specific areas discussed. Our main emphasis in this paper is to address Questions 1 and 2.

Regarding Question 1, this paper derives, for the first time, the complete characterization of Gorbunov and Pinsker [3] nonanticipatory epsilon entropy of multivariate Gaussian Markov sources with letter-by-letter and block letter square-error fidelity criteria. The realizations of the reproduction processes that induce the optimal test channel distributions, which achieve the nonanticipatory epsilon entropy, are parametrized by a tuple of matrices that satisfy specific properties. These are feedback realizations, which are equivalent to innovations encoders with side information the past reproductions, followed by Gaussian noise channels, followed by decoders with side information the past reproductions. The feedback realizations fill the gap of the partial characterization of nonanticipatory epsilon entropy given by Gorbunov and Pinsker [3, Theorem 5]. For the nonstationary, multivariate Gaussian Markov sources, it is shown that certain matrices that define the realizations of the reproduction processes and induce the optimal test channels, admit a spectral representations with respect to the same unitary matrices. This structural property leads to characterizations of the nonanticipatory epsilon entropy by sequential multivariate optimization problems in time, with respect to (w.r.t.) the canonical correlation coefficients of a tuple of multivariate jointly Gaussian random processes. Further, for the letter-by-letter square-error criterion, an upper bound on the characterization of the nonanticipatory epsilon entropy is derived that is easily computed sequentially in time, based on a water-filling calculation in spatial dimensions, which is further shown to be analogous to the water-filling solution of the classical RDF of multivariate memoryless Gaussian sources with square-error fidelity.

To the best of the author’s knowledge, current knowledge on the complete characterization of the nonanticipative epsilon entropy, is restricted to scalar Gaussian Markov sources, given by Gorbunov and Pinsker [3, Example 2] for the letter-by-letter square-error fidelity criterion, and in [30] for the block letter square-error criterion. For multivariate Gaussian Markov sources with a letter-by-letter square-error distortion criterion, a partial characterization of the nonanticipatory epsilon entropy is given by Gorbunov and Pinsker in [3, Theorem 5]; the term “partial” refers to the fact that the matrices which define the realization of the test channel distribution remain to date unspecified. It appears from the literature that the only other research paper that attempted to characterize the nonanticipatory epsilon entropy for multivariate Gaussian Markov sources with letter-by-letter and block letter square-error fidelity criteria is [13]; see also discussion in [31]. However, the analysis in [13] produces a water-filling solution, and realizations of the test channel distributions, which are suboptimal, and hence these correspond to upper bounds on the nonanticipatory epsilon entropy. Due to the lack of an efficient characterization of the nonanticipatory epsilon entropy, for the case of multivariate Gaussian Markov sources with a letter-by-letter distortion criterion, a numerical approach is developed in [32], that makes use of semidefinite programming. However, the semidefinite programming approach does not answer the question of existence of a realization that induces the optimal test channel distribution. In particular, in the application of the algorithm, the joint distribution of the source and its reproduction, should be such that one of its marginal distributions is the source distribution, which is not easy to check.

Regarding Question 2, this paper formulates a state dependent variant of Gorbunov and Pinsker [3] nonanticipatory epsilon entropy, which is analogous to the Wyner and Ziv [4] and Wyner [5] RDF of memoryless sources. The main features of this new formulation, is the use of a test channel distribution w.r.t. an auxiliary random process (i.e., state process), subject to a causality constraint and a conditional independence condition to ensure the encoder does not use feedback from the past reproductions, unlike Gorbunov’s and Pinsek’s nonanticipatory epsilon entropy. Although, the state dependent nonanticipatory epsilon entropy is larger or equal to the Gorbunov and Pinsker nonanticipatory epsilon entropy, sufficient conditions are identified for the two to coincide. Moreover, it is shown that the operational rate of zero-delay codes is bounded below by the state dependent nonanticipatory epsilon entropy rate. It is also shown that, for multivariate Gaussian sources with square-error fidelity criterion, the two nonanticipatory epsilon entropies coincide, and moreover a nonfeedback realization is constructed of the auxiliary random process and the reproduction process that induce the optimal test channel distribution of the state dependent Gorbunov’s and Pinsek’s nonanticipatory epsilon entropy.

A. Notation

\[ \mathbb{R} \triangleq (-\infty, \infty), \mathbb{Z} \triangleq \{\ldots, -1, 0, 1, \ldots\}, \mathbb{Z}_0 \triangleq \{0, 1, 2, \ldots\}, \mathbb{N} \triangleq \{1, 2, \ldots\}, \mathbb{N}^p \triangleq \{1, \ldots, n\}, n \in \mathbb{N}. \]

For any matrix \( A \in \mathbb{R}^{p \times m} \), \( p, m \in \mathbb{N} \times \mathbb{N} \), we denote its transpose by \( A^t \), and for \( m = p \), we denote its trace by \( \text{tr}(A) \), and the matrix with diagonal entries \( A_{ii}, i = 1, \ldots, p \) and zero elsewhere by \( \text{diag}(A) \). \( S_{++}^{p \times p} \) 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
semi-definite (resp. positive definite). For \( x \in \mathbb{R} \), we define \( \{x^n\} = \max\{1, x\} \).

\( \{(X_n, B(X_n)) : n \in \mathbb{Z}\} \) denotes a measurable space, where \( X_n \) are confined to complete separable metric spaces or Polish space, and \( B(X_n) \) the Borel \( \sigma \)-algebras of subsets of \( X_n \). Points in the product space \( X_n \times \mathbb{R}^n \) are denoted by \( x_n \in (\ldots, x_{-1}, x_0, x_1, \ldots) \in \mathbb{X}^\mathbb{Z} \), and their restrictions to finite coordinates for any \( (m, n) \in \mathbb{Z} \times \mathbb{Z} \) are denoted by \( \{x_m^n \} \). Hence, \( B(X_n) = \otimes x_n \in \mathbb{Z} \) denotes the \( \sigma \)-algebra on \( \mathbb{X}^\mathbb{Z} \) generated by cylinder sets \( \{x_0, x_1, x_2, \ldots\} \). Thus, \( B(X_n) \) denote the \( \sigma \)-algebras of cylinder sets in \( X_n \), with bases over \( A_j \in B(X_n) \), \( j \in \mathbb{Z} \).

Given a probability space \( (\Omega, \mathcal{F}, \mathbb{P}) \) and a RV \( X : (\Omega, \mathcal{F}) \rightarrow (X, B(X)) \), we denote by \( \mathbb{P}\{X \in dx\} = \mathbb{P}(dx) \) the probability measure induced by \( X \) on \( (X, B(X)) \) (i.e., probability distribution if \( X = \mathbb{R}^n \)). Given another RV \( Y : (\Omega, \mathcal{F}) \rightarrow (Y, B(Y)) \), we denote by

\[
\mathbb{P}\{Y \in dy \mid X = x\} = \mathbb{P}(dy|x) \quad (I.1)
\]

the conditional probability measure of random variable (RV) \( Y \) for a fixed \( x \). The joint measure of \((X, Y)\), for a Borel set, \( G \in B(X \times Y) \), is

\[
\mathbb{P}\{(X, Y) \in G\} = \mathbb{P}(X, Y) = (\mathbb{P}(X) \otimes \mathbb{P}(Y))(G) = \mathbb{P}(X)(G) = \Delta X \mathbb{P}(X)(dx) \quad (I.2)
\]

where \( G_x \triangleq \{y \in Y \mid (x, y) \in G\} \) is the \( x \)-section of the set \( G \). That is, \( \otimes \) denotes the compounded probability operator in (I.2), and the notation, \( \mathbb{P}(X, Y)(dx, dy) = \mathbb{P}(X)(dx) \otimes \mathbb{P}(Y)(dy|x) \), is used.

Consider three random variables (RVs) \( X : \Omega \rightarrow X, Y : \Omega \rightarrow Y, \) and \( W : \Omega \rightarrow W \) defined on some probability space \((\Omega, \mathcal{F}, \mathbb{P})\). We say that RVs \((X, Y)\) are conditionally independent given RV \( W \) if \( \mathbb{P}(X, Y \mid W) = \mathbb{P}(X \mid W) \mathbb{P}(Y \mid W) \) a.s., (almost surely); the specification a.s. is often omitted. We often denote this conditional independence by the Markov Chain (MC) \( Y \leftarrow W \rightarrow X \).

Consider a tuple of jointly Gaussian finite-dimensional RVs \( X : \Omega \rightarrow \mathbb{R}^n, Y : \Omega \rightarrow \mathbb{R}^n, (n_x, n_y) \in \mathbb{N} \times \mathbb{N} \). The covariance matrix of the tuple \((X, Y)^T : \Omega \rightarrow \mathbb{R}^{n_x+n_y} \) is denoted by,

\[
K(X,Y) \triangleq \begin{pmatrix}
K_X & K_{X,Y} \\
K_{Y,X} & K_Y
\end{pmatrix} \in \mathbb{R}^{(n_x+n_y) \times (n_x+n_y)},
\]

\[
K(X,Y) \succeq 0.
\]

The reader should distinguish the covariance matrices \( K(X,Y) \) and \( K_{X,Y} \in \mathbb{R}^{n_x \times n_y} \).

**B. The Gorburnov and Pinsky Nonanticipatory Epsilon Entropy and Generalizations**

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

\[
X_t = A_{t-1}X_{t-1} + B_{t-1}W_t, \quad X_0 = x_0, \quad t = 1, \ldots, n, \quad (I.3)
\]

where

(i) \( A_t \in \mathbb{R}^{p \times p}, B_t \in \mathbb{R}^{p \times q}, t = 0, \ldots, n - 1, \) are non-random matrices;

(ii) \( \{W_t : t = 1, \ldots, n\} \) is an \( \mathbb{R}^q \)-valued independent Gaussian process, \( N(0, K_{W_t}), K_{W_t} \geq 0, \) independent of \( X_0; \)

(iii) \( X_0 \in \mathbb{R}^p \) is Gaussian \( N(0, K_{X_0}), K_{X_0} \geq 0 \).

A time-invariant Gaussian Markov process \( X_n \) corresponds to the restriction \( (A_t, B_t, K_{W_t}) = (A, B, K_W), \forall t, \) i.e., \( X_n \) is not necessarily stationary or asymptotically stationary.

We evaluate the reproduction \( Y_n \triangleq \{Y_0, Y_1, \ldots, Y_n\}, Y_i : \Omega \rightarrow \mathbb{Y}_i \subseteq \mathbb{R}^p, t = 0, 1, \ldots, n, \) of \( X^n \triangleq \{X_0, X_1, \ldots, X_n\}, \) w.r.t. the average of a block symbol square error fidelity criterion, \( d_{0,n}(\cdot, \cdot), \) defined by

\[
\frac{1}{n+1} \mathbb{E}\{d_{0,n}(X^n, Y^n)\} \leq D, \quad d_{0,n}(x^n, y^n) \triangleq \sum_{t=0}^n ||x_t - y_t||^2_{\mathbb{R}^p}, \quad D \in [0, \infty). \quad (I.4)
\]

We also consider the average of symbol-by-symbol square error fidelity criterion defined by

\[
\mathbb{E}\{d_t(X_t, Y_t)\} \leq D_t, \quad d_t(x_t, y_t) \triangleq ||x_t - y_t||^2_{\mathbb{R}^p}, \quad D_t \in [0, \infty), \quad t = 0, \ldots, n. \quad (I.5)
\]

Given the fixed source probability distribution \( \mathbb{P}(X^n), \) and a conditional probability distribution \( \mathbb{P}(Y^n \mid X^n), \) define the joint distribution of \((X^n, Y^n)\), by \( \mathbb{P}(X^n, Y^n) = \mathbb{P}(Y^n \mid X^n) \otimes \mathbb{P}(X^n), \) with marginal distributions \( \mathbb{P}(X^n), \mathbb{P}(Y^n). \) Let \( I(X^n, Y^n) \) denote the mutual information between \( X^n \) and its reproduction \( Y^n, \) defined by [7]

\[
I(X^n, Y^n) \triangleq \int \log \left( \frac{\mathbb{P}(X^n, Y^n)}{\mathbb{P}(X^n) \mathbb{P}(Y^n)} \right) \mathbb{P}(X^n, Y^n) \in [0, \infty] \quad (I.6)
\]

\[
= \int \log \left( \frac{\mathbb{P}(Y^n \mid X^n) \mathbb{P}(X^n)}{\mathbb{P}(Y^n) \mathbb{P}(X^n)} \right) \mathbb{P}(Y^n \mid X^n) \mathbb{P}(X^n) \quad (I.7)
\]

\[
= \int \log \left( \frac{\mathbb{P}(Y^n \mid X^n) \mathbb{P}(X^n)}{\mathbb{P}(Y^n)} \right) \mathbb{P}(Y^n \mid X^n) \mathbb{P}(X^n) \quad (I.8)
\]

where \( \mathbb{P}(Y^n \mid X^n) \) is the distribution defined by the product of the two distributions, and (I.8) follows because

\[
\mathbb{P}(Y^n \mid X^n) \otimes \mathbb{P}(X^n) = \mathbb{P}(Y^n \mid X^n) - \text{a.s.}
\]

(by the Radon-Nikodym derivative theorem).

**(P1) The nonanticipatory epsilon entropy w.r.t. the block symbol fidelity criterion** [2], is defined by

\[
R_n^{GP}(D) \triangleq \inf_{P_{X^n, Y^n}} \mathbb{E}\{d_{0,n}(X^n, Y^n)\} \leq D \quad (I.9)
\]

\[\text{Gorburnov and Pinsky also considered double-sided sources processes, i.e., } X^n \triangleq \{\ldots, X_{-1}, X_0, X_1, \ldots, X_n\}, n = 0, 1, \ldots.\]
(P2) The nonanticipatory epsilon entropy w.r.t. the symbol-by-symbol fidelity criterion [2], is defined by
\[
R_n^{GP}(D_0, \ldots, D_n) \triangleq \inf_{P_{X^n,Y^n}: P_{Y^n|X^n}=P_{Y^n|X_t}} \mathbb{E}\{d_t(X_t,Y_t)\} \leq D_t, t=0, \ldots, n.
\]
(I.10)

In (P1) and (P2) the infimum is taken over all joint distributions \( P_{X^n,Y^n} \) of the joint process \((X^n,Y^n)\) such that following conditions hold:

(GP-1) \( \int_{y^n} P_{X^n,Y^n}(dx^n,dy^n) = P_{X^n}(dx^n) \), i.e., the joint distribution has an \( X^n \)-marginal that corresponds to the fixed distribution of the source \( X^n \), and

(GP-2) for each \( t \in \{0, \ldots, n\} \), the RV \( Y^t \) is conditionally independent of \( X_{t+1}^n \) conditioned on \( X^t \), that is, \( P_{Y^t|X^t} = P_{Y^t|X_t} \), equivalently expressed as a Markov chain (\( \leftrightarrow \))

\[
X^n_{t+1} \leftrightarrow X^t \leftrightarrow Y^t, \quad t = 0, \ldots, n.
\]
(I.11)

Conditional independence (I.11) is known as the “causality condition” of the reproduction distribution. The underlying assumption in [3] is that the infimums in (I.9) and (I.10) exist and they are finite.

In the literature, the nonanticipatory epsilon entropy is often called sequential RDF, because the nonanticipative constraint described under (GP-2) above, implies [11], \( P_{Y_{1:t+1}|X_{1:t}} = P_{Y_{1:t+1}|X_1} \) and \( P_{X_{1:t+1}|Y_{1:t} = y} = P_{X_{1:t+1}} \), \( t = 0, \ldots, n \), and hence \( R_n^{GP}(D) \) can be expressed sequentially, as a special case of directed information, as follows:

\[
R_n^{GP}(D) = R_n^{na}(D) \triangleq \inf_{P_{X^n,Y^n}=\otimes_{t=0}^n(P_{Y_{t+1}|X_{1:t}} \otimes P_{X_{t+1}|X_t})} \sum_{t=0}^n I(X^t; Y^t|Y^{t-1}),
\]
(I.12)

The reproduction distribution is \( P_{Y_t|Y^{t-1},X_t}, t = 0, \ldots, n \). Hence, at each time \( t \) the probability distribution of \( Y_t \) depends on \( X^t \) and also the past reproductions \( Y^{t-1} \). For application area 3) this is described in Fig. 2. Next, we introduce the state dependent variant of Gorbunov and Pinsker nonanticipatory epsilon entropy, which imposes the requirement that the test channel distribution \( P_{Y_t|Y^{t-1},X_t}, t = 0, \ldots, n \) is induced by a state dependent reproduction process \( Y^n \), as shown in Fig. 3. This is motivated by the Wyner and Ziv [4] and [5] RDF of memoryless sources with side information only at the decoder.

(P3) The nonanticipatory epsilon entropy with state dependent reproduction process and block symbol fidelity criterion. Consider a sequence of states or auxiliary RVs \( S^n \triangleq \{ S_0, S_1, \ldots, S_n \}, S_t : \Omega \rightarrow S_t \triangleq \mathbb{R}^m, t = 0, 1, \ldots, n \), such that at each time \( t \), the reproduction process is state dependent, \( Y_t = f_s(S_t, Y^{t-1}) \) for some measurable functions \( f_s(\cdot), t = 0, \ldots, n \). Define the state dependent variant of the Gorbunov and Pinsker nonanticipatory epsilon entropy, by the following optimization problem\(^6\)

\[
R_n^{GPS}(D) \triangleq \inf_{M_n^{GPS}(D)} \sum_{t=0}^n I(X^t; S_t|S^{t-1}, Y^{t-1})
\]
(I.13)

where \( M_n^{GPS}(D) \) is specified by the set

\[
M_n^{GPS}(D) \triangleq \left\{ S_t : \Omega \rightarrow S_t : \right. \left. P_{S_t|X^t,Y^{t-1},S^{t-1}} = P_{S_t|X^t}, \quad \right. \left. P_{X^t|S^{t-1},Y^{t-1}} = P_{X^t|S^{t-1}}, \quad \right. \left. \exists \text{ meas. funct. } f_t : S_0 \times \ldots S_t \times \mathbb{Y}_0 \times \ldots \mathbb{Y}_{t-1} \rightarrow \mathbb{Y}_t, \quad \right. \left. Y_t = f_t(S^t, Y^{t-1}), \quad t = 0, \ldots, n, \right. \left. \frac{1}{n+1} \mathbb{E}\{d_{0,n}(X^n,Y^n)\} \leq D \right\}.
\]
(I.14)

In (P3) the optimization is over the sequence of auxiliary RVs \( S^n \), such that there exist measurable functions \( f_t(\cdot), t = 0, \ldots, n \), and the induced joint distribution \( P_{X^n,Y^n,S^n} \) of the joint process \((X^n,Y^n,S^n)\) satisfies the following conditions:

(GPS-1) \( \int_{x^n,y^n} P_{X^n,Y^n,S^n}(dx^n,dy^n,ds^n) = P_{X^n}(dx^n) \), i.e., the joint distribution has \( X^n \)-marginal the fixed distribution of the source \( X^n \), and

(GPS-2) for each \( t \in \{0,1,\ldots,n\} \), the conditional independence conditions the joint distribution is

\[
P_{X^t,S^t,Y^t} = P_{Y^t|Y^{t-1}} \otimes P_{S_t|S^{t-1}} \otimes P_{X_t|X^{t-1}} \otimes P_{X^{t-1}|S^{t-1},Y^{t-1}}, \quad Y_t = f_t(S^t, Y^{t-1}).
\]
(I.15)

The test channel is \( P_{S_t|S^{t-1}} \) of Fig. 3. The state dependent nonanticipatory epsilon entropy involves the state or auxiliary RVs; the reproduction process \( Y^n \) is analogous to the Wyner-Ziv [4], [5] RDF with side information \( Y^{t-1} \) at the decoder but not the encoder. We show in subsequent parts of the paper that in general, \( R_n^{GPS}(D) \geq R_n^{GP}(D) \), and equality holds if at each time \( t \), and for fixed \((s^{t-1}, y^{t-1})\), the map \( f_t(s_t, s^{t-1}, y^{t-1}) = y_t \) uniquely defines \( s_t \) for all \( t = 0, \ldots, n \), i.e., sequentially, \( y_0 \) uniquely defines \( s_0 \), \( y_1 \) and \( s_0, y_0 \) uniquely defines \( s_1 \), etc. The above variation is motivated by the fundamental differences of the feedback and nonfeedback realizations, Fig. 2 and Fig. 3, described under applications area 3).

C. Main Contributions

We analyze the Gorbunov and Pinsker nonanticipative epsilon entropy, and its state dependent variant (I.13), and we introduce new methods to completely characterize these RDFs, for sources of Definition 1 and fidelity criteria defined by (I.4), (I.5). The highlights of the paper are the following:

1) Realizations of \( Y^n \) that induce the optimal reproduction distributions \( P_{Y^n|X^n,S^n}, \ n = 0, 1, \ldots, \ R_n^{GP}(D) \) and \( R_n^{GPS}(D) \). We make use of properties of mean-square estimation theory, achievable lower bounds on mutual information, and the weak stochastic realization theory of Gaussian processes, to derive structural properties

---

\(^6\)Our formulation of the infimum follows closely Wyner’s [5] analog of memoryless sources; some authors formulate Wyner’s problem by optimizing over the test channel and the function \( f_s(\cdot) \).
of the realizations of $Y^n$ that induce the optimal reproduction distribution (or test channel). The specification of the parameters of the realization completes and simplifies the prior characterizations given in [3].

2) Spectral Representations of the matrices of the realizations of optimal reproduction distributions of $R_n^{GP} (D)$ and $R_n^{GPS} (D_0, \ldots, D_n)$. We show a structural property, which states that certain matrices that parametrize the realizations of the reproduction process and produce the optimal test channel distribution, admit spectral representations w.r.t. unitary matrices. This structural property is shown by making use of Hadamard’s inequality to derive achievable lower bounds on the mutual information $I(X^n; Y^n)$.

3) Complete Sequential characterizations of $R_n^{GP} (D)$ and $R_n^{GPS} (D_0, \ldots, D_n)$ in time of Gaussian Markov processes via sequential water-filling. Characterization of nonanticipatory epsilon entropy by a sequential multivariate optimization problem, which is expressed in terms of the canonical correlation coefficients of a tuple of multivariate jointly Gaussian random processes, which is equivalent to optimization problems that lead to water-filling. Further, an upper bound is proposed on $R_n^{GP} (D_0, \ldots, D_n)$ which is easy to compute using pre-processing and post-processing of the source $X^n$ and its reproduction $Y^n$ by a unitary transformation, and sequential in time, water-filling in spatial dimensions, based on parallel additive Gaussian noise channels. Simulations show the upper bound is tight.

4) Equality of $R_n^{GP} (D) = R_n^{GPS} (D)$. We show that in general, $R_n^{GPS} (D) \geq R_n^{GP} (D)$, while equality holds if certain conditions are satisfied. For the example of Gaussian Markov source with square-error fidelity, we show equality $R_n^{GP} (D) = R_n^{GPS} (D)$ by constructing a realization of the RVs $(S^n, Y^n)$, in the set $M_n^{GPS} (D)$ defined by (I.14). Due to space limitations, the consequences of the state dependent Gorbunov and Pinsker nonanticipatory epsilon entropy $R_n^{GPS} (D)$ are discussed with respect to application area 3) and 1) but not 2).

D. Organization

The paper is organized as follows. In Section II, we review an alternative equivalent definition of the Gorbunov and Pinsker nonanticipatory epsilon entropy, often called nonanticipative RDF, and use this to prove the main theorem on the state dependent nonanticipatory epsilon entropy. In Section III, we derive the main results for the complete characterization of Gorbunov and Pinsker nonanticipatory epsilon entropy of multivariate Gaussian Markov sources with letter-by-letter and block letter square-error fidelity criteria. The realizations of the reproduction processes that induce the optimal test channel distributions, which achieve the nonanticipatory epsilon entropy, are parametrized by a tuple of matrices that satisfy specific properties. In Section IV, we derive the corresponding results for the state dependent Gorbunov and Pinsker nonanticipatory epsilon entropy. In Section V, we briefly discuss prior applications of Gorbunov and Pinsker nonanticipatory epsilon entropy and message generation rates, to the rate loss of zero-delay codes. In this section we derive a new lower bound on the operation rate of zero-delay codes using the state dependent nonanticipatory epsilon entropy and message generation rates, and illustrate fundamental differences between encoders that require knowledge of past reproductions and those that do not. It is then apparent that the state dependent Gorbunov and Pinsker nonanticipatory epsilon entropy is more suitable for such applications, since no feedback is required to the encoder. A numerical example is given in Section VI. Finally, in Section VII, we draw conclusions and discuss open problems on this topic.

II. NONANTICIPATIVE RDF, NONANTICIPATORY EPSILON ENTROPY, AND STATE DEPENDENT NONANTICIPATORY EPSILON ENTROPY

In this section, we establish the connections between nonanticipatory epsilon entropy, nonanticipative RDF, and the state dependent nonanticipatory epsilon entropy. The main result of this section is Theorem 1, which establishes $R_n^{GPS} (D) \geq R_n^{GP} (D)$ and identifies sufficient conditions such that $R_n^{GPS} (D) = R_n^{GP} (D)$.

A. Equivalence of Nonanticipatory Epsilon Entropy and Nonanticipative RDF

Definition 2 (Source and Reproduction Distributions): Let $x^n \triangleq \{x_0, x_1, \ldots, x_n\} \in X^n \triangleq \times_{i=0}^n X_i$ denote a sequence generated by the source and let $y^n \triangleq \{y_0, y_1, \ldots, y_n\} \in Y^n \triangleq \times_{i=0}^n Y_i$ denote its reproduction sequence.

(a) The source generates sequences from the set of distributions that satisfy conditional independence, i.e.,

$$M_{[0,n]} \triangleq \{ \mathcal{M}_{X_i | X_{i-1}, t = 0, \ldots, n} : \mathcal{P}_{X_i | Y_{t-1}, x_i, t = 0, \ldots, n} = \mathcal{M}_i (dx_i | x_{i-1}) = a.a. \}

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

$$Q_{[0,n]} \triangleq \{ \mathcal{P}_{Y_i | Y_{t-1}, x_i} = \mathbf{Q}_{Y_i | Y_{t-1}, x_i} = \mathcal{Q}_i (dy_i | y_{t-1}, x_t), y_{t-1}, x_t \in Y_{t-1} \times X_t, t = 0, \ldots, n \},$$

where $\mathcal{Q}_{Y_i | Y_{t-1}, x_i} = \mathcal{M}_{X_0}$ is the initial distribution of $X_0$, and a.a. means almost all.

The definitions of $M_{[0,n]}$ and $Q_{[0,n]}$, presuppose that the initial states $\{X^{-1}, Y^{-1}\}$ generate the trivial information, that is, if $\sigma \{X^{-1}, Y^{-1}\}$ is the $\sigma$-algebra generated by $\{X^{-1}, Y^{-1}\}$, then $\sigma \{X^{-1}, Y^{-1}\} = \{\emptyset, \Omega\}$.

For each $t = 0, 1, \ldots$, we introduce the space $\mathcal{G}^t$ of admissible source and reproduction histories up to time $t$, as follows. Define

$$\mathcal{G}^t \triangleq X_0 \times Y_0 \times \ldots \times X_{t-1} \times Y_{t-1} \times X_t \times Y_t, t = 0, 1, \ldots.$$ (II.2)

A typical element of $\mathcal{G}^t$ is $(x_t, y_t) \triangleq (x_0, y_0, \ldots, x_t, y_t)$. Let $(\Omega, \mathcal{F})$ be a measurable space consisting of the
canonical sample space $\Omega \triangleq \mathbb{G}^\infty$, and $\mathcal{F}$ its corresponding $\sigma$-algebra $\mathcal{F} \triangleq \mathcal{B}(\mathbb{G}^\infty)$. Given the elements of $\mathcal{M}_{[0,n]}$, $\mathcal{Q}_{[0,n]}$, i.e., $M_{X|X^{-t}}$, $Q_{Y|Y^{-t},X_{t}} : t = 0, \ldots$, then by Ionescu-Tulcea theorem, there exists a unique probability measure $\mathbb{P}^Q$ on $(\mathbb{G}^\infty, \mathcal{B}(\mathbb{G}^\infty))$, which is supported on $\mathbb{G}^\infty$, that is, $\mathbb{P}^Q(\mathbb{G}^\infty) = 1$, carrying the sequence of RVs $\{X_t, Y_t : t = 0, 1, \ldots\}$, and defined by

$$\mathbb{P}^Q(dx_0, dy_0, dx_1, dy_1, \ldots, dx_{n-1}, dy_{n-1}, dx_n, dy_n) = M_{X_0} \otimes Q_{Y|X_0} \otimes M_{X_1|X_0} \otimes \cdots \otimes Q_{Y_{n-1}|Y_{n-2}, X_{n-1}} \otimes M_{X_n|X_{n-1}} \otimes Q_{Y_n|Y_{n-1}, X_n}, n = 0, 1, \ldots$$  \hspace{1cm} (II.3)

The conditional distribution of $Y_t$ given $Y^{t-1}$ is then defined by

$$\mathbb{P}\left\{Y_t \in dy_t | Y^{t-1} = y^{t-1}\right\} = \mathbb{P}^Q_{Y_t|Y^{t-1}} = \mathbb{P}^Q Y_t \triangleq \int_{X_t} Q_{Y_t|Y^{t-1}, x^{t-1}} \otimes M_{X_t|x^{t-1}} \otimes Q_{Y_{t+1}|Y^t, x^t}, t = 1, \ldots, n$$  \hspace{1cm} (II.4)

Next, we define the nonanticipative RDF. Then, we invoke a lemma to deduce its equivalence to the Gorbunov and Pinsker nonanticipatory epsilon entropy, which is defined in the Introduction.

**Definition 3 (Nonanticipative RDF):** Consider the source and reproduction distributions of Definition 2. The information measure is

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

$$= \mathbb{E}^Q \left\{ \sum_{t=1}^{n} \log \left( \frac{Q_{Y_t|Y^{t-1}, X^t}}{\mathbb{P}^Q_{Y_t|Y^{t-1}}} \right) \right\} \in [0, \infty]$$  \hspace{1cm} (II.5)

$$= \mathbb{E}^Q \left\{ \log \left( \frac{Q_{Y_0|X_0}}{\mathbb{P}^Q_{Y_0}} \right) + \sum_{t=1}^{n} \log \left( \frac{Q_{Y_t|Y^{t-1}, X^t}}{\mathbb{P}^Q_{Y_t|Y^{t-1}}} \right) \right\},$$

where $I(X_t; Y_t|Y^{t-1})$ denotes the conditional mutual information between $X^t$ and $Y_t$ conditioned on $Y^{t-1}$, and the value $+\infty$ is allowed (i.e., for RVs which are continuous-valued).

(a) The nonanticipative RDF of the source subject to a block letter distortion is defined by

$$R^a_n(D) \triangleq \inf_{Q_{[0,n]}(D)} \mathbb{E}^Q \left\{ \sum_{t=0}^{n} \log \left( \frac{Q_{Y_t|Y^{t-1}, X^t}}{\mathbb{P}^Q_{Y_t|Y^{t-1}}} \right) \right\} \in [0, \infty]$$  \hspace{1cm} (II.6)

where

$$Q_{[0,n]}(D) \triangleq \left\{ Q_{Y_t|Y^{t-1}, X^t}, t = 0, \ldots, n : \frac{1}{n+1} \mathbb{E}^Q \left\{ d_{0,n}(X^n, Y^n) \right\} \leq D \right\}$$  \hspace{1cm} (II.7)

and $d_{0,n} : \mathbb{G}^n \times \mathbb{G}^n \to [0, \infty)$ is a measurable function. If the infimum does not exists then $R^a_n(D) = +\infty$.

(b) The nonanticipative RDF of the source subject to a letter-by-letter distortion is defined by

$$R_n^a(D_0, \ldots, D_n) \triangleq \inf_{Q_{[0,n]}(D_0, \ldots, D_n)} \mathbb{E}^Q \left\{ \sum_{t=0}^{n} \log \left( \frac{Q_{Y_t|Y^{t-1}, X^t}}{\mathbb{P}^Q_{Y_t|Y^{t-1}}} \right) \right\} \in [0, \infty]$$  \hspace{1cm} (II.8)

where

$$Q_{[0,n]}(D_0, \ldots, D_n) \triangleq \left\{ Q_{Y_t|Y^{t-1}, X^t}, t = 0, \ldots, n : \mathbb{E}^Q \left\{ d_t(X_t, Y_t) \right\} \leq D_t, t = 0, \ldots, n \right\} \in [0, \infty]$$

If the infimum does not exist, then $R_n^a(D_0, \ldots, D_n) = +\infty$. If two-sided processes are considered, then $(x^n, y^n)$ are replaced by $(x^{-\infty}_n, y^{n\infty}_n)$, and information measure $I(X_0; Y_0) + \sum_{t=1}^{n} I(X^t; Y^{t-1})$ is replaced by $I(X^{-\infty}; Y^0|Y^{-\infty}) + \sum_{t=1}^{n} I(X^t; Y^{t-1})$, where the distribution of the RVs $(X^{-\infty}, Y^{-\infty})$ is $P_{X^{-\infty}, X^{-\infty}} \otimes P_{Y^\infty}$. It should be mentioned that existence of the infimum is shown in [33], under appropriate conditions of weak compactness of probability measures.

**Remark 1:** Gorbunov and Pinsker [2] also analyzed a variation of nonanticipatory epsilon entropy, when the conditional independence (I.11) is replaced by

$$X_{t+1}^n \leftrightarrow X_{t-\infty} \leftrightarrow Y_{t-\infty}, t = 0, \ldots, n.$$  \hspace{1cm} (II.10)

Thus, if the $\sigma$-algebra generated by $(X^{-1}, Y^{-1})$ is $\sigma\{X^{-1}, Y^{-1}\} = \{\Omega, \varnothing\}$, i.e., the trivial $\sigma$-algebra, then (II.10) reduces to (I.11).

From the next lemma follows that $R_n^{GP}(D) = R_n^a(D)$ and $R_n^{GP}(D_0, \ldots, D_n) = R_n^a(D_0, \ldots, D_n)$.

**Lemma 1 (Conditional Independence Conditions):** Consider the distributions of Definition 2. Then, for any $n = 0, 1, \ldots$, the following statements are equivalent.

**MC1:** $P_{Y^n|X^n} = \bigotimes_{t=0}^{n} P_{Y_t|Y^{t-1}, X^t}$

**MC2:** $Y_t \leftrightarrow (X^t, Y^{t-1}) \leftrightarrow (X_{t+1}, X_{t+2}, \ldots, X_n)$ forms a MC, for each $t = 0, 1, \ldots, n$.

**MC3:** $Y^t \leftrightarrow X^t \leftrightarrow X_{t+1}$ forms a MC, for each $t = 0, 1, \ldots, n$.

**MC4:** $X_{t+1}^n \leftrightarrow X^t \leftrightarrow Y^t$ forms a MC, for each $t = 0, 1, \ldots, n$.

**Proof:** The equivalence of MC1-MC3 is shown in [11]. The equivalence of MC1-MC3 to MC4 is done similarly to [11].

**B. State Dependent Nonanticipatory Epsilon Entropy**

Now, we turn our attention to the state dependent nonanticipatory epsilon entropy, with state or auxiliary random process, $S^n$, i.e., $R_n^{GPS}(D)$, defined by (I.13), to prove the upper bound $R_n^{GPS}(D) \geq R_n^a(D) \equiv R_n^{GP}(D)$ and to identify a fundamental property for the inequality to hold with equality.

**Theorem 1 (Properties of State Dependent Nonanticipatory Epsilon Entropy):** Consider $R_n^{GPS}(D)$ defined by (I.13). Then, the following hold.
(a) The following identity holds.
\[ R_{n}^{GPS}(D) \triangleq \inf_{M_{[0,n]}^{GPS}(D)} \sum_{t=0}^{n} I(X_t;S_t|S^{t-1},Y^{t-1}) \] (II.11)

(b) For each element of the set \( M_{[0,n]}^{GPS}(D) \) the following inequality holds,
\[ \sum_{t=0}^{n} I(X_t;S_t|S^{t-1},Y^{t-1}) \geq \sum_{t=0}^{n} I(X_t;Y_t|S^{t-1},Y^{t-1}). \] (II.12)

Moreover, if for each \( t \), the function \( f_t(\cdot) \), is such that for fixed \((s^{t-1},y^{t-1})\) the map \( f_t(\cdot,s^{t-1},y^{t-1}) : S_t \to Y_t \), \( f_t(s_t,s^{t-1},y^{t-1}) = y_t \) is invertible and measurable (i.e., for fixed \((s^{t-1},y^{t-1})\), knowledge of \( y_t \) uniquely defines \( s_t \)) for all \( t = 0,\ldots,n \), then the equality holds,
\[ \sum_{t=0}^{n} I(X_t;S_t|S^{t-1},Y^{t-1}) = \sum_{t=0}^{n} I(X_t;Y_t|S^{t-1},Y^{t-1}) = \sum_{t=0}^{n} I(X_t;Y_t|Y^{t-1}). \] (II.13)

(c) The inequality holds,
\[ R_{n}^{GPS}(D) \geq R_{n}^{na}(D) \equiv R_{n}^{GP}(D). \] (II.14)

Moreover, inequality in (II.14) holds with equality, if the map \( f_t(\cdot) : S_t \to Y_t \), \( f_t(s_t,s^{t-1},y^{t-1}) = y_t \) is invertible and measurable for all \( t = 0,\ldots,n \).

Proof: See Appendix A.

Remark 2: For the Gaussian Markov process \( X^n \) of Definition 1, and a square-error distortion criterion, in subsequent sections, we give a realization of \((S^n,Y^n)\) such that the equality holds, \( R_{n}^{GPS}(D) = R_{n}^{na}(D) \).

III. NONANTICIPATIVE RDF OF MULTIVARIATE GAUSSIAN MARKOV SOURCES WITH SQUARE ERROR FIDELITY

In this section, we derive the characterizations of nonanticipative RDF for the time-varying multivariate Gaussian Markov source of Definition 1. The main results of this section are:

1) Theorem 4, which applies Proposition 1 and Theorem 2, and a data processing inequality to identify sufficient conditions, in terms of mean-square estimation theory, to show that a Markov Gaussian joint distribution \( P_{X^n,Y^n}(dx^n,dy^n) \) achieves a lower bound on the nonanticipative RDF \( R_{n}^{na}(D) \).

2) Theorem 5, which identifies the Markov parametric realization of the joint process \((X^n,Y^n)\), which achieves the lower bound of the nonanticipative RDF derived in Theorem 4, Theorem 5 and Corollary 1, complete Gorbunov’s and Pinsker’s partial characterization in [3, Theorem 5].

3) Theorem 6, which applies Hadamard’s inequality to identify certain structural properties of the parametric realization of the joint process \((X^n,Y^n)\) that achieves a lower bound on the nonanticipative RDF, and relates this achievable lower bound to the canonical correlation coefficients of a tuple of multivariable jointly Gaussian random processes.

4) Theorem 7 which characterizes \( R_{n}^{na}(D) \) as an optimization problem that involves the canonical correlation coefficients.

5) Theorem 8, which gives an upper bound on the complete characterization of \( R_{n}^{na}(D) \), and Proposition 2, which gives an upper bound solution for \( R_{n}^{na}(D_0,\ldots,D_n) \), based on sequential water-filling across the dimensions of the source \( X^n \).

In subsequent sections, we make use of the following definitions of conditional means and error covariances, from mean-square estimation theory.

\[ \hat{X}_t|_t \triangleq E\{X_t|Y^t\}, t = 0,\ldots,n, \]
\[ \hat{X}_{t-1} \triangleq E\{X_{t-1}|Y^{t-1}\}, t = 0,\ldots,n, \]
\[ \hat{X}_{0|0} \triangleq E\{X_0\} = 0, \]
\[ \Sigma_t \triangleq E\{(X_t - \hat{X}_t|_t)(X_t - \hat{X}_t|_t)^\top\}, t = 0,\ldots,n, \]
\[ \Sigma_t \triangleq E\{(X_t - \hat{X}_{t-1})(X_t - \hat{X}_{t-1})^\top\}, t = 1,\ldots,n, \]
\[ \Sigma_0 \triangleq E\{(X_0 - \hat{X}_{0|0})(X_0 - \hat{X}_{0|0})^\top\} = K_{X_0}. \]

The derivations and structural properties of the realization of the optimal reproduction distribution are new and never appeared in the literature. These are also applicable to other RDF problems of Gaussian multivariate sources with square-error fidelity.

A. Preliminary Results on Mean-Square Estimation

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

Proposition 1: Consider a pair of RVs \( X = (X_1,\ldots,X_k)^\top : \Omega \to \mathbb{R}^k \) and \( Y = (Y_1,\ldots,Y_l)^\top : \Omega \to \mathbb{R}^l \), defined on some probability space \((\Omega,\mathcal{F},\mathbb{P})\). Let \( G \subseteq \mathcal{F} \) be a sub-\( \sigma \)-algebra.\(^8\) Assume the conditional distribution of \((X,Y)\) conditioned on \( G \), i.e., \( P(dx,dy|G) \in \mathbb{P} - a.s. \) (almost surely) Gaussian, with conditional means
\[ \mu_{X|G} \triangleq E\{X|G\}, \quad \mu_{Y|G} \triangleq E\{Y|G\} \] (III.1)
and conditional covariances
\[ Q_{X|G} \triangleq \text{cov}\{X,X|G\}, \]
\[ Q_{X,Y|G} \triangleq \text{cov}\{X,Y|G\}, \]
\[ Q_{Y|G} \triangleq \text{cov}\{Y,Y|G\} \]
\[ = E\{X - E\{X|G\}\}(Y - E\{Y|G\})^\top|G \geq 0. \]

Then, the vectors of conditional expectations \( \mu_{X,Y|G} \triangleq E\{X,Y|G\} \) and matrices of conditional covariances
\(^8\)In our application of the proposition, we often we replace \( G \) by a RV, i.e., the \( \sigma \)-algebra generated by the RV.
where notation \((\cdot)\dagger\) denotes pseudoinverse of the matrix \((\cdot)\).

If \(\mathcal{G}\) is the trivial information, i.e., \(\mathcal{G} = \{\Omega, \emptyset\}\), then \(\mathcal{G}\) is removed from the above expressions, and (III.2), (III.3) degenerate to the well-known conditional mean of mean-square estimation theory.

From Proposition 1 then follows Theorem 2 (below).

**Theorem 2 (Equivalent Statements):** Consider the statement of Proposition 1 and assume the inverse of \(Q_{Y|\mathcal{G}}\) exists. Any two of the following three conditions imply the third:

1. Condition 1: \(Q_{X,Y|\mathcal{G}}Q_{Y|\mathcal{G}}^{-1} = I\).
2. Condition 2: \(\mu_{X|\mathcal{G}} = \mu_{Y|\mathcal{G}}\).
3. Condition 3: \(\mu_{X|Y,\mathcal{G}} = Y - a.s.\).

**Proof:** The statements are easily verified by substituting any two conditions in (III.2) and (III.3) in order to retrieve the third condition.

---

**B. Characterization of \(R_n^{an}(D)\) of Multivariate Gaussian Markov Sources With Block Letter Square Error Distortion**

Now, we prepare to derive the complete characterization of the nonanticipative epsilon entropy, i.e., of \(R_n^{\text{GP}}(D) = R_n^{an}(D)\).

Suppose \((X^n, Y^n)\) is jointly Gaussian. By an application of (III.2), with \(X = X_t, Y = Y_t, \mathcal{G} = Y^{t-1}\), then (III.7), shown at the bottom of the page, is obtained.

By an application of (III.3), with \(X = X_t, Y = Y_t, \mathcal{G} = Y^{t-1}\), then (III.8) and (III.9), shown at the bottom of the page, are obtained.

Throughout the paper, we apply Proposition 1 and (III.7), (III.8), to identify the optimal realizations matrices of \(Y^n\) and their structural properties characterizing the nonanticipative RDF.

We make use of the next Theorem which follows from [13].

**Theorem 3 (\(R_n^{an}(D)\) of Gaussian Markov Processes With Block Letter Square Error Distortion):** Consider the Gaussian Markov process \(X^n \triangleq \{X_0, \ldots, X_n\}\) of Definition 1, and a

\[^9\text{If the inverse of } Q_{Y|\mathcal{G}} \text{ exists then } (Q_{Y|\mathcal{G}})^\dagger = (Q_{Y|\mathcal{G}})^{-1}.\]

block letter square error distortion criterion \(d_0, n(x^n, y^n) \triangleq \sum_{t=0}^n ||x_t - y_t||^2\).

(a) The minimum of the nonanticipative RDF \(R_n^{an}(D)\) is achieved by a Gaussian joint distribution defined by

\[
\P^Q(dx^n, dy^n) = \M_0(dx_0) \otimes Q_0(dy_0|x_0) \otimes \bigotimes_{t=1}^n \left( Q_t(dy_t|y^{t-1}, x_t) \otimes \M_t(dx_t|x_{t-1}) \right),
\]

and this joint distribution is induced by the process \(X^n\) of Definition 1 and the realization of the reproduction process, parametrized by matrices \((H_t, K_{V_t})\) as follows.

\[
Y_t = \left\{ \begin{array}{ll}
H_t X_t + (I - H_t) A_{t-1} \hat{X}_{t-1|t-1} + V_t, & t = 1, \ldots, n, \\
H_0 X_0 + V_0, & t = 0,
\end{array} \right.
\]

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

\[
\cov\left( X_t, X_t \big| Y^t \right) = \Sigma_t
\]

\[
= \E \left\{ \left( X_t - \E\{X_t|Y^t\} \right) \left( X_t - \E\{X_t|Y^t\} \right)^T \right\},
\]

\[
\cov\left( X_t, X_t \big| Y^{t-1} \right) = \Sigma_t^{-1}
\]

\[
= \E \left\{ \left( X_t - \E\{X_t|Y^{t-1}\} \right) \left( X_t - \E\{X_t|Y^{t-1}\} \right)^T \right\}
\]

given by

\[
\Sigma_t = \Sigma_t^{-1} - \Sigma_t^{-1} H_t^T \left[ H_t \Sigma_t^{-1} H_t^T + K_{V_t} \right]^{-1} \Sigma_t^{-1} H_t^T, t = 0, \ldots, n,
\]

\[
\Sigma_t^{-1} = A_{t-1} \Sigma_{t-1} A_{t-1}^T + B_{t-1} K_{W_t} B_{t-1}^T, t = 1, \ldots, n,
\]

with \(\Sigma_0 = K_{X_0}\) and where the average distortion constraint is satisfied.

Moreover, \(\hat{X}_{t|t}\) satisfies the Kalman-filter recursion

\[
\hat{X}_{t|t} = A_{t-1} \hat{X}_{t-1|t-1} + \Sigma_t H_t^T \left( H_t \Sigma_t^{-1} H_t^T + K_{V_t} \right)^{-1} (Y_t - A_{t-1} \hat{X}_{t-1|t-1}),
\]

\[
= : e_2(\hat{X}_{t-1|t-1}, Y_t), t = 1, \ldots, n, \hat{X}_{0|0} \text{ given. (III.15)}
\]

(b) The nonanticipative RDF \(R_n^{an}(D)\) is parametrized by \((H_t, K_{V_t})\) of the realization of part (a), and given by (III.16)–(III.18), shown at the bottom of the next page.

\[
\hat{X}_{t|t} \triangleq \E^Q \left\{ X_t \big| Y_t, Y^{t-1} \right\}
\]

\[
= \E^Q \left\{ X_t \big| Y^{t-1} \right\} + \cov\left( X_t, Y_t \big| Y^{t-1} \right) \left\{ \cov\left( Y_t, Y_t \big| Y^{t-1} \right) \right\}^{-1} \left\{ Y_t - \E^Q \left\{ Y_t \big| Y^{t-1} \right\} \right\}.
\]

\[
\cov\left( X_t, X_t \big| Y^t \right) = \cov\left( X_t, X_t \big| Y^{t-1} \right) - \cov\left( X_t, Y_t \big| Y^{t-1} \right) \left\{ \cov\left( Y_t, Y_t \big| Y^{t-1} \right) \right\}^{-1} \cov\left( X_t, Y_t \big| Y^{t-1} \right)^T.
\]

\[
= \Sigma_t \quad \text{if all conditional covariances in (III.8) are nonrandom.}
\]
Alternatively, $R_n^{na}(D)$ is equivalently given by (III.19) and (III.20), shown at the bottom of the page, where $Q^{H,K_{V_{t}}}_{[0,n]}(D)$ is defined by $Q^{H,K_{V_{t}}}_{[0,n]}(D)$ with $\Sigma_t^{-} \succeq \Sigma$, replaced by $H_t \Sigma_t^{-} H_t^\top \succeq 0$.

Proof: Structures (III.10), (III.11) are obtained in [13]. (III.12)-(III.15) follow from (III.7), (III.8) and the calculation of $\text{cov} \left(X_t, X_t^\top \right)$ using the parametric realization. Part (b) follows from part (a) by simple calculations. ■

Before we proceed, we digress to recall certain properties of mutual information of a tuple of scalar-valued jointly Gaussian RVs, which are further elaborated in the multivariate case by Gelfand and Yaglom [34] (see eqn(2.8)’ and Chapter II), citing Kolmogorov for their origin.

Remark 3 (On the Mutual Information of a Tuple of Jointly Gaussian RVs): Consider a tuple of scalar-valued jointly Gaussian RVs $(X,Y)$, with zero mean $E(X) = E(Y) = 0$, and correlation coefficient, $E(XY) = \rho \in [-1,1]$, and covariance matrix,

$$K_{(X,Y)} = \begin{bmatrix}
\sigma_X^2 & \rho \sigma_X \sigma_Y \\
\rho \sigma_X \sigma_Y & \sigma_Y^2
\end{bmatrix}$$

(III.21)

Then, by Gelfand and Yaglom [34], mutual information, for any $\rho \in [-1,1]$, is given by

$$I(X;Y) = \frac{-1}{2} \ln \left(1 - \rho^2\right)$$

$$= \begin{cases} 
+\infty, & \text{if } \rho \in (-1,1), \\
-\frac{1}{2} \ln \left(1 - \rho^2\right), & \text{if } \rho \in (-1,1), \rho \neq 0, \\
0, & \text{if } \rho = 0.
\end{cases}$$

(III.22)

In particular, (III.22) shows that for Gaussian RVs $(X,Y)$, then $I(X;Y) = +\infty$ if and only if the correlation coefficient takes the values $\rho = +1, -1$. In the case, $X = Y$ with probability one, then $I(X;Y) = +\infty$. For a tuple of multivariate jointly RVs, the analog of (III.22) is discussed in Gelfand and Yaglom [34] (see eqn(2.8)’ and Chapter II). The full specification is also found in [35], [36].

Remark 4: In view of Remark 3, without loss of generality we assume $\Sigma_t^{-} \succeq 0, t = 0, \ldots, n$. If it is not the case, then at least one of the eigenvalues of $\Sigma_t^{-}$ is zero which then implies it needs to be removed, because it will give a value of $\frac{1}{2} \sum_{t=0}^{n} \log \left\{\frac{\Sigma_t^{-}}{|\Sigma_t^{-}|}\right\} = +\infty$. Also, necessarily $\Sigma_t^{-} \succeq 0, t = 0, \ldots, n$, for $R_n^{na}(D) \in [0, \infty)$, i.e., to be finite. If $K_{X_0} = \Sigma_0^{-} > 0$ and $B_{t-1}K_{V_{t}}B_{t-1}^\top > 0, t = 1, \ldots, n$ then $\Sigma_t^{-} \succeq 0, t = 0, \ldots, n$. Indeed, any Gaussian RV $X : \Omega \to \mathbb{R}^n, n_x \in \mathbb{N}$, with covariance $K_X \succeq 0$, which does not satisfy $K_{X_0} > 0$, can be can transform using a linear basis transformation into the form,

$$\begin{bmatrix} K_{11} & 0 \\ 0 & 0 \end{bmatrix} \in \mathbb{R}^{n_x \times n_x},$$

$$p_{11} = \text{rank}(K_X) = \text{rank}(K_{11}), \ K_{11} \in \mathbb{R}^{p_{11} \times p_{11}}.$$
is Markov. We show existence of such parameters, by construction in Theorem 5, which completely characterizes the optimization problem of $R_{\alpha}^n(D)$ in much simpler form, and completes Gorbunov’s and Pinsker’s partial characterization in [3, Theorem 5].

Theorem 4 ($R_{\alpha}^n(D)$ of Gaussian Markov Processes With Block Letter Square Error Distortion): Consider the statement of Theorem 3(a), with $\Sigma_t^+ \in S_++^{p \times p}$, $t = 0, \ldots, n$. Then, the following statements hold.

(a) The realization of the reproduction distribution of Theorem 3(a) satisfies

$$Q_{Y_{t-1}}(X_t) = P\left(dy_t|y_{t-1}, \hat{X}_{t-1}|x_t\right)$$

(III.23a)

$$Q_{Y_0}|X_0 = Q_{Y_0}|X_0$$

(III.23b)

while the pay-off satisfies

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

(III.24a)

$$= \mathbb{E}^Q \left\{ \log \left( \frac{Q_{Y_0}|X_0}{P_{Y_0}^Q} \right) + \sum_{t=1}^n \log \left( \frac{Q_{Y_t|Y_{t-1}, X_t}}{P_{Y_t|Y_{t-1}}^Q} \right) \right\}$$

(III.24b)

$$= \mathbb{E}^Q \left\{ \log \left( \frac{Q_{Y_0}|X_0}{P_{Y_0}^Q} \right) + \sum_{t=1}^n \log \left( \frac{Q_{Y_t|Y_{t-1}, X_t}}{P_{Y_t|Y_{t-1}}^Q} \right) \right\}$$

(III.24c)

where

$$P_{Y_t|Y_{t-1}}^Q = \int Q_{Y_t|Y_{t-1}, X_t} \otimes P_{X_t|Y_{t-1}}^Q, \quad t = 1, \ldots, n$$

(III.25a)

$$P_{Y_0}^Q = \int Q_{Y_0|X_0} \otimes M_{X_0}.$$  

(III.25b)

(b) For any joint distribution $P_{X^n, Y^n}$ induced by the Gaussian Markov source $X^n$ and the realization of $Y^n$ given in Theorem 3(a), then the following inequality holds.

$$I(X_0; Y_0) + \sum_{t=1}^n I(X_t; Y_t|Y_{t-1}, \hat{X}_{t-1}|x_t) \geq I(X_0; \hat{X}_0|x_0) + \sum_{t=1}^n I(X_t; \hat{X}_t|Y_{t-1}, \hat{X}_{t-1}|x_t).$$

(III.26)

(c) If there exists $(H_t, K_{V_t}) \in \mathbb{R}^{p \times p} \times S_+^{p \times p}, t = 0, \ldots, n$, such that $\hat{X}_{t|t} = Y_t - a.s., t = 0, \ldots, n$, then the inequality (III.26) holds with equality, and

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

(III.27)

where $Y^n$ is Markov, the test channel distribution satisfies

$$Q_{Y_t|Y_{t-1}, X_t} = Q_{Y_t|\hat{X}_{t-1}|x_t} = Q_{Y_t|Y_{t-1}, X_t}, \quad t = 1, \ldots, n$$

$$Q_{Y_0}|X_0 = Q_{Y_0}|X_0.$$  

(III.28)

and

$$P_{Y_t|Y_{t-1}} = P_{Y_t|Y_{t-1}}$$

$$= \int Q_{Y_t|X_t, Y_{t-1}} \otimes P_{X_t|Y_{t-1}}, \quad t = 1, \ldots, n.$$  

Moreover, the nonanticipative RDF is given by

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

$$= \inf_{Q_{X^n, Y^n}^M} \mathbb{E}^M \left\{ \log \left( \frac{Q_{Y_0|X_0}^M}{P_{Y_0}^M} \right) + \sum_{t=1}^n \log \left( \frac{Q_{Y_t|Y_{t-1}, X_t}^M}{P_{Y_t|Y_{t-1}}^M} \right) \right\}$$

(III.29)

where

$$Q_{X^n, Y^n}^M(D) \triangleq \left\{ (H_t, K_{V_t}) \in \mathbb{R}^{p \times p} \times S_+^{p \times p}, t = 0, \ldots, n : \right.$$  

$$\hat{X}_{t|t} = Y_t - a.s., t = 0, \ldots, n,$$

$$\frac{1}{n+1} \mathbb{E}^M \left\{ \sum_{t=0}^{n} ||X_t - Y_t||^2_2 \right\} \leq D \right\}$$

(III.30)

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

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

(III.31)

$$Y_t = H_t X_t + (I - H_t) A_{t-1} Y_{t-1} + V_t,$$

(III.32)

$$Y_0 = H_0 X_0 + V_0.$$  

(III.33)

Further, the joint distribution of the process $(X^n, Y^n)$ is given by

$$P_{X^n, Y^n}^M = \mathbb{E}^M \left\{ P_{X^n, Y^n}^M \right\}$$

(III.34)

$$= \mathbb{E}^M \left\{ M_{X_0} \otimes Q_{X^n, Y^n}^M \otimes_{t=1}^{n} \left( Q_{Y_t|Y_{t-1}, X_t}^M \otimes M_{X_t|X_{t-1}} \right) \right\}.$$  

Proof: See Appendix B.}

Theorem 5 (below) establishes existence of the matrices $(H_t, K_{V_t}) \in \mathbb{R}^{p \times p} \times S_+^{p \times p}, t = 0, \ldots, n$, such that $\hat{X}_{t|t} = Y_t - a.s., t = 0, \ldots, n$ holds, i.e., such that Theorem 4(c) holds, by constructing these matrices.

Theorem 5 (Complete Characterization of $R_{\alpha}^n(D)$ of Gaussian Markov Processes With Block Symbol Square Error Distortion): Consider the statement of Theorem 3(a), with $\Sigma_t^+ \in S_+^{p \times p}, t = 0, \ldots, n$. Then, the following statements hold.

(a) Suppose there exists $(H_t, K_{V_t}) \in \mathbb{R}^{p \times p} \times S_+^{p \times p}, t = 0, \ldots, n$ such that (III.35), shown at the bottom of the next page, holds.

Then, the statements of Theorem 4(c) hold.
Moreover, if \( \text{cov} \left( Y_t, Y_t | Y^{t-1} \right) \geq 0 \) then Conditions 1 and 2 stated below are sufficient for the equalities \( \bar{X}_{t|t} = Y_t - a.s., t = 0, \ldots, n \) to hold.

**Condition 1:** \( \text{cov} \left( X_t, Y_t | Y^{t-1} \right) = \text{cov} \left( Y_t, Y_t | Y^{t-1} \right) \), \( t = 0, \ldots, n \). 

\( \Rightarrow I_t = \left( \Sigma_t \right)^{-1} \Sigma_t \) \( \Rightarrow K_{Yt} = K_{Yt}^{-1} \) \( = \Sigma_t - \Sigma_t \left( \Sigma_t^{-1} \right)^{-1} \Sigma_t \geq 0 \), \( t = 0, \ldots, n \), are given by

\[ H_t \Sigma_t^{-1} = \Sigma_t H_t = \Sigma_t, \quad \Sigma_t = \Sigma_t^{-1} = 0 \quad (\text{III.41a}) \]

**Condition 2:** \( E_t^Q \left\{ X_t | Y^{t-1} \right\} = E_t^Q \left\{ Y_t | Y^{t-1} \right\}, t = 0, \ldots, n \). 

\[ Y_t = H_t X_t + \left( I - H_t \right) A_{t-1} Y_{t-1} + V_t, t = 1, \ldots, n. \]  

(III.38)

\[ Y_0 = H_0 X_0 + V_0. \]  

(III.40)

where \( \left( H_t, K_{Yt} \right), t = 0, \ldots, n \), are defined below, satisfies \( \bar{X}_{t|t} = Y_t - a.s., t = 0, \ldots, n \).

Further, the representation of the reproduction process \( Y^n \), with \( \left( H_t, K_{Yt} \right), t = 0, \ldots, n \), defined below, satisfies \( \bar{X}_{t|t} = Y_t - a.s., t = 0, \ldots, n \).

\[ Y_t = H_t X_t + \left( I - H_t \right) A_{t-1} Y_{t-1} + V_t, t = 1, \ldots, n, \]  

(III.38)

\[ Y_0 = H_0 X_0 + V_0. \]  

(III.40)

where the constraint set \( \mathcal{Q}^M_{[0,n]} \) is specified by

\[ \mathcal{Q}^M_{[0,n]} \left( D \right) = \left\{ \Sigma_t \in S_+^{n \times n}, t = 0, \ldots, n : \right. \]

\[ \left. \Sigma_t \leq A_{t-1} \Sigma_{t-1} A_{t-1}^T + B_{t-1} K_{Wt} B_{t-1}^T, \quad \Sigma_0 \leq \Sigma_0 = K_{Xo}, \quad \right\} \quad \text{tr} \left( \Sigma_t \right) \leq D, \]  

(III.44)

**Proof:** See Appendix C.

**Remark 5:** We observe that the realization of Theorem 5 is such that

\[ E_t^Q \left\{ X_t | Y^{t-1} \right\} = E_t^Q \left\{ Y_t | Y^{t-1} \right\} = A_{t-1} Y_{t-1}, t = 1, \ldots, n, \]  

(III.40)

\[ E_t^Q \left\{ Y_t | Y^{t-1} \right\} = E_t^Q \left\{ Y_t | Y^{t-1} \right\} = A_{t-1} Y_{t-1}, t = 1, \ldots, n, \]

and similarly for \( t = 0 \). In view of the above properties, we can apply pre-processing after the source generates \( X^n = x^n \) and post-processing prior to the reconstruction process \( Y^n = y^n \), to give an alternative equivalent characterization of the \( R_{n}^a \) of Theorem 5, in terms of translated RVs, and a single-letter information theoretic measure.

**Corollary 1** (Equivalent Characterization of \( R_{n}^a \) of Gaussian Markov Processes With Block Letter Square Error Distortion): Consider the Gaussian Markov processes with block letter square error distortion of Theorem 5. Define

\[ Z_t \triangleq X_t - Y_t, \quad \tilde{X}_t \triangleq X_t - A_{t-1} Y_{t-1} \]

\[ \tilde{Y}_t \triangleq Y_t - A_{t-1} Y_{t-1}, \quad \tilde{X}_0 = X_0, \quad \tilde{Y}_0 = Y_0 \]  

(III.45)

for \( t = 1, \ldots, n \), where \( E_t^Q \left\{ X_t | Y^{t-1} \right\} = E_t^Q \left\{ Y_t | Y^{t-1} \right\} = E_t^Q \left\{ X_t | Y^{t-1} \right\} = E_t^Q \left\{ Y_t | Y^{t-1} \right\} = A_{t-1} Y_{t-1} \).

(a) The representation of Theorem 5.(a) is equivalently expressed as

\[ X_t = Y_t + Z_t, \quad t = 1, \ldots, n, \]  

(III.46)

\[ Z_t = \left( I - H_t \right) A_{t-1} Z_{t-1} + \left( I - H_t \right) B_{t-1} W_t - V_t, \]  

(III.47)

\[ Z_0 = X_0 - Y_0, \]  

(III.48)

\[ Y_t = H_t A_{t-1} Z_{t-1} + A_{t-1} Y_{t-1} + H_t B_{t-1} W_t + V_t, \]  

(III.49)

(b) The representation of \( Y^n \) in part (a) induces a distribution \( Q_{Y^n | Y^{t-1}} X_n = Q_{Y^n | Y^{t-1}} X_n = Q_{Y^n | Y^{t-1}} X_n = Q_{Y^n | Y^{t-1}} X_n \) \( t = 0, \ldots, n \), which achieves the nonanticipative RDF \( R_{n}^a \) of (III.43).

(c) The complete characterization of the nonanticipative RDF is given by (III.43), shown at the bottom of the page.
where $W_t$ is independent of $Z_{t-1}$ for $t = 0, \ldots, n$ and (III.41a), (III.41b) hold.

(b) The processes $(\tilde{X}^n, \tilde{Y}^n, Z^n)$ satisfy the following recursions.

\[ \tilde{X}_t = A_{t-1}Z_{t-1} + B_{t-1}W_t, \tilde{X}_0 = X_0, t = 2, \ldots, n, \quad (III.51) \]

$Z_t$ satisfies (III.47),

\[ \tilde{Y}_t = H_tA_{t-1}Z_{t-1} + H_tB_{t-1}W_t + V_t, t = 1, \ldots, n, \quad (III.52) \]

\[ = H_t\tilde{X}_t + V_t, \quad (III.54) \]

\[ \tilde{Y}_0 = H_0\tilde{X}_0 + V_0, \quad (III.55) \]

where $W_t$ is independent of $Z_{t-1}$, $(H_t, K_{V_t}), t = 0, \ldots, n$, are given by (III.41), and

\[ K_{(\tilde{X}_t, \tilde{Y}_t)} = E^{Q^M}\left\{ \begin{pmatrix} \tilde{X}_t \\ \tilde{Y}_t \end{pmatrix} \begin{pmatrix} \tilde{X}_t \\ \tilde{Y}_t \end{pmatrix}^T \right\} = \begin{pmatrix} \Sigma_{\tilde{X}_t/\tilde{Y}_t} & \Sigma_{\tilde{X}_t/\tilde{Y}_t}^T \\ \Sigma_{\tilde{Y}_t/\tilde{X}_t}^T & \Sigma_{\tilde{Y}_t/\tilde{X}_t} \end{pmatrix} \]

(III.56)

Further, the pay-off and average block letter square error distortion of the characterization of nonanticipative RDF are equivalently expressed as

\[ I(X_0; Y_0) + \sum_{t=1}^n I(X_t; Y_t | Y_{t-1}) = I(\tilde{X}_0; \tilde{Y}_0) + \sum_{t=1}^n I(\tilde{X}_t; \tilde{Y}_t) \]

\[ = E^{Q^M}\left\{ \log \frac{Q^M_{Y_t|X_t}}{P^M_{Y_t|X_t}} \right\} + \sum_{t=1}^n \log \frac{Q^M_{Y_t|X_t}}{P^M_{Y_t|X_t}} \right\} \]

\[ = \frac{1}{2} \log \left\{ \left| \frac{\Sigma_{\tilde{X}_t/\tilde{Y}_t}}{\Sigma_{\tilde{Y}_t/\tilde{X}_t}} \right| \right\}^+ + \sum_{t=1}^n \log \left| A_t \right| + \sum_{t=1}^n \log \left| A_t^{-1} \right| \]

\[ = E^{Q^M}\left\{ \sum_{t=0}^n \|X_t - Y_t\|^2 \right\} = E^{Q^M}\left\{ \sum_{t=0}^n \|Z_t\|^2 \right\} \]

\[ = \sum_{t=0}^n \tr(\Sigma_t), \quad (III.59) \]

where $A_t = A_{t-1}\Sigma_{t-1}^{-1}A_{t-1}^T + B_{t-1}K_{W_t}B_{t-1}^T \geq \Sigma_t$, $t = 1, \ldots, n$, $K_{X_0} = \Sigma_0$, $K_{V_t} = \Sigma_t$, that is, $P_{Y_t|X_{t-1}}^{Q^M} = P_{Y_t|X_{t-1}}^{Q^M} - a.s.$ and

\[ P_{Y_t|X_{t-1}}^{Q^M} = Q^M_{Y_t|X_{t-1}} - a.s. \]

\[ \text{Proof:} \quad \text{(a) By representation of Theorem 5 and the definition of } Z^n, \text{the equivalence is obtained.} \]

\[ \text{(b) Recursions (III.51)-(III.55) follow from part (a). By a property of mutual information then} \]

\[ I(X_0; Y_0) + \sum_{t=1}^n I(X_t; Y_t | Y_{t-1}) \]

\[ = I(\tilde{X}_0; \tilde{Y}_0) + \sum_{t=1}^n I(X_t - A_{t-1}Y_{t-1}; Y_t - A_{t-1}Y_{t-1} | Y_{t-1}) \]

In view of the above identities, the equivalent characterization follows from Theorem 5(b).

Remark 6 (On Corollary 1): It is obvious from Corollary 1, that the translated realization in (III.54) corresponds to the memoryless realization, while the pay-off and average distortions (III.58) and (III.59), remain invariant w.r.t. the translation. In Fig. 4, we show a block diagram of the realization for $Y^n$ of nonanticipative RDF $R_{0,\infty}^n(D)$ for multivariate Gaussian sources, which is equivalent to an innovations encoder in which the past reproductions are fed back through a noiseless channel and used for the encoding.

In the next theorem, we identify structural properties of the realization coefficients $\{H_t, K_{V_t}, \Sigma_t, \Sigma_t^\circ \}$; these are further used to obtain the complete characterization of the nonanticipative RDF.

Theorem 6 (Achievable Lower Bound-Structural Properties of Realization Of Theorem 5): Consider the characterization of nonanticipative RDF given in Theorem 5, and assume $\Sigma_t^\circ \in S_{++}^{p\times p}, t = 0, \ldots, n$. Then, the following statements hold.

(a) For any $\{\Sigma_t^\circ, \Sigma_t^\circ \}$ that belongs to the constraint set $Q_{[0,n]}^M(D)$ then

\[ \frac{1}{2} \sum_{t=0}^n \log \left\{ \left| \frac{\Sigma_t}{\Sigma_t^\circ} \right| \right\}^+ + \frac{1}{2} \log \left(\frac{1}{I - (A_t)^{-1}U_t(\Sigma_t^\circ - \Sigma_t)U_t^T}\right)^+, \Sigma_t^\circ \geq \Sigma_t \]

Fig. 4. Block diagram of the feedback realization of $Y^n$ of nonanticipative RDF $R_{0,\infty}^n(D)$ for multivariate Gaussian sources.
where $\Lambda_t$ is the diagonal matrix of the spectral representation of the symmetric positive definite matrix $\Sigma_t^-$, that is,

\[
(\Sigma_t^-)^{-1} = U_t \Lambda_t^{-1} U_t^T, \quad \Lambda_t^{-1} = \text{diag}\left\{ \frac{1}{\lambda_{t,p}}, \ldots, \frac{1}{\lambda_{t,1}} \right\},
\]

(III.63a)

\[
U_t U_t^T = I, \quad U_t^T U_t = I,
\]

(III.64a)

\[
\frac{1}{\lambda_{t,p}} \geq \frac{1}{\lambda_{t,p-1}} \geq \ldots \geq \frac{1}{\lambda_{t,1}} > 0, \quad t = 0, \ldots, n.
\]

(III.64b)

\[
(b) \text{ Perform the spectral representation of } \Sigma_t^- - \Sigma_t \succeq 0 w.r.t. \text{ the unitary matrix } V_t, \text{ for } t = 0, \ldots, n, \text{ that is,}
\]

\[
\Sigma_t^- - \Sigma_t = V_t D_{\Sigma_t^- - \Sigma_t} V_t^T,
\]

(III.65a)

\[
D_{\Sigma_t^- - \Sigma_t} = \text{diag}\{\sigma_{t,1}, \ldots, \sigma_{t,p}\}, \quad V_t V_t^T = I, \quad V_t^T V_t = I,
\]

(III.65b)

\[
\sigma_{t,i} \geq \sigma_{t,2} \geq \ldots \geq \sigma_{t,p} \geq 0, \quad t = 0, \ldots, n.
\]

(III.65c)

\[
\sigma_{t,i} > 0, \quad i = 1, \ldots, p_{t,12}, \quad t = 0, \ldots, n.
\]

(III.65d)

where $p_{t,12}$ is the number of positive singular values, that depends also on $D \in [0, \infty)$, in addition to the indicated $t$.

The inequality in (III.63) holds with equality if and only if the spectral representation of $\Sigma_t^- - \Sigma_t \succeq 0$ is expressed w.r.t. the same unitary matrix of $\Sigma_t^-$, that is,

\[
U_t = V_t, \quad t = 0, \ldots, n.
\]

(III.66)

(c) If (III.66) holds, then the lower bound of part (a) is achievable, and given by

\[
\sum_{t=0}^{n} \log \left\{ \frac{1}{|\Sigma_t^-|} \right\}^+, \quad I \succeq (\Lambda_t)^{-1} D_{\Sigma_t^- - \Sigma_t} \succeq 0
\]

(III.67a)

\[
= \sum_{t=0}^{n} \log \left\{ \frac{1}{|I - (\Lambda_t)^{-1} D_{\Sigma_t^- - \Sigma_t}|} \right\}^+, \quad I \succeq (\Lambda_t)^{-1} D_{\Sigma_t^- - \Sigma_t} \succeq 0
\]

(III.67b)

\[
= \sum_{t=0}^{n} \log \left\{ \frac{1}{\prod_{i=1}^{p} \left(1 - \frac{1}{\lambda_{t,i}^{-1}} \sigma_{t,i} \right)} \right\}^+, \quad 1 \geq \lambda_{t,i}^{-1} \sigma_{t,i} \geq 0, \quad i = 1, \ldots, p
\]

(III.67c)

\[
= \sum_{t=0}^{n} 2I(\tilde{X}_t; \tilde{Y}_t)
\]

(III.67d)

where we have (III.68) and (III.69), shown at the bottom of the page, and where $d_{t,i} \in (0, 1)$ are the canonical correlation coefficients of the $p_{t,12}$-correlated components $X_{t}^c \in \mathbb{R}^{p_{t,12}}$ and $Y_{t}^c \in \mathbb{R}^{p_{t,12}}$ of the transformed vectors of $X_t$ and $Y_t$, by two linear nonsingular transformations $(T_1, T_2)$, $\tilde{X}_{t} \mapsto T_1 \tilde{X}_t, \tilde{Y}_t \mapsto T_2 \tilde{Y}_t$, into their canonical variable form (see [35], [36] for full specification), i.e., the correlated components of $T_1 \tilde{X}_t$ and $T_2 \tilde{Y}_t$ are $X_{t}^c \in \mathbb{R}^{p_{t,12}}$ and $Y_{t}^c \in \mathbb{R}^{p_{t,12}}$, respectively, and

\[
d_{t,i} = d_{t,i}(\tilde{X}_{t}^c, \tilde{Y}_{t}^c) = \frac{\mathbb{E}\left\{ \tilde{X}_{t,i}^c \tilde{Y}_{t,i}^c \right\}}{\sqrt{\mathbb{E}\left\{ \tilde{X}_{t,i}^c \right\}^2 \mathbb{E}\left\{ \tilde{Y}_{t,i}^c \right\}^2}}, \quad i = 1, \ldots, p_{t,12}, \quad t = 0, \ldots, n.
\]

(III.70)

Proof: See Appendix D.

By Theorem 6 and in view of Remark 3, if for some $t$ and $D \in [0, \infty)$, $d_{t,i}^2 = 1$ then this should be removed from (III.67), because it gives a value of $+\infty$. In fact, $d_{t,i}^2 = 1$ (resp. $d_{t,i}^2 = 0$) are the canonical correlation coefficients that correspond to the identical (resp. independent) components of the transformed vectors $\tilde{X}_t \in \mathbb{R}^p$ and $\tilde{Y}_t \in \mathbb{R}^p$ into their canonical variable form (see [35], [36]). In view this we have the following main characterization.

\textbf{Theorem 7 (Complete Characterization of $R_n^{na}(D)$).} Consider the Gaussian Markov process $X^n \triangleq \{X_0, \ldots, X_n\}$ of Definition 1, and a block symbol square-error distortion

\[
d_{0,n}(x^n, y^n) \triangleq \sum_{t=0}^{n} ||x_t - y_t||^2, \quad \text{and assume } \Sigma_t^- \in S_{++}^{p \times p}, \quad t = 0, \ldots, n.
\]

Then, the following statements hold.

(a) The characterization of RDF $R_n^{na}(D)$ is given by the solution of the optimization problem,

\[
R_n^{na}(D) = \inf_{0 \leq d_{t,i}^2 \leq 1, t = 0, \ldots, n} \sum_{t=0}^{n} \frac{1}{2} \sum_{i=1}^{p} \log \left( 1 - d_{t,i}^2 \right), \quad \sum_{t=0}^{n} \frac{1}{2} \sum_{i=1}^{p} \log \left( 1 - d_{t,i}^2 \right) \leq D.
\]

(III.71)

\[
d_{t,i}^2 \triangleq \lambda_{t,i}^{-1} \sigma_{t,i}, \quad \forall D \in (0, D_{\text{max}}] \subseteq (0, \infty).
\]

(III.72)

(b) Let $(H_t^*, K_{V_t}, \Sigma_t^*, \Sigma_t^{* -})$, $t = 0, \ldots, n$ (which depends on $D$), denote the optimal values of $R_n^{na}(D)$ of part (a). Then, the realization $Y^n = Y^{*+}$ given below achieves $R_n^{na}(D)$.

\[
Y_t^* = H_t^* X_t + (I - H_t^*) A_{t-1} Y_{t-1}^* + V_t^*,
\]

(III.73)

\[
= H_t^* A_{t-1} (X_{t-1} - Y_{t-1}^*) + A_{t-1} Y_{t-1}^* + H_t^* B_{t-1} W_t + V_t^*, \quad t = 1, \ldots, n,
\]

(III.74)

\[
Y_0^* = H_0^* X_0 + V_0^*,
\]

(III.75)

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

\[
H_t^* \Sigma_t^{-} \triangleq \Sigma_t^* - \Sigma_t \succeq 0, \quad t = 0, \ldots, n,
\]

(III.76a)

\[
K_{V_t} \triangleq (K_{V_t})^T \triangleq \Sigma_t^{-} H_t^* - H_t^* \Sigma_t^{-} H_t^* \quad \text{is positive semi-definite,}
\]

(III.76b)

\[
\Sigma_t^{* -} \triangleq A_{t-1} \Sigma_t^{* -} A_{t-1}^T + B_{t-1} K_{W_t} B_{t-1}^T, \quad t = 0, \ldots, n.
\]

(III.76c)
with $\Sigma_{0}^{\ast\ast} = KX_{0}$ and where the following property holds: 
$\{\Sigma_{t}^{\ast\ast}, H_{t}^{\ast\ast}, \Sigma_{t}^{\ast\ast} \}$ have a spectral decomposition w.r.t. the same unitary matrix $U_{t}^{\ast\ast} U_{t}^{\ast\ast T} = I$, as given in Theorem 6. Moreover, the optimal test channel satisfies the following properties.

\begin{align}
(i) \quad & P_{X_{t}|Y_{t}^{n-1}} = P_{X_{t}|Y_{t}^{n-1}} \quad \text{(III.77)} \\
(ii) \quad & E\{X_{t}^{T} Y_{t}^{n-1} \} = E\{X_{t}^{T} Y_{t}^{n-1} \} = Y_{t}^{n-1} \quad \text{(III.78)} \\
& \implies E\{X_{t}^{T} Y_{t}^{n-1} \} = E\{Y_{t}^{n-1} \} \quad \text{(III.79)}
\end{align}

The realization of the optimal $Y_{t}^{*,n}$ that achieves $R_{n}^{a}(D)$ is shown in Fig. 5.

**Proof:** Follows from Theorem 6 while $D_{\max}$ is shown in Lemma 4.

Next, we show how to treat the case $\Sigma_{t}^{\ast\ast} \geq 0$ but not $\Sigma_{t}^{\ast\ast} > 0$ for any $t = 0, \ldots, n$.

**Lemma 2:** On the Structure of Covariance Matrix of Realization of Corollary 1 Consider Corollary 1 and the covariance matrix $K_{\{X_{t}^{\ast}, Y_{t}^{\ast}\}}$ of the vector $(X_{t}^{\ast}, Y_{t}^{\ast})^{T}$, for $t = 0, \ldots, n$, specified by (III.56), where $\Sigma_{t}^{\ast\ast} \geq 0$ but not $\Sigma_{t}^{\ast\ast} > 0$ for any $t = 0, \ldots, n$. Then, the following statement holds. The pseudoinverse of $\Sigma_{t}^{\ast\ast}$ $(\Sigma_{t}^{\ast\ast})^{1}$, has the same decomposition with $\Sigma_{t}^{\ast\ast}$, with the difference that some of the eigenvalues are zero, i.e.,

\begin{align}
(\Sigma_{t}^{\ast\ast})^{1} &= U_{t}A_{t}^{1}U_{t}^{T}, \\
A_{t}^{1} &= \text{diag}\{\lambda_{t,1}, \ldots, \lambda_{t,1}\}, \\
U_{t}U_{t}^{T} &= I, \\
U_{t}^{T}U_{t} &= I, \\
\lambda_{t,p}^{1} &\geq \lambda_{t,p-1}^{1} \geq \ldots \geq \lambda_{t,1}^{1} \geq 0, \\
\lambda_{t,1}^{1} &> 0, i = p, p-1, \ldots, p_{t,\lambda}, t = 0, \ldots, n,
\end{align}

where $\lambda_{t,i}^{1} = \lambda_{t,i}^{-1}$ for $i = p, p-1, \ldots, p_{t,\lambda}, t = 0, \ldots, n$, and where $p_{t,\lambda}$ depends on $D$. Therefore, all the aforementioned results hold with the appropriate changes.

**Proof:** Follows from properties of pseudoinverse of positive semidefinite matrices.

The next lemma gives a further, generally sub-optimal structural properties of the realization matrices $\Sigma_{t}^{\ast\ast}$, $\Sigma_{t}$, $H_{t}$, $K_{V_{t}}$, $t = 0, \ldots, n$, of Theorem 5, which imply $\Sigma_{t}^{\ast\ast}$, $\Sigma_{t}$, $H_{t}$, $K_{V_{t}}$ have spectral decomposition w.r.t. the same unitary matrix $U_{t}U_{t}^{T} = I$, for $t = 0, \ldots, n$.

**Lemma 3 ( Generally Sub-Optimal Structural Properties of Realization Of Theorem 5):** Consider the optimal reproduction distribution, and its realization in Theorem 5.

Suppose $\Sigma_{t}^{\ast\ast}$ defined by (III.42) commutes with $\Sigma_{t}$, that is,

$$\Sigma_{t}^{\ast\ast}\Sigma_{t} = \Sigma_{t}\Sigma_{t}^{\ast\ast}, \quad t = 0, \ldots, n, \quad \text{(III.80)}$$

where

$$\Sigma_{t}^{\ast\ast} = A_{t-1}\Sigma_{t-1}A_{t-1}^{T} + B_{t-1}K_{W_{t}}B_{t-1}^{T}, \quad t = 1, \ldots, n,$$

and $\Sigma_{0}^{\ast\ast} = KX_{0}$. Then,

$$H_{t}^{\ast\ast} = I - \Sigma_{t}^{\ast\ast}(\Sigma_{t}^{-1})^{-1} = H_{t}^{\ast\ast} \in S^{p \times p}_{+},$$

$$K_{V_{t}}^{\ast\ast} = \Sigma_{t}^{\ast\ast}H_{t}^{\ast\ast} = \Sigma_{t}^{\ast\ast}H_{t} = K_{V_{t}}^{\ast\ast} \in S^{p \times p}_{+}, \quad t = 0, \ldots, n, \quad \text{(III.81)}$$

that is, $H_{t}^{\ast\ast}$ is also a symmetric positive semi-definite matrix, and moreover $\{\Sigma_{t}, \Sigma_{t}^{\ast\ast}, H_{t}, K_{V_{t}}^{\ast\ast}\}$ have a spectral decomposition w.r.t. the same unitary matrix $U_{t}U_{t}^{T} = I$, $U_{t}^{T}U_{t} = I$, for $t = 0, \ldots, n$.

**Proof:** See Appendix E.

With the aid of Lemma 3 we can obtain an upper bound on the characterization of the nonanticipative RDF $R_{n}^{a}(D)$, where the matrices $\{H_{t}, K_{V_{t}}, \Sigma_{t}, \Sigma_{t}^{\ast\ast}\}$ are diagonalized w.r.t. to the same unitary matrix.

**Theorem 8 (Upper Bound on the Characterization of $R_{n}^{a}(D)$):** Consider the Gaussian Markov process $X^{n} \triangleq \{X_{0}, \ldots, X_{n}\}$ of Definition 1, and a block symbol square-error distortion $d_{0,n}(x^{n}, y^{n}) = \sum_{t=0}^{n} ||x_{t-1} - y_{t}||^{2}$, and assume $\Sigma_{t}^{\ast\ast} \in S^{p \times p}_{+}, t = 0, \ldots, n$.

An upper bound on the RDF $R_{n}^{a}(D)$ is given by the solution of the optimization problem

$$R_{n}^{a,up}(D) = \inf_{D_{\max}, \Sigma_{t}, \Sigma_{t}^{\ast\ast}} \frac{1}{2} \sum_{t=0}^{n} \log \frac{A_{t}}{A_{t}}, \quad \text{(III.82)}$$

where

$$\Sigma_{t}^{\ast\ast} = A_{t-1}\Sigma_{t-1}A_{t-1}^{T} + B_{t-1}K_{W_{t}}B_{t-1}^{T}, \quad \text{(III.83)}$$

$$\Lambda_{t} = \text{diag}\{\delta_{t,1}, \ldots, \delta_{t,p}\},$$

$$\Sigma_{t} = U_{t}\Delta_{t}U_{t}^{T}, \quad \Delta_{t} = \text{diag}\{\delta_{t,1}, \ldots, \delta_{t,p}\}. \quad \text{(III.84)}$$

**Proof:** Follows from Theorem 5 and Lemma 3.

**C. Characterization of $R_{\max}(D_{0}, \ldots, D_{n})$ of Multivariate Gaussian Markov Sources With Letter-by-Letter Square Error Distortion**

Clearly, the realizations and optimal test channels of $R_{n}^{a}(D)$ also hold for the characterization of $R_{n}^{a}(D_{0}, D_{1}, \ldots, D_{n})$ as given in the next corollary.

**Corollary 2 (Complete Characterization of $R_{n}^{a}(D_{0}, D_{1}, \ldots, D_{n})$):** Consider the Gaussian Markov process $X^{n} \triangleq \{X_{0}, \ldots, X_{n}\}$ of Definition 1, and a letter-by-letter square-error distortion $d_{t}(x_{t}, y_{t}) = ||x_{t} - y_{t}||^{2}$, $t = 0, \ldots, n$, and assume $\Sigma_{t}^{\ast\ast} \in S^{p \times p}_{+}, t = 0, \ldots, n$. Then, the following hold.
(a) $R_n^{\alpha}(D_0, D_1, \ldots, D_n)$ is given by the solution of the optimization problem
\[
R_n^{\alpha}(D_0, D_1, \ldots, D_n) = \inf_{D \in [0, D_{\max}]} \frac{1}{n} \sum_{t=0}^{n} \log \left| \frac{\Sigma_t}{\Sigma_{t-1}} \right|^{+},
\]
where the constraint set $Q_{[0,n]}(D_0, D_1, \ldots, D_n)$ is specified by
\[
\forall M \in Q_{[0,n]}(D_0, D_1, \ldots, D_n)
\]
\[
\bigcup_{t=0}^{n} \left\{ \Sigma_t \in S_{+}^{p \times p}, \ t = 0, \ldots, n : \
\Sigma_t \preceq \Sigma_{t-1} \triangleq A_{t-1}^* \Sigma_{t-1} A_{t-1} + B_{t-1} K W B_{t-1}^\top, \ 
\Sigma_0 \preceq K_{X_0}, \ t = 1, \ldots, n, \ \text{tr}(\Sigma_t) \leq D_t, \ t = 0, \ldots, n \right\}.
\]

(b) The realization of $Y^n$ which induces the optimal test channel of the RDF $R_n^{\alpha}(D_0, D_1, \ldots, D_n)$ is given in Theorem 5 (a), (b).

(c) The analog of Theorem 7 also holds for $R_n^{\alpha}(D_0, D_1, \ldots, D_n)$.

\textbf{Proof:} Follows from Theorem 5 and Theorem 6. \hfill \blacksquare

Next lemma is useful for the identification of the value of $D_{\max}$, if such exists in $[0, \infty)$, beyond which the value of $R_n^{\alpha}(D) = 0$.

\textbf{Lemma 4:} Consider the characterization of nonanticipative RDF given in Theorem 5, (III.43). Then, the following statements hold.

(a) Let $P : \{0, 1, \ldots, n\} \rightarrow \mathbb{R}^{p \times p}$ denote the covariance of the Gaussian Markov process $\{X_t : t = 0, \ldots, n\}$ defined by the representation (I.3), i.e., $P_t \triangleq \text{cov}(X_t, X_t)$. Then, $P_t$ satisfies the recursion
\[
P_t = A_{t-1} P_{t-1} A_{t-1}^\top + B_{t-1} K W B_{t-1}^\top, \ t = 1, \ldots, n,
\]
where $P_0$ is given.

(b) For any element of the set $Q_{[0,n]}(D)$ defined by (III.44), and $\Sigma_t, t = 0, \ldots, n$ defined by (III.42), the following inequality holds:
\[
P_t \succeq \Sigma_t \succeq 0, \ t = 1, \ldots, n.
\]

(c) For a fixed $D \in [0, D_{\max}] \subseteq [0, \infty)$, then a necessary condition for $R_n^{\alpha}(D) \in [0, \infty)$ is $\Sigma_t \succeq \Sigma_t, t = 0, \ldots, n$.

(d) If at some $t \in \{1, \ldots, n\}$, $H_t = 0$, equivalently, $\Sigma_t = \Sigma_t^-$, then the following hold:
\[
Y_t = A_{t-1} Y_{t-1}, \ 
\Sigma_t^+ = A_t \Sigma_t A_t^\top + B_{t-1} K W B_{t-1}^\top, \ 
I(X_t; Y_t|Y_{t-1}) = 0.
\]

For $D \in [D_{\max}, \infty]$ then $R_n^{\alpha}(D_{\max}) = 0$, where $D_{\max}$ is given by
\[
D_{\max} \triangleq \frac{1}{n+1} \left\{ \text{tr} \left( \Sigma_0^- \right) + \frac{n}{n-1} \text{tr} \left( A_{t-1} \Sigma_t^- A_{t-1}^\top + B_{t-1} K W B_{t-1}^\top \right) \right\}.
\]

\textbf{D. Simple Upper Bound Solution of $R_n^{\alpha}(D_0, \ldots, D_n)$ of Gaussian Markov Sources Using Water-Filling}

Now, we turn our attention to optimization problem of $R_n^{\alpha}(D_0, \ldots, D_n)$ of Corollary 2, to give a simple upper bound based on the solution of a sequential water-filling over the spatial dimensions.

\textbf{Proposition 2:} Consider the complete characterization of $R_n^{\alpha}(D_0, D_1, \ldots, D_n)$ of Corollary 2.

An upper bound on $R_n^{\alpha}(D_0, \ldots, D_n)$ is determined as follows:
\[
R_n^{\alpha}(D_0, \ldots, D_n) \leq \sum_{t=0}^{n} R_t^{\alpha,up}(D_t),
\]
where
\[
R_t^{\alpha,up}(D_t) = \frac{1}{2} \log \left| \frac{\Sigma_t^\top}{\Sigma_t} \right| + \frac{1}{2} \sum_{j=1}^{p} \log \left| \frac{\lambda_{t,j}^*}{\theta_{t,j}} \right|^{+},
\]
and where $U_t, t = 0, \ldots, n$ are unitary matrices such that
\[
\Sigma_t^\top = A_{t-1} \Sigma_t^+ A_{t-1}^\top + B_{t-1} K W B_{t-1}^\top
\]
\[
= U_t A_t^\top U_t^\top, \quad \Sigma_t^* = U_t \Delta_t^* U_t^\top, \quad \Sigma_t^\top = K_{X_0}, \quad A_t^\top \triangleq \text{diag}(A_{t,1}^*, \ldots, A_{t,p}^*),
\]
\[
\Delta_t^\top \triangleq \text{diag}(d_{t,1}^*, \ldots, d_{t,p}^*), \ t = 0, \ldots, n.
\]
with diagonal entries in decreasing order, and for $t = 1, \ldots, n$, where the constants $\theta_{t} > 0$, $t = 0, \ldots, n$ are
uniquely determined by
\[
\sum_{j=1}^{p} \min \{ \lambda_{0,j}, \theta_{0} \} = D_{0}, \tag{III.97}
\]
and
\[
\sum_{j=1}^{p} \min \{ \lambda_{t,j}, \theta_{t} \} = D_{t}, \quad t = 1, \ldots, n. \tag{III.98}
\]

Proof: Follows directly from Corollary 2, i.e., by considering the optimization problem \( R_{n}^{a}(D_{0}, \ldots, D_{n}) \), and solving the optimization sequentially forward in time, starting at stage \( t = 0 \) to obtain the solution of \( \inf_{\Delta_{0}} \operatorname{tr} \left( \Delta_{0} \right) \leq D_{0} \frac{1}{2} \log \left| \lambda_{0} \right| \) and then apply the optimal value of this stage \( \Delta_{0}^{*} \) to solve the optimization problem at stage \( t = 1 \), \( \inf_{\Delta_{1}} \operatorname{tr} \left( \Delta_{1} \right) \leq D_{1} \frac{1}{2} \log \left| \lambda_{1} \right| \), etc.

IV. STATE DEPENDENT NONANTICIPATORY EPSILON ENTROPY OF MULTIVARIATE GAUSSIAN MARKOV SOURCES WITH SQUARE ERROR FIDELITY

Now, we turn our attention to the nonanticipative epsilon entropy \( R_{n}^{GPS}(D) \) defined by (I.13)-(I.15). By Theorem 1, in general, \( R_{n}^{GPS}(D) \geq R_{n}^{a}(D) \).

In Theorem 9, we construct the auxiliary RVs \( S^{n} \) and \( Y_{t} = f_{t}(S^{t}, Y^{t-1}) \), \( t = 0, \ldots, n \), such that the equality \( R_{n}^{GPS}(D) = R_{n}^{a}(D) \) holds. Similar results also hold for the nonanticipative analog \( R_{n}^{a}(D_{0}, \ldots, D_{n}) \) with side information only to the decoder, i.e., defined similar to (I.14).

Theorem 9 (Complete Characterization of \( R_{n}^{GPS}(D) \)): Consider the Gaussian Markov process \( X^{n} \) of Definition 1, and block letter square error distortion function \( d_{n,n}(x^{n}, y^{n}) \triangleq \sum_{t=0}^{n} \| x_{t} - y_{t} \|^{2} \), and the state dependent nonanticipative epsilon entropy \( R_{n}^{GPS}(D) \) defined by (I.13)-(I.15). Consider also \( R_{n}^{a}(D) \) is given in Theorem 8(a). Then, the following statements hold.

(a) \( R_{n}^{GPS}(D) = R_{n}^{a}(D) \) and it is achieved by the realization \( Y_{t}^{*} = f_{t}^{*}(S_{t}^{*}, Y_{t-1}^{*}) \), \( t = 0, \ldots, n \), i.e.,

\[
Y_{t}^{*} = f_{t}^{*}(S_{t}^{*}, Y_{t-1}^{*}) = (I - H_{t}^{*}) A_{t-1} Y_{t-1}^{*} + S_{t}^{*}, \quad Y_{0}^{*} = S_{0}^{*},
\]

\[
= A_{t-1} Y_{t-1}^{*} + S_{t}^{*} - S_{t|t-1}^{*}, \tag{IV.3}
\]

\[
S_{t}^{*} = H_{t}^{*} (A_{t-1} X_{t-1} + B_{t-1} W_{t}) + V_{t}^{*},
\]

\[
= H_{t}^{*} (S_{t}^{*} + V_{t}^{*}), \tag{IV.5}
\]

\[
S_{t|t-1}^{*} = E \{ S_{t}^{*} | Y_{t-1}^{*} \} = H_{t} E \{ X_{t} | Y_{t-1}^{*} \} \tag{IV.6}
\]

\[
= H_{t} A_{t-1} Y_{t-1}^{*}, \tag{IV.7}
\]

\[
P_{S_{t}^{*}|X_{t},S_{t-1}^{*},\ldots,Y_{t-1}^{*}} = P_{S_{t}^{*}|X_{t}}, \tag{IV.8}
\]

\[
\{ \Sigma_{t}^{*}, \Sigma_{t}^{*}, H_{t}^{*}, K_{V_{t}^{*}} \} \text{ satisfy (III.76).} \tag{IV.9}
\]

and where the following property holds: \( \{ \Sigma_{t}^{*}, \Sigma_{t}^{*}, H_{t}^{*}, K_{V_{t}^{*}} \} \) have a spectral decomposition w.r.t. the same unitary matrix \( U_{t}^{*} U_{t}^{*} = I, U_{t}^{*}^{*} U_{t}^{*} = I \).

Moreover, the following structural properties hold:

(i) \( P_{X_{t}|Y_{t}^{*},Y_{t-1}^{*}} = P_{X_{t}|Y_{t}^{*}} = P_{X_{t}|Y_{t}^{*}} \), \( Y_{t}^{*} = f_{t}^{*}(S_{t}^{*}, Y_{t-1}^{*}) \) of part (a) that achieves \( R_{n}^{GPS}(D) = R_{n}^{a}(D) \) is realized by the nonfeedback scheme depicted in Fig. 6.

Proof: (a) The stated realization is an alternative representation of the realization of Theorem 7(b). Since \( Y_{t}^{*} = f_{t}^{*}(S_{t}^{*}, Y_{t-1}^{*}) = (I - H_{t}^{*}) A_{t-1} Y_{t-1}^{*} + S_{t}^{*} \), where \( S_{t}^{*} \) is specified by (IV.4)-(IV.9), then the test channel satisfies \( P_{S_{t}^{*}|X_{t},S_{t-1}^{*},\ldots,Y_{t-1}^{*}} = P_{S_{t}^{*}|X_{t}} \), as required by the definition of \( R_{n}^{GPS}(D) \). Moreover, since for each \( t \), the tuple \( (Y_{t}^{*}, Y_{t-1}^{*}) \) uniquely defines \( S_{t}^{*} \) for all \( t = 0, \ldots, n \), then Theorem 1(c), holds. (IV.10), (IV.11) are easily verified.

(b) This follows directly from the optimal realization of part (a), with state dependent \( S_{t}^{*} \), reproduction \( Y_{t}^{*} \). ■

Remark 7: Theorem 9 addressed Question 2; it states that there is no loss of performance if side information of past reproductions is available to the decoder and not to the encoder, compared to the Gorbunov and Pinsker nonanticipatory epsilon entropy, which uses side information at both the encoder and the decoder.

The connection of Theorem 9 to recursive causal filters, implemented using Fig. 3, is now obvious.

V. ACHIEVABILITY OF NONANTICIPATORY EPSILON ENTROPY AND MESSAGE GENERATION RATES WITH NONCAUSAL, CAUSAL AND ZERO DELAY CODES SUBJECT TO RATE LOSS

In this section, we describe the connection of the characterizations of nonanticipatory epsilon entropy and message generation rates to rate loss of causal and zero delay codes w.r.t. the OPTA by noncausal codes, and the construction of encoders and decoders which achieve such rate loss.

A. Asymptotic Analysis

First, we recall that Gorbunov and Pinsker [2] also analyzed, under certain conditions, the nonanticipatory
message generation rates,

\[ R^{GP}(D) \triangleq \lim_{n \to \infty} \frac{1}{n+1} R_n^{GP}(D), \]  
(V.1a)

\[ R^{GP}(D_0), ..., D_{\infty}) \triangleq \lim_{n \to \infty} \frac{1}{n+1} R_n^{GP}(D_0), ..., D_{n}) \]  
(V.1b)

provided \( R_n^{GP}(D) \in [0, \infty) \), \( R_n^{GP}(D_0), ..., D_{n}) \in [0, \infty) \) (i.e., existence of optimal test channels is assumed) and the limits exist.

Theorem 10 is analogous to a theorem of Gorbunov and Pinski [2].

**Theorem 10 (Asymptotic Limit):** Consider the nonanticipatory RDF \( R_n^{na}(D) \) of Definition 3, and assume the infimum exists and \( R_n^{na}(D) \triangleq \lim_{n \to \infty} \frac{1}{n+1} R_n^{na}(D) \) exists. Then the inequality (V.2), shown at the bottom of the page, holds, where the infimum is taken over all test channels \( Q_{Y_t|Y_ {t-1}, X_t}, t = 0, 1, ... \), such that the limits on the right hand side exist and the infimum also exists.

A similar inequality holds for \( R^{GP}(D_0), ..., D_{\infty}) \triangleq \lim_{n \to \infty} \frac{1}{n+1} R_n^{GP}(D_0), ..., D_{n}) \).

**Proof:** The proof follows from [2, Theorem 1].

By definition,

\[ \frac{1}{n+1} R_n^{na}(D) \leq \frac{1}{n+1} \mathbb{E}^Q \left\{ \sum_{t=0}^{n} \log \left( \frac{Q_{Y_t|Y_ {t-1}, X_t}}{P_{Y_t|Y_ {t-1}}} \right) \right\} \]  
(V.3)

for all elements of \( Q_{[0,n]}(D) \). From (V.3), then

\[ R^{na}(D) \triangleq \lim_{n \to \infty} \frac{1}{n+1} R_n^{na}(D) \]  

\[ \leq \lim_{n \to \infty} \frac{1}{n+1} \mathbb{E}^Q \left\{ \sum_{t=0}^{n} \log \left( \frac{Q_{Y_t|Y_ {t-1}, X_t}}{P_{Y_t|Y_ {t-1}}} \right) \right\}, \]  

where the limit of the right hand side is taken over all test channels \( Q_{Y_t|Y_ {t-1}, X_t}, t = 0, 1, ... \), such that \( \lim_{n \to \infty} \frac{1}{n+1} \mathbb{E}^Q \left\{ \sum_{t=0}^{n} \log \left( \frac{Q_{Y_t|Y_ {t-1}, X_t}}{P_{Y_t|Y_ {t-1}}} \right) \right\} \leq D \) is satisfied and the limit of the right hand side exists. Finally, taking the infimum over all test channels \( Q_{Y_t|Y_ {t-1}, X_t}, t = 0, 1, ... \), such that the two limits exists in the right hand side of (V.2) exist, we obtain the inequality.

**Remark 8:** Theorem 10 is not very useful, because it presupposes existence of the limits, which may not exist, especially if the sequence of test channels \( Q_{Y_t|Y_ {t-1}, X_t}, t = 0, 1, ... \) is time-varying.

### B. Noncausal Codes

Fig. 7 illustrates the operational definition of an achievable rate of noncausal codes, denoted by \( (n, |\mathcal{M}_n|, \Delta X), n = 0, 1, ... \), and defined as follows.

\[ R^{na}(D) \leq R^{na, +}(D) \triangleq \inf_{\lim_{n \to \infty} \frac{1}{n+1} \mathbb{E}^Q \left\{ \sum_{t=0}^{n} \log \left( \frac{Q_{Y_t|Y_ {t-1}, X_t}}{P_{Y_t|Y_ {t-1}}} \right) \right\} \leq D} \lim_{n \to \infty} \frac{1}{n+1} \mathbb{E}^Q \left\{ \sum_{t=0}^{n} \log \left( \frac{Q_{Y_t|Y_ {t-1}, X_t}}{P_{Y_t|Y_ {t-1}}} \right) \right\} \]  
(V.2)
is the optimal performance theoretically attainable (OPTA) by the noncausal codes.

In general, $R_{n}^{GP}(D) = R_{n}^{a}(D) \leq R_{n}^{GPS}(D)$ are upper bound on Shannon’s classical RDF [37], $R_{n}(D)$,

$$R_{n}(D) \leq R_{n}^{GP}(D) \leq R_{n}^{GPS}(D) \quad (V.11)$$
due to the additional nonanticipative or causality constraint (GP-2), i.e., (I.11), which is imposed on the optimal reproduction distribution $P_{Y|X}^{*}$ of $R_{n}^{GP}(D)$, and Theorem 1. Noncausal codes are often non-desirable, because they experience significant delays, which makes them unsuitable for delay sensitive applications.

C. Causal Codes

Causal codes is a sub-class of noncausal codes, introduced by Neuhoff and Gilbert [14], to circumvent the long coding and decoding delays. The encoder accepts $x^\infty \triangleq \{x_{0}, x_{2}, \ldots\}$ and produces a sequence of bits $y_{0}, y_{1}, y_{2}, \ldots$. The decoder accepts $y_{0}, y_{1}, y_{2}, \ldots$, and generates the reproductions $y^\infty \triangleq \{y_{0}, y_{1}, y_{2}, \ldots\}$ of $x^\infty$. The cascade of the encoder and the decoder, called causal reproduction coders is a system that characterizes the operation of the source code from $x^\infty$ to $y^\infty$. The reproduction coder is specified by the sequence of maps $\{f_{0}, f_{1}, f_{2}, \ldots\}$ such that the reproduction of $x_{n}$ by $y_{n}$ is determined by $y_{n} = f_{n}(x^\infty)$ for $n = 0, 1, 2, \ldots$.

1) Causal Reproduction Coder: The cascade of the encoder-decoder pair, is called a causal reproduction coder, if the sequence of measurable maps $\{f_{n}(\cdot) : n = 0, 1, \ldots\}$ satisfies [14]

$$f_{n}(x^\infty) = f_{n}(\hat{x}^\infty) \quad \text{if} \quad x^{n} = \hat{x}^{n}, \quad n = 0, 1, \ldots \quad (V.12)$$

At each $n$, a causal reproduction coder has the property that, at each $t = 0, 1, \ldots, n$, all the bits from which the decoder determines $y_{t}$ are functionally determined by $x_{0}, \ldots, x_{t}$, and assumes $\{y_{0}, y_{1}, \ldots, y_{t-1}\}$ are already generated, i.e., the reproduction symbol $y_{t}$ of $x_{t}$ depends on the present and past source symbols $\{x_{0}, \ldots, x_{t}\}$ but not on the future source symbols $\{x_{t+1}, x_{t+2}, \ldots\}$. Further, the encoder-decoder pair is defined w.r.t. uniquely decodable codes. It also follows from (V.12) that the encoder-decoder pair is causal if conditions (GP-2) stated in (I.11) holds, i.e., $X^\infty_{n+1} \leftrightarrow X^{n} \leftrightarrow Y^{n}, \quad n = 0, 1, \ldots$.

2) Operational Causal Rates: Let $l_{k}(X^\infty)$ denote the total number of bits received by the decoder at the time it produces the output sequence $y_{k}$ (under the assumption $y^{k-1}$ is already produced), given that $b_{0}, b_{1}, b_{2}, \ldots$ is the binary sequence produced by the encoder in response to the source sequence $x^\infty$. The operational rate of the causal reproduction coders is defined by [14]

$$r_{c}^{\text{op}^{+}}(D) \triangleq \inf_{Y_{n}} \lim_{n \to \infty} \frac{1}{n+1} \mathbb{E}\left\{l_{n}(X^\infty)\right\} \quad (V.13)$$

provided the limit exists. Neuhoff and Gilbert showed that the OPTA by causal reproduction coders is,

$$r_{c}^{\text{op}^{+}}(D) = r_{c}^{+}(D) \quad (V.14)$$

where $D > D_{\text{min}}$ and $D_{\text{min}}$ is a given threshold. To this date the computation of expressions $r_{c}^{\text{op}^{+}}(D), r_{c}^{+}(D)$ remains a challenging problem, even for independent and identically distributed sources. In view of the difficulty of computing $r_{c}^{+}(D)$, the following bounds derived in [10], using the converse source coding theorem, are often applied to evaluate the performance of causal codes.

$$R(D) \leq R_{n}^{a} + (D) \leq r_{c}^{+}(D) \quad (V.15)$$

D. Zero-Delay Codes

The above bounds can be extended to include zero-delay codes, as follows. Let $r_{zd}^{+}(D)$ denote the OPTA by the class of zero-delay codes [15]-[18], which impose the additional constraint that for each $n = 0, 1, \ldots$, the encoder outputs are specified by the measurable maps, $(m^{n-1}, x^{n}) \mapsto m_{n} = e^{zd}(m^{n-1}, x^{n})$. At each time $n$ message $m_{n}$ is transmitted through a noiseless channel to the decoder, which is specified by the measurable maps, $f_{n}^{zd}(\cdot, m^{n-1})$, such that $m_{n} \mapsto y_{n} = f_{n}^{zd}(m_{n}, m^{n-1})$, assuming $y^{n-1}$ is already generated. The encoder-decoder pair is defined w.r.t. instantaneous codes. It is further shown in [10] that the following bounds hold:

$$R(D) \leq R_{n}^{a} + (D) \leq r_{c}^{+}(D) \leq r_{zd}^{+}(D) \quad (V.16)$$

In view of the difficulty to evaluate $r_{c}^{+}(D)$ and $r_{zd}^{+}(D)$, the emphasis is often focused on the computation of their lower bound $R_{n}^{GPS} + (D)$.

We now have the following proposition that relates the operational rate of zero-delay codes to the state state dependent Gorbunov and Pinsker nonanticipatory entropy $R_{n}^{GPS}(D)$ and its rate, and gives a tighter bound than any previously known bounds.

Proposition 3 (Operational Zero-Delay Rate and State Dependent Gorbunov and Pinsker Nonanticipatory Entropy): Consider the zero-delay code, $m_{n} = e^{zd}(m^{n-1}, x^{n})$ and $y_{n} = h^{zd}(m_{n}, y^{n-1})$, for $n = 0, 1, \ldots$, with block length distortion $d_{0,n}(x^{n}, y^{n})$, and let $l_{n}(x^{n}) = \sum_{t=0}^{n} l_{t}(m_{t})$ denote the accumulated number of bits received by the decoder at the time of reproducing $y_{n}$. 

Fig. 7. Block diagram of a simple communication system.
CHARALAMBOUS satisfies, where the optimization is over the code where the infimum in the last inequality is taken w.r.t. an arbitrary noncausal code strategy. In addition, 

\[ \lim_{n \to \infty} \frac{1}{n+1} R_{n}^{GPS} (D) \geq \lim_{n \to \infty} \frac{1}{n+1} R_{n}^{GP} (D) \]  

(V.30)

provided the infimums and limits exist.

Proof: First, it is easy to verify that the code \( e_{t}^{id} (\cdot) \), \( h_{t}^{zd} (\cdot) \), \( t = 0, \ldots, n \) satisfies the conditional independence conditions of the set \( \mathcal{M}_{[0,n]}^{GPS} (D) \). (V.17)-(V.23) follow from standard properties of entropy and mutual information. Inequality (V.24) is due to \( I(Y^{t−1}; S_{t} | S^{t−1}) \geq 0 \). Inequality (V.25) is due to the fact that randomized strategies from the set \( \mathcal{M}_{[0,n]}^{GPS} (D) \) includes deterministic code strategies (these also belong to the set \( \mathcal{M}_{[0,n]}^{GPS} (D) \)). Equality (V.26) is shown in Appendix A, i.e., in the proof of Theorem 1.(a). Taking the infimum of both sides of (V.25) or (V.26) over the distortion constraint and the code maps, we obtain the first inequality in (V.28); the second is due to Theorem 1.(c). Taking the per unit time limit on both sides of (V.25) or (V.26), following by the infimum over the code strategies gives the first inequality in (V.30), and the second is due to Theorem 1.(c).

Remark 9: Clearly, the bound in (V.30) using \( R_{n}^{GPS} (D) \) is tighter than the bound using \( R_{n}^{GS} (D) \). However, Theorem 1 gives conditions for the two lower bounds to coincide.

E. Rate Loss of Causal and Zero Delay Codes

In recent years [10, Sections II, III], the information quantities \( R_{n}^{a,+} (D) \) and \( R_{n=0,\ldots,D_{\infty}} (D_{0}, \ldots, D_{\infty}) \), are used as explained above, to quantify the rate loss of causal [14] and zero delay codes, w.r.t. the OPTA by noncausal codes \( R(D) \). Therefore, it is necessary to know the expressions for the nonanticipative message generation rates \( R_{n}^{a,+} (D) \) and \( R_{n=0,\ldots,D_{\infty}} (D_{0}, \ldots, D_{\infty}) \) of sources with memory, which to this date were unknown.

1) For high resolution (small distortion) \( D \), an optimal causal code for stationary sources with finite entropy (differential) and square error distortion, consists of uniform quantizers followed by a sequence of entropy coders (see [38]). The rate loss due to causality (compared to noncausal codes) is given by the so-called “space-filling” loss of quantizers, which is at most \( \frac{1}{2} \log_{2} \left( \frac{2\pi e}{12} \right) \approx 0.254 \) bits/sample.

2) For zero-mean stationary Gaussian sources with square-error distortion (and bounded entropy rate), the OPTA by causal codes \( r_{n}^{+} (D) \) exceeds \( R_{n}^{a,+} (D) \) by less than approximately 0.254 bits/sample [10],

\[ r_{n}^{+} (D) \leq R_{n}^{a,+} (D) + \frac{1}{2} \log_{2} \left( \frac{2\pi e}{12} \right) \text{ bits/sample, (V.31)} \]

(provided the limiting expressions are defined) but no expression is given for \( R_{n}^{a,\pm} (D) \), except for the scalar-valued Gaussian source with letter-by-letter square-error distortion which is already computed in [3, Example 2].

3) For zero-mean stationary Gaussian sources with square-error distortion (and bounded entropy rate), the OPTA by zero-delay codes exceeds \( R_{n}^{a,+} (D) \) by less than approximately
codes w.r.t. noncausal codes. Applicable to calculate the rate loss of causal and zero-delay with square-error fidelity obtained in this paper, are directly anticipatory epsilon entropy of multivariate Gaussian sources, matrices

$$I_n = 3 \text{ bits/sample, and the following bound holds [10]:}$$

$$r^{+}(D) \leq D_{na,n}^{+}(D) + \frac{1}{2} \log_2 \left( \frac{2\pi e}{D^2} \right) + 1 \text{ bits/sample.}$$

(V.32)

We conclude by stating that the expressions of nonanticipatory epsilon entropy of multivariate Gaussian sources, with square-error fidelity obtained in this paper, are directly applicable to calculate the rate loss of causal and zero-delay codes w.r.t. noncausal codes.

VI. 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 Gaussian Markov source, as given in Proposition 2. Specifically, we choose $p = 3$, i.e., a 3-dimensional source $X_t$ and a time-horizon $n = 3$. Matrices $A_t$, $B_t$ and $K_{X_0}$ are chosen randomly. Additionally, $K_{W_t}$ and $K_{X_0}$ are randomly chosen covariance matrices. The solution is shown in Fig. 8.

In Figure 9, we plot the nonanticipative RDF of our proposed upper bound and of the solution given by [12] via Semidefinite programming (SDP) (although, strictly speaking no realization is provided in [12] which achieves it).

VII. CONCLUSION

The complete characterization of the Gorbunov and Pinsker nonanticipatory epsilon entropy for nonstationary, multivariate Gaussian Markov sources with block letter and letter-by-letter square-error distortion criteria is derived, based on the structural properties of the matrices of two stochastic realizations of the reproduction process that induce the optimal test channel reproduction distribution. One of the realizations is shown to be analogous to the Wyner-Ziv and Wyner [4], [5] RDF, with the nonanticipatory epsilon entropy defined with respect to an auxiliary random variable, with past reproductions available only at the decoder as causal side information. Additionally, an upper bound is provided by a simple sequential in time, water-filling solution across the spatial dimensions. It is shown via simulations that the upper bound is tight.

APPENDIX A

PROOF OF THEOREM 1

(a) Consider an element of $\mathcal{M}_{[0, n]}^{GPS}(D)$ and let $Y_t = f_t(S_t, Y_{t-1}), t = 0, \ldots, n$. Then

$$I(X^t; S_t|S^{t-1}, Y_{t-1}) = E \left\{ \sum_{t=0}^{n} \log \left( \frac{P_{S_t|S^{t-1}, X_t, Y_{t-1}}}{P_{S_t|S^{t-1}, Y_{t-1}}} \right) \right\}$$

$$= E \left\{ \sum_{t=0}^{n} \log \left( \frac{P_{S_t|S^{t-1}, Y_{t-1}}}{P_{S_t|S^{t-1}}} \right) \right\} \text{ by def. of } \mathcal{M}_{[0, n]}^{GPS}(D)$$

$$= E \left\{ \sum_{t=0}^{n} \log \left( \frac{P_{S_t|S^{t-1}, Y_{t-1}}}{P_{S_t|S^{t-1}}} \right) \right\}$$

$$= E \left\{ \sum_{t=0}^{n} \log \left( \frac{P_{S_t|S^{t-1}, Y_{t-1}}}{P_{S_t|S^{t-1}}} \right) \right\}$$

$$= \sum_{t=0}^{n} \left( I(X^t; S_t|S^{t-1}) - I(Y^t; S_t|S^{t-1}) \right).$$

The last equation establishes (II.11).

(b) Consider an element of $\mathcal{M}_{[0, n]}^{GPS}(D)$ and let $Y_t = f_t(S_t, Y_{t-1}), t = 0, \ldots, n$. Then, $I(X^t; S_t) = I(X^t; S_t|S^{t-1}, Y_{t-1}) - I(Y^t; S_t|S^{t-1})$.

(c) Consider an element of $\mathcal{M}_{[0, n]}^{GPS}(D)$ and let $Y_t = f_t(S_t, Y_{t-1}), t = 0, \ldots, n$. Then $Y_t : \Omega \to \mathbb{Y}_t, t = 0, \ldots, n$ is invertible and measurable for all $t = 0, \ldots, n$; then $I(X^t; S_t|S^{t-1}, Y^{t-1}) = 0$, and $I(X^t; Y_{t|S^{t-1}, Y^{t-1}}) = I(X^t; Y_{t|Y^{t-1}})$, for all $t = 0, \ldots, n$. Hence, part (b) is shown.

Hence, (A.3) establishes the inequality (II.12). Suppose the map $f_t(., s^{t-1}, y^{t-1}) : S_t \to \mathbb{Y}_t, f_t(s_t, s^{t-1}, y^{t-1}) = y_t$ is invertible and measurable for all $t = 0, \ldots, n$; then $I(X^t; S_t|S^{t-1}, Y^{t-1}) = 0$, and $I(X^t; Y_{t|S^{t-1}, Y^{t-1}}) = I(X^t; Y_{t|Y^{t-1}})$, for all $t = 0, \ldots, n$. Hence, part (b) is shown.

(c) Consider an element of $\mathcal{M}_{[0, n]}^{GPS}(D)$ and let $Y_t = f_t(S_t, Y_{t-1}), t = 0, \ldots, n$. Then $Y_t : \Omega \to \mathbb{Y}_t, t = 0, \ldots, n$ is

![Figure 8. Water-filling subject to letter-by-letter distortion in time-domain for $n = 3$ time units.](image)

![Figure 9. Nonanticipative RDF for different values of letter-by-letter distortion for the upper bound proposed and its numerical computation via SDP [12].](image)
also an element of $Q_{[0,n]}(D)$ defined by (II.1). Alternatively, the infimum of the right side of the inequality in (II.12) over $M_{[0,n]}(D)$ occurs at a distribution $Q_{Y|Y^t-1, X_t}$, $t = 0, \ldots, n$. Hence, by (II.12), and $f_t(s,t-1, y^{t-1}) := S_t \rightarrow Y_t$, $f_t(s,t-1, y^{t-1}) = y_t$ is invertible and measurable for all $t = 0, \ldots, n$,

$$
\sum_{t=0}^{n} I(X^t; S^t|S^{t-1}, Y^{t-1}) \geq \sum_{t=0}^{n} I(X^t; Y^t|S^{t-1}, Y^{t-1})
$$

$$
= \sum_{t=0}^{n} I(X^t; Y^t|S^{t-1}, Y^{t-1}) \geq R^n_{\alpha}(D).
$$

Taking the infimum of the last equality over $M_{[0,n]}(D)$ then inequality (II.14) is established. Equality in (II.14) holds if the reproduction process $Y^n \in Q_{[0,n]}(D)$ which achieves $\sum_{t=0}^{n} I(X^t; Y^t|S^{t-1}, Y^{t-1}) = R^n_{\alpha}(D)$ is such that $I(X^t; S^t|S^{t-1}, Y^{t-1}) = 0$ for all $t = 0, \ldots, n$; this is ensured by the property of the map $f_t(\cdot)$.

**APPENDIX B**

**PROOF OF THEOREM 4**

(a) By the realization of Theorem 3.3(a), (III.11), then

$$
Q_{Y|Y^{t-1}, X_t} = P(dy_t|y^{t-1}, \tilde{x}_{t-1|t-1}, x_t),
$$

(B.1)

since $\tilde{x}_{t-1|t-1}$ is a measurable function of $Y^{t-1}$, and by independence of $V_t$ and $Y^{t-1}$ we get

$$
Q_{Y|Y^{t-1}, X_t} = Q_{Y|Y^{t-1}, \tilde{x}_{t-1|t-1}, x_t}.
$$

(B.2)

The above shows (III.23). By (III.23) then (III.24), (III.25) are obtained.

(b) Consider the identity (III.24d). Since $\tilde{x}_{t|t}$ satisfies the Kalman filter recursion (III.14), we have the following bounds on the pay-off (III.24d).

$$
I(X_0; Y_0) = \sum_{t=1}^{n} I(X^t; Y^t|S^{t-1}, \tilde{x}_{t-1|t-1})
$$

$$
= I(X_0; Y_0, \tilde{x}_{0|0})
$$

$$
= I(X_0; Y_0) + \sum_{t=1}^{n} I(X^t; Y^t|S^{t-1}, \tilde{x}_{t-1|t-1})
$$

$$
= \sum_{t=1}^{n} I(X_t; Y_t|S^{t-1}, \tilde{x}_{t|t}, \tilde{x}_{t-1|t-1})
$$

$$
+ \sum_{t=1}^{n} I(X_t; Y^t|S^{t-1}, \tilde{x}_{t|t}, \tilde{x}_{t-1|t-1})
$$

$$
+ \sum_{t=1}^{n} I(X^t; Y^t|S^{t-1}, \tilde{x}_{t-1|t-1})
$$

$$
= I(X_0; \tilde{x}_{0|0}) + \sum_{t=1}^{n} I(X_t; Y_t|S^{t-1}, \tilde{x}_{t|t}, \tilde{x}_{t-1|t-1})
$$

$$
\geq I(X_0; \tilde{x}_{0|0}) + \sum_{t=1}^{n} I(X_t; Y_t|S^{t-1}, \tilde{x}_{t-1|t-1}),
$$

(B.7)

where (a) is due to the property of mutual information between two RVs, $X$ and $Y$, which states $I(X; Y) = I(X; Y, g(Y))$, for any measurable function $g(\cdot)$ [39], and applied to $I(X_0; Y_0)$, $I(X_t; Y^t|S^{t-1}, \tilde{x}_{t-1|t-1})$, $t = 1, \ldots, n$, using the fact that by the Kalman recursion, (III.15), $\tilde{x}_{t|t} = e_t(\tilde{x}_{t-1|t-1}, Y_t)$; (b) is due to the chain rule of mutual information between $X$ and $(Y, Z)$, which states $I(X; Y, Z) = I(X; Z|Y) + I(X; Y)$ [39], and applied to $I(X_0; Y_0, \tilde{x}_{0|0})$, $I(X_t; Y_t, \tilde{x}_{t|t}, \tilde{x}_{t-1|t-1})$, $t = 1, \ldots, n$; (c) is due to the property of mutual information [39] $I(X_t; Y) \geq 0$, and applied to $I(X_0; Y_0, \tilde{x}_{0|0}) \geq 0$ and $I(X_t; Y^t|S^{t-1}, \tilde{x}_{t|t}, \tilde{x}_{t-1|t-1}) \geq 0$, $t = 1, \ldots, n$. Hence, we have established the inequality (III.26).

(c) By (II.7), and the property of conditional mutual information [39], $I(X; Y|Z) = 0$ if and only if $P_{Y|X, Z} = P_{Y|Z} - a.s.$, then (II.7) holds with equality if and only if $P_{X_0; Y_0|\tilde{x}_{0|0}} = 0$ and $I(X_t; Y_t|\tilde{x}_{t|t}, \tilde{x}_{t-1|t-1}) = 0$, $t = 1, \ldots, n$.

Clearly, $\tilde{x}_{t|t} = Y_t - a.s., t = 0, \ldots, n$ is sufficient for (B.8), (B.9) to hold. Substituting $\tilde{x}_{t|t} = Y_t - a.s., t = 0, \ldots, n$ into the realization (III.14) then inequality (III.26) holds with equality, and in addition (III.27) and (III.28) are easily verified. By mean-square estimation theory,

$$
E^Q\left[\sum_{t=0}^{n} ||X_t - \tilde{x}_{t|t}||^2\right] \geq E^Q\left[\sum_{t=0}^{n} ||X_t - \tilde{x}_{t|t}||^2\right].
$$

(B.10)

Moreover, if $\tilde{x}_{t|t} = Y_t - a.s., t = 0, \ldots, n$ then, (III.27) and (III.28) hold, and the inequality (B.10) holds with equalities. Hence, if $\tilde{x}_{t|t} = Y_t - a.s., t = 1, \ldots, n$, then by (III.11), the reproduction process is given by (III.32), its conditional distribution is $Q_{Y|Y^{t-1}, X_t} = Q_{Y|Y^{t-1}, Y_t}$, for $t = 1, \ldots, n$, $Q_{Y_0|X_0} = Q_{Y_0|X_0}$, By (III.25) then $P_{Y_t|Y^{t-1}} = P_{Y_t|Y^{t-1}, Y_t} - a.s., t = 0, \ldots, n$. Thus, if $\tilde{x}_{t|t} = Y_t - a.s., t = 0, \ldots, n$ then we also obtain $I(X_0; \tilde{x}_{0|0}) = I(X_0; Y_0)$, from which all statements of part (c) follow.

**APPENDIX C**

**PROOF OF THEOREM 5**

(a) First, we verify that if there exist $(H_t, K_{t}) \in R^{p \times p} \times S^{p \times p}$, $t = 0, \ldots, n$ such that (III.35) hold, then the statements of Theorem 4.3(c) hold. By (III.35), $\tilde{x}_{t|t} = Y_t - a.s., t = 0, \ldots, n$. Moreover, by (III.11), letting $g_t(\tilde{x}_{t-1|t-1}) \triangleq (I - H_t)A_{t-1}\tilde{x}_{t-1|t-1}$, we have

$$
g_t(\tilde{x}_{t-1|t-1}) = g_t(Y_t-1)
$$

$$\triangleq I - H_t A_{t-1} Y_{t-1}, \quad t = 1, \ldots, n
$$

$$\Rightarrow Y_t = H_t A_{t-1} (X_{t-1} - Y_{t-1})
$$

$$+ A_{t-1} Y_{t-1} + H_t B_{t-1} W_t + V_t, \quad t = 1, \ldots, n,
$$

$$Y_0 = H_0 X_0 + V_0.
$$

(C.3)

By (C.2), (C.3) and the definition of $X^n$, follow all statements of Theorem 4.3(c), by simple calculations. Next we
proceed to identify $(H_t, K_{V_i}) \in \mathbb{R}^{p \times p} \times S^+_p, t = 0, \ldots, n$ such that (III.35) holds.

Consider the realization (III.11). Then following preliminary calculations are easily verified.

\[
\begin{align*}
\text{cov} \left( X_t, X_t \big| Y^{t-1} \right) &= \Sigma_t^r \\
&= A_{t-1} \Sigma_{t-1} A_{t-1}^t + B_{t-1} K_{W_t} B_{t-1}^t, \quad t = 1, \ldots, n, \quad (C.4) \\
\text{cov} \left( Y_t, Y_t \big| Y^{t-1} \right) &= A_{t-1} \Sigma_{t-1} (H_t A_{t-1})^t + B_{t-1} K_{W_t} B_{t-1}^t H_t^t, \quad t = 1, \ldots, n, \quad (C.5) \\
\text{cov} \left( Y_t, Y_{t'} \big| Y^{t-1} \right) &= H_t \Sigma_t^r H_t^t + K_{V_t}, \quad t = 1, \ldots, n. 
\end{align*}
\]

\[
\begin{align*}
\text{cov} \left( X_0, X_0 \right) &= \Sigma_0^r = K_{X_0}, \\
\text{cov} \left( X_0, Y_0 \right) &= K_{X_0} H_0^t, \\
\text{cov} \left( Y_0, Y_0 \right) &= H_0 \Sigma_0^r H_0^t + K_{V_0}, \\
\mathbb{E}^Q \left\{ X_t \big| Y^{t-1} \right\} &= \mathbb{E}^Q \left\{ Y_t \big| Y^{t-1} \right\} \\
= A_{t-1} \tilde{X}_{t-1|t-1}, \quad t = 0, \ldots, n \quad \text{i.e., (III.37) holds.} 
\end{align*}
\]

(C.11)

Special Case. First we consider the special case \( \text{cov} \left( Y_t, Y_t \big| Y^{t-1} \right) > 0 \), which implies (III.36). (III.37) are sufficient for \( \tilde{X}_{t|i} = Y_i - a.s., t = 0, \ldots, n \) to hold. First, we determine an equation for \( \{K_{V_t} : t = 0, \ldots, n\}. \)

If (III.36) holds then by (C.5), (C.6):

\[
K_{V_t} = (A_{t-1} \Sigma_{t-1} A_{t-1}^t + B_{t-1} K_{W_t} B_{t-1}^t) H_t^t - H_t \Sigma_t^r H_t^t, \\
\Rightarrow K_{V_t} = (I - H_t) \Sigma_t^r H_t^t \quad \text{provided } K_{V_t} = K_{V_t}^r. 
\]

At \( t = 0 \) then \( K_{V_0} = (I - H_0) \Sigma_0^r H_0^t. \)

Now, we determine \( H_t : t = 0, \ldots, n \).

If (III.36) holds then by (III.8):

\[
\Sigma_t = \Sigma_t^r - \left( \Sigma_t^r H_t^t \right)^t, \quad t = 0, \ldots, n. 
\]

\[
\Rightarrow H_t \Sigma_t^r = \Sigma_t - \Sigma_t^r \quad \text{and } H_t \Sigma_t = (H_t \Sigma_t^r)^t \\
\Rightarrow H_t = I - \Sigma_t (\Sigma_t^r)^{-1} \text{ by } \Sigma_t^r > 0. 
\]

(C.18)

Substituting (C.18) into (C.14) then

\[
K_{V_t} = \Sigma_t H_t^t = \Sigma_t - \Sigma_t (\Sigma_t^r)^{-1} \Sigma_t \quad \text{and } K_{V_t} = K_{V_t}^r. 
\]

At \( t = 0 \) we also have \( K_{V_0} = \Sigma_0 - \Sigma_0 (\Sigma_0^r)^{-1} \Sigma_0 \).

Since \( K_{V_t}, t = 0, \ldots, n \) is a covariance matrix, necessary \( \Sigma_t H_t^t = H_t \Sigma_t \geq 0. \) Hence, we have shown (III.38)-(III.42), and that \( \tilde{X}_{t|i} = Y_i - a.s., t = 0, \ldots, n. \)

The general case \( \text{cov} \left( Y_t, Y_t \big| Y^{t-1} \right) \geq 0 \) is done similarly, by using (III.35) and properties of pseudo-inverse. This completes the derivation of (a).

(b) This follows from part (a).

(c) Finally, we show (III.43) as follows. By the realization in part (a), then the following conditional independence conditions hold:

\[
P_{X_t|Y^t} = P_{Q_t|X_t}, \quad P_{Q_t|Y^t-1} = P_{Q_t|X_{t-1}}, \quad t = 0, \ldots, n. 
\]

(C.20)

Hence,

\[
I(X_0; Y_0) + \sum_{t=1}^n I(X_t; Y_t|Y^{t-1}) \\
= \mathbb{E}^{Q_t} \left\{ \log \left( \frac{P_{X_0|Y_0}}{P_{X_0}} \right) \right\} + \mathbb{E}^{Q_t} \left\{ \sum_{t=1}^n \log \left( \frac{P_{X_t|Y_t|Y^{t-1}}}{P_{X_t|Y^{t-1}}} \right) \right\} \\
= I(X_0; Y_0) + \sum_{t=1}^n I(X_t; Y_t|Y_{t-1}). 
\]

Thus, we have constructed a joint distribution \( P^Q(dx^n, dy^n) \) whose marginal is \( P(dx^n) = \lim_{n \to \infty} P(dx|X_{n-1}) \) that satisfies \( \frac{1}{n} \mathbb{E}^{Q_t} \left\{ \sum_{t=0}^n \|X_t - Y_t\|^2 \right\} \leq D, \) and achieves the lower bound

\[
I(X_0; Y_0) + \sum_{t=1}^n I(X_t; Y_t|Y_{t-1}) \\
\geq I(X_0; Y_0) + \sum_{t=1}^n I(X_t; Y_t|Y_{t-1}). 
\]

(C.21)

APPENDIX D

PROOF OF THEOREM 6

(a) Consider Theorem 5. Then the following hold.

\[
I(X_t; Y_t|Y_{t-1}) = \frac{1}{2} \log \left( \frac{\Sigma_t^r}{\Sigma_t} \right)^+ + \frac{1}{2} \log \left( \frac{\Sigma_t}{\Sigma_t - \Sigma_t^r} \right)^+ \quad (D.1) \\
= \frac{1}{2} \log \left( \frac{1}{(\Sigma_t^r)^{-1} (\Sigma_t - \Sigma_t^r + \Sigma_t)} \right)^+ \quad (D.2) \\
= \frac{1}{2} \log \left( \frac{1}{(I - (\Sigma_t)^{-1} (\Sigma_t^r - \Sigma_t)) \Sigma_t} \right)^+ \quad (D.3) \\
= \frac{1}{2} \log \left( \frac{1}{(I - (\Lambda_t)^{-1} U_t^t (\Sigma_t^r - \Sigma_t) U_t)} \right)^+ \quad (D.4) \\
= \frac{1}{2} \log \left( \frac{1}{(I - (\Lambda_t)^{-1} U_t^t (\Sigma_t^r - \Sigma_t) U_t) 1_{\tilde{X}_{t|i}}} \right)^+ \\
= \frac{1}{2} \log \left( \frac{1}{\Pi_{i=1}^n (1 - \lambda_{i,t}^{-1} U_t^t (\Sigma_t^r - \Sigma_t) U_t)} \right)^+, \quad (D.5) \\
\]

where (a) is due to the spectral representation \( \Sigma_t^r = U_t \Lambda_t U_t^t, \) \( \Lambda_t \) a diagonal matrix, with entries \( \lambda_{i,t}, i = 1, \ldots, p, U_t^t U_t = U_t^t U_t = I, \) (b) is due to properties of determinant, and (c) is due to Hadamard’s inequality, i.e., for any positive-semidefinite matrix \( A, \) with diagonal entries \( A_{ii}, \) then \( |A| \leq \prod_{i=1}^n |A_{ii}|, \) and equality holds if and only if \( A \) is diagonal.

(b) The lower bound in (D.6) is achieved if and only if the \( \Sigma_t - \Sigma_t^r \geq 0 \) has a spectral representation \( \Sigma_t - \Sigma_t^r = U_t D_t \Sigma_t - \Sigma_t U_t^t \), where \( D_t \Sigma_t - \Sigma_t \) is diagonal with non-negative entries, for \( t = 0, \ldots, n, \) that is, and if only if \( \Sigma_t - \Sigma_t^r \) and \( \Sigma_t \) have the same spectral representation with the same unitary matrix, or equivalently if and only if \( \Sigma_t (\Sigma_t - \Sigma_t) = \Sigma_t (\Sigma_t^r - \Sigma_t^r) \) (i.e., they commute [40]) for \( t = 0, \ldots, n. \)

(c) This follows from parts (a), (b) and simple calculations.
APPENDIX E

PROOF OF LEMMA 3

We apply the following results from linear algebra [40]. Given two matrices \( A \in \mathcal{S}^{k \times k} \), \( B \in \mathcal{S}^{k \times k} \), then the following statements are equivalent: (1) \( AB \) is normal, (2) \( AB \geq 0 \), where \( AB \) normal means \( (AB)(AB)^\dagger = (AB)^\dagger(AB) \). Note that \( AB \) is normal if and only if only if \( AB = BA \).

Let \( A = U_A D_A U_A^T, B = U_B D_B U_B^T, U_A U_A^T = I, U_B U_B^T = I \), i.e., there exists a spectral representation of \( A, B \) in terms of unitary matrices \( U_A, U_B \) and diagonal matrices \( D_A, D_B \). Then, \( AB \geq 0 \) if and only if the matrices \( A \) and \( B \) commute, i.e., \( AB = BA \), and \( A \) and \( B \) commute if and only if \( U_A = U_B \). Suppose (III.80) holds. Letting \( A_t = \Sigma_t^+, B_t = \Sigma_t \), then \( A_t = U_A D_A U_A^T, B_t = U_B D_B U_B^T, U_A U_A^T = I, U_B U_B^T = I \), \( U_A D_A U_A^T = I, U_A U_A^T = U_B \). Since \( \Sigma_t^{-1} = A_t^{-1} = U_A D_A^{-1} U_A^T \), then \( \Sigma_t (\Sigma_t^{-1}) = (\Sigma_t^{-1})^{-1} \Sigma_t \), i.e., they commute. Hence, \( \Sigma_t^{-1} \) and \( \Sigma_t \) commute.

From the definition of covariance matrix of innovations process \( \{v_t : t = 0, \ldots, n\} \), given by (III.41) then we must have \( K_{V_t} = \Sigma_t H_t^2 = K_t^\dagger = H_t \Sigma_t \), Hence, \( K_{V_t} = \Sigma_t H_t \). As a result, \( \Sigma_t, \Sigma_t^{-1}, H_t, K_{V_t} \) are all elements of \( \mathcal{S}^{n \times p}_+ \) having a spectral decomposition w.r.t. the same unitary matrix \( U_t U_t^T = I \), for \( t = 0, \ldots, n \).

APPENDIX F

PROOF OF LEMMA 4

(a) The fact that \( P_t \) satisfies the recursion follows by calculating the expectation of \( X_t X_t^T, t = 0, \ldots, n \), since the mean of \( X_t \) is zero for \( t = 0, \ldots, n \).

(b) By definition \( P_0 = K_0 \). Moreover,

\[
\Sigma_t = E^Q \left\{ (X_t - Y_t)(X_t - Y_t)^\dagger \right\} = E^Q \left\{ X_t X_t^T \right\} - E^Q \left\{ Y_t Y_t^T \right\} = P_t - E^Q \left\{ Y_t Y_t^T \right\}.
\]

Hence,

\[
P_t - \Sigma_t^{-1} = A_{t-1}(P_{t-1} - \Sigma_{t-1}) A_{t-1}^T, \quad t = 1, \ldots, n
\]

(c) The condition \( \Sigma_t^{-1} \geq \Sigma_t \) is necessary, otherwise

\[
I(X_t; Y_t | Y_{t-1}) = \frac{1}{2} \log \frac{\Sigma_t}{\Sigma_t^{-1}} \in [0, \infty] \text{ is not defined, and hence } R_{0,n}^I(D) \in [0, \infty] \text{ is also not defined. If } \Sigma_t \text{ is not full rank for some } t, \text{ then } R_{0,n}^I(D) = \infty.
\]

(d) If at some \( t \in \{1, 2, \ldots, n\} \), \( \Sigma_t = \Sigma_t^{-1} \) then \( H_t = 0 \), and \( Y_t = A_{t-1} Y_{t-1} \) which implies

\[
E\left[ \left\| X_t - Y_t \right\|^2 \right] = tr \left\{ \Sigma_t \right\} = tr \left\{ A_{t-1} \Sigma_{t-1} A_{t-1}^T + B_{t-1} K W_t B_{t-1}^T \right\} = D_{t, \max} \implies I(X_t; Y_t | Y_{t-1}) = 0.
\]
[17] T. Linder and G. Lugosi, “A zero-delay sequential scheme for lossy coding of individual sequences,” *IEEE Trans. Inf. Theory*, vol. 47, no. 6, pp. 2533–2538, Sep. 2001.

[18] E. Akyol, K. B. Viswanatha, K. Rose, and T. A. Ramstad, “On zero-delay source-channel coding,” *IEEE Trans. Inf. Theory*, vol. 60, no. 12, pp. 7473–7489, Dec. 2014.

[19] R. Zamir and M. Feder, “On universal quantization by randomized uniform/lattice quantizers,” *IEEE Trans. Inf. Theory*, vol. 38, no. 2, pp. 428–436, Mar. 1992.

[20] R. Zamir and M. Feder, “On lattice quantization noise,” *IEEE Trans. Inf. Theory*, vol. 47, no. 6, pp. 2533–2538, Sep. 2001.

[21] R. Zamir and M. Feder, “On lattice quantization noise,” *IEEE Trans. Inf. Theory*, vol. 47, no. 6, pp. 2533–2538, Sep. 2001.

[22] R. Zamir, “Rate-cost tradeoffs in control,” in Proc. 54th Annu. Allerton Conf. Commun., Control, Comput., Sep. 2016, pp. 1157–1164.

[23] C. D. Charalambous and P. A. Stavrou, “Optimization of directed information and relations to filtering theory,” in Proc. European Control Conf. (ECC), Jun. 2014, pp. 1385–1390.

[24] R. S. Bucy, “Distortion rate theory and filtering,” in *IEEE Trans. Inf. Theory*, vol. IT-28, no. 2, pp. 336–340, Mar. 1982.

[25] P. E. Caines, *Linear Stochastic Systems* (Wiley Series in Probability and Statistics). New York, NY, USA: Wiley, 1988.

[26] R. J. Elliott, L. Aggoun, and J. B. Moore, *Hidden Markov Models: Estimation and Control*. New York, NY, USA: Springer-Verlag, 1995.

[27] V. Kostina and B. Hassibi, “Rate-cost tradeoffs in control,” in Proc. 54th Annu. Allerton Conf. Commun., Control, Comput., Sep. 2016, pp. 1157–1164.

[28] 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 Syst. Lett.*, vol. 2, no. 1, pp. 175–180, Jan. 2018.

[29] P. A. Stavrou, T. Tanaka, and S. Tatikonda, “The time-invariant multidimensional Gaussian sequential rate-distortion problem revisited,” *IEEE Trans. Autom. Control*, vol. 65, no. 5, pp. 2245–2249, May 2020.

[30] T. Tanaka, K. H. Johansson, T. Oechtering, H. Sandberg, and M. Skoglund, “Rate of prefix-free codes in LQG control systems,” in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2016, pp. 2399–2403.

[31] C. D. Charalambous and P. A. Stavrou, “Directed information on abstract spaces: Properties and variational equalities,” *IEEE Trans. Inf. Theory*, vol. 62, no. 11, pp. 6019–6052, Nov. 2016.

[32] M. I. Gelfand and M. Yaglom, “Calculation of the amount of information about a random function contained in another such function,” *Ann. Math. Soc. Transl. Ser. 2*, pp. 199–246, 1959.

[33] C. D. Charalambous and J. H. van Schuppen, “Characterization of conditional independence and weak realizations of multivariate Gaussian random variables: Applications to networks,” 2020, arXiv:2001.06824.

[34] C. D. Charalambous and J. H. van Schuppen, “Characterization of conditional independence and weak realizations of multivariate Gaussian random variables: Applications to networks,” in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2020, pp. 2444–2449.

[35] C. D. Charalambous and J. H. van Schuppen, “Characterization of conditional independence and weak realizations of multivariate Gaussian random variables: Applications to networks,” in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2020, pp. 2444–2449.

[36] C. D. Charalambous and J. H. van Schuppen, “Characterization of conditional independence and weak realizations of multivariate Gaussian random variables: Applications to networks,” in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2020, pp. 2444–2449.

[37] C. E. Shannon, “Coding theorems for a discrete source with a fidelity criterion,” *IRE Int. Conv. Rec.*, vol. 4, no. 3, pp. 142–163, 1959.

[38] T. Linder and R. Zamir, “Causal coding of stationary sources and individual sequences with high resolution,” in *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 662–680, Feb. 2006.

[39] M. Pinsker, *Information and Information Stability of Random Variables and Processes*, A. Feinstein, Ed. San Francisco, CA, USA: Holden-Day, 1964.

[40] R. A. Horn and C. R. Johnson, *Matrix Analysis*, 2nd ed. New York, NY, USA: Cambridge Univ. Press, 2013.