

Asymptotic Reverse-Waterfilling Characterization of Nonanticipative Rate Distortion Function of Vector-Valued Gauss-Markov Processes with MSE Distortion

Photios A. Stavrou, Themistoklis Charalambous, Charalambos D. Charalambous, Sergey Loyka, and Mikael Skoglund

Abstract—We analyze the asymptotic nonanticipative rate distortion function (NRDF) of vector-valued Gauss-Markov processes subject to a mean-squared error (MSE) distortion function. We derive a parametric characterization in terms of a reverse-waterfilling algorithm, that requires the solution of a matrix Riccati algebraic equation (RAE). Further, we develop an algorithm reminiscent of the classical reverse-waterfilling algorithm that provides an upper bound to the optimal solution of the reverse-waterfilling optimization problem, and under certain cases, it operates at the NRDF. Moreover, using the characterization of the reverse-waterfilling algorithm, we derive the analytical solution of the NRDF, for a simple two-dimensional parallel Gauss-Markov process. The efficacy of our proposed algorithm is demonstrated via an example.

I. INTRODUCTION

Gorbunov and Pinsker in [1], [2] introduced a variant of the classical rate distortion function (RDF), called *nonanticipatory ϵ -entropy*. The authors envisioned that such information theoretic measure is directly related the design of real-time communication systems that can process information with minimum coding delays. In [1], the authors also characterized the nonanticipatory epsilon entropy for a time-varying and stationary scalar-valued Gauss-Markov process, subject to a pointwise or per-letter mean squared-error (MSE) distortion function, in terms of a reverse-waterfilling algorithm at each time instant. Tatikonda *et al.* in [3] revisited the nonanticipatory epsilon entropy, for time-invariant scalar and vector-valued Gauss-Markov processes subject to pointwise MSE distortion function, under the name “sequential RDF”, and identified connections between unstable eigenvalues of linear Gaussian control systems and the minimum rate requirements to stabilize such systems, when feedback is applied through a limited rate channel (memoryless). Derpich and Østergaard in [4] characterized variants of the nonanticipatory ϵ -entropy for stationary

scalar-valued Gaussian autoregressive models with pointwise MSE distortion function. Tanaka *et al.* in [5] revisited the sequential RDF of a vector-valued Gauss-Markov process subject to a pointwise MSE distortion function and applied semidefinite programming (under certain assumptions) to compute its optimal value numerically.

Recently, Stavrou *et al.* in [6] characterized the nonanticipatory ϵ -entropy under the name nonanticipative rate distortion function (NRDF), for a time-varying scalar-valued Gauss-Markov process subject to a total (or average in time) MSE distortion function and characterized it via the solution of a reverse-waterfilling problem, similar to the well-known reverse-waterfilling of classical RDF [7]. The extension of [6] to time-varying \mathbb{R}^P -valued Gauss-Markov processes is considered in [8].

A. Motivation and Contributions

Recently, the authors in [9] provided a counterexample, showing that the reverse-waterfilling algorithm the way it was suggested in [3, Equation (15)] to solve optimally the sequential RDF of a vector-valued Gauss-Markov process is incorrect. As a consequence, the parametric or analytical solution for this problem is still open. The only existing knowledge about the problem is that it is semidefinite representable and, thus, its solution can only be found numerically (see [5]). Although this is an important step, it lacks the insight of the parametric reverse-waterfilling solution that identifies the realization of the optimal distribution that achieves the NRDF. This paper analyzes the per unit time limit of finite-time horizon characterization of NRDF, in an effort to characterize the per unit time infinite horizon NRDF through a reverse-waterfilling algorithm. The main contributions of the paper are the following.

- 1) The characterization of the per unit time infinite horizon NRDF for a vector-valued Gauss-Markov process subject to a MSE distortion, via a reverse-waterfilling algorithm, expressed in terms of a matrix RAE.
- 2) An iterative scheme that computes the reverse-waterfilling optimization problem optimally at high rates, whereas at the entire rate-distortion region it computes an upper bound.
- 3) A closed form expression of the characterization of the per unit time infinite horizon NRDF for a two-dimensional parallel Gauss-Markov process subject to a MSE distortion.

P. A. Stavrou and M. Skoglund received funding by the KAW Foundation and the Swedish Foundation for Strategic Research.

P. A. Stavrou and M. Skoglund are with the Department of Information Science and Engineering, KTH Royal Institute of Technology, Sweden, *email*: {fstavrou, skoglund}@kth.se.

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

C. D. Charalambous is with the Department of Electrical and Computer Engineering, University of Cyprus, Cyprus, *email*: chadcha@ucy.ac.cy

S. Loyka is with the School of Electrical Engineering and Computer Science, University of Ottawa, Ontario, Canada, K1N 6N5., *email*: sergey.loyka@ottawa.ca.

II. PROBLEMS STATEMENT AND PRELIMINARY RESULTS

Notation: \mathbb{R} denotes the set of real numbers, \mathbb{N}_0 the set of nonnegative integers, and $\mathbb{N}_0^n \triangleq \{0, \dots, n\}$, $n \in \mathbb{N}_0$. Let \mathcal{X} be a complete separable metric space, and $\mathcal{B}_{\mathcal{X}}$ be the Borel σ -algebra on \mathcal{X} . $(\Omega, \mathcal{F}, \mathcal{P})$ denotes a probability space and $X : (\Omega, \mathcal{F}) \mapsto (\mathcal{X}, \mathcal{B}_{\mathcal{X}})$ a random variable (RV), $\mathbf{P}_X(dx) \equiv \mathbf{P}(dx)$ is the probability distribution induced by X on $(\mathcal{X}, \mathcal{B}_{\mathcal{X}})$. The conditional distribution of another RV Y given $X = x$ is denoted by $\mathbf{P}_{Y|X}(dy|X = x) \equiv \mathbf{P}(dy|x)$. $X^n = (X_0, \dots, X_n)$ denotes a sequence of RVs with convention $X^{-1} = (X_{-\infty}, \dots, X_{-1})$. Unless otherwise stated, the notation $\Sigma \succ 0$ (respectively, $\Sigma \succeq 0$) means symmetric positive-definite matrix (respectively, symmetric positive-semidefinite matrix). The statement $\Lambda \succeq \Sigma$ ($\Lambda \succ \Sigma$) means that $\Lambda - \Sigma$ is a positive semidefinite (definite) matrix. The determinant, diagonal and trace of a square matrix $\Sigma \in \mathbb{R}^{p \times p}$ are denoted by $|\Sigma|$, $\text{diag}(\mu_{\Sigma,i})$, and $\text{trace}(\Sigma)$, respectively, where $\mu_{\Sigma,i}$ is the i th eigenvalue of matrix Σ . I is the identity matrix.

In rate distortion theory it is desirable to reproduce sequences of symbols $X^n = x^n$ generated by a source, via their reproduction symbols $Y^n = y^n$, subject to a pre-specified fidelity. The distribution of the source is fixed and given by

$$\mathbf{P}(dx^n) \triangleq \prod_{t=0}^n \mathbf{P}(dx_t|x^{t-1}). \quad (1)$$

The convention is that at $t = 0$, $\mathbf{P}(dx_0|x^{-1}) = \mathbf{P}(dx_0)$. Following [1], the channel that is used to reproduce $Y^n = y^n$ from $X^n = x^n$ is described by the conditional distribution

$$\mathbf{P}(dy^n|x^n) \triangleq \prod_{t=0}^n \mathbf{P}(dy_t|y^{t-1}, x^t), \quad (2)$$

and it is subject to a design or is found by an optimization problem. Here the convention is $\mathbf{P}(dy_0|y^{-1}, x^0) = \mathbf{P}(dy_0|x_0)$. By (1) and (2), the joint distribution is $\mathbf{P}(dx^n, dy^n) \triangleq \mathbf{P}(dx^n) \otimes \mathbf{P}(dy^n|x^n)$ while $\mathbf{P}(dy_t|y^{t-1})$ is induced by the joint distribution $\mathbf{P}(dx^n, dy^n)$.

The NRDF of the source distribution (1) is defined through the mutual information between X^n and Y^n , i.e., $I(X^n; Y^n)$, subject to a distortion or fidelity of reproducing $X^n = x^n$ by $Y^n = y^n$ based on (2). Given (1) and (2), the mutual information is defined by

$$I(X^n; Y^n) = \sum_{t=0}^n \mathbf{E} \left\{ \log \left(\frac{\mathbf{P}(dY_t|Y^{t-1}, X^t)}{\mathbf{P}(dY_t|Y^{t-1})} \right) \right\}. \quad (3)$$

Next, we introduce the finite-time horizon NRDF and its per unit time limit, henceforth, called asymptotic NRDF of a time-invariant vector-valued Gauss-Markov process subject to a MSE distortion function. This definition was introduced in [1] and further analyzed in [3].

Definition 1: (Asymptotic Gaussian NRDF with MSE distortion) Let X_t be the time-invariant \mathbb{R}^p -valued Gauss-Markov process

$$X_{t+1} = AX_t + W_t, \quad t \in \mathbb{N}_0^n, \quad (4)$$

where $A \in \mathbb{R}^{p \times p}$ is a deterministic matrix, $X_0 \sim \mathcal{N}(0; \Sigma_{X_0})$ is the initial state with $\Sigma_{X_0} \succ 0$, and $W_t \in \mathbb{R}^p \sim \mathcal{N}(0; \Sigma_W)$,

$\Sigma_W \succ 0$, is a white Gaussian noise process independent of X_0 . The finite-time horizon NRDF is defined by

$$R_{0,n}^{\text{na}}(D) \triangleq \inf_{\mathbf{P}(dy^n|x^n)} \frac{1}{n+1} I(X^n; Y^n). \quad (5)$$

$$\text{s.t.} \quad \frac{1}{n+1} \sum_{t=0}^n \mathbb{E} \|X_t - Y_t\|^2 \leq D \quad (6)$$

assuming existence of a finite solution. The per unit time limit, called asymptotic (5) is defined by

$$R^{\text{na}}(D) \triangleq \lim_{n \rightarrow \infty} R_{0,n}^{\text{na}}(D), \quad (7)$$

provided that the limit exists and it is finite.

An *upper bound* on $R^{\text{na}}(D)$ is the following expression:

$$\widehat{R}^{\text{na}}(D) \triangleq \inf_{\mathbf{P}(dy^\infty|x^\infty)} \lim_{n \rightarrow \infty} \frac{1}{n+1} I(X^n; Y^n), \quad (8)$$

$$\text{s.t.} \quad \frac{1}{n+1} \sum_{t=0}^n \mathbb{E} \|X_t - Y_t\|^2 \leq D$$

provided the limit exists and it is finite, where $\mathbf{P}(dy^\infty|x^\infty)$ denotes the sequence of conditional probability distributions $\mathbf{P}(dy_t|y^{t-1}, x^t)$, $t \in \mathbb{N}_0$. It should be mentioned that for stationary Gauss-Markov process $R^{\text{na}}(D) = \widehat{R}^{\text{na}}(D)$ the above expressions coincide as shown in [1, Theorem 4].

It is well-known that the optimization problem of (7) is convex with respect to the set of test channels $\mathbf{P}(dy^n|x^n)$ that satisfy the average (over time) MSE distortion, for $D \in (0, D_{\text{max}}) \subseteq (0, \infty)$, and there exists an optimal solution characterizing it under general source distributions and distortion functions (e.g., [10]). By [8, Theorem 4.1], the optimal “test channel” corresponding to $R_{0,n}(D)$ is of the form

$$\mathbf{P}^*(dy_t|y^{t-1}, x^t) = \mathbf{P}^*(dy_t|y^{t-1}, x_t), \quad t \in \mathbb{N}_0, \quad (9)$$

that is non necessarily time-invariant, while the corresponding joint process $\{(X_t, Y_t) : t \in \mathbb{N}_0\}$ is not necessarily stationary. Further, the joint process $\{(X_t, Y_t) : t \in \mathbb{N}_0\}$ is jointly Gaussian [8, §5] (this is also shown in [3]).

It should be mentioned that in [1, Example 1] the authors considered a scalar-valued Gauss-Markov process subject to a pointwise MSE distortion, and derived the parametric expression of the optimal distribution that corresponds to $R_{0,n}^{\text{na}}(D)$ and, its corresponding closed form solution. In what follows, we generalize the parametric characterization of $R_{0,n}^{\text{na}}(D)$ given in [1, Example 1] to an \mathbb{R}^p -valued Gauss-Markov process, using a slightly different approach that utilizes Kalman-filtering as an intermediate step.

Lemma 1: [8, Lemma 5.2](Realization of optimal reproduction distribution) Consider Definition 1. Then, the following statements hold.

(a) Any candidate of the optimal reproduction distribution $\{\mathbf{P}(dy_t|y^{t-1}, x_t) : t \in \mathbb{N}_0^n\}$ is realized by the recursion

$$Y_t = H_t \left(X_t - \widehat{X}_{t|t-1} \right) + \widehat{X}_{t|t-1} + V_t, \quad t \in \mathbb{N}_0^n \quad (10)$$

where $\widehat{X}_{t|t-1} \triangleq \mathbf{E}\{X_t|Y^{t-1}\}$, $\widehat{X}_{0|-1} = \mathbf{E}\{X_0\}$, $\{V_t \in \mathbb{R}^p \sim N(0, \Sigma_{V_t}) : t \in \mathbb{N}_0^n\}$ is an independent Gaussian process independent of $\{W_t : t \in \mathbb{N}_0^{n-1}\}$ and X_0 , and $\{H_t \in$

$\mathbb{R}^{p \times p} : t \in \mathbb{N}_0^n$ are time-varying deterministic functions. That is, the distribution $\mathbf{P}(dy_t|y^{t-1}, x_t)$ is parametrized by $\{H_t, K_{V_t}\}$.

Moreover, the innovations process $\{\nu_t \in \mathbb{R}^p : t \in \mathbb{N}_0^n\}$ of (10) is the orthogonal process defined by

$$\begin{aligned} \nu_t &\triangleq Y_t - \mathbf{E}\{Y_t|Y^{t-1}\} \\ &= Y_t - \widehat{X}_{t|t-1} = H_t \left(X_t - \widehat{X}_{t|t-1} \right) + V_t, \end{aligned}$$

where $\nu_t \sim \mathcal{N}(0; \Sigma_{\nu_t})$, $\Sigma_{\nu_t} = H_t \Sigma_{t|t-1} H_t^\top + \Sigma_{V_t}$ and $\Sigma_{t|t-1} \triangleq \mathbf{E}\left\{ (X_t - \widehat{X}_{t|t-1})(X_t - \widehat{X}_{t|t-1})^\top | Y^{t-1} \right\}$.

(b) Let $\widehat{X}_{t|t} \triangleq \mathbf{E}\{X_t|Y^t\}$ and $\Sigma_{t|t} \triangleq \mathbf{E}\left\{ (X_t - \widehat{X}_{t|t})(X_t - \widehat{X}_{t|t})^\top | Y^t \right\}$. Then, $\{\widehat{X}_{t|t-1}, \Sigma_{t|t-1} : t \in \mathbb{N}_0^n\}$ satisfy the following \mathbb{R}^p -valued filtering recursions:

$$\widehat{X}_{t|t-1} = A\widehat{X}_{t-1|t-1}, \quad (11a)$$

$$\Sigma_{t|t-1} = A\Sigma_{t-1|t-1}A^\top + \Sigma_W, \quad \Sigma_{0|-1} = \Sigma_{X_0}, \quad (11b)$$

$$\widehat{X}_{t|t} = \widehat{X}_{t|t-1} + K_t \nu_t, \quad (11c)$$

$$K_t = \Sigma_{t|t-1} H_t^\top \Sigma_{\nu_t}^{-1} \text{ (Kalman Gain)}, \quad (11d)$$

$$\Sigma_{t|t} = \Sigma_{t|t-1} - \Sigma_{t|t-1} H_t^\top \Sigma_{\nu_t}^{-1} H_t \Sigma_{t|t-1}, \quad (11e)$$

where $\Sigma_{t|t} = \Sigma_{t|t}^\top \succeq 0$ and $\Sigma_{t|t-1} = \Sigma_{t|t-1}^\top \succ 0$.

(c) $R_{0,n}^{\text{na}}(D)$ is given by

$$\begin{aligned} R_{0,n}^{\text{na}}(D) &= \inf_{H_t \succeq 0, \Sigma_{V_t} \succeq 0, t \in \mathbb{N}_0^n} \frac{1}{2} \frac{1}{n+1} \sum_{t=0}^n \log \frac{|\Sigma_{t|t-1}|}{|\Sigma_{t|t}|}, \\ \text{s.t.} \quad &H_t \text{ is not necessarily symmetric} \\ \frac{1}{n+1} \sum_{t=0}^n \text{trace} \left((I - H_t) \Sigma_{t|t-1} (I - H_t)^\top + \Sigma_{V_t} \right) &\leq D \end{aligned} \quad (12)$$

for some $D \in [0, \infty)$.

The next theorem is the analogue of [1, Theorem 5] but for total MSE distortion function.

Theorem 1: [8, Theorem 5.3] (Alternative characterization of (12)) The following alternative characterization of Definition 1, (5) holds.

(a) The optimal reproduction distribution is realized by

$$Y_t = H_t X_t + (I - H_t) A Y_{t-1} + V_t, \quad t \in \mathbb{N}_0^n, \quad (13)$$

where

$$H_t \triangleq I - \Sigma_{t|t} \Sigma_{t|t-1}^{-1} \succeq 0, \quad \Sigma_{t|t} \succeq 0, \quad \Sigma_{t|t-1} \succ 0, \quad (14a)$$

$$\Sigma_{V_t} \triangleq \Sigma_{t|t} H_t^\top \succeq 0, \quad (14b)$$

$$\Sigma_{t|t-1} = A \Sigma_{t-1|t-1} A^\top + \Sigma_W, \quad \Sigma_{0|-1} = \Sigma_{X_0}. \quad (14c)$$

(b) Moreover, the above realization yields

$$\widehat{X}_{t|t} = Y_t, \quad \widehat{X}_{t|t-1} = A Y_{t-1}, \quad (15)$$

$$\mathbf{P}(dy_t|y^{t-1}, x_t) = \mathbf{P}(dy_t|y_{t-1}, x_t), \quad (16)$$

(c) The characterization of $R_{0,n}^{\text{na}}(D)$ is

$$R_{0,n}^{\text{na}}(D) = \inf_{\Sigma_{t|t} \succeq 0, t \in \mathbb{N}_0^n} \frac{1}{2} \frac{1}{n+1} \sum_{t=0}^n \log \frac{|\Sigma_{t|t-1}|}{|\Sigma_{t|t}|}, \quad (17a)$$

$$\text{s.t. } \Sigma_{t|t-1} \text{ satisfies (11b), } \Sigma_{t|t-1} \succ 0 \quad (17b)$$

$$0 \preceq \Sigma_{t|t} \preceq \Sigma_{t|t-1} \quad (17c)$$

$$\frac{1}{n+1} \sum_{t=0}^n \text{trace}(\Sigma_{t|t}) \leq D \quad (17d)$$

for some $D \in [0, \infty)$.

Next, we discuss about some of the implications of Theorem 1.

Remark 1: (1) Theorem 1, (a) provides the complete realization of the Gaussian NRDF of \mathbb{R}^p -valued Gauss-Markov processes. This result complements the realization obtained for the Gaussian NRDF in [1, Theorem 5] in the sense that the structure of matrices (H_t, Σ_{V_t}) is also provided. (2) The constraint (17c) ensures a finite rate (existence of solution) in (17a) if $\Sigma_{t|t} \succ 0$. Hence, a property of the solution is that $0 \prec \Sigma_{t|t}$, otherwise the rate is infinite.

III. MAIN RESULTS

In this section, we study the asymptotic limit of the corresponding finite time horizon optimization problem of Theorem 1. Before we proceed, we need the following theorem.

Theorem 2: Consider the sequence of NRDF of Definition 1. Then the following hold.

(a) The sequence $\{R_{0,n}^{\text{na}}(D) : n = 0, 1, \dots\}$ is sub-additive, and the limit exists, i.e.,

$$\lim_{n \rightarrow \infty} R_{0,n}^{\text{na}}(D) = \inf_n R_{0,n}^{\text{na}}(D) \quad (18)$$

and further it is finite.

(b) The limits

$$\begin{aligned} \lim_{t \rightarrow \infty} \Sigma_{t|t} &= \Delta, \quad \lim_{t \rightarrow \infty} \Sigma_{t|t-1} = \Lambda, \quad \lim_{t \rightarrow \infty} H_t = H, \\ \lim_{t \rightarrow \infty} \Sigma_{V_t} &= \Sigma_V \end{aligned} \quad (19)$$

exist and (Λ, H, Σ_V) satisfy the steady state versions of the equations Theorem 1.

Proof: (a) Sub-additivity is well-known, and shown under various conditions (see, e.g., [2, Lemma 1]). The fact that the limit is finite follows from existence of the minimizing distribution from [10, Theorem 15], since all conditions hold for the source and distortion function of this paper, then the limit is finite $\Sigma_{t|t} \succ 0$. (b) Consider the characterization of Theorem 1, and for $n \in \mathbb{N}_0$ let $\mathbf{P}^{(n)}(dy_t|y_{t-1}, x_t)$ denote the sequence of Gaussian distributions generated from the sequence $\Sigma_{t|t}^{(n)}, t \in \mathbb{N}_0^n$. Then the sequence of Gaussian distributions is tight, hence relatively compact. Using this, we can further show the limiting distribution is the one parametrized the limits (19), i.e., it is closed. ■

First, we give the asymptotic characterization of $R^{\text{na}}(D)$ of \mathbb{R}^p -valued Gauss-Markov process with MSE distortion.

Lemma 2: (1) The asymptotic optimal reproduction distribution is realized by

$$Y_t = HX_t + (I - H)AY_{t-1} + V_t^c, \quad (20)$$

where

$$H \triangleq I - \Delta\Lambda^{-1} \succeq 0, \quad \Sigma_V \triangleq \Delta H^T \succeq 0, \quad (21)$$

$$\Lambda = A\Delta A^T + \Sigma_W, \quad \Delta \succeq 0, \quad \Lambda \succ 0. \quad (22)$$

(2) For $D > 0$, $R^{\text{na}}(D)$ is given by

$$R^{\text{na}}(D) = \inf_{\Delta \succ 0: \text{trace}(\Delta) \leq D} \frac{1}{2} \log \frac{|\Lambda|}{|\Delta|}, \quad (23a)$$

$$\text{s. t. } 0 \prec \Delta \preceq \Lambda \quad (23b)$$

$$\Lambda = A\Delta A^T + \Sigma_W \quad (23c)$$

Proof: This follows from Theorem 2. ■

Next, we provide the solution to the asymptotic characterization of $R^{\text{na}}(D)$ of \mathbb{R}^p -valued Gauss-Markov process with MSE distortion.

Theorem 3: The parametric solution subject to a reverse-waterfilling algorithm that corresponds to (23) is the following.

$$R^{\text{na}}(D) = \frac{1}{2} \log \frac{|\Lambda|}{|\Delta|}, \quad (24)$$

such that

$$\Delta = \begin{cases} \Delta_1^*, & \text{if } \Delta \prec \Lambda, \\ \Delta_2^*, & \text{if } \Delta \preceq \Lambda \text{ and } \Delta \not\prec \Lambda, \\ \Lambda, & \text{otherwise} \end{cases} \quad (25)$$

where $\Delta_1^* \succ 0$ is the solution of the Riccati equation

$$\left(-\frac{I}{2}\right) \Delta_1^* + \Delta_1^* \left(-\frac{I}{2}\right) - \Delta_1^* B \Delta_1^* + \frac{1}{2\theta} I = 0, \quad (26)$$

$$B \triangleq A^T \Sigma_W^{-1} A,$$

and $\Delta_2^* \succ 0$ is the solution of the Riccati equation

$$\left(-\frac{I}{2}\right) \Delta_2^* + \Delta_2^* \left(-\frac{I}{2}\right) - \Delta_2^* B \Delta_2^* + \Upsilon^{-1} = 0, \quad (27)$$

$$\Upsilon \triangleq 2(\theta I + F_2 - A^T F_2 A), \quad \Upsilon = \Upsilon^T \succ 0, \quad (28)$$

with the Lagrangian variables $\theta > 0$ and $F_2 = F_2^T \succeq 0$ chosen such that

$$\text{trace}(\Delta) = D, \quad F_2(\Delta - \Lambda) = 0. \quad (29)$$

Proof: The proof follows similar steps to the derivation of [8, Appendix A] with a few necessary changes. The augmented Lagrange functional can be formulated as follows:

$$\begin{aligned} \mathcal{L}(\Delta, \Lambda, \theta, F_1, F_2) = & \frac{1}{2} [\log |\Sigma_W| + \log |B\Delta + I| - \log |\Delta|] \\ & + \theta (\text{trace}(\Delta) - D) - \text{trace}(F_1 \Delta_t) + \text{trace}(F_2 \Delta) \\ & - \text{trace}(F_2 (A\Delta A^T + \Sigma_W)), \end{aligned} \quad (30)$$

where $\theta \in [0, \infty)$ is a Lagrange multiplier for the distortion constraint $\text{trace}(\Delta) \leq D$, and $F_j \succeq 0$, $j = 1, 2$ are the Lagrange multiplier symmetric matrices responsible for

$\Delta \succ 0, \Lambda \succeq \Delta$. By KKT conditions, we differentiate (30) to obtain

$$\frac{1}{2}(I + B\Delta^*)^{-1}B - \frac{1}{2}\Delta^{*, -1} - F_1 + F_2 + \theta I - A^T F_2 A = 0. \quad (31)$$

The analysis of the individual cases as these arise from the KKT conditions gives (25) with (26)-(29). Note that if any other case occurs, the problem is trivial since the rate is always zero. ■

Remark 2: (1) The reverse-waterfilling algorithm of Theorem 3 is based on the Riccati equations (26), (27). (26) allocates the distortion when there is no reverse-waterfilling in dimension, i.e., when $\Delta \prec \Lambda$. On the contrary, (27) takes care the distortion allocation at each dimension when some dimensions are inactive. This corresponds to $\Delta \preceq \Lambda$. Unfortunately, finding F_2 in (27) is very hard and for this reason in the sequel we propose a suboptimal numerical scheme to solve the reverse-waterfilling optimization problem of Theorem 3. (2) Equation (27) (and its special case (26)) are cast as continuous-time algebraic Riccati equations (CARE). Also, for (27) observe that the pair $(-\frac{I}{2}, A)$ is always stabilizable and the pair $(-\frac{I}{2}, \Upsilon^{-1})$ is always detectable. This means that the solution of (27) (or (26)) is unique (for details on sufficient conditions of unique positive definite solutions of CARE see, for instance, [11]).

Remark 3: (Simultaneous diagonalization of Δ, Λ) The symmetric positive definite matrices (Δ, Λ) can be jointly diagonalized using a version of the cogredient diagonalization of [12, Theorem 8.3.1]. Then, there exists a single non-singular matrix $S \in \mathbb{R}^{p \times p}$ such that $S\Delta S^T = \text{diag}(\mu_{\Delta,i})$, $S\Lambda S^T = \text{diag}(\mu_{\Lambda,i})$.

The next proposition leverages Remark 3 to arrive at a good approximation to the reverse-waterfilling algorithm of Theorem 3.

Proposition 1: (Upper bound to the reverse-waterfilling solution of Theorem 3) An upper bound to the solution of (24), (25) is given by

$$\Delta \triangleq \min \{\Delta^\dagger, \Lambda\}, \quad (32)$$

where $\min \{\Delta^\dagger, \Lambda\} \triangleq S^{-1} \text{diag}(\min \{\mu_{\Delta^\dagger,i}, \mu_{\Lambda,i}\}) (S^T)^{-1}$ (for details, see [13, Section II.A]), and $\Delta^\dagger \succ 0$, is the unique solution of

$$\left(-\frac{I}{2}\right) \Delta^\dagger + \Delta^\dagger \left(-\frac{I}{2}\right) - \Delta^\dagger B \Delta^\dagger + \frac{1}{2\theta} I = 0, \quad (33)$$

Λ is given by (23c) using Δ , and with $\theta \in (0, \infty)$ chosen to satisfy $\text{trace}(\Delta) = D$. This upper bound gives the optimal solution for the case where $\Delta \prec \Lambda$.

Proof: If $\Delta \prec \Lambda$, then, (33) corresponds to the optimal solution of (26). However, if $\Delta \preceq \Lambda$ the solution is not necessarily the optimal one, hence it serves as an upper bound. Note that by choosing Δ as in (32), we ensure that $\Delta \preceq \Lambda$. ■

We wish to remark that the extension of Theorem 3 and Proposition 1 to the non-asymptotic regime appears in [8].

Based on the solution provided in Proposition 1, we propose Algorithm 1, based on bisection method, to solve Theorem 3 numerically to a predefined approximation.

Algorithm 1 \mathbb{R}^p -valued reverse-waterfilling algorithm of Proposition 1

Initialize: distortion level D ; error tolerance ϵ ; nominal minimum and maximum value $\theta^{\min} = 0$ and $\theta^{\max} = \frac{p}{2D}$ (where p denotes the number of dimensions); initial variance $\Lambda = \Sigma_{X_0}$ of the initial state X_0 , values of A and Σ_W of (4).

Set $\theta = 1/2D$; flag = 0.

while flag = 0 **do**

 Compute Δ as follows:

 Compute Δ^\dagger according to (33).

 Compute Δ according to (32).

 Compute Λ according to (23c).

if $\theta^{\max} - \theta^{\min} \geq \epsilon$ **then**

if $\text{trace}(\Delta) - D \geq \epsilon$ **then**

 Set $\theta^{\min} = \theta$.

else

 Set $\theta^{\max} = \theta$.

end if

 Compute $\theta = \frac{\theta^{\min} + \theta^{\max}}{2}$.

else

 flag \leftarrow 1

end if

end while

Output: Δ , Λ , for a distortion level D .

It should be noted that although Algorithm 1 is suboptimal in general, it is much faster compared to the existing optimal numerical approach via semidefinite programming (SDP) (see [5]). To illustrate this point, in Table I, we provide an example where we compare the mean and the standard deviation of the computational time needed for each of the two numerical methods to execute over a sample of 1000 instances. Algorithm 1 is using an error tolerance $\epsilon = 10^{-9}$.

Solver	Mean	Standard Deviation
SDP [5]	0.697	0.0498
Algorithm 1 ($\epsilon = 10^{-9}$)	0.031	0.0057

TABLE I: Comparison of the computational time needed between SDP and Algorithm 1 for 1000 instances.

Our results demonstrate that, despite its sub-optimality, Algorithm 1 is much faster (more than 20 times) and the fact that we can allow for different levels of tolerance (and thus make it faster or slower accordingly), makes it more preferable to be implemented in modern delay-constrained and computationally-limited devices than the optimal but much slower SDP.

IV. EXAMPLES

In this section, we give two examples to demonstrate the validity of our framework. The results are compared to the

optimal numerical solution of [5, Equation (27)] obtained via semidefinite programming.

Example 1: Consider a \mathbb{R}^4 -valued time-invariant Gauss-Markov process with parameters

$$A = \begin{bmatrix} 0.5508 & 0.8929 & 0.0515 & 0.6491 \\ 0.7081 & 0.8963 & 0.4408 & 0.2785 \\ 0.2909 & 0.1256 & 0.0299 & 0.6763 \\ 0.5108 & 0.2072 & 0.4568 & 0.5909 \end{bmatrix},$$

$$\Sigma_W = \begin{bmatrix} 0.0240 & 0 & 0 & 0 \\ 0 & 0.6931 & 0 & 0 \\ 0 & 0 & 0.3064 & 0 \\ 0 & 0 & 0 & 0.6724 \end{bmatrix}.$$

To compute $R^{\text{na}}(D)$, we run Algorithm 1 for error tolerance $\epsilon = 10^{-9}$ and initial $\theta = \frac{1}{2D}$, for distortion levels D ranging in the interval $[0.1, 1.8]$ with a step size of 0.1. Then, we use the same parameters (A, Σ_W) and the same distortion levels D in the SDP algorithm of [5, Equation (27)]. In Fig. IV.1 we illustrate a comparison between Algorithm 1 with the SDP algorithm. At high rates, where the condition $\Delta \prec \Lambda$ is always satisfied, Algorithm 1 performs optimally. On the other hand, at low rates the algorithm is suboptimal because the reverse-waterfilling kicks in and (33) is not the optimal equation to solve this case as we know from (27).

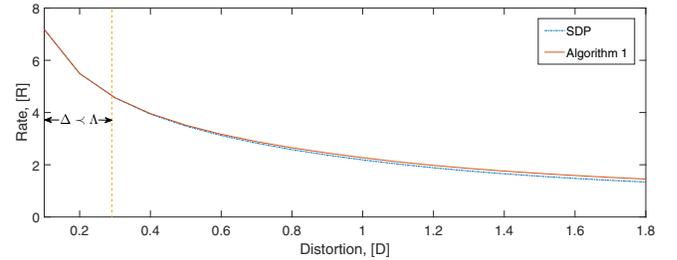


Fig. IV.1: Comparison of algorithm 1 to the optimal numerical solution obtained via SDP algorithm.

Example 2: We consider a \mathbb{R}^2 -valued time-invariant Gauss-Markov process with parameters

$$(A, \Sigma_W) = \left(\begin{bmatrix} \alpha & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} \sigma_{W_1}^2 & 0 \\ 0 & \sigma_{W_2}^2 \end{bmatrix} \right),$$

where $a \in \mathbb{R} \setminus \{0\}$, $\sigma_{W_i} > 0$, $i = 1, 2$. In this example, we use Theorem 3, to derive a closed form solution for $R^{\text{na}}(D)$ for this specific class of Gaussian sources. Following Theorem 3, we observe that in (25) we need to solve two cases, i.e., when $\Delta \prec \Lambda$ that corresponds to the Riccati equation (26), and, when $\Delta \preceq \Lambda$ that corresponds to the Riccati equation (27). We refer to the former as *full-rank solution* because $\mu_{\Delta-\Lambda, i} < 0$, for $i = 1, 2$ and to the latter as *rank-deficient solution* because $\mu_{\Delta-\Lambda, i} \leq 0$, for $i = 1, 2$. We note that the last case in (25) requires zero rate, hence for this case the problem is trivial.

Full-rank Solution ($\Delta \prec \Lambda$): Upon solving (26) we obtain

$$\mu_{\Delta, 1} = -\frac{1}{2\beta^2} \pm \frac{1}{2\beta^2} \sqrt{1 + \frac{2\beta^2}{\theta}}, \quad \mu_{\Delta, 2} = \frac{1}{\theta}.$$

Note that one of the solutions of $\mu_{\Delta,1}$ is rejected because it is negative hence the resulting matrix Δ gives eigenvalues:

$$\mu_{\Delta,1} = \frac{1}{2\beta^2} \left(\sqrt{1 + \frac{2\beta^2}{\theta}} - 1 \right), \quad \mu_{\Delta,2} = \frac{1}{2\theta}. \quad (34)$$

From (23c) we obtain Λ with eigenvalues:

$$\mu_{\Lambda,1} = \alpha^2 \left(\frac{1}{2\beta^2} \left(\sqrt{1 + \frac{2\beta^2}{\theta}} - 1 \right) \right) + \sigma_{W_1}^2, \quad \mu_{\Lambda,2} = \sigma_{W_2}^2. \quad (35)$$

Now, we use the left hand side (LHS) equation in (29) to obtain an additional equation which is used to find θ . By substituting (34) in (29) we obtain

$$\frac{1}{2\beta^2} \left(\sqrt{1 + \frac{2\beta^2}{\theta}} - 1 \right) + \frac{1}{2\theta} = D. \quad (36)$$

The solution of (36) gives

$$\theta = \begin{cases} \frac{\beta^2}{2(1+\beta^2 D) \left(1 + \sqrt{\frac{1}{1+\beta^2 D}} \right)}, & \text{(rejected),} \\ \frac{\beta^2}{2(1+\beta^2 D) \left(1 - \sqrt{\frac{1}{1+\beta^2 D}} \right)}, & \end{cases} \quad (37)$$

Both solutions in (37) are positive. However, from the Lagrange duality theorem [14] we know that the chosen θ is the one that results in greater rates. In this case, the second solution gives greater rates, whereas the first solution gives lower rates hence the first solution is rejected.

Next, we compute the individual rates and the total sum rate over both dimensions.

Clearly, $\mu_{\Delta,1} \triangleq \frac{1}{2\beta^2} \left(\sqrt{1 + \frac{2\beta^2}{\theta}} - 1 \right)$, $\mu_{\Delta,2} \triangleq \frac{1}{2\theta}$, $\mu_{\Lambda,1} \triangleq \alpha^2 \left(\frac{1}{2\beta^2} \left(\sqrt{1 + \frac{2\beta^2}{\theta}} - 1 \right) \right) + \sigma_{W_1}^2$, $\mu_{\Lambda,2} \triangleq \sigma_{W_2}^2$. Then, from (24) we obtain

$$\begin{aligned} R^{\text{na}}(D) &= \frac{1}{2} \log \frac{|\Lambda|}{|\Delta|} = \frac{1}{2} \sum_{i=1}^2 \log \left(\frac{\mu_{\Lambda,i}}{\mu_{\Delta,i}} \right) \\ &= \frac{1}{2} \left\{ \log \left(\alpha^2 + \frac{\sigma_{W_1}^2}{\mu_{\Delta,1}} \right) + \log \left(\frac{\sigma_{W_2}^2}{\mu_{\Delta,2}} \right) \right\} \end{aligned} \quad (38)$$

$$\begin{aligned} &\stackrel{(a)}{=} \frac{1}{2} \left\{ \log \left(\alpha^2 + \frac{2\alpha^2}{\sqrt{1 + 4(1 + \beta^2 D) \left(1 - \sqrt{\frac{1}{1 + \beta^2 D}} \right)} - 1} \right) \right. \\ &\quad \left. + \log \left(\frac{\sigma_{W_2}^2 \beta^2}{(1 + \beta^2 D) \left(1 - \sqrt{\frac{1}{1 + \beta^2 D}} \right)} \right) \right\}, \end{aligned} \quad (39)$$

where (a) follows if we substitute $\mu_{\Delta,1}$, $\mu_{\Delta,2}$, and θ in (38). Recall that (39) holds if and only if $\mu_{\Delta,i} < \mu_{\Lambda,i}$, $i = 1, 2$, which means that

$$\mu_{\Delta,1}^{\max} \equiv \mu_{\Lambda,1} = \begin{cases} \alpha^2 \mu_{\Delta,1} + \sigma_{W_1}^2, & \text{if } |\alpha| \geq 1, \\ \frac{\sigma_{W_1}^2}{1 - \alpha^2}, & \text{if } -1 < \alpha < 1, \end{cases}, \quad (40)$$

$$\mu_{\Delta,2}^{\max} \equiv \mu_{\Lambda,2} = \sigma_{W_2}^2. \quad (41)$$

Rank-Deficient Solution ($\Delta \preceq \Lambda$): this case corresponds to solving (27) such that the right hand side (RHS) equation

of (29) is satisfied. First, observe that because we consider a \mathbb{R}^2 -valued parallel Gaussian source, then from (29) we can show that F^2 , Δ , Λ commute hence, they are jointly diagonalizable matrices. This implies the study of the following two cases:

Case 1: $\mu_{\Delta-\Lambda,1} > 0$ and $\mu_{\Delta-\Lambda,2} = 0$ which in turn means that $\mu_{\Delta,1} < \mu_{\Lambda,1}$ and $\mu_{\Delta,2} = \mu_{\Lambda,2}$. Since $\mu_{\Delta,2} = \mu_{\Lambda,2}$ then, from the RHS equation of (29) we require $\mu_{F_2,2} > 0$ whereas $\mu_{F_2,1} = 0$.

Case 2: $\mu_{\Delta-\Lambda,1} = 0$ and $\mu_{\Delta-\Lambda,2} > 0$ which in turn means that $\mu_{\Delta,1} = \mu_{\Lambda,1}$ and $\mu_{\Delta,2} < \mu_{\Lambda,2}$. Since $\mu_{\Delta-\Lambda,1} = 0$ we require $\mu_{F_2,1} > 0$ whereas $\mu_{F_2,2} = 0$.

Case 1: Upon solving (28) we obtain:

$$\Upsilon = \begin{bmatrix} \theta & 0 \\ 0 & \mu_{F_2,2} + \theta \end{bmatrix}. \quad (42)$$

Moreover, by solving (27) we obtain:

$$\mu_{\Delta,1} = -\frac{1}{2\beta^2} \pm \frac{1}{2\beta^2} \sqrt{1 + \frac{2\beta^2}{\theta}}, \quad \mu_{\Delta,2} = \frac{1}{2(\mu_{F_2,2} + \theta)}.$$

Again, one of the solutions of $\mu_{\Delta,1}$ is rejected because it is negative hence the resulting matrix Δ has eigenvalues:

$$\mu_{\Delta,1} = \frac{1}{2\beta^2} \left(\sqrt{1 + \frac{2\beta^2}{\theta}} - 1 \right), \quad \mu_{\Delta,2} = \frac{1}{2(\mu_{F_2,2} + \theta)},$$

Additionally, using (23c) we obtain the following eigenvalues for Λ :

$$\mu_{\Lambda,1} = \frac{\alpha^2}{2\beta^2} \left(\sqrt{1 + \frac{2\beta^2}{\theta}} - 1 \right) + \sigma_{W_1}^2, \quad \mu_{\Lambda,2} = \sigma_{W_2}^2.$$

Since $\mu_{\Delta,2} = \mu_{\Lambda,2}$, then, $\frac{1}{2(\mu_{F_2,2} + \theta)} = \sigma_{W_2}^2$, that is, $\mu_{\Delta,2} = \mu_{\Lambda,2} = \sigma_{W_2}^2$. Hence, the LHS term of (29) gives

$$\frac{1}{2\beta^2} \left(\sqrt{1 + \frac{2\beta^2}{\theta}} - 1 \right) = D - \sigma_{W_2}^2. \quad (43)$$

The solution of (43) gives

$$\theta = \frac{1}{2(D - \sigma_{W_2}^2)(\beta^2(D - \sigma_{W_2}^2) + 1)}. \quad (44)$$

From (44) we obtain for Δ the eigenvalues $\mu_{\Delta,1} = \frac{1}{2\beta^2} \left(\sqrt{1 + 4\beta^2(D - \sigma_{W_2}^2)(\beta^2(D - \sigma_{W_2}^2) + 1)} - 1 \right)$, $\mu_{\Delta,2} = \sigma_{W_2}^2$, and for Λ the eigenvalues $\mu_{\Lambda,1} = \frac{\alpha^2}{2\beta^2} \left(\sqrt{1 + 4\beta^2(D - \sigma_{W_2}^2)(\beta^2(D - \sigma_{W_2}^2) + 1)} - 1 \right) + \sigma_{W_1}^2$, $\mu_{\Lambda,2} = \sigma_{W_2}^2$. Hence, for case 1 the sum rate is obtained as follows:

$$\begin{aligned} R^{\text{na}}(D) &= \frac{1}{2} \log \frac{|\Lambda|}{|\Delta|} = \frac{1}{2} \sum_{i=1}^2 \log \left(\frac{\mu_{\Lambda,i}}{\mu_{\Delta,i}} \right), \\ &= \frac{1}{2} \log \left(\alpha^2 \right. \\ &\quad \left. + \frac{2\alpha^2}{\sqrt{1 + 4\beta^2(D - \sigma_{W_2}^2)(\beta^2(D - \sigma_{W_2}^2) + 1)} - 1} \right). \end{aligned} \quad (45)$$

Recall that (45) holds if and only if $\mu_{\Delta,1} < \mu_{\Lambda,1}$ where $\mu_{\Lambda,1}$ is given from (40).

Case 2: Upon solving (28) we obtain:

$$\Upsilon = \begin{bmatrix} \mu_{F_2,1}(1 - \alpha^2) + \theta & 0 \\ 0 & \theta \end{bmatrix}.$$

Moreover, by solving (27), we obtain for matrix Δ :

$$\begin{aligned} \mu_{\Delta,1} &= -\frac{1}{2\beta^2} \pm \frac{1}{2\beta^2} \sqrt{1 + \frac{2\beta^2}{\mu_{F_2,1}(1 - \alpha^2) + \theta}}, \\ \mu_{\Delta,2} &= \frac{1}{2\theta}. \end{aligned} \quad (46)$$

Similar as before, one of the solutions of $\mu_{\Delta,1}$ is rejected because it is negative, hence giving $\mu_{\Delta,1} = \frac{1}{2\beta^2} \left(\sqrt{1 + \frac{2\beta^2}{\mu_{F_2,1}(1 - \alpha^2) + \theta}} - 1 \right)$.

Also, from (23c) we obtain:

$$\Lambda = \begin{bmatrix} \frac{\alpha^2}{2\beta^2} \left(\sqrt{1 + \frac{2\beta^2}{\mu_{F_2,1}(1 - \alpha^2) + \theta}} - 1 \right) + \sigma_{W_1}^2 & 0 \\ 0 & \sigma_{W_2}^2 \end{bmatrix}. \quad (47)$$

Similar to case 1, using the fact that $\mu_{\Delta,1} = \mu_{\Lambda,1}$, then, $\mu_{\Delta,1} = \frac{\sigma_{W_1}^2}{1 - \alpha^2}$. Hence, from the LHS of (29) we obtain

$$\mu_{\Delta,1} + \frac{1}{2\theta} = D \Rightarrow \theta = \frac{1}{2 \left(D - \frac{\sigma_{W_1}^2}{1 - \alpha^2} \right)},$$

which means that from (46) we obtain $\mu_{\Delta,1} = D - \frac{\sigma_{W_1}^2}{1 - \alpha^2}$. We also have from (47) that $\mu_{\Delta,2} = \sigma_{W_2}^2$. Hence, the sum rate for case 2 is obtained as follows:

$$R^{\text{na}}(D) = \frac{1}{2} \log \frac{|\Lambda|}{|\Delta|} = \frac{1}{2} \log \left(\frac{\sigma_{W_2}^2}{D - \frac{\sigma_{W_1}^2}{1 - \alpha^2}} \right). \quad (48)$$

We note that (48) holds if and only if $\mu_{\Delta,2} < \mu_{\Lambda,2}$ where $\mu_{\Lambda,2} = \sigma_{W_2}^2$.

Fig. IV.2 illustrates a comparison between the optimal closed form solution for this class of \mathbb{R}^2 -valued Gauss-Markov processes and the optimal numerical solution obtained via SDP for some random values of α , $\sigma_{W_1}^2$ and $\sigma_{W_2}^2$.

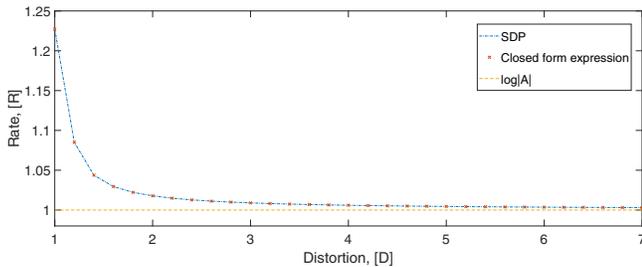


Fig. IV.2: Comparison of SDP with the closed form expression of $R^{\text{na}}(D)$ for a \mathbb{R}^2 -valued Gauss-Markov process with parameters $\alpha = 2$, $\sigma_{W_1}^2 = 0.1$, $\sigma_{W_2}^2 = 1$.

V. CONCLUSIONS AND FUTURE DIRECTIONS

In this paper, we revisited the asymptotic nonanticipative or sequential rate distortion problem and derive the parametric expression of the Gaussian NRDF subject to a reverse-waterfilling algorithm. Then, we proposed an iterative algorithm reminiscent of the classical reverse-waterfilling algorithm to compute the reverse-waterfilling solution. This scheme is, in general, suboptimal but at high rates performs optimally. Moreover, we used the matrix equations of the reverse-waterfilling algorithm to obtain the optimal closed form solution of a simple \mathbb{R}^2 -valued Gauss-Markov source. Our framework is demonstrated via numerical experiments. The proposed framework is general and, therefore, it is not restricted to the analytical solution of the Gaussian NRDF derived in this paper. Instead, it can be used to obtain analytical expressions of the Gaussian NRDF for more general \mathbb{R}^p -valued Gauss-Markov processes. To the best of our knowledge, this is the first work where analytical solutions of the Gaussian NRDF beyond the scalar case are provided.

As an ongoing research, we investigate schemes that may lead to the optimal evaluation of the reverse-waterfilling algorithm of Theorem 3 for the whole rate-distortion region.

REFERENCES

- [1] A. K. Gorbunov and M. S. Pinsker, "Prognostic epsilon entropy of a Gaussian message and a Gaussian source," *Problems Inf. Transmiss.*, vol. 10, no. 2, pp. 93–109, 1974.
- [2] —, "Nonanticipatory and prognostic epsilon entropies and message generation rates," *Problems Inf. Transmiss.*, vol. 9, no. 3, pp. 184–191, 1973.
- [3] S. Tatikonda, A. Sahai, and S. Mitter, "Stochastic linear control over a communication channel," *IEEE Trans. Autom. Control*, vol. 49, pp. 1549–1561, 2004.
- [4] M. S. Derpich and J. Østergaard, "Improved upper bounds to the causal quadratic rate-distortion function for Gaussian stationary sources," *IEEE Trans. Inf. Theory*, vol. 58, no. 5, pp. 3131–3152, May 2012.
- [5] T. Tanaka, K. K. Kim, P. A. Parrilo, and S. K. Mitter, "Semidefinite programming approach to Gaussian sequential rate-distortion trade-offs," *IEEE Trans. Autom. Control*, vol. 62, no. 4, pp. 1896–1910, April 2017.
- [6] P. A. Stavrou, T. Charalambous, and C. D. Charalambous, "Finite-time nonanticipative rate distortion function for time-varying scalar-valued Gauss-Markov sources," *IEEE Control Systems Letters*, vol. 2, no. 1, pp. 175–180, Jan 2018.
- [7] T. Berger, *Rate Distortion Theory: A Mathematical Basis for Data Compression*. Englewood Cliffs, NJ: Prentice-Hall, 1971.
- [8] P. A. Stavrou, T. Charalambous, C. D. Charalambous, and S. Loyka, "Optimal estimation via nonanticipative rate distortion function and applications to time-varying Gauss-Markov processes," *SIAM J. on Control Optim.*, 2018, (to appear).
- [9] P. A. Stavrou, T. Tanaka, and T. Tatikonda, "The time-invariant multidimensional Gaussian sequential rate-distortion problem revisited," *IEEE Trans. Autom. Control*, 2017, (under review). [Online]. Available: <https://arxiv.org/pdf/1711.09853>
- [10] 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.
- [11] W. M. Wonham, "On a matrix Riccati equation of stochastic control," *SIAM Journal on Control*, vol. 6, no. 4, pp. 681–697, 1968.
- [12] D. S. Bernstein, *Matrix Mathematics: Theory, Facts, and Formulas*, 2nd ed. Princeton University Press, 2011.
- [13] A. Zahedi, J. Østergaard, S. H. Jensen, P. A. Naylor, and S. Bech, "Source coding in networks with covariance distortion constraints," *IEEE Trans. Signal Proc.*, vol. 64, no. 22, pp. 5943–5958, Nov 2016.
- [14] D. G. Luenberger, *Optimization by Vector Space Methods*. John Wiley & Sons, Inc., New York, 1969.