

# Asymptotic Reverse Waterfilling Algorithm of NRDF for Certain Classes of Vector Gauss-Markov Processes

Photios A. Stavrou and Mikael Skoglund

**Abstract**—In this paper, we revisit the asymptotic reverse-waterfilling characterization of the nonanticipative rate distortion function (NRDF) derived for a time-invariant multidimensional Gauss-Markov processes with mean-squared error (MSE) distortion in [1]. We show that for certain classes of time-invariant multidimensional Gauss-Markov processes, the specific characterization behaves as a reverse-waterfilling algorithm obtained in *matrix form* ensuring that the numerical approach of [1, Algorithm 1] is optimal. In addition, we give an equivalent characterization that utilizes the *eigenvalues of the involved matrices* reminiscent of the well-known reverse-waterfilling algorithm in information theory. For the latter, we also propose a novel numerical approach to solve the algorithm optimally. The efficacy of our proposed iterative scheme compared to similar existing schemes is demonstrated via experiments. Finally, we use our new results to derive an analytical solution of the asymptotic NRDF for a correlated time-invariant two-dimensional Gauss-Markov process.

**Index Terms**—Nonanticipative rate distortion function, vector-valued Gauss-Markov processes, orthogonal matrices, commute by pairs, simultaneous diagonalization, reverse-waterfilling.

## I. INTRODUCTION

Tatikonda *et al.* in [2] studied the fundamental limitations of a multidimensional closed-loop control system when a communication link intervenes between a stochastic linear fully observable time-invariant plant driven by a Gaussian noise process and a controller whereas the performance criterion is the classical linear quadratic cost. First, the authors of [2] gave sufficient conditions for their problem to ensure the control-theoretic separation principle between the feedback controller and the state estimation when the channel is either noiseless digital or noisy (and memoryless). Then, provided these conditions hold, they showed that the optimal cost of control can be cast by two terms, a full-knowledge cost due to the fully observable state and a communication cost that can be computed by a state estimation problem. To compute the communication cost, they introduced a Gaussian sequential rate distortion function (SRDF) defined by minimizing a variant of directed information [3] subject to a *pointwise* mean squared error (MSE) distortion function, a definition that is attributed to Gorbunov and Pinsker in [4]. This information measure when solved, it ensures a lower bound on the communication cost irrespectively of whether the channel is noiseless or noisy.

For time-invariant scalar-valued Gauss-Markov processes, the authors of [2] found an explicit expression of a lower bound on the optimal communication cost by computing in closed form the expression of the asymptotic Gaussian SRDF (with an unbounded distortion budget for unstable processes). The same expression (with a bounded distortion budget) was already derived for stationary Gauss-Markov processes in [5, Eq. (1.43)] (see also the more recent work in [6, Theorem 3]). For vector-valued Gauss-Markov processes, [2] showed that the lower bound on the optimal communication cost can be obtained subject to a reverse-waterfilling algorithm. Unfortunately, as shown in [7], [8] the reverse-waterfilling algorithm proposed in [2] gives a suboptimal solution to the SRDF. Fortunately, a way to compute optimally the lower bound on the communication cost for vector-valued Gauss-Markov processes was recently proposed by Tanaka *et al.* in [9]. Specifically, the authors in [9] proved that the SRDF for multidimensional Gauss-Markov processes with MSE distortion is semidefinite representable and thus readily computable. Although this numerical approach is an important step, it lacks the insight of the parametric reverse-waterfilling solution that may lead to analytical solutions for this problem beyond scalar-valued processes. Moreover, semidefinite programming (SDP) is a computationally expensive choice to be implemented in modern delay-constrained and computationally-limited devices and it does not adapt well to high dimensional systems. Therefore, understanding the correct reverse-waterfilling characterization that completely solves the multidimensional case in [2] is still an important open task to be accomplished.

Toward this end, Stavrou *et al.* in [10] characterized the SRDF 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 dynamic reverse-waterfilling problem similar to the well-known reverse-waterfilling of classical RDF [11, Theorem 10.3.3]. The authors of [12] characterized the NRDF for time-varying vector-valued Gauss-Markov process subject to total and pointwise MSE distortion function via the solution of a dynamic reverse-waterfilling problem and they gave an iterative approach that solves the problem optimally only when all dimensions of the multi-input multi-output (MIMO) system are active. This result was restricted to time-invariant processes in [1].

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.

### A. Contributions

In this paper, we use the asymptotic reverse-waterfilling characterization of [1, Theorem 3] to obtain the following new results.

- (1) We show that for certain classes of time-invariant vector-valued Gauss-Markov processes, [1, Theorem 3] simplifies to a reverse-waterfilling algorithm obtained in matrix form (see Theorem 2). Moreover, this reverse-waterfilling algorithm has an equivalent form that amounts only to the eigenvalues of the involved matrices (we refer to this algorithm as the “*eigenvalue form*”) and is similar to the classical reverse-waterfilling algorithm in information theory (see Corollary 1);
- (2) We construct an iterative scheme that computes optimally the reverse-waterfilling algorithm obtained in eigenvalue form (see Algorithm 1);
- (3) We derive a closed form expression of the asymptotic NRDF for a two-dimensional correlated Gauss-Markov process subject to a MSE distortion (see §IV).

*Discussion of the results.* (1) The two equivalent forms of the reverse-waterfilling algorithm, i.e., the matrix form and the eigenvalue form, are obtained because in our analysis we prove that for the specific classes of time-invariant Gauss-Markov processes, all matrices commute by pairs hence they are simultaneously diagonalizable by an orthogonal matrix (an information lossless operation).<sup>1</sup> (2) The iterative scheme proposed herein is shown to execute much faster compared to similar numerical approaches that solve the same problem, e.g., [9], [1] and adapts extremely well to high dimensional systems. (3) To the best of the authors knowledge, this is the first non-trivial example of computing asymptotic NRDF beyond scalar-valued Gauss-Markov processes.

The rest of the paper is structured as follows. In §II, we give preliminaries and known results. In §III, we derive our new results and in §IV we apply our results to compute analytically an example. We draw conclusions and future directions in §V.

## II. PRELIMINARIES AND KNOWN 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  $\sigma$ -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,  $\mathbf{P}_X(dx) \equiv \mathbf{P}(dx)$  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  $X^{-1} = (X_{-\infty}, \dots, X_{-1})$ . We denote  $\Sigma \succ 0$  (resp.,  $\Sigma \succeq 0$ ) the symmetric positive-definite matrix (respectively, symmetric positive-semidefinite matrix). When we write  $\Sigma \succ \Sigma'$  (resp.  $\Sigma \succeq \Sigma'$ ) we mean that  $\Sigma - \Sigma' \succ 0$  (resp.  $\Sigma - \Sigma' \succeq 0$ ). The determinant, trace and eigenvalues of a square matrix  $\Sigma \in \mathbb{R}^{p \times p}$  are denoted by  $|\Sigma|$ ,  $\text{trace}(\Sigma)$ , and  $\text{diag}(\mu_{\Sigma,i})$ , respectively, where  $\mu_{\Sigma,i}$  denotes the  $i^{\text{th}}$  eigenvalue of matrix

<sup>1</sup>For details on the definition of commuting matrices, orthogonal matrices and the theorem that connects commuting matrices and simultaneously diagonalizable matrices see Appendix A.

$\Sigma$ .  $I_p$  is the  $\mathbb{R}^p$ -valued identity matrix.  $\mathbf{E}\{\cdot\}$  denotes the expectation operator and  $\|\cdot\|$  denotes the Euclidean norm.

In rate distortion theory one is interested to reproduce sequences of symbols  $X^n = x^n$  generated by a source, by their reproduction symbols  $Y^n = y^n$ , subject to a 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 at  $t = 0$ ,  $\mathbf{P}(dx_0|x^{-1}) = \mathbf{P}(dx_0)$ . Following [4], the channel that is used to reproduce  $X^n = x^n$  generated by a source, by the reproduction symbols  $Y^n = y^n$ , is described by the conditional distribution<sup>2</sup>

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

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

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

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 [4] and further analyzed in [2].

**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 the infimum exists and it is finite. 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 both infimum and limit exist and they are finite.

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

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

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

<sup>2</sup>The notation  $\mathbf{P}(\cdot||\cdot)$  is adopted from [13].

provided the infimum and the limit exist and 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 or time-invariant Gauss-Markov process it is shown in [4, Theorem 4] (see also [8, Theorem 1]) that  $R^{\text{na}}(D) = \widehat{R}^{\text{na}}(D)$ .

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  $D \in (0, D_{\max}) \subseteq (0, \infty)$ , and there exists an optimal solution characterizing it under general source distributions and distortion functions (see, e.g., [14]). By [12], 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, by [15] the joint process  $\{(X_t, Y_t) : t \in \mathbb{N}_0\}$  is jointly Gaussian (this is also shown in [2]). Moreover, [12] showed that for a multidimensional Gauss-Markov processes with average (over time) *total* MSE distortion constraint, the conditionally Gaussian optimal minimizer of a jointly Gaussian process is of the form

$$\mathbf{P}^*(dy_t|y^{t-1}, x^t) = \mathbf{P}^*(dy_t|y_{t-1}, x_t), \quad t \in \mathbb{N}_0. \quad (10)$$

A similar result for *per-letter* MSE distortion is derived in [5, Theorem 5].

The next result is derived in [16, Theorem 3].

**Lemma 1:** Suppose that any  $\mathbf{P}^*(dy_t|y_{t-1}, x_t)$  is time-invariant and there output distribution  $\mathbf{P}^*(dy_t|y_{t-1})$  is also time-invariant with a unique invariant distribution. Then, the following holds.

(1) For  $D > 0$ , if  $R^{\text{na}}(D) < \infty$ , then, it is given by

$$R^{\text{na}}(D) = \inf_{\Delta \succ 0: \text{trace}(\Delta) \leq D} \frac{1}{2} \log \frac{|\Lambda|}{|\Delta|}, \quad (11)$$

s.t.  $0 < \Delta \preceq \Lambda$

(2) The asymptotic optimal reproduction distribution is realized by

$$Y_t = HX_t + (I - H)AY_{t-1} + V_t^c, \quad (12)$$

where

$$H \triangleq I - \Delta \Lambda^{-1} \succeq 0, \quad \Sigma_V \triangleq \Delta H^T \succeq 0, \quad (13)$$

$$\Lambda = A\Delta A^T + \Sigma_W, \quad \Delta \succeq 0, \quad \Lambda \succ 0. \quad (14)$$

Next, we state the solution of (11) when solving the KKT conditions [17]. This result is derived in [1, Theorem 3] and it is a consequence of the derivation in [12, Appendix].

**Theorem 1:** (Reverse-waterfilling characterization)

The parametric solution subject to a reverse-waterfilling algorithm that corresponds to (11) is the following.

$$R^{\text{na}}(D) = \frac{1}{2} \log \frac{|\Lambda|}{|\Delta|}, \quad (15)$$

such that

$$\Delta = \begin{cases} \Delta_1^*, & \text{if } \Delta \prec \Lambda, \\ \Delta_2^*, & \text{if } \Delta \preceq \Lambda, \\ \Lambda, & \text{otherwise} \end{cases} \quad (16)$$

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

$$\left(-\frac{I}{2}\right) \Delta_1^* + \Delta_1^* \left(-\frac{I}{2}\right) - \Delta_1^* B \Delta_1^* + \frac{1}{2\theta} I = 0, \quad (17)$$

$$B \triangleq A^T \Sigma_W^{-1} A, \quad (18)$$

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

$$\left(-\frac{I}{2}\right) \Delta_2^* + \Delta_2^* \left(-\frac{I}{2}\right) - \Delta_2^* B \Delta_2^* + \Upsilon^{-1} = 0, \quad (19)$$

$$\Upsilon \triangleq 2(\theta I + F_2 - A^T F_2 A), \quad \Upsilon = \Upsilon^T \succ 0, \quad (20)$$

with  $\theta > 0$  and  $F_2 = F_2^T \succeq 0$  such that

$$\text{trace}(\Delta) = D, \quad F_2(\Delta - \Lambda) = 0. \quad (21)$$

**Remark 1:** (Comments on Theorem 1)

The characterization of Theorem 1 is derived based on the algebraic Riccati equations (17), (19). (17) allocates the distortion when there is no reverse-waterfilling in dimension, i.e., when  $\Delta \prec \Lambda$ . On the contrary, (19) takes care the distortion allocation at each dimension when some dimensions or all are inactive. This corresponds to  $\Delta \preceq \Lambda$ . Unfortunately, finding  $F_2$  in (19) is very challenging, in general. Therefore, in [1, Algorithm 1], the authors proposed a numerical approach to compute (17) optimally, i.e.,  $\Delta \prec \Lambda$ .

In this paper, we show that for certain classes of multivariate Gauss-Markov processes, the characterization in Theorem 1 is independent of (19) and consequently of the condition  $F_2(\Delta - \Lambda) = 0$ . Hence, for these classes of sources, the numerical approach derived in [1, Algorithm 1] is optimal.

### III. NEW RESULTS

In this section, we derive our main results.

First, we prove the following proposition.

**Proposition 1:** (Matrix structure of  $(A, \Sigma_W)$ )

Suppose that in (4) the matrix structure of the pair  $(A, \Sigma_W)$  satisfies one of the following three cases:

- (i)  $A = \alpha I_p$  (scalar matrix) and  $\Sigma_W \succ 0$ ;
- (ii)  $A$  is symmetric and  $\Sigma_W = \sigma_w^2 I_p$  (scalar matrix), where  $\sigma_w^2 > 0$ .
- (iii)  $A = \Sigma_W \succ 0$ ;

Then, in (11)  $(A, \Delta, \Sigma_W)$  commute by pairs which means that  $\Lambda$  and  $\Delta$  commute.<sup>3</sup>

*Proof:* See Appendix B. ■

Proposition 1 essentially reveals structural conditions on (4) such that all matrices of the optimization problem in (11) commute by pairs. Using this result, we now prove the following main result.

**Theorem 2:** (Asymptotic reverse-waterfilling algorithm: matrix form)

Suppose that the conditions of Proposition 1 hold. Then, the characterization of Theorem 1 can be simplified as the follows.

$$R^{\text{na}}(D) = \frac{1}{2} \log \frac{|\Lambda|}{|\Delta|}, \quad (22)$$

<sup>3</sup>The definition of commuting matrices and its generalization of commuting matrices by pairs is given in Definition 2 in Appendix A.

such that

$$\Delta = \begin{cases} \Delta^*, & \text{if } \Delta \prec \Lambda, \\ \Lambda, & \text{otherwise} \end{cases} \quad (23)$$

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

$$\left(-\frac{I}{2}\right) \Delta^* + \Delta^* \left(-\frac{I}{2}\right) - \Delta^* B \Delta^* + \frac{1}{2\theta} I = 0, \quad (24)$$

with  $\theta > 0$  such that  $\text{trace}(\Delta) = D$ .

*Proof:* Under the conditions of Proposition 1, we can reformulate Theorem 1 as follows:

$$R^{\text{na}}(D) = \sum_{i=1}^p R_i^{\text{na}}, \quad R_i^{\text{na}} \triangleq \frac{1}{2} \log \left( \frac{\mu_{\Lambda,i}}{\mu_{\Delta,i}} \right). \quad (25)$$

$$\mu_{\Delta,i} = \begin{cases} \mu_{\Delta_1^*,i}, & \text{if } \mu_{\Delta,i} < \mu_{\Lambda,i}, \\ \mu_{\Delta_2^*,i} & \text{if } \mu_{\Delta,i} \leq \mu_{\Lambda,i}, \\ \mu_{\Lambda,i}, & \text{otherwise} \end{cases}, \quad \forall i, \quad (26)$$

where  $\mu_{\Lambda,i} = \mu_{A^2,i} \mu_{\Delta,i} + \mu_{\Sigma_W,i}$ ,  $\mu_{\Delta_1^*,i} > 0$  is computed from the quadratic equation

$$\mu_{\Delta_1^*,i} \mu_{B,i} + \mu_{\Delta_1^*,i} - \frac{1}{2\theta} = 0, \quad \forall i, \quad (27)$$

with  $\mu_{B,i} \triangleq \frac{\mu_{A^2,i}}{\mu_{\Sigma_W,i}}$ , and  $\mu_{\Delta_2^*,i} > 0$  is computed from the quadratic equation

$$\mu_{\Delta_2^*,i} \mu_{B,i} + \mu_{\Delta_2^*,i} + \frac{1}{2(\theta + (1 - \mu_{A^2,i})\mu_{F_2,i})} = 0, \quad \forall i, \quad (28)$$

with  $\theta > 0$  and  $\mu_{F_2,i} \geq 0$  such that

$$\sum_{i=1}^p \mu_{\Delta,i} = D, \quad \mu_{F_2,i}(\mu_{\Delta,i} - \mu_{\Lambda,i}) = 0, \quad \forall i. \quad (29)$$

The above equivalent characterization of Theorem 1 follows because all matrices  $(A, \Delta, \Sigma_W, \Lambda, F_2)$  commute by pairs and they are simultaneously diagonalizable by an orthogonal matrix.<sup>4</sup> Clearly, from (25) and the third constraint of (26) we can easily see that whenever  $\mu_{\Lambda,i} \leq \mu_{\Delta,i}$ , for any  $i$ ,  $R_i^{\text{na}} = 0$ . This also means that the second condition of (26) is always satisfied once the other two conditions are imposed in (26). Hence without any loss of optimality we can remove the second condition of (26) and consequently (28). This also means that the condition  $\mu_{F_2,i}(\mu_{\Delta,i} - \mu_{\Lambda,i}) = 0, \forall i$  is always satisfied and can be removed.

Hence the resulting asymptotic reverse-waterfilling characterization of (25)-(29) without (28) can also be written as an equivalent asymptotic reverse-waterfilling characterization in matrix form given by (22)-(24). This completes the derivation. ■

A way to solve optimally Theorem 2 is already proposed in [1, Algorithm 1]. That numerical scheme is shown to operate faster compared to the SDP algorithm proposed in [9, Eq. (27)] because it only schedules the rate-distortion allocation based on the already solved optimization problem in (11) (obtained in the form of an algebraic Riccati equation)

whereas SDP solver uses interior point methods to solve the optimization problem (11) which are computationally much more expensive.

Although the characterization of Theorem 2 is attractive, one may claim that it is not the same as the classical form of a reverse-waterfilling algorithm, see, e.g., [11, Theorem 10.3.3]. For this reason, we reformulate the asymptotic reverse-waterfilling algorithm of Theorem 2 into an equivalent form that is similar to the classical form of a reverse-waterfilling algorithm in information theory. We refer to this form of the reverse-waterfilling algorithm as the *eigenvalue form*.

**Corollary 1:** (Asymptotic reverse waterfilling algorithm: eigenvalue form)

Under the conditions of Proposition 1, the parametric solution subject to an asymptotic reverse-waterfilling algorithm that corresponds to (11) is the following.

$$R^{\text{na}}(D) = \frac{1}{2} \sum_{i=1}^p \log \left( \frac{\mu_{\Lambda,i}}{\mu_{\Delta,i}} \right), \quad (30)$$

where  $\mu_{\Lambda,i} = \mu_{A^2,i} \mu_{\Delta,i} + \mu_{\Sigma_W,i}$ , and  $\mu_{\Delta,i}$  is computed by the reverse-waterfilling algorithm

$$\mu_{\Delta,i} = \begin{cases} \mu_{\Delta^*,i}, & \text{if } \mu_{\Delta,i} < \mu_{\Lambda,i}, \\ \mu_{\Lambda,i}, & \text{otherwise} \end{cases}, \quad \forall i, \quad (31)$$

where  $\mu_{\Delta^*,i} > 0$  is given by

$$\mu_{\Delta^*,i} = \frac{1}{2\mu_{B,i}} \left( \sqrt{1 + \frac{2\mu_{B,i}}{\theta}} - 1 \right), \quad \text{if } \mu_{B,i} \neq 0, \quad \forall i, \quad (32)$$

$$\mu_{\Delta^*,i} = \frac{1}{2\theta}, \quad \text{if } \mu_{B,i} = 0$$

where  $\mu_{B,i} \triangleq \frac{\mu_{A^2,i}}{\mu_{\Sigma_W,i}}$  and  $\theta > 0$  is chosen such that  $\sum_{i=1}^p \mu_{\Delta,i} = D$ .

*Proof:* The proof is immediate from the derivation of Theorem 2. Note that (32) is the positive solution of the quadratic equation (27). ■

In Algorithm 1 we propose an iterative scheme to solve optimally the reverse-waterfilling algorithm of Corollary 1. This approach is similar to the dynamic reverse-waterfilling algorithm proposed in [18, Algorithm 1] (see also [10, Algorithm 1]) but instead of the time index we now have the spatial (or dimension) index.

*Comparison of existing algorithms that solve optimally (11):* In what follows, we demonstrate a simulation experiment where we consider matrix structure for  $(A, \Sigma_W)$  within the two classes of vector-valued time-invariant Gauss-Markov processes introduced in Proposition 1 and a given distortion level  $D$ . Then, we solve (11) using the eigenvalue form of Algorithm 1, the equivalent matrix form that can be solved using [1, Algorithm 1] and the SDP approach of [9, Equation (27)]. We try three distinct experiments where for each we change the number of dimensions of the system. Our aim is twofold. First, to study if Algorithm 1 is faster than [1, Algorithm 1] that in turn is known to be much faster than [9, Equation (27)]. Second, to study its scalability by increasing the dimensions of the system.

To illustrate this point, in Table I, we provide for the first two examples the mean and the standard deviation of

<sup>4</sup>This follows from Theorem 3 in Appendix A.



---

**Algorithm 1** Reverse-waterfilling algorithm of Corollary 1
 

---

**Initialize:** number of spatial components  $p$ ; distortion level  $D$ ; error tolerance  $\epsilon$ ; nominal minimum and maximum value  $\theta^{\min} = 0$  and  $\theta^{\max} = \frac{p}{2D}$ ; pick an initial variance for  $\mu_{\Delta,1}$ ; pick the matrix structure of  $(A, \Sigma_W)$  in (4) and their corresponding eigenvalues  $\{(\mu_{A,i}, \mu_{\Sigma_W,i}) : i \in \mathbb{N}_1^p\}$ . Set  $\theta = p/2D$ ; flag = 0.

**while** flag = 0 **do**

  Compute  $\mu_{\Delta,i} \forall i$  as follows:

**for**  $i = 1 : p$  **do**

    Compute  $\mu_{\Delta^*,i}$  according to (32).

    Compute  $\mu_{\Delta,i}$  according to (31).

**end for**

**if**  $\sum_{i=1}^p \mu_{\Delta,i} - D \geq \epsilon$  **then**

    Set  $\theta^{\min} = \theta$ .

**else**

    Set  $\theta^{\max} = \theta$ .

**end if**

**if**  $\theta^{\max} - \theta^{\min} \geq \frac{\epsilon}{2}$  **then**

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

**else**

    flag  $\leftarrow 1$

**end if**

**end while**

**Output:**  $\{\mu_{\Delta,i} : i \in \mathbb{N}_1^p\}$ ,  $\{\mu_{\Lambda,i} : i \in \mathbb{N}_1^p\}$ , for a given distortion level  $D$ .

---

the computational time needed for each of the three optimal numerical methods to execute over a sample of 1000 instances. All algorithms operate using an error tolerance  $\epsilon = 10^{-9}$ . For the last example we could not extract conclusive results for SDP because to execute over a sample of 1000 instances it takes days. For the first experiment we assume  $10 \times 10$

Solver (Numb. dims. $p = 10$ )	Mean	Stand. Deviation
SDP [9] (by default $\epsilon = 10^{-9}$ )	0.701	0.0859
Algorithm 1 [1] ( $\epsilon = 10^{-9}$ )	0.031	0.0038
Algorithm 1 ( $\epsilon = 10^{-9}$ )	$5.24 \times 10^{-4}$	$2.52 \times 10^{-4}$
Solver (Numb. dims. $p = 50$ )	Mean	Stand. Deviation
SDP [9] (by default $\epsilon = 10^{-9}$ )	138.017	5.86
Algorithm 1 [1] ( $\epsilon = 10^{-9}$ )	0.8317	0.0617
Algorithm 1 ( $\epsilon = 10^{-9}$ )	$8.14 \times 10^{-4}$	$4.59 \times 10^{-4}$
Solver (Numb. dims. $p = 150$ )	Mean	Stand. Deviation
SDP [9] (by default $\epsilon = 10^{-9}$ )	non-conclusive	non-conclusive
Algorithm 1 [1] ( $\epsilon = 10^{-9}$ )	4.551	0.1736
Algorithm 1 ( $\epsilon = 10^{-9}$ )	0.002	$5.19 \times 10^{-4}$

TABLE I: Comparison of the computational time needed between SDP in [9], [1, Algorithm 1] and Algorithm 1 for 1000 instances.

symmetric matrices. Our results demonstrate that Algorithm 1 is extremely fast operating around 1300 times faster than SDP and is significantly faster than [1, Algorithm 1] (around 60 times). By increasing the number of dimensions in the optimization problem (we assume  $50 \times 50$  symmetric matrices), the results are even more impressive. Algorithm 1 is still extremely fast operating around 1000 times faster than [1, Algorithm 1]

and more than 140000 times faster than SDP. By assuming a relatively high-dimensional system (i.e.,  $150 \times 150$  symmetric matrices) we observe that Algorithm 1 is significantly faster than [1, Algorithm 1] (more than 2000 times).

Overall, the previous experiments show that Algorithm 1 is an elegant and more preferable choice to be implemented in modern delay-constrained and computationally-limited devices like the architectures of networked control systems, compared to SDP or [1, Algorithm 1]. It is very fast and adapts extremely well in high dimensional systems (it is scalable).

Note that both Algorithm 1 and [1, Algorithm 1] can allow for different levels of tolerance (and thus become faster or slower accordingly). The choice of  $\epsilon = 10^{-9}$  is made to achieve the same accuracy with the SDP solver.

#### IV. EXAMPLE WITH ANALYTICAL SOLUTION

In this section, we give an example in which we derive an analytical solution of  $R^{\text{na}}(D)$  for a correlated time-invariant  $\mathbb{R}^2$ -valued Gauss-Markov process. This is feasible because we choose  $(A, \Sigma_W)$  to be in the class of matrices that satisfy the conditions of Proposition 1. We compare our analytical solution to the optimal numerical solution of [9, Equation (27)].

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

$$(A, \Sigma_W) = \left( \begin{bmatrix} 1 & -0.5 \\ -0.5 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right). \quad (33)$$

Clearly, the matrix structure in (33) is in the class of matrices that satisfy the conditions of Proposition 1. We will use Theorem 1, to derive a closed form solution for  $R^{\text{na}}(D)$  for this specific correlated time-invariant Gauss-Markov source. We solve the full-rank case, i.e., when  $\Delta \prec \Lambda$  that corresponds to the algebraic Riccati equation (17), and, the rank-deficient case, i.e., when  $\Delta \preceq \Lambda$  that corresponds to the algebraic Riccati equation (19). We will show that (19) is never used to obtain the optimal solution something expected from Theorem 2. We note that the last case in (16) requires zero rate, hence for this case the problem's solution is trivial.

**Full-rank Solution ( $\Delta \prec \Lambda$ ):** First, we use the spectral representation or eigenvalue decomposition of the real symmetric matrix  $\Delta_1^* \equiv \Delta$  of (17) which means that for an orthogonal matrix  $U \in \mathbb{R}^{p \times p}$ , then,  $U^T \Delta U = \text{diag}(\mu_{\Delta,i})$ . This transforms (17) to the following matrix equation:

$$\text{diag}(\mu_{\Delta,i}) U^T B U \text{diag}(\mu_{\Delta,i}) + \text{diag}(\mu_{\Delta,i}) - \frac{1}{2\theta} I = 0. \quad (34)$$

Observe that in (34), the equation is satisfied if and only if  $U^T B U$  is diagonal. Since,

$$B = \begin{bmatrix} 1.25 & -1 \\ -1 & 1.25 \end{bmatrix} \quad (35)$$

we can obtain an orthogonal matrix  $U$  of the form:

$$U = \begin{bmatrix} -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix}. \quad (36)$$

From (34), we observe that  $U$  is the orthogonal matrix that simultaneously diagonalizes  $(A, \Sigma_W, \Delta)$  and also  $\Lambda$  (This is

expected from Proposition 1). Clearly, (34) can be reformulated in the equivalent *eigenvalue form* of (27) as expected from the derivation of Theorem 2 (see Corollary 1). In this example, we put eigenvalues in a matrix form for better illustration of our results.

By substituting (35), (36) in (34) we obtain:

$$\Delta = \begin{bmatrix} \mu_{\Delta,1} & 0 \\ 0 & \mu_{\Delta,2} \end{bmatrix}, \quad (37)$$

where

$$\mu_{\Delta,1} = -\frac{2}{9} \pm \frac{2}{9} \sqrt{1 + \frac{9}{2\theta}}, \quad (38a)$$

$$\mu_{\Delta,2} = -2 \pm 2 \sqrt{1 + \frac{1}{2\theta}}. \quad (38b)$$

Note that one of the solutions of (38a), (38b) is rejected because it is negative hence the resulting matrix  $\Delta$  is:

$$\Delta = \begin{bmatrix} \frac{2}{9} \left( \sqrt{1 + \frac{9}{2\theta}} - 1 \right) & 0 \\ 0 & 2 \left( \sqrt{1 + \frac{1}{2\theta}} - 1 \right) \end{bmatrix}. \quad (39)$$

In order to find  $\Lambda$  we use (14). This means that we need the eigenvalue decomposition of matrix  $A$ . This gives:

$$U^T A U = \begin{bmatrix} \frac{3}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix}. \quad (40)$$

Using (39) and (40) in (14) we obtain:

$$\Lambda = \begin{bmatrix} \left(\frac{3}{2}\right)^2 \mu_{\Delta,1} + 1 & 0 \\ 0 & \left(\frac{1}{2}\right)^2 \mu_{\Delta,2} + 1 \end{bmatrix}. \quad (41)$$

Now, we use the left hand side (LHS) equation in (21) to obtain an additional equation which is used to find  $\theta$ . By substituting (39) in (21) we obtain:

$$\frac{2}{9} \left( \sqrt{1 + \frac{9}{2\theta}} - 1 \right) + 2 \left( \sqrt{1 + \frac{1}{2\theta}} - 1 \right) = D. \quad (42)$$

Equation (42) gives two positive solutions. Using Lagrange duality theorem [19] we choose the one that results into greater rates and discard the one that gives the lower rates. This chosen  $\theta$  is as follows:

$$\theta = \frac{1}{\frac{D^2}{2} + D(-\mu_{\Delta,1} + 2) - \mu_{\Delta,1}(-\frac{\mu_{\Delta,1}}{2} + 2)}, \quad (43)$$

where

$$\mu_{\Delta,1} = \frac{\sqrt{(D+4)(9D+4)} - (D+4)}{8}, \quad (44a)$$

$$\mu_{\Delta,2} = \frac{(9D+4) - \sqrt{(D+4)(9D+4)}}{8}. \quad (44b)$$

Using (44) we can compute  $\Lambda$  in (41) which is given by:

$$\mu_{\Lambda,1} = \frac{9}{4} \left( \frac{\sqrt{(D+4)(9D+4)} - (D+4)}{8} \right) + 1, \quad (45a)$$

$$\mu_{\Lambda,2} = \frac{1}{4} \left( \frac{(9D+4) - \sqrt{(D+4)(9D+4)}}{8} \right) + 1. \quad (45b)$$

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

$$R^{\text{na}}(D) = \frac{1}{2} \log \frac{|\Lambda|}{|\Delta|} = \frac{1}{2} \sum_{i=1}^2 \log \left( \frac{\mu_{\Lambda,i}}{\mu_{\Delta,i}} \right) \quad (46)$$

$$\stackrel{(a)}{=} \frac{1}{2} \left\{ \log \left( \frac{9}{4} + \frac{1}{\mu_{\Delta,1}} \right) + \log \left( \frac{1}{4} + \frac{1}{\mu_{\Delta,2}} \right) \right\} \quad (47)$$

$$\stackrel{(b)}{=} \frac{1}{2} \left\{ \log \left( \frac{9}{4} + \frac{8}{\sqrt{(D+4)(9D+4)} - (D+4)} \right) + \log \left( \frac{1}{4} + \frac{8}{(9D+4) - \sqrt{(D+4)(9D+4)}} \right) \right\}, \quad (48)$$

where (a) follows from (39), (41); (b) follows by substituting (44), (45) in (47). Recall that (48) holds if and only if  $\mu_{\Delta,i} < \mu_{\Lambda,i}$ ,  $i = 1, 2$ , which means that

$$\mu_{\Delta,1}^{\max} \equiv \mu_{\Lambda,1} = \frac{9}{4} \mu_{\Delta,1} + 1, \quad |\mu_{\Lambda,1}| = 1.5 > 1, \quad (49)$$

$$\mu_{\Delta,2}^{\max} \equiv \mu_{\Lambda,2} = \frac{1}{1 - \mu_{\Delta,2}} = \frac{4}{3}, \quad |\mu_{\Lambda,2}| = 0.5 < 1. \quad (50)$$

**Rank-Deficient Solution ( $\Delta \preceq \Lambda$ ):** Again, we need to solve (19) such that the right hand side (RHS) equation of (21) is satisfied. First, observe that the RHS equation of (21) implies that  $\Delta - \Lambda \preceq 0$  and  $F_2 = F_2^T \succeq 0$  commute because they give zero matrix which is a symmetric matrix. Moreover, from Proposition 1 we have that  $(\Lambda, \Delta, F_2, A, \Sigma_W)$  commute by pairs, hence they are simultaneously diagonalizable by an orthogonal matrix  $U$  similar to (34). 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 (21) 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 (20) we obtain:

$$\Upsilon = \begin{bmatrix} \theta & 0 \\ 0 & \theta + \frac{3}{4} \mu_{F_2,2} \end{bmatrix}. \quad (51)$$

Moreover, by solving (19) we obtain:

$$\Delta = \begin{bmatrix} \frac{2}{9} \left( \sqrt{1 + \frac{9}{2\theta}} - 1 \right) & 0 \\ 0 & 2 \left( \sqrt{1 + \frac{1}{2(\theta + \frac{3}{4} \mu_{F_2,2})}} - 1 \right) \end{bmatrix}. \quad (52)$$

Additionally, using (14) we obtain  $\Lambda$  as follows:

$$\Lambda = \begin{bmatrix} \frac{9}{4} \mu_{\Delta,1} + 1 & 0 \\ 0 & \frac{4}{3} \end{bmatrix}. \quad (53)$$

where  $\mu_{\Lambda,2}$  is obtained because  $\mu_{\Delta,2} = \mu_{\Lambda,2}$  (from (50)). We observe that in this case, the term  $\mu_{\Delta,2} = 2 \left( \sqrt{1 + \frac{1}{2(\theta + \frac{3}{4} \mu_{F_2,2})}} - 1 \right)$  in (52) is not required. This is

expected because of our main result in Theorem 2. Hence, the LHS term of (21) gives:

$$\frac{2}{9} \left( \sqrt{1 + \frac{9}{2\theta}} - 1 \right) = D - \frac{4}{3}. \quad (54)$$

The solution of (54) gives

$$\theta = \frac{1}{\frac{9D^2}{2} - 10D + \frac{48}{9}}. \quad (55)$$

Substituting (55) and the fact that  $\mu_{\Delta,2} = \mu_{\Lambda,2}$  in (52) we obtain:

$$\Delta = \begin{bmatrix} \frac{2}{9} \left( \sqrt{1 + \frac{(9D-8)(9D-12)}{4}} - 1 \right) & 0 \\ 0 & \frac{4}{3} \end{bmatrix}. \quad (56)$$

Hence, for case 1 the sum rate is obtained as follows:

$$\begin{aligned} R^{\text{na}}(D) &= \frac{1}{2} \log \frac{|\Lambda|}{|\Delta|} = \frac{1}{2} \sum_{i=1}^2 \log \left( \frac{\mu_{\Lambda,i}}{\mu_{\Delta,i}} \right), \\ &\stackrel{(c)}{=} \frac{1}{2} \left\{ \log \left( \frac{9}{4} + \frac{1}{\mu_{\Delta,1}} \right) + \underbrace{\log \left( \frac{\mu_{\Lambda,2}}{\mu_{\Delta,2}} \right)}_{=0} \right\} \\ &= \frac{1}{2} \log \left( \frac{9}{4} + \frac{1}{\frac{2}{9} \left( \sqrt{1 + \frac{(9D-8)(9D-12)}{4}} - 1 \right)} \right). \end{aligned} \quad (57)$$

where (c) follows from (53) and (56). Recall that (57) holds if and only if  $\mu_{\Delta,1} < \mu_{\Lambda,1}$ .

Case 2: Upon solving (20) we obtain:

$$\Upsilon = \begin{bmatrix} \theta - \frac{5}{4}\mu_{F_2,1} & 0 \\ 0 & \theta \end{bmatrix}. \quad (58)$$

Moreover, by solving (19), we obtain:

$$\Delta = \begin{bmatrix} \frac{2}{9} \left( \sqrt{1 + \frac{9}{2(\theta - \frac{5}{4}\mu_{F_2,1})}} - 1 \right) & 0 \\ 0 & 2 \left( \sqrt{1 + \frac{1}{2\theta}} - 1 \right) \end{bmatrix}. \quad (59)$$

However, for this case we can observe that  $\mu_{\Lambda,1}$  can never be equal to  $\mu_{\Delta,1}$  because the source is unstable (see, eq. (49)). Moreover, even if one assumes that  $\mu_{\Lambda,1} = \mu_{\Delta,1}$  this will mean that  $\mu_{\Delta,1} < 0$  which is incorrect. Hence, Case 2 can never be observed in this example.

In Fig. IV.1 we illustrate a comparison between the closed form solution for this  $\mathbb{R}^2$ -valued Gauss-Markov process and the optimal numerical solution obtained via SDP for  $D \in [0.5, 10]$ .

*Discussion.* The analytical solution of  $R^{\text{na}}(D)$  is one small step towards more general analytical solutions within the classes of time-invariant multidimensional Gauss-Markov processes considered in this note. From the steps involved to compute  $R^{\text{na}}(D)$  for the specific vector source, it is clear that the characterization of Theorem 1, (19) is not needed (matrix  $F_2$  is never used). This is expected from our Theorem 2.

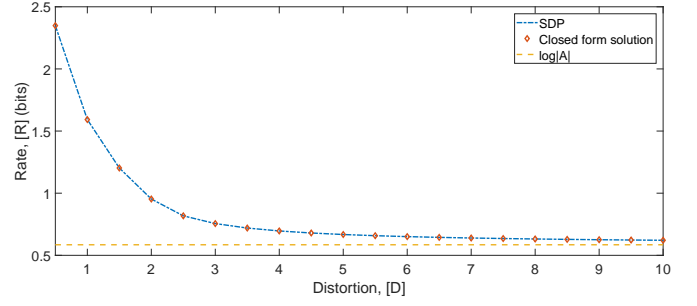


Fig. IV.1: Comparison of SDP with the closed form expression of  $R^{\text{na}}(D)$ .

## V. CONCLUSIONS AND FUTURE DIRECTIONS

In this paper, we considered certain classes of time-invariant multidimensional Gauss-Markov process which ensure that all matrices in the optimization problem (11) commute by pairs and thus they are simultaneously diagonalizable. As a result of this feature, we showed that the asymptotic reverse-waterfilling characterization of [1] can be simplified considerably and it is equivalent to a reverse-waterfilling algorithm obtained only by the eigenvalues of the involved matrices. For the latter algorithm, we proposed an iterative approach to compute optimally the optimization problem. We showed via experiments that the specific algorithm is extremely fast and much more scalable compared to the other existing algorithms in the literature that solve the same problem. Finally, we used our new results to obtain the closed form expression of NRDF for a correlated time-invariant  $\mathbb{R}^2$ -valued Gauss-Markov source.

Our ongoing research will focus on deriving more general analytical expressions than the one derived in this paper and on studying similar algorithms for the more general case of partially observable Gauss-Markov processes with MSE distortion function. Possible extension of Proposition 1 to time-varying systems is also under consideration.

## APPENDIX A

### USEFUL DEFINITIONS AND THEOREMS

In what follows, we state a few important definitions and a theorem that we use throughout the paper.

**Definition 2:** (Commuting matrices) [20, p. 5] Two  $p \times p$  matrices  $A, B$  commute if  $AB = BA$ . More generally, the collection of  $p \times p$  matrices  $(A_1, \dots, A_k)$  commute by pairs if  $A_i A_j = A_j A_i$ , for  $j > i$ ,  $i = 1, 2, \dots, k$ .

**Definition 3:** (Product of symmetric matrices) [20, Section 1.3 (6)] The production  $AB$  of two square symmetric matrices  $A, B$  is itself symmetric if and only if  $A$  and  $B$  commute.

**Definition 4:** (Orthogonal matrix) [20, Section 8.4 (a)] A square (say  $p \times p$ ) matrix  $U$  is said to be orthogonal if  $U^T U = U U^T = I_p$ .

**Theorem 3:** (Commuting matrices are simultaneously diagonalizable) [20, Theorem 21.13.1] If a collection of  $p \times p$  matrices  $(A_1, \dots, A_k)$  commute by pairs, then they can be simultaneously diagonalized with an orthogonal matrix  $U \in \mathbb{R}^{p \times p}$ , namely, there exists an orthogonal matrix  $U$  and diagonal matrices, say  $D_1, \dots, D_k$  such that for  $i = 1, \dots, k$

$U^T A_i U = D_i$ . If a collection of  $p \times p$  matrices  $(A_1, \dots, A_k)$  are *simultaneously diagonalizable*, then, they *commute by pairs*.

## APPENDIX B PROOF OF PROPOSITION 1

Recall that  $\Lambda$  and  $\Delta$  commute if  $\Lambda\Delta = \Delta\Lambda$  which further means that

$$(A\Delta A^T + \Sigma_W)\Delta = \Delta(A\Delta A^T + \Sigma_W), \quad (60)$$

from (14).

First, we consider case (i). Then, the LHS in (60) can be reformulated as follows:

$$\begin{aligned} & (A\Delta A^T + \Sigma_W)\Delta \\ & \stackrel{(a)}{=} V \text{diag}(\mu_{A,i}) V^T U \text{diag}(\mu_{\Delta,i}) U^T V \text{diag}(\mu_{A^T,i}) V^T \\ & \quad U \text{diag}(\mu_{\Delta,i}) U^T + S \text{diag}(\mu_{\Sigma_W,i}) S^T U \text{diag}(\mu_{\Delta,i}) U^T \\ & \stackrel{(b)}{=} U \text{diag}(\mu_{A^2,i}) \text{diag}(\mu_{\Delta^2,i}) U^T \\ & \quad + U \text{diag}(\mu_{\Sigma_W,i}) \text{diag}(\mu_{\Delta,i}) U^T \\ & = U \text{diag}((\mu_{A^2,i}\mu_{\Delta,i} + \mu_{\Sigma_W,i})\mu_{\Delta,i}) U^T, \end{aligned}$$

where (a) follows because  $(A, \Delta, \Sigma_W)$  are all real symmetric matrices and we can write their spectral representation (or eigenvalue decomposition), i.e., for  $A = A^T = V \text{diag}(\mu_{A,i}) V^T$ ,  $\Sigma_W = S \text{diag}(\mu_{\Sigma_W,i}) S^T$ , and  $\Delta = U \text{diag}(\mu_{\Delta,i}) U^T$  where  $(V, S, U)$  are the corresponding orthogonal matrices,<sup>5</sup> respectively; (b) the first part follows because  $A$  is a scalar matrix and commutes with every other matrix of the same dimensions (in this case  $\Delta$ ) and therefore we can choose  $V = U$  and the second part follows because  $U$  is a design parameter (because  $\Delta$  is a design parameter) and we can choose it such that  $U = S$ . The RHS in (60) follows similarly giving the exact same result. This proves (i).

Next, we consider case (ii). Then, the LHS in (60) can be reformulated as follows:

$$\begin{aligned} & (A\Delta A^T + \Sigma_W)\Delta \\ & = V \text{diag}(\mu_{A,i}) V^T U \text{diag}(\mu_{\Delta,i}) U^T V \text{diag}(\mu_{A^T,i}) V^T \\ & \quad U \text{diag}(\mu_{\Delta,i}) U^T + S \text{diag}(\mu_{\Sigma_W,i}) S^T U \text{diag}(\mu_{\Delta,i}) U^T \\ & \stackrel{(c)}{=} V \text{diag}(\mu_{A^2,i}) \text{diag}(\mu_{\Delta^2,i}) V^T \\ & \quad + U \text{diag}(\mu_{\Sigma_W,i}) \text{diag}(\mu_{\Delta,i}) U^T \\ & \stackrel{(d)}{=} U \text{diag}((\mu_{A^2,i}\mu_{\Delta,i} + \mu_{\Sigma_W,i})\mu_{\Delta,i}) U^T, \end{aligned}$$

where (c) follows using a similar argument to (b), i.e., the first term because  $U$  is a design parameter and we can choose it such that  $U = V$  and the second term because  $\Sigma_W$  is a scalar matrix and commutes with every other matrix of the same dimensions (in this case  $\Delta$ ) and therefore we can choose  $S = U$ ; (d) follows because  $U = V$  (by design). The RHS follows similarly, thus completing the proof for (ii).

Consider case (iii). Then, the LHS in (60) can be reformulated as follows:

$$\begin{aligned} & (A\Delta A^T + \Sigma_W)\Delta \\ & \stackrel{(e)}{=} V \text{diag}(\mu_{A,i}) V^T U \text{diag}(\mu_{\Delta,i}) U^T V \text{diag}(\mu_{A^T,i}) V^T \\ & \quad U \text{diag}(\mu_{\Delta,i}) U^T + V \text{diag}(\mu_{\Sigma_W,i}) V^T U \text{diag}(\mu_{\Delta,i}) U^T \\ & \stackrel{(f)}{=} U \text{diag}((\mu_{A^2,i}\mu_{\Delta,i} + \mu_{\Sigma_W,i})\mu_{\Delta,i}) U^T, \end{aligned}$$

where (e) follows because  $(A, \Sigma_W)$  have the same spectral representation; (f) because we can choose the design parameter  $U$  such that  $U = V$ . The RHS term in (60) follows similarly. This completes the proof for case (iii).

## REFERENCES

- [1] P. A. Stavrou, T. Charalambous, C. D. Charalambous, S. Loyka, and M. Skoglund, "Asymptotic reverse-waterfilling characterization of nonanticipative rate distortion function of vector-valued Gauss-Markov processes with MSE distortion," in *Proc. IEEE Conf. Decision Control*, Dec 2018, pp. 14–20.
- [2] S. Tatikonda, A. Sahai, and S. Mitter, "Stochastic linear control over a communication channel," *IEEE Trans. Autom. Control*, vol. 49, pp. 1549–1561, 2004.
- [3] J. L. Massey, "Causality, feedback and directed information," in *International Symposium on Information Theory and its Applications (ISITA '90)*, Nov. 27–30 1990, pp. 303–305.
- [4] 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.
- [5] —, "Nonanticipatory and prognostic epsilon entropies and message generation rates," *Problems Inf. Transmiss.*, vol. 9, no. 3, pp. 184–191, 1973.
- [6] 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.
- [7] V. Kostina and B. Hassibi, "Rate-cost tradeoffs in control," *IEEE Transactions on Automatic Control*, vol. 64, no. 11, pp. 4525–4540, Nov 2019.
- [8] P. A. Stavrou, T. Tanaka, and S. Tatikonda, "The time-invariant multidimensional Gaussian sequential rate-distortion problem revisited," *IEEE Trans. Autom. Control*, vol. 65, pp. 1–1, May 2020.
- [9] 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.
- [10] P. A. Stavrou, T. Charalambous, and C. D. Charalambous, "Finite-time nonanticipative rate distortion function for time-varying scalar-valued Gauss-Markov sources," *IEEE Control Systems Letters*, vol. 2, no. 1, pp. 175–180, Jan 2018.
- [11] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. John Wiley & Sons, Inc., Hoboken, New Jersey, 2006.
- [12] P. A. Stavrou, T. Charalambous, C. D. Charalambous, and S. Loyka, "Optimal estimation via nonanticipative rate distortion function for time-varying Gauss-Markov processes," *SIAM J. on Control Optim.*, vol. 56, no. 5, pp. 3731–3765, May 2018.
- [13] G. Kramer, "Directed information for channels with feedback," Ph.D. dissertation, ETH Zurich, 1998.
- [14] 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.
- [15] S. Ihara, *Information theory - for Continuous Systems*. World Scientific, 1993.
- [16] P. A. Stavrou, J. Østergaard, and C. D. Charalambous, "Zero-delay rate distortion via filtering for vector-valued Gaussian sources," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 5, pp. 841–856, Oct 2018.
- [17] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York, NY, USA: Cambridge University Press, 2004.
- [18] P. A. Stavrou, M. Skoglund, and T. Tanaka, "Sequential source coding for stochastic systems subject to finite rate constraints," *arxiv.org*, 2019. [Online]. Available: <https://arxiv.org/abs/1906.04217>
- [19] D. G. Luenberger, *Optimization by Vector Space Methods*. John Wiley & Sons, Inc., New York, 1969.
- [20] D. A. Harville, *Matrix Algebra From a Statistician's Perspective*. Springer-Verlag, 1997.

<sup>5</sup>The definition of an orthogonal matrix is given by Definition 4 in Appendix A.