

## Article

# Bounds on the Sum-Rate of MIMO Causal Source Coding Systems with Memory under Spatio-Temporal Distortion Constraints

Photios A. Stavrou <sup>1,\*</sup> , Jan Østergaard <sup>2</sup>  and Mikael Skoglund <sup>1</sup> 

<sup>1</sup> Department of Intelligent Systems, Division of Information Science and Engineering, KTH Royal Institute of Technology, 11428 Stockholm, Sweden; skoglund@kth.se

<sup>2</sup> Section on Signal and Information Processing, Department of Electronic Systems, Aalborg University, 9000 Aalborg, Denmark; jo@es.aau.dk

\* Correspondence: fstavrou@kth.se

Received: 18 June 2020; Accepted: 27 July 2020; Published: 30 July 2020



**Abstract:** In this paper, we derive lower and upper bounds on the OPTA of a two-user multi-input multi-output (MIMO) causal encoding and causal decoding problem. Each user's source model is described by a multidimensional Markov source driven by additive i.i.d. noise process subject to three classes of spatio-temporal distortion constraints. To characterize the lower bounds, we use state augmentation techniques and a data processing theorem, which recovers a variant of rate distortion function as an information measure known in the literature as nonanticipatory  $\epsilon$ -entropy, sequential or nonanticipative RDF. We derive lower bound characterizations for a system driven by an i.i.d. Gaussian noise process, which we solve using the SDP algorithm for all three classes of distortion constraints. We obtain closed form solutions when the system's noise is possibly non-Gaussian for both users and when only one of the users is described by a source model driven by a Gaussian noise process. To obtain the upper bounds, we use the best linear forward test channel realization that corresponds to the optimal test channel realization when the system is driven by a Gaussian noise process and apply a sequential causal DPCM-based scheme with a feedback loop followed by a scaled ECDQ scheme that leads to upper bounds with certain performance guarantees. Then, we use the linear forward test channel as a benchmark to obtain upper bounds on the OPTA, when the system is driven by an additive i.i.d. non-Gaussian noise process. We support our framework with various simulation studies.

**Keywords:** bounds; causal coding; one-shot information theory; convex programming; estimation; spatial distortion constraints; temporal distortion constraints; multi-user rate distortion theory

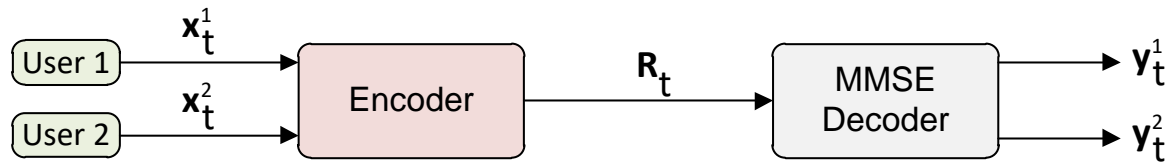
## 1. Problem Statement

We consider the two-user causal encoding and causal decoding setup illustrated in Figure 1. In this setup, users 1 and 2 are modeled by the following discrete-time time-invariant multidimensional Markov processes:

$$\begin{aligned} \mathbf{x}_{t+1}^1 &= A_1 \mathbf{x}_t^1 + \mathbf{w}_t^1, \quad t = 0, 1, \dots \\ \mathbf{x}_{t+1}^2 &= A_2 \mathbf{x}_t^2 + \mathbf{w}_t^2, \end{aligned} \quad (1)$$

where  $\mathbf{x}_t^1 \in \mathbb{R}^{p_1}$ ,  $\mathbf{x}_t^2 \in \mathbb{R}^{p_2}$ , with  $p_1$  not necessarily equal to  $p_2$ ,  $(A_1, A_2)$  are known constant matrices of appropriate dimensions and  $(\mathbf{w}_t^1, \mathbf{w}_t^2)$  are additive i.i.d. possibly non-Gaussian noise processes with zero mean and covariance matrix  $\Sigma_{\mathbf{w}^i} \succ 0, i = 1, 2$ , independent of  $\mathbf{x}_0^i, i = 1, 2$  and from each other for all  $t \geq 0$ . The initial states  $\mathbf{x}_0^i, i = 1, 2$  are given by  $\mathbf{x}_0^i \sim (0; \Sigma_{\mathbf{x}_0^i}), i = 1, 2$ . Finally, we restrict the

eigenvalues of  $(A_1, A_2)$  to be within the unit circle, which means that each user's system model in (1) is asymptotically stable (i.e., asymptotically stationary).



**Figure 1.** System model. The encoder receives information from two users that do not interact from a dynamical system perspective, but they are allowed to allocate bits between them and across the dimensions. The compression is done causally whereas the clocks of the encoder and the decoder are assumed to be synchronized.

The goal is to cast the performance of the setup in Figure 1 under various distortion metrics when the encoder compress information causally whereas the lossless compression between the encoder and the decoder is done in one shot assuming their clocks are synchronized.

First, we apply state space augmentation [1] to the state-space models in (1) to transform them into a single augmented state-space model as follows:

$$\mathbf{x}_{t+1} = A\mathbf{x}_t + \mathbf{w}_t, \quad (2)$$

where  $\mathbf{x}_{t+1} = [\mathbf{x}_{t+1}^1, \mathbf{x}_{t+1}^2]^T \in \mathbb{R}^{p_1+p_2}$ ,  $A$  is a block diagonal matrix and  $\mathbf{w}_t$  is an additive i.i.d. possibly non-Gaussian noise process such that  $\mathbf{w}_t \sim (0; \Sigma_{\mathbf{w}})$  where  $(A, \Sigma_{\mathbf{w}})$  are of the form

$$\begin{bmatrix} A_1 & \mathbf{0} \\ \mathbf{0} & A_2 \end{bmatrix} \in \mathbb{R}^{(p_1+p_2) \times (p_1+p_2)}, \quad \begin{bmatrix} \Sigma_{\mathbf{w}^1} & \mathbf{0} \\ \mathbf{0} & \Sigma_{\mathbf{w}^2} \end{bmatrix} \in \mathbb{R}^{(p_1+p_2) \times (p_1+p_2)}. \quad (3)$$

We note that the operation in (3) can be mathematically denoted as  $A = A_1 \oplus A_2$  and similarly,  $\Sigma_{\mathbf{w}} = \Sigma_{\mathbf{w}^1} \oplus \Sigma_{\mathbf{w}^2}$  (see the notation section for “ $\oplus$ ”).

**System Operation:** The encoder at each time instant  $t$  observes the augmented state  $\mathbf{x}_t$  and generates the data packet  $\ell_t \in \{1, \dots, 2^{R_t}\}$  of instantaneous rate  $R_t$ . At time  $t$ , the packet  $\ell_t$  is sent over a *noiseless channel* with rate  $R_t$ . The decoder at each time  $t$ , receives  $\ell_t$  to construct an estimate  $\mathbf{y}_t$  of  $\mathbf{x}_t$ . We assume that the clocks of the encoder and decoder are synchronized. Formally, the encoder ( $\mathcal{E}$ ) and the decoder ( $\mathcal{D}$ ) are specified by a sequence of measurable functions  $\{(f_t, g_t) : t \in \mathbb{N}_0\}$  as follows:

$$\begin{aligned} \mathcal{E} : \ell_t &= f_t(\ell^{t-1}, \mathbf{x}_t), \quad \ell_{-\infty}^{-1} = \emptyset, \\ \mathcal{D} : \mathbf{y}_t &= g_t(\ell^t). \end{aligned} \quad (4)$$

### 1.1. Generalizations

It should be noted that the setup in Figure 1 can be generalized to any finite number of users. The only change will appear in the number of state-space equations and the dimension of the vectors and matrices in the augmented state-space representation of (2).

Next, we explain the setup of two users that are correlated (in states). In such scenario, users 1 and 2 are modeled by the following discrete-time time-invariant multidimensional Markov processes:

$$\begin{aligned} \mathbf{x}_{t+1}^1 &= A_{11}\mathbf{x}_t^1 + A_{12}\mathbf{x}_t^2 + \mathbf{w}_t^1, \quad t = 0, 1, \dots \\ \mathbf{x}_{t+1}^2 &= A_{22}\mathbf{x}_t^2 + A_{21}\mathbf{x}_t^1 + \mathbf{w}_t^2, \end{aligned} \quad (5)$$

where  $\mathbf{x}_t^1 \in \mathbb{R}^{p_1}, \mathbf{x}_t^2 \in \mathbb{R}^{p_2}$ , with  $p_1$  not necessarily equal to  $p_2$ ,  $(A_{11}, A_{12}, A_{21}, A_{22})$  are known constant matrices of appropriate dimensions whereas all the other assumptions remain similar to the user models described in (1). The single augmented state-space model now is obtained as follows:

$$\mathbf{x}_{t+1} = \hat{A}\mathbf{x}_t + \mathbf{w}_t, \quad (6)$$

where  $\hat{A}$  is a block matrix of the form

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \in \mathbb{R}^{(p_1+p_2) \times (p_1+p_2)}, \quad (7)$$

where  $A_{11}, A_{22}$  are square matrices but  $(A_{12}, A_{21})$  may be rectangular matrices (if  $p_1 \neq p_2$ ). We will not consider this case in our paper because it is straightforward by replacing everywhere matrix  $A$  with matrix  $\hat{A}$ . Clearly, this case can be generalized to any finite number of users with appropriate modifications on the state-space models.

### 1.2. Distortion Constraints

In this work we consider three types of distortion constraints. These are articulated as follows:

- (i) a per-dimension (spatial) distortion constraint on the asymptotically averaged total (across the time) MMSE covariance matrix;
- (ii) an asymptotically averaged total (across the time and across the space) distortion constraint;
- (iii) a covariance matrix distortion constraint.

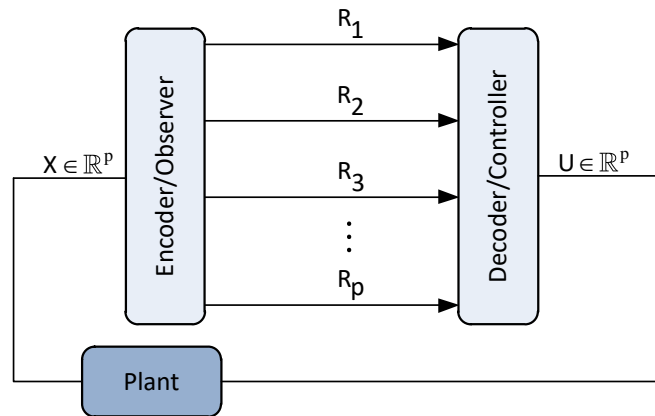
Next, we give the definition of each distortion constraint and explain some of their utility in multi-user systems.

A per-dimension (spatial) distortion constraint imposed on the covariance distortion matrix  $\Sigma_\Delta \triangleq \limsup_{n \rightarrow \infty} \frac{1}{n+1} \sum_{t=0}^n \mathbb{E} \{(\mathbf{x}_t - \mathbf{y}_t)(\mathbf{x}_t - \mathbf{y}_t)^T\}$ , where  $\Sigma_\Delta \succeq 0$ , is defined as follows:

$$\Sigma_{\Delta ii} \leq D_{ii}, \quad i = 1, \dots, p, \quad (8)$$

where  $D_{ii} \in [0, D_{ii}^{\max}]$  are given diagonal entries of the positive semidefinite matrix  $\hat{D} \succeq 0$ , with  $\text{trace}(\hat{D}) \equiv D$ ,  $D \in [0, D_{\max}]$ . Note that under this distortion constraint, it trivially holds that  $\limsup_{n \rightarrow \infty} \frac{1}{n+1} \sum_{t=0}^n \mathbb{E} \{ \|\mathbf{x}_t - \mathbf{y}_t\|_2^2 \} \leq D$ .

**Utility:** The choice of per-dimension distortion constraints is arguably more realistic in various network systems. For instance, one use of such hard constraints can be found in multivariable feedback control systems also called centralized multi-input multi-output (MIMO) systems [2] (see Figure 2). In such networks, it may be the case that one wishes to minimize the temporally total performance criterion or satisfy some total fidelity constraint. However, in addition it is always required that the resource allocation to the different nodes (or variables) to never exceed certain performance thresholds when the demands in data transmission within the communication link allows only limited rate. Nonetheless, the problem is that variables interact. Some variables could be considered more important for certain applications according to the demands of the system or the quality of service, which is why they need hard constraints.



**Figure 2.** Centralized multivariable multi-input multi-output (MIMO) control system.

An asymptotically averaged total (across the time and space) distortion constraint is defined as follows:

$$\limsup_{n \rightarrow \infty} \frac{1}{n+1} \sum_{t=0}^n \mathbb{E} \left\{ \|\mathbf{x}_t - \mathbf{y}_t\|_2^2 \right\} \leq D_{\tau}, \quad (9)$$

where  $D_{\tau} \in [0, D_{\tau}^{\max}]$ , with  $D_{\tau}$  not necessarily equal to  $D$ .

**Utility:** The asymptotically averaged total (across the time and space) distortion constraint ensure shared or allocated distortion arbitrarily among the transmit dimensions. The combination of the per-dimension distortion constraint with the averaged total distortion constraints ensure a total allocated distortion budget in the system that depends on the allowable (by design) distortion budget at each dimension (or user).

A covariance matrix distortion constraint is a generalization of the per-dimension distortion constraint defined by

$$\Sigma_{\Delta} \preceq D_{\text{cov}}, \quad (10)$$

where  $D_{\text{cov}} \succeq 0$ .

**Utility:** During the recent years, there has been a shift from conventional MSE distortion constraints (scalar-valued target distortions) to covariance matrix distortions in the areas of multiterminal and distributed source coding [3–7] and signal processing [7–9]. Nonetheless, the argument for considering covariance distortion constraints despite its difficulty is its generality and the flexibility in formulating new problems. For instance, one practical example would be wireless adhoc microphones, that transmit to receiver(s) over an MIMO channel. In such setups, perhaps the receiver(s) need to do beam forming or some multi-channel Wiener filtering variants. In both cases, one needs to know the covariance matrix of e.g., the error signal (covariance distortion matrix) to perform the desired signal enhancement. For example, if the quality of one of the signals is too bad, this could harm the overall signal enhancement, and one therefore need to trade-off the bits correctly among the microphones. In an adhoc microphone array, the different signals are naturally correlated, which adds an interesting interplay between them that goes beyond MSE distortion.

### 1.3. Operational Interpretations

In this subsection, we use the three types of distortion constraints introduced in Section 1.2 to define the corresponding operational definitions for which we study lower and upper bounds in this paper.

**Definition 1** (Causal RDF subject to (8)). *The operational causal RDF under per-dimension distortion constraints is defined as follows:*

$$R_{\text{pd}}^c(D) \triangleq \inf_{\substack{(f_t, g_t): t \in \mathbb{N}_0 \\ \Sigma_{\Delta_{ii}} \leq D_{ii}, i=1, \dots, p}} \limsup_{n \rightarrow \infty} \frac{1}{n+1} \sum_{t=0}^n R_t, \quad (11)$$

where  $D_{ii} \in [0, D_{ii, \max}]$  and  $D \in [0, D_{\max}]$ .

**Definition 2** (Causal RDF subject to (8) and (9)). *The operational causal RDF under joint per-dimension and asymptotically averaged total distortion constraints is defined as follows:*

$$R_{\text{joint}}^c(D^*) \triangleq \inf_{\substack{(f_t, g_t): t \in \mathbb{N}_0 \\ \Sigma_{\Delta_{ii}} \leq D_{ii}, i=1, \dots, p, \\ \limsup_{n \rightarrow \infty} \frac{1}{n+1} \sum_{t=0}^n \mathbb{E}\{\|\mathbf{x}_t - \mathbf{y}_t\|_2^2\} \leq D^*}} \limsup_{n \rightarrow \infty} \frac{1}{n+1} \sum_{t=0}^n R_t, \quad (12)$$

where  $D^* = \min\{D_{\tau}, D\}$ .

**Interplay between Definitions 1 and 2.** Clearly, Definition 1 is a lower bound to Definition 2 because its constraint set of feasible solutions is larger. Note that, in general, the asymptotically averaged total distortion constraint in (12) is active when  $D_{\tau} \leq D$ , otherwise, it is a trivial constraint and (12) is equivalent to the optimization problem of (11). This observation will be shown via a simulation study in the sequel of the paper.

**Definition 3** (Causal RDF subject to (10)). *The operational causal RDF under covariance matrix distortion constraints is defined as follows:*

$$R^c(D_{\text{cov}}) \triangleq \inf_{\substack{(f_t, g_t): t \in \mathbb{N}_0 \\ \Sigma_{\Delta} \preceq D_{\text{cov}}}} \limsup_{n \rightarrow \infty} \frac{1}{n+1} \sum_{t=0}^n R_t, \quad (13)$$

where  $D_{\text{cov}} \succeq 0$ .

**Literature Review.** In information theory, causal coding and causal decoding also termed zero-delay coding (see, e.g., [10–14]) does not rely on the traditional construction based on random codebooks that in turn allows asymptotically large source vector dimensions [15] to establish achievability of a certain (non-causal) rate-distortion performance. Indeed, the optimal rate-distortion performance for causal source coding (with the clocks of the encoder and decoder to be synchronized), is hard to compute and often bounds are derived in the literature. For example, lower and upper bounds on the operational causal RDF subject to solely the distortion constraint in (9) (or the more stringent per instant distortion constraint  $\mathbb{E}\{\|\mathbf{x}_t - \mathbf{y}_t\|_2^2\} \geq D_t, \forall t$ ) are already studied extensively for various special cases of the setup of Figure 1, see, e.g., [11,14,16,17] and the references therein. In this work, we study new problems related to the causal RDF for the general multi-user source coding setup of Figure 1 under various classes of distortion constraints that their utility (partly explained in Section 1.2) has not been studied in the literature so far. These bounds are established using tools from information theory, convex optimization and causal MMSE estimation.

#### 1.4. Contributions

In this paper we obtain the following results for the setup of Figure 1.

- Characterization and computation of the true lower bounds on (11)–(13) when the users' augmented source model is driven by a Gaussian noise process (Lemma 3, Theorem 2).

- Analytical lower bounds on (11)–(13) when the users' augmented source model is driven by additive i.i.d. noise process (including both additive Gaussian and non-Gaussian noise) (Theorem 3). As a consequence, we also obtain analytical lower bounds when only one of the users' source model is driven by a Gaussian noise process (Corollary 2).
- Characterizations and computation of achievable bounds on (11)–(13) when the users' augmented source model is driven by a Gaussian noise process (Theorem 4).
- Characterizations of achievable bounds on (11)–(13), when the users' augmented source model is driven by additive i.i.d. non-Gaussian noise process (Theorem 5).

**Machinery and tools.** The information theoretic rate distortion definitions that are used to obtain the lower bounds in this paper are derived using a data processing theorem (Theorem 1) that reveals the “suitable” information measure to use. The derivation of the steady-state characterization of the lower bounds in Lemma 3 is derived using inequalities from matrix algebra and a convexity argument that allows the use of Jensen's inequality. To derive lower bounds beyond additive i.i.d. Gaussian noise process, we use the fact that the characterizations of the lower bounds for the Gaussian case are in fact the characterizations obtained for the best linear coding policies (Lemma 4) hence these can serve as a benchmark to derive lower bounds beyond Gaussian noise process by leveraging certain trace/determinant inequalities and most importantly Minkowski's determinant inequality ([18], Exercise 12.13) and EPI [19]. The upper bounds on the OPTA when the system's noise is Gaussian, are derived using a causal sequential DPCM-based scheme with feedback loop which is equivalent to the scheme first derived in [14], followed by an ECDQ scheme that uses vector quantization. The upper bounds on the OPTA, when the system's noise is additive i.i.d. non-Gaussian are obtained using precisely the same trick that is used to obtain the lower bounds, i.e., we use the linear test channel realization that achieves similar upper bounds for the Gaussian case and then, using an SLB type concept (Theorem 5) we obtain the desired results.

The paper is structured as follows. In Section 2 we characterize and compute lower bounds on the OPTA of (11)–(13). In Section 3 we characterize and compute achievable coding scheme on the OPTA of (11)–(13). We draw conclusions and future directions in Section 4.

## 2. Lower Bounds

In this section, we first choose a suitable information measure that will be used to derive a lower bound on Definitions 1–3. This information measure is a variant of directed information subject to some conditional independence constraints. Then, we obtain lower bounds on Definitions 1–3 for jointly Gaussian Markov processes and for Markov processes driven by additive i.i.d. possibly non-Gaussian noise process.

First, we write the joint distribution of the communication system of Figure 1, i.e., from the two users described by the augmented state  $\{\mathbf{x}_t : t \in \mathbb{N}_0^n\}$  to the augmented output of the MMSE decoder  $\{\mathbf{y}_t : t \in \mathbb{N}_0^n\}$ . In particular, the joint distribution induced by the joint process  $\{(\mathbf{x}_t, \ell_t, \mathbf{y}_t) : t \in \mathbb{N}_0^n\}$  admits the following decomposition:

$$\begin{aligned} \mathbf{P}(d\mathbf{x}^n, d\ell^n, d\mathbf{y}^n) &= \otimes_{t=0}^n \mathbf{P}(d\mathbf{y}_t, d\ell_t, d\mathbf{x}_t | \mathbf{y}^{t-1}, \ell^{t-1}, \mathbf{x}^{t-1}) \\ &= \otimes_{t=0}^n \mathbf{P}(d\mathbf{y}_t | \mathbf{y}^{t-1}, \mathbf{x}^t, \ell^t) \otimes \mathbf{P}(d\ell_t | \ell^{t-1}, \mathbf{y}^{t-1}, \mathbf{x}^t) \otimes \mathbf{P}(d\mathbf{x}_t | \mathbf{x}^{t-1}, \mathbf{y}^{t-1}, \ell^{t-1}) - a.s. \\ &= \otimes_{t=0}^n \mathbf{P}(d\mathbf{y}_t | \mathbf{y}^{t-1}, \ell^t) \otimes \mathbf{P}(d\ell_t | \ell^{t-1}, \mathbf{y}^{t-1}, \mathbf{x}^t) \otimes \mathbf{P}(d\mathbf{x}_t | \mathbf{x}^{t-1}) - a.s., \end{aligned} \quad (14)$$

which means that the augmented state “source” process  $\mathbf{x}_t$ , and the decoder's output process  $\mathbf{y}_t$  satisfy the following conditional independence constraints:

$$\mathbf{P}(d\mathbf{x}_t | \mathbf{x}^{t-1}, \mathbf{y}^{t-1}, \ell^{t-1}) = \mathbf{P}(d\mathbf{x}_t | \mathbf{x}^{t-1}) - a.s., \quad (15)$$

$$\mathbf{P}(d\mathbf{y}_t | \mathbf{y}^{t-1}, \ell^t, \mathbf{x}^t) = \mathbf{P}(d\mathbf{y}_t | \mathbf{y}^{t-1}, \ell^t) - a.s. \quad (16)$$

For (14) we state the following technical remark.

**Remark 1** (Trivial initial information). In (14) we assume that the joint distribution  $\mathbf{P}(dx^{-1}, d\ell^{-1}, dy^{-1})$  generates trivial information. This means that  $\mathbf{P}(dx_0|x^{-1}, y^{-1}, \ell^{-1}) = \mathbf{P}(dx_0)$ ,  $\mathbf{P}(dy_0|y^{-1}, x^0, \ell^0) = \mathbf{P}(dy_0|x_0, \ell_0)$  and  $\mathbf{P}(d\ell_0|\ell^{-1}, y^{-1}, x^0) = \mathbf{P}(d\ell_0|x_0)$ .

We next prove a data processing theorem.

**Theorem 1** (Data processing theorem). Provided the decomposition of the joint distribution in (14) holds, the augmented state-space representation of the system in Figure 1 admits the following data processing inequalities:

$$I(\mathbf{x}^n; \mathbf{y}^n) \stackrel{\text{(ii)}}{\leq} I(\mathbf{x}^n; \ell^n | \mathbf{y}^{n-1}) \stackrel{\text{(i)}}{\leq} \sum_{t=0}^n R_t, \quad (17)$$

where

$$\begin{aligned} I(\mathbf{x}^n; \ell^n | \mathbf{y}^{n-1}) &\triangleq \sum_{t=0}^n I(\mathbf{x}^t; \ell_t | \ell^{t-1}, \mathbf{y}^{t-1}), \\ I(\mathbf{x}^n; \mathbf{y}^n) &\triangleq \sum_{t=0}^n I(\mathbf{x}^t; \mathbf{y}_t | \mathbf{y}^{t-1}), \end{aligned} \quad (18)$$

assuming  $I(\mathbf{x}^t; \ell_t | \ell^{t-1}, \mathbf{y}^{t-1}) < \infty, \forall t$ , and  $I(\mathbf{x}^t; \mathbf{y}_t | \mathbf{y}^{t-1}) < \infty, \forall t$ .

**Proof.** We first prove (i).

$$\begin{aligned} \sum_{t=0}^n R_t &\geq \sum_{t=0}^n H(\ell_t | \ell^{t-1}) \\ &\stackrel{(a)}{\geq} \sum_{t=0}^n H(\ell_t | \ell^{t-1}, \mathbf{y}^{t-1}) \\ &\stackrel{(b)}{\geq} \sum_{t=0}^n \left[ H(\ell_t | \ell^{t-1}, \mathbf{y}^{t-1}) - H(\ell_t | \ell^{t-1}, \mathbf{y}^{t-1}, \mathbf{x}^t) \right] \\ &\stackrel{(c)}{=} \sum_{t=0}^n I(\mathbf{x}^t; \ell_t | \ell^{t-1}, \mathbf{y}^{t-1}) \\ &\equiv I(\mathbf{x}^n; \ell^n | \mathbf{y}^{n-1}), \end{aligned}$$

where (a) follows because conditioning reduces entropy [19]; (b) follows because of the non-negativity of the discrete entropy [19]; (c) follows by definition.

Next, we prove (ii). This can be shown as follows:

$$\begin{aligned} I(\mathbf{x}^t; \ell_t | \ell^{t-1}, \mathbf{y}^{t-1}) - I(\mathbf{x}^t; \mathbf{y}_t | \mathbf{y}^{t-1}) &\stackrel{(d)}{=} I(\mathbf{x}^t; \ell_t, \mathbf{y}_t | \ell^{t-1}, \mathbf{y}^{t-1}) - I(\mathbf{x}^t; \mathbf{y}_t | \mathbf{y}^{t-1}) \\ &\stackrel{(e)}{=} I(\mathbf{x}^t; \ell^t | \mathbf{y}^t) - I(\mathbf{x}^t; \ell^{t-1} | \mathbf{y}^{t-1}), \\ &\stackrel{(f)}{=} I(\mathbf{x}^t; \ell^t | \mathbf{y}^t) - I(\mathbf{x}^{t-1}; \ell^{t-1} | \mathbf{y}^{t-1}), \end{aligned} \quad (19)$$

where (d) follows from an adaptation of ([20], Lemma 3.3) to processes, i.e.,  $I(\mathbf{x}^t; \ell_t, \mathbf{y}_t | \ell^{t-1}, \mathbf{y}^{t-1}) = I(\mathbf{x}^t; \ell_t | \ell^{t-1}, \mathbf{y}^{t-1}) + I(\mathbf{x}^t; \mathbf{y}_t | \ell^t, \mathbf{y}^{t-1})$  and the second term is zero because of the conditional independence constraint (16); (e) follows by the chain rule of conditional mutual information (again an adaptation of ([20], Lemma 3.3)) which decomposes the conditional mutual information in two different ways, i.e.,

$$\begin{aligned} I(\mathbf{x}^t; \ell^t, \mathbf{y}_t | \mathbf{y}^{t-1}) &= I(\mathbf{x}^t; \ell^{t-1} | \mathbf{y}^{t-1}) + I(\mathbf{x}^t; \ell_t, \mathbf{y}_t | \ell^{t-1}, \mathbf{y}^{t-1}) \\ &= I(\mathbf{x}^t; \ell^t | \mathbf{y}^t) + I(\mathbf{x}^t; \mathbf{y}_t | \mathbf{y}^{t-1}); \end{aligned}$$

(f) follows because an adaptation of ([20], Lemma 3.3) can be applied to  $I(\mathbf{x}^t; \ell^{t-1} | \mathbf{y}^{t-1})$  as follows

$$\begin{aligned} I(\mathbf{x}^t; \ell^{t-1} | \mathbf{y}^{t-1}) &= I(\mathbf{x}_t, \mathbf{x}^{t-1}; \ell^{t-1} | \mathbf{y}^{t-1}) \\ &= I(\mathbf{x}_t; \ell^{t-1} | \mathbf{x}^{t-1}, \mathbf{y}^{t-1}) + I(\mathbf{x}^{t-1}; \ell^{t-1} | \mathbf{y}^{t-1}) \\ &\stackrel{(g)}{=} I(\mathbf{x}^{t-1}; \ell^{t-1} | \mathbf{y}^{t-1}), \quad \forall t, \end{aligned}$$

where (g) follows because  $I(\mathbf{x}_t; \ell^{t-1} | \mathbf{x}^{t-1}, \mathbf{y}^{t-1}) = 0$ . This can be shown as follows.

$$\begin{aligned} I(\mathbf{x}_t; \ell^{t-1} | \mathbf{x}^{t-1}, \mathbf{y}^{t-1}) &= h(\mathbf{x}_t | \mathbf{x}^{t-1}, \mathbf{y}^{t-1}) - h(\mathbf{x}_t | \mathbf{x}^{t-1}, \mathbf{y}^{t-1}, \ell^{t-1}) \\ &= h(\mathbf{x}_t | \mathbf{x}^{t-1}, \mathbf{y}^{t-1}) - h(\mathbf{x}_t | \mathbf{x}^{t-1}) \leq 0, \quad \forall t, \end{aligned} \quad (20)$$

where each  $h(\cdot)$  is assumed to be finite for any  $t$ , and (20) follows from the conditional independence constraint in (15). From the non-negativity of conditional mutual information [19], the result follows.

Finally, from (19) we have

$$\begin{aligned} \sum_{t=0}^n \left[ I(\mathbf{x}^t; \ell^t | \mathbf{y}^t) - I(\mathbf{x}^{t-1}; \ell^{t-1} | \mathbf{y}^{t-1}) \right] &= I(\mathbf{x}_0; \ell_0 | \mathbf{y}_0) + I(\mathbf{x}^1; \ell^1 | \mathbf{y}^1) - I(\mathbf{x}_0; \ell_0 | \mathbf{y}_0) + \dots \\ &\quad \dots + I(\mathbf{x}^n; \ell^n | \mathbf{y}^n) - I(\mathbf{x}^{n-1}; \ell^{n-1} | \mathbf{y}^{n-1}) \end{aligned} \quad (21)$$

$$= I(\mathbf{x}^n; \ell^n | \mathbf{y}^n) \geq 0, \quad (22)$$

where (22) follows by applying the method of differences in (21). The result follows because (22) is by definition non-negative. We note that if  $I(\mathbf{x}_0; \ell_0 | \mathbf{y}_0)$  also appeared in the cancellations, then, this would have been the telescopic sum of the series. This completes the derivation.  $\square$

We note that Theorem 1 is different from the data processing theorem derived in ([21], Lemma 1) in that we assume the conditional independence constraint (15) instead of the conditional independence constraint  $\mathbf{P}(dx_t | x^{t-1}, y^{t-1}, \ell^{t-1}) = \mathbf{P}(dx_t | x^{t-1}, y^{t-1})$  a.s., i.e., the source process is not allowed to have access via feedback to the previous output symbols  $\mathbf{y}^{t-1}$ . This technical difference results into having the mutual information in (18) subject to conditional independence constraints, instead of the well-known directed information [22].

Before we introduce the information theoretic definitions that correspond to lower bounds on (11)–(13), we formally show the construction of  $I(\mathbf{x}^n; \mathbf{y}^n)$ .

**Source.** The augmented source process  $\{\mathbf{x}_t : t \in \mathbb{N}_0\}$  induces the sequence of conditional distributions  $\{\mathbf{P}(dx_t | x^{t-1}), t \in \mathbb{N}_0^n\}$ . Since no initial information is assumed, the distribution at  $t = 0$  is  $\mathbf{P}(dx_0)$ . In addition, by Bayes' rule we obtain  $\mathbf{P}(dx^n) \triangleq \otimes_{t=0}^n \mathbf{P}(dx_t | x^{t-1})$ .

**Reproduction or "test-channel".** The reproduction process  $\{\mathbf{y}_t : t \in \mathbb{N}_0^n\}$  parameterized by  $\mathcal{X}^t$  induces the sequence of conditional distributions, known as test-channels, by  $\{\mathbf{P}(dy_t | y^{t-1}, x^t), t \in \mathbb{N}_0^n\}$ . At  $t = 0$ , no initial state information is assumed, hence  $\mathbf{P}(dy_0 | y^{-1}, x^0) = \mathbf{P}(dy_0 | x_0)$ . In addition, by Bayes' rule we obtain  $\vec{\mathbf{Q}}(dy^n | x^n) \triangleq \otimes_{t=0}^n \mathbf{P}(dy_t | y^{t-1}, x^t)$ .

From ([23], Remark 1), it can be shown that the sequence of conditional distributions  $\{\mathbf{P}(dx_t | x^{t-1}) : t \in \mathbb{N}_0^n\}$  and  $\{\mathbf{P}(dy_t | y^{t-1}, x^t) : t \in \mathbb{N}_0^n\}$  uniquely define the family of conditional distributions on  $\mathcal{X}^n$  and  $\mathcal{Y}^n$  parameterized by  $x^n \in \mathcal{X}^n$ , respectively, given by the joint distribution

$$\mathbf{P}(dx^n, dy^n) = \mathbf{P}(dx^n) \otimes \vec{\mathbf{Q}}(dy^n | x^n). \quad (23)$$

In addition, from (23), we can uniquely define the  $\mathcal{Y}^n$ -marginal distribution by

$$\mathbf{P}(dy^n) \triangleq \int_{\mathcal{X}^n} \mathbf{P}(dx^n) \otimes \vec{\mathbf{Q}}(dy^n | x^n),$$

and the conditional distributions  $\{\mathbf{P}(dy_t | y^{t-1}) : t \in \mathbb{N}_0^n\}$ .

Given the above construction of distributions, we can formally introduce the information measure using relative entropy as follows:

$$\begin{aligned}
 I(\mathbf{x}^n; \mathbf{y}^n) &\stackrel{(a)}{=} \mathbb{D}(\mathbf{P}(dx^n, dy^n) || \mathbf{P}(dx^n) \times \mathbf{P}(dy^n)) \in [0, \infty] \\
 &\stackrel{(b)}{=} \int_{\mathcal{X}^n \times \mathcal{Y}^n} \log \left( \frac{d\vec{\mathbf{Q}}(\cdot | \mathbf{x}^n)}{d\mathbf{P}(\cdot)}(\mathbf{y}^n) \right) \mathbf{P}(dx^n, dy^n) \\
 &\stackrel{(c)}{=} \sum_{t=0}^n \mathbb{E} \left\{ \log \left( \frac{d\mathbf{P}(\cdot | \mathbf{y}^{t-1}, \mathbf{x}^t)}{d\mathbf{P}(\cdot | \mathbf{y}^{t-1})}(\mathbf{y}_t) \right) \right\} \\
 &\stackrel{(d)}{=} \sum_{t=0}^n I(\mathbf{x}^t; \mathbf{y}_t | \mathbf{y}^{t-1}),
 \end{aligned} \tag{24}$$

where (a) follows by definition of relative entropy between  $\mathbf{P}(dx^n, dy^n)$  and the product distribution  $\mathbf{P}(dx^n) \times \mathbf{P}(dy^n)$ ; (b) is due to the Radon–Nikodym derivative theorem ([23], Appendixes A and C); (c) is due to chain rule of relative entropy; (d) follows by definition.

We now state as a definition the lower bounds on (11)–(13).

**Definition 4** (Lower bounds). *Using the previous construction of distributions and the information measure of (24), we can define the following lower bounds on (11)–(13).*

(1) *The sum-rate subject to per-dimension distortion constraint is defined as follows:*

$$R_{\text{pd}}^{\text{LB}}(D) \triangleq \inf_{\substack{\mathbf{P}(dy_t | y^{t-1}, x^t): t=0, \dots, \infty \\ \Sigma_{\Delta_{ii}} \leq D_{ii}, i=1, \dots, p}} \limsup_{n \rightarrow \infty} \frac{1}{n+1} I(\mathbf{x}^n; \mathbf{y}^n), \tag{25}$$

with

$$R_{\text{pd}, [0, n]}^{\text{LB}}(D) \triangleq \inf_{\substack{\mathbf{P}(dy_t | y^{t-1}, x^t): t \in \mathbb{N}_0^n \\ \Sigma_{\Delta_{ii}, t} \leq D_{ii}, i=1, \dots, p}} I(\mathbf{x}^n; \mathbf{y}^n). \tag{26}$$

where  $\Sigma_{\Delta, t} \triangleq \frac{1}{n+1} \sum_{t=0}^n \mathbb{E} \{ (\mathbf{x}_t - \mathbf{y}_t)(\mathbf{x}_t - \mathbf{y}_t)^T \}$ ,  $\Sigma_{\Delta_{ii}, t} \triangleq \frac{1}{n+1} \sum_{t=0}^n [\mathbb{E} \{ (\mathbf{x}_t - \mathbf{y}_t)(\mathbf{x}_t - \mathbf{y}_t)^T \}]_{ii}$  and  $D \geq 0$ .

(2) *The sum-rate subject to joint distortion constraints is defined as follows:*

$$\begin{aligned}
 R_{\text{joint}}^{\text{LB}}(D^*) &\triangleq \inf_{\substack{\mathbf{P}(dy_t | y^{t-1}, x^t): t=0, \dots, \infty \\ \Sigma_{\Delta_{ii}} \leq D_{ii}, i=1, \dots, p}} \limsup_{n \rightarrow \infty} \frac{1}{n+1} I(\mathbf{x}^n; \mathbf{y}^n), \\
 &\limsup_{n \rightarrow \infty} \frac{1}{n+1} \sum_{t=0}^n \mathbb{E} \{ \|\mathbf{x}_t - \mathbf{y}_t\|_2^2 \} \leq D_T
 \end{aligned} \tag{27}$$

$$\begin{aligned}
 R_{\text{joint}, [0, n]}^{\text{LB}}(D^*) &\triangleq \inf_{\substack{\mathbf{P}(dy_t | y^{t-1}, x^t): t \in \mathbb{N}_0^n \\ \Sigma_{\Delta_{ii}, t} \leq D_{ii}, i=1, \dots, p}} I(\mathbf{x}^n; \mathbf{y}^n), \\
 &\frac{1}{n+1} \sum_{t=0}^n \mathbb{E} \{ \|\mathbf{x}_t - \mathbf{y}_t\|_2^2 \} \leq D_T
 \end{aligned} \tag{28}$$

where  $D^* = \min\{D, D_T\}$ .

(3) *The sum-rate subject to covariance matrix distortion constraints is defined as follows:*

$$R^{\text{LB}}(D_{\text{cov}}) \triangleq \inf_{\substack{\mathbf{P}(dy_t | y^{t-1}, x^t): t=0, \dots, \infty \\ \Sigma_{\Delta} \preceq D_{\text{cov}}}} \limsup_{n \rightarrow \infty} \frac{1}{n+1} I(\mathbf{x}^n; \mathbf{y}^n), \tag{29}$$

$$R_{[0,n]}^{LB}(D_{\text{cov}}) \triangleq \inf_{\substack{\mathbf{P}(dy_t|y^{t-1}, x^t): t \in \mathbb{N}_0^n \\ \Sigma_{\Delta, t} \preceq D_{\text{cov}}}} I(\mathbf{x}^n; \mathbf{y}^n), \quad (30)$$

where  $D_{\text{cov}} \succeq 0$ .

Next, we stress some technical remarks related to the new information theoretic measures in Definition 4 that can be obtained using known results in the literature and some known lower bounds that use the same objective function with (26)–(30).

**Remark 2** (Comments on Definition 4). It can be shown that the infimization problems (26), (28) and (30), in contrast to their operational counterparts (11)–(13) are convex with respect to their test channel. This can be shown following, for instance, the techniques of [23]. By the structural properties of the test channel derived in ([24], Section 4), if the source is first-order Markov, i.e., with distribution  $\mathbf{P}(dx_t|x_{t-1})$ ,  $t \in \mathbb{N}_0^n$ , the test channel distribution is of the form  $\mathbf{P}(dy_t|y^{t-1}, x_t)$ ,  $t \in \mathbb{N}_0^n$ . Finally, combining this structural result, with ([25], Theorem 1.8.6), it can be shown that if  $\mathbf{x}^n$  is Gaussian then a jointly Gaussian process  $\{(\mathbf{x}_t, \mathbf{y}_t) : t \in \mathbb{N}_0\}$  achieves a smaller value of the information rates, and if  $\mathbf{x}^n$  is Gaussian and Markov, then the infimum in (26), (28) and (30) can be restricted to test channel distributions which are Gaussian, of the form  $\mathbf{P}(dy_t|y_{t-1}, x_t)$ .

We recall that when the distortion constraint set contains only (9), its finite time horizon counterpart or the per instant distortion constraint  $\mathbb{E}\{\|\mathbf{x}_t - \mathbf{y}_t\|_2^2\} \leq D_t \forall t$ , we end up having the well known nonanticipatory- $\epsilon$  entropy [26] also found in the literature as sequential or nonanticipative RDF [27,28]. Nonanticipatory- $\epsilon$  entropy received significant interest during the last twenty years in an anthology of papers (see, e.g., [11,16,24,29–31]) due to its utility in control related and delay-constrained applications. Moreover, the characterizations in (29) and (30) do not appear to be manageable to solve using standard techniques, and no closed-form statements are available for the general RDF in the literature. For this reason, we will seek only for non-tight bounds.

In view of the above, in the sequel we characterize and compute lower bounds on Definitions 1–3 for Gauss–Markov processes and for Markov models driven by additive i.i.d. noise processes.

### 2.1. Characterization and Computation of Jointly Gaussian Processes

In this section, we assume that the augmented joint process  $\{(\mathbf{x}_t, \mathbf{y}_t) : t \in \mathbb{N}_0\}$  is jointly Gaussian. We use this assumption to first characterize and then to compute optimally (26), (28) and (30).

We first use the following helpful lemma. We exclude the proof because it is already derived in other papers, see, e.g., [14,24]. The only modification is the augmented joint processes  $\{(\mathbf{x}_t, \mathbf{y}_t) : t \in \mathbb{N}_0^n\}$ .

**Lemma 1** (Realization of  $\{\mathbf{P}^*(dy_t|y^{t-1}, x_t) : t \in \mathbb{N}_0^n\}$ ). Consider the class of conditionally Gaussian test channels  $\{\mathbf{P}^*(dy_t|y^{t-1}, x_t) : t \in \mathbb{N}_0^n\}$ . Then, the following statements hold.

(1) Any candidate of  $\{\mathbf{P}^*(dy_t|y^{t-1}, x_t) : t \in \mathbb{N}_0^n\}$  can be realized by the recursion

$$\mathbf{y}_t = H_t \left( \mathbf{x}_t - \widehat{\mathbf{x}}_{t|t-1} \right) + \widehat{\mathbf{x}}_{t|t-1} + \mathbf{v}_t, \quad \widehat{\mathbf{x}}_{0|-1} = \text{given}, \quad t \in \mathbb{N}_0^n, \quad (31)$$

where  $\widehat{\mathbf{x}}_{t|t-1} \triangleq \mathbb{E}\{\mathbf{x}_t|\mathbf{y}^{t-1}\}$ ,  $\{\mathbf{v}_t \in \mathbb{R}^{p_1+p_2} \sim \mathcal{N}(0; \Sigma_{\mathbf{v}_t}) : t \in \mathbb{N}_0^n\}$  is an independent Gaussian process independent of  $\{\mathbf{w}_t : t \in \mathbb{N}_0^{n-1}\}$  and  $\mathbf{x}_0$ , and  $\{H_t \in \mathbb{R}^{(p_1+p_2) \times (p_1+p_2)} : t \in \mathbb{N}_0^n\}$  are time-varying deterministic matrices.

Moreover, the innovations process  $\{\mathbf{I}_t \in \mathbb{R}^{p_1+p_2} : t \in \mathbb{N}_0^n\}$  of (31) is the orthogonal process defined by

$$\mathbf{I}_t \triangleq \mathbf{y}_t - \mathbb{E}\{\mathbf{y}_t|\mathbf{y}^{t-1}\} = H_t \left( \mathbf{x}_t - \widehat{\mathbf{x}}_{t|t-1} \right) + \mathbf{v}_t,$$

where  $\mathbf{I}_t \sim \mathcal{N}(0; \Sigma_{\mathbf{I}_t})$ ,  $\Sigma_{\mathbf{I}_t} = H_t \Sigma_{t|t-1} H_t^T + \Sigma_{\mathbf{v}_t}$  and  $\Sigma_{t|t-1} \triangleq \mathbb{E} \left\{ (\mathbf{x}_t - \hat{\mathbf{x}}_{t|t-1})(\mathbf{x}_t - \hat{\mathbf{x}}_{t|t-1})^T | \mathbf{y}^{t-1} \right\} = \mathbb{E} \left\{ (\mathbf{x}_t - \hat{\mathbf{x}}_{t|t-1})(\mathbf{x}_t - \hat{\mathbf{x}}_{t|t-1})^T \right\}$ .

- (2) Let  $\hat{\mathbf{x}}_{t|t} \triangleq \mathbb{E} \{ \mathbf{x}_t | \mathbf{y}^t \}$  and  $\Sigma_{t|t} \triangleq \mathbb{E} \left\{ (\mathbf{x}_t - \hat{\mathbf{x}}_{t|t})(\mathbf{x}_t - \hat{\mathbf{x}}_{t|t})^T | \mathbf{y}^t \right\} = \mathbb{E} \left\{ (\mathbf{x}_t - \hat{\mathbf{x}}_{t|t})(\mathbf{x}_t - \hat{\mathbf{x}}_{t|t})^T \right\}$ . Then,  $\{ \hat{\mathbf{x}}_{t|t-1}, \Sigma_{t|t-1} : t \in \mathbb{N}_0^n \}$  satisfy the following vector-valued equations:

$$\begin{aligned} \hat{\mathbf{x}}_{t|t-1} &= A \hat{\mathbf{x}}_{t-1|t-1}, \\ \Sigma_{t|t-1} &= A \Sigma_{t-1|t-1} A^T + \Sigma_{\mathbf{w}_t}, \\ \hat{\mathbf{x}}_{t|t} &= \hat{\mathbf{x}}_{t|t-1} + N_t \mathbf{I}_t, \\ N_t &= \Sigma_{t|t-1} H_t^T \Sigma_{\mathbf{I}_t}^{-1} \text{ (Kalman Gain)}, \\ \Sigma_{t|t} &= \Sigma_{t|t-1} - \Sigma_{t|t-1} H_t^T \Sigma_{\mathbf{I}_t}^{-1} H_t \Sigma_{t|t-1}, \end{aligned} \quad (32)$$

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

- (3) Using MMSE estimation via the vector-valued KF recursions of (32), the following finite dimensional characterizations of  $R_{\text{pd},[0,n]}^{LB,G}(D)$ ,  $R_{\text{joint},[0,n]}^{LB,G}(D^*)$ ,  $R_{\text{cov},[0,n]}^{LB,G}(D_{\text{cov}})$  can be obtained:

$$R_{\text{pd},[0,n]}^{LB,G}(D) = \inf_{\substack{H_t \in \mathbb{R}^{(p_1+p_2) \times (p_1+p_2)}, \Sigma_{\mathbf{v}_t} \succeq 0, t \in \mathbb{N}_0^n \\ 0 \leq \tilde{\Sigma}_{ii,t} \leq D_{ii}, i=1, \dots, p}} \frac{1}{2} \sum_{t=0}^n \left[ \log \frac{|\Sigma_{t|t-1}|}{|\Sigma_{t|t}|} \right]^+, \quad (33)$$

$$R_{\text{joint},[0,n]}^{LB,G}(D^*) = \inf_{\substack{H_t \in \mathbb{R}^{(p_1+p_2) \times (p_1+p_2)}, \Sigma_{\mathbf{v}_t} \succeq 0, t \in \mathbb{N}_0^n \\ 0 \leq \tilde{\Sigma}_{ii,t} \leq D_{ii}, i=1, \dots, p}} \frac{1}{2} \sum_{t=0}^n \left[ \log \frac{|\Sigma_{t|t-1}|}{|\Sigma_{t|t}|} \right]^+, \quad (34)$$

$$\frac{1}{n+1} \sum_{t=0}^n \text{trace}((I_{p_1+p_2} - H_t) \Sigma_{t|t-1} (I_{p_1+p_2} - H_t)^T + \Sigma_{\mathbf{v}_t}) \leq D_T$$

$$R_{[0,n]}^{LB,G}(D_{\text{cov}}) = \inf_{\substack{H_t \in \mathbb{R}^{(p_1+p_2) \times (p_1+p_2)}, \Sigma_{\mathbf{v}_t} \succeq 0, t \in \mathbb{N}_0^n \\ \frac{1}{n+1} \sum_{t=0}^n ((I_{p_1+p_2} - H_t) \Sigma_{t|t-1} (I_{p_1+p_2} - H_t)^T + \Sigma_{\mathbf{v}_t}) \preceq D_{\text{cov}}}} \frac{1}{2} \sum_{t=0}^n \left[ \log \frac{|\Sigma_{t|t-1}|}{|\Sigma_{t|t}|} \right]^+, \quad (35)$$

where  $\tilde{\Sigma}_{ii,t} \triangleq \frac{1}{n+1} \sum_{t=0}^n \left[ (I_{p_1+p_2} - H_t) \Sigma_{t|t-1} (I_{p_1+p_2} - H_t)^T + \Sigma_{\mathbf{v}_t} \right]_{ii} \geq 0$ ,  $D \in [0, D_{\text{max}}]$  and  $D^* \in [0, D_{\text{max}}^*]$ .

We note that one can directly study the finite-dimensional characterizations of Lemma 1, (3), and try to come up with numerical solutions. However, it is much more insightful to use instead the identification of the design parameters  $\{(H_t, \Sigma_{\mathbf{v}_t}) : t \in \mathbb{N}_0^n\}$  of the test-channel realization in (31). This approach is already done in [14,24] hence we state it without a proof. Note, however, that compared to [14,24] that assume distortion constraints like (9) (or the per time instant counterpart of (9), i.e.,  $\mathbb{E} \{ \|\mathbf{x}_t - \mathbf{y}_t\|_2^2 \} \leq D_t, \forall t$ ), here we assume augmented state-space models and various spatio-temporal distortion constraints, namely, per-dimension, jointly per-dimension and averaged total distortion constraints, and covariance matrix distortion constraint.

**Lemma 2** (Alternative characterizations of (33)–(35) via system identification). Consider Lemma 1 and set  $\Delta_t \triangleq \Sigma_{t|t}$  and  $\Lambda_t \triangleq \Sigma_{t|t-1}$ . Then, the following statements hold.

- (1) The test-channel distribution  $\mathbf{P}(d\mathbf{y}_t | \mathbf{y}^{t-1}, \mathbf{x}_t)$  admits the following linear Markov additive noise realization:

$$\mathbf{y}_t = H_t \mathbf{x}_t + (I_{p_1+p_2} - H_t) A \mathbf{y}_{t-1} + \mathbf{v}_t, \mathbf{y}_{-1} = \text{given}, t \in \mathbb{N}_0^n, \quad (36)$$

where

$$H_t \triangleq I_{p_1+p_2} - \Delta_t \Lambda_t^{-1}, \Sigma_{\mathbf{v}_t} \triangleq \Delta_t H_t^T \succeq 0. \quad (37)$$

- (2) The finite-dimensional characterizations of  $R_{\text{pd},[0,n]}^{LB,G}(D)$ ,  $R_{\text{joint},[0,n]}^{LB,G}(D^*)$ ,  $R_{\text{cov},[0,n]}^{LB,G}(D_{\text{cov}})$  can be simplified to the following alternative characterizations:

$$R_{\text{pd},[0,n]}^{LB,G}(D) = \inf_{0 \leq \Delta_{ii,t} \leq D_{ii}, i=1,\dots,p, t \in \mathbb{N}_0^n} \frac{1}{2} \sum_{t=0}^n \left[ \log \frac{|\Lambda_t|}{|\Delta_t|} \right]^+, \quad (38)$$

$$R_{\text{joint},[0,n]}^{LB,G}(D^*) = \inf_{0 \leq \Delta_{ii,t} \leq D_{ii}, i=1,\dots,p, t \in \mathbb{N}_0^n, \frac{1}{n+1} \sum_{t=0}^n \text{trace}(\Delta_t) \leq D_T} \frac{1}{2} \sum_{t=0}^n \left[ \log \frac{|\Lambda_t|}{|\Delta_t|} \right]^+, \quad (39)$$

$$R_{[0,n]}^{LB,G}(D_{\text{cov}}) = \inf_{\frac{1}{n+1} \sum_{t=0}^n (\Delta_t) \preceq D_{\text{cov}}} \frac{1}{2} \sum_{t=0}^n \left[ \log \frac{|\Lambda_t|}{|\Delta_t|} \right]^+, \quad (40)$$

where  $\Delta_{ii,t}$  is defined precisely as  $\tilde{\Sigma}_{ii,t}$ .

Next, we give some technical remarks related to Lemma 2.

**Remark 3** (Special case and technical remarks).

- (1) Clearly, if in the forward test-channel realization with additive noise, we assume that the block diagonal matrix  $A = \mathbf{0}$  (null matrix), then, we recover the classical forward test-channel realization for vector memoryless Gaussian source subject to a MSE distortion (see, e.g., ([32], Chapter 4.5), ([15], Chapter 9.7)) given by

$$\mathbf{y}_t = H_t \mathbf{x}_t + \mathbf{v}_t, t \in \mathbb{N}_0^n, \quad (41)$$

and the coefficients in (37) give

$$H_t \triangleq I_{p_1+p_2} - \Delta_t \Sigma_{\mathbf{w}}^{-1} \succeq 0, \Sigma_{\mathbf{v}_t} \triangleq \Delta_t H_t^T \succeq 0. \quad (42)$$

Moreover, the characterizations in (38)–(40) change in that  $\Lambda_t = \Sigma_{\mathbf{w}}$ . Clearly, (42) can be seen as reverse-waterfilling design parameters.

- (2) Compared to (1), we note that  $H_t$  in (37) should not be confused with a positive semidefinite matrix defined in the usual quadratic form [33] but instead it can possibly be a non-symmetric matrix which however contains only real non-negative eigenvalues. This observation is important because it means that in general the design variables  $(\Delta_t, \Lambda_t)$  do not commute like in the classical reverse-waterfilling problems for memoryless multivariate Gaussian random variables or in i.i.d. processes (see, e.g., [19,25]).
- (3) For jointly Gaussian processes, the linear forward realization in (36) is the optimal realization among all realizations for this problem because the KF is the optimal causal MMSE estimator. Beyond Gaussian processes, and when the noise is zero-mean, uncorrelated and white (in our setup these properties hold), the optimal realization for Gaussian processes becomes the best linear realization (see, e.g., ([34], §3.2) or ([35], p. 130)) and similarly the corresponding characterizations in (38)–(40) are the best linear characterizations. By saying “best-linear” realization and characterizations, respectively, we mean that there may be non-linear realizations and hence non-linear-based characterizations that outperform the best linear.
- (4) The characterization of (39) is different from the characterization obtained in ([16], Theorem 1, (25e)) that uses weighted distortion constraints. The former optimization problem imposes hard constraints whereas the latter imposes soft constraints via weights. Nonetheless, an interesting open question is whether there

exists a set of weights, which will give the same per dimension distortion when imposed as a weighted total distortion constraint.

- (5) It should also be stressed that the per-dimension constraints on the diagonal entries of  $\Delta_t$  are not the same as having constraints on the eigenvalues of  $\Delta_t$ . This further means that even for this class of distortion constraints, it is still possible to have rate-distortion resource allocation (i.e., a type of reverse-waterfilling optimization).

**Remark 4** (Convexity). The optimization problems in (38) and (39) are convex because the objective function is linear and the constraints are affine and positive semi-definite. Thus, the problem can be solved numerically using convex programming software (see, e.g., [36]) or the more challenging KKT conditions that are first-order necessary conditions for global optimality ([37], Chapter 5.3). The latter, will give certain non-linear matrix Riccati equations that need to be solved in order to construct a reverse-waterfilling algorithm.

**Remark 5** (Existence of Solution). A sufficient condition for existence of a solution with a finite value in (38)–(40) is to consider the strict LMI constraint  $0 \prec \Delta_t \preceq \Lambda_t$  that ensures the objective function is bounded. The strict LMI ensures that  $\Delta_t \succ 0$  which further means that  $D > 0$ ,  $D_\tau > 0$  and  $D_{\text{cov}} \succ 0$ .

In what follows, we derive lower bounds on (11)–(13).

**Lemma 3** (Steady-state lower bounds on (11)–(13)). Suppose that the conditions of Remark 5 hold. Moreover, let  $\Delta \triangleq \frac{1}{n+1} \sum_{t=0}^n \Delta_t$  for some finite  $n$ . Then, the following statements hold.

(1)

$$R_{\text{pd}}^c(D) \geq \min_{\substack{0 \prec \Delta \preceq \Lambda \\ \Delta_{ii} \leq D_{ii}, i=1,\dots,p}} \frac{1}{2} \log \frac{|\Lambda|}{|\Delta|}, \quad (43)$$

where  $\Lambda = A\Delta A^T + \Sigma_w$ .

(2)

$$R_{\text{joint}}^c(D^*) \geq \min_{\substack{0 \prec \Delta \preceq \Lambda \\ \Delta_{ii} \leq D_{ii}, i=1,\dots,p \\ \text{trace}(\Delta) \leq D}} \frac{1}{2} \log \frac{|\Lambda|}{|\Delta|}, \quad (44)$$

for some  $D^* = \min\{D_\tau, D\}$ .

(3)

$$R^c(D_{\text{cov}}) \geq \min_{\substack{0 \prec \Delta \preceq \Lambda \\ 0 \prec \Delta \preceq D_{\text{cov}}}} \frac{1}{2} \log \frac{|\Lambda|}{|\Delta|}. \quad (45)$$

**Proof.** See Appendix A.  $\square$

It should be remarked that instead of the derivation based on a convexity argument in Lemma 3, one can assume that the optimal minimizer  $\mathbf{P}(dy_t|y_{t-1}, x_t)$  that achieves (43)–(45) is time-invariant and the output distribution  $\mathbf{P}(dy_t|y_{t-1})$  is also time-invariant with a unique invariant distribution, see, e.g., ([14], Theorem 3). Moreover, the optimal linear forward test-channel that achieves the lower bounds in (43)–(45) correspond to the time-invariant realization (36), given by

$$\mathbf{y}_t = H\mathbf{x}_t + (I_{p_1+p_2} - H)A\mathbf{y}_{t-1} + \mathbf{v}_t, \quad (46)$$

whereas the corresponding time-invariant scaling coefficients in (37) are as follows

$$H \triangleq I_{p_1+p_2} - \Delta \Lambda^{-1}, \Sigma_{\mathbf{v}} \triangleq \Delta H^T \succeq 0. \quad (47)$$

From Lemma 3, the following corollary can be immediately obtained.

**Corollary 1** (Fixed design variable  $\Delta$ ). *If in Lemma 2 we assume that  $\Delta_t = \Delta, \forall t$ , then we obtain (43)–(45).*

**Proof.** This is immediate from the derivation of Lemma 3.  $\square$

In what follows, we show that the lower bounds in Lemma 3 are semidefinite representable, thus, they can be readily computed.

**Theorem 2** (Computation of the lower bounds in Lemma 3). *Consider the variable  $Q_1 \triangleq \Delta^{-1} - A^T \Sigma_{\mathbf{w}}^{-1} A$ , where  $\Delta \succ 0$ . Then, the following semidefinite programming representations hold.*

- (1) *For some  $D \equiv \text{trace}(\hat{D}) > 0$ , the lower bound in (43), denoted hereinafter by  $R_{\text{pd}}^{\text{LB}}(D)$ , is semidefinite representable as follows:*

$$R_{\text{pd}}^{\text{LB}}(D) = \min_{Q_1 \succ 0} -\frac{1}{2} \log |Q_1| + \frac{1}{2} \log |\Sigma_{\mathbf{w}}|. \quad (48)$$

$$\text{s.t. } 0 \prec \Delta \preceq \Lambda$$

$$\Delta_{ii} \leq D_{ii}, i = 1, \dots, p$$

$$\begin{bmatrix} \Delta - Q_1 & \Delta A^T \\ A\Delta & \Lambda \end{bmatrix} \succeq 0 \quad (49)$$

- (2) *For some  $D^* = \min\{D_\tau, D\} > 0$ , the lower bound in (44), denoted hereinafter by  $R_{\text{joint}}^{\text{LB}}(D^*)$ , is semidefinite representable as follows:*

$$R_{\text{joint}}^{\text{LB}}(D^*) = \min_{Q_1 \succ 0} -\frac{1}{2} \log |Q_1| + \frac{1}{2} \log |\Sigma_{\mathbf{w}}|. \quad (50)$$

$$\text{s.t. } 0 \prec \Delta \preceq \Lambda$$

$$\Delta_{ii} \leq D_{ii}, i = 1, \dots, p$$

$$\text{trace}(\Delta) \leq D_\tau$$

$$\begin{bmatrix} \Delta - Q_1 & \Delta A^T \\ A\Delta & \Lambda \end{bmatrix} \succeq 0$$

- (3) *For some  $D_{\text{cov}} \succ 0$ , the lower bound in (45), denoted hereinafter by  $R^{\text{LB}}(D_{\text{cov}})$ , is semidefinite representable as follows:*

$$R^{\text{LB}}(D_{\text{cov}}) = \min_{Q_1 \succ 0} -\frac{1}{2} \log |Q_1| + \frac{1}{2} \log |\Sigma_{\mathbf{w}}|. \quad (51)$$

$$\text{s.t. } 0 \prec \Delta \preceq \Lambda$$

$$\Delta \preceq D_{\text{cov}}$$

$$\begin{bmatrix} \Delta - Q_1 & \Delta A^T \\ A\Delta & \Lambda \end{bmatrix} \succeq 0$$

**Proof.** See Appendix B.  $\square$

Next, we stress some comments on the semidefinite representation of the lower bounds in Theorem 2.

**Remark 6** (Comments on Theorem 2).

- (1) We note that a similar characterization to the characterizations derived in Theorem 2 (subject to the distortion constraint (9) or the per-instant distortion constraint  $\mathbb{E}\{\|\mathbf{x}_t - \mathbf{y}_t\|_2^2\} \leq D_t, \forall t$ , for a special case of the setup in Figure 1) is recently derived in ([16], Equation (27)). The log-determinant convex optimization problems in Theorem 2 are widely used in systems and control theories because they are able to deal efficiently with LMIs [38].
- (2) Recently, the efficiency of SDP algorithm in solving linear and non-linear optimization problems attracted experts from the field of information theory who noticed its usefulness in solving distributed source coding problems (see, e.g., [3,39]). Such log-determinant problems when solved using the semidefinite programming method are known to have polynomial worst-case complexity (see, e.g., [40]). In addition, for an interior point method such as the SDP approach, the most computationally expensive step is the Cholesky factorization involved in the Newton steps.
- (3) On the other hand, due to its complexity, the SDP approach for high dimensional systems is often time consuming whereas for very large scale systems is occasionally impossible to obtain numerical solutions. Hence, ideally one could preferably consider alternative methods to solve a problem sacrificing for instance the optimality of the SDP algorithm but gaining in scalability and reducing the complexity. The most computationally efficient way to compute such problems and, additionally, to gain some insight from the solution is via the well-known reverse-waterfilling algorithm ([19], Theorem 10.3.3), which is however very hard to construct and compute because one needs to employ and solve complicated KKT conditions [37]. Such effort was recently made for multivariate Gauss-Markov processes under per instant, averaged total and asymptotically averaged total distortion constraints in [24,41].

Next, we perform some numerical illustrations using the semidefinite representations of Theorem 2. We also compare (48) and (50), to the known expression obtained only for the asymptotically averaged total MSE distortion constraint in ([16], Equation (27)). We note that the SDP algorithm for (48)–(51) is implemented using the CVX platform [36].

**Example 1** (Comparison of  $R_{\text{joint}}^{\text{LB}}(D^*)$ ,  $R_{\text{pd}}^{\text{LB}}(D)$  and ([16], Equation (27))). For the system in (1), we assume that user 1 is described by a  $\mathbb{R}^2$ -valued time-invariant Markov source driven by i.i.d. Gaussian noise process with parameters  $(A_1, \Sigma_{\mathbf{w}^1})$ :

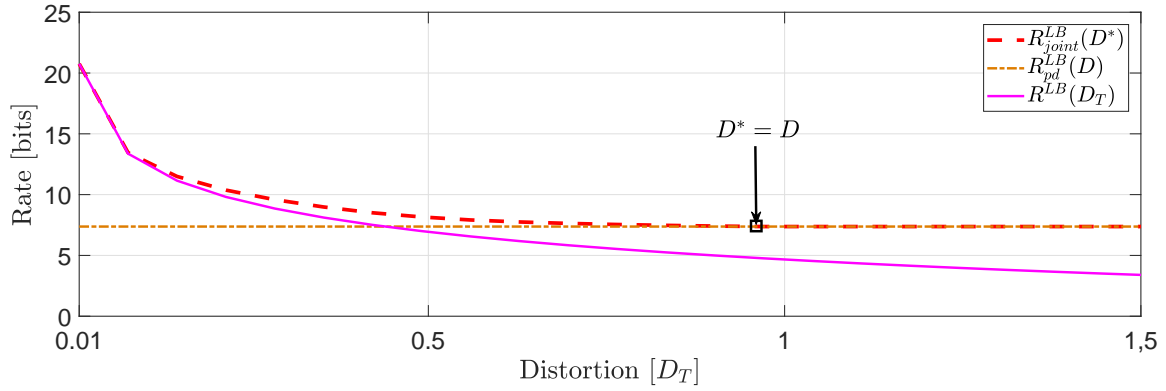
$$(A_1, \Sigma_{\mathbf{w}^1}) = \left( \begin{bmatrix} 0.5 & 0.2 \\ 0.3 & 0.6 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right), \quad (52)$$

whereas, user 2 is described by a  $\mathbb{R}^3$ -valued time-invariant Markov source driven by i.i.d. Gaussian noise process with parameters  $(A_2, \Sigma_{\mathbf{w}^2})$ :

$$(A_2, \Sigma_{\mathbf{w}^2}) = \left( \begin{bmatrix} 0.5 & 0.2 & 0.1 \\ 0.3 & 0.6 & 0.1 \\ 0.7 & 0.3 & 0.4 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0.2 & 0 \\ 0 & 0 & 0.5 \end{bmatrix} \right). \quad (53)$$

Clearly, the augmented state space model (2) generates  $A = A_1 \oplus A_2$  and  $\Sigma_{\mathbf{w}} = \Sigma_{\mathbf{w}^1}^1 \oplus \Sigma_{\mathbf{w}^2}^2$ . For this example, we assume that  $D_{\tau} = 1.5$  and  $D_{11} = 0.1, D_{22} = 0.01, D_{33} = 0.6, D_{44} = 0.15, D_{55} = 0.1$ , which implies that  $D = 0.96$ . This means that  $D^* = \min\{D_{\tau}, D\} = 0.96$ .

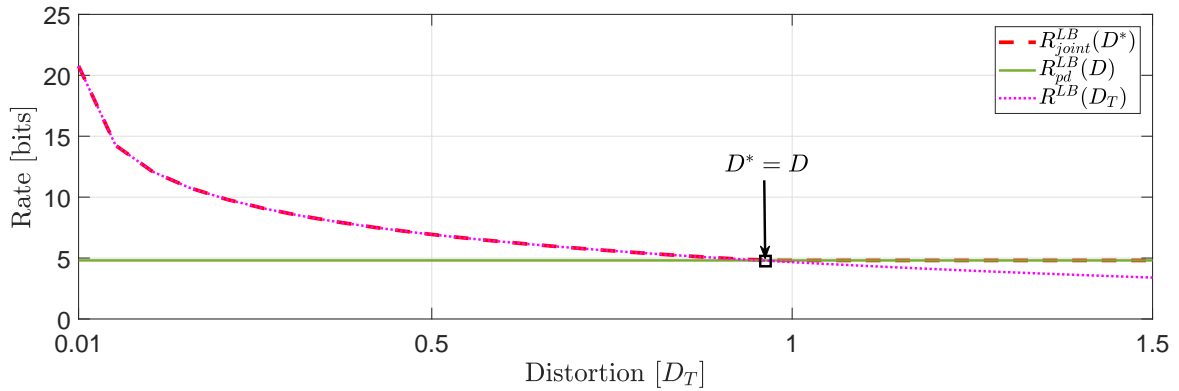
In Figure 3, we compare the numerical solutions of  $R_{\text{joint}}^{\text{LB}}(D^*)$  and  $R_{\text{pd}}^{\text{LB}}(D)$  with ([16], Equation (27)), denoted hereinafter as  $R^{\text{LB}}(D_{\tau})$ .



**Figure 3.** Comparison of  $R_{\text{joint}}^{\text{na}}(D^*)$ ,  $R_{\text{pd}}^{\text{na}}(D)$  when  $D_{ii} \neq D_{jj}$ , for  $i \neq j$ , and comparison with ([16], Equation (27)).

Based on this numerical study, we observe that for distortion levels between  $(0, D^* = D]$ ,  $R_{\text{joint}}^{\text{LB}}(D^*) \geq R_{\text{pd}}^{\text{LB}}(D)$  whereas for values of  $D_T$  greater than  $D^*$  we observe that  $R_{\text{joint}}^{\text{LB}}(D^*) = R_{\text{pd}}^{\text{LB}}(D)$  because the asymptotically averaged total MSE distortion constraint is inactive. This observation verifies our comment in Section 1.2 regarding the connection of (11) and (12). Clearly, at high rates (or high resolution) we observe that  $R_{\text{joint}}^{\text{LB}}(D^*) \approx R^{\text{LB}}(D_T)$ .

Another interesting observation (illustrated in Figure 4) that can be made, is that if in the same example we allocate the total budget of per dimension distortion equally, i.e.,  $D_{ii} = D_{jj}$ ,  $\forall i \neq j$ , we observe that for distortion levels between  $(0, D^* = D]$ ,  $R_{\text{joint}}^{\text{LB}}(D^*) = R^{\text{LB}}(D_T) \geq R_{\text{pd}}^{\text{LB}}(D)$ .



**Figure 4.** Comparison of  $R_{\text{joint}}^{\text{LB}}(D^*)$ ,  $R_{\text{pd}}^{\text{LB}}(D)$  and  $R^{\text{LB}}(D_T)$  when  $D_{ii} = D_{jj}$ ,  $\forall i \neq j$ .

**Example 2** (Covariance matrix distortion constraint). For the system in (1) we assume that user 1 is described by a  $\mathbb{R}^2$ -valued time-invariant Markov source driven by i.i.d. Gaussian noise process with parameters  $(A_1, \Sigma_{\mathbf{w}^1})$ :

$$(A_1, \Sigma_{\mathbf{w}^1}) = \left( \begin{bmatrix} 0.5 & 0.2 \\ 0.3 & 0.6 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right), \quad (54)$$

whereas, user 2 is described by an  $\mathbb{R}$ -valued time-invariant Markov source driven by i.i.d. Gaussian noise process with parameters  $(A_2, \Sigma_{\mathbf{w}^2})$ :

$$(A_2, \Sigma_{\mathbf{w}^2}) = (0.6, 2). \quad (55)$$

The augmented state space model (2) is generated by  $A = A_1 \oplus A_2$  and similarly  $\Sigma_{\mathbf{w}} = \Sigma_{\mathbf{w}}^1 \oplus \Sigma_{\mathbf{w}}^2$ . For this example, we assume a covariance matrix distortion constraint given by:

$$D_{\text{cov}} = \begin{bmatrix} 1.5 & \gamma & \gamma \\ \gamma & 1 & \gamma \\ \gamma & \gamma & 0.5 \end{bmatrix}, \quad (56)$$

where  $\gamma > 0$  is the positive correlation coefficient between the distortion matrix components (i.e., diagonal entries) and it is chosen such that  $D_{\text{cov}} \succ 0$ .

In Figure 5 we demonstrate a comparison between  $R^{LB}(D_{\text{cov}})$  evaluated for several different values of  $\gamma$ . One interesting observation that can be made is that higher distortion correlation in (56) leads to less bits with a  $\gamma_{\text{max}} \approx 0.53$ , beyond which the value of  $R^{LB}(D_{\text{cov}})$  remains unchanged. Another interesting observation is that for negative correlation  $\gamma$ , the approximation via SDP does not give a number. However, this is not the case, in general (see, e.g., ([42], Example 1)).

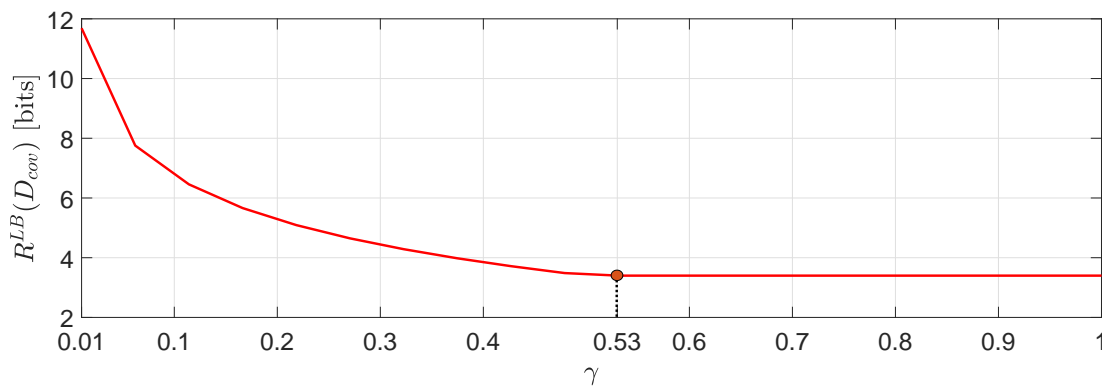


Figure 5.  $R^{LB}(D_{\text{cov}})$  as a function of  $\gamma \geq 0$ .

Using the same simulation study, we can arrive to an interesting connection between the approximation in (51) and (48). In particular, if for instance in (56) we restrict the matrix distortion constraint only to the main diagonal elements (i.e., exactly like the the per-dimension constraints) then, we obtain the plot of Figure 6 which clearly demonstrates that  $R^{LB}(D_{\text{cov}}) = R_{\text{pd}}^{LB}(D)$ . In fact, restricting the covariance matrix distortion constraint of (56) to the per dimension distortion constraint, is as if we optimize via a solution space in which  $\gamma$  is allowed to have any value in  $\mathbb{R}$ . As a result, the feasible set of solutions is larger when the constraint set is subject to per-dimension distortion constraints rather than the covariance matrix distortion constraint.

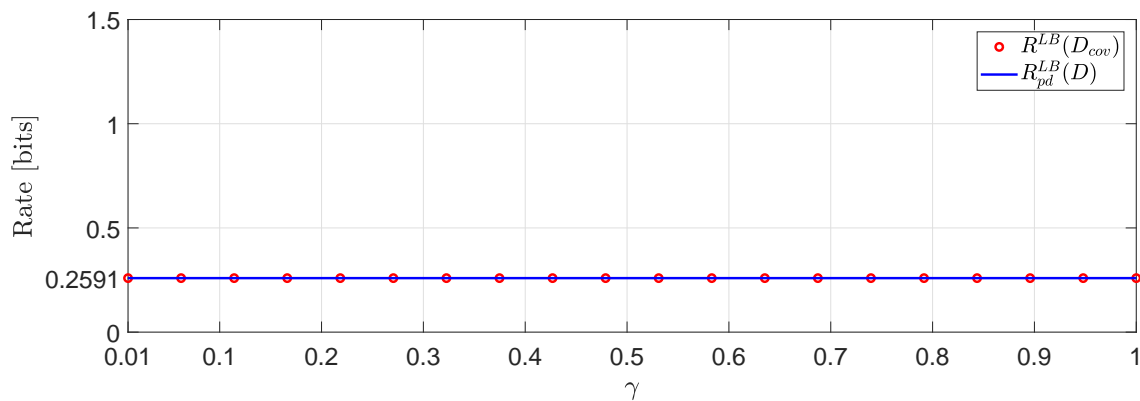


Figure 6. Comparison of  $R^{LB}(D_{\text{cov}})$  (restricted to its values in the main diagonal) and  $R_{\text{pd}}^{LB}(D)$  for certain values of  $\gamma$ .

## 2.2. Analytical Lower Bounds for Markov Sources Driven by Additive i.i.d. Noise Processes

In this subsection, we derive analytical lower bounds on (11)–(13) when the source model describing the behavior of user 1 or user 2 is driven by possibly i.i.d. non-Gaussian noise process.

We first give a lemma which will facilitate the derivation of our lower bounds. We only consider the case of RDFs subject to per-dimension distortion constraints because the other classes of distortion constraints follow similarly.

**Lemma 4** (Rate-distortion bounds). *For the augmented source model describing the behavior of users 1, 2 in (3), the following inequalities hold assuming distortion constraints in the class of (8):*

$$R_{\text{pd}}^{\text{LB}}(D) \stackrel{(a)}{\leq} R_{\text{pd}}^{\text{LB}, \text{linear}}(D) \leq R_{\text{pd}}^{\text{c}}(D), \quad (57)$$

where

$$R_{\text{pd}}^{\text{LB}, \text{linear}}(D) \triangleq \inf_{\substack{\mathbf{P}^{\text{linear}}(dy_t|y_{t-1}, x_t): t \in \mathbb{N}_0 \\ \Sigma_{\Delta_{ii}} \leq D_{ii}, i=1, \dots, p}} \limsup_{n \rightarrow \infty} \frac{1}{n+1} \sum_{t=0}^n I(\mathbf{x}_t; \mathbf{y}_t | \mathbf{y}_{t-1}), \quad (58)$$

and (a) holds with equality if the augmented state space model described in (2) is jointly Gaussian and the optimal minimizer, i.e.,  $\mathbf{P}^*(dy_t|y_{t-1}, x_t)$  of  $R_{\text{pd}}^{\text{LB}}(D)$  is conditionally Gaussian. Equality (a) holds trivially at  $D_{\max}$ .

**Proof.** The RHS inequality follows from Theorem 1 and (43) whereas the LHS inequality follows from the fact that the constraint set of  $R_{\text{pd}}^{\text{LB}}(D)$  is larger than the constraint set of  $R_{\text{pd}}^{\text{LB}, \text{linear}}(D)$  which is restricted to linear coding policies. Now, under the specific augmented source model in (3), and using Lemma 2, (1), we obtain  $R_{\text{pd}}^{\text{LB}, \text{linear}}(D)$  defined as in (58) because these are the best linear coding policies since KF algorithm is the best linear causal MSE estimator beyond additive Gaussian noise processes (see the discussion in Remark 3, (3)). Clearly, if the augmented source in (3) is jointly Gaussian and the optimal realization of  $R_{\text{pd}}^{\text{LB}}(D)$  is conditionally Gaussian, then, the system model is jointly Gaussian and the optimal policies are linear given by the forward linear test channel realization obtained in (36) hence the LHS inequality holds with equality.  $\square$

**Remark 7** (Comments on Lemma 4). *We note that Lemma 4 holds if we assume RDFs with distortion constraints in the class of (9) or (10).*

The following theorem is a major result of this paper.

**Theorem 3** (Analytical lower bounds on (11)–(13)). *Consider the source models of users 1, 2 in (1). Then, the following analytical lower bounds on (11)–(13) hold.*

(1) For  $D_{ii} > 0$ ,  $\forall i$ , we obtain

$$R_{\text{pd}}^{\text{c}}(D) \geq \frac{(p_1 + p_2)}{2} \log \left( |A^T A|^{\frac{1}{p_1 + p_2}} + \frac{(p_1 + p_2)N(\mathbf{w})}{D} \right), \quad (59)$$

where  $D \in \left( 0, \frac{(p_1 + p_2)N(\mathbf{w})}{1 - |A^T A|^{\frac{1}{p_1 + p_2}}} \right]$  with  $N(\mathbf{w}) = \frac{1}{2\pi e} 2^{\frac{2}{p_1 + p_2} h(\mathbf{w})}$  and  $h(\mathbf{w}) > -\infty$ .

(2) For  $D_r > 0$ , and  $D_{ii} > 0$ ,  $\forall i$ , we obtain

$$R_{\text{joint}}^{\text{c}}(D^*) \geq \frac{(p_1 + p_2)}{2} \log \left( |A^T A|^{\frac{1}{p_1 + p_2}} + \frac{(p_1 + p_2)N(\mathbf{w})}{D^*} \right), \quad (60)$$

where  $D^* \in \left(0, \frac{(p_1+p_2)N(\mathbf{w})}{1-|A^T A|^{\frac{1}{p_1+p_2}}}\right]$ , with  $N(\mathbf{w})$  defined as in (1) and  $h(\mathbf{w}) > -\infty$ .

(3) For  $D_{\text{cov}} \succ 0$  we obtain

$$R^c(D_{\text{cov}}) \geq \frac{(p_1+p_2)}{2} \log \left( |A^T A|^{\frac{1}{p_1+p_2}} + \frac{N(\mathbf{w})}{|D_{\text{cov}}|^{\frac{1}{p_1+p_2}}} \right), \quad (61)$$

where  $|D_{\text{cov}}| \in \left(0, \left(\frac{N(\mathbf{w})}{1-|A^T A|^{\frac{1}{p_1+p_2}}}\right)^{(p_1+p_2)}\right]$ , with  $N(\mathbf{w})$  defined as in (1) and  $h(\mathbf{w}) > -\infty$ .

**Proof.** See Appendix C.  $\square$

The following technical remarks can be made regarding Theorem 3.

**Remark 8.**

- (1) Note that if in Theorem 3 we allow  $h(\mathbf{w}) = -\infty$ , then, the analytical lower bound expressions take a negative finite value or  $-\infty$ , which cannot be the case (RDF is, by definition, non-negative). A way to include the case where  $h(\mathbf{w})$  is allowed to be  $-\infty$  in our lower bound expressions, is to set the objective functions in (59)–(61) to be  $[\log(\cdot)]^+$ . This will mean that whenever  $h(\mathbf{w}) = -\infty$ , the analytical lower bound expression will be zero.
- (2) The analytical lower bounds in (59)–(61) do not correspond to the best linear forward test channel realization of Lemma 3 (see (46)) which is also the optimal policy under the assumption of a MMSE decoder when the system's noise is purely Gaussian (see Remark 3, (3)). Moreover, it is not clear what is the realization that achieves them the same way the bounds in Lemma 3 are achieved for Gaussian processes.
- (3) If in Theorem 3 we assume that the users 1,2 have source models described by Markov processes driven by additive Gaussian noise processes then from the EPI (see, e.g., ([43], Equation (7)))  $N(\mathbf{w}) = |\Sigma_{\mathbf{w}}|^{p_1+p_2}$  and (59)–(61) change accordingly.
- (4) One can choose to further bound (61) using the inequality  $|D_{\text{cov}}|^{-\frac{1}{p_1+p_2}} \geq \frac{p_1+p_2}{\text{trace}(D_{\text{cov}})}$  obtaining a further lower bound that coincides with the lower bound in (59) (see also the discussion in Example 2). Such lower bound will mean that we extend the set of feasible solutions that correspond to the initial problem statement (13), to be similar to the initial problem statement of (11) which cannot be the case, in general. Our bound in (61) encapsulates the off diagonal elements of the distortion covariance matrix distortion  $D_{\text{cov}}$  hence it is an appropriate lower bound for the specific problem.

In what follows, we give a numerical simulation where we compare the solution of  $R_{\text{joint}}^{\text{LB}}(D^*)$  (that corresponds to the lower bound achieved by the optimal coding policies when the system is driven by additive i.i.d. Gaussian noise processes) computed via the SDP representation of (50), with the lower bound obtained in (60) when the system's noise is also Gaussian.

**Example 3** (Comparison of (50) with (60) for jointly Gaussian processes). We consider the same input data assumed in Example 1 for users 1,2. Then, we proceed to compute the true lower bound of (50) and the lower bound obtained in (60).

Our simulation study in Figure 7 shows that at high rates the performance of the two bounds is almost identical whereas to moderate and low rates we observe a gap that remains constant when  $D_{\tau} \geq D$ , i.e., when the asymptotically averaged total distortion constraint is inactive. The same behavior is expected for systems of larger dimension (larger scale optimization problems) with a possibility of an increased gap to moderate and low rates depending on the structure of the block diagonal matrices  $A$  and  $\Sigma_{\mathbf{w}}$ .

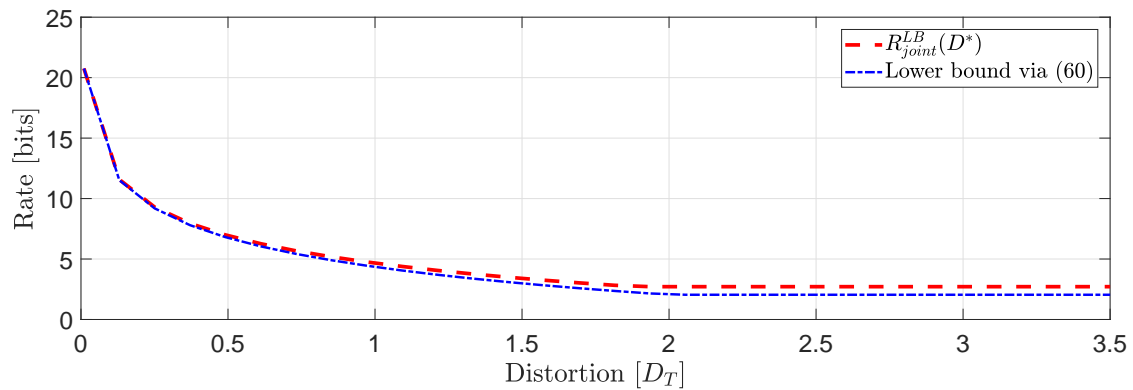


Figure 7. Comparison of the lower bound  $R_{\text{joint}}^{LB}(D^*)$  with the analytical expression of (60).

Next, we state a corollary of Theorem 3.

**Corollary 2** (Analytical bounds when users 1, 2 are not specified by the same additive noise process). Consider the source models of users 1, 2 in (1). Moreover, assume that  $\mathbf{w}_t^1 \sim \mathcal{N}(0; \Sigma_{\mathbf{w}^1})$  and  $\mathbf{w}_t^2 \sim (0; \Sigma_{\mathbf{w}^2})$  with  $\Sigma_{\mathbf{w}^1} \preceq I_{p_1}$  and  $\Sigma_{\mathbf{w}^2} \preceq I_{p_2}$  and  $h(\mathbf{w}^2) > -\infty$ . Then, the following analytical lower bounds on (11)–(13) hold.

(1) For  $D_{ii} > 0$ ,  $\forall i$ , we obtain

$$R_{\text{pd}}^c(D) \geq \frac{(p_1 + p_2)}{2} \log \left( |A^T A|^{\frac{1}{p_1 + p_2}} + \frac{(p_1 + p_2) |\Sigma_{\mathbf{w}^1}|^{\frac{1}{p_1}} N(\mathbf{w}^2)}{D} \right), \quad (62)$$

where  $D \in \left( 0, \frac{(p_1 + p_2) N(\mathbf{w}^2)}{1 - |A^T A|^{\frac{1}{p_1 + p_2}}} \right]$  with  $N(\mathbf{w}^2) = \frac{1}{2\pi e} 2^{\frac{2}{p_2} h(\mathbf{w}^2)}$ .

(2) For  $D_T > 0$ ,  $D_{ii} > 0$ ,  $\forall i$ , we obtain

$$R_{\text{joint}}^c(D^*) \geq \frac{(p_1 + p_2)}{2} \log \left( |A^T A|^{\frac{1}{p_1 + p_2}} + \frac{(p_1 + p_2) |\Sigma_{\mathbf{w}^1}|^{\frac{1}{p_1}} N(\mathbf{w}^2)}{D^*} \right), \quad (63)$$

where  $D^* \in \left( 0, \frac{(p_1 + p_2) |\Sigma_{\mathbf{w}^1}|^{\frac{1}{p_1}} N(\mathbf{w}^2)}{1 - |A^T A|^{\frac{1}{p_1 + p_2}}} \right]$ .

(3) For  $D_{\text{cov}} \succ 0$  we obtain

$$R^c(D_{\text{cov}}) \geq \frac{(p_1 + p_2)}{2} \log \left( |A^T A|^{\frac{1}{p_1 + p_2}} + \frac{|\Sigma_{\mathbf{w}^1}|^{\frac{1}{p_1}} N(\mathbf{w}^2)}{|D_{\text{cov}}|^{\frac{1}{p_1 + p_2}}} \right), \quad (64)$$

where  $|D_{\text{cov}}| \in \left( 0, \left( \frac{|\Sigma_{\mathbf{w}^1}|^{\frac{1}{p_1}} N(\mathbf{w}^2)}{1 - |A^T A|^{\frac{1}{p_1 + p_2}}} \right)^{(p_1 + p_2)} \right]$ .

**Proof.** All cases (1)–(3) follow almost identical steps with the derivation of Theorem 3. The only different but crucial step lies in (A6) where we then use the fact that

$$|\Sigma_{\mathbf{w}}|^{\frac{1}{p_1 + p_2}} \stackrel{(a)}{=} |\Sigma_{\mathbf{w}^1}|^{\frac{1}{p_1 + p_2}} |\Sigma_{\mathbf{w}^2}|^{\frac{1}{p_1 + p_2}} \stackrel{(b)}{\geq} |\Sigma_{\mathbf{w}^1}|^{\frac{1}{p_1}} |\Sigma_{\mathbf{w}^2}|^{\frac{1}{p_2}} \stackrel{(c)}{\geq} |\Sigma_{\mathbf{w}^1}|^{\frac{1}{p_1}} N(\mathbf{w}^2), \quad (65)$$

where (a) follows from properties of block diagonal matrices ([33], Section 0.9.2); (b) follows from the conditions of the corollary on the noise covariance matrices; (c) follows from the EPI ([43], Equation (7)).  $\square$

One can deduce the following for Corollary 2.

**Remark 9.** Corollary 2 will give similar analytical lower bounds (with appropriate modifications) if instead of user 1, we assume that the source model of user 2 is driven by a Gaussian noise process. The additional assumption on the covariance matrix of the noise process in both users is imposed because otherwise we cannot guarantee that the key series of inequalities (65) will be satisfied.

### 3. Upper Bounds

In this section we explain the case of encoding the augmented vector-valued Markov source modeled by (3) using a sequential causal DPCM scheme with a feedback loop followed by an ECDQ. The scheme relies on the linear forward test channel realization of the bounds in Lemma 2. The precursor of the DPCM-based scheme with feedback loop is [14] whereas ECDQ is a classical source coding approach with standard performance guarantees in information theory (see, e.g., [44]). The ECDQ scheme is utilized to bound the rate performance of the DPCM scheme. This approach furnish with an achievable (upper) bound the operation causal RDFs in (11)–(13).

#### 3.1. DPCM with Feedback Loop

First, we briefly describe the sequential causal DPCM scheme with feedback loop introduced in ([14], Figure 2) (see also [45]). Observe that because the augmented source is modeled as a first-order multidimensional Markov process, the sequential causal coding is precisely equivalent to a predictive coding paradigm (see, e.g., [14,46]).

At each time instant  $t$ , the encoder or innovations' encoder performs the linear operation

$$\hat{\mathbf{x}}_t = \mathbf{x}_t - A\mathbf{y}_{t-1}, \quad (66)$$

where at  $t = 0$  we assume initial data  $\hat{\mathbf{x}}_0 = \mathbf{x}_0$  and also  $\mathbf{y}_{t-1} \triangleq \mathbb{E}\{\mathbf{x}_{t-1}|\ell^{t-1}\}$ , i.e., an estimate of  $\mathbf{x}_{t-1}$  given the previous quantized symbols  $\ell^{t-1}$  (Note that the process  $\hat{\mathbf{x}}_t$  has a temporal correlation since it subtracts the error of  $\mathbf{x}_t$  given all previous quantized symbols  $\ell^{t-1}$  and not the infinite past of the source  $\mathbf{x}_{-\infty}^t$ . Hence,  $\hat{\mathbf{x}}_t$  is only an estimate of the true process and this causes a part of the sub-optimality of this scheme.). Then, by means of a  $\mathbb{R}^{p_1+p_2}$ -valued MMSE quantizer that operates at a rate  $R_t$ , we generate the quantized reconstruction  $\hat{\mathbf{y}}_t$  of the residual source  $\hat{\mathbf{x}}_t$  denoted by  $\hat{\mathbf{y}}_t = \mathbf{y}_t - A\mathbf{y}_{t-1}$ . Then, we send  $\ell_t$  over the channel (the corresponding data packet to  $\hat{\mathbf{y}}_t$ ). At the decoder we receive  $\ell_t$  and recover the quantized symbol  $\hat{\mathbf{y}}_t$  of  $\hat{\mathbf{x}}_t$ .

Then, we generate the estimate  $\mathbf{y}_t$  using the linear operation

$$\mathbf{y}_t = \hat{\mathbf{y}}_t + A\mathbf{y}_{t-1}. \quad (67)$$

Combining both (66) and (67), we obtain

$$\mathbf{x}_t - \mathbf{y}_t = \hat{\mathbf{x}}_t - \hat{\mathbf{y}}_t. \quad (68)$$

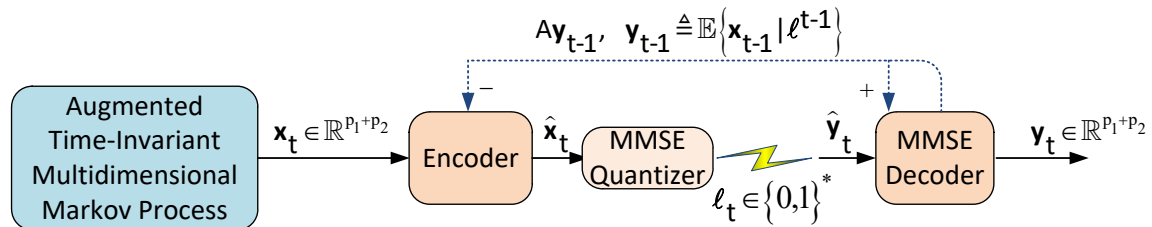
From (68), we can immediately deduce that the error between  $\mathbf{x}_t$  and  $\mathbf{y}_t$  is equal to the quantization error introduced by  $\hat{\mathbf{x}}_t$  and  $\hat{\mathbf{y}}_t$  which means that the MSE distortion at each instant of time satisfy

$$\mathbb{E}\{\|\mathbf{x}_t - \mathbf{y}_t\|_2^2\} = \mathbb{E}\{\|\hat{\mathbf{x}}_t - \hat{\mathbf{y}}_t\|_2^2\}. \quad (69)$$

In addition, the covariance matrix  $\Sigma_\Delta$  yields

$$\mathbb{E}\{(\mathbf{x}_t - \mathbf{y}_t)(\mathbf{x}_t - \mathbf{y}_t)^T\} = \mathbb{E}\{(\hat{\mathbf{x}}_t - \hat{\mathbf{y}}_t)(\hat{\mathbf{x}}_t - \hat{\mathbf{y}}_t)^T\}. \quad (70)$$

A pictorial view of the DPCM scheme with feedback loop is given in Figure 8.



**Figure 8.** DPCM scheme with feedback loop for the augmented multidimensional Markov model of (3).

### 3.2. Bounding (11)–(13) via A DPCM-based ECDQ for Gaussian Noise Processes

In this subsection, we bound the rate performance of the DPCM scheme described in Section 3.1 in the infinite time horizon, using a scheme that utilizes the steady-state linear forward test-channel realization that achieves the lower bounds of Lemma 3. Essentially, what we do in this scheme is that we replace the quantization noise with an additive Gaussian noise with the same second moments (see e.g., [47] or [44] (Chapter 5) and the references therein).

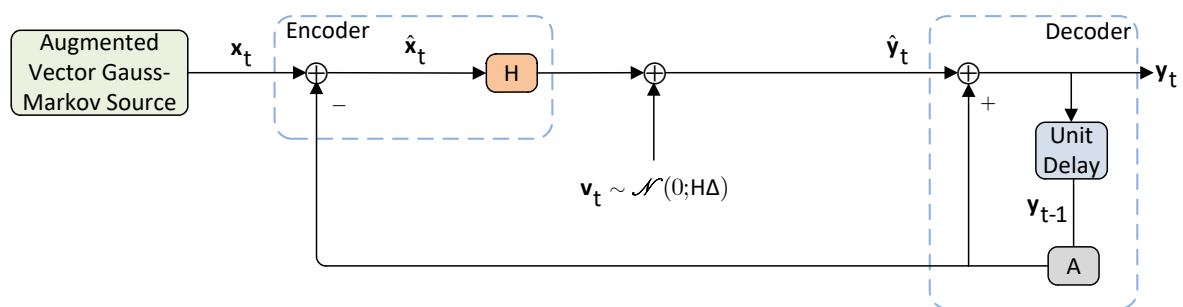
Recall that the steady-state linear forward test-channel realization of the lower bounds in Lemma 3 is written as follows:

$$\mathbf{y}_t = H\mathbf{x}_t + (I_{p_1+p_2} - H)A\mathbf{y}_{t-1} + \mathbf{v}_t, \quad (71)$$

whereas the steady-state reverse-waterfilling parameters  $(H, \Sigma_v)$  are given by

$$H \triangleq I_{p_1+p_2} - \Delta\Lambda^{-1}, \Sigma_v = H\Delta \succeq 0. \quad (72)$$

The forward test-channel realization of (71) is illustrated in Figure 9.



**Figure 9.** Forward test channel realization of (71).

Before we proceed, we point out the following important technical remarks on the realization of (71) and the coefficients (72).

**Remark 10** (Observations (71) and (72)). *The linear forward test channel realization with additive noise in (71) is equivalent to the steady-state realization in (46) because for both it can be shown that the MSE distortion constraint is achieved (i.e.,  $\mathbf{v}_t \sim \mathcal{N}(0; \Sigma_v)$ ,  $\Sigma_v = H\Delta = \Delta H^T \succeq 0$ ). Moreover, this realization is equivalent but simpler to build compared to the forward test channel realization introduced in [14] in which non-singular matrices and diagonalization by congruence is assumed (see ([14], Theorem 4)).*

In the test channel realization of Figure 9, a reverse-waterfilling in spatial dimension is possible when we assume asymptotically averaged total MSE distortion constraints similar to ([14], Theorem 40). This reverse-waterfilling is dictated by the rank of matrix  $H$ . To make this point clear, if  $H$  is full rank, then all spatial dimensions in the system are active whereas if  $H$  is rank deficient, then, some dimensions are inactive (these dimensions form the null space and the nullity of  $H$ ) and for these dimensions the rate is zero hence they can be excluded from the realization of Figure 9. In the sequel, we will present simulations where we study the reverse-waterfilling in the spatial domain under a certain distortion constraint studied in this paper.

**Pre/Post Filtered ECDQ with multiplicative factors for augmented multivariate Gauss–Markov sources and spatial reverse-waterfilling.** First, we consider a  $\text{rank}(H)$ –dimensional lattice quantizer  $Q_{\text{rank}(H)}(\cdot)$  [48] such that

$$\mathbb{E}\{\mathbf{z}_t \mathbf{z}_t^T\} = \Sigma_{\mathbf{v}^c}, \Sigma_{\mathbf{v}^c} \succ 0,$$

where  $\mathbf{z}_t \in \mathbb{R}^{\text{rank}(H)}$  is a random dither vector generated both at the encoder and the decoder independent of the input signals  $\hat{\mathbf{x}}_t$  and the previous realizations of the dither, uniformly distributed over the basic Voronoi cell of the  $\text{rank}(H)$ –dimensional lattice quantizer  $Q_{\text{rank}(H)}(\cdot)$  such that  $\mathbf{v}_t^c \sim \text{Unif}(0; \Sigma_{\mathbf{v}^c})$ . At the encoder the lattice quantizer quantize  $H\hat{\mathbf{x}}_t + \mathbf{z}_t$ , that is,  $Q_{\text{rank}(H)}(H\hat{\mathbf{x}}_t + \mathbf{z}_t)$ , where  $\hat{\mathbf{x}}_t$  is given by (66). Then, the encoder applies conditional entropy coding to the output of the quantizer and transmits the output of the entropy coder. At the decoder the coded bits are received and the output of the quantizer is reconstructed, i.e.,  $Q_{\text{rank}(H)}(H\hat{\mathbf{x}}_t + \mathbf{z}_t)$ . Then, it generates an estimate by subtracting  $\mathbf{z}_t$  from the quantizer’s output and multiplies the result by  $I_{\text{rank}(H)}$  ( $I_{\text{rank}(H)}$  denotes the identity matrix with dimensions according to the rank of  $H$ . This identity matrix can be excluded but we include it here for completeness.) as follows:

$$\mathbf{y}_t = I_{\text{rank}(H)}(Q_{\text{rank}(H)}(H\hat{\mathbf{x}}_t + \mathbf{z}_t) - \mathbf{z}_t), \quad (73)$$

Performance. The coding rate at each instant of time of the conditional entropy of the MMSE quantizer is given by

$$\begin{aligned} H(Q_{\text{rank}(H)}|\mathbf{z}_t) &= I(H\hat{\mathbf{x}}_t; H\hat{\mathbf{x}}_t + \mathbf{v}_t^c) \\ &\stackrel{(a)}{=} I(H\hat{\mathbf{x}}_t; H\hat{\mathbf{x}}_t + \mathbf{v}_t) + \mathbb{D}(\mathbf{v}_t^c || \mathbf{v}_t) - \mathbb{D}(H\hat{\mathbf{x}}_t + \mathbf{v}_t^c || H\hat{\mathbf{x}}_t + \mathbf{v}_t) \\ &\stackrel{(b)}{\leq} I(H\hat{\mathbf{x}}_t; H\hat{\mathbf{x}}_t + \mathbf{v}_t) + \mathbb{D}(\mathbf{v}_t^c || \mathbf{v}_t) \\ &\stackrel{(c)}{\leq} I(H\hat{\mathbf{x}}_t; H\hat{\mathbf{x}}_t + \mathbf{v}_t) + \frac{\text{rank}(H)}{2} \log(2\pi e G_{\text{rank}(H)}) \\ &\stackrel{(d)}{\leq} I(\mathbf{x}_t; \mathbf{y}_t | \mathbf{y}^{t-1}) + \frac{\text{rank}(H)}{2} \log(2\pi e G_{\text{rank}(H)}) \end{aligned} \quad (74)$$

where  $\mathbf{v}_t^c \in \mathbb{R}^{\text{rank}(H)}$  is the (uniform) coding noise in the ECDQ scheme and  $\mathbf{v}_t$  is the corresponding Gaussian counterpart; (a) follows because the two random vectors  $\mathbf{v}_t^c, \mathbf{v}_t$  have the same second moments hence we can use the identity  $\mathbb{D}(\mathbf{x} || \mathbf{x}') = h(\mathbf{x}') - h(\mathbf{x})$ ; (b) follows because  $\mathbb{D}(H\hat{\mathbf{x}}_t + \mathbf{v}_t^c || H\hat{\mathbf{x}}_t + \mathbf{v}_t) \geq 0$ ; (c) follows because the divergence of the coding noise from Gaussianity is less than or equal to  $\frac{\text{rank}(H)}{2} \log(2\pi e G_{\text{rank}(H)})$  [47] where  $G_{\text{rank}(H)}$  is the dimensionless normalized second moment of the lattice ([44], Definition 3.2.2); (d) follows from data processing properties, namely,  $I(\mathbf{x}_t; \mathbf{y}_t | \mathbf{y}^{t-1}) \stackrel{(*)}{=} I(\mathbf{x}_t; \mathbf{y}_t | \mathbf{y}_{t-1}) \stackrel{(**)}{=} I(\hat{\mathbf{x}}_t; \hat{\mathbf{y}}_t) \stackrel{(***)}{\geq} I(H\hat{\mathbf{x}}_t; H\hat{\mathbf{x}}_t + \mathbf{v}_t)$  where (a) follows from the realization of (71), (b) follows from the fact that  $\hat{\mathbf{x}}_t$  and  $\hat{\mathbf{y}}_t$  (obtained by (67)) are independent of  $\mathbf{y}_{t-1}$ , and (c) is a consequence of data processing inequality since  $(H\hat{\mathbf{x}}_t + \mathbf{v}_t) \leftrightarrow \hat{\mathbf{x}}_t \leftrightarrow H\hat{\mathbf{x}}_t$ . Under the assumption that

the clocks of the entropy encoder and entropy decoder in the ECDQ scheme are synchronized, then, the total coding rate is obtained as follows

$$\begin{aligned} \sum_{t=1}^n R_t &\leq \sum_{t=0}^n (H(Q_{\text{rank}(H)}|\mathbf{z}_t) \stackrel{(e)}{\leq} \sum_{t=0}^n I(\mathbf{x}_t; \mathbf{y}_t | \mathbf{y}^{t-1}) + \frac{(n+1) \text{rank}(H)}{2} \log(2\pi e G_{\text{rank}(H)}) \\ &\stackrel{(f)}{=} \frac{1}{2} \sum_{t=0}^n \log_2 \frac{|\Delta_t|}{|\Delta|} + \frac{(n+1) \text{rank}(H)}{2} \log(2\pi e G_{\text{rank}(H)}), \end{aligned} \quad (75)$$

where (e) follows from (74); (f) follows from the derivation of Lemma 2.

The previous analysis yields the following theorem.

**Theorem 4** (Achievability bound on (11)–(13)). *Suppose that  $\Delta_t = \Delta$ ,  $\forall t$  and assume that the users 1, 2 source models (1) are driven by Gaussian noise processes. Then, the augmented state space source model in (3) ensures the following achievability bounds on (11)–(13), as follows*

$$R_{\text{pd}}^c(D) \leq R_{\text{pd}}^{LB}(D) + \frac{\text{rank}(H)}{2} \log(2\pi e G_{\text{rank}(H)}) \quad (76)$$

$$R_{\text{joint}}^c(D^*) \leq R_{\text{joint}}^{LB}(D^*) + \frac{\text{rank}(H)}{2} \log(2\pi e G_{\text{rank}(H)}), \quad (77)$$

$$R^c(D_{\text{cov}}) \leq R^{LB}(D_{\text{cov}}) + \frac{\text{rank}(H)}{2} \log(2\pi e G_{\text{rank}(H)}). \quad (78)$$

**Proof.** Under the conditions of the Theorem and the ECDQ scheme that leads to (75), the RHS terms in (76)–(78) are all constants. Then, taking the limit in both sides of (76)–(78) and then the appropriate infimization (minimization) constraint sets, the result follows.  $\square$

We wish to point out the following for Theorem 4.

**Remark 11** (Comments on Theorem 4).

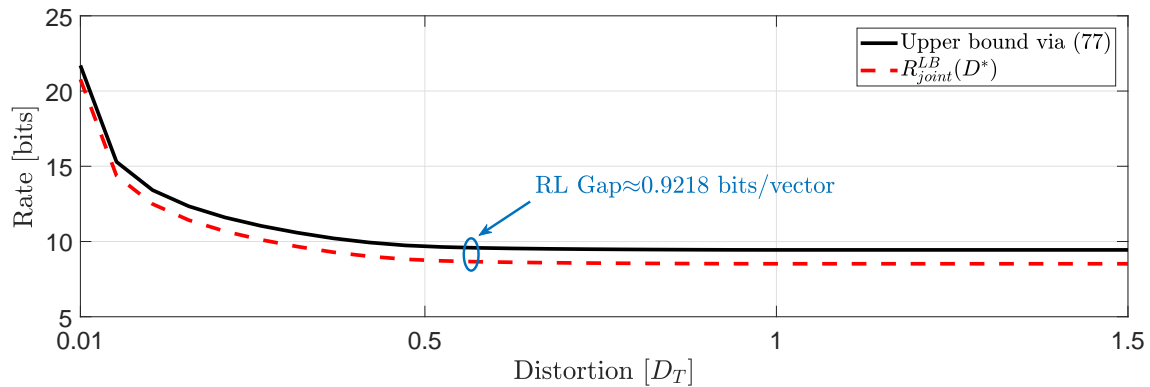
- (1) The ECDQ that leads to (75) is not the same as the standard symmetric ECDQ scheme for scalar-valued processes, i.e., when the coefficient  $H$  breaks into two pre and post additive noise channel scalings that tune the MSE distortion (see, e.g., [44,47]). In our pre-post scaled ECDQ scheme we take asymmetric coefficients based on the realization of Figure 9. This leads to a coarser lattice than the one used for the unscaled ECDQ (for details see for instance [44]).
- (2) Since the upper bound essentially relies on the obtained lower bound for all (76)–(78), this means that similar observations can be made. For instance, if  $D < D_{\tau}$ , then, (77) recovers (76), i.e., the asymptotically averaged total distortion constraint is inactive. Moreover, we cannot claim tightness of the achievability bound in (78) because the lower bound is already non-tight.

Next, we give an example where we compare the RL gap for various distortion levels of the achievability bound obtained in Theorem 4 and the lower bound obtained in Theorem 2 for the operational causal RDF with joint distortion constraints.

**Example 4** (RL gap of achievability and lower bounds). *In this example, we plot lower and upper bounds on the operational causal RDF subject to joint distortion constraints using the bounds derived in (50) and (77), respectively. We first consider the same input data assumed in Example 1 for users 1, 2. Then, we proceed to compute the lower bound via (50) and the achievability bound in (77). After the first numerical study, we consider another one for which we only change the Gaussian noise process for both users 1, 2, as follows*

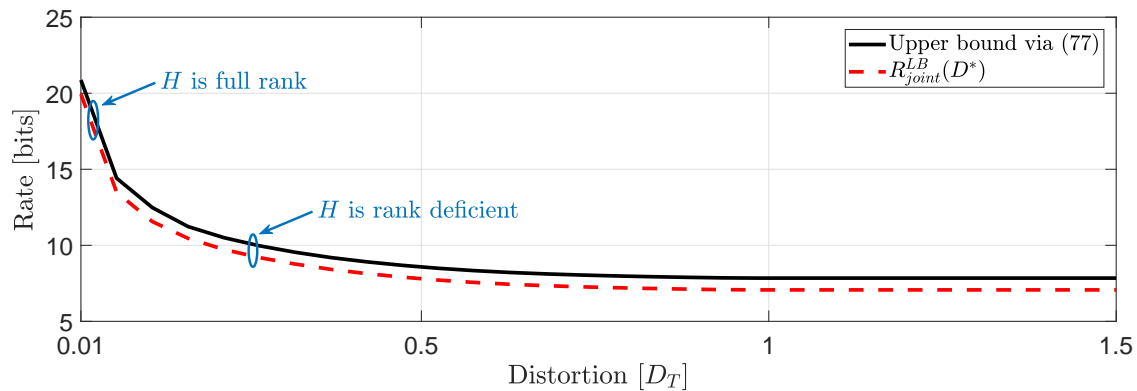
$$(\Sigma_{\mathbf{w}^1}, \Sigma_{\mathbf{w}^2}) = \left( \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}, \begin{bmatrix} 1.4039 & 0.6034 & 0.5165 \\ 0.6034 & 0.9563 & 0.7682 \\ 0.5165 & 0.7682 & 0.6620 \end{bmatrix} \right). \quad (79)$$

Using the data of Example 1, and the same  $D_T$  and  $D_{ii}$ ,  $i = 1, 2, 3, 4, 5$ , we obtain the plots of Figure 10. For this study, we have used a Schläfli lattice (for details on this lattice see, e.g., [48]).  $\tilde{D}_5$  with a dimensionless normalized second moment of the lattice  $G_5 \approx 0.0756$  bits. In this example  $H$  is always full rank and the RL gap is constant at 0.9218 bits/augmented vector.



**Figure 10.** Comparison of lower and upper bounds on  $R_{\text{joint}}^c(D)$  when  $H$  is full rank.

For the second example, we obtain the plots of Figure 11. For this study, we have used the dimensionless normalized second moment of a Schläfli lattice  $\tilde{D}_5$  for the full rank case and for the rank deficient cases a Schläfli lattice  $D_4$  with a dimensionless normalized second moment of the lattice  $G_4 \approx 0.0766$  bits. Similar to the first study, when  $H$  is always full rank, the RL gap is 0.9218 bits/augmented vector whereas for  $H$  rank deficient the RL gap is 0.7754 bits/augmented vector.



**Figure 11.** Comparison of the lower and upper bounds on  $R_{\text{joint}}^c(D)$  when  $H$  is rank deficient.

### 3.3. Bounding (11)–(13) via a DPCM-based ECDQ for Non-Gaussian Noise Processes

Similar to Lemma 4 where linear policies are the benchmark to derive lower bounds on (11)–(13), in this subsection we derive upper bounds on (11)–(13) using the linear test channel realization in Figure 9 and the DPCM-based ECDQ scheme of Sections 3.1 and 3.2.

Next, we state the following theorem.

**Theorem 5** (Achievability bound on (11)–(13) for additive non-Gaussian noise process). *Suppose that  $\Delta_t = \Delta$ ,  $\forall t$  and assume that the users 1, 2 source models (1) are driven by non-Gaussian noise processes. Then, the augmented state space source model in (3) ensures the following achievability bounds on (11)–(13), as follows*

$$R_{\text{pd}}^c(D) \leq R_{\text{pd}}^{\text{LB}, \text{linear}}(D) + \frac{\text{rank}(H)}{2} \log(2\pi e G_{\text{rank}(H)}) + \mathbb{D}(\hat{\mathbf{x}} || \hat{\mathbf{x}}^G) \quad (80)$$

$$R_{\text{joint}}^c(D^*) \leq R_{\text{joint}}^{\text{LB}, \text{linear}}(D^*) + \frac{\text{rank}(H)}{2} \log(2\pi e G_{\text{rank}(H)}) + \mathbb{D}(\hat{\mathbf{x}} || \hat{\mathbf{x}}^G), \quad (81)$$

$$R^c(D_{\text{cov}}) \leq R^{\text{LB}, \text{linear}}(D_{\text{cov}}) + \frac{\text{rank}(H)}{2} \log(2\pi e G_{\text{rank}(H)}) + \mathbb{D}(\hat{\mathbf{x}} || \hat{\mathbf{x}}^G). \quad (82)$$

where  $\mathbb{D}(\hat{\mathbf{x}} || \hat{\mathbf{x}}^G)$  is the KL divergence between the residual source  $\hat{\mathbf{x}}_t$  assuming linear policies and the Gaussian residual source  $\hat{\mathbf{x}}_t^G \sim \mathcal{N}(0; \Lambda)$  in Figure 9.

**Proof.** We only prove (80) because (81) and (82) follow similarly. In addition, in parts we sketch the derivation because it is clear from the previous results. From Lemmas 4 and 3, we can easily obtain the following lower bound (similar to SLB [32])

$$R_{\text{pd}}^c(D) \geq R_{\text{pd}}^{\text{LB}, \text{linear}}(D) - \mathbb{D}(\hat{\mathbf{x}} || \hat{\mathbf{x}}^G), \quad (83)$$

where  $\mathbb{D}(\hat{\mathbf{x}} || \hat{\mathbf{x}}^G) \geq 0$  is the discrepancy between the residual source  $\hat{\mathbf{x}}_t$  assuming linear policies and the optimal Gaussian residual source  $\hat{\mathbf{x}}_t^G \sim \mathcal{N}(0; \Lambda)$ . From (83), we obtain

$$R_{\text{pd}}^{\text{LB}, \text{linear}}(D) \leq R_{\text{pd}}^c(D) + \mathbb{D}(\hat{\mathbf{x}} || \hat{\mathbf{x}}^G). \quad (84)$$

Then, applying the DPCM-based ECDQ scheme based on the linear forward test channel realization of Figure 9 discussed in Sections 3.1 and 3.2 we obtain (76). Since the coding scheme is obtained because we have assumed linear policies,  $R_{\text{pd}}^{\text{LB}}(D)$  will be replaced by  $R_{\text{pd}}^{\text{LB}, \text{linear}}(D)$ . This completes the derivation.  $\square$

**Remark 12** (Comments on Theorem 5). *Clearly, Theorem 5 is a generalization of Theorem 4 under the assumption of the linear realization of Figure 9 with systems driven by additive i.i.d. non-Gaussian noise process. If in (80)–(82), we assume that the system is driven by additive i.i.d. Gaussian noise process, then, clearly  $\mathbb{D}(\hat{\mathbf{x}} || \hat{\mathbf{x}}^G) = 0$  and Theorem 5 recovers Theorem 4.*

#### 4. Conclusions and Future Research

In this paper, we derived bounds on the OPTA of a two-user MIMO causal encoding and causal decoding problem (assuming the clocks of the encoder and the decoder to be synchronized). In our setup, each one of the users is described by a multivariate Markov source driven by additive i.i.d. noise process (possibly non-Gaussian) subject to three classes of spatio-temporal distortion constraints.

Although not directly pursued in this paper, all the results can be easily generalized to any finite number of users in Figure 1. Moreover, as a future research we aim to study the case of separate encoding for each user which will be a generalized version of a multi-user (distributed) source coding setup. Finally, due to the lack of insight of our results (mainly because we employed the general SDP solvers to compute our bounds) it makes sense to consider more specific setups and try to solve them using KKT conditions and, then, identify structural properties of matrices  $(A, \Sigma_w)$  for which KKT conditions can be optimally solved.

**Author Contributions:** Conceptualization, P.A.S., J.Ø. and M.S.; methodology, P.A.S., J.Ø. and M.S.; software, P.A.S.; validation, P. A. S, J.Ø. and M.S.; formal analysis, P. A. S; investigation, P.A.S.; resources, M.S.; data curation, P.A.S.; writing—original draft preparation, P.A.S.; writing—review and editing, P.A.S., J.Ø. and M.S.; visualization, P.A.S.; supervision, M.S.; project administration, M.S.; funding acquisition, M.S. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by KAW Foundation and the Swedish Foundation for Strategic Research.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

## Notation

The following notation is used in this manuscript:

Symbol	Description
$\mathbb{R}$	The set of real numbers
$\mathbb{Z}$	The set of integers
$\mathbb{N}_0$	The set of natural numbers including zero
$\mathbb{N}_0^n$	The set $\{0, \dots, n\}$ where $n \in \mathbb{N}_0$
$\mathbf{x} \in \mathbb{R}$	Random variable
$\mathcal{X}$	Alphabet for the random variable $\mathbf{x}$
$\mathbf{x}_r^t$	The sequence of random variables $(\mathbf{x}_r, \mathbf{x}_{r+1}, \dots, \mathbf{x}_t), (r, t) \in \mathbb{Z} \times \mathbb{Z}, r \leq t$
$x_r^t$	Sequence of the random variables realizations, where $x_r^t \in \mathcal{X}_r^t$
$\mathcal{X}_r^t$	$\times_{k=r}^t \mathcal{X}_k$ with $\mathcal{X}_t = \mathcal{X}$
$\mathbf{P}(dx)$	The probability of the RV $\mathbf{x}$ on $\mathcal{X}$
$\mathbf{P}(dy x)$	The conditional distribution of RV $\mathbf{y}$ given $\mathbf{x} = x$
$\otimes$	compound product
$\oplus$	Direct sum
$\mathbf{x} \in \mathbb{R}^{p \times 1}$	Column vector
$\mathbf{x}^T \in \mathbb{R}^{1 \times p}$	Row vector
$K \in \mathbb{R}^{p \times p}$	Square real matrix
$K^T \in \mathbb{R}^{p \times p}$	Transpose of square real matrix
$K_{ii}$	Diagonal elements of matrix $K$
$ K $	Determinant of $K$
$\text{rank}(K)$	Rank of $K$
$\text{trace}(K)$	Trace of $K$
$\mu_{K,i}$	the $i^{\text{th}}$ eigenvalue of matrix $K$
$\Sigma_{\mathbf{x}}$	The covariance of a random vector $\mathbf{x}$
$\Sigma_{\mathbf{x}} \succ 0$	Positive definite covariance matrix $\Sigma_{\mathbf{x}}$
$\Sigma_{\mathbf{x}} \succeq 0$	Positive semidefinite covariance matrix $\Sigma_{\mathbf{x}}$
$\Sigma_{\mathbf{x}} \succeq \Sigma_{\mathbf{x}'}$	$\Sigma_{\mathbf{x}} - \Sigma_{\mathbf{x}'}$ is positive semidefinite
$\Sigma_{\mathbf{x}} \succ \Sigma_{\mathbf{x}'}$	$\Sigma_{\mathbf{x}} - \Sigma_{\mathbf{x}'}$ is positive definite
$\mathbf{0}$	Null matrix
$I_p$	Identity matrix of dimension $p$
$H(\cdot)$	Discrete entropy
$h(\cdot)$	Differential entropy
$\mathbb{D}(\mathbf{x}  \mathbf{x}')$	KL Divergence of probability distribution $\mathbf{P}(x)$ with respect to probability distribution $\mathbf{P}(x')$
$\mathbf{x} \sim \mathcal{N}(0; \Sigma)$	Gaussian random vector $\mathbf{x}$ with zero mean and covariance $\Sigma$
$\mathbf{x} \sim \text{Unif}(0; \Sigma)$	Uniformly distributed random vector $\mathbf{x}$ with zero mean and covariance $\Sigma$
$h^G(\cdot)$	Gaussian differential entropy
$R^G(\cdot)$	Gaussian information RDF
$N(\mathbf{x})$	The entropy power of random vector $\mathbf{x}$
$\ \cdot\ _2$	Euclidean norm
$\mathbb{E}\{\cdot\}$	Expectation operator
$[\cdot]^+$	$\max\{0, \cdot\}$
$A \leftrightarrow B \leftrightarrow C$	$A, B, C$ form a Markov chain

## Abbreviations

The following abbreviations are used in this manuscript:

OPTA	Optimal Performance Theoretically Attainable
RDF	Rate distortion function
DPCM	Differential pulse coded modulation
ECDQ	Entropy coded dithered quantization
MSE	Mean-squared error
MMSE	Minimum MSE
RHS	Right Hand Side
LHS	Left Hand Side
i.i.d.	Independent Identically Distributed
a.s.	almost surely
KKT	Karush Kuhn Tucker
LMI	Linear matrix inequality
SDP	Semidefinite programming
KF	Kalman filter
EP	Entropy power
EPI	Entropy power inequalities
RL	Rate loss
SLB	Shannon Lower Bound

## Appendix A

**Proof of Lemma 3.** We only prove (1) as both (2), (3), follow similarly. First, by assumptions of the theorem,  $\Delta \triangleq \frac{1}{n+1} \sum_{t=0}^n \Delta_t$  for some finite  $n$  with  $\Delta \succ 0$  (sufficient condition for existence of a finite solution) and  $B \triangleq A^T \Sigma_{\mathbf{w}}^{-1} A \succeq 0$ . Moreover,

$$\begin{aligned}
 \frac{1}{n+1} \sum_{t=0}^n R_t &\stackrel{(a)}{\geq} \frac{1}{n+1} \sum_{t=0}^n I(\mathbf{x}_t; \mathbf{y}_t | \mathbf{y}^{t-1}) \\
 &\stackrel{(b)}{=} \frac{1}{n+1} \frac{1}{2} \sum_{t=0}^n \log \frac{|\Lambda_t|}{|\Delta_t|} \\
 &\stackrel{(c)}{=} \frac{1}{n+1} \left[ \frac{1}{2} \log |\Lambda_0| - \frac{1}{2} \log |\Lambda_{n+1}| + \frac{1}{2} \sum_{t=0}^n \log \frac{|\Lambda_{t+1}|}{|\Delta_t|} \right] \\
 &\stackrel{(d)}{=} \frac{1}{n+1} \left[ \frac{1}{2} \log |\Lambda_0| - \frac{1}{2} \log |\Lambda_{n+1}| + \frac{1}{2} \sum_{t=0}^n \left[ \log |\Sigma_{\mathbf{w}}| + \log |\Delta_t^{-1} + B| \right] \right] \\
 &\stackrel{(e)}{\geq} \frac{1}{n+1} \left[ \frac{1}{2} \log |\Lambda_0| - \frac{1}{2} \log \left( \frac{(\text{trace}(A^T A) \text{trace}(\widehat{D}) + \text{trace}(\Sigma_{\mathbf{w}}))^p}{p} \right) \right] \\
 &\quad + \frac{1}{2} \frac{1}{n+1} \sum_{t=0}^n \left[ \log |\Sigma_{\mathbf{w}}| + \log |\Delta_t^{-1} + B| \right] \\
 &\stackrel{(f)}{\geq} \frac{1}{n+1} \left[ \frac{1}{2} \log |\Lambda_0| - \frac{1}{2} \log \left( \frac{(\text{trace}(A^T A) \text{trace}(\widehat{D}) + \text{trace}(\Sigma_{\mathbf{w}}))^p}{p} \right) \right] \\
 &\quad + \frac{1}{2} \log |\Sigma_{\mathbf{w}}| + \frac{1}{2} \log |\Delta^{-1} + B| \\
 &\stackrel{(g)}{\geq} \frac{1}{n+1} \left[ \frac{1}{2} \log |\Lambda_0| - \frac{1}{2} \log \left( \frac{(\text{trace}(A^T A) \text{trace}(\widehat{D}) + \text{trace}(\Sigma_{\mathbf{w}}))^p}{p} \right) \right] + \frac{1}{2} \log \frac{|\Lambda|}{|\Delta|}, \quad (\text{A1})
 \end{aligned}$$

where (a) follows from Theorem 1 (see also Remark 2); (b) follows from Lemma 1, (1), because  $h^G(\mathbf{x}_t | \mathbf{y}^{t-1}) = \frac{1}{2} \log(2\pi e)^{p_1+p_2} |\Lambda_t|$  and  $h^G(\mathbf{x}_t | \mathbf{y}^t) = \frac{1}{2} \log(2\pi e)^{p_1+p_2} |\Delta_t|$ ; (c) follows by reformulating

the additive objective; (d) follows because  $|\Lambda_{t+1}||\Delta_t^{-1}| = |A\Delta_t A^T + \Sigma_w||\Delta_t^{-1}| = |\Sigma_w(\Sigma_w^{-1}A\Delta_t A^T + I_{p_1+p_2})||\Delta_t^{-1}| \stackrel{(d1)}{=} |\Sigma_w(A^T \Sigma_w^{-1} A \Delta_t + I_{p_1+p_2})||\Delta_t^{-1}| \stackrel{(d2)}{=} |\Sigma_w||B + \Delta_t^{-1}|$  where (d1) follows from Weinstein–Aronszajn identity ([49], Corollary 18.1.2) and (d2) from standard determinant properties of square matrices of the same size; (e) follows from the inequalities

$$\begin{aligned} |\Lambda_{n+1}| &\stackrel{(e1)}{\leq} \frac{(\text{trace}(A\Delta_n A^T + \Sigma_w))^p}{p} \\ &\stackrel{(e2)}{\leq} \frac{(\text{trace}(A\Delta_n A^T) + \text{trace}(\Sigma_w))^p}{p} \\ &\stackrel{(e3)}{\leq} \frac{(\text{trace}(A^T A) \text{trace}(\Delta_n) + \text{trace}(\Sigma_w))^p}{p} \\ &\stackrel{(e4)}{\leq} \frac{(\text{trace}(A^T A) \text{trace}(\widehat{D}) + \text{trace}(\Sigma_w))^p}{p}, \end{aligned}$$

where (e1) follows because  $|K|^{\frac{1}{p}} \leq \frac{\text{trace}(K)}{p}$  for  $K \succ 0$ , (e2) follows from ([18], Ex. 12.14), (e3) follows from the cycling property of trace and ([18], Ex. 12.14), (e4) follows because  $\text{trace}(\Delta_n) \leq \text{trace}(\widehat{D})$  (by definition); (f) follows because the term  $\log |\Delta_t^{-1} + B|$  is convex with respect to  $\Delta_t$  for  $B \succeq 0$  (see, e.g., [50]) hence we can apply Jensen’s inequality ([19], Theorem 2.6.2); (g) follows because  $\log |\Sigma_w| + \frac{1}{2} \log |\Delta^{-1} + B|$  can be rearranged to  $\frac{|\Lambda|}{|\Delta|}$  where  $\Lambda = A\Delta A^T + \Sigma_w$ . Taking the limit in both sides in (A1), we observe that the first RHS term vanishes asymptotically. Finally, using the appropriate infimization constraints in both sides of the limiting objective functions in (A1), and because we assume sufficient conditions for existence of solution of the RHS term infimum is in fact minimum and the result follows.  $\square$

## Appendix B

**Proof of Theorem 2.** Similar to the proof of Lemma 3, we only prove one case, i.e., (1), as the following cases can be shown using the exact same way.

By invoking matrix determinant lemma ([49], Theorem 18.1.1) in  $R_{\text{pd}}^{LB}(D)$ , we obtain

$$R_{\text{pd}}^{\text{na}}(D) = \min_{\substack{0 \prec \Delta \preceq \Lambda \\ \Delta_{ii} \leq D_{ii}, i=1, \dots, p}} \frac{1}{2} \log |\Sigma_w| - \frac{1}{2} \log |\Delta^{-1} - A^T \Sigma_w^{-1} A|. \quad (\text{A2})$$

Next, we introduce a decision variable  $Q_1 \triangleq \Delta^{-1} - A^T \Sigma_w^{-1} A$ . Using the monotonicity of the determinant we can rewrite (A2) as

$$R_{\text{pd}}^{\text{na}}(D) = \min_{\substack{0 \prec \Delta \preceq \Lambda \\ \Delta_{ii} \leq D_{ii}, i=1, \dots, p \\ 0 \prec Q_1 \preceq (\Delta^{-1} - A^T \Sigma_w^{-1} A)^{-1}}} \frac{1}{2} \log |\Sigma_w| - \frac{1}{2} \log |Q_1|. \quad (\text{A3})$$

Applying Woodbury matrix identity ([49], Theorem 18.2.8) in the inequality constraint  $0 \prec Q_1 \preceq (\Delta^{-1} - A^T \Sigma_w^{-1} A)^{-1}$ , we obtain

$$0 \prec Q_1 \preceq \Delta - \Delta A^T (\Sigma_w + A \Delta A^T)^{-1} A \Delta. \quad (\text{A4})$$

From Theorem 2 we have  $\Lambda = A \Delta A^T + \Sigma_w$ , hence (A4) is equivalent to the LMI condition of (49).

The decision variable is convex and there exists an optimal solution because  $\Delta \succ 0$ . This completes the proof.  $\square$

## Appendix C

**Proof of Theorem 3.** We only prove in detail (1) and sketch the proof for (2), (3) because several steps are identical to (1). First note that from Lemma 3 we have the bound in (43). Next, we simply show that the objective function can be lower bounded by a constant value (independent of the constraint set).

$$\begin{aligned}
 \frac{1}{2} \log \frac{|\Lambda|}{|\Delta|} &= \frac{1}{2} \log |A\Delta A^T + \Sigma_{\mathbf{w}}| - \frac{1}{2} \log |\Delta| \\
 &= \frac{(p_1 + p_2)}{2} \log |A\Delta A^T + \Sigma_{\mathbf{w}}|^{\frac{1}{p_1+p_2}} - \frac{(p_1 + p_2)}{2} \log |\Delta|^{\frac{1}{p_1+p_2}} \\
 &\stackrel{(a)}{\geq} \frac{(p_1 + p_2)}{2} \log \left( |A\Delta A^T|^{\frac{1}{p_1+p_2}} + |\Sigma_{\mathbf{w}}|^{\frac{1}{p_1+p_2}} \right) - \frac{(p_1 + p_2)}{2} \log |\Delta|^{\frac{1}{p_1+p_2}} \\
 &\stackrel{(b)}{=} \frac{(p_1 + p_2)}{2} \log \left( |A^T A|^{\frac{1}{p_1+p_2}} |\Delta|^{\frac{1}{p_1+p_2}} + |\Sigma_{\mathbf{w}}|^{\frac{1}{p_1+p_2}} \right) - \frac{(p_1 + p_2)}{2} \log |\Delta|^{\frac{1}{p_1+p_2}} \\
 &= \frac{(p_1 + p_2)}{2} \log \left( |A^T A|^{\frac{1}{p_1+p_2}} + |\Sigma_{\mathbf{w}}|^{\frac{1}{p_1+p_2}} |\Delta|^{-\frac{1}{p_1+p_2}} \right) \tag{A5}
 \end{aligned}$$

$$\begin{aligned}
 &\stackrel{(c)}{\geq} \frac{(p_1 + p_2)}{2} \log \left( |A^T A|^{\frac{1}{p_1+p_2}} + |\Sigma_{\mathbf{w}}|^{\frac{1}{p_1+p_2}} \frac{(p_1 + p_2)}{\text{trace}(\Delta)} \right) \\
 &\stackrel{(d)}{\geq} \frac{(p_1 + p_2)}{2} \log \left( |A^T A|^{\frac{1}{p_1+p_2}} + |\Sigma_{\mathbf{w}}|^{\frac{1}{p_1+p_2}} \frac{(p_1 + p_2)}{D} \right) \tag{A6}
 \end{aligned}$$

$$\stackrel{(e)}{\geq} \frac{(p_1 + p_2)}{2} \log \left( |A^T A|^{\frac{1}{(p_1+p_2)}} + \frac{(p_1 + p_2)N(\mathbf{w})}{D} \right), \tag{A7}$$

where (a) follows from Minkowski's determinant inequality ([18], Exercise 12.13); (b) follows from standard properties of determinants for square matrices of the same size; (c) follows from the reverse application of EPI (see, e.g., ([43], Equation (7))); (d) follows because  $\text{trace}(\Delta) \leq \text{trace}(\hat{D}) \equiv D$ ; (e) follows from ([43], Equation (7)). The constant value in (A7) is well defined if  $D \in \left( 0, \frac{(p_1+p_2)N(\mathbf{w})}{1-|A^T A|^{\frac{1}{p_1+p_2}}} \right]$ ,

where  $N(\mathbf{w}) = \frac{1}{2\pi e} 2^{\frac{2}{p_1+p_2} h(\mathbf{w})}$  and  $h(\mathbf{w}) > -\infty$ .

In (2) we follow similar steps to (1) but in inequality (d) we have instead  $\text{trace}(\Delta) \leq \min\{D_{\tau}, \text{trace}(\hat{D})\} \equiv D^*$ .

In (3) we follow similar steps until (A5). Afterwards, we leverage the fact that  $|\Delta| \leq |D_{\text{cov}}|$  to obtain instead of (A6) the following:

$$\frac{(p_1 + p_2)}{2} \log \left( |A^T A|^{\frac{1}{p_1+p_2}} + \frac{|\Sigma_{\mathbf{w}}|^{\frac{1}{p_1+p_2}}}{|D_{\text{cov}}|^{\frac{1}{p_1+p_2}}} \right).$$

This completes the derivation.  $\square$

## References

1. Caines, P.E. *Linear Stochastic Systems*; Wiley Series in Probability and Statistics; John Wiley & Sons, Inc.: New York, NY, USA, 1988.
2. Skogestad, S.; Postlethwaite, I. *Multivariable Feedback Control: Analysis and Design*, 2nd ed.; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2005.
3. Wang, J.; Chen, J. Vector Gaussian Multiterminal Source Coding. *IEEE Trans. Inf. Theory* **2014**, *60*, 5533–5552. [CrossRef]
4. Ekrem, E.; Ulukus, S. An Outer Bound for the Vector Gaussian CEO Problem. *IEEE Trans. Inf. Theory* **2014**, *60*, 6870–6887. [CrossRef]
5. Oohama, Y. Indirect and Direct Gaussian Distributed Source Coding Problems. *IEEE Trans. Inf. Theory* **2014**, *60*, 7506–7539. [CrossRef]

6. Rahman, M.S.; Wagner, A.B. Rate Region of the Vector Gaussian One-Helper Source-Coding Problem. *IEEE Trans. Inf. Theory* **2015**, *61*, 2708–2728. [[CrossRef](#)]
7. Zahedi, A.; Østergaard, J.; Jensen, S.H.; Naylor, P.A.; Bech, S. Source Coding in Networks With Covariance Distortion Constraints. *IEEE Trans. Signal Proc.* **2016**, *64*, 5943–5958. [[CrossRef](#)]
8. Vaseghi, S.V. *Multimedia Signal Processing: Theory and Applications in Speech, Music and Communications*; John Wiley & Sons: Chichester, UK, 2007.
9. Zahedi, A.; Østergaard, J.; Jensen, S.H.; Bech, S.; Naylor, P. Audio coding in wireless acoustic sensor networks. *Signal Process.* **2015**, *107*, 141–152. [[CrossRef](#)]
10. Linder, T.; Lagosi, G. A zero-delay sequential scheme for lossy coding of individual sequences. *IEEE Trans. Inf. Theory* **2001**, *47*, 2533–2538. [[CrossRef](#)]
11. Derpich, M.S.; Østergaard, J. Improved Upper Bounds to the Causal Quadratic Rate-Distortion Function for Gaussian Stationary Sources. *IEEE Trans. Inf. Theory* **2012**, *58*, 3131–3152. [[CrossRef](#)]
12. Kaspi, Y.; Merhav, N. Structure Theorems for Real-Time Variable Rate Coding With and Without Side Information. *IEEE Trans. Inf. Theory* **2012**, *58*, 7135–7153. [[CrossRef](#)]
13. Linder, T.; Yüksel, S. On Optimal Zero-Delay Coding of Vector Markov Sources. *IEEE Trans. Inf. Theory* **2014**, *60*, 5975–5991. [[CrossRef](#)]
14. Stavrou, P.A.; Østergaard, J.; Charalambous, C.D. Zero-Delay Rate Distortion via Filtering for Vector-Valued Gaussian Sources. *IEEE J. Sel. Top. Signal Process.* **2018**, *12*, 841–856. [[CrossRef](#)]
15. Gallager, R.G. *Information Theory and Reliable Communication*; Wiley: New York, NY, USA, 1968.
16. Tanaka, T.; Kim, K.K.K.; Parrilo, P.A.; Mitter, S.K. Semidefinite Programming Approach to Gaussian Sequential Rate-Distortion Trade-Offs. *IEEE Trans. Autom. Control* **2017**, *62*, 1896–1910. [[CrossRef](#)]
17. Khina, A.; Kostina, V.; Khisti, A.; Hassibi, B. Tracking and Control of Gauss-Markov Processes over Packet-Drop Channels with Acknowledgments. *IEEE Trans. Control Netw. Syst.* **2019**, *6*, 549–560. [[CrossRef](#)]
18. Abadir, K.M.; Magnus, J.R. *Matrix Algebra*; Cambridge University Press: New York, NY, USA, 2005.
19. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*, 2nd ed.; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2006.
20. Wyner, A. A definition of conditional mutual information for arbitrary ensembles. *Inf. Control* **1978**, *38*, 51–59. [[CrossRef](#)]
21. Tanaka, T.; Esfahani, P.M.; Mitter, S.K. LQG Control With Minimum Directed Information: Semidefinite Programming Approach. *IEEE Trans. Autom. Control* **2018**, *63*, 37–52. [[CrossRef](#)]
22. Massey, J.L. Causality, Feedback and Directed information. In Proceedings of the International Symposium on Information Theory and its Applications (ISITA '90), Waikiki, HI, USA, 27–30 November 1990; pp. 303–305.
23. Charalambous, C.D.; Stavrou, P.A. Directed Information on Abstract Spaces: Properties and Variational Equalities. *IEEE Trans. Inf. Theory* **2016**, *62*, 6019–6052. [[CrossRef](#)]
24. Stavrou, P.; Charalambous, T.; Charalambous, C.; Loyka, S. Optimal Estimation via Nonanticipative Rate Distortion Function and Applications to Time-Varying Gauss-Markov Processes. *SIAM J. Control Optim.* **2018**, *56*, 3731–3765. [[CrossRef](#)]
25. Ihara, S. *Information Theory—For Continuous Systems*; World Scientific: Singapore, 1993.
26. Gorbunov, A.K.; Pinsker, M.S. Nonanticipatory and Prognostic Epsilon Entropies and Message Generation Rates. *Problems Inf. Transmiss.* **1973**, *9*, 184–191.
27. Tatikonda, S.C. Control Under Communication Constraints. Ph.D. Thesis, Mass. Inst. of Tech. (M.I.T.), Cambridge, MA, USA, 2000.
28. Charalambous, C.D.; Stavrou, P.A.; Ahmed, N.U. Nonanticipative Rate Distortion Function and Relations to Filtering Theory. *IEEE Trans. Autom. Control* **2014**, *59*, 937–952. [[CrossRef](#)]
29. Tatikonda, S.; Sahai, A.; Mitter, S. Stochastic linear control over a communication channel. *IEEE Trans. Autom. Control* **2004**, *49*, 1549–1561. [[CrossRef](#)]
30. Silva, E.I.; Derpich, M.S.; Østergaard, J. A Framework for Control System Design Subject to Average Data-Rate Constraints. *IEEE Trans. Autom. Control* **2011**, *56*, 1886–1899. [[CrossRef](#)]
31. Stavrou, P.A.; Charalambous, T.; Charalambous, C.D. Finite-Time Nonanticipative Rate Distortion Function for Time-Varying Scalar-Valued Gauss-Markov Sources. *IEEE Control Syst. Lett.* **2018**, *2*, 175–180. [[CrossRef](#)]
32. Berger, T. *Rate Distortion Theory: A Mathematical Basis for Data Compression*; Prentice-Hall: Englewood Cliffs, NJ, USA, 1971.

33. Horn, R.A.; Johnson, C.R. (Eds.) *Matrix Analysis*, 2nd ed.; Cambridge University Press: New York, NY, USA, 2013.
34. Anderson, B.; Moore, J. *Optimal Filtering*; Prentice-Hall: Englewood Cliffs, NJ, USA, 1979.
35. Simon, D. *Optimal State Estimation: Kalman,  $H_\infty$ , and Nonlinear Approaches*; Wiley-Interscience: Hoboken, NJ, USA, 2006.
36. Grant, M.; Boyd, S. CVX: Matlab Software for Disciplined Convex Programming, Version 2.1. Available online: <http://cvxr.com/cvx> (accessed on 1 March 2014).
37. Boyd, S.; Vandenberghe, L. *Convex Optimization*; Cambridge University Press: New York, NY, USA, 2004.
38. Boyd, S.; El Ghaoui, L.; Feron, E.; Balakrishnan, V. *Linear Matrix Inequalities in System and Control Theory*; Society for Industrial and Applied Mathematics: Philadelphia, PA, USA, 1994.
39. Wang, J.; Chen, J.; Wu, X. On the Sum Rate of Gaussian Multiterminal Source Coding: New Proofs and Results. *IEEE Trans. Inf. Theory* **2010**, *56*, 3946–3960. [[CrossRef](#)]
40. Vandenberghe, L.; Boyd, S.; Wu, S.P. Determinant maximization with linear matrix inequality constraints. *SIAM J. Matrix Anal. Appl.* **1998**, *19*, 499–533. [[CrossRef](#)]
41. Stavrou, P.A.; Charalambous, T.; Charalambous, C.D.; Loyka, S.; Skoglund, M. Asymptotic Reverse-Waterfilling Characterization of Nonanticipative Rate Distortion Function of Vector-Valued Gauss-Markov Processes with MSE Distortion. In Proceedings of the 2018 IEEE Conference on Decision and Control (CDC), Beach, FL, USA, 17–19 December 2018; pp. 14–20.
42. Stavrou, P.A.; Østergaard, J.; Skoglund, M. On Zero-delay Source Coding of LTI Gauss-Markov Systems with Covariance Matrix Distortion Constraints. In Proceedings of the European Control Conference (ECC), Limassol, Cyprus, 12–15 June 2018; pp. 3083–3088.
43. Rioul, O. Information Theoretic Proofs of Entropy Power Inequalities. *IEEE Trans. Inf. Theory* **2011**, *57*, 33–55. [[CrossRef](#)]
44. Zamir, R. *Lattice Coding for Signals and Networks*; Cambridge University Press: Cambridge, UK, 2014.
45. Fuglsig, A.J.; Østergaard, J. Zero-delay Multiple descriptions of stationary scalar Gauss-Markov sources. *Entropy* **2019**, *21*, 1185. [[CrossRef](#)]
46. Tanaka, T.; Johansson, K.H.; Oechtering, T.; Sandberg, H.; Skoglund, M. Rate of prefix-free codes in LQG control systems. In Proceedings of the 2016 IEEE International Symposium on Information Theory (ISIT), Barcelona, Spain, 10–15 July 2016; pp. 2399–2403.
47. Zamir, R.; Feder, M. Information rates of pre/post-filtered dithered quantizers. *IEEE Trans. Inf. Theory* **1996**, *42*, 1340–1353. [[CrossRef](#)]
48. Conway, J.H.; Sloane, N.J.A. *Sphere-packings, Lattices, and Groups*, 3rd ed.; Springer: New York, NY, USA, 1999.
49. Harville, D.A. *Matrix Algebra From a Statistician's Perspective*; Springer: New York, NY, USA, 1997.
50. Kim, K.K.K. Optimization and Convexity of  $\log \det(I + KX^{-1})$ . *Int. J. Contr. Autom. Syst.* **2019**, *17*, 1067–1070. [[CrossRef](#)]

