2018 European Control Conference (ECC)
June 12-15, 2018. Limassol, Cyprus

On Zero-delay Source Coding of LTI Gauss-Markov Systems with Covariance Matrix Distortion Constraints

Photios A. Stavrou, Jan Østergaard, and Mikael Skoglund

***Abstract*— We introduce a new class of zero-delay source coding problems where a vector-valued Gauss-Markov source is conveyed subject to covariance matrix distortion constraints. We address this problem by defining an information theoretic measure where we minimize mutual information subject to causality constraints and covariance matrix distortion constraints. The resulting measure serves as a lower bound to the zero-delay rate distortion function (RDF). We solve this problem by showing that it is semidefinite representable and, thus, can be computed numerically. We also show that for this new class of information measures, it is possible to have achievable rates up to a constant space-filling loss due to a vector lattice quantizer and a constant loss due to entropy coding. We corroborate our framework with illustrative simulation examples.**

I. Introduction

Real-time source coding plays a crucial role in modern infrastructure. For instance, by the end of this decade, it is expected that billions of devices will be interconnected over Internet of things (IoT) technology. In this scenario, wireless sensor networks (WSNs) will form a unified operational component [1] communicating under much more stringent requirements on latency and reliability than what current standards can guarantee. In such large-scale sensing networks, there are several observations of the target data source due to the many sensors. If the observations at different nodes are correlated, then, to reduce the communication costs it is desirable to use the correlation among the measurements. Real-time source coding is required for encoding the data collected from sensors into fewer bits in an instant manner, and hence reducing energy and bandwidth consumption.

In applications where both instantaneous encoding and decoding are required, it is common to use the term *zero-delay source coding* instead of real-time source coding. Zero-delay source coding is particularly relevant in understanding and identifying the fundamental performance limitations of closed loop control systems over limited rate communication channels. During the last decade, significant progress has been made in both control and information theory communities to address this problem. Some indicative works on zero-delay source coding in control systems include [2]–[8]. All these works consider Gaussian or continuous alphabet linear time-varying (LTV) or linear time-invariant (LTI) systems where the performance objective is minimized subject to a scalar-valued target cost constraint.

In this paper, we consider a zero-delay source coding problem where a Gaussian source modeled as a fully observable time-invariant vector-valued Gauss-Markov process is compressed subject to a covariance matrix distortion constraint. We tackle the problem, by considering a new class of information theoretic measures where the target cost constraint is no more a scalar but a matrix. Specifically, to evaluate the performance of this dynamical system, we minimize mutual information subject to a causality constraint and a covariance matrix distortion constraint. This class of information measures is a variant of the nonanticipative rate distortion function (NRDF) that for Gaussian sources is shown to be a tight achievable lower bound to the operational zero-delay RDF subject to scalar-valued distortion constraints such as the total or per-letter MSE distortion constraint (see, e.g., [7]). We show that this class of information measures is achievable for Gauss-Markov sources by invoking standard techniques on entropy coded dithered quantizer (ECDQ) [9] followed by memoryless entropy coding.

The choice of covariance distortion constraint is not accidental. 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 [10]–[12] and signal processing [13]. In general, covariance matrix distortion constraints give rise to new issues compared to the scalar-valued distortion constraints, which make the problem harder to solve (for details see, e.g., [12]). Nonetheless, the argument for considering covariance distortion constraints despite its difficulty is its generality and the flexibility in formulating new problems. For example, in many centralized wireless sensor network setups, there is a central processing unit which makes an estimate of the transmitted source data at the individual nodes using the received signals. However, the signals conveyed to the processing unit are often corrupted versions of the actual signals due to the coding noise. In such cases, one can use the distortion matrices as design parameters to control or even minimize this noise at the processing center, i.e., maximize the signal to noise ratio at the processing unit.

The paper is structured as follows. In Section II we define the zero-delay Gaussian source-coding problem subject to covariance distortion constraint while in Section III we define a lower bound to it. In Section IV we give an achievable coding scheme based on ECDQ and entropy coding. In

P. A. Stavrou and M. Skoglund received funding by the KAW Foundation and the Swedish Foundation for Strategic Research. J. Østergaard received funding by the VILLUM FONDEN Young Investigator Programme, under grant agreement No. 19005.

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

J. Østergaard is with the Department of Electronic Systems, Aalborg University, Denmark, *emails:* jo@es.aau.dk.

Section V we present numerical experiments to support our framework. We draw conclusions and future directions in Section VI.

Notation: We use the following notation. $\mathbb{R}$ denotes the set of real numbers, $\mathbb{Z}$ the set of integers, $\mathbb{N}_0$ the set of natural numbers including zero, and $\mathbb{N}_0^n \triangleq \{0,\ldots,n\}, n\in\mathbb{N}_0$. We denote a probability space by $(\Omega,\mathcal{F},\mathbb{P})$ where Ω is the sample space, $\mathcal{F}$ is the $\sigma-$algebra of Ω, and $\mathbb{P}$ the probability function. A random variable (RV) defined on some probability space $(\Omega,\mathcal{F},\mathbb{P})$, is a map $X:\Omega\longmapsto\mathcal{X}$, where $(\mathcal{X},\mathcal{B}(\mathcal{X}))$ is a measurable space. We denote a sequence of RVs by $\mathbf{x}_r^t\triangleq(\mathbf{x}_r,\mathbf{x}_{r+1},\ldots,\mathbf{x}_t),(r,t)\in\mathbb{Z}\times\mathbb{Z},t\geq r$, and their values by $x_r^t\in\mathcal{X}_r^t\triangleq\times_{k=r}^t\mathcal{X}_k$, with $\mathcal{X}_k=\mathcal{X}$, for simplicity. If $r=-\infty$ and $t=-1$, we use the notation $\mathbf{x}_{-\infty}^{-1}=\mathbf{x}^{-1}$, and if $r=0$, we use the notation $\mathbf{x}_0^t=\mathbf{x}^t$. The distribution of the RV $\mathbf{x}$ on $\mathcal{X}$ is denoted by $\mathbf{P}_{\mathbf{x}}(dx)\equiv\mathbf{P}_{\mathbf{x}}$. The conditional distribution of RV $\mathbf{y}$ given $\mathbf{x}=x$ is denoted by $\mathbf{P}_{\mathbf{y}|\mathbf{x}}(dy|\mathbf{x}=x)\equiv\mathbf{P}_{\mathbf{y}|\mathbf{x}}$. The transpose of a matrix or vector K is denoted by K^{T}. For a square matrix $K\in\mathbb{R}^{p\times p}$ with entries K_{ij} on the i^{th} row and j^{th} column, we denote by $\operatorname{diag}\{K\}$ the matrix having $K_{ii},\ i=1,\ldots,p$, on its diagonal and zero elsewhere, its trace by $\operatorname{trace}\{K\}$, and its covariance by Σ_K. We denote the determinant of K by $|K|$. We denote by $K\succ 0$ (respectively, $K\succeq 0$) a positive-definite matrix (respectively, positive-semidefinite matrix). The statement $K\succeq T$ means that $K-T$ is positive semidefinite. We denote identity matrix by $I\in\mathbb{R}^{p\times p}$. We denote the time index with "t" and the dimension index with "i". We denote by $(\cdot)^G$ any RV or a vector that is Gaussian distributed. We denote by $h(\cdot)$ the differential entropy.

II. Problem Statement

In this paper, we consider the zero-delay source coding problem illustrated in Fig. II.1. In this setting, the $p-$dimensional (vector) Gaussian source is governed by the following fully observable discrete-time LTI Gauss-Markov state-space model

$$\mathbf{x}_{t+1}=A\mathbf{x}_t+\mathbf{w}_t,\tag{1}$$

where $A\in\mathbb{R}^{p\times p}$ is a deterministic matrix, $\mathbf{x}_0\in\mathbb{R}^p\sim\mathcal{N}(0;\Sigma_{\mathbf{s}_0})$ is the initial state with $\Sigma_{\mathbf{s}_0}\succ 0$, $\mathbf{w}_t\in\mathbb{R}^p\sim\mathcal{N}(0;\Sigma_{\mathbf{w}})$, $\Sigma_{\mathbf{w}}\succ 0$, is an independent Gaussian sequence, independent of $\mathbf{x}_0$. $\mathbf{x}_t\in\mathbb{R}^p$ is the observation process (observed by possible sensors). We allow A to have eigenvalues outside the unit circle which means that $\mathbf{x}_t$ can be unstable.

The system operates as follows. At every time step $t\in\mathbb{N}_0$, the *encoder* observes the vector source $\mathbf{x}_t$ and produces a single binary codeword $\mathbf{z}_t$ from a predefined set of codewords $\mathcal{Z}_t$ of at most a countable number of codewords. Since the source is random, $\mathbf{z}_t$ and its length $\mathbf{l}_t$ (in bits) are RVs. Upon receiving $\mathbf{z}^t$, the *decoder* produces the reconstruction $\mathbf{y}_t$ of the source sample. We assume that both the encoder and decoder process information without delay and they are allowed to have infinite memory of the past. The analysis of the noiseless digital channel is restricted to the class of instantaneous variable-length binary codes $\mathbf{z}_t$. The countable set of all codewords (codebook) $\mathcal{Z}_t$ is time-varying to allow the binary representation $\mathbf{z}_t$ to be an arbitrarily long sequence.

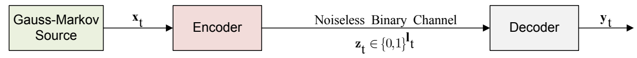

Fig. II.1: A zero-delay Gaussian source coding problem using variable-length binary codes.

Formally, the zero-delay source coding problem of Fig. II.1 can be explained as follows. Define the input and output alphabet of the noiseless digital channel by $\mathcal{M}=\{1,2,\ldots,M\}$ where $M=\max_t|\mathcal{Z}_t|$ (possibly infinite). The elements in $\mathcal{M}$ enumerate the codewords of $\mathcal{Z}_t$. The encoder is specified by the sequence of measurable functions $\{f_t:\ t\in\mathbb{N}_0\}$ with $f_t:\mathcal{M}^{t-1}\times\mathcal{X}^t\rightarrow\mathcal{M}$. At time t, the encoder transmits the message $z_t=f_t(z^{t-1},x^t)$ with $z_0=f_0(x_0)$. The decoder is specified by the sequence of measurable functions $\{g_t:\ t\in\mathbb{N}_0\}$ with $g_t:\mathcal{M}^t\rightarrow\mathcal{Y}_t$. For each $t\in\mathbb{N}_0$, the decoder generates $y_t=g_t(z^t)$ with $y_0=g_0(z_0)$.

The design in Fig. II.1 is required to operate within an asymptotic covariance matrix distortion constraint defined as follows:

$$\Sigma_{\mathbf{x}-\mathbf{y}}\triangleq\limsup_{n\longrightarrow\infty}\frac{1}{n+1}\Sigma_{\mathbf{x}^n-\mathbf{y}^n}\preceq D,\tag{2a}$$

$$\Sigma_{\mathbf{x}^n-\mathbf{y}^n}\triangleq\sum_{t=0}^n\mathbb{E}\left\{(\mathbf{x}_t-\mathbf{y}_t)(\mathbf{x}_t-\mathbf{y}_t)^{\mathrm{T}}\right\},\tag{2b}$$

where $D\succ 0$ and $D=D^{\mathrm{T}}$.

The objective is to minimize the expected average codeword length denoted by $\limsup_{n\longrightarrow\infty}\frac{1}{n+1}\sum_{t=0}^n\mathbb{E}(\mathbf{l}_t)$, over all measurable encoding and decoding functions $\{(f_t,g_t):\ t\in\mathbb{N}_0\}$. These design requirements are formally cast by the following optimization problem:

$$R_{\text{ZD}}^{\text{op}}(D)\triangleq\inf_{\substack{\{(f_t,\ g_t):\ t\in\mathbb{N}_0\}\\ \Sigma_{\mathbf{x}-\mathbf{y}}\preceq D}}\limsup_{n\longrightarrow\infty}\frac{1}{n+1}\sum_{t=0}^n\mathbb{E}(\mathbf{l}_t).\tag{3}$$

We refer to expression (3) as the *operational zero-delay Gaussian* RDF.

Alas, the solution of problem (3) is very hard to find because it is defined over all operational codes. For this reason, in the next section we introduce a lower bound to this problem which is defined based on information theoretic quantities.

III. Lower Bound on $R_{\text{ZD}}^{\text{op}}(D)$

In this section, we define a lower bound on (3) by minimizing mutual information subject to a causality constraint under the covariance matrix distortion constraint of (2). We use the term NRDF subject to covariance distortion constraints to characterize this information measure, hereinafter denoted by $R_{\text{cov}}^{\text{na}}(D)$.

First, notice that for general continuous alphabet sources, i.e., sources that are not necessarily Gaussian, it can be easily verified (it is a trivial extension of the inequalities in [7], [14]) that the following bounds hold:

$$R_{\text{cov}}(D) \le R^{\text{na}}_{\text{cov}}(D) \le R^{\text{op}}_{\text{ZD}}(D), \tag{4}$$

where $R_{\text{cov}}(D)$ denotes a variant of the classical RDF where instead of scalar-target distortion constraint we have the covariance matrix distortion constraint of (2).

Since, $R^{\text{na}}_{\text{cov}}(D)$ provides a tighter lower bound to $R^{\text{op}}_{\text{ZD}}(D)$ compared to $R_{\text{cov}}(D)$, we can use this information measure to obtain achievable bounds on (3).

We consider a data source that randomly generates sequences $\mathbf{x}_t = x_t \in \mathcal{X}_t, t \in \mathbb{N}_0^n$, that we wish to reproduce or reconstruct by $\mathbf{y}_t = y_t \in \mathcal{Y},\ t \in \mathbb{N}_0^n$, subject to the distortion criterion of (2).

Data Source. Suppose the data source generates sequences $\mathbf{x}^n = x^n, n \in \mathbb{N}_0$, randomly, according to the collection of distributions

$$\mathbf{P}_{\mathbf{x}_t|\mathbf{x}^{t-1},\mathbf{y}^{t-1}} \triangleq \mathbf{P}(dx_t|x^{t-1}), \quad t \in \mathbb{N}_0^n. \tag{5}$$

At time $t = 0$, we assume $\mathbf{P}_{\mathbf{x}_0|\mathbf{x}^{-1},\mathbf{y}^{-1}} \triangleq \mathbf{P}(dx_0)$. Thus, given the conditional distributions $\mathbf{P}(\cdot|\cdot,\cdot)$ in (5), by Bayes' rule we can formally define the joint distribution on $\mathcal{X}^n$ by $\mathbf{P}_{\mathbf{x}^n} \equiv \mathbf{P}(dx^n) \triangleq \otimes_{t=0}^n \mathbf{P}(dx_t|x^{t-1})$.

Reproduction or "test-channel". Suppose the reproduction $\mathbf{y}^n = y^n, n \in \mathbb{N}_0$ of x^n is randomly generated, according to the collection of conditional distributions, known as test-channels, by

$$\mathbf{P}_{\mathbf{y}_t|\mathbf{y}^{t-1},\mathbf{x}^t} \triangleq \mathbf{P}(dy_t|y^{t-1},x^t), \quad t \in \mathbb{N}_0^n. \tag{6}$$

At $n = 0$, we assume $\mathbf{P}_{\mathbf{y}_0|\mathbf{y}^{-1},\mathbf{x}^0}(dy_0|y^{-1},x^0) = \mathbf{P}(dy_0|x_0)$.

From [15], we know that the conditional distributions $\mathbf{P}(dy_t|y^{t-1},x^t)$ in (6), uniquely define the family of conditional distributions on $\mathcal{Y}^n$ parametrized by $x^n \in \mathcal{X}^n$, given by $\overrightarrow{\mathbf{P}}(dy^n|x^n) \triangleq \otimes_{t=0}^n \mathbf{P}(dy_t|y^{t-1},x^t)$, and vice-versa. By (5) and (6), we can uniquely define the joint distribution of $\{(\mathbf{x}^n,\mathbf{y}^n):\ t \in \mathbb{N}_0^n\}$ by

$$\mathbf{P}_{\mathbf{x}^n,\mathbf{y}^n} \equiv \mathbf{P}(dx^n,dy^n) = \mathbf{P}(dx^n) \otimes \overrightarrow{\mathbf{P}}(dy^n|x^n). \tag{7}$$

From (7), we can uniquely define the $\mathcal{Y}^n-$marginal distribution by $\mathbf{P}_{\mathbf{y}^n} \equiv \mathbf{P}(dy^n) \triangleq \int_{\mathcal{X}^n} \mathbf{P}(dx^n) \otimes \overrightarrow{\mathbf{P}}(dy^n|x^n)$, and the conditional distributions $\mathbf{P}_{\mathbf{y}_t|\mathbf{y}^{t-1}} = \mathbf{P}(dy_n|y^{n-1})$, $t \in \mathbb{N}_0^n$.

Given the above construction of distributions, we introduce the mutual information between $\mathbf{x}^n$ to $\mathbf{y}^n$ as follows:

$$\begin{aligned}
I(\mathbf{x}^n;\mathbf{y}^n) &\overset{(a)}{=} \int_{\mathcal{X}^n\times\mathcal{Y}^n} \log\left(\frac{\overrightarrow{\mathbf{P}}(\cdot|x^n)}{\mathbf{P}(\cdot)}(\mathbf{y}^n)\right)\mathbf{P}(dx^n,dy^n)\\
&\overset{(b)}{=} \sum_{t=0}^n \mathbb{E}\left\{\log\left(\frac{\mathbf{P}(\cdot|\mathbf{y}^{t-1},\mathbf{x}^t)}{\mathbf{P}(\cdot|\mathbf{y}^{t-1})}(\mathbf{y}_t)\right)\right\}\\
&\overset{(c)}{=} \sum_{t=0}^n I(\mathbf{x}^t;\mathbf{y}_t|\mathbf{y}^{t-1}),
\end{aligned}$$

where (a) is due to the Radon-Nikodym derivative theorem [16]; (b) due to chain rule of relative entropy; (c) follows by definitions of (5) and (6) which are equivalent to conditional independence (or causality constraint)

$$\mathbf{P}(dy_t|y^{t-1},x^n) = \mathbf{P}(dy_t|y^{t-1},x^t),\ \forall(x^n,y^{t-1}),\ t \in \mathbb{N}_0^n.$$

We now state the definition of NRDF subject to a covariance matrix distortion constraint.

Definition 1: (NRDF subject to a covariance matrix distortion constraint) Assume the covariance distortion constraint of (2). Then,

(1) the finite-time NRDF is defined as:

$$R^{\text{na}}_{\text{cov},0,n}(D) \triangleq \inf_{\substack{\mathbf{P}(dy_t|y^{t-1},x^t):\ t=0,\ldots,n\\ \frac{1}{n+1}\Sigma_{\mathbf{x}^n-\mathbf{y}^n}\preceq D}} \frac{1}{n+1} I(\mathbf{x}^n;\mathbf{y}^n); \tag{8}$$

(2) the per unit time asymptotic limit of (8) is defined as:

$$R^{\text{na}}_{\text{cov}}(D) = \lim_{n\longrightarrow\infty} R^{\text{na}}_{\text{cov},0,n}(D), \tag{9}$$

assuming the limit exists.

By replacing $\lim\inf$ with $\inf\lim$ in (9), then an upper bound to $R^{\text{na}}_{\text{cov}}(D)$ is obtained, defined as follows:

$$\bar{R}^{\text{na}}_{\text{cov}}(D) \triangleq \inf_{\overrightarrow{\mathbf{P}}(dy^\infty|x^\infty)} \lim_{n\longrightarrow\infty} R^{\text{na}}_{\text{cov},0,n}(D), \tag{10}$$

where $\overrightarrow{\mathbf{P}}(dy^\infty|x^\infty)$ denotes the sequence of conditional probability distributions $\{\mathbf{P}(dy_t|y^{t-1},x^t):\ t \in \mathbb{N}_0\}$.

Following [17, Theorem 6.6], if the limit in (10) exists, and the source is stationary (or asymptotically stationary), then, $R^{\text{na}}_{\text{cov}}(D) = \bar{R}^{\text{na}}_{\text{cov}}(D)$.

The optimization problem of Definition 1, in contrast to the one given in (3) is convex and there exists an optimal solution characterizing it (assuming a non-zero matrix distortion) (for details see, e.g., [15]). In addition, assuming the source model of (1) and the covariance matrix distortion of (2), then, a direct implication of [5, Theorem 1], is that the optimal "test channel" corresponding to (9) is of the form

$$\mathbf{P}^*(dy_t|y^{t-1},x^t) = \mathbf{P}^*(dy_t|y^{t-1},x_t),\ t \in \mathbb{N}_0. \tag{11}$$

In addition, the corresponding joint process $\{(\mathbf{x}_t,\mathbf{y}_t):\ t \in \mathbb{N}_0\}$ is jointly Gaussian. We note that (11) essentially parallels the prior work of [18] using, however, a different problem formulation to derive the same structural result.

A. Realization of the Optimal Test-Channel

In this section, we recall the realization scheme of [5, Theorem 2] and we give conditions to ensure asymptotic stationarity of this model. Further, we give the asymptotic values of the statistics and the minimization problem corresponding to the minimum data rate required subject to a covariance matrix distortion constraint for stabilizing this specific model in the asymptotic regime.

The authors of [5, Theorem 2] realized the optimal "test channel" of (11) with the feedback realization scheme illustrated in [5, Fig. IV.3] that corresponds to a realization of the form:

$$\mathbf{y}_t = \tilde{H}_t(\mathbf{x}_t - \widehat{\mathbf{x}}_{t|t-1}) + E_t^{\text{T}}\Theta_t\mathbf{v}_t + \widehat{\mathbf{x}}_{t|t-1}, \tag{12}$$

where $\tilde{H}_t \triangleq E_t^{\mathrm{T}} H_t E_t$ with $H_t \triangleq \Phi_t \Theta_t$; $\mathbf{v}_t$ is an independent Gaussian noise process with $N(0;\Sigma_{\mathbf{v}_t})$, $\Sigma_{\mathbf{v}_t} = \mathrm{diag}\{V_t\}$ independent of $\mathbf{x}_0$; the error $\mathbf{x}_t - \widehat{\mathbf{x}}_{t|t-1}$ is Gaussian with $N(0;\Sigma_{t|t-1})$, and $\widehat{\mathbf{x}}_{t|t-1} \triangleq \mathbb{E}\{\mathbf{x}_t|\mathbf{y}^{t-1}\}$; the error $\mathbf{x}_t - \widehat{\mathbf{x}}_{t|t}$ is Gaussian with $N(0;\Sigma_{t|t})$, and $\widehat{\mathbf{x}}_{t|t} \triangleq \mathbb{E}\{\mathbf{x}_t|\mathbf{y}^t\}$. Moreover, $\{\widehat{\mathbf{x}}_{t|t-1},\ \Sigma_{t|t-1}\}$ are given by the following Kalman filter recursions:

$$\textit{Prediction:}\ \widehat{\mathbf{x}}_{t|t-1} = A\widehat{\mathbf{x}}_{t-1|t-1}, \quad \widehat{\mathbf{x}}_{0|-1} = \mathbb{E}\{\mathbf{x}_0\}, \tag{13a}$$

$$\Sigma_{t|t-1} = A\Sigma_{t-1|t-1}A^{\mathrm{T}} + \Sigma_{\mathbf{w}}, \quad \Sigma_{0|-1} = \Sigma_{\mathbf{x}_0}, \tag{13b}$$

$$\textit{Update:}\ \widehat{\mathbf{x}}_{t|t} = \widehat{\mathbf{x}}_{t|t-1} + G_t\tilde{\mathbf{k}}_t, \quad \mathbf{x}_{0|0} = \mathbb{E}\{\mathbf{x}_0\}, \tag{13c}$$

$$\tilde{\mathbf{k}}_t \triangleq \mathbf{y}_t - \widehat{\mathbf{x}}_{t|t-1} \quad \text{(innovation)}, \tag{13d}$$

$$\Sigma_{t|t} = \Sigma_{t|t-1} - G_t S_t G_t^{\mathrm{T}}, \quad \Sigma_{0|0} = \Sigma_{\mathbf{x}_0}, \tag{13e}$$

$$G_t = \Sigma_{t|t-1}\tilde{H}_t^{\mathrm{T}} S_t^{-1} \quad \text{(Kalman Gain)},$$

$$S_t = \tilde{H}_t\Sigma_{t|t-1}\tilde{H}_t^{\mathrm{T}} + E_t^{\mathrm{T}}\Theta_t\Sigma_{\mathbf{v}_t}\Theta_t^{\mathrm{T}}E_t,$$

where $\Sigma_{t|t-1} = \Sigma_{t|t-1}^{\mathrm{T}} \succ 0$ and $\Sigma_{t|t} = \Sigma_{t|t}^{\mathrm{T}} \succ 0$ and $E_t \in \mathbb{R}^{p\times p}$ is the joint diagonalizer of $\Sigma_{t|t},\ \Sigma_{t|t-1},\ \tilde{H}_t$, that is, $\Sigma_{t|t} = E_t^{\mathrm{T}}\Delta_t E_t$, $\Sigma_{t|t-1} = E_t^{\mathrm{T}}\Lambda_t E_t$. In addition, we choose

$$H_t \triangleq I - \Delta_t\Lambda_t^{-1} \succeq 0,\ \Delta_t \triangleq \mathrm{diag}\{\delta_t\},\ \Lambda_t \triangleq \mathrm{diag}\{\lambda_t\}, \tag{14a}$$

$$\Theta_t \triangleq \sqrt{H_t\Delta_t\Sigma_{\mathbf{v}_t}^{-1}} \succeq 0,\ \Phi_t \triangleq \sqrt{H_t\Delta_t^{-1}\Sigma_{\mathbf{v}_t}} \succeq 0, \tag{14b}$$

with $\Lambda_t \succ 0,\ \Delta_t \succ 0$ such that $\Lambda_t \succeq \Delta_t \succ 0$. It is easy to verify that for the choice of (14), the following hold:

$$\widehat{\mathbf{x}}_{t|t-1} = A\mathbf{y}_{t-1},\ \widehat{\mathbf{x}}_{t|t} = \mathbf{y}_t,\ G_t = I. \tag{15}$$

By substituting (15) in (12), we can also observe that $\mathbf{P}^*(dy_t|y^{t-1},x_t) = \mathbf{P}^*(dy_t|y_{t-1},x_t)$.

We assume that $\Sigma \triangleq \lim_{t\longrightarrow\infty}\Sigma_{t|t-1} < \infty$, $\Sigma' \triangleq \lim_{t\longrightarrow\infty}\Sigma_{t|t} < \infty$. This means that $E_t \equiv E$, $H_t \equiv H$, $\Lambda_t \equiv \Lambda$, $\Delta_t \equiv \Delta$, and $\Sigma_{\mathbf{v}_t} \equiv \Sigma_{\mathbf{v}}$. The resulting asymptotically stationary feedback realization scheme is depicted in Fig. III.2.

Next, we briefly discuss the basic features of this scheme.

Encoder: Introduce the estimation error $\{\mathbf{k}_t \in \mathbb{R}^p:\ t \in \mathbb{N}_0\}$, where $\mathbf{k}_t \triangleq \mathbf{x}_t - \widehat{\mathbf{x}}_{t|t-1},\ t \in \mathbb{N}_0$ with covariance $\Sigma_{t|t-1}$, $t \in \mathbb{N}_0$. The corresponding asymptotic value Σ is diagonalized by the joint diagonalizer E (invertible matrix) such that $E^{\mathrm{T}}\Sigma E = \mathrm{diag}\{\lambda\} \triangleq \Lambda$. Φ is the asymptotic limit of diagonal matrix Φ_t.

Decoder: Analogously, we introduce the innovations process $\{\tilde{\mathbf{k}}_t:\ t \in \mathbb{N}_0\}$ defined by (13d). Θ is the asymptotic limit of diagonal matrix Θ_t.

Channel: The AWGN channel is of the form $\boldsymbol{\beta}_t = \alpha_t + \mathbf{v}_t = \Phi E\mathbf{k}_t + \mathbf{v}_t,\quad \mathbf{v}_t \sim \mathcal{N}(0;\Sigma_{\mathbf{v}})$, $\Sigma_{\mathbf{v}} = \mathrm{diag}\{V\}$, $V_{ii} = \tilde{V}$, $i = 1,\ldots,p$, $t \in \mathbb{N}_0$.

The covariance matrix $\Sigma_{\mathbf{x}_t-\mathbf{y}_t} \triangleq \mathbb{E}\{(\mathbf{x}_t-\mathbf{y}_t)(\mathbf{x}_t-\mathbf{y}_t)^{\mathrm{T}}\}$ at each t is the same as $\Sigma_{\mathbf{k}_t-\tilde{\mathbf{k}}_t} \triangleq \mathbb{E}\{(\mathbf{k}_t-\tilde{\mathbf{k}}_t)(\mathbf{k}_t-\tilde{\mathbf{k}}_t)^{\mathrm{T}}\}$.

Furthermore, the steady state values of covariance matrices in (13b), (13e) are

$$\Sigma = A\Sigma' A^{\mathrm{T}} + \Sigma_{\mathbf{w}}, \quad \Sigma' = E^{\mathrm{T}}\Delta E. \tag{16}$$

The total covariance matrix distortion constraint of the scheme in Fig. III.2 is

$$\lim_{n\longrightarrow\infty}\frac{1}{n+1}\sum_{t=0}^{n}\mathbb{E}\{(\mathbf{x}_t-\mathbf{y}_t)(\mathbf{x}_t-\mathbf{y}_t)^{\mathrm{T}}\} \preceq D. \tag{17}$$

Hence, we can now characterize $R_{\mathrm{cov}}^{\mathrm{na}}(D)$ of an LTI vector-valued Gauss-Markov source modeled as in (1), subject to the covariance matrix distortion constraint (2).

Lemma 1: (Characterization of $R_{\mathrm{cov}}^{\mathrm{na}}(D)$) Consider the asymptotically stationary feedback realization scheme of Fig. III.2. Then, the asymptotic limit $R_{\mathrm{cov}}^{\mathrm{na}}(D)$ of an LTI vector Gauss-Markov source modeled as in (1) subject to the distortion constraint (2), can be cast to the following minimization problem:

$$R_{\mathrm{cov}}^{\mathrm{na}}(D) = \min_{\substack{0\prec\Sigma'\preceq\Sigma\\ \Sigma'\preceq D}}\frac{1}{2}\log\frac{|\Sigma|}{|\Sigma'|}, \tag{18}$$

$$\equiv \min_{\substack{0\prec\Sigma'\preceq\Sigma\\ \Sigma'\preceq D}}\frac{1}{2}\log\frac{|A\Sigma'A^{\mathrm{T}}+\Sigma_{\mathbf{w}}|}{|\Sigma'|}. \tag{19}$$

Proof: The proof is straightforward and we omit it due to space limitations. ■

The minimization problem of Lemma 1 is a max-det optimization problem that is convex and can be solved via interior point methods [19]. In the next theorem, we invoke one such technique based on semidefinite programming [19] by showing that the optimization problem of Lemma 1 is semidefinite representable and hence it can be readily evaluated numerically.

Theorem 1: (Optimal Solution of $R_{\mathrm{cov}}^{\mathrm{na}}(D)$) Define $F \triangleq (\Sigma')^{-1} - A^{\mathrm{T}}\Sigma_{\mathbf{w}}^{-1}A \succ 0$ where $\Sigma' \succ 0$. Then, for $D \succ 0$, $R_{\mathrm{cov}}^{\mathrm{na}}(D)$ is semidefinite representable as follows:

$$R_{\mathrm{cov}}^{\mathrm{na}}(D) = \min_{F\succ 0} -\frac{1}{2}\log|F| + \frac{1}{2}\log|\Sigma_{\mathbf{w}}|. \tag{20a}$$

$$\text{s.t.}\ 0 \prec \Sigma' \preceq \Sigma \tag{20b}$$

$$\Sigma' \preceq D \tag{20c}$$

$$\begin{bmatrix} \Sigma' - F & \Sigma' A^{\mathrm{T}} \\ A\Sigma' & \Sigma \end{bmatrix} \succeq 0 \tag{20d}$$

Proof: The proof is omitted due to space limitations. ■

IV. Achievable Coding Scheme

In this section we explain in brief the case of encoding the vector Gaussian source modeled by (1) based on ECDQ [9] followed by memoryless entropy coding.

First, recall that the optimal realization of the "test-channel" in Fig. III.2 implies that the mutual information from $\mathbf{x}_t$ to $\mathbf{y}_t$ given $\mathbf{y}_{t-1}$ is the same as the one from $\mathbf{k}_t$ to $\tilde{\mathbf{k}}_t$, i.e.,

$$I(\mathbf{x}_t;\mathbf{y}_t|\mathbf{y}_{t-1}) = I(\mathbf{k}_t;\tilde{\mathbf{k}}_t), \quad \forall t.$$

In the encoder, the vector Gauss-Markov source $\mathbf{x}_t$ is shaped into the estimation error process $\mathbf{k}_t$ (see Section (III)) shifted and scaled by the joint diagonalizer E and the diagonal matrix Φ, respectively, before encoded by an

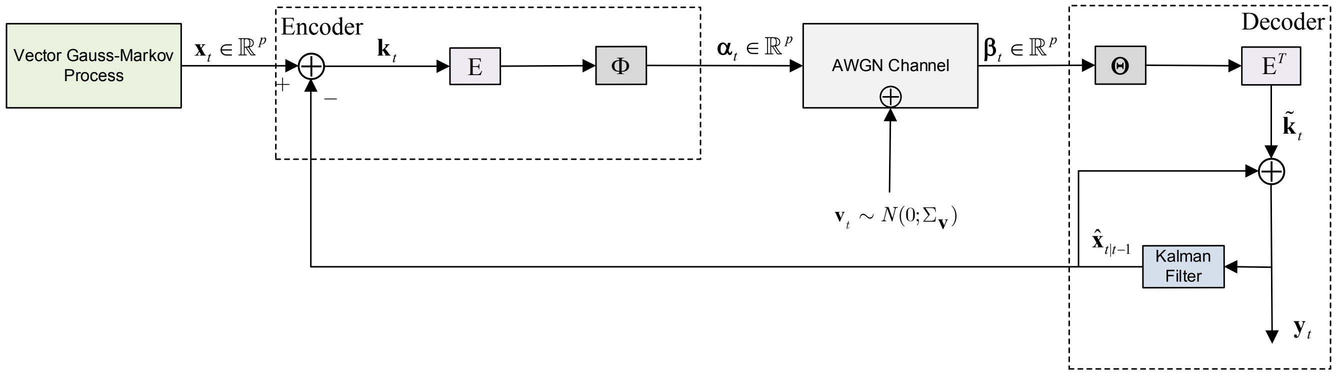

Fig. III.2: Realization of the optimal "test-channel" $\mathbf{P}^*(dy_t|y_{t-1},x_t)$ corresponding to a vector-valued Gauss-Markov source operating subject to a covariance matrix distortion constraint.

(unbounded) lattice quantizer. The reconstruction of $\mathbf{k}_t$ is the innovations process $\tilde{\mathbf{k}}_t$ given as follows:

$$\tilde{\mathbf{k}}_t = E^{\mathrm{T}}\Theta(\Phi E\mathbf{k}_t + \mathbf{v}_t),\ \ \mathbf{v}_t \sim \mathcal{N}(0;\Sigma_{\mathbf{v}}), \tag{21}$$

where $\Sigma_{\mathbf{v}} = \mathrm{diag}\{V\}$, $V_{ii} = \tilde{V}$, $\forall t$. The proposed vector scheme reconstructs $\tilde{\mathbf{k}}_t$ with a $p-$dimensional lattice quantizer Q_p, such that

$$\mathbb{E}\{\mathbf{r}_t\mathbf{r}_t^{\mathrm{T}}\} = \Sigma_{\mathbf{v}}, \tag{22}$$

where $\mathbf{r}_t$ is the dither vector which is uniformly distributed over the basic cell of the lattice. The reconstruction of $\mathbf{k}_t$ is then

$$\tilde{\mathbf{k}}_t = E^{\mathrm{T}}\Theta(Q_p(\Phi E\mathbf{k}_t + \mathbf{r}_t) - \mathbf{r}_t), \tag{23}$$

which corresponds to a coding rate expressed as the conditional entropy of the lattice quantizer given by

$$H(Q_p(\Phi E\mathbf{k}_{\mathbf{t}} + \mathbf{r}_t)|\mathbf{r}_t)\ \ \text{bits/sample}. \tag{24}$$

Entropy coding: We apply entropy coding across the vector dimension. To this end, at each time step t the output of the quantizer is subjected to the dither to generate a codeword $\mathbf{z}_t$. The decoder reproduces $\tilde{\mathbf{k}}_t$ by subtracting the dither signal $\mathbf{r}_t$ from $\tilde{\mathbf{k}}_t$. Specifically, at every time step t, we require that a message $\tilde{\mathbf{k}}_t$ is mapped into a codeword $\mathbf{z}_t \in \{0,1\}^{\mathbf{l}_t}$ designed using Huffman codes or other optimal symbol-by-symbol codes [20, Chapter 5.4].

Next, we state the main theorem of this section.

Theorem 2: (Achievable upper bound to $R^{\mathrm{op}}_{\mathrm{ZD}}(D)$) For the previously discussed coding scheme, the covariance matrix distortion constraint of the vector Gaussian source modeled by (1), is equal to the distortion matrix associated with the realization scheme of Fig. III.2. Moreover, the normalized per dimension zero-delay RDF is upped bounded as follows:

$$R^{\mathrm{op}}_{\mathrm{ZD}}(D) \leq R^{\mathrm{na}}_{\mathrm{cov}}(D) + \frac{\mathrm{rank}(H)}{2}\log_2(2\pi e G_p) + 1 \tag{25}$$

where $\mathrm{rank}(H)$ is the rank of matrix H and G_p is the normalized second moment of the lattice [9].

Proof: The derivation is quite similar to the proof of [7, Theorem 1], hence we omit it. ■

We note that for uniform scalar (i.e., $p=1$) quantization with subtractive dithering, then, $G_1 = \frac{1}{12}$. This gives the following upper bound to $R^{\mathrm{op}}_{\mathrm{ZD}}(D)$ (normalized per dimension):

$$R^{\mathrm{op}}_{\mathrm{ZD}}(D) \leq R^{\mathrm{na}}_{\mathrm{cov}}(D) + \frac{1}{2}\log_2\left(\frac{\pi e}{6}\right) + 1. \tag{26}$$

V. Numerical Example

In what follows, we present examples to illustrate various features of $R^{\mathrm{na}}_{\mathrm{cov}}(D)$.

Example 1: (Covariance matrix distortion constraint vs. MSE distortion constraint)
Consider a stable $\mathbb{R}^2$-valued Gauss-Markov source modeled by (1) with parameters:

$$(A,\ \Sigma_{\mathbf{w}}) = \left(\begin{bmatrix}0.5 & 0.2\\ 0.3 & 0.6\end{bmatrix}, \begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}\right). \tag{27}$$

We consider the following two types of distortion constraint.

(i) The covariance matrix distortion constraint given by:

$$D = \begin{bmatrix}1.5 & \gamma\\ \gamma & 1.5\end{bmatrix}, \tag{28}$$

where $\gamma \in \mathbb{R}$ is the correlation coefficient between the distortion matrix components (i.e., diagonal entries) and it is chosen such that $D \succ 0$ and $R^{\mathrm{na}}_{\mathrm{cov}}(D) < \infty$.

(ii) The "restriction" of (28) to a MSE distortion constraint given by:

$$D_{\mathrm{MSE}} = \mathrm{trace}(D) = 3. \tag{29}$$

This means that in (20c) the constraint becomes

$$\mathrm{trace}\,(\Sigma') \leq D_{\mathrm{MSE}}. \tag{30}$$

For this example, the NRDF of vector-valued Gauss-Markov sources subject to the scalar-valued MSE distortion constraint (30) is denoted by $R^{\mathrm{na}}_{\mathrm{MSE}}(D)$.

In what follows, we consider the $\mathbb{R}^2-$valued Gauss-Markov source with the parameters of (27) to evaluate $R^{\mathrm{na}}_{\mathrm{cov}}(D)$ for different values of the correlation coefficient γ in (28). Using the same source, we also evaluate $R^{\mathrm{na}}_{\mathrm{MSE}}(D)$ for the case of MSE distortion constraint of (29). The computation of both $R^{\mathrm{na}}_{\mathrm{cov}}(D)$ and $R^{\mathrm{na}}_{\mathrm{MSE}}(D)$ is established

TABLE I: $R_{\mathrm{cov}}^{\mathrm{na}}(D)$ and $R_{\mathrm{MSE}}^{\mathrm{na}}(D)$ for different values of γ.

test	γ	$R_{\mathrm{cov}}^{\mathrm{na}}(D)$ (bits/sample)	$R_{\mathrm{MSE}}^{\mathrm{na}}(D)$ (bits/sample)
1	0	0.31	0.1155
2	0.1	0.242	0.1155
3	-0.1	0.3982	0.1155
4	0.2	0.1877	0.1155
5	-0.2	0.5180	0.1155
6	0.3	0.1495	0.1155
7	-0.3	0.6936	0.1155
8	0.4	0.141	0.1155
9	-0.4	0.9881	0.1155

by invoking the SDP solver CVX platform [21] that contains built-in functions to handle log-determinant optimization problems.

In Table I we demonstrate a comparison between $R_{\mathrm{cov}}^{\mathrm{na}}(D)$ evaluated for several different values of γ and $R_{\mathrm{MSE}}^{\mathrm{na}}(D)$ evaluated using (29).

Based on the numerical results of Table I, we can observe the following:

(a) evaluating $R_{\mathrm{cov}}^{\mathrm{na}}(D)$ with negative correlation between the distortion matrix components can cause extra redundancy on the information rates, i.e., the information rates for positive γ are smaller compared to the ones for negative γ;

(b) restricting the distortion constraint of (28) to the MSE distortion constraint of (30) is as if we optimize via a solution space in which γ 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 the MSE distortion constraint rather than the covariance matrix distortion constraint. For this reason, we always have $R_{\mathrm{MSE}}^{\mathrm{na}}(D) \leq R_{\mathrm{cov}}^{\mathrm{na}}(D)$.

In Fig. V.3 we illustrate the behavior of $R_{\mathrm{cov}}^{\mathrm{na}}(D)$ as a function of γ subject to the distortion covariance matrix of (28).

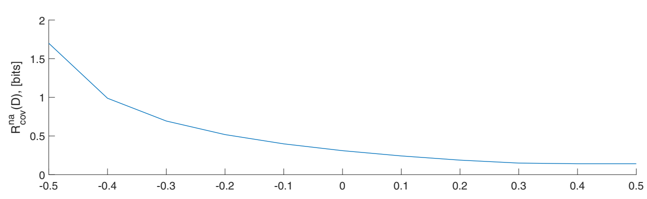

Fig. V.3: $R_{\mathrm{cov}}^{\mathrm{na}}(D)$ as a function of γ under the distortion covariance matrix of (28).

VI. Conclusions and Future Directions

We introduced a new class of zero-delay source coding problems where a vector-valued Gauss-Markov source is transmitted subject to a covariance distortion constraint. We tackled the problem by defining a lower bound to it where we minimize mutual information subject to causality constraints and covariance distortion constraints. This lower bound is solved by showing that it can be cast as a log-determinant maximization problem, thus, is semidefinite representable and can be evaluated numerically. We also showed that this class of information measures, has achievable rates up to a constant loss due to the lossy quantization and memoryless entropy coding. We exhibited our framework via numerical example. As an ongoing research we investigate the extension of this framework in systems where the source is a controlled process.

References

[1] J. Gubbi, R. Buyya, S. Marusic, and M. Palaniswami, “Internet of things (IoT): A vision, architectural elements, and future directions,” *Future Gener. Comput. Syst.*, vol. 29, no. 7, pp. 1645 – 1660, 2013.

[2] E. I. Silva, M. S. Derpich, J. Østergaard, and M. A. Encina, “A characterization of the minimal average data rate that guarantees a given closed-loop performance level,” *IEEE Trans. Autom. Control*, vol. 61, no. 8, pp. 2171–2186, Aug 2016.

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

[4] V. Kostina and B. Hassibi, “Rate-cost tradeoffs in control,” in *Proc. Annual Allerton Conf. on Comm., Control, and Comput.*, Sept 2016, pp. 1157–1164.

[5] P. A. Stavrou, T. Charalambous, and C. D. Charalambous, “Filtering with fidelity for time-varying Gauss-Markov processes,” in *Proc. IEEE Conf. Decision Control*, December 2016, pp. 5465–5470.

[6] T. Tanaka, K. K. K. Kim, P. A. Parrilo, and S. K. Mitter, “Semidefinite programming approach to Gaussian sequential rate-distortion trade-offs,” *IEEE Trans. Autom. Control*, vol. 62, no. 4, pp. 1896–1910, April 2017.

[7] P. A. Stavrou, J. Østergaard, C. D. Charalambous, and M. S. Derpich, “An upper bound to zero-delay rate distortion via Kalman filtering for vector Gaussian sources,” in *Proc. IEEE Inf. Theory Workshop*, 2017, pp. 534–538, extended version in https://arxiv.org/pdf/1701.06368v3.pdf.

[8] A. Khina, V. Kostina, A. Khisti, and B. Hassibi, “Tracking and control of GaussMarkov processes over packet-drop channels with acknowledgments,” *IEEE Trans. Control Netw. Sys. (submitted)*, 2017. [Online]. Available: http://www.its.caltech.edu/~khina/papers/journal/tracking_cones2017_2col.pdf

[9] R. Zamir, *Lattice Coding for Signals and Networks*. Cambridge: Cabridge University Press, 2014.

[10] Y. Oohama, “Indirect and direct Gaussian distributed source coding problems,” *IEEE Trans. Inf. Theory*, vol. 60, no. 12, pp. 7506–7539, Dec 2014.

[11] M. S. Rahman and A. B. Wagner, “Rate region of the vector Gaussian one-helper source-coding problem,” *IEEE Trans. Inf. Theory*, vol. 61, no. 5, pp. 2708–2728, May 2015.

[12] A. Zahedi, J. Østergaard, S. H. Jensen, P. A. Naylor, and S. Bech, “Source coding in networks with covariance distortion constraints,” *IEEE Trans. Signal Proc.*, vol. 64, no. 22, pp. 5943–5958, Nov 2016.

[13] A. Zahedi, J. Østergaard, S. H. Jensen, S. Bech, and P. Naylor, “Audio coding in wireless acoustic sensor networks,” *Signal Process.*, vol. 107, pp. 141 – 152, 2015.

[14] M. S. Derpich and J. Østergaard, “Improved upper bounds to the causal quadratic rate-distortion function for Gaussian stationary sources,” *IEEE Trans. Inf. Theory*, vol. 58, no. 5, pp. 3131–3152, May 2012.

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

[16] P. Dupuis and R. S. Ellis, *A Weak Convergence Approach to the Theory of Large Deviations*. John Wiley & Sons, Inc., New York, 1997.

[17] P. A. Stavrou, “Extremum problems of directed information,” Ph.D. dissertation, University of Cyprus, 2016.

[18] H. S. Witsenhausen, “On the structure of real-time source coders,” *Bell Syst. Tech. J.*, vol. 58, no. 6, pp. 1437–1451, July 1979.

[19] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York, NY, USA: Cambridge University Press, 2004.

[20] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. John Wiley & Sons, Inc., Hoboken, New Jersey, 2006.

[21] M. Grant and S. Boyd, “CVX: Matlab software for disciplined convex programming, version 2.1,” http://cvxr.com/cvx, Mar. 2014.