Fixed-Rate Zero-Delay Source Coding for Stationary Vector-Valued Gauss-Markov Sources

Stavrou, Fotios; Østergaard, Jan

Published in:
Proceedings - DCC 2018

DOI (link to publication from Publisher):
10.1109/DCC.2018.00034

Publication date:
2018

Document Version
Accepted author manuscript, peer reviewed version

Fixed-Rate Zero-Delay Source Coding for Stationary Vector-Valued Gauss-Markov Sources

Photios A. Stavrou* and Jan Østergaard†

*KTH Royal Institute of Technology
Dep. of Info. Sci. and Eng.
Main Campus, Osquldas väg 10, C-436
Stockholm, 100 44, Sweden
fstavrou@kth.se

†Aalborg University
Dep. Electr. Sys.
Fredrik Bajers Vej 7, B5-210
Aalborg, 9220, Denmark
jo@es.aau.dk

Abstract

We consider a fixed-rate zero-delay source coding problem where a stationary vector-valued Gauss-Markov source is compressed subject to an average mean-squared error (MSE) distortion constraint. We address the problem by considering the Gaussian nonanticipative rate distortion function (NRDF) which is a lower bound to the zero-delay Gaussian RDF. Then, we use its corresponding optimal "test-channel" to characterize the stationary Gaussian NRDF and evaluate the corresponding information rates. We show that the Gaussian NRDF can be achieved by p-parallel fixed-rate scalar uniform quantizers of finite support with dithering signal up to a multiplicative distortion factor and a constant rate penalty. We demonstrate our framework with a numerical example.

1 Introduction

Zero-delay processing of information in source coding is the decoding of each source sample at the same time instant that the source sample is encoded. Zero-delay source coding is desirable in various applications, like for instance, in signal processing [1] and in networked control systems [2, 3].

In this paper, we consider a fixed-rate zero-delay source coding problem where a vector-valued Gaussian source modeled as a stationary linear time-invariant (LTI) vector-valued Gauss-Markov process is compressed subject to a MSE distortion constraint. We tackle the problem, by considering the NRDF [4] which is known to be a tighter lower bound to the zero-delay RDF compared to the classical RDF (for details see, e.g., [5, 6]). We use the optimal test-channel that corresponds to the NRDF of the aforementioned stationary Gaussian source model under the MSE distortion constraint, to characterize the stationary Gaussian NRDF and evaluate its corresponding information rates [6,7]. The realization of the optimal test-channel enables us to show that the Gaussian NRDF can be achieved by p-parallel subtractively dithered scalar uniform quantizers (SDSUQ) of finite support up to a multiplicative factor to the MSE distortion and a constant rate penalty.

Our idea stems from the recent work of [8], where the authors used dithered fixed-rate scalar quantization of a scalar Gaussian process in combination with $\Sigma - \Delta$ modulation. They showed that for a fixed probability of overload, say $\mathbf{P}_{\text{ol}}$, and by a careful choice of bitrate and limits on the quantizer's support, it is possible to guarantee system's stability with a quantizer bitrate $R' = R + o(\log\log(\frac{1}{\mathbf{P}_{\text{ol}}}))$, where

R is the bitrate required for an unbounded quantizer without overload. Moreover, the resulting distortion due to having overload distortion in addition to the granular distortion D, would be $\frac{D(1+o(1))}{1-P_{\rm ol}}$, where the multiplicative term $o(1)\rightarrow 0$ as the length of the source sample increases since the effect of the initial state of their causal filter vanishes. It is worth emphasizing that the results of [8], which we will build upon in the sequel, were established under the assumption of i.i.d., scalar RVs obtained due to non-causal prediction of a stationary scalar Gaussian colored process. In our work, we consider *zero-delay* coding of a stationary *vector* Gaussian process with spatio-temporal correlation, i.e., with correlation within the elements of a vector as well as correlation between vectors.

The fixed-rate zero-delay source coding scheme of this paper complements the achievability approach using variable-length zero-delay source coding that was proposed in [6, Section IV] by means of an entropy coded dithered quantizer (ECDQ) [9].

Notation. Let $\mathbb{R}$ denote 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 sequence of random variables 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 $x_r^t\in{\cal X}_r^t\triangleq\times_{k=r}^t{\cal X}_k$, with ${\cal X}_k={\cal X}$, 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 $\mathbf{x}$ on ${\cal X}$ is denoted by $\mathbf{P}_{\mathbf{x}}(dx)\equiv\mathbf{P}_{\mathbf{x}}$. The conditional distribution of $\mathbf{y}$ 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 $K^{\rm T}$. For a square matrix $K\in\mathbb{R}^{p\times{p}}$ with entries K_{ij} on the $i^{\rm th}$ row and $j^{\rm th}$ column, we denote by ${\rm diag}\{K\}$ the matrix having $K_{ii},~i=1,\ldots,p$, on its diagonal and zero elsewhere, its trace by ${\rm trace}(K)$, 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".

2 Problem Formulation

In this paper, we consider a source coding problem with instantaneous encoding and decoding. We assume a stationary $\mathbb{R}^p$-valued Gaussian source governed by the following 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, $\mathbf{x}_0\in\mathbb{R}^p\sim{\cal N}(0;\Sigma_{\mathbf{x}})$ is the initial state with stationary covariance $\Sigma_{\mathbf{x}}\succ{0}$, $\mathbf{w}_t\in\mathbb{R}^p\sim{\cal N}(0;\Sigma_{\mathbf{w}})$, $\Sigma_{\mathbf{w}}\succ{0}$, is an i.i.d. Gaussian sequence, independent of $\mathbf{x}_0$. To ensure that (1) is a stationary process, we restrict the absolute values of the eigenvalues of A to be inside the unit circle.

Formally, this zero-delay source coding problem can be interpreted 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 ${\cal Z}_t$ of at most countable number of codewords. We define the input and output alphabet of the noiseless digital channel by ${\cal M}=\{1,2,\ldots,M\}$ where $M=\max_t|{\cal Z}_t|<\infty$. 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 $z_0=f_0(x_0)$. Since the source is random, $\mathbf{z}_t$ and its length $\mathbf{l}_t$ (in bits) are RVs (although the main result of this paper treats fixed-rate coding). Upon receiving $\mathbf{z}^t$, the *decoder* produces the reconstruction $\mathbf{y}_t$. 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 $y_0=g_0(z_0)$. Both the encoder and decoder process information without delay, and at each time step t, the per-letter MSE distortion needs to satisfy $\mathbb{E}\{||\mathbf{x}_t-\mathbf{y}_t||_2^2\}\leq D$, $\forall t$, where $D>0$ is the pre-specified distortion level.

The objective is to minimize the average expected 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 $\{(f_t,g_t):\ t\in\mathbb{N}_0\}$. This is formally cast by the following optimization problem:

$$R_{\mathrm{ZD}}^{\mathrm{op}}(D)\triangleq\inf_{\substack{\{(f_t,\ g_t):\ t\in\mathbb{N}_0\},\\ \mathbb{E}\{||\mathbf{x}_t-\mathbf{y}_t||_2^2\}\leq D,\ \forall t}}\limsup_{n\longrightarrow\infty}\frac{1}{n+1}\sum_{t=0}^{n}\mathbb{E}(\mathbf{l}_t). \tag{2}$$

We refer to (2) as the *operational zero-delay Gaussian* RDF. Solving (2) is very hard, since it is defined over all operational codes. Instead, we consider in the next section a lower bound to (2), which is based on information theoretic quantities.

3 A lower bound on $R_{\mathrm{ZD}}^{\mathrm{op}}(D)$

In this section, we state the definition of the NRDF, denoted by $R^{\mathrm{na}}(D)$, for the source model of (1) subject to a MSE distortion constraint [6]. After the definition, we explain the optimal test-channel that corresponds to this Gaussian source model [7].

First, notice that for general continuous alphabet sources, i.e., sources that are not necessarily Gaussian, it holds that $R(D)\overset{(a)}{\leq}R^{\mathrm{na}}(D)\overset{(b)}{\leq}R_{\mathrm{ZD}}^{\mathrm{op}}(D)$ (see [5,6]), where $R(D)$ denotes the classical RDF [10]. Note that, inequality (a) is strict, in general, and becomes equality when the source is i.i.d. or when the rate tends to infinity. In contrary, inequality (b) is strict at high rates (high resolution) due to space-filling loss and becomes equality at zero rate and at infinite dimensional vector-valued Gauss-Markov sources.

We consider a 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 a squared-error distortion $||\mathbf{x}_t-\mathbf{y}_t||_2^2$.

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

$$\mathbf{P}_{\mathbf{x}_t|\mathbf{x}^{t-1},\mathbf{y}^{t-1}}\triangleq\mathbf{P}(dx_t|x_{t-1}),\quad t\in\mathbb{N}_0^n. \tag{3}$$

At time $t=0$, we assume $\mathbf{P}(dx_0|x_{-1})\triangleq\mathbf{P}(dx_0)$. Given the conditional distributions in (3), then, 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

$x^n \equiv (x_0, \ldots, x_n)$ is randomly generated, according to the collection of conditional distributions, known also as test-channels, by

$$\mathbf{P}_{\mathbf{y}_t|\mathbf{y}^{t-1},\mathbf{x}^t} \triangleq \mathbf{P}(dy_t|y^{t-1}, x^t), \quad t \in \mathbb{N}_0^n. \tag{4}$$

At $n = 0$, we assume $\mathbf{P}(dy_0|y^{-1}, x_0) = \mathbf{P}(dy_0|x_0)$. From [11], we know that the conditional distributions $\mathbf{P}(dy_t|y^{t-1}, x^t)$ in (4), 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 (3) and (4), we can uniquely define the joint distribution of $\{(\mathbf{x}^n, \mathbf{y}^n) : \ t \in \mathbb{N}_0^n\}$ by

$$\mathbf{P}_{\mathbf{x}^n,\mathbf{y}^n} = \mathbf{P}(dx^n) \otimes \overrightarrow{\mathbf{P}}(dy^n|x^n). \tag{5}$$

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

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

where (a) is due to the Radon-Nikodym derivative theorem [12]; (b) due to chain rule of relative entropy; (c) follows by definition.

Definition 1 *(*NRDF *subject to a* MSE *distortion constraint) The* NRDF *of the stationary source model* (1) *subject to a per-letter* MSE *distortion constraint is defined as (assuming the limit exists):*

$$R^{\text{na}}(D) = \lim_{n \longrightarrow \infty} \frac{1}{n+1} \inf_{\mathbf{P}(dy_t|y^{t-1}, x^t), \ t \in \mathbb{N}_0^n: \ \mathbb{E}\{||\mathbf{x}_t - \mathbf{y}_t||^2\} \leq D, \ \forall t \in \mathbb{N}_0^n} I(\mathbf{x}^n; \mathbf{y}^n). \tag{6}$$

The optimization problem of Definition 1, in contrast to the one given in (2) is *convex* and there exists an optimal solution characterizing it (assuming a non-zero distortion) (for details see [11]). Moreover, by [6], the choice of the source model of (1) subject to a MSE distortion constraint yields an optimal "test channel" that achieves the infimum in (6) of the form

$$\mathbf{P}^*(dy_t|y^{t-1}, x^t) = \mathbf{P}^*(dy_t|y_{t-1}, x_t), \ t \in \mathbb{N}_0. \tag{7}$$

In addition, the corresponding joint process $\{(\mathbf{x}_t, \mathbf{y}_t) : \ t \in \mathbb{N}_0\}$ is jointly Gaussian.

Realization of the Optimal Test-Channel with Stationary Statistics. We will now assume that the joint process $\{(\mathbf{x}_t, \mathbf{y}_t) : \ t \in \mathbb{N}_0\}$ is stationary, so that by [6, Section IV] (or [7, Theorem 2]), we can obtain the realization of the optimal test-channel that corresponds to a stationary optimal minimizer (7). This is illustrated in Fig. 1. In this setup, the stationary output process $\mathbf{y}_t$ is of the form:

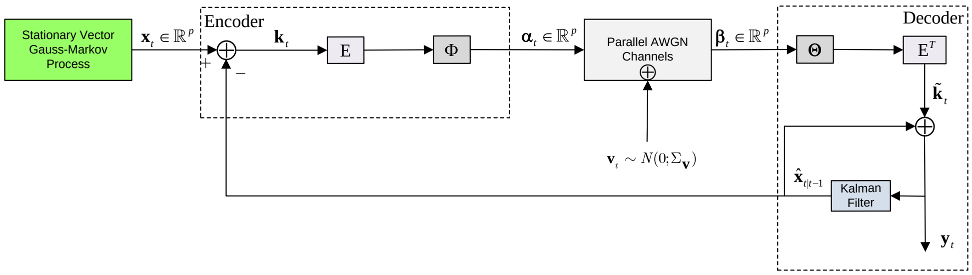

Figure 1: Realization of the optimal stationary "test-channel" $\mathbf{P}^*(dy_t|y_{t-1}, x_t)$.

$$\mathbf{y}_t = E^{\mathrm{T}} H E(\mathbf{x}_t - \widehat{\mathbf{x}}_{t|t-1}) + \widehat{\mathbf{x}}_{t|t-1} + E^{\mathrm{T}}\Theta\mathbf{v}_t, \tag{8}$$

where $H \succeq 0$ is a scaling matrix; $\mathbf{v}_t$ is an independent Gaussian noise process with $\mathcal{N}(0;\Sigma_{\mathbf{v}})$, $\Sigma_{\mathbf{v}} = \mathrm{diag}\{V\}$ independent of $\mathbf{x}_0$; the error $(\mathbf{x}_t - \widehat{\mathbf{x}}_{t|t-1}) \sim \mathcal{N}(0;\Sigma)$, where $\Sigma = \mathbb{E}\{(\mathbf{x}_t - \widehat{\mathbf{x}}_{t|t-1})(\mathbf{x}_t - \widehat{\mathbf{x}}_{t|t-1})^{\mathrm{T}}\}$, $\forall t$, and $\widehat{\mathbf{x}}_{t|t-1} \triangleq \mathbb{E}\{\mathbf{x}_t|\mathbf{y}^{t-1}\}$; the error $(\mathbf{x}_t - \widehat{\mathbf{x}}_{t|t}) \sim \mathcal{N}(0;\Sigma')$, where $\Sigma' = \mathbb{E}\{(\mathbf{x}_t - \widehat{\mathbf{x}}_{t|t})(\mathbf{x}_t - \widehat{\mathbf{x}}_{t|t})^{\mathrm{T}}\}$, $\forall t$, and $\widehat{\mathbf{x}}_{t|t} \triangleq \mathbb{E}\{\mathbf{x}_t|\mathbf{y}^t\}$. Moreover, $\{\widehat{\mathbf{x}}_{t|t-1},\ \Sigma\}$ are given by the following time-invariant filtering recursions:

$$\widehat{\mathbf{x}}_{t|t-1} = A\widehat{\mathbf{x}}_{t-1|t-1}, \quad \widehat{\mathbf{x}}_{0|-1} = \text{invariant value}, \tag{9a}$$
$$\Sigma = A\Sigma' A^{\mathrm{T}} + \Sigma_{\mathbf{w}}, \tag{9b}$$
$$\widehat{\mathbf{x}}_{t|t} = \widehat{\mathbf{x}}_{t|t-1} + G\tilde{\mathbf{k}}_t, \quad \widehat{\mathbf{x}}_{0|0} = \text{invariant value}, \tag{9c}$$
$$\tilde{\mathbf{k}}_t \triangleq \mathbf{y}_t - \widehat{\mathbf{x}}_{t|t-1} \quad \text{(innovation)}, \tag{9d}$$
$$\Sigma' = \Sigma - GSG^{\mathrm{T}}, \tag{9e}$$
$$G = \Sigma(E^{\mathrm{T}}HE)^{\mathrm{T}}S^{-1} \quad \text{(Kalman Gain)}, \quad S = (E^{\mathrm{T}}HE)\Sigma(E^{\mathrm{T}}HE)^{\mathrm{T}} + E^{\mathrm{T}}\Theta\Sigma_{\mathbf{v}}\Theta^{\mathrm{T}}E,$$

where (9b), (9e) are the steady state covariance matrices of the filter with $\Sigma = \Sigma^{\mathrm{T}} \succ 0$ and $\Sigma' = \Sigma'^{\mathrm{T}} \succ 0$. In addition, we let

$$H \triangleq I - \tilde{\Delta}\Lambda^{-1},\ \Theta \triangleq \sqrt{H\tilde{\Delta}\Sigma_{\mathbf{v}}^{-1}},\ \Phi \triangleq \Theta^{-1}H, \tilde{\Delta} \triangleq \mathrm{diag}\{\delta\},\ \Lambda_t = \mathrm{diag}\{\lambda\}, \tag{10}$$

where $\Lambda \succ 0$, $\tilde{\Delta} \succ 0$ and $E \in \mathbb{R}^{p\times p}$ is an orthogonal matrix. It is easy to verify that for the choice of (10), the output process in (8) simplifies to:

$$\mathbf{y}_t = E^{\mathrm{T}}HE\mathbf{x}_t + (I - E^{\mathrm{T}}HE)A\mathbf{y}_{t-1} + E^{\mathrm{T}}\Theta\mathbf{v}_t, \tag{11}$$

because $\widehat{\mathbf{x}}_{t|t-1} = A\mathbf{y}_{t-1}$, $\widehat{\mathbf{x}}_{t|t} = \mathbf{y}_t$, $G = I$, $\Sigma' = E^{\mathrm{T}}\tilde{\Delta}E$.

Next, we briefly explain the basic features of the realization of Fig. 1.

Encoder. We introduce the estimation error $\{\mathbf{k}_t \in \mathbb{R}^p :\ t \in \mathbb{N}_0\}$, where $\mathbf{k}_t \triangleq \mathbf{x}_t - \widehat{\mathbf{x}}_{t|t-1} \sim \mathcal{N}(0;\Sigma)$, $\forall t \in \mathbb{N}_0$. The stationary value Σ is then, diagonalized by introducing an orthogonal matrix E (invertible matrix) such that $E\Sigma E^{\mathrm{T}} = \mathrm{diag}\{\lambda\} \triangleq \Lambda$.

Decoder. We introduce the innovations process $\{\tilde{\mathbf{k}}_t :\ t \in \mathbb{N}_0\}$ defined by (9d).

Parallel AWGN *Channel.* The AWGN channel is of the form $\boldsymbol{\beta}_t = \alpha_t + \mathbf{v}_t = \Phi E\mathbf{k}_t + \mathbf{v}_t$, $\mathbf{v}_t \sim \mathcal{N}(0;\Sigma_{\mathbf{v}})$, $\Sigma_{\mathbf{v}} = \mathrm{diag}\{V\}$, $t \in \mathbb{N}_0$.

Distortion. The squared-error distortion $||\mathbf{k}_t - \tilde{\mathbf{k}}_t||_2^2$ at each t is not affected by the above processing of $\{(\mathbf{x}_t, \mathbf{y}_t) : \ t \in \mathbb{N}_0\}$, i.e., $||\mathbf{x}_t - \mathbf{y}_t||_2^2 = ||\mathbf{k}_t - \tilde{\mathbf{k}}_t||_2^2, \ \ t \in \mathbb{N}_0$. Moreover, at each t, the scheme of Fig. 1 yields a MSE distortion $\mathbb{E}\{||\mathbf{x} - \mathbf{y}||_2^2\} \equiv \mathbb{E}\{(\mathbf{x}_t - \mathbf{y}_t)^{\mathrm{T}}(\mathbf{x}_t - \mathbf{y}_t)\} = \mathrm{trace}\,\mathbb{E}\{(\mathbf{x}_t - \widehat{\mathbf{x}}_{t|t})(\mathbf{x}_t - \widehat{\mathbf{x}}_{t|t})^{\mathrm{T}}\} = \mathrm{trace}(\Sigma') \leq D$. Hence, following [6, Section IV], the stationary Gaussian NRDF subject to a MSE distortion can be expressed as follows:

$$R^{\mathrm{na}}(D) = \min_{0 \prec \Sigma' \preceq \Sigma, \ \ \mathrm{trace}(\Sigma') \leq D} \frac{1}{2} \log \frac{|\Sigma|}{|\Sigma'|}. \tag{12}$$

The max-det optimization problem of (12) is convex and can be solved numerically using, for instance, interior point methods via semidefinite programming approach [13]. In the next theorem, we give the semidefinite representation that corresponds to the optimization problem of (12).

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

$$R^{\mathrm{na}}(D) = \min_{F \succ 0} -\frac{1}{2}\log|F| + \frac{1}{2}\log|\Sigma_{\mathbf{w}}|. \tag{13a}$$

$$s.t. \ \ 0 \prec \Sigma' \preceq \Sigma, \quad \ \mathrm{trace}(\Sigma') \leq D \tag{13b}$$

$$\begin{bmatrix} \Sigma' - F & \Sigma' A^{T} \\ A\Sigma' & \Sigma \end{bmatrix} \succeq 0 \tag{13c}$$

Proof. The proof is based on reformulating (12) using matrix determinant lemma and Woodbury matrix identity to obtain the linear matrix inequality (13c). This procedure is omitted due to space limitations. ■

4 Fixed-Rate Coding Using Predictive SDSUQ of Finite Support

In this section, we leverage upon ideas in [8] in order to bound the probability of overload, denoted $\mathbf{P}_{\mathrm{ol}}$, using a scalar predictive quantizer of finite support with dithering signal. Our idea is to construct a predictive coder based on the test-channel realization of Fig. 1. With this scheme, the p elements of the $\mathbb{R}^p$-valued vector $\boldsymbol{\alpha}_t$, namely, $\boldsymbol{\alpha}_t = (\boldsymbol{\alpha}_{t,1}, \ldots, \boldsymbol{\alpha}_{t,p})$, will become mutually independent after quantization but not identically distributed. We use p-parallel SDSUQ, namely, $\underline{Q}(\cdot) \triangleq (Q_1^{\mathrm{SD}}(\cdot), \ldots, Q_p^{\mathrm{SD}}(\cdot))$, which are applied separately along the p-dimensions of $\boldsymbol{\alpha}_t$, hence substituting the p-parallel AWGN channels of Fig. 1. In particular, the i^{th} quantizer, $Q_i^{\mathrm{SD}}(\cdot)$ which is applied on the i^{th}-element $\boldsymbol{\alpha}_{t,i}$ has a bitrate R_i, 2^{R_i} quantization levels, and quantizer support $[-\frac{\Gamma_i}{2}, \frac{\Gamma_i}{2})$, where $\Gamma_i \triangleq 2^{R_i}\sqrt{12V_{ii}}$. We choose all the quantizers to have quantization step $\Delta_i = \sqrt{12V_{ii}}$. We denote by $\mathbf{r}_t$ the $\mathbb{R}^p$-valued random process of dithered signals whose individual components, denoted by $(\mathbf{r}_{t,1}, \ldots, \mathbf{r}_{t,p})$, are mutually independent across time and space, and uniformly distributed over the interval $\mathbf{r}_{t,i} \sim \mathrm{Unif}[-\frac{\Delta_i}{2}, \frac{\Delta_i}{2})$, independent of $\boldsymbol{\alpha}_{t,i}$, $\forall t, i$. The quantization scheme is depicted in Fig. 2. In order to quantize $\boldsymbol{\alpha}_{t,i}$, we add and subtract the dither signal $\mathbf{r}_{t,i}$, such that the output of the quantizer $\tilde{\boldsymbol{\beta}}_{t,i} = Q_i^{\mathrm{SD}}(\boldsymbol{\alpha}_{t,i} + \mathbf{r}_{t,i})$ and $\boldsymbol{\beta}_{t,i} = Q_i^{\mathrm{SD}}(\boldsymbol{\alpha}_{t,i} + \mathbf{r}_{t,i}) - \mathbf{r}_{t,i}$. The quantization error $\boldsymbol{\xi}_t$ in each dimension i is:

$$\boldsymbol{\xi}_{t,i} \triangleq \boldsymbol{\beta}_{t,i} - \boldsymbol{\alpha}_{t,i} = Q_i^{\mathrm{SD}}(\boldsymbol{\alpha}_{t,i} + \mathbf{r}_{t,i}) - (\boldsymbol{\alpha}_{t,i} + \mathbf{r}_{t,i}). \tag{14}$$

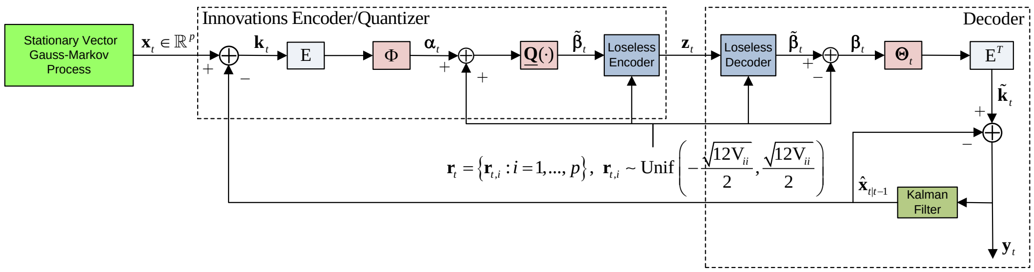

Figure 2: Proposed scheme with p-parallel predictive SDSUQ of finite support.

Clearly, if we employ the p-parallel SDSUQ at each time step t, then, it is highly probable to cause an overload error which is further amplified due to the feedback loop in our scheme. In that particular case, the system is no longer stable and the MSE distortion is expected to be significantly increased. To avoid this critical issue, we reset the system in a slightly different way than in [8, Section III], whenever an overload error appears. In particular, we initialize the memory of the Kalman filter by samples drawn from its stationary distribution. This means that *only* the current input $\mathbb{R}^p$-valued vector is dropped. This approach does not cause any temporal non-stationarities in the statistics of the system because as analyzed in Section 3, the statistics of the filter are stationary and the reproduction vector $\mathbf{y}_t \in \mathbb{R}^p$ is stationary.

Theorem 2 *At each t, we let D be the* MSE *distortion attained by the test channel of Fig. 1, and $I(\boldsymbol{\alpha}_{t,i};\boldsymbol{\alpha}_{t,i}+\mathbf{r}_{t,i})$ the scalar mutual information between the input and the output of the* AWGN *channel at each dimension. Then, at each time step t, for $0<\mathbf{P}_{\text{ol}}<1$, the p-parallel* SDSUQ *of Fig. 2 each with quantization rate per dimension $R_i=I(\boldsymbol{\alpha}_{t,i};\boldsymbol{\alpha}_{t,i}+\mathbf{r}_{t,i})+\delta(\mathbf{P}_{\text{ol}})$ attain a* MSE *distortion per dimension, hereinafter denoted by D_i, smaller than*

$$\frac{D_i}{1-\mathbf{P}_{\text{ol}}},\ \ D\triangleq\frac{1}{p}\sum_{i=1}^{p}D_i, \tag{15}$$

provided that overload did not occur. Moreover, the overload probability per dimension is smaller than $\mathbf{P}_{\text{ol}}$ with

$$\delta(\mathbf{P}_{\text{ol}})\triangleq\frac{1}{2}\log\left(-\frac{2}{3}\ln\frac{\mathbf{P}_{\text{ol}}}{2p}\right). \tag{16}$$

Proof. The key idea here, is to get into a situtation where we can closely follow the proof technique of [8]. Let $\tilde{Q}_{\sqrt{12V_{ii}}\mathbb{Z}}(x)$ denote rounding x to the nearest point in the (infinite) lattice $\sqrt{12V_{ii}}\mathbb{Z}$. It is easy to verify that for any $x\in[-\frac{\Gamma_i}{2},\frac{\Gamma_i}{2})$ we have:

$$Q_i(x)=\tilde{Q}_{\sqrt{12V_{ii}}\mathbb{Z}}\left(x+\frac{\sqrt{12V_{ii}}}{2}\right)-\frac{\sqrt{12V_{ii}}}{2}. \tag{17}$$

Applying (14), if $|\boldsymbol{\alpha}_{t,i}+\boldsymbol{\xi}_{t,i}|\leq\frac{\Gamma_i}{2}$, then overload did not occur, and we obtain:

$$\boldsymbol{\xi}_{t,i}=\tilde{Q}_{\sqrt{12V_{ii}}\mathbb{Z}}(\boldsymbol{\alpha}_{t,i}+\mathbf{r}_{t,i}+\sqrt{3V_{ii}})-(\boldsymbol{\alpha}_{t,i}+\mathbf{r}_{t,i}+\sqrt{3V_{ii}}). \tag{18}$$

Dealing with the overload event of the finite support scalar quantizer is quite complicated. For this reason, we first consider a *reference system* with an infinite-support scalar quantizer , i.e., $R_i = \infty$, and analyze its performance. Now, if the magnitude of the i^{th}-input to the i^{th} unbounded quantizer never exceeds $\frac{\Gamma_i}{2}$ for $i = 1, \ldots, p$, then, clearly the reference system is equivalent to the original system where we have assumed a finite-support scalar quantizer. Thus, it suffices to find the average distortion of the reference system and the probability that the i^{th} input to its i^{th} quantizer exceeds $\frac{\Gamma_i}{2}$ for $i = 1, \ldots, p$.

Next, we will assume that the quantization noise for each $i = 1, \ldots, p$, is given by (18) regardless of whether or not $|\boldsymbol{\alpha}_{t,i} + \mathbf{r}_{t,i}| \leq \frac{\Gamma_i}{2}$, and then, we will seek for the probability of an overload error.

By assumption, $\boldsymbol{\xi}_t = (\boldsymbol{\xi}_{t,1}, \ldots \boldsymbol{\xi}_{t,p})$ is an i.i.d. sequence of RVs uniformly distributed over the interval $\boldsymbol{\xi}_{t,i} \sim \left[\frac{-\sqrt{12V_{ii}}}{2}, \frac{\sqrt{12V_{ii}}}{2}\right)$ independent of $\boldsymbol{\alpha}_t = (\boldsymbol{\alpha}_{t,1}, \ldots, \boldsymbol{\alpha}_{t,p})$. Moreover, for each i, $\boldsymbol{\xi}_{t,i}$ has zero mean and variance V_{ii}. For this reason, the p-parallel SDSUQ depicted in Fig. 2 (with unbounded support) is equivalent to the test-channel of Fig. 1 with $\boldsymbol{\xi}_{t,i} \sim \text{Unif}\left[\frac{-\sqrt{12V_{ii}}}{2}, \frac{\sqrt{12V_{ii}}}{2}\right)$ instead of $\mathbf{v}_{t,i} \sim \mathcal{N}(0; V_{ii})$, where $\mathbf{v}_t = (\mathbf{v}_{t,1}, \ldots, \mathbf{v}_{t,p})$. Hence, the average MSE distortion attained by the reference p-parallel SDSUQ of Fig. 2 is precisely the same as the one attained by the analysis for Fig. 1 and is equal to D (see Section 3).

In what follows, we concentrate into analyzing the probability of an overload error within a vector of length p, as a function of R_i and $I(\boldsymbol{\alpha}_{t,i}; \boldsymbol{\alpha}_{t,i} + \boldsymbol{\xi}_{t,i})$. This event translates into an event at the reference system where some input to the quantizer will exceeds $\frac{\Gamma_i}{2}$ in magnitude within block p. Our aim is to upper bound the probability of this event. We define the overload event $OL_i \triangleq \{|\boldsymbol{\alpha}_{t,i} + \boldsymbol{\xi}_{t,i}| > \Gamma_i/2\}$ and the event $OL \triangleq \cup_{i=1}^{p} OL_i$. By the union bound, we know that

$$\mathbf{P}_{\text{ol}} = Pr(OL) \leq \sum_{i=1}^{p} Pr(OL_i). \tag{19}$$

The RV $\boldsymbol{\alpha}_{t,i} + \boldsymbol{\xi}_{t,i}$ is a linear combination of the Gaussian RVs $\{(\mathbf{x}_{0,i}, \mathbf{x}_{1,i} \ldots, \mathbf{x}_{t,i}) : t \in \mathbb{N}_0\}$ and the independent uniform RVs $(\boldsymbol{\xi}_{0,i}, \boldsymbol{\xi}_{1,i}, \ldots, \boldsymbol{\xi}_{t,i})$. Using [14, Lemma 4], we can bound by its variance the probability of a RV of this type to exceed a certain threshold. Applying this bound to $\boldsymbol{\alpha}_{t,i} + \boldsymbol{\xi}_{t,i}$ yields:

$$\begin{aligned} Pr\left(OL_i\right) &\leq \exp\left\{-\frac{\Gamma_i^2}{8\mathbb{E}(\boldsymbol{\alpha}_{t,i} + \boldsymbol{\xi}_{t,i})^2}\right\} \stackrel{(a)}{=} 2\exp\left\{-\frac{12V_{ii}2^{2R_i}}{8\left(\mathbb{E}(\boldsymbol{\alpha}_{t,i})^2 + \mathbb{E}(\boldsymbol{\xi}_{t,i})^2\right)}\right\}, \\ &= 2\exp\left\{-\frac{12V_{ii}2^{2R_i}}{8V_{ii}\left(1 + \frac{\mathbb{E}(\boldsymbol{\alpha}_{t,i})^2}{V_{ii}}\right)}\right\} = 2\exp\left\{-\frac{3}{2}2^{2\left(R_i - \frac{1}{2}\log\left(1 + \frac{\mathbb{E}(\boldsymbol{\alpha}_{t,i})^2}{V_{ii}}\right)\right)}\right\} \\ &\stackrel{(b)}{=} 2\exp\left\{-\frac{3}{2}2^{2(R_i - I(\boldsymbol{\alpha}_{t,i};\boldsymbol{\alpha}_{t,i} + \boldsymbol{\xi}_{t,i}))}\right\} \end{aligned} \tag{20}$$

where (a) follows from the definition of Γ_i and the fact that $\boldsymbol{\alpha}_{t,i}$ and $\boldsymbol{\xi}_{t,i}$ are independent $\forall i$; (b) follows from the definition of $I(\boldsymbol{\alpha}_{t,i}; \boldsymbol{\alpha}_{t,i} + \boldsymbol{\xi}_{t,i})$ when $\boldsymbol{\alpha}_{t,i}$ is a Gaussian

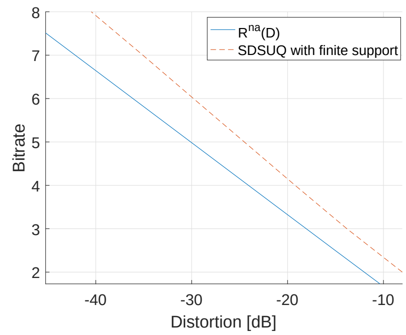

Figure 3: Comparison between $R^{\rm na}(D)$ and the zero-delay operational rates.

RV, and the fact that $\mathbb{E}(\boldsymbol{\xi}_{t,i})^2 = \mathbb{E}(\mathbf{v}_{t,i})^2$ which implies $\mathbb{E}(\boldsymbol{\alpha}_{t,i})^2 = \mathbb{E}(\boldsymbol{\alpha}^{\rm Gaussian}_{t,i})^2$. By substituting (20) into (19) we obtain:

$$\mathbf{P}_{\rm ol} \le 2\sum_{i=1}^{p} \exp\left\{-\frac{3}{2}2^{2(R_i - I(\boldsymbol{\alpha}_{t,i};\boldsymbol{\alpha}_{t,i}+\boldsymbol{\xi}_{t,i}))}\right\} \overset{(c)}{\le} 2p\exp\left\{-\frac{3}{2}2^{2\bar{\delta}_i}\right\}, \tag{21}$$

where (c) follows from the fact that $\bar{\delta} \triangleq \min_i\{\delta_i\}$ with $\delta_i \triangleq R_i - I(\boldsymbol{\alpha}_{t,i};\boldsymbol{\alpha}_{t,i}+\boldsymbol{\xi}_{t,i})$. Until now, we have shown that the reference system, i.e., the one with infinite-support quantizer, achieves D_i for each $i^{\rm th}$ SDSUQ, that results into obtaining the total MSE distortion normalized per dimension $D = \frac{1}{p}\sum_{i=1}^{p} D_i$ which corresponds to the total MSE distortion of the test-channel in Fig. 1. Based on our previous analysis, the probability that the input samples for each of the p-parallel SDSUQ exceeds $\frac{\Gamma_i}{2}$ in magnitude of block length p, is bounded by (21). For our original system whose $i^{\rm th}$ quantizer has finite support of $[-\frac{\Gamma_i}{2}, \frac{\Gamma_i}{2})$, this means that the overload probability is also upper bounded by (21). Furthermore, D_i of the original system if overload did not occur is the same as that of reference system conditioned on the event that OL did not occur. We denote this event by $D_i^{\overline{OL}}$ and the average distortion conditioned on the event that OL did occur by D_i^{OL}. For the reference system, we have that:

$$D_i = Pr(\overline{OL})D_i^{\overline{OL}} + Pr(OL)D_i^{OL} \ge Pr(\overline{OL})D_i^{\overline{OL}} \Rightarrow D_i^{\overline{OL}} \le \frac{D_i}{1-\mathbf{P}_{\rm ol}},\ \forall i.$$

Thus, we have shown that the p-parallel SDSUQ system illustrated in Fig. 2, whose $i^{\rm th}$ quantizer has limited support $[-\frac{\Gamma_i}{2}, \frac{\Gamma_i}{2})$, with $R_i = I(\boldsymbol{\alpha}_{t,i};\boldsymbol{\alpha}_{t,i}+\mathbf{r}_{t,i}) + \delta(\mathbf{P}_{\rm ol})$ achieves the same average MSE normalized per dimension $D = \frac{1}{p}\sum_{i=1}^{p} D_i$ as the test-channel of Fig. 1 where D_i is up to a multiplicative factor $\frac{1}{1-\mathbf{P}_{\rm ol}}$ with block error probability less than or equal to $2p\exp\left\{-\frac{3}{2}2^{2\bar{\delta}_i}\right\}$. Thus, (12) characterizes the rate-distortion tradeoff achieved by a system with p-parallel SDSUQ where for each dimension $i = 1, \ldots, p$, this tradeoff is achieved up to a factor $\frac{1}{1-\mathbf{P}{\rm ol}}$ and a constant penalty $\delta(\mathbf{P}_{\rm ol})$, that depends on the overload probability. To clarify this, for any $\mathbf{P}_{\rm ol} \in (0,1)$, taking the constant penalty of (16) at each dimension i, ensures that the overload probability is smaller than $\mathbf{P}_{\rm ol}$. ∎

Example 1 *We consider an* $\mathbb{R}^{10}$*-valued source modeled as in* (1)*, where* A *has* i.i.d. *Gaussian elements with* $A_{ij} \sim \mathcal{N}(0; \frac{1}{p^2})$*, and* $\Sigma_{\mathbf{w}} = I$*. We consider distortions*

$D \in [0.0003, 0.0009, 0.0045, 0.07, 0.31, 0.92]$ *and* $V_{ii} \in [1, 1, 1.2, 1.4, 1.7, 1.7]$ *corresponding to* $R_i = [8, 7, 6, 4, 3, , 2]$, $\Gamma_i = [13.9, 27.7, 66.5, 310.4, 753.8, 1507.6]$, *and* $\Delta_i = [3.46, 3.46, 4.15, 4.85, 5.89, 5.89]$, $\forall i = 1, \ldots, 10$. *Fig. 3 compares* $R^{\rm na}(D)$ *and the corresponding operational rates both normalized per dimension. We note that* $R^{\rm na}(D)$ *is evaluated by invoking the* SDP *solver of CVX platform [15] while the operational rates are evaluated using a* SDSUQ *of finite support in each dimension based on the previously mentioned values of* $R_i,\ \Gamma_i,\ \Delta_i$. *According to Fig. 3, at high rates, there is a loss of approximately 1 bit/dimension between the* $R^{\rm na}(D)$ *and the zero-delay operational rates whereas at low rates this rate-loss mitigates to 0.5 bits/dimension.*

References

[1] Y. Huang and J. Benesty (eds.), "Audio signal processing for next generation multimedia communication systems," 2004.

[2] 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.

[3] 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.

[4] A. K. Gorbunov and M. S. Pinsker, "Prognostic epsilon entropy of a Gaussian message and a Gaussian source," *Problems Inf. Transmiss.*, vol. 10, no. 2, pp. 93–109, 1974.

[5] 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.

[6] 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.

[7] 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.

[8] O. Ordentlich and U. Erez, "Performance analysis and optimal filter design for sigma-delta modulation via duality with DPCM," in *Proc. IEEE Int. Symp. Inf. Theory*, June 2015, pp. 321–325.

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

[10] T. Berger, *Rate Distortion Theory: A Mathematical Basis for Data Compression.* Englewood Cliffs, NJ: Prentice-Hall, 1971.

[11] 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.

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

[13] L. Vandenberghe, S. Boyd, and S.-P. Wu, "Determinant maximization with linear matrix inequality constraints," *SIAM J. Matrix Anal. and Applic*, vol. 19, no. 2, pp. 499–533, 1998.

[14] O. Ordentlich and U. Erez, "Precoded integer-forcing universally achieves the MIMO capacity to within a constant gap," *IEEE Trans. Inf. Theory*, vol. 61, no. 1, 2015.

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