Sequential Coding of Gauss–Markov Sources With Packet Erasures and Feedback

Anatoly Khina, Victoria Kostina, Ashish Khisti, and Babak Hassibi

***Abstract*—We consider the problem of sequential transmission of Gauss–Markov sources. We show that in the limit of large spatial block lengths, greedy compression with respect to the squared error distortion is optimal; that is, there is no tension between optimizing the distortion of the source in the current time instant and that of future times. We then extend this result to the case where at time t a random compression rate r_t is allocated independently of the rate at other time instants. This, in turn, allows us to derive the optimal performance of sequential coding over packet-erasure channels with instantaneous feedback. For the case of packet erasures with delayed feedback, we connect the problem to that of compression with side information that is known at the encoder and may be known at the decoder — where the most recent packets serve as side information that may have been erased, and demonstrate that the loss due to a delay by one time unit is rather small.**

I. Introduction

Sequential coding of sources is increasingly finding applications, such as real-time video streaming, and cyberphysical and networked control. Such systems use compressed packet-based transmission and strive to achieve minimum distortion for the given compression rates.

The mathematical framework for this setting was provided by Viswanathan and Berger [1] for two time steps, and for more steps (or *sources*) in [2], [3]. A similar control-theoretic framework was also studied by Tatikonda [4], who noticed the connection to the early work of Gorbunov and Pinsker [5].

For the special case of Gauss–Markov sources, an explicit expression for the achievable sum-rate for given distortions was derived in [2] via the paradigms of predictive coding and differential pulse-code modulation (DPCM) [6, Ch. 6] and extended for three (general) jointly Gaussian sources, in [7].

In practice packet-based protocols are prone to erasures and possible delays. The case of sequential coding in the presence of packet erasures was treated for various erasure models. An approach that trades between the performance given all previously sent packets and the performance given only the last packet was proposed in [8]. For random independent identically distributed (i.i.d.) packet erasures, a hybrid between pulse-code modulation (PCM) and DPCM, termed leaky DPCM, was proposed in [9] and analyzed for the case of very low erasure probability in [10]. The scenario in which the erasures occur in bursts was considered in [11].

All of these works assume no feedback is available at the encoder, namely that the encoder does not know whether transmitted packets successfully arrived to the decoder.

In this paper, we first consider the problem of sequential coding of Gauss–Markov sources and determine the rate–distortion region for large spatial blocks (*frames*). Specifically, we show that greedy quantization that optimizes the distortion for each time is also optimal for minimizing the distortion of future time instants. This insight allows us to extend the result to the case where the rate r_t available for the transmission of the packet at time t is determined just prior to its transmission.

The packet-erasure channel with instantaneous output feedback (ACK/NACK) can be viewed as a special case of the above noiseless channel with random rate allocation, with $r_t = 0$ corresponding to a packet-erasure event [12]. The optimal rate–distortion region of sequential coding of Gauss–Markov sources in the presence of packet erasures and instantaneous output feedback thereby follows as a consequence.

We further tackle the delayed feedback setting, in which the encoder does not know whether the most recently transmitted packets arrived or not. By viewing these recent packets as side information (SI) that is available at the encoder and possibly at the decoder, and leveraging the results of Kaspi [13] along with their specialization for the Gaussian case [14], we adapt our transmission scheme to the case of delayed feedback. We provide a detailed description of the proposed scheme for the case where the feedback is delayed by one time unit and demonstrate that the loss due to the delay is small.

We conclude the paper by discussing the cases of large-feedback delays, scalar (fixed- and variable-length) sequential coding and an application to networked control.

II. Problem Statement

We now present the model of the source, channel, and the causal encoder and decoder used in this work.

$\|\cdot\|$ denotes the Euclidean norm. Random variables are denoted by lower-case letters with temporal subscripts ($a_t, \hat{\hat{a}}_t$), and random vectors ("frames") of length N by boldface lower-case letters ($\boldsymbol{a}, \hat{\hat{\boldsymbol{a}}}_t$). We denote temporal sequences by $\boldsymbol{a}^t \triangleq (\boldsymbol{a}_1, \ldots, \boldsymbol{a}_t)$. $\mathbb{N}$ is the set of natural numbers. All other notations represent deterministic scalars.

The communication spans the time interval $[1, T]$, $T \in \mathbb{N}$.

Source: Consider a Gauss–Markov source $\{\boldsymbol{s}_t\}$, whose outcomes are vectors ("*frames*") of length N with i.i.d. samples along the spatial dimension, that satisfy the temporal Markov relation (we assume $\boldsymbol{s}_0 = 0$ for convenience):

A. Khina, V. Kostina, and B. Hassibi are with the Department of Electrical Engineering, California Institute of Technology, Pasadena, CA 91125, USA. E-mails: {*khina, vkostina, hassibi*}*@caltech.edu*

A. Khisti is with the Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON M5S 3G4, Canada. E-mail: *akhisti@comm.utoronto.ca*

The work of A. Khina has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 708932. The work of V. Kostina was supported in part by the National Science Foundation under Grant CCF-1566567. Ashish Khisti was supported by the Canada Research Chairs Program. The work of B. Hassibi was supported in part by the National Science Foundation under grants CNS-0932428, CCF-1018927, CCF-1423663 and CCF-1409204, by a grant from Qualcomm Inc., by NASA's Jet Propulsion Laboratory through the President and Director's Fund, by King Abdulaziz University, and by King Abdullah University of Science and Technology.

$$\boldsymbol{s}_t = \alpha_t \boldsymbol{s}_{t-1} + \boldsymbol{w}_t, \qquad t = 1, \ldots, T\,, \tag{1}$$

where $\{\alpha_t\}$ are known coefficients, and the entries of $\{\boldsymbol{w}_t\}$ are i.i.d. along the spatial dimension, Gaussian and mutually independent across time of zero mean and variances $\{W_t\}$.

Denote by $S_t \triangleq \frac{1}{N}\mathbb{E}\left[\|\boldsymbol{s}_t\|^2\right]$ the average power of the entries of vector $\boldsymbol{s}_t$. Then, we obtain the following recursive relation (with $S_0 = 0$):

$$S_t = \alpha_t^2 S_{t-1} + W_t, \qquad t = 1, \ldots, T\,. \tag{2}$$

Channel: At time t, a packet $f_t \in \left\{1, 2, \ldots, 2^{NR_t}\right\}$ is sent over a noiseless channel of finite rate R_t.

Causal encoder: Sees $\boldsymbol{s}_t$ at time t and applies a causal function $\mathscr{E}_t$ to $\boldsymbol{s}^t$, to generate the packet $f_t \in \left\{1, 2, \ldots, 2^{NR_t}\right\}$:

$$f_t = \mathscr{E}_t\left(\boldsymbol{s}^t\right). \tag{3}$$

Causal decoder: Applies a causal function $\mathscr{D}_t$ to the received packets f^t, to construct an estimate $\hat{\boldsymbol{s}}_t$ of $\boldsymbol{s}_t$, at time t:

$$\hat{\boldsymbol{s}}_t = \mathscr{D}_t\left(f^t\right). \tag{4}$$

Distortion: The quadratic distortion at time t is defined as

$$D_t \triangleq \frac{1}{N}\mathbb{E}\left[\|\boldsymbol{s}_t - \hat{\boldsymbol{s}}_t\|^2\right]. \tag{5}$$

For the special case of an *asymptotically stationary* process:

$$\alpha_t \equiv \alpha, \qquad W_t \equiv W, \qquad t = 1, \ldots, T\,, \tag{6}$$

the source power S_t and distortion D_t converge to

$$S_\infty \triangleq \lim_{T\to\infty} S_t = \frac{W}{1-\alpha^2}\,, \qquad D_\infty \triangleq \lim_{T\to\infty} D_t\,,$$

respectively (assuming that the latter limit exists).

Definition (Distortion-rate region)**.** The *distortion–rate region* is the closure of all achievable distortion tuples $D^T \triangleq (D_1, \ldots, D_T)$ for a rate tuple $R^T \triangleq (R_1, \ldots, R_T)$, for any N, however large; its inverse is the *rate–distortion region*.

III. The Distortion–Rate Region

The optimal achievable distortions for given rates for the model of Sec. II are provided in the following theorem.

Theorem 1 (Distortion–rate region)**.** *The distortion–rate region of sequential coding for a rate tuple R^T is given by all distortion tuples D^T that satisfy $D_t \geq D_t^*$ with $D_0^* = 0$ and*

$$D_t^* = \left(\alpha_t^2 D_{t-1}^* + W_t\right) 2^{-2R_t}, \qquad t = 1, \ldots, T\,. \tag{7}$$

Remark 1. The setting of Th. 1 is referred to as "causal encoder–causal decoder" in [2], where for the case of Gauss-Markov sources an explicit expression is provided only for the sum-rate. Torbatian and Yang [7] extend the sum-rate result to the case of three jointly Gaussian sources (not necessarily Markovian). Our work, on the other hand, fully characterizes the rate–distortion region for Gauss–Markov sources.

A. Achievable

We construct an inner bound using the optimal greedy scheme, which amounts to the classical causal DPCM scheme. In this scheme all the quantizers are assumed to be minimum mean square error (MMSE) quantizers. We note that the quantized values of such quantizers are uncorrelated with the resulting quantization errors.

Scheme (DPCM)**.**

Encoder. At time t:

- Generates the prediction error
$$\tilde{\boldsymbol{s}}_t \triangleq \boldsymbol{s}_t - \alpha_t \hat{\boldsymbol{s}}_{t-1}\,, \tag{8}$$
where $\hat{\boldsymbol{s}}_{t-1}$, defined in (4), is the previous source reconstruction at the decoder, and $\hat{\boldsymbol{s}}_0 = 0$. A linear recursive relation for $\hat{\boldsymbol{s}}_t$ is provided in the sequel in (9).[1]
- Generates $\hat{\tilde{\boldsymbol{s}}}_t$, the quantized reconstruction of the prediction error $\tilde{\boldsymbol{s}}_t$, by quantizing $\tilde{\boldsymbol{s}}_t$ using the MMSE quantizer of rate R_t and frame length N.
- Sends $f_t = \hat{\tilde{\boldsymbol{s}}}_t$ over the channel.

Decoder. At time t:

- Receives f_t.
- Recovers the reconstruction $\hat{\tilde{\boldsymbol{s}}}_t$ of the prediction error $\tilde{\boldsymbol{s}}_t$.
- Generates an estimate $\hat{\boldsymbol{s}}_t$ of $\boldsymbol{s}_t$:
$$\hat{\boldsymbol{s}}_t = \alpha_t \hat{\boldsymbol{s}}_{t-1} + \hat{\tilde{\boldsymbol{s}}}_t\,. \tag{9}$$

Performance analysis. First note that the error between $\boldsymbol{s}_t$ and $\hat{\boldsymbol{s}}_t$, $\boldsymbol{e}_t \triangleq \boldsymbol{s}_t - \hat{\boldsymbol{s}}_t$, is equal to $\boldsymbol{e}_t = \tilde{\boldsymbol{s}}_t - \hat{\tilde{\boldsymbol{s}}}_t$ by (8), (9); thus, the distortion (5) is also the distortion in reconstructing $\tilde{\boldsymbol{s}}_t$.

This, along with (1) and (8) means that $\tilde{\boldsymbol{s}}_t = \alpha_t \boldsymbol{e}_{t-1} + \boldsymbol{w}_t$.

Since $\boldsymbol{w}_t$ is independent of $\boldsymbol{e}_{t-1}$, the average power of the entries of $\tilde{\boldsymbol{s}}_t$ is equal to $\tilde{S}_t = \alpha_t^2 D_{t-1} + W_t$.

Using the property that the rate–distortion function under mean square error distortion of a source with a given average variance is upper bounded by that of an i.i.d. Gaussian source with the same variance (see, e.g., [15, pp. 338–339]), we obtain $D_t \leq \left(\alpha_t^2 D_{t-1} + W_t\right) 2^{-2R_t}$, and hence (7) is achievable within an arbitrarily small $\epsilon > 0$, for a sufficiently large N. ■

B. Impossible (Converse)

Let $N \in \mathbb{N}$. We shall prove

$$\begin{aligned} D_t &\geq 2^{-2R_t}\mathbb{E}_{\check{f}^{t-1}}\left[\mathscr{N}\left(\boldsymbol{s}_t \middle| f^{t-1} = \check{f}^{t-1}\right)\right] && \text{(10a)} \\ &\geq D_t^*, \qquad t = 1, \ldots, T\,, && \text{(10b)} \end{aligned}$$

by induction, where the sequence $\{D_t^*\}$ is defined in (7),

$$\mathscr{N}\left(\boldsymbol{s}_t\right) \triangleq \frac{1}{2\pi\mathrm{e}} 2^{\frac{2}{N}h(\boldsymbol{s}_t)}, \quad \mathscr{N}\left(\boldsymbol{s}_t \middle| f^k = \check{f}^k\right) \triangleq \frac{1}{2\pi\mathrm{e}} 2^{\frac{2}{N}h\left(\boldsymbol{s}_t \middle| f^k = \check{f}^k\right)}$$

denote the entropy-power (EP) and conditional EP of $\boldsymbol{s}_t$ given $f^k = \check{f}^k$, the expectation $\mathbb{E}_{\check{f}^{t-1}}[\cdot]$ is w.r.t. $\check{f}^{t-1}$, and the random vector $\check{f}^t$ is distributed the same as f^t.

Basic step $(t = 1)$**.** Since $\boldsymbol{s}_0 = 0$, and the vector $\boldsymbol{w}_1$ consists of i.i.d. Gaussian entries of variance W_1, (10b) is satisfied with equality. To prove (10a), we use the fact that the optimal achievable distortion D_1 for a Gaussian source ($\boldsymbol{s}_1 = \boldsymbol{w}_1$) with i.i.d. entries of power W_1 and rate R_1 is dictated by its rate–distortion function [15, Ch. 10.3.2]: $D_1 \geq W_1 2^{-2R_1}$.

Inductive step. Let $k \geq 2$ and suppose (10) is true for $t = k-1$. We shall now prove that it holds also for $t = k$.

$$\begin{aligned} D_k &= \frac{1}{N}\mathbb{E}\left[\mathbb{E}\left[\|\boldsymbol{s}_k - \hat{\boldsymbol{s}}_k\|^2 \middle| f^{k-1}\right]\right] && \text{(11a)} \\ &= \frac{1}{N}\mathbb{E}_{\check{f}^{k-1}}\left[\mathbb{E}\left[\|\boldsymbol{s}_k - \hat{\boldsymbol{s}}_k\|^2 \middle| f^{k-1} = \check{f}^{k-1}\right]\right] && \text{(11b)} \\ &\geq \mathbb{E}_{\check{f}^{k-1}}\left[\mathscr{N}\left(\boldsymbol{s}_k \middle| f^{k-1} = \check{f}^{k-1}\right) 2^{-2R_k}\right] && \text{(11c)} \\ &= \mathbb{E}_{\check{f}^{k-1}}\left[\mathscr{N}\left(\alpha_k \boldsymbol{s}_{k-1} + \boldsymbol{w}_k \middle| f^{k-1} = \check{f}^{k-1}\right)\right] 2^{-2R_k} && \text{(11d)} \\ &\geq \Big\{\mathbb{E}_{\check{f}^{k-2}}\left[\mathbb{E}_{\check{f}_{k-1}}\left[\mathscr{N}\left(\alpha_k \boldsymbol{s}_{k-1} \middle| f^{k-1} = \check{f}^{k-1}\right) \middle| \check{f}^{k-2}\right]\right] \\ &\qquad + \mathscr{N}\left(\boldsymbol{w}_k\right)\Big\} 2^{-2R_k} && \text{(11e)} \end{aligned}$$

[1] $\hat{\boldsymbol{s}}_{t-1} = \mathbb{E}\left[\boldsymbol{s}_{t-1} \middle| f^{t-1}\right]$ and $\alpha_t \hat{\boldsymbol{s}}_{t-1} = \mathbb{E}\left[\boldsymbol{s}_t \middle| f^{t-1}\right]$ are the MMSE estimators of $\boldsymbol{s}_{t-1}$ and $\boldsymbol{s}_t$, respectively, given all the past channel outputs.

$$\geq \Big\{\alpha_k^2 \mathbb{E}_{\check{f}^{k-2}}\left[\mathscr{N}\left(\boldsymbol{s}_{k-1}\middle| f^{k-2}=\check{f}^{k-2},\, f_{k-1}\right)\right] + W_k\Big\} 2^{-2R_k} \tag{11f}$$

$$\geq \Big\{\alpha_k^2 \mathbb{E}_{\check{f}^{k-2}}\left[\mathscr{N}\left(\boldsymbol{s}_{k-1}\middle| f^{k-2}=\check{f}^{k-2}\right)\right] 2^{-2R_{k-1}} + W_k\Big\} 2^{-2R_k} \tag{11g}$$

$$\geq 2^{-2R_k}\left(\alpha_k^2 D_{k-1}^* + W_k\right) \tag{11h}$$

$$= D_k^*, \tag{11i}$$

where (11a) follows from (5) and the law of total expectation, (11b) holds since f^{k-1} and $\check{f}^{k-1}$ have the same distribution, (11c) follows by bounding from below the inner expectation (conditional distortion) by the rate–distortion function and the Shannon lower bound [15, Ch. 10] — this also proves (10a), (11d) is due to (1), (11e) follows from the EP inequality [15, Ch. 17], (11f) holds since $\boldsymbol{w}_k$ is Gaussian, the scaling property of differential entropies and Jensen's inequality:

$$\mathbb{E}_{\check{f}_{k-1}}\left[2^{\frac{2}{N}h\left(\boldsymbol{s}_{k-1}\middle|f^{k-1}=\check{f}^{k-1}\right)}\middle|\check{f}^{k-2}\right] \geq 2^{\frac{2}{N}h\left(\boldsymbol{s}_{k-1}\middle|f^{k-2}=\check{f}^{k-2},\, f_{k-1}\right)},$$

(11g) follows from the following standard set of inequalities:

$$\begin{aligned} NR_{k-1} &\geq H\left(f_{k-1}\middle|f^{k-2}=\check{f}^{k-2}\right) \\ &\geq I\left(\boldsymbol{s}_{k-1}; f_{k-1}\middle|f^{k-2}=\check{f}^{k-2}\right) \\ &= h\left(\boldsymbol{s}_{k-1}\middle|f^{k-2}=\check{f}^{k-2}\right) - h\left(\boldsymbol{s}_{k-1}\middle|f^{k-2}=\check{f}^{k-2}, f_{k-1}\right), \end{aligned}$$

(11h) is by the induction hypothesis, and (11i) holds by the definition of $\{D_t^*\}$ (7) — which also proves (10b). ■

C. Asymptotically Stationary Sources

For the asymptotically stationary source in (6), the steady-state average distortion is as follows (formally proved in [16]).

Corollary 1 (Steady state performance with fixed-rate budget)**.** *For the fixed-parameter* (6) *fixed rate budget $R_t \equiv R$ setting:*

$$D_\infty^* \triangleq \lim_{t\to\infty} D_t^* = \frac{W 2^{-2R}}{1-\alpha^2 2^{-2R}}\,.$$

Interestingly, the optimal steady-state distortion achievable with fixed-rate budget (1) is in fact optimal even if we loosen this restriction to a total rate-budget constraint. While this result can be concluded from a classical formula of Gorbunov and Pinsker [5, (1.43)], we provide a simple standalone proof in [16]. The same conclusion holds if the frame entries are correlated Gaussians, as recently proved by by Tanaka [17].

IV. Random-Rate Budgets

In this section we generalize the results of Sec. III to random rates $\{r_t\}$ that are independent of each other and of $\{\boldsymbol{w}_t\}$. r_t is revealed to the encoder just before the transmission at time t.

Theorem 2 (Distortion–rate region)**.** *The distortion–rate region of sequential coding with independent rates r^T is given by all distortion tuples D^T that satisfy $D_t \geq D_t^*$ with ($D_0^* = 0$):*

$$D_t^* = \left(\alpha_t^2 D_{t-1}^* + W_t\right)\mathbb{E}\left[2^{-2r_t}\right], \qquad t = 1,\ldots,T\,. \tag{13}$$

Proof: *Achievable.* Since the achievability scheme in Th. 1 does not use the knowledge of future transmission rates to encode and decode the packet at time t, we have

$$d_t \triangleq \frac{1}{N}\mathbb{E}\left[\left\|\boldsymbol{s}_t - \hat{\boldsymbol{s}}_t\right\|^2 \middle| r^T\right] \tag{14a}$$

$$= \frac{1}{N}\mathbb{E}\left[\left\|\boldsymbol{s}_t - \hat{\boldsymbol{s}}_t\right\|^2 \middle| r^t\right] \tag{14b}$$

$$\leq (\alpha_t^2 d_{t-1} + W_t) 2^{-2r_t} + \epsilon, \tag{14c}$$

for any $\epsilon > 0$, however small, and large enough N.

Taking an expectation of (14c) with respect to r^t and using the independence of r^{t-1} and r_t, we obtain (13).

Impossible. Revealing the rates to the encoder and the decoder prior to the start of transmission can only improve the distortion. Thus, the distortions $\{d_t\}$ conditioned on $\{r_t\}$ (14a) are bounded from below as in Th. 1; by taking the expectation w.r.t. $\{r_t\}$, we attain the desired result. ■

For the special case of an asymptotically stationary source (6), the steady-state distortion is given as follows (again, see [16], for a formal proof).

Corollary 2 (Steady state)**.** *For the fixed-parameter* (6) *setting with i.i.d. rates $\{r_t\}$:*

$$D_\infty^* \triangleq \lim_{t\to\infty} D_t^* = \frac{BW}{1-\alpha^2 B}, \qquad B \triangleq \mathbb{E}\left[2^{-2r_t}\right].$$

V. Packet Erasures with Instantaneous Feedback

An important special case of the model of Sec. IV is that of packet erasures [12]. Since a packet erasure at time t can be viewed as $r_t = 0$, and assuming that the encoder sends packets of fixed rate R and is cognizant of any packet erasures instantaneously, the packet erasure channel can be cast as the random rate channel of Sec. IV with

$$r_t = b_t R = \begin{cases} R, & b_t = 1 \\ 0, & b_t = 0 \end{cases} \tag{15}$$

where $\{b_t\}$ are the packet-erasure events, such that $b_t = 1$ corresponds to a successful arrival of the packet f_t at time t, and $b_t = 0$ means it was erased. We further denote by $g_t \triangleq b_t f_t$ the received output where $g_t = 0$ corresponds to an erasure, and otherwise $g_t = f_t$. We assume that $\{b_t\}$ are i.i.d. according to a $\mathscr{Ber}(\beta)$ distribution for $\beta \in [0,1]$.

Remark 2. We shall concentrate on the case of packets of fixed rate R to simplify the subsequent discussion. This way the only randomness in rate comes from the packet-erasure effect. Nevertheless, all the results that follow can be easily extended to random/varying rate allocations to which the effect of packet erasures $\{b_t\}$ is added in the same manner as in (15).

Corollary 3 (Distortion–rate region)**.** *The distortion–rate region with packet erasures and instantaneous feedback is given as in Th. 2 with $B = 1 - \beta\left(1 - 2^{-2R}\right)$.*

Corollary 4 (Steady state)**.** *The steady-state distortion is given as in Corol. 2 with $B = 1 - \beta\left(1 - 2^{-2R}\right)$.*

Remark 3. This scenario can be extended to the case of multiple packets per frame by determining the probability distribution of the rate; see [16] for further details.

VI. Packet Erasures with Delayed Feedback

In this section we consider the case of i.i.d. packet erasures with a delayed-by-one output feedback, i.e., the case where at time t, the encoder does know whether the last packet arrived or not (does not know b_{t-1}), but knows the erasure pattern of all preceding packets (knows b^{t-2}). The encoder (3) and decoder (4) mappings can be written as [recall that $g_t \triangleq b_t f_t$]:

$$f_t = \mathscr{E}_t\left(\boldsymbol{s}^t, g^{t-2}\right), \qquad \hat{\boldsymbol{s}}_t = \mathscr{D}_t\left(g^t\right).$$

To that end, we recall the following result by Perron *et al.* [14, Th. 2], which is a specialization to the jointly Gaussian case of the result by Kaspi [13, Th. 1], who established the rate–distortion region of lossy compression with two-sided SI where the SI may or may not be available at the decoder.[2,3]

Theorem 3 ([14])**.** *Let $\boldsymbol{s}$ be an i.i.d. zero-mean Gaussian source of power S, which is jointly Gaussian with SI $\boldsymbol{y}$, which is available at the encoder and satisfies $\boldsymbol{s} = \boldsymbol{y} + \boldsymbol{z}$ where $\boldsymbol{z}$ is an i.i.d. Gaussian noise of power Z that is independent of $\boldsymbol{y}$. Denote by $\hat{\boldsymbol{s}}^+$ and $\hat{\boldsymbol{s}}^-$ the reconstructions of $\boldsymbol{s}$ with and without the SI $\boldsymbol{y}$, and by D^+ and D^- their mean squared error distortion requirements, respectively. Then, the smallest rate required to achieve these distortions is given by*

$$R^{\text{Kaspi}}(S, Z, D^-, D^+)$$
$$= \begin{cases} 0, & D^- \geq S \text{ and } D^+ \geq Z \\ \frac{1}{2}\log\left(\frac{S}{D^-}\right), & D^- < S \text{ and } D^+ \| S \geq D^- \| Z \\ \frac{1}{2}\log\left(\frac{Z}{D^+}\right), & D^+ < Z \text{ and } D^- \geq D^+ + S - Z \\ \frac{1}{2}\log\left(\frac{S}{D^- - \Delta^2}\right), & \begin{cases} D^- < S \text{ and } D^+ \| S < D^- \| Z \\ \text{and } D^- < D^+ + S - Z \end{cases} \end{cases}$$

where $a \| b \triangleq \frac{ab}{a+b}$ denotes the harmonic mean of a and b, and

$$\Delta \triangleq \frac{\sqrt{(S-Z)(S-D^-)}D^+ - \sqrt{(Z-D^+)(D^- - D^+)}S}{\sqrt{Z}\,(S-D^+)}.$$

Remark 4. Surprisingly, as observed by Perron *et al.* [14], if the SI signal $\boldsymbol{y}$ is not available at the encoder — a setting considered in [18] and [13, Th. 2] — the required rate can be strictly higher than that in Th. 3. This is in stark contrast to the case where the SI is not available at the encoder and the case where the SI is always available at the decoder (the "Wyner–Ziv Problem") [15, Ch. 15.9]. Knowing the SI at the encoder allows to (anti-)correlate the noise $\boldsymbol{z}$ with the quantization error — a thing that is not possible when the SI is not available at the encoder, as the two noises must be independent in that case. This allows for some improvement, though a modest one, as implied by the results for the dual channel problem [20].

In our case, at time t, the previous packet f_{t-1} will serve as the SI. Note that it is always available to the encoder; the decoder may or may not have access to it, depending whether the previous packet arrived or not. Since the feedback is delayed, during the transmission of the current packet f_t the encoder does not know whether the previous packet was lost.

The tradeoff between D^+ and D^- for a given rate R will be determined by the probability of a successful packet arrival β.

Scheme (Kaspi-based)**.**

Encoder. At time t:

- Generates the prediction error $\tilde{\boldsymbol{s}}_t \triangleq \boldsymbol{s}_t - \alpha_t \alpha_{t-1} \hat{\boldsymbol{s}}_{t-2}$.
- Generates f_t by quantizing the prediction error $\tilde{\boldsymbol{s}}_t$ as in Th. 3, where f_{t-1} is available as SI at the encoder and possibly at the decoder (depending on b_{t-1}) using the optimal quantizer of rate R and frame length N that minimizes the averaged over b_{t-1} distortion:
$$D_t^{\text{Weighted}} = \beta D_t^+ + (1-\beta) D_t^-\,; \qquad (16)$$
more precisely, since the encoder does not know (b_{t-1}, b_t) at time t:
  - Denote the reconstruction of $\tilde{\boldsymbol{s}}_t$ at the decoder from f_t and g^{t-1} — namely given that $b_t = 1$ — by $Q_t(\tilde{\boldsymbol{s}}_t)$, and the corresponding distortion, averaged over b_{t-1}, by D_t^{Weighted}.
  - Denote the reconstruction from f_t and g^{t-2} — namely given that $b_t = 1$ and $b_{t-1} = 0$ — by $Q_t^-(\tilde{\boldsymbol{s}}_t)$, and the corresponding distortion by D_t^-.
  - Denote the reconstruction from (f_{t-1}, f_t) and g^{t-2} — namely given that $b_t = 1$ and $b_{t-1} = 1$ — by $Q_t^+(\tilde{\boldsymbol{s}}_t)$, and the corresponding distortion by D_t^+.

  Then, the encoder sees $\alpha_t Q_{t-1}(\tilde{\boldsymbol{s}}_{t-1})$ as possible SI available at the decoder to minimize D_t^{Weighted} as in (16).
- Sends f_t over the channel.

Decoder. At time t:

- Receives g_t.
- Generates a reconstruction $\hat{\tilde{\boldsymbol{s}}}_t$ of the prediction error $\tilde{\boldsymbol{s}}_t$:
$$\hat{\tilde{\boldsymbol{s}}}_t = \begin{cases} Q_t^+(\tilde{\boldsymbol{s}}_t), & b_t = 1, b_{t-1} = 1 \\ Q_t^-(\tilde{\boldsymbol{s}}_t), & b_t = 1, b_{t-1} = 0 \\ 0, & b_t = 0 \end{cases} \qquad (17)$$
- Generates an estimate $\hat{\boldsymbol{s}}_t$ of $\boldsymbol{s}_t$: $\hat{\boldsymbol{s}}_t = \alpha_t \hat{\boldsymbol{s}}_{t-1} + \hat{\tilde{\boldsymbol{s}}}_t$.

This scheme is the optimal greedy scheme whose performance is stated next, in the limit of large N.

Theorem 4. *Let $\epsilon > 0$, however small. Then, for a large enough N, the expected distortion of the scheme at time $t \in [2, T]$ given $(b_1, \ldots, b_t)$ satisfies the recursion*

$$D_t \leq \begin{cases} D_t^+ + \epsilon, & b_t = 1, b_{t-1} = 1 \\ D_t^- + \epsilon, & b_t = 1, b_{t-1} = 0 \\ \alpha_t^2 D_{t-1} + W + \epsilon, & b_t = 0 \end{cases}$$
$$D_1 = D_1^+ = D_1^- = W_t 2^{-b_1 2R} + \epsilon,$$

where D_t^+ and D_t^- are the distortions that minimize (16)*, such that the rate of Th. 3 satisfies*

$$R^{\text{Kaspi}}(\alpha_t D_{t-1}^- + W, \alpha_t D_{t-1}^+ + W, D_t^-, D_t^+) = R.$$

The proof is again the same as that of Ths. 1 and 2, with $\hat{\tilde{\boldsymbol{s}}}_t$ generated as in (17).

Remark 5. Here, in contrast to the case of instantaneous feedback, evaluating the distortions $\{D_t\}$ in explicit form (recall Corol. 3) is more challenging. We do it numerically, instead.

Somewhat surprisingly, the loss in performance of the Kaspi-based scheme due to the feedback delay is rather small compared to the scenario in Sec. V where the feedback is available instantaneously, for all values of β.[4] This is demonstrated in Fig. 1, where the perfomances of these schemes are compared along with the performances of the following three simple schemes for $\alpha_t \equiv 0.7, W \equiv 1, \beta = 0.5, R = 2$:

- **No prediction:** A scheme that uses no prediction at all, as if the source samples were independent. This scheme achieves a distortion of $D_t = \beta S_t 2^{-2R} + (1-\beta) S_t$, where S_t is the power of the entries of $\boldsymbol{s}_t$ as given in (2).
- **Assumes worst case (WC):** Since at time t the encoder does not know b_{t-1}, a "safe" way would be to work as if $b_{t-1} = 0$. This achieves a distortion of

[2] We use a backward channel to represent the SI $\boldsymbol{s} = \boldsymbol{y} + \boldsymbol{z}$, as opposed to the forward channel $\boldsymbol{y} = \boldsymbol{s} + \boldsymbol{z}$ used in [14], [18].

[3] Kaspi's result [13, Th. 1] can also be viewed as a special case of [18] with some adjustments; see [19].

[4] For β values close to 0 or 1, the loss becomes even smaller as in these cases using the scheme of Sec. V that assumes that the previous packet arrived or was erased, respectively, becomes optimal.

$$D_t = \left[\alpha^4 D_{t-2} + (1+\alpha^2)W\right]\left[\beta 2^{-2R} + (1-\beta)^2\right] + \beta(1-\beta)(\alpha^2 D_{t-1} + W), \qquad t = 2, \ldots, T,$$
$$D_0 = 0, \qquad D_1 = W 2^{-2R}.$$

- **Assumes best case (BC):** The optimistic counterpart of the previous scheme is that which always works as if $b_{t-1} = 1$. This scheme achieves a distortion of

$$D_t = \beta\left[\alpha^2 D_{t-1|t-2} 2^{-2R} + W\right)\left[\beta 2^{-2R} + (1-\beta)\right] + (1-\beta)\left[\alpha^2 D_{t-1|t-2} + W\right], \quad t = 2, \ldots, T,$$
$$D_{t-1|t-2} \triangleq \alpha^2 D_{t-2} + W, \qquad t = 2, \ldots, T,$$
$$D_0 = 0, \quad D_1 = W 2^{-2R}.$$

VII. Discussion

A. Feedback with Larger Delays

To extend the delayed feedback scheme of Sec. VI to larger delays, a generalization of Th. 3 is needed. Unfortunately, the optimal rate–distortion region for more than two decoders remains an open problem and is only known for the ("degraded") case when the source and the possible SIs form a Markov chain. Nonetheless, achievable regions for multiple decoders have been proposed in [18], which can be used for the construction of schemes that accommodate larger delays.

B. Scalar Sequential Coding

In this paper we derived lower bounds and proved that they are tight for large values of N. In the case of scalar fixed-length quantization, the design and analysis of good schemes are more involved. For treatment for the case of log-concave distributions (Gaussian included), see [21]. Alternatively, by relaxing the rate constraint to hold only on average, one may invoke scalar ECDQ [22, Ch. 5] to attain [16]:

$$D_t^{\mathrm{ECDQ}} \leq \frac{\pi e}{12}\left(\alpha_t^2 D_{t-1} + W_t\right)\mathbb{E}\left[2^{-2r_t}\right].$$

C. Non-Gaussian

The lower bounds in this work can be extended to the case of a non-Gaussian driving process $\boldsymbol{w}_t$ in a straightforward fashion, with the variance of $\boldsymbol{w}_t$ in (13) replaced by its entropy power (recall that the two are equal in the Gaussian case).

D. Networked Control

The results in this work (including those in the discussion) can be easily adapted to the case of packet-based networked control and provide upper and lower bounds on the linear quadratic regulator (LQR) cost reminiscent of [23], by employing the control-theoretic separation principle [24]; see [16].

Acknowledgment

The authors thank Y. Su for valuable discussions.

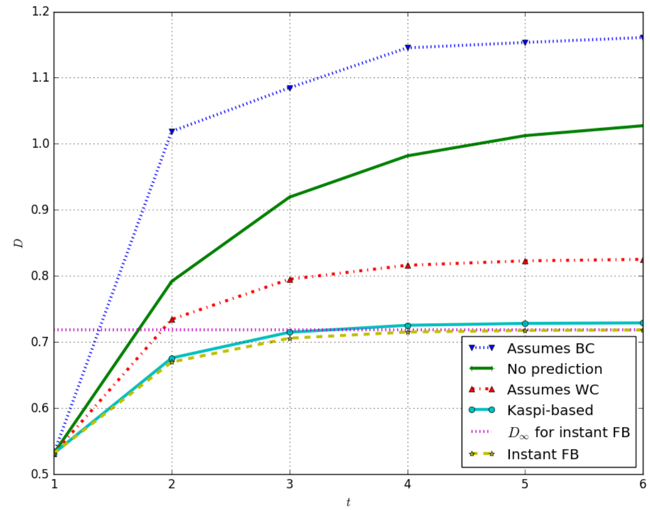

Fig. 1. Distortions D_t as a function of the time t of the various schemes presented in this section, along with that of the instantaneous-feedback scheme of Sec. V, for $\alpha = 0.7$, $W = 1$, $\beta = 0.5$ and $R = 2$.

References

[1] H. Viswanathan and T. Berger, "Sequential coding of correlated sources," *IEEE Trans. Inf. Theory*, vol. 46, no. 1, pp. 236–246, Jan. 2000.

[2] N. Ma and P. Ishwar, "On delayed sequential coding of correlated sources," *IEEE Trans. Inf. Theory*, vol. 57, no. 6, pp. 3763–3782, 2011.

[3] E.-H. Yang, L. Zheng, and D.-K. He, "Rate distortion theory for causal video coding: Characterization, computation algorithm, and comparison," *IEEE Trans. Inf. Theory*, vol. 57, no. 8, pp. 5258–5280, 2011.

[4] S. C. Tatikonda, "Control under communication constraints," Ph.D. dissertation, Cambridge, CA, USA, Sep. 2000.

[5] A. K. Gorbunov and M. S. Pinsker, "Prognostic epsilon entropy of a Gaussian message and a Gaussian source," *Problemy Pered. Info. (Problems of Info. Trans.)*, vol. 10, no. 2, pp. 5–25, 1974.

[6] N. S. Jayant and P. Noll, *Digital Coding of Waveform*. Englewood Cliffs, NJ: Prentice-Hall, 1984.

[7] M. Torbatian and E.-H. Yang, "Causal coding of multiple jointly Gaussian sources," in *Proc. Annual Allerton Conf. on Comm., Control, and Comput.*, Monticello, IL, USA, Oct. 2012, pp. 2060–2067.

[8] L. Song, J. Chen, J. Wang, and T. Liu, "Gaussian robust sequential and predictive coding," *IEEE Trans. Inf. Theory*, vol. 59, no. 6, pp. 3635–3652, June 2013.

[9] H. C. Huang, W. H. Peng, and T. Chiang, "Advances in the scalable amendment of H.264/AVC," *IEEE Comm. Magazine*, vol. 45, no. 1, pp. 68–76, Jan. 2007.

[10] Y.-Z. Huang, Y. Kochman, and G. W. Wornell, "Causal transmission of colored source frames over packet erasure channel," in *Proc. Data Comp. Conf. (DCC)*, Snowbird, UT, USA, Mar. 2010, pp. 129–138.

[11] F. Etezadi, A. Khisti, and M. Trott, "Zero-delay sequential transmission of Markov sources over burst erasure channels," *IEEE Trans. Inf. Theory*, vol. 60, no. 8, pp. 4584–4613, Aug. 2014.

[12] P. Minero, M. Franceschetti, S. Dey, and G. N. Nair, "Data rate theorem for stabilization over time-varying feedback channels," *IEEE Trans. Autom. Control*, vol. 54, no. 2, pp. 243–255, Feb. 2009.

[13] A. H. Kaspi, "Rate–distortion when side-information may be present at the decoder," *IEEE Trans. Inf. Theory*, vol. 40, no. 6, pp. 2031–2034, Nov. 1994.

[14] E. Perron, S. Diggavi, and I. E. Telatar, "On the role of encoder side-information in source coding for multiple decoders," in *Proc. IEEE Int. Symp. on Inf. Theory (ISIT)*, Seattle, WA, USA, July 2006, pp. 331–335.

[15] T. M. Cover and J. A. Thomas, *Elements of Information Theory, Second Edition*. New York: Wiley, 2006.

[16] A. Khina, V. Kostina, A. Khisti, and B. Hassibi, "Sequential coding of Gauss–Markov sources with and without packet erasures," *CoRR*, 2017. [Online]. Available: arxiv.org/abs/1702.01779

[17] T. Tanaka, "Semidefinite representation of sequential rate–distortion function for stationary Gauss–Markov processes," in *IEEE Conf. Control App. (CCA)*, Sydney, NSW, Australia, Sep. 2015, pp. 1217–1222.

[18] C. Heegard and T. Berger, "Rate–distortion when side information may be absent," *IEEE Trans. Inf. Theory*, vol. 31, pp. 727–734, Nov. 1985.

[19] A. Khina and U. Erez, "Source coding with composite side information at the decoder," in *Proc. IEEE Conv. Electrical and Electron. Engineers in Israel (IEEEI)*, Eilat, Israel, Nov. 2012.

[20] R. Zamir and U. Erez, "A Gaussian input is not too bad," *IEEE Trans. Inf. Theory*, vol. 50, no. 6, pp. 1362–1367, Jun. 2004.

[21] A. Khina, Y. Nakahira, Y. Su, and B. Hassibi, "Algorithms for optimal control with fixed-rate feedback," in *Proc. IEEE Conf. Decision and Control (CDC)*, Dec. 2018.

[22] R. Zamir, *Lattice coding for signals and networks*. Cambridge: Cambridge University Press, 2014.

[23] V. Kostina and B. Hassibi, "Rate–cost tradeoffs in control," in *Proc. Annual Allerton Conf. on Comm., Control, and Comput.*, Monticello, IL, USA, Sep. 2016, pp. 1157–1164.

[24] S. Tatikonda, A. Sahai, and S. K. Mitter, "Stochastic linear control over a communication channel," *IEEE Trans. Autom. Control*, vol. 49, no. 89, pp. 1549–1561, Sep. 2004.