Gauss–Markov Source Tracking with Side Information: Lower Bounds

Omri Lev and Anatoly Khina
{omrilev@mail,anatolyk@eng}.tau.ac.il
School of Electrical Engineering, Tel Aviv University, Tel Aviv, Israel

***Abstract*—We consider the problem of causal source coding and causal decoding of a Gauss–Markov source, where the decoder has causal access to a side-information signal. We define the information causal rate–distortion function with causal decoder side information and prove that it bounds from below its operational counterpart.**

I. Introduction

Motivated by recent advances in tracking and control over networks [1]–[17], we consider the setting where a decoder observes the system state corrupted by noise via an internal sensor, while it also receives quantized descriptions of the observations of the state from an external sensor over a rate-limited link.

We focus in this paper on the tracking (estimation) problem of a Gauss–Markov source over a rate-limited channel, i.e., causal encoding and decoding of the source; we view the internal noisy measurements of the state as side information that is available to the decoder but not to the encoder.

The idea of causal rate–distortion function (CRDF) was introduced in [18], where [4], [19], [20] (see also [15], [21]) drew the connection between the CRDF and tracking of a Gauss–Markov source over rate-limited links with causal encoding and decoding. Recently, two notable efforts have been made in determining bounds on the performance of these settings in the presence of decoder SI [22], [23], which provide a comprehensive set of definitions and bounds for this problem, by relying on the seminal work of Wyner and Ziv [24], [25] for rate–distortion with non-causal SI at the decoder. However, since the technique of Wyner and Ziv relies on non-causal knowledge of the SI at the decoder, applying it for scenarios with causal SI imposes an additional slack when used to bound from below the operational CRDF with (*causal*) SI, on top of the existing gap between the information and operational CRDFs without SI that stems from the causal encoding restriction [26], [27].

Our goal in this paper is twofold: first, providing short proofs of the lower bounds in [23] via a simple observation; secondly, deriving a tighter lower bound on the performance of causal source coding with decoder SI that is strictly higher than the bounds in [22], [23]. To derive the latter, we build on the work of Weissman and El Gamal [28] for rate—distortion with *causal* SI and extend their results for CRDFs.

This work 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 O. Lev was supported in part by the Yitzhak and Chaya Weinstein Research Institute for Signal Processing.

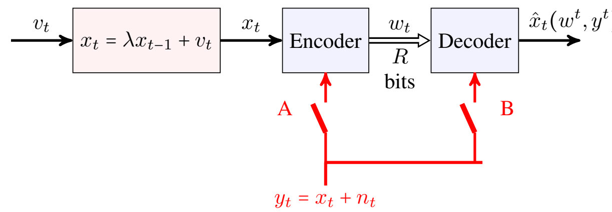


Fig. 1: Scalar tracking system with driving WGN. The channel is a bit pipe with instantaneous rate constraint R_t. The presence of the SI in the encoder/decoder is according to the state of switch A/B, respectively. We assume that the SI is the original source x_t after passing through a Gaussian channel.

As a by product, we settle a conjecture in the negative by Stavrou and Skoglund [23] regarding the optimality of Wyner–Ziv-type CRDF bounds for causal tracking over additive white Gaussian noise (AWGN) channels, by proving that an adaptation of our new lower bound is strictly higher for this setting. The rest of the paper is organized as follows. In Sec. II, we formulate the problem of tracking a Gauss–Markov source over a rate-limited link for several different SI scenarios. We review classical results and tools that are used throughout this work in Sec. III. We review the CRDF scenario without SI in Sec. IV, and with two-sided SI Sec. V. We provide simple proofs for the existing results along with new tighter bounds on the CRDF with decoder SI in Sec. VI. We evaluate the expression of the new bound for a Gaussian test channel in Sec. VII. Finally, we conclude the paper with Sec. VIII by discussing future research directions.

II. Problem Statement

In this section, we formalize the tracking setting treated in this work, depicted in Fig. 1.

Source. The source is generated by a first order Gauss–Markov model with zero initial condition ($x_0 = 0$):[1]

$$x_t = \lambda x_{t-1} + v_t, \qquad t = 1, \ldots, T, \tag{1}$$

where $x_t \in \mathbb{R}$ is the source sample at time t; v_t is the system disturbance at time t, whose temporal entries are independent and identically distributed (i.i.d.) zero-mean Gaussian of variance $\sigma_v^2 > 0$; the eigenvalue $\lambda \in \mathbb{R}$ is fixed and known.

Encoder. Observes the state x_t at time t and generates a packet $a_t \in \left\{1, \ldots, 2^{R_t}\right\}$ of rate R_t.

[1]The assumption $x_0 = 0$ can be easily replaced with a Gaussian x_0 that is independent of the system-disturbance sequence $\{v_t\}$.

Channel. At time t, a packet $a_t \in \{1, 2, \dots, 2^{R_t}\}$ is sent over a noiseless channel with rate R_t. The packets are subject to an average-rate constraint:[2]

$$\frac{1}{T} \sum_{t=1}^{T} R_t \le R.$$

Side information. The SI is a noisy version of the current source sample x_t, and is given by

$$y_t = x_t + n_t,$$

where n_t is zero-mean Gaussian of variance σ_n^2, independent of x^t,[3] and its temporal entries are i.i.d.

Decoder. At time t, receives the packet a_t and constructs an estimate $\hat{x}_t$ of x_t.

Distortion. The average quadratic distortion at time t is defined as

$$D_t = \mathbb{E}\left[(x_t - \hat{x}_t)^2\right], \tag{2}$$

and the average-stage distortion is defined as

$$D = \frac{1}{T} \sum_{t=1}^{T} D_t. \tag{3}$$

Definition 1 (Operational causal rate–distortion function)**.** The operational causal rate–distortion function (CRDF) $R_{c,op}(D)$ is defined as the infimum of all achievable average rates R, $\frac{1}{T}\sum_{t=1}^{T} R_t = R$, subject to an average distortion constraint $\frac{1}{T}\sum_{t=1}^{T} D_t \le D$.

Different scenarios for the availability of the SI may be considered, corresponding to different states of switches A and B in Fig. 1:

- *No SI (A open, B open).* The encoder applies a *causal function* $\mathcal{F}_t$ to the source history x^t, to generate the packet $a_t \in \{1, \dots, 2^{R_t}\}$: $a_t = \mathcal{F}_t(x^t)$, whereas the decoder applies a *causal function* $\mathcal{G}_t$ to the sequence of received packets a^t, to construct an estimate $\hat{x}_t$ of x_t: $\hat{x}_t = \mathcal{G}_t(a^t)$.
- *Two-sided SI (A closed, B closed).* Here, both the encoder and the decoder have access to the SI and hence $a_t = \mathcal{F}_t(x^t, y^t)$ and $\hat{x}_t = \mathcal{G}_t(a^t, y^t)$.
- *Decoder SI (A open, B closed).* Here, only the decoder has access to the SI. Thus, $a_t = \mathcal{F}_t(x^t)$ and $\hat{x}_t = \mathcal{G}_t(a^t, y^t)$.

III. Background

A. Batch Rate–Distortion

In this section we review classical results from information theory on lossy compression. The standard mode of operation assumes batch operation over long blocks ($T \to \infty$): The encoder observes a long block of source samples x^T, and maps them together to a (single) packet a; the decoder recovers the estimates $\hat{x}^T$ of the the entire sequence upon receiving a, i.e., in a non-causal fashion [cf. (2)].

Within this framework, information theory discriminates between four different scenarios of the availability of SI and its nature, which we present next for the commonly-considered case of an *i.i.d.* Gaussian source, corresponding to taking $\lambda = 0$ in (1):

- *No SI.* This is the classical rate–distortion scenario [29], [30, Ch. 10]. For which the rate–distortion function (RDF) is equal to
$$R(D) = \frac{1}{2}\log^+ \frac{\sigma_v^2}{D},$$
where $\log^+(x) \triangleq \max\{\log x, 0\}$.
- *Two-sided SI.* This scenario can be recast as that of no SI with additional conditioning, as both the encoder and the decoder know the SI. Thus, conditional RDF amounts to
$$R^{\text{both}}(D) = \frac{1}{2}\log^+ \frac{\sigma_{v|y}^2}{D} = \frac{1}{2}\log^+ \frac{\sigma_v^2 \| \sigma_n^2}{D}, \tag{4}$$
where $\sigma_{a|b}^2$ denotes the conditional variance of a given b, and $a\|b \triangleq ab/(a+b)$.
- *Decoder non-causal SI.* Here, for the reconstruction of x_t ($t \in \{1, \dots, T\}$), the decoder may use the entire side information sequence y^T in addition to a, whereas the encoder is oblivious of y^T. Surprisingly, a classical result due to Wyner [24] (an adaptation to the Gaussian case of a result by Wyner and Ziv [25]) states that, for an i.i.d. Gaussian source, the RDF for this scenario, R^{NC}, coincides with that of (4), i.e., $R^{\text{NC}}(D) \equiv R^{\text{both}}(D)$.
- *Decoder causal SI.* This scenario is identical to the previous one except that now, for the reconstruction $\hat{x}_t$ of x_t at time t, in addition to a, the decoder may use only the *causal history* of the SI y^t. Weissman and El Gamal [28] have shown that the RDF for this scenario is given by[4]
$$R^{\text{C}}(D) = \inf_{\substack{P(w|x):\, y \to x \to w, \\ \mathbb{E}\left[(x - \hat{x}(w,y))^2\right] \le D}} I(x; w) \tag{5}$$
and is higher than (4). Furthermore, it is bounded from above by
$$R^{\text{C}}(D) \le \text{c.e.}\left\{\frac{1}{2}\log^+\left(\frac{\sigma_v^2}{D} - \frac{\sigma_v^2}{\sigma_n^2}\right)\right\} \triangleq \text{c.e.}\left\{r(D)\right\}.$$
where c.e. denotes the *convex envelope* operation, and is manifested by a straight line between the points $(D_c, r(D_c))$ and $(\sigma_n^2 \| \sigma_v^2, 0)$ in the regime $D \in (D_c, D_{\max})$, where D_c is the solution to the equation $r(D_c) = (D_c - \sigma_v^2 \| \sigma_n^2) \frac{d}{dD} r(D)\big|_{D = D_c}$; the convex envelope comes into play only when $D_c < \sigma_n^2 \| \sigma_v^2$, i.e., only when $\sigma_n^2 < \sigma_v^2$.

Remark 1. The RDFs for the different scenarios serve as an outer bound for finite T and are attainable only in the limit of $T \to \infty$. However, as have been proved by Zamir and Linder [26], even in the limit of $T \to \infty$ (and even for i.i.d. Gaussian sources) they are not attainable, in general (although they can be approached up to a fixed additive loss [31, Ch. 5]). Finally, note that for the batch setting these results may be extended beyond the i.i.d. setting ($\lambda \neq 0$); see [32], [33].

Remark 2. When the side information is known to both the encoder and the decoder, it turns out that the RDFs coincide

[2] This is a more lenient constraint than the fixed-rate constraint. Consequently, our lower bounds are valid for both scenarios, although they might be too optimistic for the latter.

[3] We denote temporal sequences by $a^t \triangleq (a_1, \dots, a_t)$.

[4] $a \to b \to c$ denotes a Markov chain, i.e., given b, a is independent of c.

for the cases when the SI is known causally and non-causally. Therefore, we do not distinguish between these two scenarios.

B. Directed Information

The *Directed Information* (DI) notion, introduced by Massey [34], is the causal counterpart of the classical *Mutual Information* MI and is defined as follows.

Definition 2 (DI)**.** The DI between x^T and y^T is defined as

$$I\left(x^T \to y^T\right) = \sum_{t=1}^{T} I\left(x^t; y_t \middle| y^{t-1}\right) \tag{6a}$$

$$= \mathbb{D}\left(P(y^T \Uparrow x^T) \middle\| P_{y^T} \middle| P_{x^T}\right), \tag{6b}$$

where $I(\cdot;\cdot|\cdot)$ denotes the conditional MI, $\mathbb{D}(\cdot\|\cdot|\cdot)$ is the conditional *Kullback–Leibler divergence*, and

$$P(y^T \Uparrow x^T) \triangleq \prod_{t=1}^{T} P\left(y_t | y^{t-1}, x^t\right)$$

is the *causally conditional probability kernel* [35, Ch. 3], [22].

Clearly, $0 \le I\left(x^T \to y^T\right) \le I\left(x^T; y^T\right)$, and for a sequence of independent pairs $\{(x_t, y_t)\}_{t=1}^T$, the DI and the MI coincide (see [35, Ch. 3] for further details).

The causally conditional DI is defined next and allows, in turn, to derive a chain-rule and a *Data-Processing Inequality* (DPI) for DIs.

Definition 3. The *causally conditional DI* is defined as

$$I\left(x^T \to y^T \Uparrow z^T\right) \triangleq \sum_{t=1}^{T} I\left(x^t; y_t \middle| y^{t-1}, z^t\right),$$

and its lagged-by-one variant—as

$$I\left(x^T \to y^T \Uparrow z^{T-1}\right) \triangleq \sum_{t=1}^{T} I\left(x^t; y_t \middle| y^{t-1}, z^{t-1}\right). \tag{7}$$

Theorem 1 (Chain rule for DIs [34], [35, Ch. 3])**.**

$$I\left(\left(x^T, y^T\right) \to z^T\right) = I\left(x^T \to z^T\right) + I\left(y^T \to z^T \Uparrow x^T\right),$$

$$I\left(x^T \to \left(y^T, z^T\right)\right) = I\left(x^T \to y^T \Uparrow z^{T-1}\right) + I\left(x^T \to z^T \Uparrow y^T\right). \tag{8a}$$

Theorem 2 (DPI for DIs [1], [21])**.** *Let u^T, a^T, x^T satisfy the Markov relations $\left(x_t, a^{t-1}\right) \to \left(a^t, u^{t-1}\right) \to u_t$ for all $t \in \{1, 2, \ldots, T\}$. Then,*

$$I\left(x^T \to u^T\right) \le I\left(x^T \to a^T \Uparrow u^{T-1}\right).$$

IV. No SI

In this section we review known results for the scenario where SI is available to neither the encoder nor the decoder, corresponding to switches A and B being open in Fig. 1.

Definition 4 ([18])**.** The information CRDF of a Gaussian source $\{x_t\}$ (without SI) is defined as

$$R_c(D) = \overline{\lim_{T\to\infty}} \inf_{\substack{P(\hat{x}^T \Uparrow x^T), \\ \frac{1}{T}\sum_t \mathbb{E}\left[\|x_t - \hat{x}_t\|^2\right] \le D}} \frac{1}{T} I\left(x^T \to \hat{x}^T\right) \tag{9a}$$

$$= \frac{1}{2}\log^+ \frac{\lambda^2 D + \sigma_v^2}{D}. \tag{9b}$$

(9b) is derived in [3], [5], [18], [36].

Theorem 3. *The operational CRDF (without SI), $R_{c,op}(D)$, is bounded from below by the information CRDF (without SI)* (9)*: $R_c(D) \le R_{c,op}(D)$.*

For a detailed proof see [3], [5].

Remark 3. As mentioned in Rem. 1, equality in the lower bound of Th. 3 *cannot* be achieved, in general. Nonetheless, it can be mimicked up to a finite loss via entropy-coded dithered quantization [1], [5], [21]. Note, however, that this bound may become loose in the low-rate regime.

V. Two-Sided SI

We now treat the two-sided SI scenario, i.e., the scenario in which the SI is available to both the encoder and the decoder, corresponding to both switches A and B being closed in Fig. 1.

Definition 5 (Information CRDF with two-sided SI [22])**.** The information CRDF with two-sided SI of a Gaussian source $\{x_t\}$ with a jointly Gaussian SI $\{y_t\}$ that is known to both the encoder and the decoder is defined as

$$R_c^{\text{both}}(D) = \overline{\lim_{T\to\infty}} \inf_{\substack{P(\hat{x}^T \Uparrow x^T, y^T), \\ \frac{1}{T}\sum_t \mathbb{E}\left[\|x_t - \hat{x}_t\|^2\right] \le D}} \frac{1}{T} I\left(x^T \to \hat{x}^T \Uparrow y^T\right) \tag{10a}$$

$$= \frac{1}{2}\log^+ \frac{\sigma_n^2 \| (\lambda^2 D + \sigma_v^2)}{D}. \tag{10b}$$

Theorem 4. *The operational CRDF with two-sided SI, $R_{c,op}^{\text{both}}(D)$, is bounded from below by the information CRDF with two-sided SI* (10)*: $R_c^{\text{both}}(D) \le R_{c,op}^{\text{both}}(D)$.*

The setting with two-sided SI is equivalent to the no SI setting, w.r.t. to a (Gaussian) source that is equal to x_t given y^t. This simple observation allows a simple adaptation of the proof without SI to that of Th. 4.

Proof: By looking at the equivalent source $x_t | y^t$, the problem is equivalent to the no SI setting (9a), with the variance of the prediction error of x_t given $\hat{x}^{t-1}, y^t$ being

$$\sigma^2_{x_t|y^t, \hat{x}^{t-1}} = \sigma^2_{y_t|x_t} \middle\| \sigma^2_{x_t|y^{t-1}, \hat{x}^{t-1}} = \sigma_n^2 \| \left(\lambda^2 D_{t-1} + \sigma_v^2\right).$$

Plugging it in [5, Eq. (18)] gives rise to

$$R_c^{both}(D) = \frac{1}{T}\sum_{t=1}^{T} \frac{1}{2}\log\left(\sigma_n^2 \| \left(\lambda^2 D_{t-1} + \sigma_v^2\right)\right) - \frac{1}{2}\log D_t .$$

By applying Jensen's inequality and taking $T \to \infty$ (i.e., repeating steps (18d),(18e) of [5]) we arrive at the desired result:

$$R_c^{both}(D) = \frac{1}{2}\log \frac{\sigma_n^2 \| \left(\lambda^2 D + \sigma_v^2\right)}{D},$$

with D being the average-stage distortion (3). Using Th. 3 we conclude that $R_c^{both}(D) \le R_{c,op}^{\text{both}}(D)$. ■

VI. Causal Rate–Distortion With Decoder SI

In this section, we treat the more involved scenario where the SI is known only to the decoder while the encoder is oblivious of the SI, corresponding to switch A being open and B begin closed in Fig. 1.

We start by presenting a naïve lower bound.

Lemma 1. *The operational CRDF with decoder SI, $R_{c,op}^{dec}(D)$, is bounded from below by the information CRDF with two-sided SI* (10a)*: $R_c^{both}(D) \leq R_{c,op}^{dec}(D)$.*

Proof: Making the SI available (as a "genie") may only improve performance, and thus $R_{c,op}^{\mathrm{both}}(D) \leq R_{c,op}^{dec}(D)$. Using Th. 4, the result follows. ∎

Remark 4. Beyond the loss mentioned in Rems. 1 and 3 due to the causal encoding, the lower bound in Lem. 1 is known to be loose even for the batch memoryless RDF setting [28] due to the causal access to the SI at the decoder (see also [37]).

Definition 6 (Information CRDF with decoder SI)**.** The information CRDF with decoder SI of a Gaussian source $\{x_t\}$ with a jointly Gaussian SI $\{y_t\}$ that is known to the decoder is defined as

$$R_c^{\mathrm{dec}}(D) = \overline{\lim_{T\to\infty}} \inf_{\substack{P(w^T \Uparrow x^T), \{\hat{x}_t(w^t, y^t)\}: \\ \frac{1}{T}\sum_t \mathbb{E}\left[\|x_t - \hat{x}_t\|^2\right] \leq D, \\ (y^t, x^{t-1}) \to (x^t, w^{t-1}) \to w_t}} \frac{1}{T} I\left(x^T \to w^T\right). \quad (11)$$

Theorem 5. *The operational CRDF with decoder SI, $R_{c,op}^{dec}(D)$, is bounded from below by the information CRDF with decoder SI* (11)*: $R_c^{\mathrm{dec}}(D) \leq R_{c,op}^{dec}(D)$.*

Proof: We assume that the average distortion is equal to (or lower than) D and bound the average rate R (recall Def. 1):

$$\begin{aligned}
TR &\geq H\left(a^T\right) && \text{(12a)}\\
&= \sum_{t=1}^{T} H\left(a_t \middle| a^{t-1}\right) && \text{(12b)}\\
&\geq \sum_{t=1}^{T} H\left(a_t \middle| a^{t-1}, w^{t-1}\right) && \text{(12c)}\\
&\geq \sum_{t=1}^{T} H\left(a_t \middle| a^{t-1}, w^{t-1}\right) - H\left(a_t \middle| a^{t-1}, w^{t-1}, x^t\right) && \text{(12d)}\\
&= \sum_{t=1}^{T} I\left(x^t; a_t \middle| a^{t-1}, w^{t-1}\right) && \text{(12e)}\\
&= I\left(x^T \to a^T \Uparrow w^{T-1}\right) && \text{(12f)}\\
&\geq I\left(x^T \to w^T\right), && \text{(12g)}\\
&\geq T R_c^{\mathrm{dec}}(D), && \text{(12h)}
\end{aligned}$$

where (12a) follows from the problem statement, (12b) is due to the chain rule for entropies, (12c) holds since conditioning does not increase entropy, (12d) follows from the non-negativity of entropy, (12e) and (12f) are by the definition of the conditional MI and lagged-by-one DI (7), respectively, (12g) follows from the DPI for DIs of Th. 2 for x^t, a^t, w^t satisfying the Markov relations

$$\left(x_t, a^{t-1}\right) \to \left(a^t, w^{t-1}\right) \to w_t \quad (13)$$

for all $t \in \{1, 2, \ldots, T\}$ ($a_t \triangleq 0$, $w_0 \triangleq 0$), and (12h) follows from (11) for x^t, w^t satisfying the distortion and Markov constraints in (11).[5] ∎

[5]If w_t satisfies (13) it also satisfies the Markov constraint in (11).

Remark 5 (SI causality). Kostina and Hassibi [22, Def. 3] defined the (information) CRDF with decoder SI as

$$R_c^{\mathrm{KH}}(D) \triangleq \overline{\lim_{T\to\infty}} \inf_{\substack{P(w^T \Uparrow x^T), \{\hat{x}_t(w^t, y^t)\}: \\ \frac{1}{T}\sum_t \mathbb{E}\left[\|x_t - \hat{x}_t\|^2\right] \leq D \\ (y^t, x^{t-1}) \to (x^t, w^{t-1}) \to w_t}} \frac{1}{T} I\left(x^T \to w^T \Uparrow y^T\right) = \quad (14a)$$

$$\overline{\lim_{T\to\infty}} \inf_{\substack{P(w^T \Uparrow x^T), \{\hat{x}_t(w^t, y^t)\}: \\ \frac{1}{T}\sum_t \mathbb{E}\left[\|x_t - \hat{x}_t\|^2\right] \leq D \\ (y^t, x^{t-1}) \to (x^t, w^{t-1}) \to w_t}} \frac{1}{T} \left\{ I\left(x^T \to w^T\right) - I\left(y^T \to w^T\right) \right\}. \quad (14b)$$

and prove that $R_c^{\mathrm{KH}}(D) = R_c^{\mathrm{both}}(D)$ in the Gaussian case [22, Thm. 8].

This definition can be viewed as an adaptation of the batch RDF with decoder *non-causal* SI, $R^{\mathrm{NC}}(D)$. Indeed, as $R^{\mathrm{NC}}(D) = R^{\mathrm{both}}(D)$ in the Gaussian (batch) case, no improvement beyond the naïve bound of Lem. 1 is offered by (14) for bounding the CRDF with decoder SI.

Instead, we argue that better bounds result by relying on the technique of Weissman and El Gamal for batch RDF with decoder *causal* SI, $R^{\mathrm{C}}(D)$. By comparing (11) with (14b) the difference between the two bounds is $\frac{1}{T} I\left(y^T \to w^T\right) \geq 0$; as we shall claim in the sequel in Lem. 2, $I\left(y^T \to w^T\right) > 0$ in the Gaussian case, meaning that the bound offered by Th. 5 is strictly better than that of [22], [23].

Remark 6. We note that without the Markov chain constraint in (14a) we could choose w^t to be the $\hat{x}^t$ that minimize (10a). Thus, in general, the inequality $R_c^{\mathrm{KH}}(D) \geq R_c^{both}(D)$ holds.

Lemma 2. *$R_c^{dec}(D) > R_c^{\mathrm{KH}}(D)$ whenever $R_c^{dec}(D) > 0$, and $R_c^{dec}(D) = R_c^{\mathrm{KH}}(D) = 0$ whenever $R_c^{dec}(D) = 0$.*

Proof sketch: The statement for $R_c^{dec}(D) = 0$ trivially follows from the non-negativity of the MI (see also Rem. 5). Assume $R_c^{dec}(D) > 0$. Denote by w_*^T the w^T that achieves the infimum in (11). Consider the following two cases.

Case 1. w_*^T is jointly Gaussian with x^T (and y^T) under the limit superior in (11). Then, $\overline{\lim_{T\to\infty}} \frac{1}{T} I\left(y^T \to w^T\right) > 0$ in (14b) [5], [22], and hence $R_c^{dec}(D) > R_c^{\mathrm{KH}}(D)$.

Case 2. w_*^T is not jointly Gaussian with x^T and y^T under the limit superior in (11). Denote by w_G^T a jointly Gaussian vector with x^T and y^T that has the same joint second-order statistics with them as w_*^T. Then, we have

$$\begin{aligned}
I\left(x^T \to w_*^T\right) &\geq I\left(x^T \to w_*^T\right) - I\left(y^T \to w_*^T\right) && \text{(15a)}\\
&= I\left(x^T \to w_*^T \Uparrow y^T\right) && \text{(15b)}\\
&> I\left(x^T \to w_G^T \Uparrow y^T\right) && \text{(15c)}
\end{aligned}$$

where (15a) follows from the non-negativity of the DI, (15b) is according to (14b) and (15c) is from the uniqueness of the Gaussian solution of the problem (14) [22]. Evaluating (15) in $\overline{\lim_{T\to\infty}}$ yields the required result. ∎

Corollary 1. *The following relations hold when $R_c^{dec}(D) > 0$:*

$$R_{c,op}^{dec}(D) \overset{(a)}{\geq} R_c^{dec}(D) \overset{(b)}{>} R_c^{\mathrm{KH}}(D) \overset{(c)}{=} R_c^{\mathrm{both}}(D).$$

Proof: Steps (a), (b), and (c) follow from Th. 5, Lem. 2, and [22, Thm. 8] (see also Rem. 5), respectively. ■

Corollary 2. *The minimum distortion $D^{\mathrm{dec}}_{c,op}(D)$ of causal tracking of a Gauss–Markov source with causal SI over a memoryless channel with capacity C is bounded from below by*

$$D^{dec}_{c,op}(C) \geq \left(R^{dec}_c\right)^{-1}(C) > \left(R^{\mathrm{KH}}_c\right)^{-1}(C) = \left(R^{\mathrm{both}}_c\right)^{-1}(C)$$

Proof: The proof is a simple adaptation of [38, Thm. 2], [39, Thm. 1], which are in turn an adaptation of the necessity proof of the source–channel separation principle [40, Thm 3.7]; we outline it next. Denote the channel input and output at time t by a_t and b_t, respectively. Then, we have

$$TR^{dec}_c(D) \overset{(a)}{\leq} I\left(x^T \to b^T\right) \overset{(b)}{\leq} I\left(x^T; b^T\right) \overset{(c)}{\leq} TC$$

where (a) is due to Def. 6 and noting that b^T satisfies the conditions of w^T in (11), (b) holds since the DI is bounded from above by the MI, and (c) is due to [38, Eq. (31)]. The proof then follows from Corol. 1, by inverting the RDFs and invoking their monotonicity [40, Ch. 3]. ■

VII. Numerical Simulations

We have seen in Lem. 2 that $R^{dec}_c(D)$ gives a strictly tighter lower bound than that of $R^{both}_c(D)$ of Lem. 1 [and that of (14)] on the operational CRDF with decoder SI. Unfortunately, carrying out the optimization in (11) and finding an explicit solution is difficult and is yet to be determined even for the simpler memoryless batch, in which it reduces to the single-letter optimization problem in (5).

Following [28], we consider a Gaussian test channel—$w_t = x_t + z_t$, where z_t is a zero-mean AWGN of variance σ_z^2 in lieu of the infimum in (11) and evaluate the expression for this choice. We shall further show that Gaussian test channels are suboptimal meaning that Case 2 prevails in the proof of Lem. 2. We denote the minimum mean square errors (MMSEs) given w^t and given (y^t, w^t) by

$$D_t = \mathbb{E}\left[\left(x_t - \hat{x}_t(y^t, w^t)\right)^2\right], \quad \tilde{D}_t = \mathbb{E}\left[\left(x_t - \hat{x}_t(w^t)\right)^2\right].$$

First, note that R_1 equals the channel capacity of a power constrained AWGN channel [41]:

$$R_1 = I(x_1; w_1) = \frac{1}{2}\log\left(1 + \frac{\sigma_v^2}{\sigma_z^2}\right), \tag{16}$$

and $D_1 = \sigma_v^2 \| \sigma_n^2 \| \sigma_z^2$. By substituting it in (16), we arrive at

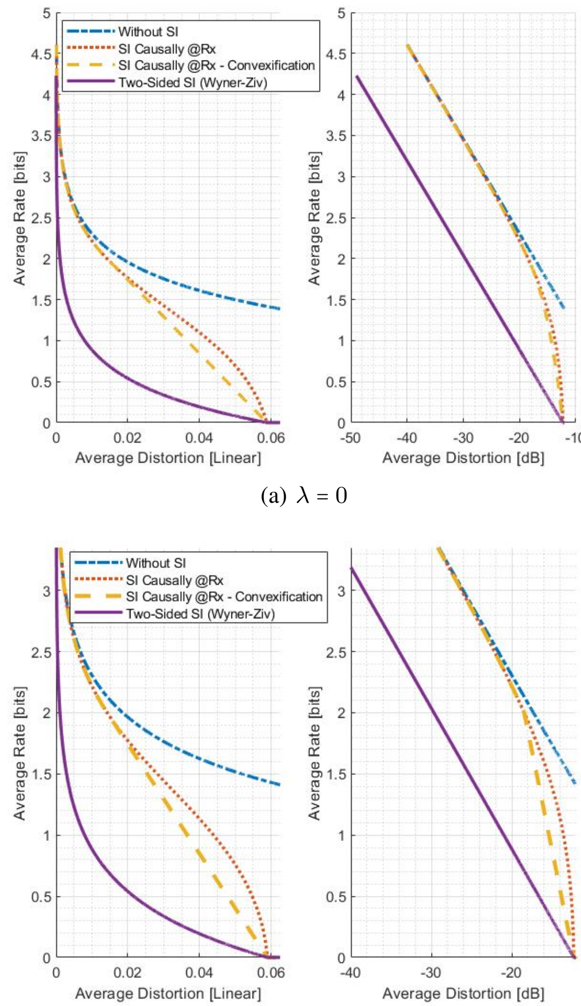
$$R_1 = \frac{1}{2}\log\left(\frac{\sigma_v^2}{D_1} - \frac{\sigma_v^2}{\sigma_n^2}\right). \tag{17}$$

Since rate–distortion curves must be convex and non-negative [41, Ch. 10], we clip R_1 of (17) at 0 and take its lower convex envelope to be the rate–distortion curve $R_1(D_1)$.

By putting forth the the process dynamics (1) and pedestrian MMSE estimation arguments we arrive at

$$D_{t+1} = \sigma_n^2 \| \sigma_z^2 \| \left(\lambda^2 D_t + \sigma_v^2\right), \quad \tilde{D}_{t+1} = \sigma_z^2 \| \left(\lambda^2 \tilde{D}_t + \sigma_v^2\right). \tag{18}$$

By defining $R_t(\tilde{D}_t) \triangleq \frac{1}{2}\log\left(\lambda^2 + \frac{\sigma_v^2}{\tilde{D}_t}\right)$ for $t > 1$, (17), and



(a) $\lambda = 0$

(b) $\lambda = 0.9$

Fig. 2: Information average rate versus the average distortion for no SI, two-sided SI, and causal decoder SI with a Gaussian test channel $w_t = x_t + n_t$ with $\sigma_n = 1/4$ for $\lambda = 0, 0.9$. We use a uniform distortion allocation $D_1 = \cdots = D_T = D$ in all the curves and $T = 2048$.

using [5, Proof of Corol. 2], [22, Thm. 2], we have ($\tilde{D}_0 = 0$)

$$I\left(x^T \to w^T\right) = \frac{1}{2}\log\left(\frac{\sigma_v^2}{\tilde{D}_1}\right) + \frac{1}{2}\sum_{t=2}^{T}\log\left(\lambda^2 + \frac{\sigma_v^2}{\tilde{D}_t}\right) \triangleq \sum_{t=1}^{T} R_t(\tilde{D}_t). \tag{19}$$

Using the definition of $R_t(\tilde{D}_t)$ and (18), we obtain

$$\tilde{D}_{t+1} = \sigma_z^2 \| \sigma_v^2 \left(1 - \lambda^2 2^{-2R_t}\right)^{-1}. \tag{20}$$

And by equating (20) with $\tilde{D}_{t+1}$ of the definition of $R_{t+1}(\tilde{D}_{t+1})$ we attain

$$\sigma_z^2 = \frac{\sigma_v^2}{2^{2R_{t+1}} - 1 - \lambda^2\left(1 - 2^{-2R_t}\right)}. \tag{21}$$

Substituting (21) into the recurssion of D_{t+1} (18) we arrive at the recursive description:

$$R_{t+1} = \frac{1}{2}\log\left(\frac{\sigma_v^2}{D_{t+1}} - \frac{\sigma_v^2}{\sigma_n^2} - \frac{\sigma_v^2}{\lambda^2 D_t + \sigma_v^2} + \lambda^2\left(1 - 2^{-2R_t}\right) + 1\right). \tag{22}$$

By substituting (22) and (17) into (19) we get an expression for the average rate.

The steady-state solution for (22) is given by

$$R = \frac{1}{2}\log^+\left(\lambda^2 + \frac{\sigma_v^2}{\tilde{D}}\right),$$

where $\tilde{D}$ is the positive solution of the quadratic equation

$$\lambda^2\tilde{D}^2 + \left[\sigma_v^2 + \left(1-\lambda^2\right)\sigma_z^2\right]\tilde{D} - \sigma_v^2\sigma_z^2 = 0,$$

whereas the distortion is given by the positive solution of the quadratic equation

$$\lambda^2 D^2 + \left[\sigma_v^2 + \left(1-\lambda^2\right)\left(\sigma_z^2 \| \sigma_n^2\right)\right] D - \sigma_v^2\left(\sigma_z^2 \| \sigma_n^2\right) = 0.$$

This curve is not convex meaning that the optimal test channel in (11) is not Gaussian. Consequently, by convexifying (corresponding to time-sharing with $R = 0$), we improve this curve.

We evaluate numerically the total rate using the recursion (22) for $\lambda = 0, 0.9$, $\sigma_n = 1/4, \sigma_v = 1$ and compare it to $R_c^{\mathrm{KH}}(D) \equiv R_c^{\mathrm{both}}(D)$ of (10b), (14) and $R_c(D)$ of (9b); clearly, the resulting curve lies between the two.

VIII. Future work

Interesting research directions, that are currently under investigation, are deriving an explicit expression for $R_c^{dec}(D)$ of (6) and determining the exact improvement compared to $R_c^{both}(D)$ of (10), and extending our results to vector systems.

References

[1] E. I. Silva, M. S. Derpich, and J. Østergaard, "A framework for control system design subject to average data-rate constraints," *IEEE Trans. Auto. Cont.*, vol. 56, no. 8, pp. 1886–1899, Aug. 2011.

[2] A. Sahai and S. K. Mitter, "The necessity and sufficiency of anytime capacity for stabilization of a linear system over a noisy communication link—part I: Scalar systems," *IEEE Trans. Inf. Theory*, vol. 52, no. 8, pp. 3369–3395, Aug. 2006.

[3] 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.

[4] S. Tatikonda, A. Sahai, and S. K. Mitter, "Stochastic linear control over a communication channel," *IEEE Trans. Auto. Cont.*, vol. 49, no. 9, pp. 1549–1561, Sep. 2004.

[5] A. Khina, V. Kostina, A. Khisti, and B. Hassibi, "Tracking and control of Gauss–Markov processes over packet-drop channels with acknowledgments," *IEEE Trans. Cont. Network Sys.*, vol. 6, pp. 549–560, 2019.

[6] P. A. Stavrou, J. Østergaard, C. D. Charalambous, and M. Derpich, "An upper bound to zero-delay rate distortion via Kalman filtering for vector Gaussian sources," in *Proc. IEEE Info. Theory Workshop (ITW)*, Kaohsiung, Taiwan, 2017, pp. 534–538.

[7] R. Bansal and T. Başar, "Simultaneous design of measurement and control strategies for stochastic systems with feedback," *Automatica*, vol. 25, no. 5, pp. 679–694, Sep. 1989.

[8] J. S. Freudenberg, R. H. Middleton, and V. Solo, "Stabilization and disturbance attenuation over a Gaussian communication channel," *IEEE Trans. Auto. Cont.*, vol. 55, no. 3, pp. 795–799, Mar. 2010.

[9] G. N. Nair and R. J. Evans, "Stabilizability of stochastic linear systems with finite feedback data rates," *SIAM J. Control and Opt.*, vol. 43, no. 2, pp. 413–436, 2004.

[10] A. S. Matveev and A. V. Savkin, "An analogue of shannon information theory for detection and stabilization via noisy discrete communication channels," *SIAM journal on Control and Optimization*, vol. 46, no. 4, pp. 1323–1367, Sep. 2007.

[11] A. Khina, Y. Nakahira, Y. Su, and B. Hassibi, "Algorithms for optimal control with fixed-rate feedback," in *Proc. IEEE Conference on Decision and Control (CDC)*, Melbourne, VIC, Australia, Dec. 2017, pp. 6015–6020.

[12] S. Yüksel, "Jointly optimal LQG quantization and control policies for multi-dimensional systems," *IEEE Trans. Auto. Cont.*, vol. 59, no. 6, pp. 1612–1617, Jun. 2014.

[13] S. Yüksel, "Stochastic stabilization of noisy linear systems with fixed-rate limited feedback," *IEEE Trans. Auto. Cont.*, vol. 55, no. 12, pp. 2847–2853, Dec. 2010.

[14] S. Yüksel and T. Başar, *Stochastic Networked Control Systems: Stabilization and Optimization Under Information Constraints*. Boston: Birkhäuser, 2013.

[15] V. Kostina and B. Hassibi, "Rate–cost tradeoffs in control," *IEEE Trans. Auto. Cont.*, vol. 64, no. 11, pp. 4525–4540, Nov. 2019.

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

[17] A. Khina, E. Riedel Gårding, G. M. Pettersson, V. Kostina, and B. Hassibi, "Control over Gaussian channels with and without source–channel separation," *IEEE Trans. Auto. Cont.*, vol. 64, no. 9, pp. 3690–3705, Sep. 2019.

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

[19] S. C. Tatikonda, "Control under communication constraints," Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, CA, USA, Sep. 2000.

[20] C. D. Charalambous, C. K. Kourtellaris, and C. Hadjicostis, "Optimal encoder and control strategies in stochastic control subject to rate constraints for channels with memory and feedback," in *Proc. IEEE Conf. Decision and Control and Euro. Control (CDC-ECC)*, Orlando, FL, USA, Dec. 2011, pp. 4522–4527.

[21] T. Tanaka, P. M. Esfahani, and S. K. Mitter, "LQG control with minimum directed information: Semidefinite programming approach," *IEEE Trans. Auto. Cont.*, vol. 63, no. 1, pp. 37–52, Jan. 2018.

[22] V. Kostina and B. Hassibi, "Rate-cost tradeoffs in scalar LQG control and tracking with side information," in *Proc. Allerton Conf. on Comm., Control, and Comput.*, Monticello, IL, USA, Oct. 2018, pp. 421–428.

[23] P. Stavrou and M. Skoglund, "The zero-delay RDF for Gaussian processes with encoder-decoder causal side information," Apr. 2019, qC 20190425. [Online]. Available: http://www.diva-portal.org/smash/get/diva2:1306744/FULLTEXT01.pdf

[24] A. D. Wyner, "The Rate–Distortion function for source coding with side information at the decoder—II: General sources," *Information and Control*, vol. 38, pp. 60–80, 1978.

[25] A. D. Wyner and J. Ziv, "The Rate–Distortion function for source coding with side information at the decoder," *IEEE Trans. Inf. Theory*, vol. 22, no. 1, pp. 1–10, Jan. 1976.

[26] T. Linder and R. Zamir, "Causal coding of stationary sources and individual sequences with high resolution." *IEEE Trans. Inf. Theory*, vol. 52, pp. 662–680, Feb. 2006.

[27] D. Neuhoff and R. Gilbert, "Causal source codes," *IEEE Trans. Inf. Theory*, vol. 28, no. 5, pp. 701–713, Sep. 1982.

[28] T. Weissman and A. El Gamal, "Source coding with limited-look-ahead side information at the decoder," *IEEE Trans. Inf. Theory*, vol. 52, pp. 5218–5239, Dec. 2006.

[29] C. E. Shannon, "Coding theorems for a discrete source with a fidelity criterion," in *Institute of Radio Engineers, International Convention Record*, vol. 7, 1959, pp. 142–163.

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

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

[32] R. Zamir, Y. Kochman, and U. Erez, "Achieving the Gaussian Rate-Distortion function by prediction," *IEEE Trans. Inf. Theory*, vol. 54, pp. 3354–3364, July 2008.

[33] Y. Kochman and R. Zamir, "Analog matching of colored sources to colored channels," *IEEE Trans. Inf. Theory*, vol. 57, no. 6, pp. 3180–3195, June 2011.

[34] J. Massey, "Causality, feedback and directed information," in *Proc. IEEE Int. Symp. Info. Theory and Its Apps. (ISITA)*, Waikiki, HI, USA, Nov. 1990, pp. 303–305.

[35] G. Kramer, *Directed information for channels with feedback*. Lausanne, Switzerland: Hartung-Gorre, Jan. 1998.

[36] 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.

[37] X. Chen and E. Tuncel, "Zero-delay joint source–channel coding using hybrid digital–analog schemes in the Wyner–Ziv setting," *IEEE Trans. Comm.*, vol. 62, no. 2, pp. 726–735, Feb. 2014.

[38] 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.

[39] N. Merhav and S. Shamai, "On joint source–channel coding for the Wyner–Ziv source and the Gel'fand–Pinsker channel," *IEEE Trans. Inf. Theory*, vol. 49, pp. 2844–2855, Nov. 2003.

[40] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge University Press, 2011.

[41] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.