Fundamental limits of distributed tracking

Victoria Kostina, Babak Hassibi

***Abstract*—Consider the following communication scenario. An n-dimensional source with memory is observed by K isolated encoders via parallel channels, who causally compress their observations to transmit to the decoder via noiseless rate-constrained links. At each time instant, the decoder receives K new codewords from the observers, combines them with the past received codewords, and produces a minimum-distortion estimate of the latest block of n source symbols. This scenario extends the classical one-shot CEO problem to multiple rounds of communication with communicators maintaining memory of the past.**

We prove a coding theorem showing that the minimum asymptotically (as $n \to \infty$) achievable sum rate required to achieve a target distortion is equal to the directed mutual information from the observers to the decoder minimized subject to the distortion constraint and the separate encoding constraint. For the Gauss-Markov source observed via K parallel AWGN channels, we solve that minimal directed mutual information problem, thereby establishing the minimum asymptotically achievable sum rate. Finally, we explicitly bound the rate loss due to a lack of communication among the observers; that bound is attained with equality in the case of identical observation channels.

The general coding theorem is proved via a new nonasymptotic bound that uses stochastic likelihood coders and whose asymptotic analysis yields an extension of the Berger-Tung inner bound to the causal setting. The analysis of the Gaussian case is facilitated by reversing the channels of the observers.

***Index Terms*—CEO problem, Berger-Tung bound, distributed source coding, causal rate-distortion theory, Gauss-Markov source, LQG control.**

I. Introduction

We set up the causal CEO (chief executive or estimation officer) problem as follows. An information source $\{X_i\}$ outputs $X_i \in \mathcal{A}^n$ at time i; it is observed by K encoders through K noisy channels; at time i, kth encoder sees Y_i^k generated according to $P_{Y_i^k|X_1,\ldots,X_i,Y_1^k,\ldots,Y_{i-1}^k}$. See Fig. 1. The encoders (observers) communicate to the decoder (CEO) via their separate noiseless rate-constrained links. At each time i, kth observer forms a codeword based on the observations it has seen so far, i.e., $Y_1^k, \ldots, Y_i^k$. The decoder at time i chooses $\hat{X}_i \in \hat{\mathcal{A}}^n$ based on the codewords it received thus far. The goal is to minimize the average distortion

$$\frac{1}{t}\sum_{i=1}^{t} \mathbb{E}\left[\mathsf{d}(X_i, \hat{X}_i)\right], \tag{1}$$

where t is the *time horizon* over which the source is being tracked, and $\mathsf{d}\colon \mathcal{A}^n \times \hat{\mathcal{A}}^n \mapsto \mathbb{R}_+$ is the distortion measure. Encoding and decoding operations leverage memory of the past but cannot look in the future. In this causal setting no delay is allowed in producing $\hat{X}_i$.

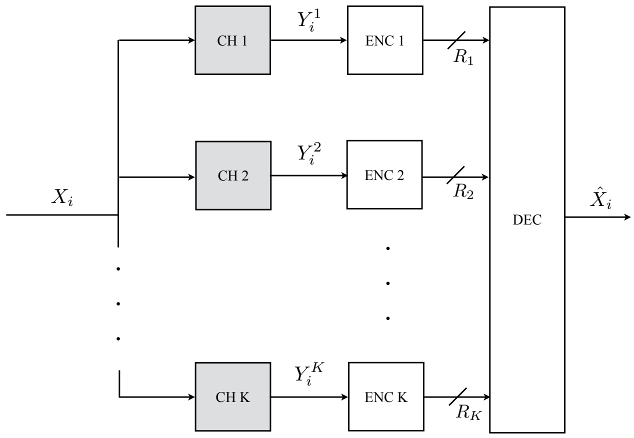

Fig. 1. The causal CEO problem.

In the classical setting with $t = 1$, the CEO problem was first introduced by Berger et al. [1] for a finite alphabet source. In the classical Gaussian CEO problem, a Gaussian source is observed via Gaussian channels and reproduced under mean-square error (MSE) distortion. The Gaussian CEO problem was studied by Viswanathan and Berger [2], who proved an achievability bound on the rate-distortion dimension for the case of K identical Gaussian channels, by Oohama [3], who derived the sum-rate rate-distortion region for that special case, by Prabharan et al. [4], who determined the full Gaussian CEO rate region, by Chen et al. [5], who proved that the minimum sum rate is achieved via waterfilling, by Behroozi and Soleymani [6] and by Chen and Berger [7], who showed rate-optimal successive coding schemes. Wagner et al. [8] found the rate region of the distributed Gaussian lossy compression problem by coupling it to the Gaussian CEO problem. Wagner and Anantharam [9] showed an outer bound to the rate region of the multiterminal source coding problem that is tighter than the Berger-Tung outer bound [10], [11]. Wang et al. [12] showed a simple converse on the sum rate of the vector Gaussian CEO problem. Concurrently, Ekrem and Ulukus [13] and Wang and Chen [14] showed an outer bound to the rate region of the vector Gaussian CEO problem that is tight in some cases and not tight in others and that particularizes the outer bound in [9] to the Gaussian case. Courtade and Weissman [15] determined the distortion region of the distributed source coding and the CEO problem under logarithmic loss.

None of the above results directly apply to the causal tracking problem in Fig. 1 because of the causality constraint in encoding the observations and in producing $\hat{X}_i$ in (1). The most basic scenario of source coding with causality

The authors are with California Institute of Technology (e-mail: vkostina@caltech.edu, hassibi@caltech.edu). This work was supported in part by the National Science Foundation (NSF) under grants CCF-1751356 and CCF-1817241. The work of Babak Hassibi was supported in part by the NSF 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, and by King Abdullah University of Science and Technology.

constraints is that of a single observer directly seeing the information source [16]. The causal rate-distortion function for the Gauss-Markov source was computed by Gorbunov and Pinsker [17]. The link between the minimum attainable linear quadratic Gaussian (LQG) control cost and the causal rate-distortion function is elucidated in [18]–[20]. A semidefinite program to compute the causal rate-distortion function for vector Gauss-Markov sources is provided in [21]. The remote Gaussian causal rate-distortion function, which corresponds to setting $K = 1$ in Fig. 1, is computed in [20]. The causal rate-distortion function of the Gauss-Markov source with Gaussian side observation available at decoder (the causal counterpart of the Wyner-Ziv setting) is computed in [22]. That causal Wyner-Ziv setting can be viewed a special case of our causal CEO problem (2), (3) with two observers, with the second observer enjoying an infinite rate. Stability of linear Gaussian systems with multiple isolated observers is investigated in [23].

The first contribution of this paper is a characterization the minimum asymptotically achievable (as $n \to \infty$) sum rate $R_1 + \ldots + R_K$ required to achieve a given average distortion (1) in the causal distributed tracking setting of Fig. 1. Provided that the components of each $X_i \in \mathcal{A}^n$ are i.i.d. (X_i can still depend on $X_1, \ldots, X_{i-1}$), the channels act on each of those components independently, and the distortion measure is separable, that minimum sum rate is equal to the directed mutual information from the observers to the decoder minimized subject to the distortion constraint and the separate encoding constraint.

The second contribution of the paper is an evaluation of the minimum sum rate for the causal Gaussian CEO problem. In that scenario, the source is an n-dimensional Gauss-Markov source,

$$X_{i+1} = aX_i + V_i, \tag{2}$$

k-th observer sees

$$Y_i^k = X_i + W_i^k, \quad k = 1, \ldots, K, \tag{3}$$

where X_1 and $\{V_i, W_i^1, W_i^2, \ldots, W_i^K\}_{i=1}^T$ are independent Gaussian vectors of length n; $V_i \sim \mathcal{N}(0, \sigma_{\mathsf{V}}^2 \mathsf{I})$; $W_i^k \sim \mathcal{N}(0, \sigma_{\mathsf{W}_k}^2 \mathsf{I})$. The distortion measure is mean-square error (MSE)

$$\mathsf{d}\left(X_i, \hat{X}_i\right) = \|X_i - \hat{X}_i\|^2. \tag{4}$$

We characterize the minimum sum rate as a convex optimization problem over K parameters; an explicit formula is given in the case of identical observation channels.

The third contribution of the paper is a bound on the rate loss due to a lack of communication among the different encoders in the causal Gaussian CEO problem: as long as the target distortion is not too small, the rate loss is bounded above by $K - 1$ times the difference between the remote and the direct rate-distortion functions. The bound is attained with equality if the observation channels are identical, indicating that among all possible observer channels with the same error in estimation $\{X_i\}$ from $\{Y_j^k\}_{j \leq i, k=1,\ldots,K}$, the identical channels case is the hardest to compress.

The rest of the paper is organized as follows. In Section II, we consider the general (non-Gaussian) causal CEO problem and present our main coding theorem establishing that the minimum sum rate is given by the directed mutual information minimized subject to a distortion constraint under separate encoding (Theorem 1). In Section III, we characterize the causal Gaussian CEO sum rate - distortion function (Theorem 2). In Section IV, we bound the rate loss due to isolated observers (Theorem 3). Most proofs are relegated to the extended version [24].

Notation: Logarithms are natural base. For a natural number M, $[M] \triangleq \{1, \ldots, M\}$. Notation $X \leftarrow Y$ reads "replace X by Y". We indicate the temporal index in the subscript and the spatial index in the superscript: $Y_{[t]}^k$ is the temporal vector $(Y_1^k, \ldots, Y_t^k)$; $Y_i^{[K]}$ is the spatial vector $(Y_i^1, \ldots, Y_i^K)^\mathsf{T}$; $Y_{[t]}^{[K]} \triangleq (Y_{[t]}^1, \ldots, Y_{[t]}^K)$. We use the following shorthand notation for causally conditional [25] probability kernels:

$$P_{Y_{[t]} \| X_{[t]}} \triangleq \prod_{i=1}^t P_{Y_i | Y_{[i-1]}, X_{[i]}}. \tag{5}$$

For a random vector X with i.i.d. components, X denotes a random variable distributed the same as each component of X.

II. Sum rate distortion function via directed information

A. Operational problem setting

A causal CEO code is formally defined as follows.

Definition 1 (Causal CEO code)**.** *Consider a discrete-time random process $\{X_i\}_{i=1}^t$ on $\mathcal{X}$, observed by K causal observers via the channels*

$$P_{Y_{[t]}^k \| X_{[t]}} \colon \mathcal{X}^{\otimes t} \mapsto \mathcal{Y}^{\otimes t}, \quad k \in [K]. \tag{6}$$

Let $\mathsf{d} \colon \mathcal{X} \times \hat{\mathcal{X}} \mapsto \mathbb{R}_+$ be the distortion measure.

A causal CEO code consists of:

a) K causal encoding policies

$$P_{B_{[t]}^k \| Y_{[t]}^k} \colon \mathcal{Y}^{\otimes t} \mapsto \prod_{i=1}^t [M_i^k], \quad k \in [K], \tag{7}$$

b) a decoding policy

$$P_{\hat{X}_{[t]}^{[K]} \| B_{[t]}^{[K]}} \colon \prod_{i=1}^t [M_i^k] \mapsto \hat{\mathcal{X}}^{\otimes t}. \tag{8}$$

If the encoding and decoding policies satisfy

$$\frac{1}{t} \sum_{i=1}^t \mathbb{E}\left[\mathsf{d}\left(X_i, \hat{X}_i\right)\right] \leq d, \tag{9}$$

we say that they form an $(M_{[t]}^{[K]}, d)$ average distortion *code.*

If the encoding and decoding policies satisfy

$$\mathbb{P}\left[\bigcup_{i=1}^t \left\{\mathsf{d}\left(X_i, \hat{X}_i\right) > d_i\right\}\right] \leq \epsilon, \tag{10}$$

we say that they form an $(M_{[t]}^{[K]}, d_{[t]}, \epsilon)$ excess distortion *code.*

The probability measure in (9) and (10) is generated by the joint distribution $P_{X_{[t]}} P_{Y_{[t]}^{[K]} \| X_{[t]}} P_{\hat{X}_{[t]}^{[K]} \| B_{[t]}^{[K]}} \prod_{k=1}^K P_{B_{[t]}^k \| Y_{[t]}^k}$.

A distortion measure $\mathsf{d}_n\colon \mathcal{A}^n \times \hat{\mathcal{A}}^n \mapsto \mathbb{R}_+$ is called *separable* if

$$\mathsf{d}_n(x, \hat{x}) = \frac{1}{n} \sum_{i=1}^{n} \mathsf{d}(x(i), \hat{x}(i)), \tag{11}$$

where $\mathsf{d}\colon \mathcal{A} \times \hat{\mathcal{A}} \mapsto \mathbb{R}_+$, and $x(i)$, $\hat{x}(i)$ denote the i-th components of vectors $x \in \mathcal{A}^n$ and $\hat{x} \in \hat{\mathcal{A}}^n$, respectively.

Definition 2 (Operational sum rate - distortion function)**.** *Consider a discrete-time random process $\{X_i\}_{i=1}^t$ on $\mathcal{X} = \mathcal{A}^n$ equipped with a separable distortion measure, observed by K causal observers via the channels (6).*

The rate-distortion tuple $\left(R^{[K]}, d\right)$ is asymptotically achievable *at time horizon t if for $\forall \gamma > 0$, $\exists n_0 \in \mathbb{N}$ such that $\forall n \geq n_0$, an $\left(M_{[t]}^{[K]}, d + \gamma\right)$ average distortion causal CEO code exists, where*

$$\frac{1}{nt} \sum_{i=1}^{t} \log M_i^k \leq R^k, \quad k \in [K]. \tag{12}$$

The sum rate - distortion pair (R, d) is asymptotically achievable if a rate-distortion tuple $\left(R^{[K]}, d\right)$ with

$$\sum_{k=1}^{K} R^k \leq R \tag{13}$$

is asymptotically achievable.

The causal sum rate - distortion function at time horizon t is defined as follows:

$$R_{t\,\mathrm{CEO}}(d) \triangleq \inf \Big\{ R\colon\ (R, d) \text{ is achievable at time horizon } t \text{ in the CEO problem.} \Big\} \tag{14}$$

B. Main coding theorem

Given a distribution $P_{X_{[t]}}$ and a causal kernel $P_{Y_{[t]} \| X_{[t]}}$, the directed mutual information is defined as [26]

$$I\left(X_{[t]} \to Y_{[t]}\right) \triangleq \sum_{i=1}^{t} I\left(X_{[i]}; Y_i | Y_{[i-1]}\right). \tag{15}$$

Theorem 1 (Main coding theorem)**.** *Consider a discrete-time random process $\{X_i\}_{i=1}^t$ on $\mathcal{X} = \mathcal{A}^n$ equipped with a separable distortion measure, observed by K causal observers via the channels (6) with $\mathcal{Y} = \mathcal{B}^n$ and*

$$P_{X_i | X_{[i-1]}} = P_{\mathsf{X}_i | \mathsf{X}_{[i-1]}}^{\otimes n} \tag{16}$$

$$P_{Y_i^k | X_{[i]}, Y_{[i-1]}^k} = P_{\mathsf{Y}_i^k | \mathsf{X}_{[i]}, \mathsf{Y}_{[i-1]}^k}^{\otimes n}. \tag{17}$$

Suppose further that for some $p > 1$, there exists a vector $\hat{\mathsf{x}}_{[t]}$ such that

$$\mathbb{E}\left[\left(\frac{1}{t} \sum_{i=1}^{t} \mathsf{d}(\mathsf{X}_i, \hat{\mathsf{x}}_i)\right)^p\right]^{\frac{1}{p}} \leq d_p < \infty. \tag{18}$$

The causal sum rate - distortion function is given by

$$R_{t\,\mathrm{CEO}}(d) = \inf_{\substack{P_{\mathsf{U}_{[t]}^{[K]} \| \mathsf{Y}_{[t]}^{[K]}} = \prod_{k=1}^K P_{\mathsf{U}_{[t]}^k \| \mathsf{Y}_{[t]}^k}, \\ P_{\hat{\mathsf{X}}_{[t]} \| \mathsf{U}_{[t]}^{[K]}}\colon \\ \frac{1}{t}\sum_{i=1}^t \mathbb{E}\left[\mathsf{d}\left(\mathsf{X}_i, \hat{\mathsf{X}}_i\right)\right] \leq d}} I\left(\mathsf{Y}_{[t]}^{[K]} \to \mathsf{U}_{[t]}^{[K]}\right). \tag{19}$$

Proof. The converse follows via standard data processing and single-letterization arguments. To prove the achievability, we show a nonasymptotic bound for causal distributed lossy source coding that can be viewed as an extension of the nonasymptotic Berger-Tung bound by Yassaee et al. [27], [28] to the setting with $K > 2$ sources and $t > 1$ time instances. We view the horizon-t causal coding problem as a multiterminal coding problem in which at each step coded side information from past steps is available, and we use a stochastic likelihood coder (SLC) by Yassaee et al. [27], [28] to perform encoding operations. The SLC-based encoder mimics the operation of the joint typicality encoder while admitting sharp nonasymptotic bounds on its performance. Unfortunately, the SLC-based decoder of [27], [28] is ill-suited to the case $K > 2$. We propose a novel decoder that falls into the class of generalized likelihood decoders [29] and uses K different threshold tests depending on the point of the rate-distortion region the code is operating at. An asymptotic analysis of our nonasymptotic bound yields an extension of the Berger-Tung bound [10], [11] to the causal coding setting. See [24] for details. □

Theorem 1 establishes the operational meaning of the minimal directed mutual information in (19). Note that $R_{t\,\mathrm{CEO}}(d)$ is a convex function of d. (Convexity can be proven in the standard way, using time sharing between kernels achieving different d_j's and the convexity of mutual information in those kernels.)

III. Gaussian sum rate - distortion function

A. Problem setup

In this section, we focus on the scenario of the Gauss-Markov source in (2) observed through the Gaussian channels in (3) under MSE distortion (4). Given an encoding policy in Definition 1, the optimal decoding policy $P_{\hat{X}_{[t]}^{[K]} \| B_{[t]}^{[K]}}$ that achieves the minimum of $\mathbb{E}\left[\|X_i - \hat{X}_i\|^2\right]$ is $\hat{X}_i = \mathbb{E}\left[X_i | B_{[i]}^{[K]}\right]$.

For simplicity we focus on the infinite time-horizon limit.

$$R_{\mathrm{CEO}}(d) \triangleq \limsup_{t \to \infty} R_{t\,\mathrm{CEO}}(d). \tag{20}$$

In other words, $R_{\mathrm{CEO}}(d)$ is the infimum of R's such that $\forall \gamma > 0$, $\exists t_0 \geq 0$ such that $\forall t \geq t_0$, $\exists n_0 \in \mathbb{N}$ such that $\forall n \geq n_0$, an $\left(M_{[t]}^{[K]}, d + \gamma\right)$ average distortion causal CEO code exists with $M_{[t]}^{[K]}$ satisfying (12) and (13).

Taking the limit $t \to \infty$ simplifies the solution of many minimal directed mutual information problems ([20, Th. 9], [22, Th. 6, Th. 7], [30, Th. 1], [31, Th. 2]) by eliminating the transient effects due to the starting location X_1 of the process $\{X_i\}$ that is being transmitted. In this steady state regime, the optimal rate allocation across time is uniform (i.e., $\log M_1^k = \ldots = \log M_t^k$ in (12)). Furthermore, $R_{t\,\mathrm{CEO}}(d)$ approaches its steady-state value (20) as $O\left(\frac{1}{t}\right)$.

Notation: For a random process $\{\mathsf{X}_i\}$ on $\mathbb{R}$, its stationary variance (can be $+\infty$) is denoted by

$$\sigma_{\mathsf{X}}^2 \triangleq \limsup_{i \to \infty} \frac{1}{n} \mathbb{E}\left[\mathsf{X}_i^2\right]. \tag{21}$$

The minimum MSE (MMSE) in the estimation of X_i from $\mathsf{Y}_{[i]}^{[K]}$ is denoted by

$$\sigma^2_{\mathsf{X}_i|\mathsf{Y}_{[i]}^{[K]}} \triangleq \mathbb{E}\left[\left(\mathsf{X} - \mathbb{E}\left[\mathsf{X}_i|\mathsf{Y}_{[i]}^{[K]}\right]\right)^2\right], \tag{22}$$

and the steady-state causal MMSE by

$$\sigma^2_{\mathsf{X}\|\mathsf{Y}^{[K]}} \triangleq \limsup_{i\to\infty} \sigma^2_{\mathsf{X}_i|\mathsf{Y}_{[i]}^{[K]}}. \tag{23}$$

B. Gaussian sum rate - distortion function

In Theorem 2, the causal sum rate - distortion function is expressed as a convex optimization problem over parameters $\{d_k\}_{k=1}^K$ that determine the individual rates of the transmitters and that correspond to the MSE achievable at the decoder in estimation of $\{X_i\}_{i=1}^t$ provided that it correctly decoded the codewords from k-th transmitter.

Theorem 2. *For all $\sigma^2_{\mathsf{X}\|\mathsf{Y}^{[K]}} < d < \sigma^2_{\mathsf{X}}$, the causal sum rate - distortion function (20) for the Gauss-Markov source in (2) observed through the Gaussian channels in (3) is given by*

$$R_{\text{CEO}}(d) = \frac{1}{2}\log\frac{\bar{d}}{d} + \min_{\{d_k\}_{k=1}^K} \sum_{k=1}^K \frac{1}{2}\log\frac{\bar{d}_k - \sigma^2_{\mathsf{X}\|\mathsf{Y}^k}}{d_k - \sigma^2_{\mathsf{X}\|\mathsf{Y}^k}}\frac{d_k}{\bar{d}_k}, \tag{24}$$

where

$$\bar{d} \triangleq a^2 d + \sigma_V^2, \tag{25}$$

$$\bar{d}_k \triangleq a^2 d_k + \sigma_V^2, \tag{26}$$

and the minimum is over d_k, $k \in [K]$, that satisfy

$$\frac{1}{d} \le \frac{1}{\sigma^2_{\mathsf{X}\|\mathsf{Y}^{[K]}}} - \sum_{k=1}^K \left(\frac{1}{\sigma^2_{\mathsf{X}\|\mathsf{Y}^k}} - \frac{1}{d_k}\right), \tag{27}$$

$$\sigma^2_{\mathsf{X}\|\mathsf{Y}^k} \le d_k \le \sigma^2_{\mathsf{X}}, \tag{28}$$

Proof. We break up the minimal directed mutual information problem in Theorem 1 into subproblems, and we use the tools we developed in [22] to evaluate the causal rate-distortion functions for each subproblem. As it turns out, the additive white Gaussian kernels

$$\mathsf{U}_i^{k\star} = \bar{\mathsf{X}}_i^k + \mathsf{Z}_i^k, \quad \mathsf{Z}_i \sim \mathcal{N}(0, \sigma^2_{\mathsf{Z}^k}), \tag{29}$$

where

$$\bar{\mathsf{X}}_i^k \triangleq \mathbb{E}\left[\mathsf{X}_i|\mathsf{Y}_{[i]}^k\right] \tag{30}$$

is the MMSE estimate of X_i given $\mathsf{Y}_{[i]}^k$, attain the minimal directed mutual information in Theorem 1. To link the parameters of the subproblems together to obtain the solution of the original problem, we extend the proof technique by Wang et al. [12], developed for the case $t = 1$, to $t > 1$. Converting the "forward channels" from $\mathsf{X}_{[t]}$ to observations $\mathsf{Y}_{[t]}^k$ into the "backward channels" from MMSE estimates $\bar{\mathsf{X}}_{[t]}^k$ to $\mathsf{X}_{[t]}$ is key to that extension. See [24] for details. □

If the source is observed directly by one or more of the encoders, say if $\sigma^2_{\mathsf{X}\|\mathsf{Y}^1} = 0$, then $d_1 = d$, $d_2 = \ldots = d_K = \sigma^2_{\mathsf{X}}$ is optimal, and (24) reduces to the causal rate-distortion function [17, eq. (1.43)] (and e.g. [18], [32, Th. 3], [20, (64)]), [22, Th. 6]):

$$R(d) = \frac{1}{2}\log\frac{\bar{d}}{d}. \tag{31}$$

The sum over $k \in [K]$ in (24) is thus the penalty due to the encoders not observing the source directly and not communicating with each other.

If the observation channels satisfy

$$\sigma^2_{\mathsf{X}\|\mathsf{Y}^1} = \ldots = \sigma^2_{\mathsf{X}\|\mathsf{Y}^K}, \tag{32}$$

we can explicitly write the sum rate - distortion function $R_{\text{CEO}}^{K-\text{sym}}(d)$ for this symmetrical scenario.

Corollary 1. *If in the scenario of Theorem 2 the observation channels satisfy (32), the causal sum rate - distortion function (20) is given by*

$$R_{\text{CEO}}^{K-sym}(d) = \frac{1}{2}\log\frac{\bar{d}}{d} + \frac{K}{2}\log\frac{\bar{d}_1 - \sigma^2_{\mathsf{X}\|\mathsf{Y}^1}}{d_1 - \sigma^2_{\mathsf{X}\|\mathsf{Y}^1}}\frac{d_1}{\bar{d}_1}, \tag{33}$$

where d_1 satisfies

$$\frac{1}{d} = \frac{1}{\sigma^2_{\mathsf{X}\|\mathsf{Y}^{[K]}}} - \frac{K}{\sigma^2_{\mathsf{X}\|\mathsf{Y}^1}} + \frac{K}{d_1}. \tag{34}$$

Proof. It suffices to show that the minimum in (24) is attained by $d_1 = \ldots = d_K$. Since each of the terms in the sum in (24) is a convex function of d_k, applying Jensen's inequality concludes the proof. □

Think now of adding identical observers by letting $K \to \infty$ in (32). Since $\sigma^2_{\mathsf{X}\|\mathsf{Y}^{[K]}} \to 0$, had the observers communicated with each other, they could recover the source exactly, and they could operate at sum rate (31) in the limit. As the following result demonstrates, $\lim_{K\to\infty} R_{\text{CEO}}^{K-\text{sym}}(d)$ is actually strictly greater than (31), thus a nonvanishing penalty due to separate encoding is present in this regime. See Section IV for a more thorough discussion on the loss due to separate encoding.

Corollary 2.

$$\lim_{K\to\infty} R_{\text{CEO}}^{K-\text{sym}}(d) = \frac{1}{2}\log\frac{\bar{d}}{d} + \frac{1}{2}\frac{\frac{1}{d} - \frac{1}{\bar{d}}}{\frac{1}{\sigma^2_{\mathsf{X}\|\mathsf{Y}^1}} - \frac{1}{\sigma^2_{\mathsf{X}}}}. \tag{35}$$

Proof. [24]. □

Corollary 2 extends the result of Oohama [3, Cor. 1] to causal compression, and recovers it if $a = 0$.

Considering a scenario where the encoders and the decoder do not keep any memory of past observations and codewords, we may invoke the results on the classical Gaussian CEO problem in [4], [5] to express the minimum achievable sum rate as

$$R_{\text{CEO}}^{\text{no memory}}(d) = \frac{1}{2}\log\frac{\sigma^2_{\mathsf{X}}}{d} + \min_{\{d_k\}_{k=1}^K} \sum_{k=1}^K \frac{1}{2}\log\frac{\sigma^2_{\mathsf{X}} - \sigma^2_{\mathsf{X}|\mathsf{Y}^k}}{d_k - \sigma^2_{\mathsf{X}|\mathsf{Y}^k}}\frac{d_k}{\sigma^2_{\mathsf{X}}}, \tag{36}$$

where the minimum is over

$$\frac{1}{d} \le \frac{1}{\sigma^2_{\mathsf{X}|\mathsf{Y}^{[K]}}} - \sum_{k=1}^{K} \left(\frac{1}{\sigma^2_{\mathsf{X}|\mathsf{Y}^k}} - \frac{1}{d_k} \right), \tag{37}$$

$$\sigma^2_{\mathsf{X}|\mathsf{Y}^k} \le d_k \le \sigma^2_{\mathsf{X}}. \tag{38}$$

Here $\sigma^2_{\mathsf{X}|\mathsf{Y}^k} \triangleq \lim_{i\to\infty} \sigma_{\mathsf{X}_i|\mathsf{Y}^k_i}$ and $\sigma^2_{\mathsf{X}|\mathsf{Y}^{[K]}} \triangleq \lim_{i\to\infty} \sigma^2_{\mathsf{X}_i|\mathsf{Y}^{[K]}_i}$ denote the stationary MMSE achievable in the estimation of X_i from Y^k_i and $\mathsf{Y}^{[K]}_i$ respectively, i.e., without memory of the past.

If $a = 0$, the observed process (2) becomes a stationary memoryless Gaussian process, the predictive MMSEs reduce to the variance of X_i: $\bar{d} = \bar{d}_k = \sigma^2_{\mathsf{X}} = \sigma^2_{\mathsf{V}}$; similarly, $\sigma^2_{\mathsf{X}|\mathsf{Y}^k} = \sigma^2_{\mathsf{X}\|\mathsf{Y}^k}$ and $\sigma^2_{\mathsf{X}|\mathsf{Y}^{[K]}} = \sigma^2_{\mathsf{X}\|\mathsf{Y}^{[K]}}$, and the result of Theorem 2 coincides with the classical Gaussian CEO sum rate - distortion function (36). This shows that if the source is memoryless, asymptotically there is no benefit in keeping the memory of previously encoded estimates as permitted by Definition 1. Classical codes that forget the past after encoding the current block of length n perform just as well.

If $|a| > 1$, the benefit due to memory is infinite: indeed, since the source is unstable, $\sigma^2_{\mathsf{X}} = \infty$, while $\bar{d} < \infty$. If $|a| < 1$, that benefit is finite and is characterized by the discrepancy between the stationary variance of the process $\{\mathsf{X}_i\}_{i=1}^{\infty}$ $\sigma^2_{\mathsf{X}} = \frac{\sigma^2_{\mathsf{V}}}{1-a^2}$ and the steady-state predictive MMSE $\bar{d} < \sigma^2_{\mathsf{X}}$, as well as that between $\sigma^2_{\mathsf{X}|\mathsf{Y}^k}$ and $\sigma^2_{\mathsf{X}\|\mathsf{Y}^k}$.

IV. Loss due to isolated observers

Unrestricted communication among the encoders is equivalent to having one encoder that sees all the observation processes $\left\{Y_i^{[K]}\right\}$. It is also equivalent to allowing joint encoding policies $P_{B_{[t]}^{[K]}\|Y_{[t]}^{[K]}}$ in lieu of independent encoding policies $\prod_{k=1}^{K} P_{B_{[t]}^k\|Y_{[t]}^k}$ in Definition 1.

The lossy compression setup in which the encoder has access only to a noise-corrupted version of the source has been referred to as "remote", "indirect", or "noisy" rate-distortion problem [33]–[36]. A causal setting was considered in [20, Th. 5–8, Cor. 1].

We denote the joint encoding counterpart of the operational fundamental limit $R_{\text{CEO}}(d)$ (20) by $R_{\text{rm}}(d)$ (remote).

The following result is a corollary to Theorem 2.

Corollary 3 (Causal remote rate-distortion function)**.** *For all $\sigma^2_{\mathsf{X}\|\mathsf{Y}^{[K]}} < d < \sigma^2_{\mathsf{X}}$, the sum rate - distortion function with joint encoding for the Gauss-Markov source in (2) observed through the Gaussian channels in (3) is given by*

$$R_{\text{rm}}(d) = \frac{1}{2} \log \frac{\bar{d} - \sigma^2_{\mathsf{X}\|\mathsf{Y}^{[K]}}}{d - \sigma^2_{\mathsf{X}\|\mathsf{Y}^{[K]}}}. \tag{39}$$

Proof. [24]. □

The loss due to isolated encoders is bounded as follows.

Theorem 3 (Loss due to isolated observers)**.** *Consider the causal Gaussian CEO problem (2), (3). Assume that target distortion d satisfies*

$$\frac{1}{d} \ge \frac{1}{\sigma^2_{\mathsf{X}\|\mathsf{Y}^{[K]}}} + \frac{K}{\sigma^2_{\mathsf{X}}} - \min_{k\in[K]} \frac{K}{\sigma^2_{\mathsf{X}\|\mathsf{Y}^k}}. \tag{40}$$

Then, the rate loss due to isolated observers is bounded as

$$R_{\text{CEO}}(d) - R_{\text{rm}}(d) \le (K-1)\left(R_{\text{rm}}(d) - R(d)\right), \tag{41}$$

with equality if and only if $\sigma^2_{\mathsf{X}\|\mathsf{Y}^k}$ are all the same.

Proof. We upper-bound $R_{\text{CEO}}(d)$ by waterfilling over d_k's. While for $t = 1$ waterfilling is optimal [5], it is only suboptimal if $t > 1$ due to the memory of the past steps at the encoders and the decoder. This is unsurprising as for the same reason waterfilling cannot be applied to solve the vector Gaussian rate-distortion problem for $t > 1$ [20, Remark 2]. See [24] for details. □

Theorem 3 parallels the corresponding result for the classical Gaussian CEO problem [37, Cor. 1], and recovers it if $a = 0$. It's interesting that in both cases, the rate loss is bounded above by $K-1$ times the difference between the remote and the direct rate-distortion functions.

V. Conclusion

In this paper, we set up the causal CEO problem (Definition 1, Definition 2) and we prove that the sum rate - distortion function is given by the directed mutual information from the encoders to the decoder minimized subject to the distortion constraint and the separate encoding constraint (Theorem 1). The proof of the direct coding theorem hinges upon an SLC-based nonasymptotic bound that extends [28, Th. 6] to the case with $K > 2$ observers and $t > 1$ time steps. An asymptotic analysis of that bound leads to an extension of the Berger-Tung inner bound [10], [11] to $t > 1$ time steps.

By solving the minimal directed mutual information problem in Theorem 1, we characterize the Gaussian sum rate - distortion function as a convex optimization problem over K parameters (Theorem 2). We give an explicit formula in the identical-channels case (Corollary 1) and study its asymptotic behavior as $K \to \infty$ (Corollary 2). We derive the causal Gaussian remote rate-distortion function as a corollary to Theorem 2 with $K = 1$ (Corollary 3). Using a suboptimal waterfilling allocation over the K optimization parameters in Theorem 2, we upper-bound the rate loss due to separated observers (Theorem 3).

As future work, it will be interesting to determine the full rate-distortion region of the causal Gaussian CEO problem as opposed to the sum rate we found in this paper. Certain causal multiterminal source coding problems also appear within reach in view of the result in [8] and the applicability of our nonasymptotic achievability bound to multiterminal source coding. Finally, computing the rate-distortion region for Gaussian processes beyond the Gauss-Markov source with i.i.d. components would be an important advance.

References

[1] T. Berger, Z. Zhang, and H. Viswanathan, "The CEO problem [multiterminal source coding]," *IEEE Transactions on Information Theory*, vol. 42, no. 3, pp. 887–902, 1996.

[2] H. Viswanathan and T. Berger, "The quadratic Gaussian CEO problem," *IEEE Transactions on Information Theory*, vol. 43, no. 5, pp. 1549–1559, 1997.

[3] Y. Oohama, "The rate-distortion function for the quadratic Gaussian CEO problem," *IEEE Transactions on Information Theory*, vol. 44, no. 3, pp. 1057–1070, May 1998.

[4] V. Prabhakaran, D. Tse, and K. Ramachandran, "Rate region of the quadratic Gaussian CEO problem," in *Proceedings 2004 International Symposium on Information Theory*, June 2004, p. 119.

[5] J. Chen, X. Zhang, T. Berger, and S. B. Wicker, "An upper bound on the sum-rate distortion function and its corresponding rate allocation schemes for the CEO problem," *IEEE Journal on Selected Areas in Communications*, vol. 22, no. 6, pp. 977–987, 2004.

[6] H. Behroozi and M. R. Soleymani, "Optimal rate allocation in successively structured Gaussian CEO problem," *IEEE Transactions on Wireless Communications*, vol. 8, no. 2, pp. 627–632, 2009.

[7] J. Chen and T. Berger, "Successive Wyner–Ziv coding scheme and its application to the quadratic Gaussian CEO problem," *IEEE Transactions on Information Theory*, vol. 54, no. 4, pp. 1586–1603, 2008.

[8] A. B. Wagner, S. Tavildar, and P. Viswanath, "Rate region of the quadratic Gaussian two-encoder source-coding problem," *IEEE Transactions on Information Theory*, vol. 54, no. 5, pp. 1938–1961, 2008.

[9] A. B. Wagner and V. Anantharam, "An improved outer bound for multiterminal source coding," *IEEE Transactions on Information Theory*, vol. 54, no. 5, pp. 1919–1937, 2008.

[10] T. Berger, *Multi-terminal source coding. Chapter in The Information Theory Approach to Communications*. Springer-Verlag, 1978.

[11] S.-Y. Tung, "Multiterminal source coding," Ph.D. dissertation, School of Electrical Engineering, Cornell University, 1978.

[12] J. Wang, J. Chen, and X. Wu, "On the sum rate of Gaussian multiterminal source coding: New proofs and results," *IEEE Transactions on Information Theory*, vol. 56, no. 8, pp. 3946–3960, 2010.

[13] E. Ekrem and S. Ulukus, "An outer bound for the vector Gaussian CEO problem," *IEEE Transactions on Information Theory*, vol. 60, no. 11, pp. 6870–6887, 2014.

[14] J. Wang and J. Chen, "Vector Gaussian multiterminal source coding," *IEEE Transactions on Information Theory*, vol. 60, no. 9, pp. 5533–5552, 2014.

[15] T. A. Courtade and T. Weissman, "Multiterminal source coding under logarithmic loss," *IEEE Transactions on Information Theory*, vol. 60, no. 1, pp. 740–761, Jan. 2014.

[16] A. Gorbunov and M. S. Pinsker, "Nonanticipatory and prognostic ϵ-entropies and message generation rates," *Problemy Peredachi Informatsii*, vol. 9, no. 3, pp. 12–21, 1973.

[17] ——, "Prognostic epsilon entropy of a Gaussian message and a Gaussian source," *Problemy Peredachi Informatsii*, vol. 10, no. 2, pp. 5–25, 1974.

[18] S. Tatikonda, A. Sahai, and S. Mitter, "Stochastic linear control over a communication channel," *IEEE Transactions on Automatic Control*, vol. 49, no. 9, pp. 1549–1561, 2004.

[19] E. Silva, M. Derpich, J. Ostergaard, and M. Encina, "A characterization of the minimal average data rate that guarantees a given closed-loop performance level," *IEEE Transactions on Automatic Control*, 2016.

[20] V. Kostina and B. Hassibi, "Rate-cost tradeoffs in control," *IEEE Transactions on Automatic Control*, Apr. 2019.

[21] T. Tanaka, K.-K. K. Kim, P. A. Parrilo, and S. K. Mitter, "Semidefinite programming approach to Gaussian sequential rate-distortion tradeoffs," *IEEE Transactions on Automatic Control*, vol. 62, no. 4, pp. 1896–1910, 2017.

[22] V. Kostina and B. Hassibi, "Rate-cost tradeoffs in scalar LQG control and tracking with side information," in *2018 56th Annual Allerton Conference on Communication, Control, and Computing*, Monticello, IL, Oct. 2018, pp. 421–428.

[23] A. P. Johnston and S. Yüksel, "Stochastic stabilization of partially observed and multi-sensor systems driven by unbounded noise under fixed-rate information constraints," *IEEE Transactions on Automatic Control*, vol. 59, no. 3, pp. 792–798, 2014.

[24] V. Kostina and B. Hassibi, "Fundamental limits of distributed tracking," *arXiv preprint arXiv:1910.02534*, May 2020.

[25] G. Kramer, "Directed information for channels with feedback," Ph.D. dissertation, ETH Zurich, 1998.

[26] J. Massey, "Causality, feedback and directed information," in *Proc. Int. Symp. Inf. Theory Applic.(ISITA-90)*, 1990, pp. 303–305.

[27] M. H. Yassaee, M. R. Aref, and A. Gohari, "A technique for deriving one-shot achievability results in network information theory," in *Proceedings 2013 IEEE International Symposium on Information Theory*, Istanbul, Turkey, July 2013.

[28] ——, "A technique for deriving one-shot achievability results in network information theory," 2013.

[29] N. Merhav, "The generalized stochastic likelihood decoder: Random coding and expurgated bounds," *IEEE Transactions on Information Theory*, vol. 63, no. 8, pp. 5039–5051, 2017.

[30] T. Tanaka, "Semidefinite representation of sequential rate-distortion function for stationary Gauss-Markov processes," in *Proceedings 2015 IEEE Conference on Control Applications (CCA)*, Sep. 2015, pp. 1217–1222.

[31] N. Guo and V. Kostina, "Optimal causal rate-constrained sampling of the Wiener process," in *Proceedings 57th Annual Allerton Conference on Communication, Control and Computing,*, Monticello, IL, Sep. 2019.

[32] M. S. Derpich and J. Ostergaard, "Improved upper bounds to the causal quadratic rate-distortion function for Gaussian stationary sources," *IEEE Transactions on Information Theory*, vol. 58, no. 5, pp. 3131–3152, May 2012.

[33] R. Dobrushin and B. Tsybakov, "Information transmission with additional noise," *IRE Transactions on Information Theory*, vol. 8, no. 5, pp. 293 –304, Sep. 1962.

[34] T. Berger, *Rate distortion theory*. Prentice-Hall, Englewood Cliffs, NJ, 1971.

[35] H. S. Witsenhausen, "Indirect rate distortion problems," *IEEE Transactions on Information Theory*, vol. 26, no. 5, pp. 518–521, Sep. 1980.

[36] V. Kostina and S. Verdú, "Nonasymptotic noisy lossy source coding," *IEEE Transactions on Information Theory*, vol. 62, no. 11, pp. 6111–6123, Nov. 2016.

[37] V. Kostina, "Rate loss in the Gaussian CEO problem," in *Proceedings 2019 IEEE Information Theory Workshop*, Visby, Gotland, Sweden, Aug. 2019.