Fundamental limits of distributed tracking

Victoria Kostina

***Abstract*—Consider the following communication scenario. A Gauss-Markov source is observed by K isolated observers via independent AWGN 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 mean-square error running estimate of the source. This is a causal version of the Gaussian CEO problem. We determine the minimum asymptotically achievable sum rate required to achieve a given mean-square error, which is stated as an optimization problem over K parameters. We give an explicit expression for the symmetrical case, and compute the limit of the sum rate as $K \to \infty$, which turns out to be finite and nontrivial. Furthermore, using a suboptimal waterfilling allocation among the K parameters, we explicitly bound the loss due to a lack of cooperation among the observers; that bound is attained with equality in the symmetrical case.**

***Index Terms*—Linear stochastic control, LQG control, remote control, rate-distortion tradeoff, causal rate-distortion theory, Gauss-Markov source, real-time tracking, Gaussian CEO problem.**

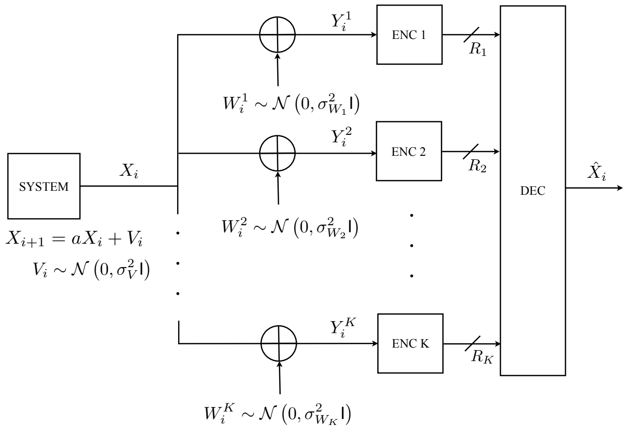

Fig. 1. The causal CEO problem.

I. Introduction

We set up the causal Gaussian CEO (chief executive or estimation officer) problem as follows. The Gauss-Markov source

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

is observed by K noisy observers; the k-th observer sees

$$Y_i^k = X_i + W_i^k, \quad k = 1, \ldots, K, \tag{2}$$

where $\{V_i, W_i^1, W_i^2, \ldots, W_i^K\}_{i=1}^T$ are independent Gaussian vectors of length n; $V_i \sim \mathcal{N}(0, \sigma_V^2 \mathsf{I})$; $W_i^k \sim \mathcal{N}(0, \sigma_{W_k}^2 \mathsf{I})$; X_1 is Gaussian and independent of $\{V_i, W_i^1, W_i^2, \ldots, W_i^K\}_{i=1}^T$. The observers communicate to the decoder (the CEO, chief executive or estimation officer) via their separate rate-constrained channels. The goal of the decoder at time i is to causally choose $\hat{X}_i$ based on the information received up to that point so that the long-term average mean-square error (MSE)

$$\frac{1}{t} \sum_{i=1}^{t} \mathbb{E}\left[\|X_i - \hat{X}_i\|^2\right] \tag{3}$$

is minimized; t is the time horizon. There is a feedback from the decoder to each of the observers. See Fig. 1. We call the causal Gaussian CEO problem *symmetrical* if $\sigma_{W_1}^2 = \ldots = \sigma_{W_K}^2$.

In the classical (noncausal) setting, the CEO problem was first introduced by Berger et al. [1] in the context of K noisy observations of a discrete source. The quadratic Gaussian CEO problem was studied by Viswanathan and Berger [2], who proved an achievability bound on the rate-distortion dimension for the symmetric case, 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. Wang et al. [9] showed a simple converse on the sum rate of the Gaussian CEO problem. Most recently, Courtade and Weissman [10] determined the distortion region of the distributed source coding and the CEO problem under logarithmic loss.

In the case of a single noiseless observer, the Gaussian *causal rate-distortion function*, introduced by Gorbunov and Pinsker [11], describes the fundamental operational limits of causal tracking. The link between the minimum attainable linear quadratic Gaussian (LQG) control cost and the causal rate-distortion function is elucidated in [12]–[14]. A semidefinite program to compute the causal rate-distortion function for vector Gauss-Markov sources is provided in [15]. The noisy causal rate-distortion function, which corresponds to $K = 1$ in (2), is computed in [14]. The causal rate-distortion function of the Gauss-Markov source with Gaussian side observation available at decoder (causal Wyner-Ziv) is computed in [16]. That causal Wyner-Ziv setting can be viewed a special case of our causal CEO problem (1), (2) with two observers, with the second observer enjoying an infinite rate. Stability of linear Gaussian systems with multiple isolated observers was investigated in [17].

In this paper, we characterize the minimum asymptoti-

The author is with California Institute of Technology (e-mail: vkostina@caltech.edu). This work was supported in part by the National Science Foundation (NSF) under grants CCF-1751356 and CCF-1817241.

cally achievable sum rate $R_1+\ldots+R_K$ required to achieve a given mean-square error (3) in the causal distributed tracking setting of Fig. 1. We recover the noisy and noiseless causal rate-distortions as special cases. We study the rate loss due to a lack of cooperation between the different encoders: as long as the target distortion is not too small, the rate loss is bounded above by $K-1$ times the difference between the noisy and noiseless rate-distortion functions. This parallels the corresponding result for the classical Gaussian CEO problem [18, Cor. 1]. In the symmetrical case, we compute $\lim_{K\to\infty}\sum_{k=1}^K R_k$, extending the result of Oohama [3, Cor. 1] to the causal setting and recovering it if $a=0$.

Notation: Logarithms are natural base. For a natural number M, $[M] \triangleq \{1,\ldots,M\}$. Notation $X \hookleftarrow Y$ reads "replace X by Y". We indicate the temporal index in the subscript and the spacial 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 $(Y_i^1,\ldots,Y_i^K)^\mathsf{T}$; $Y_{[t]}^{[K]} \triangleq (Y_{[t]}^1,\ldots,Y_{[t]}^K)$. $\mathcal{D}$ denotes delay by one, i.e. $\mathcal{D}X_{[t]} = (0, X_1,\ldots,X_{t-1})$. For a random vector X with i.i.d. components, X denotes a random variable distributed the same as each component of X. For a random vector X, its normalized (by dimension) variance is denoted by σ_X^2; its normalized conditional variance given a random object Y is denoted by $\sigma_{X|Y}$. For a random process $\{X_i\}$ on $\mathbb{R}^n$, its normalized stationary variance (can be $+\infty$) is denoted by

$$\sigma_X^2 \triangleq \limsup_{i\to\infty} \frac{1}{n}\mathbb{E}\left[X_i^2\right]. \tag{4}$$

The normalized minimum MSE (MMSE) in the estimation of X_i from $Y_{[i]}^{[K]}$ is denoted by

$$\sigma^2_{X_i|Y_{[i]}^{[K]}} \triangleq \frac{1}{n}\mathbb{E}\left[\left(X-\mathbb{E}\left[X_i|Y_{[i]}^{[K]}\right]\right)^\mathsf{T}\left(X-\mathbb{E}\left[X_i|Y_{[i]}^{[K]}\right]\right)\right], \tag{5}$$

and the normalized steady-state causal MMSE by

$$\sigma^2_{X\|Y^{[K]}} \triangleq \limsup_{i\to\infty} \sigma^2_{X_i|Y_{[i]}^{[K]}}. \tag{6}$$

We use the following shorthand notation for causally conditional probability kernels [19]:

$$P_{Y_{[t]}\|X_{[t]}} \triangleq \prod_{i=1}^t P_{Y_i|Y_{[i-1]},X_{[i]}}. \tag{7}$$

II. Results and Discussion

A. Minimum achievable sum rate in the causal Gaussian CEO problem

A causal CEO code is formally defined as follows.

Definition 1 (Causal CEO code)**.** *An (R,d,t) causal CEO code for a discrete-time random process on $\mathbb{R}^n$ $\{X_i\}_{i=1}^t$, observed through $P_{Y_{[t]}^{[K]}\|X_{[t]}}$ consists of*

a) K encoding policies

$$P_{L_{[t]}^k\|Y_{[t]}^k,\mathcal{D}U_{[t]}^k}, \quad k=1,\ldots,K, \tag{8}$$

that satisfy the sum rate constraint

$$\frac{1}{nt}\sum_{i=1}^t\sum_{i=1}^K \log|\mathcal{L}_i^k| \le R, \tag{9}$$

where $\mathcal{L}_i^k$ is the alphabet of L_i^k, and that give rise to

$$P_{L_{[t]}^{[K]}\|Y_{[t]}^{[K]},\mathcal{D}U_{[t]}^{[K]}} = \prod_{k=1}^K P_{L_{[t]}^k\|Y_{[t]}^k,\mathcal{D}U_{[t]}^k}; \tag{10}$$

b) a decoding policy $P_{U_{[t]}^{[K]}\|L_{[t]}^{[K]}}$*, that, together with the MMSE estimator*

$$\hat{X}_i = \mathbb{E}\left[X_i|U_{[i]}^{[K]}\right], \quad i=1,\ldots,t \tag{11}$$

satisfies the average MSE constraint

$$\frac{1}{nt}\sum_{i=1}^t \mathbb{E}\left[\|X_i-\hat{X}_i\|^2\right] \le d. \tag{12}$$

One can think of L_i^k as the summary of the codeword that k-th encoder intends to transmit at time i, and of U_i^k as the decoder's reconstruction of that codeword. The decoder reconstructs U_i^k by jointly processing all the summaries received from all the encoders up to that time, i.e., $L_{[i]}^{[K]}$. That process is not error-free, and to avoid error propagation Definition 1 allows noiseless feedback from the decoder back to each of the encoders containing the decoder's estimate U_i^k of that encoder's intended codeword.

The asymptotically achievable rate region of the causal CEO problem is defined as follows.

Definition 2 (Asymptotic sum rate - distortion function)**.** *The rate-distortion pair (R,d) is achievable if for any $\gamma>0$, there exists an $(R+\gamma,d+\gamma,t)$ causal CEO code for all t and n large enough. The causal sum rate - distortion function is defined as follows:*

$$R_{\mathrm{CEO}}(d) \triangleq \inf\Big\{R\colon (R,d) \text{ is achievable in the CEO problem}\Big\} \tag{13}$$

If there were no communication rate constraints, at time i, the decoder would have directly observed $Y_i^{[K]}$, and it could form the minimum MSE (MMSE) estimate

$$\bar{X}_i \triangleq \mathbb{E}\left[X_i|Y_{[i]}^{[K]}\right]. \tag{14}$$

of X_i based on those noisy observations. Similarly, kth encoder MMSE estimate based on the observations it has seen far is given by

$$\bar{X}_i^k \triangleq \mathbb{E}\left[X_i|Y_{[i]}^k\right]. \tag{15}$$

Both (14) and (15) can be computed using the Kalman filter. The steady-state causal estimation MMSEs $\sigma^2_{X\|Y}$ and $\sigma^2_{X\|Y^k}$ can be computed by solving the corresponding Riccati recursions.

In Theorem 1, the causal sum rate - distortion function is expressed as a convex optimization problem over parameters $d_k = \sigma^2_{X\|U^k}$, which determine the individual rates of transmitters, to achieve a target distortion $d = \sigma^2_{X\|U^{[K]}}$. Note that

$$\sigma^2_{X||\mathcal{D}U^{[K]}} = a^2 d + \sigma^2_V, \tag{16}$$

$$\sigma^2_{X||\mathcal{D}U^k} = a^2 d_k + \sigma^2_V. \tag{17}$$

Theorem 1. *For all $d > \sigma^2_{X\|Y^{[K]}}$, the causal sum rate - distortion function for the source in (1) observed through the channels in (2) is given by*

$$\begin{aligned} R_{\text{CEO}}(d) &= \frac{1}{2}\log\frac{\sigma^2_{X\|\mathcal{D}U^{[K]}}}{d} \\ &+ \min_{\{d_k\}_{k=1}^K}\sum_{k=1}^K \frac{1}{2}\log\frac{\sigma^2_{X\|\mathcal{D}U^k}-\sigma^2_{X\|Y^k}}{d_k-\sigma^2_{X\|Y^k}}\frac{d_k}{\sigma^2_{X\|\mathcal{D}U^k}}, \end{aligned} \tag{18}$$

where the minimum is over d_k, $k = 1, \ldots, K$, that satisfy

$$\frac{1}{d} \le \frac{1}{\sigma^2_{X\|Y^{[K]}}} - \sum_{k=1}^K\left(\frac{1}{\sigma^2_{X\|Y^k}}-\frac{1}{d_k}\right), \tag{19}$$

$$\sigma^2_{X\|Y^k} \le d_k \le \sigma^2_X. \tag{20}$$

Proof. Section III-D. □

To approach the sum rate in Theorem 1 in practice, one could find the optimal d_k's by solving the convex optimization problem (18), (19), (20). These parameters determine the rate of each encoder. At time i, the kth encoder quantizes the difference between $\mathbb{E}\left[X_i|U^k_{[i-1]}\right]$ and (15), and composes a summary of the quantized index using binning. The decoder performs joint decoding of $U_i^{[K]}$ using all received summaries and computes (11) using the Kalman filter.

B. Loss due to isolated observers

It is interesting to compare the minimum sum rate in (18) to what would be achievable had the encoders acted cooperatively. Full cooperation is equivalent to having one encoder that simultaneously observes all the observation processes $\{Y_i^{[K]}\}$. The corresponding code and the information-theoretic limit are formally defined as follows.

Definition 3 (Causal noisy rate-distortion code)**.** *An (R, d, t) causal CEO noisy rate-distortion code for a discrete-time random process on $\mathbb{R}^n$ $\{X_i\}_{i=1}^t$, observed through $P_{Y_{[t]}\|X_{[t]}}$, consists of*

a) an encoding policy *$P_{L_{[t]}\|Y_{[t]},\mathcal{D}U_{[t]}}$ that satisfies the rate constraint $\frac{1}{nt}\sum_{i=1}^t \log|\mathcal{L}_i| \le R$, where $\mathcal{L}_i$ is the alphabet of L_i;*

b) a decoding policy *$P_{U_{[t]}\|L_{[t]}}$ that, together with the MMSE estimator $\hat{X}_i = \mathbb{E}\left[X_i|U_{[i]}\right]$, $i = 1, \ldots, t$, satisfies the average MSE constraint (12).*

To particularize Definition 3 to our scenario with full cooperation, consider X_i in (1) and $Y_i \leftarrow Y_i^{[K]}$ in (2). This is equivalent to dropping the product constraint on the combined encoding policy $P_{L_{[t]}^{[K]}\|Y_{[t]}^{[K]},\mathcal{D}U_{[t]}^{[K]}}$ in (10) in Definition 1.

Definition 4 (Causal noisy rate-distortion function)**.** *The rate-distortion pair (R, d) is achievable if for any $\gamma > 0$, there exists an $(R+\gamma, d+\gamma, t)$ noisy causal code for all t and n large enough. The causal noisy rate-distortion function is defined as follows:*

$$R_{\text{noisy}}(d) \triangleq \inf\Big\{R\colon (R,d) \textit{ is achievable in the causal noisy rate-distortion problem}\Big\} \tag{21}$$

As a simple corollary to Theorem 1, we can record the following.

Theorem 2 (Noisy causal rate-distortion function)**.** *For all $d > \sigma^2_{X\|Y^{[K]}}$, the causal noisy rate-distortion function for the source in (1) based on the observations (2) is given by*

$$R_{\text{noisy}}(d) = \frac{1}{2}\log\frac{\sigma^2_{X\|\mathcal{D}U}-\sigma^2_{X\|Y^{[K]}}}{d-\sigma^2_{X\|Y^{[K]}}}, \tag{22}$$

where

$$\sigma^2_{X||\mathcal{D}U} = a^2 d + \sigma^2_V. \tag{23}$$

Previously, $R_{\text{noisy}}(d)$ was computed in [14] in a different form using a different method; both forms are equivalent (Appendix A). If the source is observed directly, $\sigma^2_{X\|Y^{[K]}} = 0$, and (22) reduces to noiseless causal rate-distortion function [11] (and e.g. [12], [20, Th. 3], [14, (64)]):

$$R(d) = \frac{1}{2}\log\frac{\sigma^2_{X\|\mathcal{D}U}}{d}. \tag{24}$$

The loss due to isolated encoders is bounded as follows.

Theorem 3 (Loss by separation)**.** *In the causal Gaussian CEO problem (1), (2), for d small enough so that*

$$\frac{1}{d} > \frac{1}{\sigma^2_{X\|Y^{[K]}}} + \frac{K}{\sigma^2_X} - \min_k \frac{K}{\sigma^2_{X\|Y^k}}, \tag{25}$$

the loss due to separation is bounded as

$$R_{\text{CEO}}(d) - R_{\text{noisy}}(d) \le (K-1)\left(R_{\text{noisy}}(d) - R(d)\right), \tag{26}$$

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

Proof. Section III-G. □

Theorem 3 parallels the corresponding result for the classical Gaussian CEO problem [18, 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 noisy and the noiseless rate-distortion functions.

C. A large number of identical observers

Denote the sum rate - distortion function of a symmetrical causal Gaussian CEO problem by $R^{K-\text{sym}}_{\text{CEO}}(d)$. As a simple corollary to Theorem 3, we note that

$$\begin{aligned} R^{K-\text{sym}}_{\text{CEO}}(d) &= \frac{1}{2}\log\frac{\sigma^2_{X\|\mathcal{D}U^{[K]}}}{d} \\ &+ \frac{K}{2}\log\frac{\sigma^2_{X\|\mathcal{D}U^1}-\sigma^2_{X\|Y^1}}{d_1-\sigma^2_{X\|Y^1}}\frac{d_1}{\sigma^2_{X\|\mathcal{D}U^1}}, \end{aligned} \tag{27}$$

where d_1 satisfies

$$\frac{1}{d} = \frac{1}{\sigma^2_{X\|Y^{[K]}}} - \frac{K}{\sigma^2_{X\|Y^1}} + \frac{K}{d_1}. \tag{28}$$

It turns out that the limit $\lim_{K\to\infty} R^{K-\text{sym}}_{\text{CEO}}(d)$ is finite and nontrivial.

Theorem 4. *In the symmetrical causal Gaussian CEO problem* (1), (2),

$$\lim_{K\to\infty} R^{K-\text{sym}}_{\text{CEO}}(d) = \frac{1}{2}\log\frac{\sigma^2_{X\|\mathcal{D}U^{[K]}}}{d} + \frac{1}{2}\frac{\frac{1}{d} - \frac{1}{\sigma^2_{X\|\mathcal{D}U^{[K]}}}}{\frac{1}{\sigma^2_{X\|Y^1}} - \frac{1}{\sigma^2_X}}. \tag{29}$$

Proof. Section III-H. □

The second term in (29) is the penalty due to isolated encoders. Theorem 4 extends the result of Oohama [3, Cor. 1] to causal noisy compression of the Gauss-Markov source, and recovers it if $a = 0$.

D. The effect of memory

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

$$R^{\text{no memory}}_{\text{CEO}}(d) = \frac{1}{2}\log\frac{\sigma^2_X}{d} + \min_{\{d_k\}_{k=1}^K} \sum_{k=1}^K \frac{1}{2}\log\frac{\sigma^2_X - \sigma^2_{X|Y^k}}{d_k - \sigma^2_{X|Y^k}}\frac{d_k}{\sigma^2_X}, \tag{30}$$

where the minimum is over

$$\frac{1}{d} \leq \frac{1}{\sigma^2_{X|Y^{[K]}}} - \sum_{k=1}^K \left(\frac{1}{\sigma^2_{X|Y^k}} - \frac{1}{d_k}\right), \tag{31}$$

$$\sigma^2_{X|Y^k} \leq d_k \leq \sigma^2_X. \tag{32}$$

Here $\sigma^2_{X|Y^k} = \lim_{i\to\infty} \sigma_{X_i|Y_i^k}$ ($\sigma^2_{X|Y^{[K]}} = \lim_{i\to\infty} \sigma^2_{X_i|Y_i^{[K]}}$) denotes the stationary MMSE achievable in the estimation of X_i from Y_i^k ($Y_i^{[K]}$), i.e., without memory of the past.

If $a = 0$, the observed process (1) becomes a stationary memoryless Gaussian process, the predictive MMSEs reduce to the variance of X_i: $\sigma^2_{X\|\mathcal{D}U^{[K]}} = \sigma^2_{X\|\mathcal{D}U^k} = \sigma^2_X = \sigma^2_V$; similarly, $\sigma^2_{X|Y^k} = \sigma^2_{X\|Y^k}$ and $\sigma^2_{X|Y^{[K]}} = \sigma^2_{X\|Y^{[K]}}$, and the result of Theorem 1 coincides with the classical Gaussian CEO sum rate - distortion function (30). This shows that if the source is memoryless, 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, $\sigma^2_X = \infty$, while $\sigma^2_{X\|\mathcal{D}U^{[K]}} < \infty$. If $|a| < 1$, that benefit is characterized by the discrepancy between the stationary variance of the process $\{X\}_{i=1}^\infty$ $\sigma^2_X = \frac{\sigma^2_V}{1-a^2}$ and the steady-state predictive MMSE $\sigma_{X\|\mathcal{D}U} < \sigma^2_X$, as well as that between $\sigma^2_{X|Y^k}$ and $\sigma^2_{X\|Y^k}$.

E. The effect of observation noise

If there is no observation noise, i.e. $\sigma^2_{W_k} = 0$, then the sum rate in Theorem 1 collapses to the causal rate-distortion function of the Gauss-Markov process (24). There is no penalty due to isolated encoders in that case.

III. Proofs

A. Preliminaries

We prepare some notation and tools. Given a joint distribution of random vectors $X_{[t]}$ and $Y_{[t]}$, the directed mutual information is defined as [21]

$$I(X_{[t]} \to Y_{[t]}) \triangleq \sum_{i=1}^t I(X_{[i]}; Y_i | Y_{[i-1]}). \tag{33}$$

Causally conditioned directed information is defined as

$$I(X_{[t]} \to Y_{[t]} \| Z_{[t]}) \triangleq \sum_{i=1}^t I(X_{[i]}; Y_i | Y_{[i-1]}, Z_{[i]}). \tag{34}$$

Lemma 1 ([19, (3.14)–(3.16)]). *Directed information chain rule [19]:*

$$I((X_{[t]}, Y_{[t]}) \to Z_{[t]}) = I(X_{[t]} \to Z_{[t]}) + I(Y_{[t]} \to Z_{[t]} \| X_{[t]}). \tag{35}$$

We also use the following notation. For random processes $\{X_i\}_{i=1}^\infty$, $\{Y_i\}_{i=1}^\infty$, $\{Z_i\}_{i=1}^\infty$,

$$\underline{I}(X \to Y) \triangleq \liminf_{t\to\infty} \frac{1}{t} I(X_{[t]} \to Y_{[t]}), \tag{36}$$

$$\underline{I}(X \to Y \| Z) \triangleq \liminf_{t\to\infty} \frac{1}{t} I(X_{[t]} \to Y_{[t]} \| Z_{[t]}). \tag{37}$$

B. General converse and achievability bounds

Theorem 5. *The causal sum rate - distortion function for the source in* (1) *observed through the channels in* (2) *is bounded below as,*

$$R_{\text{CEO}}(d) \geq \inf \underline{I}\left(\bar{\mathsf{X}}^{[K]} \to \mathsf{U}^{[K]}\right), \tag{38}$$

where

$$\bar{\mathsf{X}}^k_i \triangleq \mathbb{E}\left[\mathsf{X}_i | \mathsf{Y}^k_{[i]}\right], \tag{39}$$

and the infimum is over sequences of single-letter causal kernels

$$P_{\mathsf{U}^k_{[t]} \| \bar{\mathsf{X}}^k_{[t]}}, \quad k = 1, \ldots, K, \tag{40}$$

$t = 1, 2, \ldots$, *giving rise to*

$$P_{\mathsf{U}^{[K]}_{[t]} \| \bar{\mathsf{X}}^{[K]}_{[t]}} = \prod_{k=1}^K P_{\mathsf{U}^k_{[t]} \| \bar{\mathsf{X}}^k_{[t]}}, \tag{41}$$

that satisfy

$$\sigma^2_{\mathsf{X}\|\mathsf{U}^{[K]}} \leq d. \tag{42}$$

Proof. The proof follows standard arguments: for any (R, d, t) code in Definition 1,

$$nR \geq \sum_{k=1}^{K} H(U_{[t]}^k) \tag{43}$$

$$\geq H(U_{[t]}^{[K]}) \tag{44}$$

$$\geq I\left(\bar{X}_{[t]}^{[K]} \to U_{[t]}^{[K]}\right). \tag{45}$$

Minimizing the right-hand side over all causal kernels (8) that satisfy (10) and (12) (ignoring the possibility of decoding error in reconstructing U_i^k) and applying standard single-letterization arguments, we obtain (38). □

Theorem 6. *The causal sum rate - distortion function for the source in* (1) *observed through the channels in* (2) *is bounded above as,*

$$R_{\text{CEO}}(d) \leq \inf \underline{I}\left(\mathsf{X}^{[K]} \to \mathsf{U}^{[K]}\right), \tag{46}$$

where the infimum is over sequences of single-letter causal kernels (40) *giving rise to* (41) *that satisfy* (42)*.*

Proof outline. This result is a generalization of Berger-Tung inner bound to our causal setting. For $i = 1$, k-th codebook is drawn i.i.d. from $P_{\mathsf{U}_1^k}^n$. Each encoder uses typicality encoding with binning to produce a summary of the chosen codeword, in isolation. The decoder uses joint typicality decoding to jointly recover (with high probability) the codewords $U_1^{[K]}$ sent by all encoders and, using those, to produce an estimate of the source, $\hat{X}_1$, using Lemma 2. For $t > 1$, at each $1 < i \leq t$, the k-th encoder has access to side information consisting of the previous decoded outputs of kth encoder, $U_{[i-1]}^k$, and the decoder has access to $U_{[i-1]}^{[K]}$. Encoding and decoding for time steps $1 < i \leq t$ is done similarly to step 1, except codebooks at each time step $1 < i \leq t$ are generated conditionally on the value of $U_{[i-1]}^k$, and decoding is performed taking into account $U_{[i-1]}^{[K]}$. Using an extension of Markov lemma ([22, Lemma 12.1], [3, Lemma 5]), one can show that $U_{[i-1]}^K$ is jointly typical with $\bar{X}_{[i-1]}^{[K]}$. Writing out the bound on the sum rate (similar to [22, Sec. 12.1.2]), one arrives at (46). □

C. MMSE estimation lemmas

We record several elementary estimation lemmas.

Lemma 2. *Let $X \sim \mathcal{N}\left(0, \sigma_X^2 \mathsf{I}\right)$, and let*

$$Y_k = X + W_k, \; k = 1, \ldots, K, \tag{47}$$

where $W_k \sim \mathcal{N}\left(0, \sigma_{W_k}^2 \mathsf{I}\right)$, $W_k \perp W_j$, $j \neq k$. Then, the MMSE estimate and the normalized estimation error of X given $Y_{[K]}$ are given by

$$\mathbb{E}\left[X | Y_{[K]}\right] = \sum_{k=1}^{K} \frac{\sigma_{X|Y_{[K]}}^2}{\sigma_{W_k}^2} Y_k, \tag{48}$$

$$\frac{1}{\sigma_{X|Y_{[K]}}^2} = \frac{1}{\sigma_X^2} + \sum_{k=1}^{K} \frac{1}{\sigma_{W_k}^2}. \tag{49}$$

Proof. The result is well known; for completeness, we provide a proof in Appendix B. □

Remark 1. We may use Lemma 2 to derive the Kalman filter for the estimation of X_i (1) given the history of observations $Y_{[i]} = Y_{[i]}^{[K]}$ (2):

$$\bar{X}_i = a\bar{X}_{i-1} + \sum_{k=1}^{K} \frac{\sigma_{X_i|Y_{[i]}}^2}{\sigma_{W_k}^2}\left(Y_i^k - a\bar{X}_{i-1}\right), \tag{50}$$

$$\frac{1}{\sigma_{X_i|Y_{[i]}^{[K]}}^2} = \frac{1}{\sigma_{X_i|Y_{[i-1]}^{[K]}}^2} + \sum_{k=1}^{K} \frac{1}{\sigma_{W_k}^2}. \tag{51}$$

where $\bar{X}_i$ was defined in (14). Equation (50) is nothing more than the Kalman filter recursion with Kalman filter gain equal to the row vector $\sigma_{X_i|Y_{[i]}}^2 \left(\frac{1}{\sigma_{W_1}^2}, \ldots, \frac{1}{\sigma_{W_K}^2}\right)$, and (51) is the corresponding Riccati recursion for the MSE.

The following result records a simple corollary to Lemma 2.

Lemma 3. *Let $X \sim \mathcal{N}\left(0, \sigma_X^2 \mathsf{I}\right)$, $W \sim \mathcal{N}\left(0, \sigma_W^2 \mathsf{I}\right)$, $W \perp X$, and let*

$$Y = X + W. \tag{52}$$

Then,

$$\sigma_{X|Y}^2 = \sigma_X^2\left(1 - \frac{\sigma_X^2}{\sigma_Y^2}\right) \tag{53}$$

Proof. Equality (53) follows from

$$\sigma_Y^2 = \sigma_X^2 + \sigma_W^2, \tag{54}$$

$$\frac{1}{\sigma_{X|Y}^2} = \frac{1}{\sigma_X^2} + \frac{1}{\sigma_W^2}, \tag{55}$$

where (55) is a particularization of (49). □

Remark 2. Dropping the assumptions of Gaussianity but keeping those of uncorrelatedness in Lemmas 2–3, relations (49) and (53) continue to hold replacing the normalized conditional variances $\sigma_{X|Y_{[K]}}^2$ and $\sigma_{X|Y}^2$ with the MMSEs achieved by the optimal *linear* estimator.

The following result is also a corollary to Lemma 2.

Lemma 4. *Let $\bar{X}_k$ and W'_k be Gaussian random variables, $\{\bar{X}_k\}_{k=1}^K \perp \{W'_j\}_{j=1}^K$, such that $W_k \perp W_j$, $j \neq k$, and*

$$X = \bar{X}_k + W'_k. \tag{56}$$

Then, the MMSE estimate and the estimation error $\sigma_{W'}^2 = \sigma_{X|\bar{X}_{[K]}}^2$ of X given the vector $\bar{X}_{[K]}$ are given by

$$\mathbb{E}\left[X | \bar{X}_{[K]}\right] = \sum_{k=1}^{K} \frac{\sigma_{W'}^2}{\sigma_{W'_k}^2} \bar{X}_k, \tag{57}$$

$$\frac{1}{\sigma_{W'}^2} = \sum_{k=1}^{K} \frac{1}{\sigma_{W'_k}^2} - \frac{K-1}{\sigma_X^2} \tag{58}$$

Proof. Notice that (56) with $\bar{X}_k = \mathbb{E}[X|Y_k]$ and $W_k' \sim \mathcal{N}(0, \sigma^2_{X|Y_k})$ is just another way to write (47). Reparameterizing (48) and (49) accordingly, one recovers (57) and (58). ☐

Lemma 4 converts the "forward channels" from X to observations Y_k (47) into the "backward channels" from estimates $\bar{X}_k$ to X (56). While both representations are equivalent, (56) is often easier to write down and to work with. Backward channel representations find widespread use in rate-distortion theory [24].

D. Proof of Theorem 1: converse

The proof is inspired by the converse technique developed by Wang et al. [9] for the noncausal Gaussian CEO problem. We also use the tools developed in [16].

Starting with the data processing converse in Theorem 5, we write

$$
\begin{aligned}
&R_{\text{CEO}}(d)\\
&\geq \inf_{\substack{(41):\\(42)\text{ holds}}} \underline{I}\left(\bar{\mathsf{X}}^{[K]}_{[t]} \to \mathsf{U}^{[K]}_{[t]}\right) && (59)\\
&= \inf_{\substack{(41):\\(42)\text{ holds}}} \underline{I}((\mathsf{X}, \bar{\mathsf{X}}^{[K]}) \to \mathsf{U}^{[K]}) && (60)\\
&= \inf_{\substack{(41):\\(42)\text{ holds}}} \left\{\underline{I}(\mathsf{X} \to \mathsf{U}^{[K]}) + \underline{I}(\bar{\mathsf{X}}^{[K]} \to \mathsf{U}^{[K]} \| \mathsf{X})\right\} && (61)\\
&= \inf_{\substack{(41):\\(42)\text{ holds}}} \left\{\underline{I}(\mathsf{X} \to \mathsf{U}^{[K]}) + \sum_{k=1}^{K} \underline{I}(\bar{\mathsf{X}}^k \to \mathsf{U}^k \| \mathsf{X})\right\} && (62)\\
&\geq \inf_{\{\rho_k\}_{k=1}^K \in \mathcal{R}} \left\{\frac{1}{2}\log\frac{\sigma^2_{\mathsf{X}\|\mathcal{D}\mathsf{U}^{[K]}}}{d} + \sum_{k=1}^{K}\frac{1}{2}\log\frac{\sigma^2_{\bar{\mathsf{X}}^k\|\mathsf{X},\mathcal{D}\mathsf{U}^k}}{\rho_k}\right\}, && (63)
\end{aligned}
$$

where

- (59) is by Theorem 5;
- (60) is due to the chain rule of directed information (35), and $I(\mathsf{X}_{[t]} \to \mathsf{U}^{[K]}_{[t]} \| \bar{\mathsf{X}}^{[K]}_{[t]}) = 0$;
- (61) is by the chain rule of directed information (35);
- (62) is due to (41);
- in (63), we denoted
$$\rho_k \triangleq \sigma^2_{\bar{\mathsf{X}}^k\|\mathsf{X},\mathsf{U}^k}; \tag{64}$$
(63) bounds each directed information in (62) by the corresponding rate-distortion function: the first term is the point-to-point causal Gaussian rate-distortion function [11, eq. (1.43)], the remaining K terms are the causal rate-distortion functions for processes $\bar{\mathsf{X}}^k$ with causal Gaussian side information [16, Th. 7]
$$\mathsf{X}_i = \bar{\mathsf{X}}^k_i + \mathsf{W}^{k\,\prime}_i, \tag{65}$$
where $\mathsf{W}^{k\,\prime}_i$ are independent, $\mathsf{W}^{k\,\prime}_i \sim \mathcal{N}\left(0, \sigma^2_{\mathsf{X}_i|\mathsf{Y}^k_{[i]}}\right)$, and $\mathsf{W}^{k\,\prime}_i \perp \bar{\mathsf{X}}^k_i$.

Note that $\sigma^2_{\mathsf{X}\|\mathcal{D}\mathsf{U}^{[K]}}$ is uniquely determined by d via (16), and $\sigma^2_{\bar{\mathsf{X}}^k\|\mathsf{X},\mathcal{D}\mathsf{U}^k}$ is similarly uniquely determined by ρ_k (see (82) and (84) below). Thus (63) reduces the minimization over causal kernels (41) to that over scalar parameters $\rho_1, \ldots, \rho_K$. We proceed to establish a connection between d and $\rho_1, \ldots, \rho_K$, that is, to identify the optimization region $\mathcal{R} \subseteq \mathbb{R}^K_+$ in (63).

Invoking Lemma 4 with $X \leftarrow \mathsf{X}_i$, $\bar{X}_k \leftarrow \bar{\mathsf{X}}^k_i$, $W_k' \leftarrow \mathsf{W}^{k\,\prime}_i$, we conclude

$$
\begin{aligned}
\bar{\mathsf{X}}_i &\triangleq \mathbb{E}\left[\mathsf{X}_i|\mathsf{Y}^{[K]}_{[i]}\right] && (66)\\
&= \sum_{k=1}^{K}\frac{\sigma^2_{\mathsf{X}_i\|\mathsf{Y}^{[K]}_{[i]}}}{\sigma^2_{\mathsf{X}_i\|\mathsf{Y}^k_{[i]}}}\bar{\mathsf{X}}^k_i, && (67)
\end{aligned}
$$

which implies in particular

$$
\begin{aligned}
\mathbb{E}\left[\bar{\mathsf{X}}_i|\mathsf{X}_{[i]}, \mathsf{U}^K_{[i]}\right] &= \sum_{k=1}^{K}\frac{\sigma^2_{\mathsf{X}\|\mathsf{Y}^{[K]}}}{\sigma^2_{\mathsf{X}\|\mathsf{Y}^k}}\mathbb{E}\left[\bar{\mathsf{X}}^k_i|\mathsf{X}_{[i]}, \mathsf{U}^K_{[i]}\right] && (68)\\
&= \sum_{k=1}^{K}\frac{\sigma^2_{\mathsf{X}\|\mathsf{Y}^{[K]}}}{\sigma^2_{\mathsf{X}\|\mathsf{Y}^k}}\mathbb{E}\left[\bar{\mathsf{X}}^k_i|\mathsf{X}_{[i]}, \mathsf{U}^k_{[i]}\right]. && (69)
\end{aligned}
$$

It follows that steady-state causal MMSE in estimating $\bar{\mathsf{X}}_i$ from $\mathsf{X}_{[i]}$ and $\mathsf{U}^{[K]}_{[i]}$ satisfies

$$\sigma^2_{\bar{\mathsf{X}}\|\mathsf{X},\mathsf{U}^{[K]}} = \sum_{k=1}^{K}\frac{\sigma^4_{\mathsf{X}\|\mathsf{Y}}}{\sigma^4_{\mathsf{X}\|\mathsf{Y}^k}}\rho_k. \tag{70}$$

Denote

$$\hat{\mathsf{X}}_i \triangleq \mathbb{E}\left[\mathsf{X}_i|\mathsf{U}^{[K]}_{[i]}\right]. \tag{71}$$

Observe that

$$
\begin{aligned}
\sigma^2_{\bar{\mathsf{X}}_i|\mathsf{X}_{[i]},\mathsf{U}^{[K]}_{[i]}} &= \sigma^2_{\bar{\mathsf{X}}_i - \mathbb{E}\left[\bar{\mathsf{X}}_i|\mathsf{X}_{[i]},\mathsf{U}^{[K]}_{[i]}\right]} && (72)\\
&= \sigma^2_{\bar{\mathsf{X}}_i - \mathsf{X}_i - \mathbb{E}\left[\bar{\mathsf{X}}_i - \mathsf{X}_i|\mathsf{X}_{[i]},\mathsf{U}^{[K]}_{[i]}\right]} && (73)\\
&\leq \sigma^2_{\bar{\mathsf{X}}_i - \mathsf{X}_i - \mathbb{E}\left[\bar{\mathsf{X}}_i - \mathsf{X}_i|\mathsf{X}_i - \hat{\mathsf{X}}_i\right]} && (74)\\
&= \sigma^2_{\mathsf{X}_i - \bar{\mathsf{X}}_i|\mathsf{X}_i - \hat{\mathsf{X}}_i} && (75)\\
&\leq \tilde{\sigma}^2_{\mathsf{X}_i - \bar{\mathsf{X}}_i|\mathsf{X}_i - \hat{\mathsf{X}}_i}, && (76)
\end{aligned}
$$

where $\tilde{\sigma}_{X|Y}$ denotes the MMSE achievable in estimation of X given Y, where the estimator is constrained to be a linear function of Y. Now, we apply Lemma 3 with $X \leftarrow \mathsf{X}_i - \bar{\mathsf{X}}_i$, $Y \leftarrow \mathsf{X}_i - \hat{\mathsf{X}}_i$, $W \leftarrow \bar{\mathsf{X}}_i - \hat{\mathsf{X}}_i$ (see Remark 2; the assumption $W \perp X$ is easily verified directly) to establish

$$\lim_{i\to\infty}\tilde{\sigma}^2_{\mathsf{X}_i - \bar{\mathsf{X}}_i|\mathsf{X}_i - \hat{\mathsf{X}}_i} = \sigma^2_{\mathsf{X}\|\mathsf{Y}}\left(1 - \frac{\sigma^2_{\mathsf{X}\|\mathsf{Y}}}{d}\right), \tag{77}$$

which, together with (70) and (76), means

$$\frac{1}{d} \leq \frac{1}{\sigma^2_{\mathsf{X}\|\mathsf{Y}}} - \sum_{k=1}^{K}\frac{\rho_k}{\sigma^4_{\mathsf{X}\|\mathsf{Y}_k}}. \tag{78}$$

Also, note that

$$0 \leq \rho_k \leq \sigma^2_{\bar{\mathsf{X}}^k\|\mathsf{X}}. \tag{79}$$

We conclude that

$$\mathcal{R} = \{\rho_k,\ k = 1, \ldots, K\colon\ (78) \text{ and } (79) \text{ are satisfied}\}. \tag{80}$$

It remains to clarify how the form in (18), (19), (20), parameterized in terms of

$$d_k \triangleq \sigma^2_{\mathsf{X}\|\mathsf{U}^k} \tag{81}$$

rather than ρ_k, is obtained. An application of Lemma 3 with $X \leftrightarrow \mathsf{X}_i - \bar{\mathsf{X}}_i^k$, $Y \leftrightarrow \mathsf{X}_i - \hat{\mathsf{X}}_i^k$, $W \leftrightarrow \bar{\mathsf{X}}_i^k - \hat{\mathsf{X}}_i^k$ leads to

$$\rho_k = \sigma^2_{\mathsf{X}\|\mathsf{Y}^k}\left(1 - \frac{\sigma^2_{\mathsf{X}\|\mathsf{Y}^k}}{d_k}\right). \tag{82}$$

Plugging (82) into (78) leads to (19). Applying Lemma 3 with $X \leftrightarrow \mathsf{X}_i - \bar{\mathsf{X}}_i^k$, $Y \leftrightarrow \mathsf{X}_i$, $W \leftrightarrow \bar{\mathsf{X}}_i^k$, we express

$$\sigma^2_{\bar{\mathsf{X}}^k\|\mathsf{X}} = \sigma^2_{\mathsf{X}\|\mathsf{Y}^k}\left(1 - \frac{\sigma^2_{\mathsf{X}\|\mathsf{Y}^k}}{\sigma^2_{\mathsf{X}}}\right), \tag{83}$$

which, together with (82), implies the equivalence of (79) and (20). Finally, applying Lemma 3 with $X \leftrightarrow \mathsf{X}_i - \bar{\mathsf{X}}_i^k$, $Y \leftrightarrow \mathsf{X}_i - a\hat{\mathsf{X}}_{i-1}^k$, $W \leftrightarrow \bar{\mathsf{X}}_i^k - a\hat{\mathsf{X}}_{i-1}^k$, we express

$$\sigma^2_{\bar{\mathsf{X}}^k\|\mathsf{X},\mathcal{D}\mathsf{U}^k} = \sigma^2_{\mathsf{X}\|\mathsf{Y}^k}\left(1 - \frac{\sigma^2_{\mathsf{X}\|\mathsf{Y}^k}}{\sigma^2_{\mathsf{X}\|\mathcal{D}\mathsf{U}^k}}\right). \tag{84}$$

Plugging (82) and (84) into (63), we conclude the equivalence of (63) and (18). ☐

E. Proof of Theorem 1: achievability

With the bound in Theorem 6, it suffices to prove the existence of K causal kernels in (40) such that (42) holds and equality in (63) is achieved. Let

$$\mathsf{U}_i^k = \bar{\mathsf{X}}_i^k + \mathsf{Z}_i^k, \quad k = 1, \ldots, K, \tag{85}$$

where $\mathsf{Z}_i \sim \mathcal{N}\left(0, \sigma^2_{\mathsf{Z}_k}\right)$, satisfy both conditions for equality in [16, Th. 2] (for the first term in (63)) and [16, Th. 3] (for the K terms in the sum in (63)). This means that (85) achieves equality in (63). To check that the resulting ρ_k fall within $\mathcal{R}$ in (80), note that (79) is satisfied trivially, while (78) is attained with equality because in the Gaussian case (85), equality holds in (74) and (76). ☐

F. A suboptimal waterfilling solution

First, we present an upper bound to $R_{\text{CEO}}(d)$ obtained by waterfilling over d_k's. While optimal in the classical Gaussian CEO problem [5], such waterfilling is only suboptimal in the causal Gaussian CEO problem. This is unsurprising as a similar dichotomy is observed between causal and noncausal Gaussian point-to-point rate-distortion functions [14, Remark 2].

Theorem 7. *The causal sum rate - distortion function in the causal Gaussian CEO problem (1), (2) is bounded by*

$$R_{\text{CEO}}(d) \le \frac{1}{2}\log\frac{\sigma^2_{X\|\mathcal{D}U^{[K]}}}{d} + \frac{1}{2}\sum_{k=1}^K \left|\log\left(\frac{1}{\sigma^2_{X\|Y^k}} - \frac{1}{\sigma^2_{X\|\mathcal{D}U^k}}\right)\lambda\right|_+, \tag{86}$$

where λ is the solution to

$$\frac{1}{\lambda}\sum_{k=1}^K 1\left\{\frac{1}{\lambda} < \frac{1}{\sigma^2_{X\|Y^k}} - \frac{1}{\sigma^2_{X\|\mathcal{D}U^k}}\right\} + \sum_{k=1}^K\left(\frac{1}{\sigma^2_{X\|Y^k}} - \frac{1}{\sigma^2_X}\right)1\left\{\frac{1}{\lambda} \ge \frac{1}{\sigma^2_{X\|Y^k}} - \frac{1}{\sigma^2_{X\|\mathcal{D}U^k}}\right\} = \frac{1}{\sigma^2_{X\|Y}} - \frac{1}{d}, \tag{87}$$

with equality if all $\sigma^2_{X\|Y^k}$ are equal.

Proof. Consider the following choice of ρ_k in (78):

$$\rho_k = \min\left\{\frac{\sigma^4_{X\|Y_k}}{\lambda},\ \sigma^2_{\bar{X}^k\|X}\right\}. \tag{88}$$

This choice is feasible because with λ in (87) and ρ_k in (88), equality in (78) is attained (see (83)). Furthermore, (63) evaluates to (86) (using (84)).

To show equality in the symmetrical case, it suffices to show that in that case, the infimum in (63) is attained by $\rho_1 = \ldots = \rho_K$. Writing $\bar{\mathsf{X}}_i^k$ in a manner similar to (104) in Appendix A below, and using (55) and (65), we may express

$$\frac{1}{\sigma^2_{\bar{\mathsf{X}}^k\|\mathsf{X},\mathcal{D}\mathsf{U}^k}} = \frac{1}{\sigma^2_{X\|Y^k}} + \frac{1}{a^2\rho_k + \sigma^2_{X\|\mathcal{D}Y^k} - \sigma^2_{X\|Y^k}}. \tag{89}$$

The sum in (63) can then be bounded as (we replace each of $\sigma^2_{X\|Y^k}$ by $\sigma^2_{X\|Y^1}$ to emphasize that they are all the same)

$$\frac{1}{2}\sum_{k=1}^K \log\frac{\left(\frac{1}{\sigma^2_{X\|Y^1}} + \frac{1}{a^2\rho_k + \sigma^2_{X\|\mathcal{D}Y^1} - \sigma^2_{X\|Y^1}}\right)^{-1}}{\rho_k} \ge \frac{K}{2}\log\frac{\left(\frac{1}{\sigma^2_{X\|Y^1}} + \frac{1}{a^2\frac{1}{K}\sum_{k=1}^K\rho_k + \sigma^2_{X\|\mathcal{D}Y^1} - \sigma^2_{X\|Y^1}}\right)^{-1}}{\frac{1}{K}\sum_{k=1}^K \rho_k}, \tag{90}$$

where the inequality follows from the convexity in ρ_k of each summand. ☐

G. Proof of Theorem 3

Using Theorem 2 (or Theorem 1) and (62), we can compute

$$R_{\text{noisy}}(d) - R(d) = \min_{\substack{P_{\mathsf{U}^{[K]}\|\bar{\mathsf{X}},\,\mathcal{D}\mathsf{U}^{[K]}}:\\ (42)\text{ holds}}} \underline{I}(\bar{\mathsf{X}}^{[K]} \to \mathsf{U}^{[K]}\|\mathsf{X}) \tag{91}$$

$$= \frac{1}{2}\log\frac{\frac{1}{\sigma^2_{X\|Y^{[K]}}} - \frac{1}{\sigma^2_{X\|\mathcal{D}U^{[K]}}}}{\frac{1}{\sigma^2_{X\|Y^{[K]}}} - \frac{1}{d}}. \tag{92}$$

On the other hand, using Theorem 7,

$$R_{\text{CEO}}(d) - R(d)$$

$$= \inf_{\substack{(41):\\(42)\text{ holds}}} \sum_{k=1}^K I(\bar{\mathsf{X}}^k \to \mathsf{U}^k\|\mathsf{X}) \tag{93}$$

$$\le \frac{1}{2}\sum_{k=1}^K\left|\log\left(\frac{1}{\sigma^2_{X\|Y^k}} - \frac{1}{\sigma^2_{X\|\mathcal{D}U^k}}\right)\lambda\right|_+ \tag{94}$$

$$= \frac{1}{2}\sum_{k=1}^K \log\left(\frac{1}{\sigma^2_{X\|Y^k}} - \frac{1}{\sigma^2_{X\|\mathcal{D}U^k}}\right)\lambda \tag{95}$$

$$\le \frac{K}{2}\log\sum_{k=1}^K\left(\frac{1}{\sigma^2_{X\|Y^k}} - \frac{1}{\sigma^2_{X\|\mathcal{D}U^k}}\right)\frac{\lambda}{K} \tag{96}$$

$$= \frac{K}{2}\log\left(\frac{1}{\sigma^2_{X\|Y^{[K]}}} - \frac{1}{\sigma^2_{X\|\mathcal{D}U^{[K]}}}\right)\frac{\lambda}{K}, \tag{97}$$

where

- (93) is due to (62);
- (94) is by Theorem 7;
- (95) holds because with (25), the waterfilling solution in Theorem 7 results in all active transmitters; note that in that case, (87) reduces to

$$\lambda = K\left(\frac{1}{\sigma^2_{X\|Y^{[K]}}} - \frac{1}{d}\right)^{-1}; \tag{98}$$

- (96) is by Jensen's inequality, since $\log$ is concave;
- (97) is due to

$$\frac{1}{\sigma^2_{X\|Y^{[K]}}} = \sum_{k=1}^{K}\frac{1}{\sigma^2_{X\|Y^k}} - \frac{K-1}{\sigma^2_X}, \tag{99}$$

$$\frac{1}{\sigma^2_{X\|\mathcal{D}U^{[K]}}} = \sum_{k=1}^{K}\frac{1}{\sigma^2_{X\|\mathcal{D}U^k}} - \frac{K-1}{\sigma^2_X}, \tag{100}$$

which holds by Lemma 4 even if the source is nonstationary (that is, $|a| \geq 1$ and $\sigma^2_X = \infty$), as a simple limiting argument taking $\frac{K-1}{\sigma^2_X}$ to 0 confirms.

Combining (92), (97) and (98), one obtains (26). To establish condition for equality, note that '$=$' holds in (94) in the symmetrical case by Theorem 7, and that '$=$' holds in (96) only in the symmetrical case due to the strict concavity of the $\log$. □

H. Proof of Theorem 4

Plugging

$$\frac{1}{\sigma^2_{X\|Y^{[K]}}} = \frac{K}{\sigma^2_{X\|Y^1}} - \frac{K-1}{\sigma^2_X}, \tag{101}$$

which particularizes (99) to the symmetrical case, into (28), and eliminating d_1 from (27), one readily verifies that

$$R^{K-\mathrm{sym}}_{\mathrm{CEO}}(d) - R(d) = \frac{1}{2}\frac{\frac{1}{d} - \frac{1}{\sigma^2_{X\|\mathcal{D}U^{[K]}}}}{\frac{1}{\sigma^2_{X\|Y^1}} - \frac{1}{\sigma^2_X}} + O\left(\frac{1}{K}\right), \tag{102}$$

and (29) follows. □

Appendix A
Two Equivalent Representations of $R_{\mathrm{noisy}}(d)$

In this appendix, we verify that (22) coincides with the lower bound on the causal noisy rate-distortion function derived in [14]. Indeed, [14, Cor. 1 and Th. 9] imply

$$R_{\mathrm{noisy}}(d) \geq \frac{1}{2}\log\left(a^2 + \frac{\sigma^2_{X\|\mathcal{D}Y^{[K]}} - \sigma^2_{X\|Y^{[K]}}}{d - \sigma^2_{X\|Y^{[K]}}}\right). \tag{103}$$

Here, $\sigma^2_{X\|\mathcal{D}Y^{[K]}} - \sigma^2_{X\|Y^{[K]}}$ is the variance of the innovations of the Gauss-Markov process $\{\bar{X}_i\}$, i.e.

$$\bar{X}_{i+1} = a\bar{X}_i + \bar{W}_i, \tag{104}$$

$\bar{W}_i \sim \mathcal{N}(0, \sigma^2_{X\|\mathcal{D}Y^{[K]}} - \sigma^2_{X\|Y^{[K]}})$. The form in (103) leads to that in (22) via (23) and

$$\sigma^2_{X\|\mathcal{D}Y^{[K]}} = a^2\sigma^2_{X\|Y^{[K]}} + \sigma^2_V. \tag{105}$$

□

Appendix B
Proof of Lemma 2

For jointly Gaussian random vectors X, Y,

$$\mathbb{E}\left[X|Y=y\right] = \mathbb{E}\left[X\right] + \Sigma_{XY}\Sigma_{YY}^{-1}\left(y - \mathbb{E}\left[Y\right]\right), \tag{106}$$

$$\mathrm{Cov}[X|Y] = \Sigma_{XX} - \Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX}. \tag{107}$$

Denote for brevity

$$\Sigma_W \triangleq \begin{bmatrix} \sigma^2_{W_1} & & 0 \\ & \ddots & \\ 0 & & \sigma^2_{W_K} \end{bmatrix} \tag{108}$$

For notational simplicity, we consider only the case $n = 1$: X is a scalar and $Y = Y_{[K]}$ is a vector, and

$$\Sigma_X = \sigma^2_X \tag{109}$$

$$\Sigma_{YY} = \Sigma_W + \begin{bmatrix}1\\ \vdots \\ 1\end{bmatrix}\sigma^2_X\begin{bmatrix}1 & \cdots & 1\end{bmatrix} \tag{110}$$

$$\Sigma_{XY} = \sigma^2_X\begin{bmatrix}1 & \ldots & 1\end{bmatrix} \tag{111}$$

Using the matrix inversion lemma, we compute readily

$$\begin{aligned}\mathrm{Cov}[X|Y]^{-1} &= \Sigma_{XX}^{-1} - \Sigma_{XX}^{-1}\Sigma_{XY} \\ &\quad \left(\Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY} - \Sigma_{YY}\right)^{-1}\Sigma_{YX}\Sigma_{XX}^{-1} \qquad (112)\\ &= \Sigma_{XX}^{-1} + \Sigma_{XX}^{-1}\Sigma_{XY}\Sigma_W^{-1}\Sigma_{YX}\Sigma_{XX}^{-1} \qquad (113)\\ &= \frac{1}{\sigma^2_X} + \frac{1}{\sigma^2_{W_1}} + \ldots + \frac{1}{\sigma^2_{W_K}}, \qquad (114)\end{aligned}$$

which shows (49). To show (48), we apply the matrix inversion lemma to Σ_{YY} to write:

$$\Sigma_{YY}^{-1} = \Sigma_W^{-1} - \Sigma_W^{-1}\begin{bmatrix}1\\ \vdots \\ 1\end{bmatrix}\sigma^2_{X|Y_{[K]}}\begin{bmatrix}1 & \ldots & 1\end{bmatrix}\Sigma_W^{-1}$$

It's easy to verify that

$$\sigma^2_X\begin{bmatrix}1 & \ldots & 1\end{bmatrix}\left(\mathrm{I}\frac{1}{\sigma^2_{X|Y_{[K]}}} - \Sigma_W^{-1}\begin{bmatrix}1\\ \vdots \\ 1\end{bmatrix}\begin{bmatrix}1 & \ldots & 1\end{bmatrix}\right)$$

$$= \begin{bmatrix}1 & \ldots & 1\end{bmatrix}, \tag{115}$$

so

$$\mathbb{E}\left[X|Y=y\right] = \Sigma_{XY}\Sigma_{YY}^{-1}y \tag{116}$$

$$= \begin{bmatrix}1 & \ldots & 1\end{bmatrix}\Sigma_W^{-1}\sigma^2_{X|Y_{[K]}}y, \tag{117}$$

which is equivalent to (48). □

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] 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.

[10] 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.

[11] A. Gorbunov and M. S. Pinsker, "Prognostic epsilon entropy of a Gaussian message and a Gaussian source," *Problemy Peredachi Informatsii*, vol. 10, no. 2, pp. 5–25, 1974.

[12] 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.

[13] 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.

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

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

[16] 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.

[17] 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.

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

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

[20] 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.

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

[22] A. El Gamal and Y.-H. Kim, *Network information theory*. Cambridge university press, 2011.

[23] D. P. Bertsekas, *Dynamic programming and optimal control*. Athena Scientific Belmont, MA, 1995, vol. 1.

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