Bounds on the Benefits of Interaction in Distributed Source Coding for Function Computation

Abdellatif Zaidi † ‡ Mohamed Ali Moussa † Roy Timo ƚ

† Université Paris-Est Marne La Vallée, Champs-sur-Marne 77454, France
‡ Mathematical and Algorithmic Sciences Lab., France Research Center, Huawei Technologies
ƚ Technische Universität München, Germany
{abdellatif.zaidi@u-pem.fr, mohamed-ali.moussa@u-pem.fr, roy.timo@tum.de}

Abstract— **In this paper, we study a setting in which two terminals A and B respectively observe, or measure, two memoryless, possibly statistically dependent, sources X and Y; and they interact bidirectionally in the aim of computing, at terminal B, a function $f_B(X,Y)$ of the two sources. Essentially, we establish upper bounds on the maximum gain that can be brought up by the interaction, in terms of minimum sum rate improvement for a given average distortion. In particular, we show that this gain is bounded by the redundancy of the one-message minimal rate for computing the function $f_B(X,Y)$ in the case in which Terminal A does not know the side information Y over the one-message minimal rate for computing the same function with the same tolerance but with Terminal A informed about the side information Y. That is, the redundancy of the one-message Wyner-Ziv rate-distortion function for function computation over the one-message conditional rate-distortion function for function computation. In the special case of lossy source reproduction, the bound reduces to the rate loss of the Wyner-Ziv problem as studied by Zamir in the case of a difference distortion measure. In the case of lossy function computation, we use this bound to establish an alternate bound that is generally easier to compute. Furthermore, we also apply the results to some important special cases, thus allowing us to gain some fundamental insights on the benefits of the interaction for both lossless and lossy function computations in these cases.**

I. Introduction

In many realworld distributed systems, such as data centers, peer-to-peer networks, and sensor networks, the task of the network is to compute a function, which can then utilized to make a decision or coordinate some action. The traditional approach to perform computations in networks consists in transferring the required data that is generated, or measured, by the nodes to a single decision-making node (e.g., fusion center) which has the task of performing the desired computations. This technique has the advantage of being rather easy to implement and analyze, but is suboptimal in general. Additional gains, that are substantial in some cases, can be made possible if the nodes are allowed to interact, i.e., the information can flow from sources towards destinations and back possibly over multiple rounds.

The benefits of interaction for two-way source coding are studied in [1] for lossy source reproduction, and in [2] for function computation. Important advances on the role of nodes' interaction for function computation have been made recently by Ma and Ishwar, for a two-terminal problem in [3], [4] and for larger collocated networks in [5].

In this work, we consider a setting in which two terminals A and B respectively observe, or measure, two memoryless sources $\mathbf{X}$ and $\mathbf{Y}$; and they interact bidirectionally in the aim of computing, at terminal B, a component-wise function $f_B^{(n)}(\mathbf{X},\mathbf{Y}) = (f_B(X_1,Y_1),\dots,f_B(X_n,Y_n))$ of the two sources. The sources are assumed to be possibly statistically dependent, through a memoryless joint distribution $P_{X,Y}(x,y)$. The computation is to be performed in a lossless or lossy manner, depending on the scenario. The system model is shown in Figure 1. We measure the performance in terms of the minimum number of bits per source sample to be exchanged in t bidirectional rounds in order to compute the function $f_B^{(n)}(\mathbf{X},\mathbf{Y})$ to within some prescribed distortion D_B. That is, the minimum sum-rate $R_{\text{sum,t}} = R_1+\dots+R_t$, where R_i, $i=1,\dots,t$, is the rate, in bits per source sample, in round i. For this model, Ma and Ishwar have shown that the interaction is useless in terms of the minimum sum-rate if the goal is lossless source reproduction, whereas it is useful for both lossy source reproduction and general function computation [3], [4].

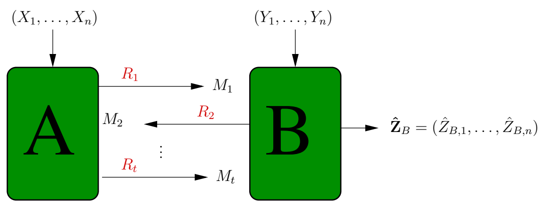

Fig. 1. Interactive distributed source coding with t alternating messages for the computation of function $f_B^{(n)}(\mathbf{X},\mathbf{Y})$ at terminal B.

Investigating closely the role of the interaction, it can be seen that, intuitively, in the cases in which the interaction is useful the improvement is enabled precisely by the fact that each node furnishes to the other node a (partial) description of the observed source, which can then be utilized to progressively learn and/or refine some information about the desired function computation. Based on this observation, we first study the setting in which Terminal A also knows the source $\mathbf{Y}$; and we show that, as expected in this case, the interaction is completely useless in terms of the minimum sum rate that is required to compute the function $f_B^{(n)}(\mathbf{X},\mathbf{Y})$ to within some prescribed average distortion D_B. That is, the one-message rate is optimal in this case. This result, which can also be inferred from Kaspi's results in [1] for the cases of lossless and lossy source reproduction, is shown here directly to also hold in the cases of lossless and lossy function computation.

Next, turning to the case in which Terminal A does *not* know the side information $\mathbf{Y}$, we establish an upper bound on the maximum gain that can be brought up by the interaction in this case, in terms of the maximum sum rate improvement for the lossy computation of function $f_B^{(n)}(\mathbf{X},\mathbf{Y})$. We express this bound in terms of quantities that do not depend on the interaction and can be computed easily in many cases. Specifically, the maximum gain of the interaction in terms of minimum sum rate is shown to be bounded by the redundancy of the one-message minimal rate for computing the function $f_B^{(n)}(\mathbf{X},\mathbf{Y})$ in the case in which Terminal A does not know the side information $\mathbf{Y}$ over the one-message minimal rate for computing the same function with the same tolerance but with Terminal A informed about the side information $\mathbf{Y}$. That is, the redundancy of the one-message Wyner-Ziv rate-distortion function for function computation over the one-message conditional rate-distortion function for function computation. In the special case of *lossy source reproduction*, the bound reduces to the rate loss of the Wyner-Ziv problem. This rate loss is studied by Zamir in [6] in the case of a difference distortion measure; and, so, the advantage of the interaction in terms of minimum sum rate can be bounded by the *minimax capacity* bound of [6] in this case. An important implication of this is in showing that the loss due to the lack of the interaction, in terms of minimum sum rate, is bounded.

Furthermore, we also apply the results to some important special cases, thus allowing us to gain some fundamental insights on the benefits of the interaction for both lossless and lossy function computations in these cases. In some cases, such as those of lossless source reproduction, lossy reproduction of Gaussian sources as well as lossy reproduction of binary sources with erased side information, the bound is zero, thus establishing that the interaction is completely useless. In other cases, such as lossy sources reproduction (general sources, not necessarily Gaussian), as well as lossless and lossy function computation, the approach provides computable bounds on the maximum gain that can be hoped for with infinite number of interaction rounds. For example, we obtain that the ultimate gain that can be enabled by the interaction (in terms of minimum sum rate) cannot exceed 0.22 bits per sample for *any* binary sources with Hamming distance, and 1/2 bit per sample for any continuous sources with squared error distortion.

II. Problem Setting and Definitions

Consider the two-terminal interactive distributed source coding problem with alternating messages shown in Figure 1. The sources $\mathbf{X} = (X_1,\ldots,X_n)$ and $\mathbf{Y} = (Y_1,\ldots,Y_n)$ are two statistically dependent discrete memoryless stationary sources with elements taking values in finite alphabets $\mathcal{X}$ and $\mathcal{Y}$, respectively, with joint probability mass function (PMF) $P_{X,Y}(x,y)$ that captures the statistical dependencies among their samples. Terminal A observes the source $\mathbf{X} = (X_1,\ldots,X_n)$ and does not compute any function; and Terminal B observes the source $\mathbf{Y} = (Y_1,\ldots,Y_n)$ and wants to compute a sample-wise function $\mathbf{Z}_B = f_B^{(n)}(\mathbf{X},\mathbf{Y}) = (f_B(X_1,Y_1),\ldots,f_B(X_n,Y_n))$ of the sources $\mathbf{X}$ and $\mathbf{Y}$ to within some prescribed average distortion level D_B. That is, with $f_B : \mathcal{X}\times\mathcal{Y} \longrightarrow \mathcal{Z}_B$ denoting the function of interest at Terminal B, the desired outputs at this location is the vector $\hat{\mathbf{Z}}_B = (\hat{Z}_{B,1},\ldots,\hat{Z}_{B,n})$ such that $\mathbb{E}[d_B^{(n)}(\mathbf{X},\mathbf{Y},\hat{\mathbf{Z}}_B)] \leq D_B + \epsilon$, where

$$d_B^{(n)}(\mathbf{X},\mathbf{Y},\hat{\mathbf{Z}}_B) \triangleq \frac{1}{n}\sum_{i=1}^{n} d_B(X_i,Y_i,\hat{Z}_{B,i}) \tag{1}$$

with, for $i = 1,\ldots,n$, $Z_i = f_B(X_i,Y_i)$, $\hat{Z}_{B,i}$ is the estimate of the i-th element of the vector $\mathbf{Z}_B$ and $d_B : \mathcal{X}\times\mathcal{Y}\times\mathcal{Z}_B \longrightarrow \mathbb{R}^+$ is some general single-letter distortion function (note that this measure can depend arbitrarily on X and Y). To achieve the goal of computing the desired function at location B, the two terminals exchange coded messages, alternately.

***Definition** 1:* A two-terminal interactive distributed source code for function computation with initial location A and parameters $(t,n,|\mathcal{M}_1|,\ldots,|\mathcal{M}_t|)$ is the tuple $(\phi_1,\ldots,\phi_t,\psi_B)$ of t block encoding functions $\phi_1,\ldots,\phi_t$ and one block decoding function ψ_B, of blocklength n, such that for $j = 1,\ldots,t$, we have

$$\text{Enc. round } j \;\; \phi_j : \begin{cases} \mathcal{X}^n\times\mathcal{M}_1\times\ldots\times\mathcal{M}_{j-1} \to \mathcal{M}_j, & \text{if } j \text{ is odd} \\ \mathcal{Y}^n\times\mathcal{M}_1\times\ldots\times\mathcal{M}_{j-1} \to \mathcal{M}_j, & \text{if } j \text{ is even} \end{cases} \tag{2a}$$

$$\text{Dec. at Terminal B} \quad \psi_B : \mathcal{Y}^n\times\mathcal{M}_1\times\ldots\times\mathcal{M}_t \longrightarrow \mathcal{Z}_B^n \tag{2b}$$

The output of the encoding function ϕ_j in round j is called the j-th message M_j, and t is the number of rounds or exchanged messages. For $j = 1,\ldots,t$, the coding rate in round j, in bits per source sample, is given by $R_j = (1/n)\log_2(|\mathcal{M}_j|)$.

***Definition** 2:* A rate-distortion tuple $(\mathbf{R},D_B) = (R_1,\ldots,R_t,D_B)$ is admissible for t-message interactive function computation with initial location A if for any $\epsilon > 0$, there exist a sufficiently large n and an interactive distributed source code with initial location A and parameters $(t,n,|\mathcal{M}_1|,\ldots,|\mathcal{M}_t|)$ satisfying

$$\frac{1}{n}\log_2|\mathcal{M}_j| \leq R_j + \epsilon,\; j = 1,\ldots,t \tag{3}$$

$$\mathbb{E}[d_B^{(n)}(\mathbf{X},\mathbf{Y},\hat{\mathbf{Z}}_B)] \leq D_B + \epsilon. \tag{4}$$

The set of all admissible rate-distortion tuples is called the rate-distortion region for t-message interactive computation of function $f_B(\cdot)$ with initial location A, and is denoted by $\mathcal{RD}_t^{\mathrm{A}}(f_B)$. For a given distortion D_B, the set of all admissible rate t-tuples $\mathbf{R} = (R_1,\ldots,R_t)$ such that $(\mathbf{R},D_B) \in \mathcal{RD}_t^{\mathrm{A}}(f_B)$ be denoted by $\mathcal{R}_t^{\mathrm{A}}(f_B,D_B)$. The sum-rate-distortion function $R_{\mathrm{sum},t}^{\mathrm{A}}(f_B,D_B)$ is given by the minimization, over all t-tuples $(R_1,\ldots,R_t)$ of $\mathcal{R}_t^{\mathrm{A}}(f_B,D_B)$, of the sum $(R_1+\ldots+R_t)$, i.e.,

$$R_{\mathrm{sum},t}^{\mathrm{A}}(f_B,D_B) = \min_{\mathbf{R}\in\mathcal{R}_t^{\mathrm{A}}(f_B,D_B)} \sum_{i=1}^{t} R_i. \tag{5}$$

If Terminal B initiates the procedure of messages exchange, the definitions are similar to the above. Finally, in the limit as t goes to infinity, both $R_{\mathrm{sum},t}^{\mathrm{A}}(f_B,D_B)$ and $R_{\mathrm{sum},t}^{\mathrm{B}}(f_B,D_B)$

converge to the infinite-message sum-rate-distortion function $R_{\text{sum},\infty}(f_B, D_B)$ [3].

Due to space limitation, the results of this paper are either outlined only or mentioned without proofs. Detailed proofs for the model of this paper can be found in [7].

In the settings in which the interaction is useful, the two-message interactive source code with initial location B *strictly* outperforms the one-message code with initial location A, i.e., $R^{\text{B}}_{\text{sum},2}(f_B, D_B) < R^{\text{A}}_{\text{sum},1}(f_B, D_B)$. In such cases, the intuition for the reason for which the interaction helps, in terms of the minimum required sum rate to compute the desired function, is that in the first round Terminal B provides to Terminal A a partial description of the source $\mathbf{Y}$ that is observes, which is then utilized by this terminal to get a sense of the function to be computed at the other side. This may suggest that had Terminal A known the side information $\mathbf{Y}$, then the interaction would be completely useless. Based on this observation, in the next section we study the setting, shown in Figure 2, in which terminal A also knows the side information $\mathbf{Y}$; and we show that, as expected in this case, the interaction is useless in terms of the minimum sum rate that is required to compute the function $\mathbf{Z}_B = f_B^{(n)}(\mathbf{X}, \mathbf{Y})$ to within some distortion D_B.

III. Case in Which Terminal A Knows Side Information Y^n

Consider the model shown in Figure 2. Here, Terminal A knows both memoryless sources $\mathbf{X}$ and $\mathbf{Y}$, and does not compute any function; and Terminal B knows the memoryless source $\mathbf{Y}$ and wants to compute the function $\mathbf{Z}_B = f_B^{(n)}(\mathbf{X}, \mathbf{Y})$ to within some prescribed distortion D_B.

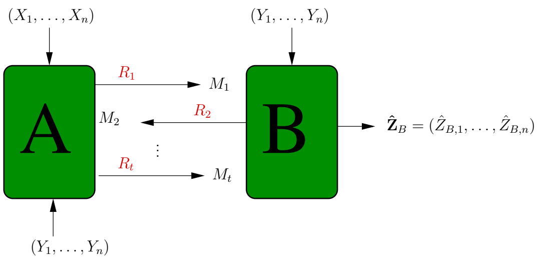

Fig. 2. Interactive distributed source coding with t alternating messages. Both terminals know the side information Y^n.

For a given average distortion D_B, the minimum sum-rate for a t-message distributed code for the setup shown in Figure 2 can be obtained by applying a slight modification of the results of [3] to setups with encoder side information, by considering that Terminal A observes the source pair $\tilde{X} = (X, Y)$ and Terminal B observes the source Y and wishes to compute the function $\tilde{f}_B(\tilde{X}, Y) = f_B(X, Y)$, where the function $\tilde{f}_B : \mathcal{X}\times\mathcal{Y}\times\mathcal{Y} \longrightarrow \mathcal{Z}_B$ is defined such that $\forall (x, y) \in \mathcal{X}\times\mathcal{Y}$, $\tilde{f}_B(\tilde{X}, Y) = f_B(X, Y)$. Specifically, assuming for example that the first message is sent by Terminal A, the minimum sum rate of a t-message interactive distributed code is given by

$$\begin{aligned} R^{\text{A}}_{\text{sum},t}(f_B, D_B) &= \min_{U^t} \; [I(U^t; \tilde{X}|Y) + I(U^t; Y|\tilde{X})] \\ &= \min_{U^t} \; I(U^t; X|Y) \end{aligned} \tag{6}$$

where the second equality follows by substituting $\tilde{X} = (X, Y)$; and the minimization is taken under all $U^t = (U_1, \ldots, U_t)$ satisfying $U_i \leftrightarrow (Y, U^{i-1}) \leftrightarrow X$ forms a Markov chain for i even, and such that $\mathbb{E}[d_B(X, Y, \hat{g}_B(Y, U^t)] \leqslant D_B$ for some deterministic function $\hat{g}_B$.

The minimum sum-rate for a given average distortion D_B as given by the expression (6) does not indicate explicitly, however, whether or not the interaction is *strictly* useful for the model shown in Figure 2. In what follows, we show that it is not. Specifically, the interaction is completely useless for the setup shown in Figure 2 in the sense that $R_{\text{sum},\infty}(f_B, D_B) = R^{\text{A}}_{\text{sum},1}(f_B, D_B)$. We note that for the cases of lossy and lossless source reproduction this result can also be inferred from [1]. Furthermore, it will be shown that the smallest admissible rate that allows Terminal B to compute the desired function $\mathbf{Z}_B$ to within the prescribed average distortion D_B with just *one iteration*, i.e., $R^{\text{A}}_{\text{sum},1}(f_B, D_B)$, is given by the conditional rate distortion function $R_{X|Y}(f_B, D_B)$ for computing the function f_B. The following theorem states the result.

***Theorem** 1:* For the model shown in Figure 2 with the memoryless source Y known to both Terminals A and B, the interaction is useless in terms of minimum required sum-rate for computing the function $f_B(X, Y) = Z_B$ at Terminal B with average distortion D_B, i.e.,

$$R_{\text{sum},\infty}(f_B, D_B) = R_{\text{sum},1}(f_B, D_B). \tag{7}$$

Moreover, $R_{\text{sum},\infty}(f_B, D_B)$ is given by

$$R_{\text{sum},\infty}(f_B, D_B) = \min_{P_{\hat{Z}_B|X,Y},\, \mathbb{E}[d_B(X,Y,\hat{Z}_B)] \leq D_B} I(\hat{Z}_B; X|Y), \tag{8}$$

where the minimization is over all conditional $P_{\hat{Z}_B|X,Y}$ such that $\mathbb{E}[d_B(X, Y, \hat{Z}_B)] \leq D_B$.

IV. Bounds on the Benefit of Interaction

Let us now turn to the model shown in Figure 1, i.e., with the side information $\mathbf{Y}$ known only at Terminal B. We denote by $R^{\text{WZ}}_{X|Y}(f_B, D_B)$ the smallest admissible *one-message* rate that allows Terminal B to compute the desired function $\mathbf{Z}_B = f_B^{(n)}(\mathbf{X}, \mathbf{Y})$ to within some average distortion D_B. That is, $R^{\text{WZ}}_{X|Y}(f_B, D_B)$ is the Wyner-Ziv rate that is required to compute the desired function with distortion D_B for the same setup but *without* interaction. In the next section, we show that the improvement that can be brought up by the interaction for the model of Figure 1, in terms of minimum required sum-rate, cannot exceed the difference $(R^{\text{WZ}}_{X|Y}(f_B, D_B) - R_{X|Y}(f_B, D_B))$.

A. Bound on the Ultimate Minimum Sum-Rate

Consider the model shown in Figure 1. For a given average distortion D_B, $R_{\text{sum},t}(f_B, D_B)$ is a non-increasing function of t; and the ultimate gain that can be allowed by the interaction, in terms of minimum required sum-rate, is given by $R_{\text{sum},1}(f_B, D_B) - R_{\text{sum},\infty}(f_B, D_B)$. The following theorem provides an upper bound on this gain.

***Theorem** 2:* For the model shown in Figure 1, we have

$$0 \leq R_{\text{sum},1}(f_B, D_B) - R_{\text{sum},\infty}(f_B, D_B) \leq R^{\text{WZ}}_{X|Y}(f_B, D_B) - R_{X|Y}(f_B, D_B). \tag{9}$$

Remark 1: The bound (9) is also valid for any finite t. That is, for given t, the improvement, in terms of minimum sum-rate for a given average distortion D_B, of any t-message code over the one-message code, or the case of no interaction, is such that

$$0 \le R_{\mathrm{sum},1}(f_B, D_B) - R_{\mathrm{sum},t}(f_B, D_B) \le R^{WZ}_{X|Y}(f_B, D_B) - R_{X|Y}(f_B, D_B). \tag{10}$$

The bound of Theorem 2 can be applied to provide interesting insights on the role of interaction and its ultimate benefits, in terms of minimum sum rate, in many lossless and lossy interactive distributed source coding settings, as we shall see in the following two sections.

B. Application to Lossless Function Computation

In [4], Ma and Ishwar have shown that the interaction is useless in terms of the minimum sum-rate if the goal is pure source-reproduction. This result can also be obtained from Theorem 2 by setting $D_B = 0$ and $f_B(X, Y) = X$. In this case, the RHS of (9) is zero since $R^{WZ}_{X|Y}(X, 0) = R_{X|Y}(X, 0) = H(X|Y)$ by Slepian-Wolf coding.

The interaction is beneficial, however, for general lossless function computation as shown through some striking examples in [4]. Evaluating the bound of Theorem 2 with the choice $D_B = 0$, we obtain a bound on the ultimate benefit of the interaction in this case. Specifically, we have $R_{X|Y}(f_B, 0) = H(Z_B|Y)$; and the term $R^{WZ}_{X|Y}(f_B, 0)$ in the RHS of (9) is given, in this case, by the conditional graph entropy $H_{\mathcal{G}}(X|Y)$ of X given Y, where $\mathcal{G}$ is the *characteristic graph* of the triple (X, Y, f_B), i.e., the graph $\mathcal{G}$ is such that the set of nodes of this graph is the support of X and there is an edge between two distinct nodes x and x' iff there exists a symbol y such that $p(x, y), p(x', y) > 0$ and $f_B(x, y) \neq f_B(x', y)$. That is, with $\Gamma(\mathcal{G})$ denoting the collection of maximally independent sets of the graph $\mathcal{G}$ [8], we have

$$H_{\mathcal{G}}(X|Y) = \min \; I(W; X|Y) \tag{11}$$

where the minimization is over all conditional pmfs $p_{W|X}(w|x)$ such that $W \in \Gamma(\mathcal{G})$ and $p_{W|X}(w|x) = 0$ if $x \notin w$.

Summarizing, in the case of lossless function computation, we get

$$R_{\mathrm{sum},1}(f_B, 0) - R_{\mathrm{sum},t}(f_B, 0) \le H_{\mathcal{G}}(X|Y) - H(Z_B|Y). \tag{12}$$

For many examples, the RHS of (12) can be computed easily.

***Example** 1:* Let $X \sim$ Uniform$\{1, \ldots, L\}$ and $Y \sim$ Bern(p), $p \in [0, 1]$, such that X and Y are independent. Also, let $f_B(X, Y) = X{\cdot}Y$ (real-valued product). This is an expanded version of [9, Example 8] that was also considered in [4] to show that the interaction is beneficial in this case. For this example, it is easy to see that the characteristic graph $\mathcal{G}$ is *complete*, and so $H_{\mathcal{G}}(X|Y) = H(X|Y) = \log_2(L)$ bits per sample. Also, we have $H(X{\cdot}Y|Y) = p\log_2(L)$ bits per sample. Thus, using (12) we get that the improvement due to the interaction, in terms of minimum sum-rate, in this case is bounded from above by $(1-p)\log_2(L)$ bits per sample. (The application of (12) also yields that the two-message code of [4] is *at most* at $h_2(p)$ bit per sample from the ultimate minimum sum rate $R_{\mathrm{sum},\infty}(f_B, 0)$). □

C. Application to Lossy Source Reproduction

The case of lossy reproduction of the source X at Terminal B can be obtained as a special case by setting $f_B(X, Y) = X$. In the case, the bound of Theorem 2 reduces to

$$0 \le R_{\mathrm{sum},1}(X, D_B) - R_{\mathrm{sum},\infty}(X, D_B) \le R^{WZ}_{X|Y}(X, D_B) - R_{X|Y}(X, D_B), \tag{13}$$

where $R^{WZ}_{X|Y}(X, D_B)$ is the rate-distortion function for coding the memoryless source X with side information Y available at the decoder (i.e., the Wyner-Ziv rate as established by Wyner and Ziv in [10]) and $R_{X|Y}(X, D_B)$ is the conditional rate distortion function. Thus, Theorem 2 bounds the benefit in terms of minimum sum rate that can be brought up by the interaction in this case by the rate loss in the Wyner-Ziv problem (w.r.t. to the conditional rate distortion problem).

For some cases, the RHS of (13) is zero and, so, the above means that the interaction is completely useless in terms of minimum sum rate in these cases (see below for some important examples). In other cases, the RHS of (13) gives a computable bound on the rate improvement that can be brought up by the interaction.

***Example** 2 (Binary sources with erased side information):* Let $X \sim$ Bern$(1/2)$, Y be an erased version of X, i.e., $P_Y(y = x) = 1 - P_Y(y = \epsilon) = p$ and the distortion measure is a Hamming distance, i.e., $d_B(x, \hat{x}) = x \oplus \hat{x}$, where $\oplus$ denotes the addition modulo-2. For this example, the Wyner-Ziv rate-distortion function $R^{WZ}_{X|Y}(X, D_B)$ and the conditional rate-distortion function $R_{X|Y}(X, D_B)$ coincide [11] and are given by

$$R^{WZ}_{X|Y}(X, D_B) = R_{X|Y}(X, D_B) = \begin{cases} p(1 - h_2(\frac{D_B}{p})) & \text{if} \quad D_B \leqslant \frac{p}{2} \\ 0 & \text{otherwise.} \end{cases} \tag{14}$$

Thus, the interaction is not useful in terms of minimum sum-rate for a given distortion D_B in this case; and, the optimal rate $R_{\mathrm{sum},\infty}(X, D_B) = R^{WZ}_{X|Y}(X, D_B)$ as given by the RHS of (14). □

***Example** 3 (Gaussian sources with quadratic distortion measure):* Let (X, Y) be a memoryless jointly Gaussian pair of sources with, say, $\mathbb{E}[X] = \mathbb{E}[Y] = 0$, $\mathbb{E}[X^2] = \mathbb{E}[Y^2] = P$ and correlation coefficient $\rho = \mathbb{E}[XY]/P$. Also, let the squared error distortion measure $d_B(x, \hat{x}) = (x - \hat{x})^2$. For this example, the Wyner-Ziv rate $R^{WZ}_{X|Y}(X, D_B)$ is the same as the rate when the side information Y is available at both the encoder and the decoder. Thus, the interaction is useless and one round Wyner-Ziv coding is optimal, i.e.,

$$R_{\mathrm{sum},\infty}(X, D_B) = \begin{cases} \frac{1}{2}\log_2\left(\frac{(1-\rho^2)P}{D_B}\right) & \text{if} \quad 0 \le (1-\rho^2)P \le D_B \\ 0 & \text{otherwise.} \end{cases} \tag{15}$$

The result on the non-utility of the interaction for this example is also stated in [8, Example 20.4]; but it is proved here differently, i.e., as a special case of a larger class of models for which the interaction does not help in terms of the minimum sum-rate. □

***Example** 4 (Doubly Symmetric Binary Sources):* Let (X, Y) be a DSBS(p), $p \in [0, 1/2]$, and $d_B(\cdot)$ the Hamming distortion measure. For this example, for $D_B \neq 0$ the interaction is beneficial if $p \neq 1/2$. The Wyner-Ziv rate-distortion function is given by

$$R^{WZ}_{X|Y}(X, D_B) = \begin{cases} g(D_B) & \text{for } 0 \le D_B \le D'_c \\ (p - D_B)g'(D'_c) & \text{for } D'_c \le D_B \le p \\ 0 & \text{for } D_B > p \end{cases} \tag{16}$$

where $g(D_B) = h_2(p * D_B) - h_2(D_B)$, g' denotes the derivative of g, D'_c is the solution of the equation $g(D'_c)/(p - D'_c) = g'(D'_c)$. Also, if the side information is available at both terminals the rate-distortion function is

$$R_{X|Y}(X, D_B) = \begin{cases} h_2(p) - h_2(D_B) & \text{for } 0 \leq D_B \leq p \\ 0 & \text{for } D_B > p. \end{cases} \tag{17}$$

Thus, the bound (13) computed for this example is given by the redundancy of the RHS of (16) over that of (17). □

The Minimax Capacity Bound

If the distortion measure $d_B(\cdot)$ is a difference distortion measure, i.e., there exists a measure $d(\cdot)$ such that $d_B(x, \hat{x}) = d(x - \hat{x})$ for all $(x, \hat{x}) \in \mathcal{X}^2$, it is shown in [6, Theorem 1] that

$$R^{WZ}_{X|Y}(X, D_B) - R_{X|Y}(X, D_B) \leq C_{\mathcal{X}}(D_B) \tag{18}$$

where $C_{\mathcal{X}}(D_B)$ is defined as

$$C_{\mathcal{X}}(D_B) = \inf_{N \in \mathcal{X}\,:\, \mathbb{E}[d(N)] \leq D_B} C(D_B, N), \tag{19}$$

and $C(D_B, N)$ denotes the capacity of the additive noise channel $W \longrightarrow W + N$ under the distortion measure $d_B(\cdot)$, with W and N independent, i.e.,

$$C(D_B, N) = \sup_{W \in \mathcal{X}\,:\, \mathbb{E}[d(W)] \leq D_B} I(W; W + N). \tag{20}$$

Combining equations (13) and (18) we get the following alternate bound on the gain in terms of minimum sum-rate that can be brought up by the interaction in the case of lossy source reproduction,

$$R_{\text{sum},1}(X, D_B) - R_{\text{sum},t}(X, D_B) \leqslant C_{\mathcal{X}}(D_B). \tag{21}$$

The main importance of the minimax capacity bound (21) is in showing that, like for the case of no interaction [6], the rate loss of the uninformed encoder is also bounded in the case of interactive coding. Another appreciable property of the bound (21) is that it does not depend on the actual distributions of the sources, but only on their alphabets. Also, it computes easily. For example, for the DSBS(1/3) sources of Example 4 with Hamming distortion measure, we have

$$C_{\mathcal{X}}(D_B) = h_2(D_B * D_B) - h_2(D_B). \tag{22}$$

Then, since $C_{\mathcal{X}}(D_B) \leq 0.22$ bit/sample for $0 \leq D_B \leq 1/2$, this implies that, in the case of lossy source reproduction, the ultimate gain that can be enabled by the interaction (in terms of minimum sum rate) cannot exceed 0.22 bits/sample for *any* binary sources with Hamming distance. Similarly, using (21) we obtain that the benefit of the interaction in terms of minimum sum rate cannot exceed 1/2 bit/sample for continuous sources with squared error distortion. Although this bound is not tight for Gaussian sources (see Example 3), its importance is that it applies in general irrespective to the joint distribution of the continuous pair (X, Y).

V. Lossy Function Computation

Consider again the model of Figure 1. We assume here that the goal is lossy computation at Terminal B of a general function that can depend on *both* sources, i.e., not necessarily lossy reproduction of the source X. In this case, the minimax capacity bounding approach of the last section, which is applicable in the case of lossy source reproduction, does not extend easily to the setup of lossy function computation. In what follows, we establish a bound on the bound of Theorem 2 that holds under some conditions.

***Proposition** 1:* For lossy function computation, the gain of the interaction in terms of minimum sum-rate for a given average distortion D_B is bounded as

$$R_{\text{sum},1}(f_B, D_B) - R_{\text{sum},\infty}(f_B, D_B) \leq \min \left\{ R^{WZ}_{X|Y}\left(X, \Psi^{-1}(D_B)\right) \right\} - R_{X|Y}(f_B, D_B), \tag{23}$$

where the minimization is over strictly increasing functions $\Psi(\cdot)$ for which there exists some measure $d(\cdot)$ such that

$$\mathbb{E}[d_B(X, Y, \hat{Z}_B)] \leq \Psi\left(\mathbb{E}[d(X, Y, \hat{X})]\right). \tag{24}$$

Remark 2: The condition (24) holds in many cases. For example, for binary sources with the Hamming distortion measure it is easily observed that

$$\mathbb{E}[d_H(X, Y, \hat{Z})] \leq \mathbb{E}[d_H(X, Y, \hat{X})]. \tag{25}$$

***Example** 5 (Lossy computation of XOR function of DSBS):* Let (X, Y) be a DSBS(p), $p \in [0, 1/2]$, and the distortion measure $d_B(\cdot)$ be such that $d_B(x, y, \hat{z}) = d_H(x \oplus y, \hat{z})$, where $d_H(\cdot)$ is the Hamming distortion measure. Assume that Terminal B wishes to compute $Z_B = X \oplus Y$ to within some average distortion D_B. The conditional rate-distortion function to convey $Z_B = X \oplus Y$ to Terminal B can be shown easily to be given by

$$R_{X|Y}(f_B, D_B) = \min_{\left\{P_{\hat{Z}_B|Y,Z_B},\, \mathbb{E}[d_H(Z_B, \hat{Z}_B)] \leq D_B\right\}} I(\hat{Z}_B; Z_B|Y) \tag{26}$$

Thus, using (25) of Remark 2, and observing that $(Z_B, Y) \sim$ DSBS(1/2) here, the bound (23) for this example yields

$$R_{\text{sum},1}(f_B, D_B) - R_{\text{sum},\infty}(f_B, D_B) \leq R^{WZ}_{X|Y}(X, D_B) - R_{Z_B|Y}(Z_B, D_B) \tag{27}$$

which can be computed as the redundancy of the RHS of (16) over that of (17). □

References

[1] A. H. Kaspi, "Two-way source coding with a fidelity criterion," *IEEE Trans. Inf. Theory*, vol. 31, pp. 735–740, 1985.

[2] A. Orlisky, "Worst case interactive communication ii: Two messages are not optimal," *IEEE Trans. Inf. Theory*, vol. 37, pp. 995–1005, 1991.

[3] N. Ma and P. Ishwar, "The infinite message limit of two terminal interactive source coding," *IEEE Trans. Inf. Theory*, vol. 59, pp. 4071–4094, Jul. 2013.

[4] ——, "Some results on distributed source coding for interactive function computation," *IEEE Trans. Inf. Theory*, vol. 57, pp. 6180–6195, 2011.

[5] ——, "Interactive source coding for interactive function computation in collocated networks," *IEEE Trans. Inf. Theory*, vol. 58, pp. 4289–4305, 2012.

[6] R. Zamir, "The rate loss in the wyner-ziv problem," *IEEE Trans. Inf. Theory*, vol. 42, pp. 2073–2084, 1996.

[7] A. Zaidi, M.-A. Moussa, and R. Timo, "Two terminal interactive source coding for function computation with remote sources," *Draft. Available at http://www-syscom.univ-mlv.fr/~zaidi/publications/draft-isc.pdf*, 2015.

[8] A. E. Gamal and Y.-H. Kim, *Network information theory*. Cambridge University Press, 2011.

[9] A. Orlisky and J. Roche, "Coding for computing," *IEEE Trans. Inf. Theory*, pp. 903–917, 2001.

[10] 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, pp. 1–10, Jan. 1976.

[11] S. Diggavi, E. Perron, and E. Telatar, "Lossy source coding with gaussian or erased side-information," in *Proc. IEEE Int. Symp. Information Theory*, 2009, pp. 1035–1039.