Distributed Source Coding Via Successive Refinement*

Jun Chen Toby Berger

School of Electrical and Computer Engineering
Cornell University
Ithaca, NY 14853
jc353@cornell.edu
berger@ece.cornell.edu

Abstract

We show that any point in the Berger-Tung rate region can be achieved via successive Wyner-Ziv coding. We generalize the concept of successive refinement from single source coding to multisource coding, where we refer to it as distributed successive refinement. The quadratic Gaussian CEO problem is used as an example to illustrate some interesting aspects of distributed successive refinement.

I. Introduction

The problem of distributed source coding has assumed renewed interest in recent years. Many practical compression schemes have been proposed for Slepian-Wolf coding (e.g., [1] and the references therein) and Wyner-Ziv coding (e.g., [2] and the references therein), whose performances are close to the fundamental theoretical bounds [3, 4]. Therefore, it is of interest to reduce the general distributed source coding problem to these well-studied cases.

Given M i.i.d. discrete sources $X_1, X_2, \cdots, X_M$, the Slepian-Wolf rate region is the union of all the rate vectors $(R_1, R_2, \cdots, R_M)$ satisfying

$$\sum_{i \in \mathcal{A}} R_i \geq H(X_{\mathcal{A}} | X_{\mathcal{I}_M \setminus \mathcal{A}}) \quad \forall \mathcal{A} \in \mathcal{I}_M,$$

where $\mathcal{I}_M = \{1, 2, \cdots, M\}$ and $X_{\mathcal{A}} = \{X_i\}_{i \in \mathcal{A}}$. The Slepian-Wolf reigon is a contra-polymatroid with $M!$ vertices, where π is a permutation of $\mathcal{I}_M$, the coordinates of vertex $\mathcal{R}_\pi = (R_1, \cdots, R_M)$ are given by

$$\begin{aligned} R_{\pi(1)} &= H(X_{\pi(1)}), \\ R_{\pi(i)} &= H(H_{\pi(i)} | X_{\pi(1)}, \cdots, X_{\pi(i-1)}) \quad i = 2, \cdots, M. \end{aligned}$$

It is known that vertices of the Slepian-Wolf region can be achieved with a complexity significantly lower than that of a general point. It was observed in [5] that, by splitting a source into two virtual sources, one can reduce the problem of coding an arbitrary

*This research is supported in part by NSF Grant No. CCR-9980616, ARO P-40116-PH-MUR, DAAD 19-99-1-0215.

point in a M-dimensional Slepian-Wolf region to that of coding a vertex of a $(2M-1)$-dimensional Slepian-Wolf region. This source-splitting idea has a dual in the problem of coding for multiple access channels, which is referred as rate-splitting [6, 7]. We shall show that a similar conclusion also holds in the general distributed lossy source coding. Our approach is based on successive refinement instead of source splitting, since it does not require the existence of the common randomness at the encoders and decoder (used in [5][7] to generate a time sharing variable).

In Section II we prove that the Berger-Tung coding scheme can be reduced to a low complexity successive Wyner-Ziv coding scheme. The duality between superposition coding for multiaccess communication and successive Wyner-Ziv coding is briefly discussed. In Section III we introduce a concept called distributed successive refinement and then apply it to the quadratic Gaussian CEO problem. We conclude the paper in Section IV.

II. Main Result

A general achievable rate region in distributed source coding is the Berger-Tung region[8, 9]. For different models, the Berger-Tung region has different forms. In this paper we use the model of the CEO problem[10], though many of our results also hold for other models. The following is a brief description of the CEO problem.

$\{X(t)\}_{t=1}^{\infty}$ is the target data sequence that is of interest to the decoder. This data sequence cannot be observed directly. M encoders are deployed, which observe independently corrupted versions of $\{X(t)\}_{t=1}^{\infty}$. The data rate at which encoder i ($i = 1, 2, \cdots, M$) may communicate information about its observations to the decoder is limited to R_i bits per second. The encoders are not permitted to communicate with each other; i.e., encoder i has to send data based solely on its own noisy observations $\{Y_i(t)\}_{t=1}^{\infty}$. Finally, the decision $\{\hat{X}(t)\}_{t=1}^{\infty}$ is computed from the combined data at the decoder so that a desired fidelity can be satisfied.

Definition 1. An M-tuple of rates $(R_1, \cdots, R_M)$ is said to be D-achievable if for any $\epsilon > 0$, there exists an n_0 such that for $n > n_0$ there exist encoders:

$$f_i^{(n)} : \mathcal{Y}_i^n \to \{1, 2, \cdots, 2^{n(R_i+\epsilon)}\} \quad i = 1, \cdots, M$$

and a decoder:

$$g^{(n)} : \{1, 2, \cdots, 2^{n(R_i+\epsilon)}\} \times \cdots \times \{1, 2, \cdots, 2^{n(R_M+\epsilon)}\} \to \mathcal{X}^n$$

for which $\frac{1}{n}E[\sum_{t=1}^{n} d(X(t), \hat{X}(t))] \leq D + \epsilon$, where $\hat{X}^n = g^{(n)}(f_1^{(n)}(Y_1^n), \cdots, f_M^{(n)}(Y_M^n))$ and $d(\cdot,\cdot) : \mathcal{X} \times \mathcal{X} \to [0, d_{\max}]$ is a given distortion measure. We let $\mathcal{R}(D)$ denote the set of all D-achievable rate tuples.

Definition 2. Let

$$\mathcal{R}(W_1, \cdots, W_M) = \left\{(R_1, \cdots, R_M) : \sum_{i\in\mathcal{A}} R_i \geq I\left(Y_{\mathcal{A}}; W_{\mathcal{A}} | W_{\mathcal{I}_M \setminus \mathcal{A}}\right), \forall \mathcal{A} \in \mathcal{I}_M\right\}.$$

The Berger-Tung rate region with respect to distortion D is

$$\mathcal{R}_{BT}(D) = \text{conv}\left(\bigcup_{(W_1,\cdots,W_M)\in\mathcal{W}_D} \mathcal{R}(W_1, \cdots, W_M)\right),$$

where $\mathcal{W}_D$ is the set of all $(W_1, \cdots, W_M)$ satisfying the following properties:

(i) $W_i \to Y_i \to \left(X, Y_{\mathcal{I}_M\setminus\{i\}}, W_{\mathcal{I}_M\setminus\{i\}}\right) \quad \forall i \in \mathcal{I}_M$.

(ii) There exists a function:

$$f: \mathcal{W}_1 \times \cdots \times \mathcal{W}_M \to \mathcal{X}$$

such that $Ed(X, \hat{X}) \leq D$, where $\hat{X} = f(W_1, \cdots, W_M)$.

It was shown in [8, 9] that $\mathcal{R}_{BT}(D) \subseteq \mathcal{R}(D)$. Computing the Berger-Tung rate region involves complicated optimization and convexification. Hence we shall only focus on $\mathcal{R}(W_1, \cdots, W_M)$. We will see later that for the quadratic Gaussian case, the properties of Berger-Tung rate region are determined completely by those of $\mathcal{R}(W_1, \cdots, W_M)$.

It was proved in [11, 12] that if $W_i \to Y_i \to \left(X, Y_{\mathcal{I}_M\setminus\{i\}}, W_{\mathcal{I}_M\setminus\{i\}}\right)$ for all $i \in \mathcal{I}_M$ (which we will assume throughout this paper) then $\mathcal{R}(W_1, \cdots, W_M)$ is a contra-polymatroid with $M!$ vertices, where π is a permutation on $\mathcal{I}_M$, the coordinates of vertex $\mathcal{R}_\pi = (R_1, \cdots, R_M)$ are given by

$$\begin{aligned} R_{\pi(1)} &= I(Y_{\pi(1)}; W_{\pi(1)}), \\ R_{\pi(i)} &= I(Y_{\pi(i)}; W_{\pi(i)} | W_{\pi(1)}, \cdots, W_{\pi(i-1)}) \quad i = 2, \cdots, M. \end{aligned}$$

The convex region bounded by these $M!$ vertices is called the dominant face of $\mathcal{R}(W_1, \cdots, W_M)$ on which we have $\sum_{i=1}^{M} R_i = I\left(Y_{\mathcal{I}_M}; W_{\mathcal{I}_M}\right)$. For each of these $M!$ vertices, there exists a low-complexity successive Wyner-Ziv coding scheme described as follows:

(i) Encoder $\pi(1)$ employs conventional lossy source coding. Encoder $\pi(i)$ $(i = 2, \cdots, M)$ employs Wyner-Ziv coding with $W_{\pi(1)}, \cdots, W_{\pi(i-1)}$ being the side information at the decoder.

(ii) The decoder first decodes the codeword $W_{\pi(1)}$ from encoder $\pi(1)$, then successively decodes the codeword $W_{\pi(i)} (i = 2, \cdots, M)$ from encoder $\pi(i)$ with side information $W_{\pi(1)}, \cdots, W_{\pi(i-1)}$, $2 \leq i \leq M$.

Draper and Wornell [13] investigated this type of successive coding strategies in the context of data fusion in tree-structured sensor networks.

Rate tuples on the dominant face other than these $M!$ vertices were previously known to be attainable only by one of two methods. The first method known to achieve these difficult rate tuples was time sharing among vertices. This approach can require as many as M distributed source codes, each distributed source codes requiring M individual codes (one conventional lossy source code and $M - 1$ Wyner-Ziv codes). Thus, this scheme requires on the order of M^2 individual codes. The second approach to achieve these rate tuples is joint decoding of all users. This is difficult to implement in practice, since random codes have a decoding complexity of the order of $2^{nI\left(Y_{\mathcal{I}_M}; W_{\mathcal{I}_M}\right)}$, where n is the block length.

One of our main results is that any rate tuple in $\mathcal{R}(W_1, \cdots, W_M)$ can be achieved by a low complexity successive Wyner-Ziv coding scheme with one conventional lossy source coding and at most $2M - 2$ Wyner-Ziv codes. Without loss of generality, we need to consider only rate tuples on the dominant face of $\mathcal{R}(W_1, \cdots, W_M)$.

Definition 3. A well-ordered permutation of $\{W_1, W_1' \cdots, W_M, W_M'\}$ is a permutation such that W_i appears after W_i' for all $i \in \mathcal{I}_M$.

Theorem 1. For any rate tuple $(R_1, \cdots, R_M)$ on the dominant face of $\mathcal{R}(W_1, \cdots, W_M)$, there exist $(W_1', \cdots, W_M')$ satisfying $W_i' \to W_i \to (X, Y_{\mathcal{I}_M}, W_{\mathcal{I}_M \setminus i}, W'_{\mathcal{I}_M \setminus i})$ and a well-ordered permutation σ of $\{W_1, W_1' \cdots, W_M, W_M'\}$ such that

$$R_i = I(Y_i; W_i' | \{W_i'\}^-) + I(Y_i; W_i | \{W_i\}^-) \quad \forall i \in \mathcal{I}_M,$$

where $\{W_i'\}^-$ denotes all the random variables before W_i' in the permutation σ and $\{W_i\}^-$ is similarly defined. Furthermore, there exists at least one $i \in \mathcal{I}_M$ such that $W_i = W_i'$.

Proof: Due to the limitation of space, we prove the theorem only for the case $M = 2$. Extension to the general case follows the approach similar to that in [6].

For any rate tuple (R_1, R_2) on the dominant face of $\mathcal{R}(W_1, W_2)$, we have $I(Y_1; W_1|W_2) \leq R_1 \leq I(Y_1; W_1)$, $I(Y_2; W_2|W_1) \leq R_2 \leq I(Y_2; W_2)$, $R_1 + R_2 = I(Y_1, Y_2; W_1, W_2)$. We can always find a W_1' with $W_1' \to W_1 \to (X, Y_1, Y_2, W_2)$ such that $R_2 = I(Y_2; W_2|W_1')$. Now

$$\begin{aligned} R_1 &= I(Y_1, Y_2; W_1, W_2) - R_2 \\ &= I(Y_1, Y_2; W_1, W_2) - I(Y_2; W_2|W_1') \\ &= I(Y_1; W_1') + I(Y_1; W_1|W_1', W_2). \end{aligned}$$

Let $\sigma = (W_1', W_2', W_2, W_1)$ and set $W_2' = W_2$. The proof is complete. ∎

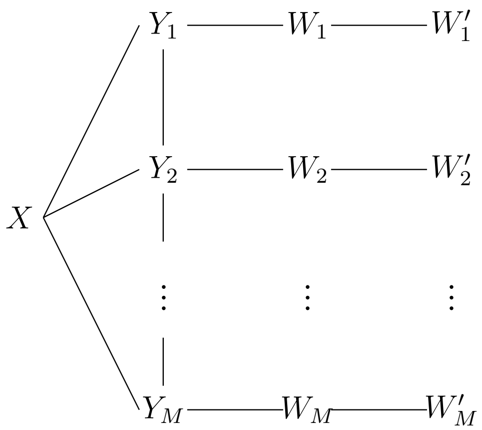

Figure 1: Successive Wyner-Ziv coding

Intuitively, we can view W_i' as a coarse quantization of Y_i and view W_i as a fine quantization of Y_i (see Fig.1). Theorem 1 suggests the following successive Wyner-Ziv coding scheme:

Encoder i first splits R_i into two pieces, $r_i^1 = I(Y_i; W_i'|\{W_i'\}^-)$ and $r_i^2 = I(Y_i; W_i|\{W_i\}^-)$. Then it uses a Wyner-Ziv code with rate r_i^1 to convey W_i' to the decoder which has side information $\{W_i'\}^-$ and uses Wyner-Ziv code with rate r_i^2 to convey the refinement W_i to decoder which has side information $\{W_i\}^-$. The decoder recovers the $\{W_1', W_1, \cdots, W_M', W_M\}$ successively in the same order as the permutation σ. We can see that this scheme requires one conventional lossy source code (reduced from a Wyner-Ziv code, since at first the decoder has no side information) and at most $2M - 2$ Wyner-Ziv codes.

This successive Wyner-Ziv coding scheme has a dual in the multiple access communication (see Fig.2).

Encoder i first generates a low rate code X_i' and then generates a high rate code X_i by superposition. Only high rate code X_i is transmitted. Decoder successively recovers

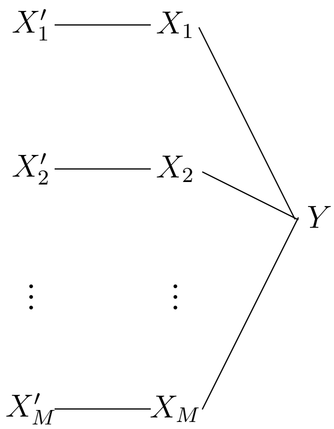

Figure 2: Multiaccess communication via superposition

the codewords according to a well-ordered permutation σ' of $(X_1', X_1, \cdots, X_M', X_M)$ such that

$$R_i = I(X_i'; Y|\{X_i'\}^-) + I(X_i; Y|\{X_i\}^-) \quad \forall i \in \mathcal{I}_M.$$

Our superposition scheme is similar to the rate-splitting scheme introduced in [6]. Actually in [6] the output of the splitting function is exactly a superposition code. The slight difference between these two schemes is that we don't impose any independence condition. Note:The independence condition is introduced in [6] to ensure that after rate-splitting, the multiaccess channel with M independent input data streams becomes a multiaccess channel with $2M-1$ independent input data streams. Since with the exception of one user, each of the other users can control 2 inputs, no independence condition is necessary for the 2 inputs controlled by one user. After removing this restriction, the rate-splitting scheme in [6] and our superposition scheme become equivalent.

III. Distributed Successive Refinement

Now we introduce a concept called distributed successive refinement which can be viewed as a generalization of the single source successive refinement [14] to the setting of the CEO problem.

Definition 4. $D^*(R_1, \cdots, R_M) = \min\{D : (R_1, \cdots, R_M) \in \mathcal{R}(D)\}$.

Definition 5. We say there exists a successive refinement scheme from rate tuple $(R_1, \cdots, R_M)$ to rate tuple $(R_1 + \triangle R_1, \cdots, R_M + \triangle R_M)$ if for any $\epsilon > 0$, there exists an n_0 such that for $n > n_0$ there exist encoders:

$$\begin{aligned} f_{1,i}^{(n)} &: \mathcal{Y}_i^n \to \{1, 2 \cdots, 2^{n(R_i+\epsilon)}\} \qquad i = 1, \cdots, M, \\ f_{2,i}^{(n)} &: \mathcal{Y}_i^n \to \{1, 2 \cdots, 2^{n(\triangle R_i+\epsilon)}\} \quad i = 1, \cdots, M, \end{aligned}$$

and decoders:

$$\begin{aligned} g_1^{(n)} &: \{1, 2 \cdots, 2^{n(R_1+\epsilon)}\} \times \cdots \times \{1, 2 \cdots, 2^{n(R_M+\epsilon)}\} \to \mathcal{X}^n, \\ g_2^{(n)} &: \{1, 2 \cdots, 2^{n(R_1+\epsilon)}\} \times \{1, 2 \cdots, 2^{n(\triangle R_1+\epsilon)}\} \times \cdots \times \{1, 2 \cdots, 2^{n(R_M+\epsilon)}\} \\ \times &\ \{1, 2 \cdots, 2^{n(\triangle R_M+\epsilon)}\} \to \mathcal{X}^n \end{aligned}$$

such that

$$
\begin{aligned}
\frac{1}{n}E[\sum_{t=1}^{n} d(X(t),\hat{X}_1(t))] &\leq D^*(R_1,\cdots,R_M)+\epsilon,\\
\frac{1}{n}E[\sum_{t=1}^{n} d(X(t),\hat{X}_2(t))] &\leq D^*(R_1+\triangle R_1,\cdots,R_M+\triangle R_M)+\epsilon,
\end{aligned}
$$

where

$$
\begin{aligned}
\hat{X}_1^n &= g_1^{(n)}(f_{1,1}^{(n)}(Y_1^n),\cdots,f_{1,M}^{(n)}(Y_M^n)),\\
\hat{X}_2^n &= g_2^{(n)}(f_{1,1}^{(n)}(Y_1^n),f_{2,1}^{(n)}(Y_1^n),\cdots,f_{1,M}^{(n)}(Y_M^n),f_{2,M}^{(n)}(Y_M^n)),
\end{aligned}
$$

and $\triangle R_i \geq 0$ $(i=1,\cdots,M)$.

Now we proceed to discuss distributed refinement for the quadratic Gaussian CEO problem. For simplicity, we only consider the case $M=2$; all our results can be extended to the general M case in a straightforward manner [15]. For the the quadratic Gaussian CEO problem [16], we have $Y_i = X+N_i$, where $X\sim\mathcal{N}(0,\sigma_X^2)$, $N_i\sim\mathcal{N}(0,\sigma_{N_i}^2)$ $(i=1,2)$ and X,N_1,N_2 are all independent. The squared error distortion measure is used. Without loss of generality, we suppose $\sigma_{N_1}^2\leq\sigma_{N_2}^2$. Let the auxiliary random variables $W_i=Y_i+T_i$, where $T_i\sim\mathcal{N}(0,\sigma_{T_i}^2)$ and X,N_1,N_2,T_1,T_2 are all independent. We can compute the Berger-Tung rate region explicitly as follows:

$$
\mathcal{R}_{BT}(D)=\bigcup_{(\sigma_{T_1}^2,\sigma_{T_2}^2)\in\Lambda(D)}\mathcal{R}_{BT}(\sigma_{T_1}^2,\sigma_{T_2}^2)
$$

where

$$
\Lambda(D) = \left\{(\sigma_{T_1}^2,\sigma_{T_2}^2):\frac{1}{D}\leq\frac{1}{\sigma_X^2}+\frac{1}{\sigma_{N_1}^2+\sigma_{T_1}^2}+\frac{1}{\sigma_{N_2}^2+\sigma_{T_2}^2}\right\},
$$

$$
\begin{aligned}
\mathcal{R}_{BT}(\sigma_{T_1}^2,\sigma_{T_2}^2)=\Bigg\{(R_1,R_2): R_1 &\geq \frac{1}{2}\log\frac{(\sigma_X^2+\sigma_{N_1}^2+\sigma_{T_1}^2)(\sigma_X^2+\sigma_{N_2}^2+\sigma_{T_2}^2)-\sigma_X^4}{\sigma_X^2\sigma_{T_1}^2+\sigma_{N_2}^2\sigma_{T_1}^2+\sigma_{T_1}^2\sigma_{T_2}^2}\\
R_2 &\geq \frac{1}{2}\log\frac{(\sigma_X^2+\sigma_{N_1}^2+\sigma_{T_1}^2)(\sigma_X^2+\sigma_{N_2}^2+\sigma_{T_2}^2)-\sigma_X^4}{\sigma_X^2\sigma_{T_2}^2+\sigma_{N_1}^2\sigma_{T_2}^2+\sigma_{T_1}^2\sigma_{T_2}^2}\\
R_1+R_2 &\geq \frac{1}{2}\log\frac{(\sigma_X^2+\sigma_{N_1}^2+\sigma_{T_1}^2)(\sigma_X^2+\sigma_{N_2}^2+\sigma_{T_2}^2)-\sigma_X^4}{\sigma_{T_1}^2\sigma_{T_2}^2}\Bigg\}.
\end{aligned}
$$

Recently by developing the method of Oohama [17], Prabhakaran *et al.*[18] have shown that for the quadratic Gaussian CEO problem, the Berger-Tung rate region is tight, i.e., $\mathcal{R}(D)=\mathcal{R}_{BT}(D)$.

Let $\partial\mathcal{R}(D)$ denote the boundary of $\mathcal{R}(D)$. We have $D^*(R_1,\cdots,R_M)=\min\{D:(R_1,\cdots,R_M)\in\partial\mathcal{R}(D)\}$. In general, $\partial\mathcal{R}(D)$ can be partitioned into 3 pieces as shown in Fig.3. Any point on part a can be represented as $(I(Y_1;W_1),I(Y_2;W_2|W_1))$ for a unique (W_1,W_2). Any point on part c can be represented as $(I(Y_1;W_1|W_2),I(Y_2;W_2))$ for a unique (W_1,W_2). For all the points on part b, we have $R_1+R_2=I(Y_1,Y_2;W_1,W_2)$, $I(Y_1;W_1|W_2)\leq R_1\leq I(Y_1;W_1)$, $I(Y_2;W_2|W_1)\leq R_2\leq I(Y_2;W_2)$ with

$$
\sigma_{T_i}^2=\begin{cases}\dfrac{\sigma_{N_i}^2\left(\frac{1}{D_0(L)}-\frac{1}{D}\right)}{\frac{L}{\sigma_{N_L}^2}+\frac{1}{D}-\frac{1}{D_0(L)}} & i\leq L\\ \infty & i>L\end{cases}\tag{1}
$$

where $L = \max\left\{k : k/\sigma^2_{N_k} + 1/D - 1/D_0(k), 1 \le k \le 2\right\}$, $D_0(k) = (1/\sigma^2_X + \sum_{i=1}^{k} 1/\sigma^2_{N_i})^{-1}$, and D is the solution to the equation

$$R_1 + R_2 = \frac{1}{2}\log^+\left(\frac{\sigma^2_X}{D}\prod_{i=1}^{L}\left(\frac{L}{\sigma^2_{N_i}\left(\frac{1}{D_0(L)} - \frac{1}{D}\right)}\right)\right),$$

which is the minimum achievable distortion with rate pair (R_1, R_2). Hence, a unique (W_1, W_2) is associated with each point $(R_1, R_2) \in \partial\mathcal{R}(D)$. Furthermore, since

$$\bigcup_{D_0(2)\le D\le\sigma^2_X} \partial\mathcal{R}(D) = \{(R_1, R_2) : R_1 \ge 0, R_2 \ge 0)\},$$

we can conclude that any rate pair (R_1, R_2) is associated with a unique (W_1, W_2), and thus a unique $(\sigma^2_{T_1}, \sigma^2_{T_2})$, which we shall denote by $(W_1(R_1, R_2), W_2(R_1, R_2))$ and $(\sigma^2_{T_1}(R_1, R_2), \sigma^2_{T_2}(R_1, R_2))$, respectively. Moreover, $(\frac{1}{\sigma^2_X} + \frac{1}{\sigma^2_{N_1}+\sigma^2_{T_1}(R_1,R_2)} + \frac{1}{\sigma^2_{N_2}+\sigma^2_{T_2}(R_1,R_2)})^{-1}$ is the minimum distortion achievable with the rate pair (R_1, R_2) .

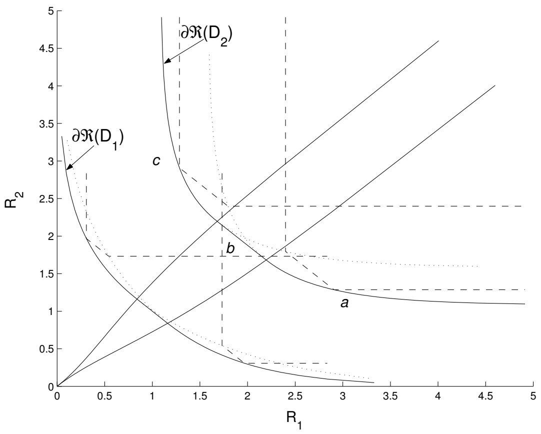

Figure 3: Rate region for the quadratic Gaussian CEO problem

Lemma 1. $(\sigma^2_{T_1}(R_1, R_2), \sigma^2_{T_2}(R_1, R_2)) \ge (\sigma^2_{T_1}(R'_1, R'_2), \sigma^2_{T_2}(R'_1, R'_2))$ if $(R'_1, R'_2) \ge (R_1, R_2)$. Here $(a, b) \ge (a', b')$ means $a \ge a'$ and $b \ge b'$.

Proof: As shown in Fig.4, we partition the region $\{(R_1, R_2) : R_1 \ge 0, R_2 \ge 0)\}$ into three subregions: A,B,C. Any point in region A is in part a of $\partial\mathcal{R}(D)$ for some D, region B and C are similarly defined. The boundary shared by two adjacent regions belongs to both of them.

Let $(W_1, W_2) = (W_1(R_1, R_2), W_2(R_1, R_2))$, $(W'_1, W'_2) = (W_1(R'_1, R'_2), W_2(R'_1, R'_2))$. Let $(\sigma^2_{T_1}, \sigma^2_{T_2}) = (\sigma^2_{T_1}(R_1, R_2), \sigma^2_{T_2}(R_1, R_2))$, $(\sigma^2_{T'_1}, \sigma^2_{T'_2}) = (\sigma^2_{T_1}(R'_1, R'_2), \sigma^2_{T_2}(R'_1, R'_2))$. If both (R_1, R_2) and (R'_1, R'_2) are in region A, then we have $R_1 = I(Y_1; W_1), R_2 = I(Y_2; W_2|W_1)$ and $R'_1 = I(Y_1; W'_1), R_2 = I(Y_2; W'_2|W'_1)$. Since $R'_1 \ge R_1$, it follows that $\sigma^2_{T_1} \ge \sigma^2_{T'_1}$. Now we have $I(Y_2; W'_2|W_1) \ge I(Y_2; W'_2|W'_1) \ge I(Y_2; W_2|W_1)$, which implies $\sigma^2_{T_2} \ge \sigma^2_{T'_2}$. It can be proved in a similar way for the case when both (R_1, R_2) and (R'_1, R'_2) are in region C.

Next consider the case when both (R_1, R_2) and (R'_1, R'_2) are in region B. Let D be the minimum distortion that can be achieved by (R_1, R_2) and D' be the minimum distortion that can be achieved by (R'_1, R'_2). Since $D \ge D'$, it follows from (1) that $(\sigma^2_{T_1}, \sigma^2_{T_2}) \ge (\sigma^2_{T'_1}, \sigma^2_{T'_2})$.

Now consider the general case when (R_1, R_2) and (R_1', R_2') are in different regions, say (R_1, R_2) is in region A and (R_1', R_2') is in region B. Suppose the line segment that connects (R_1, R_2) and (R_1', R_2') intersects the boundary of A and B at (R_1'', R_2''). Let $(\sigma^2_{T_1''}, \sigma^2_{T_2''}) = (\sigma^2_{T_1}(R_1'', R_2''), \sigma^2_{T_2}(R_1'', R_2''))$. We have $(\sigma^2_{T_1}, \sigma^2_{T_2}) \geq (\sigma^2_{T_1''}, \sigma^2_{T_2''})$ since both (R_1, R_2) and (R_1'', R_2'') are in region A and $(R_1'', R_2'') \geq (R_1, R_2)$. Similarly, $(\sigma^2_{T_1''}, \sigma^2_{T_2''}) \geq (\sigma^2_{T_1'}, \sigma^2_{T_2'})$ since both (R_1'', R_2'') and (R_1', R_2') are in region B and $(R_1', R_2') \geq (R_1'', R_2'')$. Hence we have $(\sigma^2_{T_1}, \sigma^2_{T_2}) \geq (\sigma^2_{T_1'}, \sigma^2_{T_2'})$. The other cases can be treated in a similar way. ■

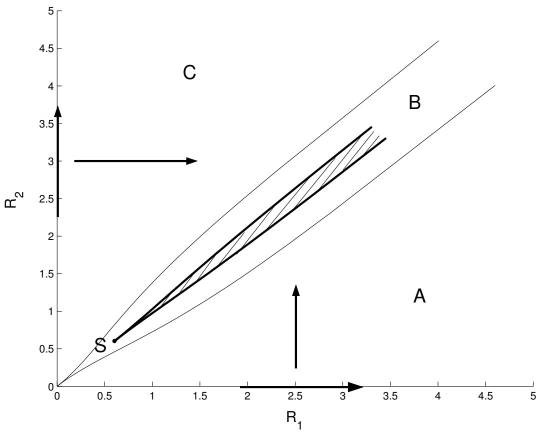

Figure 4: Distributed successive refinement for the quadratic Gaussian CEO problem

Let $(W_1, W_2) = (W_1(R_1, R_2), W_2(R_1, R_2))$, $(W_1', W_2') = (W_1(R_1', R_2'), W_2(R_1', R_2'))$. Let $(\sigma^2_{T_1}, \sigma^2_{T_2}) = (\sigma^2_{T_1}(R_1, R_2), \sigma^2_{T_2}(R_1, R_2))$, $(\sigma^2_{T_1'}, \sigma^2_{T_2'}) = (\sigma^2_{T_1}(R_1', R_2'), \sigma^2_{T_2}(R_1', R_2'))$. If $(R_1', R_2') \geq (R_1, R_2)$, by Lemma 1, we have $(\sigma^2_{T_1}, \sigma^2_{T_2}) \geq (\sigma^2_{T_1'}, \sigma^2_{T_2'})$. Let $(\hat{W}_1, \hat{W}_2) = (W_1', W_2') + (\triangle T_1, \triangle T_2)$, where $\triangle T_1 \sim \mathcal{N}(0, \sigma^2_{T_1} - \sigma^2_{T_1'}), \triangle T_2 \sim \mathcal{N}(0, \sigma^2_{T_2} - \sigma^2_{T_2'})$ are mutually independent and independent of everything else. Note: (W_1, W_2) equals $(\hat{W}_1, \hat{W}_2)$ in distribution. We can regard $(\hat{W}_1, \hat{W}_2)$ as a copy of (W_1, W_2) in the probability space in which (W_1', W_2') is constructed.

Theorem 2. There exists a distributed successive refinement scheme from (R_1, R_2) to (R_1', R_2') if (R_1', R_2') is on the line segment that connects $(R_1 + I(Y_1; W_1'|\hat{W}_1, \hat{W}_2), R_2 + I(Y_2; W_2'|W_1', \hat{W}_2))$ and $(R_1 + I(Y_1; W_1'|\hat{W}_1, W_2'), R_2 + I(Y_2; W_2'|\hat{W}_1, \hat{W}_2))$.

Proof: With rate pair (R_1, R_2), encoder 1 and encoder 2 can convey $(\hat{W}_1, \hat{W}_2)$ to the decoder. With additional rate pair $(I(Y_1; W_1'|\hat{W}_1, \hat{W}_2), I(Y_2; W_2'|W_1', \hat{W}_2))$, encoder 1 and encoder 2 can convey (W_1', W_2') to the decoder via successive refinement. Similarly, with additional rate $(I(Y_1; W_1'|\hat{W}_1, W_2'),\ I(Y_2; W_2'|\hat{W}_1, \hat{W}_2))$, encoder 1 and encoder 2 can also convey (W_1', W_2') to the decoder via successive refinement. Therefore by timesharing[1], with any additional rate pair on the line segment that connects $(I(Y_1; W_1'|\hat{W}_1, \hat{W}_2), I(Y_2; W_2'|W_1', \hat{W}_2))$ and $(I(Y_1; W_1'|\hat{W}_1, W_2'), I(Y_2; W_2'|\hat{W}_1, \hat{W}_2))$, encoder

[1]By an argument similar to that of Theorem 1, this timesharing can be replaced by a successive Wyner-Ziv coding scheme.

1 and encoder 2 can convey $(\hat{W}_1, \hat{W}_2)$ to decoder. Furthermore, it's easy to check that

$$\begin{aligned}
& R_1 + R_2 + I(Y_1; W_1'|\hat{W}_1, \hat{W}_2) + I(Y_2; W_2'|W_1', \hat{W}_2) \\
= \; & I(Y_1, Y_2; W_1, W_2) + I(Y_1; W_1'|\hat{W}_1, \hat{W}_2) + I(Y_2; W_2'|W_1', \hat{W}_2) \\
= \; & I(Y_1, Y_2; \hat{W}_1, \hat{W}_2) + I(Y_1; W_1'|\hat{W}_1, \hat{W}_2) + I(Y_2; W_2'|W_1', \hat{W}_2) \\
= \; & I(Y_1, Y_2; W_1', W_2') = R_1' + R_2'.
\end{aligned}$$

Similarly, we have $R_1 + R_2 + I(Y_1; W_1'|\hat{W}_1, \hat{W}_2) + I(Y_2; W_2'|W_1', \hat{W}_2) = R_1' + R_2'$. The proof is complete. ■

As illustrated in Fig.4, there exists a distributed successive refinement scheme from point s to any point in the dark region. The following results can be easily reduced from Theorem 2. The proofs are omitted.

Corollary 1. If both (R_1, R_2) and (R_1', R_2') are in region A, then there exists a distributed successive refinement scheme from (R_1, R_2) to (R_1', R_2') if $R_1 = R_1'$ and $R_2' \geq R_2$. If both (R_1, R_2) and (R_1', R_2') are in region C, then there exists a distributed successive refinement scheme from (R_1, R_2) to (R_1', R_2') if $R_1' \geq R_1$ and $R_2 = R_2'$.

Remark: If $R_2 = \infty$ and $\sigma_{N_1}^2 = 0$, then the quadratic Gaussian CEO problem becomes the Wyner-Ziv problem for a jointly Gaussian source. Corollary 1 implies the successive refinability for the Wyner-Ziv problem of jointly Gaussian source [19], which further includes the successive refinability of the Gaussian source as a special case [14].

Corollary 2. Suppose $\sigma_{N_1}^2 = \sigma_{N_2}^2$. Then there exists a distributed successive refinement scheme from (R, R) to (R', R') if $R' \geq R$.

IV. Conclusion

We discussed two closely related aspects of distributed source coding. Theorem 1 implies that any coding scheme in the Berger-Tung rate region can be realized via successive Wyner-Ziv coding. Roughly speaking, a high complexity distributed source code can be decomposed into low complexity codes. Theorem 2 implies, at leat for the quadratic Gaussian case, that a high rate distributed source code can be built upon a low rate code via distributed successive refinement without any sacrifice of compression and estimation efficiency.

References

[1] T. P. Coleman, A. H. Lee, M. Médard and M. Effros, "On some new approaches to practical Slepian-Wolf compression inspired by channel coding," *Proc. DCC04*, Snowbird, UT, Mar. 2004.

[2] S. Cheng and Z. Xiong, "Successive refinement for the Wyner-Ziv problem and layered code design," *Proc. DCC04*, Snowbird, UT, Mar. 2004.

[3] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Info. Theory*, vol.IT-19, pp. 471-480, Jul. 1973.

[4] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Info. Theory*, vol. 22, no. 1, pp. 1-10, Jan. 1976.

[5] B. Rimoldi and R. Urbanke, "Asynchronous Slepian-Wolf coding via source-splitting, in *Proc. ISIT*, June 29-July 4, 1997, Ulm, Germany, p. 271.

[6] A. J. Grant, B. Rimoldi, R. L. Urbanke, and P. A. Whiting, "Rate-splitting multiple access for discrete memoryless channels," *IEEE Trans. Inform. Theory* vol. 47, no. 3, pp. 873-890, Mar. 2001.

[7] B. Rimoldi, "Generalized time sharing: a low-complexity capacity-achieving multiple-access technique," *IEEE Trans. Inform. Theory*, vol. 47, no. 6, pp. 2432-2442, Sept. 2001.

[8] T. Berger, "Multiterminal source coding," in The Information Theory Approach to Communications (CISM Courses and Lectures, no. 229), G. Longo, Ed. Vienna/New York: Springer-Verlag, 1978, pp. 171-231.

[9] S. Y. Tung, "Multiterminal source coding," Ph.D. dissertation, School of Electrical Engineering, Cornell Univ., Ithaca, NY, May 1978.

[10] T. Berger, Z. Zhang, and H. Viswanathan, "The CEO problem," *IEEE Trans. Inform. Theory*, vol. 42, pp. 887-902, May 1996.

[11] P. Viswanath, "Sum rate of multiterminal gaussian source coding," submitted to *IEEE Trans. on Information Theory*, in DIMACS Series in Discrete Mathematics and Theoretical Computer Science.

[12] 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 J. Select. Areas Commun.*, vol. 22, pp. 977-987, Aug. 2004.

[13] S. C. Draper and G. W. Wornell, "Side information aware coding strategies for sensor networks," *IEEE J. Select. Areas Commun.*, vol. 22, pp. 966-976, Aug. 2004.

[14] W. H. R. Equitz and T. M. Cover, "Successive refinement of information," *IEEE Trans. on Inform. Theory*, vol. 37, pp. 269-274, Mar. 1991.

[15] J. Chen and T. Berger, "Successive Wyner-Ziv coding scheme and its application to the quadratic gaussian CEO problem," submitted *IEEE Trans. on Inform Theory.*

[16] H. Viswanathan and T. Berger, "The quadratic Gaussian CEO problem," *IEEE Trans. Inform. Theory*, vol. 43, pp. 1549-1559, Sept. 1997.

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

[18] V. Prabhakaran, D. Tse and K. Ramchandran, "Rate region of the quadratic gaussian CEO problem," *Proc. ISIT*, June 27-July 2, 2004, Chicago, USA, pp. 119.

[19] Y. Steinberg and N. Merhav, "On successive refinement for the Wyner-Ziv problem," submitted to IEEE Trans. Information Theory, Mar. 2003.