Successive Refinement of Vector Sources Under Individual Distortion Criteria

Jayanth Nayak, Ertem Tuncel, Deniz Gündüz, Elza Erkip

Abstract

The well-known successive refinement problem is extended to vector sources where individual distortion constraints are posed on every vector component. This extension is then utilized for the derivation of a necessary and sufficient condition for *vector successive refinability*. For 2-D vector Gaussian and binary symmetric sources, it is shown that the successive refinability is *not* granted everywhere, unlike in the 1-D case for these source distributions. Moreover, the behavior of these sources with respect to successive refinability is shown to exhibit remarkable similarity. For the Gaussian case, the rate loss at the second stage when the first stage achieves the optimal rate-distortion performance is also analyzed, and it is shown that the rate loss can be as high as 0.161 bits in an appropriately defined "non-degenerate" refinement scenario. On the other hand, in the degenerate case which corresponds to what is known as sequential coding of correlated sources, it is shown that the rate loss can be as high as 0.5 bits.

I. Introduction

Multiresolution coding, or successive refinement, refers to coding of a source in multiple stages, where at each stage the quality of reconstruction is improved. This source coding scenario has been extensively studied in the last two decades by Koshelev [5], Equitz and Cover [4], Rimoldi [9], Effros [1], [2], Lastras and Berger [6], Tuncel and Rose [10], [11], [12], and many others. Among the most popular questions is whether a given source is successively refinable at the prescribed distortion levels, i.e., whether it is possible to achieve the single resolution rate distortion function at every stage. In [4], it is shown that successive refinability is granted between all distortion levels for Gaussian, Laplacian, and Bernoulli sources under square error, absolute error, and Hamming distortion measures, respectively. Another celebrated result is that the *rate loss*, i.e., the amount of excess rate one needs to expend on top of the rate-distortion function, can be universally bounded for all sources under square error distortion [6].

In this paper, we extend the multiresolution scenario to vector sources where individual distortion constraints are posed on every vector component. This regime of source coding for single resolution coding is analyzed by Xiao and Luo in [15], where they characterize the behavior of the rate-distortion function for a jointly Gaussian vector source under individual square error distortion criteria. Unlike the

Jayanth Nayak and Ertem Tuncel are with University of California, Riverside, CA. Email: {jnayak,ertem}@ee.ucr.edu. Deniz Gündüz is with Princeton University, Princeton, NJ, and with Stanford University, Stanford, CA. E-mail: dgunduz@princeton.edu. Elza Erkip is with Polytechnic Institute of New York University, Brooklyn, NY. E-mail: elza@poly.edu

This work was partly presented at the 2007 Information Theory Workshop, Lake Tahoe, CA.

scalar rate-distortion function for Gaussian sources, the rate-distortion function for vector Gaussian sources does not have a uniform behavior for all distortion vectors.

While deriving a single-letter characterization of all achievable rate-distortion tuples in the multiresolution extension is a rather straightforward task, the *computation* of the rate-distortion region is not as easy in general. This is evidenced by the treatment of the special case of successive coding [7], [8], which proved difficult to solve even for jointly Gaussian sequences under square error distortion measure. Specifically, the approach in [7], [8] was to employ a generalized Shannon lower bound and rigorously show its tightness. Presumably, the same approach for the more general successive refinement problem would be even harder to follow.

In this paper, we instead tackle the issues of (i) successive refinability, and (ii) rate loss in the second stage when the first stage is optimal. We first use a straightforward extension of the Markovity condition derived in [4] to investigate whether vector successive refinability holds for two interesting cases: (i) 2-D Gaussian vectors with square error distortion criterion on each vector component, and (ii) 2-D binary symmetric vectors with Hamming distortion criterion on each component. Surprisingly, we observe that, unlike in the scalar case, successive refinability is *not* granted everywhere (i.e., from any distortion vector in the first stage to any distortion vector in the second stage) for these two examples. Also, the behavior of the two source distributions with respect to successive refinability under the corresponding distortion measures exhibit remarkable similarity.

We then turn to the rate loss problem for jointly Gaussian sources. We first tackle the *non-degenerate* refinement case, i.e., when the achieved distortion levels cannot be reduced without increasing the rate in either stage. In this “fair” scenario, we are able to observe a maximum rate loss of 0.161 bits. We then analyze the rate loss in a degenerate regime which corresponds to what is known as sequential coding of correlated sources [14]. We show that the rate loss in this case can be as high as 0.5 bits. The rate loss in the sequential coding problem was previously bounded *universally* by 1 bit [6]. Our result implies that this bound cannot be reduced by more than 0.5 bits.

In a related work, L multiple descriptions of a vector Gaussian source for individual and central receivers are studied by Wang and Viswanath [13]. One might initially perceive our scenario (for the quadratic Gaussian case) as a special case of the problem discussed therein. As usual, the presumptive specialization would be obtained by choosing a dummy decoder at the terminal receiving only the refinement information. However, instead of individual distortion criteria on every source component, Wang and Viswanath consider *covariance distortion constraints*, i.e., they analyze the case where the time-averaged covariance of the reconstruction error vector is "less than" a prescribed distortion matrix (in the sense of a positive semidefinite ordering). Though it is tempting to think that choosing a diagonal distortion matrix reduces the scenario of [13] to ours, this choice introduces extra constraints on the off-diagonal entries of the error covariance matrix, and therefore the two distortion regimes are different. Moreover, the requirement that this diagonal distortion matrix must be less than the covariance of the source, as assumed for all distortion matrices in [13], is limiting in the sense that one cannot observe the interesting non-uniform rate-distortion behavior mentioned above.

We begin by introducing the preliminaries and background in the next section. We then investigate the behavior of 2-D Gaussian sources under square error distortion and of 2-D binary sources under Hamming distortion with respect to successive refinability in Sections III and IV, respectively. Section V is devoted to the analysis of the rate loss for 2-D Gaussian sources. Section VI summarizes the results and concludes the paper.

II. Preliminaries and Background

Let a stationary and memoryless source produce the vector sequence $\{\mathbf{X}(t)\}_{t=0}^{\infty}$ according to a probability distribution $p_{\mathbf{X}}$, where $\mathbf{X}(t) = [X_1(t)\ X_2(t)\ \ldots\ X_N(t)]^T \in \mathcal{X}^N$ for some underlying alphabet $\mathcal{X}$. Denote the reconstruction of this sequence as $\{\hat{\mathbf{X}}(t)\}_{t=0}^{\infty}$ where $\hat{\mathbf{X}}(t) \in \hat{\mathcal{X}}^N$. Also denote length-$n$ blocks of the sequence and its reconstruction as $\mathbf{X}(1..n)$ and $\hat{\mathbf{X}}(1..n)$, respectively. Similar notation applies to the ith component of the sequence and its reconstruction: $X_i(1..n)$ and $\hat{X}_i(1..n)$. We shall also use $\mathbf{X}$ as a generic random vector representing one sample of the source sequence.

Let $d_i : \mathcal{X} \times \hat{\mathcal{X}} \to [0, \infty)$ with $i = 1, 2, \ldots, N$ be the single-letter distortion measures for individual vector components, i.e.,

$$d_i(X_i(1..n), \hat{X}_i(1..n)) = \frac{1}{n} \sum_{t=1}^{n} d_i(X_i(t), \hat{X}_i(t)) \ .$$

We first repeat the definition of the single-stage rate-distortion problem and its single-letter characterization, which appeared in [15], then define the multi-stage problem, and discuss its connections to similar problems in the literature.

A. The single-stage problem

Definition 1: A rate-distortion pair $(R, \mathbf{D})$ is *achievable* if for any $\epsilon > 0$ and sufficiently large n there exist an encoder

$$f : \mathcal{X}^{Nn} \rightarrow \{1, 2, \ldots, \lfloor 2^{nR} \rfloor\}$$

and N decoders

$$g_i : \{1, 2, \ldots, \lfloor 2^{nR} \rfloor\} \rightarrow \hat{\mathcal{X}}^n$$

such that

$$E\{d_i(X_i(1..n), \hat{X}_i(1..n))\} \leq D_i + \epsilon$$

for $i = 1, 2, \ldots, N$, where

$$\hat{X}_i(1..n) = g_i(f(\mathbf{X}(1..n))) \ .$$

For a fixed distortion vector $\mathbf{D}$, the minimum rate R for which $(R, \mathbf{D})$ is achievable is denoted by $R(\mathbf{D})$.

The block-diagram of the single-stage system is shown in Figure 1. One can observe that, as pointed out in [7], [8], the single-stage rate-distortion problem is a special case of the more general robust descriptions problem in [3]. More specifically, in the robust descriptions scenario, the rate-distortion function for source S and its reconstructions $\hat{S}_1, \hat{S}_2, \ldots, \hat{S}_N$ is given by

$$R(\mathbf{D}) = \min_{E\{\rho_i(S,\hat{S}_i)\} \leq D_i, \ \forall i} I(S; \hat{S}_1, \ldots, \hat{S}_N)$$

Substituting $S = \mathbf{X}$, $\hat{S}_i = \hat{X}_i$, and $\rho_i(S, \hat{S}_i) = d_i(X_i, \hat{X}_i)$ yields the following theorem as a special case.

Theorem 1 ([15]):

$$R(\mathbf{D}) = \min_{E\{\mathbf{d}(\mathbf{X},\hat{\mathbf{X}})\} \leq \mathbf{D}} I(\mathbf{X}; \hat{\mathbf{X}})$$

where

$$\mathbf{d}(\mathbf{X}, \hat{\mathbf{X}}) = [d_1(X_1, \hat{X}_1) \ \ d_2(X_2, \hat{X}_2) \ \ \ldots \ \ d_N(X_N, \hat{X}_N)]^T$$

and $\mathbf{a} \leq \mathbf{b}$ is a shorthand notation for $a_i \leq b_i$, $i = 1, 2, \ldots, N$.

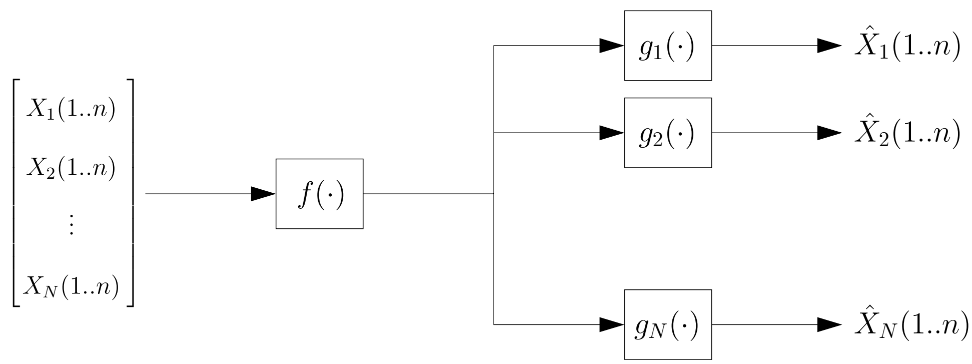

Fig. 1. Single-stage vector source coding with individual distortion criteria.

The jointly Gaussian vector source under square error distortion measure is analyzed by Xiao and Luo in [15]. We now summarize their solution in the 2-D case. Without loss of generality, assume that the covariance matrix of the source zero-mean jointly Gaussian $\mathbf{X}$ is given by

$$\mathbf{C}_{\mathbf{X}} = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix},$$

with $0 < \rho < 1$. For convenience, we will use the notation

$$\delta_i \triangleq 1 - D_i$$

when necessary. Define three regions in the unit square on the $\mathbf{D}$-plane as

$$\begin{aligned} \mathcal{D}_1 &= \{\mathbf{D} : \rho^2 \le \delta_1 \delta_2\}, \\ \mathcal{D}_2 &= \mathcal{D}_1^c \cap \left\{\mathbf{D} : \rho^2 \le \min\left(\frac{\delta_1}{\delta_2}, \frac{\delta_2}{\delta_1}\right)\right\}, \\ \mathcal{D}_3 &= \mathcal{D}_1^c \cap \mathcal{D}_2^c, \end{aligned}$$

where the superscript c denotes the complement of a set. Figure 2 illustrates these three regions. Whenever necessary, we will use the notation $\mathcal{D}_j(\rho)$, $j = 1, 2, 3$ to make explicit the dependence of these regions on ρ. The rate-distortion function is given by[1]

$$R(\mathbf{D}) = \begin{cases} \frac{1}{2}\log\frac{1-\rho^2}{D_1 D_2} & \mathbf{D} \in \mathcal{D}_1 \\ \frac{1}{2}\log\frac{1-\rho^2}{D_1 D_2 - (\rho - \sqrt{\delta_1\delta_2})^2} & \mathbf{D} \in \mathcal{D}_2 \\ \frac{1}{2}\log\frac{1}{\min(D_1, D_2)} & \mathbf{D} \in \mathcal{D}_3 \end{cases}. \quad (1)$$

The region $\mathcal{D}_3$ is essentially a degenerate one in the sense that if, for example, $\rho^2 > \frac{\delta_2}{\delta_1}$, or equivalently, $D_2 > 1 - \rho^2(1 - D_1)$, we have $R(\mathbf{D}) = \frac{1}{2}\log\frac{1}{D_1}$. The fact that this coincides with the rate-distortion function

[1] All logarithms in this paper are base 2.

of a scalar Gaussian source implies that the optimal coding strategy is to ignore $X_2(1..n)$ and optimally compress $X_1(1..n)$. The other component $X_2(1..n)$ can then be optimally estimated from $\hat{X}_1(1..n)$ resulting in $D_2 = 1 - \rho^2(1 - D_1)$. Similarly for the case $D_1 > 1 - \rho^2(1 - D_2)$.

In [15], the optimal forward test channel is characterized to be a simple additive Gaussian noise channel. However, we will be extensively using the optimal *backward* test channel, which, in the non-degenerate region $\mathcal{D}_1 \cup \mathcal{D}_2$, is given by

$$\mathbf{X} = \hat{\mathbf{X}}_* + \mathbf{Z},$$

where both $\hat{\mathbf{X}}_*$ and $\mathbf{Z}$ are Gaussian vectors independent of each other. Also

$$\begin{aligned} \mathbf{C}_{\hat{\mathbf{X}}_*} &= \begin{bmatrix} \delta_1 & \rho \\ \rho & \delta_2 \end{bmatrix}, \\ \mathbf{C}_{\mathbf{Z}} &= \begin{bmatrix} D_1 & 0 \\ 0 & D_2 \end{bmatrix} \end{aligned}$$

for $\mathbf{D} \in \mathcal{D}_1$, and

$$\begin{aligned} \mathbf{C}_{\hat{\mathbf{X}}_*} &= \begin{bmatrix} \delta_1 & \sqrt{\delta_1\delta_2} \\ \sqrt{\delta_1\delta_2} & \delta_2 \end{bmatrix}, \\ \mathbf{C}_{\mathbf{Z}} &= \begin{bmatrix} D_1 & \rho - \sqrt{\delta_1\delta_2} \\ \rho - \sqrt{\delta_1\delta_2} & D_2 \end{bmatrix} \end{aligned}$$

for $\mathbf{D} \in \mathcal{D}_2$. Note that when $\mathbf{D} \in \mathcal{D}_2$, $\mathbf{C}_{\hat{\mathbf{X}}_*}$ is in fact singular, and hence $\hat{\mathbf{X}}_*$ degenerates to a distribution on the line

$$\hat{X}_2 = \sqrt{\frac{\delta_2}{\delta_1}}\hat{X}_1 \ .$$

Observing that $\rho_{\mathbf{Z}}$, the correlation coefficient of the backward channel noise $\mathbf{Z}$, is given by

$$\rho_{\mathbf{Z}} = \frac{E\{Z_1 Z_2\}}{\sqrt{E\{|Z_1|^2\}E\{|Z_2|^2\}}} = \begin{cases} 0 & \mathbf{D} \in \mathcal{D}_1 \\ \frac{\rho - \sqrt{\delta_1\delta_2}}{\sqrt{D_1 D_2}} & \mathbf{D} \in \mathcal{D}_2 \end{cases} ,$$

or equivalently by

$$\rho_{\mathbf{Z}} = \frac{\max\{0, \rho - \sqrt{\delta_1\delta_2}\}}{\sqrt{D_1 D_2}} \ ,$$

one can rewrite (1) more concisely as

$$R(\mathbf{D}|\rho) = \frac{1}{2}\log\frac{1-\rho^2}{D_1 D_2(1-\rho_{\mathbf{Z}}^2)} \tag{2}$$

whenever $\mathbf{D} \in \mathcal{D}_1(\rho) \cup \mathcal{D}_2(\rho)$, where the dependence of $R(\mathbf{D})$ on ρ is also made explicit.

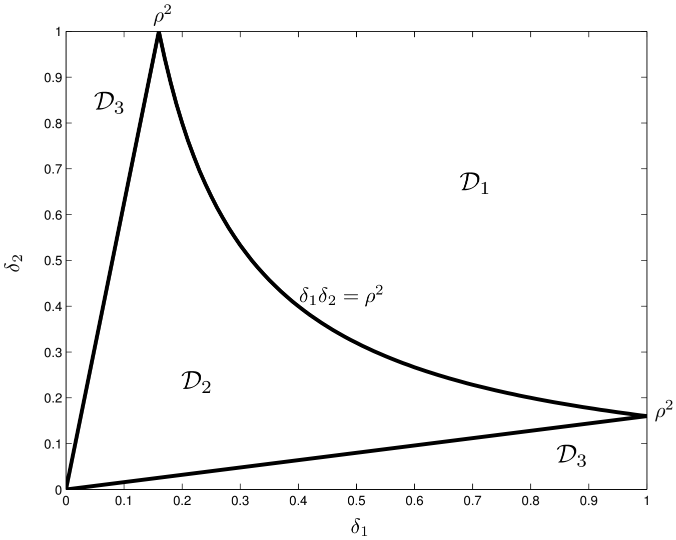

Fig. 2. The three regions $\mathcal{D}_1$, $\mathcal{D}_2$, and $\mathcal{D}_3$ for $\rho = 0.4$.

B. The multi-stage extension

Since subscripts are reserved for vector components, we use superscript l on encoders, decoders, rates, distortion levels, and random variables to indicate stage l. Exceptions to this rule are expressions such as ρ^2, $\rho^2_{\mathbf{Z}}$, $\rho^2_{\mathbf{Z}^1}$, etc., which are to be understood as the square of the respective correlation coefficient.

Definition 2: A rate-distortion $2L$-tuple $(R^1, \ldots, R^L, \mathbf{D}^1, \ldots, \mathbf{D}^L)$ is *successively achievable* if for any $\epsilon > 0$ and sufficiently large n there exist L encoders

$$f^l : \mathcal{X}^{Nn} \rightarrow \{1, 2, \ldots, \lfloor 2^{nR^l} \rfloor\}$$

and NL decoders

$$g_i^l : \{1, 2, \ldots, \lfloor 2^{nR^1} \rfloor\} \times \cdots \times \{1, 2, \ldots, \lfloor 2^{nR^l} \rfloor\} \rightarrow \hat{\mathcal{X}}^n$$

such that

$$E\{d_i(X_i(1..n), \hat{X}_i^l(1..n))\} \leq D_i^l + \epsilon$$

for $i = 1, 2, \ldots, N$ and $l = 1, 2, \ldots, L$, where

$$\hat{X}_i^l(1..n) = g_i^l(f^1(\mathbf{X}(1..n)), \ldots, f^l(\mathbf{X}(1..n))) \ .$$

The following lemma provides a single-letter characterization of the multiresolution rate-distortion function for vector sources. We omit the proof, since it is a straightforward extension of the proofs in [5], [9].

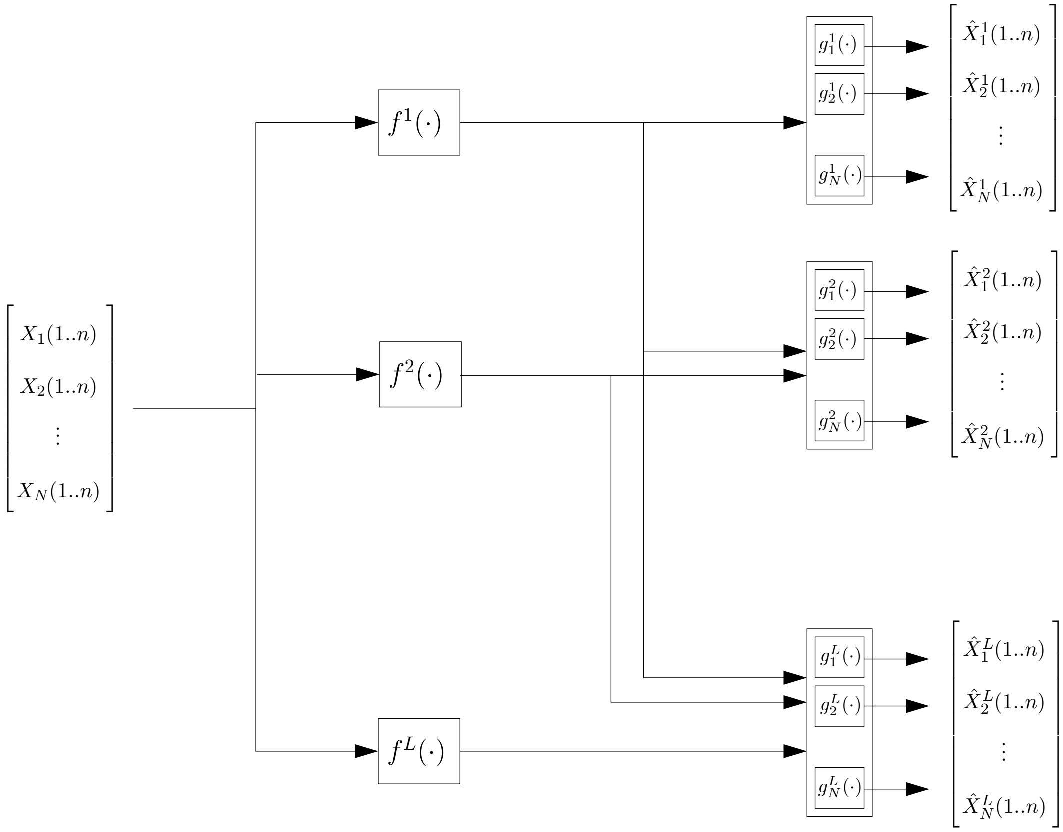

Fig. 3. Multiple-stage vector source coding with individual distortion criteria.

Lemma 1: A rate-distortion tuple $(R^1, \ldots, R^L, \mathbf{D}^1, \ldots, \mathbf{D}^L)$ is successively achievable if and only if there exist auxiliary vectors $\hat{\mathbf{X}}^1, \hat{\mathbf{X}}^2, \ldots, \hat{\mathbf{X}}^L$ satisfying

$$\begin{aligned} I(\mathbf{X}; \hat{\mathbf{X}}^1, \ldots, \hat{\mathbf{X}}^l) &\leq \sum_{k=1}^{l} R^k \\ E\{\mathbf{d}(\mathbf{X}, \hat{\mathbf{X}}^l)\} &\leq \mathbf{D}^l \end{aligned}$$

for all $l = 1, 2, \ldots, L$.

Figure 3 depicts encoder-decoder structure for the multi-stage scenario.

C. Successive refinability for vector sources

As in multi-stage scalar source coding, an important question is whether one can achieve the single-stage rate-distortion function at all stages.

Definition 3: The source $\mathbf{X}$ is said to be successively refinable at the distortion point $(\mathbf{D}^1, \ldots, \mathbf{D}^L)$ with

$$\mathbf{D}^1 \geq \mathbf{D}^2 \geq \ldots \geq \mathbf{D}^L$$

if the rate-distortion tuple $(R^1, \ldots, R^L, \mathbf{D}^1, \ldots, \mathbf{D}^L)$ satisfying

$$\sum_{k=1}^{l} R^k = R(\mathbf{D}^l)$$

for $l = 1, 2, \ldots, L$ is successively achievable.

For any scheme successively achieving $(R^1, \ldots, R^L, \mathbf{D}^1, \ldots, \mathbf{D}^L)$, the associated rate loss at stage l is given by

$$S^l = \sum_{k=1}^{l} R^k - R(\mathbf{D}^l) \ .$$

For the purposes of this paper, the following corollary to Lemma 1 is essential.

Corollary 1: The source $\mathbf{X}$ is successively refinable at $(\mathbf{D}^1, \mathbf{D}^2)$ with $\mathbf{D}^1 \geq \mathbf{D}^2$ if and only if the optimal vectors $\hat{\mathbf{X}}_*^l$ achieving $\{\mathbf{D}^l, R(\mathbf{D}^l)\}$ for $l = 1, 2$ satisfy the Markov chain

$$\mathbf{X} - \hat{\mathbf{X}}_*^2 - \hat{\mathbf{X}}_*^1 \ .$$

Proof: If $(\hat{\mathbf{X}}^1, \hat{\mathbf{X}}^2)$ achieves $\{R(\mathbf{D}^1), R(\mathbf{D}^2) - R(\mathbf{D}^1), \mathbf{D}^1, \mathbf{D}^2\}$, we have

$$R(\mathbf{D}^1) \overset{(a)}{\geq} I(\mathbf{X}; \hat{\mathbf{X}}^1) \overset{(b)}{\geq} R(\mathbf{D}^1)$$

and

$$R(\mathbf{D}^2) \overset{(c)}{\geq} I(\mathbf{X}; \hat{\mathbf{X}}^1, \hat{\mathbf{X}}^2) \overset{(d)}{\geq} I(\mathbf{X}; \hat{\mathbf{X}}^2) \overset{(e)}{\geq} R(\mathbf{D}^2)$$

where (a) and (c) follow from Lemma 1; (d) follows from the chain rule; and (b) and (e) follow from the definition of $R(\mathbf{D})$. Thus, we must have $\hat{\mathbf{X}}^1 = \hat{\mathbf{X}}_*^1$, $\hat{\mathbf{X}}^2 = \hat{\mathbf{X}}_*^2$, and $\mathbf{X} - \hat{\mathbf{X}}_*^2 - \hat{\mathbf{X}}_*^1$.

Conversely, if $\mathbf{X} - \hat{\mathbf{X}}_*^2 - \hat{\mathbf{X}}_*^1$, the choice $\hat{\mathbf{X}}^1 = \hat{\mathbf{X}}_*^1$ and $\hat{\mathbf{X}}^2 = \hat{\mathbf{X}}_*^2$ satisfies $(a) - (e)$ with equality. ■

D. Relation to the successive coding and the sequential coding problems

In this work, we consider $N = 2$ and $L = 2$. For this special case, two problems intimately related to vector successive refinement have been discussed in the literature: successive coding [7], [8] and sequential coding [14] of vector sources, as illustrated in Figure 4 and Figure 5, respectively. In both problems, the decoders g_2^1 and g_1^2 we have in the successive refinement problem do not exist. This is because the first and the second stage descriptions are to be used exclusively for the reconstructions of $X_1(1..n)$ and $X_2(1..n)$, respectively. The difference between the two problem settings is that, in the successive coding problem the whole source $\mathbf{X}(1..n)$ is available at both of the encoders, whereas in the sequential coding problem

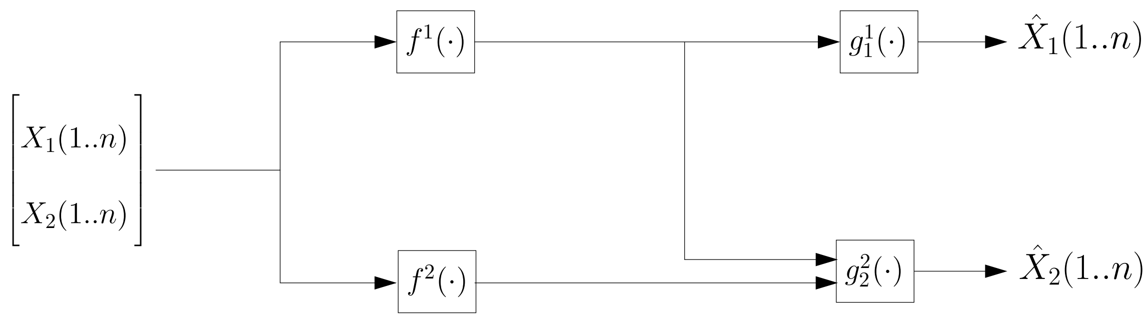

Fig. 4. Successive coding of vector sources as a special case of successive refinement with individual distortion criteria.

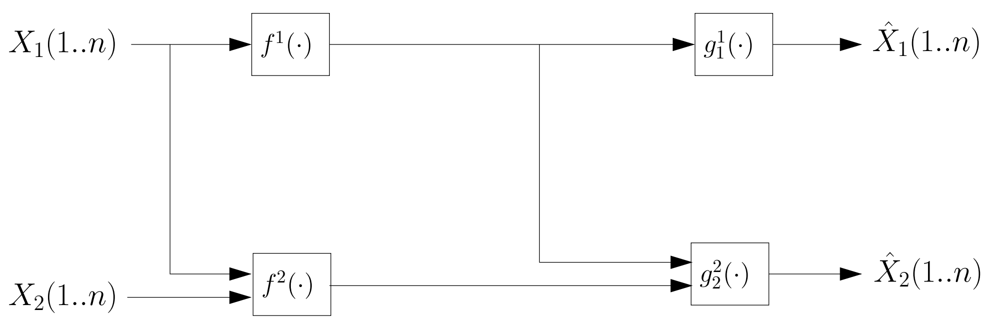

Fig. 5. Sequential coding of vector sources. For Gaussian sources and square error distortion measure, this becomes a special case of successive refinement with individual distortion criteria.

$X_2(1..n)$ is not available at the first stage encoder. In the sequential coding problem, most exciting results have been reported for the minimum achievable total rate $R^{tot} = R^1 + R^2$ (See [14, Sections V and VI] and [6, Section V]).

It is easy to see that the successive coding problem is a special case of the vector successive refinement problem: simply set $\mathbf{D}^1 = [D_1^1 \ \infty]^T$ and $\mathbf{D}^2 = [\infty \ D_2^2]^T$ and let g_2^1 and g_1^2 output arbitrary sequences. On the other hand, the connection between sequential coding and vector successive refinement is less direct due to the unavailability of $X_2(1..n)$ at the first stage encoder f^1. However, it was shown in [14, Theorem 4] that for jointly Gaussian sources and square error distortion measure, the minimum total rate R^{tot} is achieved by optimally compressing $X_1(1..n)$ in the first stage, i.e., $R^1 = R(D_1^1)$. One could thus emulate optimal sequential coding using a vector successive refinement coder with first stage rate $R^1 = R(D_1^1)$, since with this rate constraint the first stage encoder cannot utilize $X_2(1..n)$ even if it has access to it. In other words, for jointly Gaussian sources and square error distortion measure, a total rate R^{tot} is achievable in the sequential coding problem if and only if $(R(\mathbf{D}^1), R^{tot} - R(\mathbf{D}^1), \mathbf{D}^1, \mathbf{D}^2)$ is achievable in the vector successive refinement problem, where, again, $\mathbf{D}^1 = [D_1^1 \ \infty]^T$ and $\mathbf{D}^2 = [\infty \ D_2^2]^T$.

This connection between the two problems allows us to precisely compute the minimum total rate loss

$$S^{tot} \triangleq R^{tot} - R(\mathbf{D})$$

with $\mathbf{D} = [D_1^1 \;\; D_2^2]^T$. This definition of rate loss is made in [6], where the authors also show that under the square error distortion measure, $S^{tot} \leq 1$ universally, i.e., for all sources. Here we show that, surprisingly, the rate loss can be arbitrarily close to 0.5 bits even for jointly Gaussian sources.

E. Connection with the vector Gaussian multiple descriptions problem

In [13], Wang and Viswanath consider the vector Gaussian L-description problem, where there are L side receivers and one central receiver. When $L = 2$, for which the complete rate region is characterized in [13], the only *structural* difference between this scenario and the successive refinement problem is the existence of an additional set of N decoders which receive only the refinement information:

$$g_i^0 : \{1, 2, \ldots, \lfloor 2^{nR^2} \rfloor\} \to \hat{\mathcal{X}}^n$$

with

$$\hat{X}_i^0(1..n) = g_i^0(f^2(\mathbf{X}(1..n)))$$

for $i = 1, 2, \ldots, N$. As usual, if g_i^0 are chosen as dummy decoders that disregard the received information and simply output $\hat{\mathbf{X}}^0(t) = E\{\mathbf{X}(t)\}$ for $1 \leq t \leq n$, the problem reduces to that of successive refinement.

However, this observation still does not render our scenario a special case of that in [13]. Because, in [13] instead of individual distortion criteria, a covariance distortion constraint is considered, whereby it is required that

$$\frac{1}{n}\sum_{t=1}^{n} E\left\{\left(\mathbf{X}(t) - \hat{\mathbf{X}}^l(t)\right)\left(\mathbf{X}(t) - \hat{\mathbf{X}}^l(t)\right)^T\right\} \preceq \mathbf{K}^l$$

for $l = 1, 2$, with $\mathbf{K}^l$ denoting the prescribed covariance matrix at stage l, and $\preceq$ indicating "less than or equal to" in the sense of a positive definite ordering. In this setting, even a diagonal choice for $\mathbf{K}^l$ with

$$\mathbf{K}_{i,i}^l = D_i^l$$

is more constraining than the individual distortion constraints we have in Definition 2. Moreover, the assumption in [13] that $\mathbf{K}^l \preceq \mathbf{C}_{\mathbf{X}}$ is limiting for our purposes. For instance, when $N = 2$ and $\mathbf{K}^l$ is diagonal, this implies that

$$\begin{aligned} 0 &\leq \det\left(\mathbf{C_X} - \mathbf{K}^l\right) \\ &= (1-D_1^l)(1-D_2^l) - \rho^2 \end{aligned}$$

or, equivalently, $\mathbf{D}^l \in \mathcal{D}_1$. In fact, as we discuss in the next section, $\mathbf{D}^1, \mathbf{D}^2 \in \mathcal{D}_1$ is a less interesting case in which successive refinability is always granted.

III. Vector Successive Refinability for 2-D Gaussian Sources

We now investigate whether conditions in Corollary 1 are satisfied for 2-D Gaussian sources under individual square error distortion criteria. It turns out that for certain values of $\mathbf{D}^1$ and $\mathbf{D}^2$, successive refinability is not granted.

We consider three sub-cases: (i) from $\mathcal{D}_1$ to $\mathcal{D}_1$, (ii) from $\mathcal{D}_2$ to $\mathcal{D}_2$, and (iii) from $\mathcal{D}_2$ to $\mathcal{D}_1$, and investigate whether the source is successively refinable in each of these cases. In all three cases, letting $\mathbf{X} = \hat{\mathbf{X}}_*^1 + \mathbf{Z}^1 = \hat{\mathbf{X}}_*^2 + \mathbf{Z}^2$, the Markovity condition reduces to

$$\mathbf{Z}^1 = \mathbf{Z}^2 + \mathbf{N}$$

for a Gaussian vector $\mathbf{N}$ independent of $\mathbf{Z}^2$. This, in turn, holds if and only if $\mathbf{C}_{\mathbf{Z}^1} \succeq \mathbf{C}_{\mathbf{Z}^2}$.

Case i: $\mathbf{D}^1, \mathbf{D}^2 \in \mathcal{D}_1$. Since

$$\mathbf{C}_{\mathbf{Z}^1} - \mathbf{C}_{\mathbf{Z}^2} = \begin{bmatrix} \delta_1^2 - \delta_1^1 & 0 \\ 0 & \delta_2^2 - \delta_2^1 \end{bmatrix} \tag{3}$$

is positive semi-definite for all $\mathbf{D}^2 \leq \mathbf{D}^1$, successive refinability is granted everywhere in this case.

Case ii: $\mathbf{D}^1, \mathbf{D}^2 \in \mathcal{D}_2$. We now have

$$\mathbf{C}_{\mathbf{Z}^1} - \mathbf{C}_{\mathbf{Z}^2} = \begin{bmatrix} \delta_1^2 - \delta_1^1 & \sqrt{\delta_1^2\delta_2^2} - \sqrt{\delta_1^1\delta_2^1} \\ \sqrt{\delta_1^2\delta_2^2} - \sqrt{\delta_1^1\delta_2^1} & \delta_2^2 - \delta_2^1 \end{bmatrix} .$$

Since we assume $\mathbf{D}^2 \leq \mathbf{D}^1$, we have $\mathbf{C}_{\mathbf{Z}^1} \succeq \mathbf{C}_{\mathbf{Z}^2}$ if and only if

$$\begin{aligned} 0 &\leq \det(\mathbf{C}_{\mathbf{Z}^1} - \mathbf{C}_{\mathbf{Z}^2}) \\ &= (\delta_1^2 - \delta_1^1)(\delta_2^2 - \delta_2^1) - \delta_1^2\delta_2^2 - \delta_1^1\delta_2^1 + 2\sqrt{\delta_1^2\delta_2^2\delta_1^1\delta_2^1} \\ &= -\delta_1^1\delta_2^2 - \delta_1^2\delta_2^1 + 2\sqrt{\delta_1^2\delta_2^2\delta_1^1\delta_2^1} . \end{aligned}$$

This is equivalent to having

$$\frac{\delta_1^1\delta_2^2+\delta_1^2\delta_2^1}{2} \le \sqrt{\delta_1^2\delta_2^2\delta_1^1\delta_2^1},$$

which reveals that the Markovity condition is satisfied if and only if the arithmetic mean of $\delta_1^1\delta_2^2$ and $\delta_1^2\delta_2^1$ is *less than or equal to* their geometric mean of the same. But since the arithmetic mean cannot be less than the geometric mean, and the two are equal if and only if the arguments are identical, we have $\delta_1^1\delta_2^2 = \delta_1^2\delta_2^1$, or equivalently,

$$\frac{\delta_2^1}{\delta_1^1} = \frac{\delta_2^2}{\delta_1^2}\ . \tag{4}$$

Hence, the points (δ_1^1, δ_2^1) and (δ_1^2, δ_2^2) must lie on a straight line passing through the origin in the $\delta_1 - \delta_2$ plane.

<u>*Case iii:*</u> $\mathbf{D}^1 \in \mathcal{D}_2$ and $\mathbf{D}^2 \in \mathcal{D}_1$. In this case, we assume without loss of generality that $\delta_2^1 = \nu\delta_1^1$ with $\rho^2 \le \nu \le \frac{1}{\rho^2}$ and $\rho \ge \sqrt{\nu}\delta_1^1$. Then

$$\begin{aligned}\mathbf{C}_{\mathbf{Z}^1} - \mathbf{C}_{\mathbf{Z}^2} &= \begin{bmatrix} \delta_1^2 - \delta_1^1 & \rho - \sqrt{\delta_1^1\delta_2^1} \\ \rho - \sqrt{\delta_1^1\delta_2^1} & \delta_2^2 - \delta_2^1 \end{bmatrix} \\ &= \begin{bmatrix} \delta_1^2 - \delta_1^1 & \rho - \sqrt{\nu}\delta_1^1 \\ \rho - \sqrt{\nu}\delta_1^1 & \delta_2^2 - \nu\delta_1^1 \end{bmatrix}.\end{aligned}$$

The Markovity condition then reduces to

$$(\delta_1^2 - \delta_1^1)(\delta_2^2 - \nu\delta_1^1) \ge \left(\rho - \sqrt{\nu}\delta_1^1\right)^2\ . \tag{5}$$

As a sanity check, this region should include all $(\delta_1^2, \delta_2^2) \in \mathcal{D}_1$ such that $\delta_1^2 \ge \frac{\rho}{\sqrt{\nu}}$ and $\delta_2^2 \ge \rho\sqrt{\nu}$. This follows from the analysis of the previous cases and the observation that the point $(\frac{\rho}{\sqrt{\nu}}, \rho\sqrt{\nu})$ is simultaneously on the line $\delta_2^2 = \nu\delta_1^2$ and on the common boundary of $\mathcal{D}_1$ and $\mathcal{D}_2$. More specifically, according to (4) one can first successively refine without rate loss from $\mathbf{D}^1$ to any intermediate point $\mathbf{D}^0 \in \mathcal{D}_2$ on the line $\delta_2^0 = \nu\delta_1^0$, including $(\frac{\rho}{\sqrt{\nu}}, \rho\sqrt{\nu})$. It then follows from (3) that one can do the same from $\mathbf{D}^0$ to any $\mathbf{D}^2 \le \mathbf{D}^0$. Indeed, the point $\delta_1^2 = \frac{\rho}{\sqrt{\nu}}$ and $\delta_2^2 = \rho\sqrt{\nu}$ satisfies (5) with equality, and the above claim is corroborated. However, it is also clear from (5) that the successive refinability region is not limited to that rectangular region.

Figure 6 shows the successive refinability region for several choices of (δ_1^1, δ_2^1).

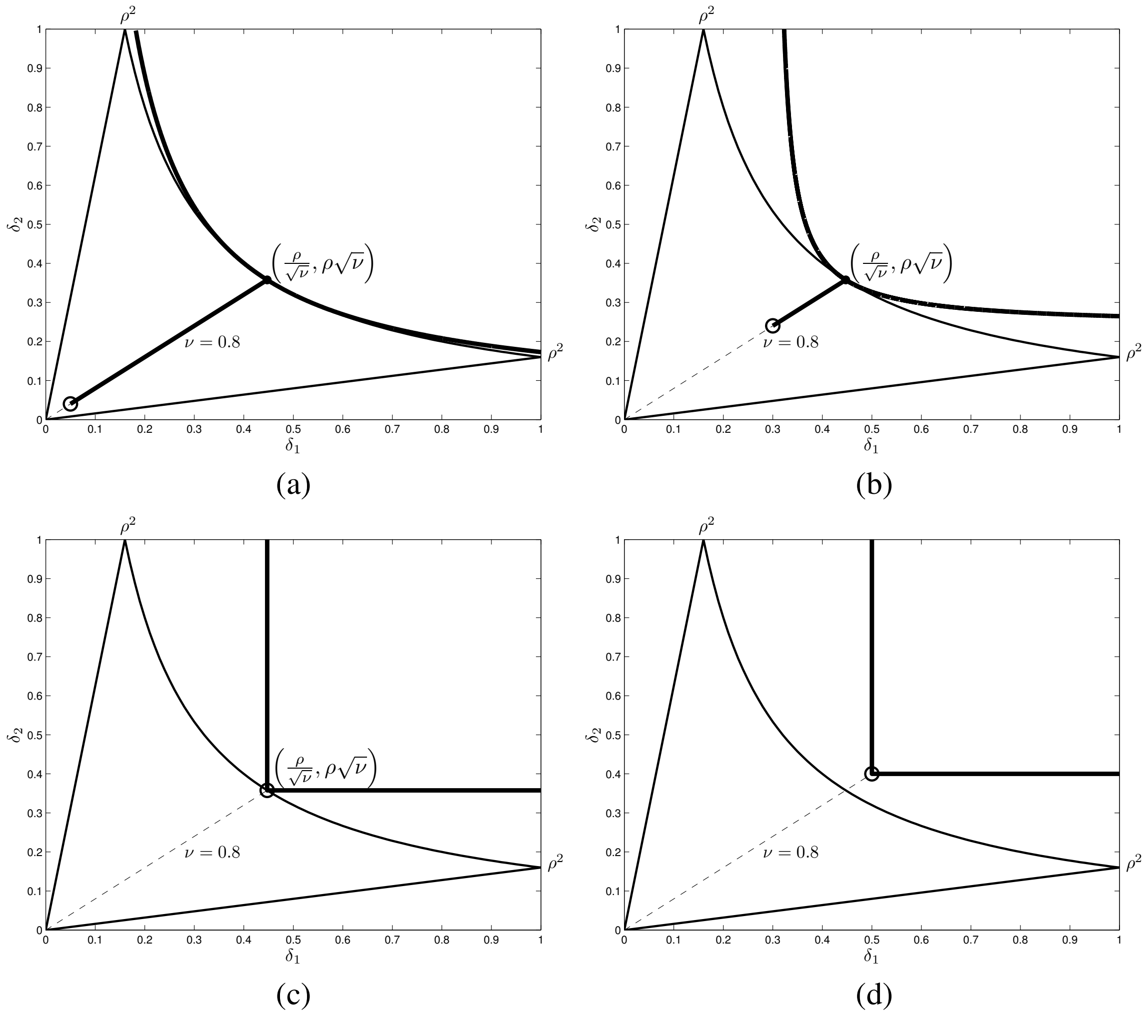

Fig. 6. The successive refinability region in the (δ_1, δ_2)-plane for Gaussian sources with and several (δ_1^1, δ_2^1) pairs (indicated using $\circ$) satisfying $\delta_2^1 = \nu\delta_1^1$ and $\nu = 0.8$. We set $\rho = 0.4$. The particular choices of δ_1^1 for (a)-(d) are 0.05, 0.3, $\frac{\rho}{\sqrt{\nu}} = 0.4472$, and 0.5, respectively. When $\mathbf{D}^1 \in \mathcal{D}_2$ as in (a) and (b), the region of $\mathbf{D}^2$ for which successive refinability holds constitutes a line in $\mathcal{D}_2$ (indicated in bold) and everywhere above the bold curve that lies in $\mathcal{D}_1$. When $\mathbf{D}^1 \in \mathcal{D}_1$ as in (c) and (d), successive refinability holds for all (δ_1^2, δ_2^2) in the rectangular region above the bold lines.

IV. Vector Successive Refinability for Binary Symmetric Sources

In this section, we investigate successive refinability of 2-D binary symmetric sources under the Hamming distortion measure. The behavior of these sources regarding successive refinability exhibits remarkable similarity with that of Gaussian sources under square error distortion.

Let the probability mass function (p.m.f.) of the source be given by

$$\mathbf{P_X} = \begin{bmatrix} p & \frac{1}{2} - p \\ \frac{1}{2} - p & p \end{bmatrix}, \tag{6}$$

where we assume $\frac{1}{4} \leq p \leq \frac{1}{2}$ without loss of generality. If $p < \frac{1}{4}$, then one could switch the roles of 0 and 1 in X_1 (or X_2) and obtain the current form. For this family of sources, we define

$$\delta_i \triangleq 1 - 2D_i$$

and $\delta_i^j \triangleq 1 - 2D_i^j$ for proper i, j. Note that since one can achieve $D_i = \frac{1}{2}$ for $i = 1, 2$ even with zero rate, we need only consider the unit square $\{(\delta_1, \delta_2) : 0 < \delta_1 \leq 1 \text{ and } 0 < \delta_2 \leq 1\}$.

We first compute the rate-distortion function and the optimal test channels for the single-stage problem.

Theorem 2: The rate-distortion function for a 2-D binary symmetric source with a p.m.f. given in (6) is characterized by

$$R(\mathbf{D}) = \begin{cases} H(\mathbf{X}) - \mathcal{H}(D_1) - \mathcal{H}(D_2) & \mathbf{D} \in \mathcal{E}_1 \\ H(\mathbf{X}) - \mathcal{H}(2p) - 2p\mathcal{H}\left(\frac{D_1+D_2+2p-1}{4p}\right) - (1-2p)\mathcal{H}\left(\frac{D_1-D_2+1-2p}{2(1-2p)}\right) & \mathbf{D} \in \mathcal{E}_2 \\ 1 - \mathcal{H}(\min\{D_1, D_2\}) & \mathbf{D} \in \mathcal{E}_3 \end{cases}$$

where $\mathcal{H}(\alpha) \triangleq -\alpha \log \alpha - (1-\alpha)\log(1-\alpha)$, and

$$\begin{aligned} \mathcal{E}_1 &= \{\mathbf{D} : 4p - 1 \leq \delta_1\delta_2\}, \\ \mathcal{E}_2 &= \mathcal{E}_1^c \cap \left\{\mathbf{D} : 4p - 1 \leq \min\left(\frac{\delta_1}{\delta_2}, \frac{\delta_2}{\delta_1}\right)\right\}, \\ \mathcal{E}_3 &= \mathcal{E}_1^c \cap \mathcal{E}_2^c \,. \end{aligned}$$

Also, in the non-degenerate region $\mathbf{D} \in \mathcal{E}_1 \cup \mathcal{E}_2$, the optimal backward channel is always of the form $\mathbf{X} = \hat{\mathbf{X}}_* + \mathbf{Z}$ where $\mathbf{P}_{\hat{\mathbf{X}}_*}$ and $\mathbf{P}_{\mathbf{Z}}$ are given as

$$\mathbf{P}_{\hat{\mathbf{X}}_*} = \begin{bmatrix} q & \frac{1}{2} - q \\ \frac{1}{2} - q & q \end{bmatrix} \tag{7}$$

with

$$q = \frac{1}{4}\left(1 + \frac{4p-1}{\delta_1\delta_2}\right), \tag{8}$$

and

$$\mathbf{P}_{\mathbf{Z}} = \begin{bmatrix} (1-D_1)(1-D_2) & (1-D_1)D_2 \\ D_1(1-D_2) & D_1D_2 \end{bmatrix} \tag{9}$$

for $\mathbf{D} \in \mathcal{E}_1$, and

$$\mathbf{P}_{\hat{\mathbf{X}}_*} = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix} \tag{10}$$

and

$$\mathbf{P}_{\mathbf{Z}} = \frac{1}{2}\begin{bmatrix} 2 - D_1 - D_2 - (1-2p) & D_2 - D_1 + (1-2p) \\ D_1 - D_2 + (1-2p) & D_1 + D_2 - (1-2p) \end{bmatrix} \tag{11}$$

for $\mathbf{D} \in \mathcal{E}_2$.

The proof is given in the Appendix.

The partitioning of the unit square with respect to different rate-distortion behaviors is *exactly* as shown in Figure 2 where $\mathcal{D}_1$, $\mathcal{D}_2$, $\mathcal{D}_3$, and ρ^2 play the roles of $\mathcal{E}_1$, $\mathcal{E}_2$, $\mathcal{E}_3$, and $4p-1$, respectively. Similar to the Gaussian case, $\mathcal{E}_3$ is degenerate in the sense that if, for example, $4p-1 > \frac{\delta_2}{\delta_1}$, or equivalently $D_2 > 1-2p+(4p-1)D_1$, D_2 can be further reduced to $D_2 = 1-2p+(4p-1)D_1$ without increasing $R(\mathbf{D})$. Similarly for the case $D_1 > 1-2p+(4p-1)D_2$.

As in the Gaussian problem, we now consider successive refinement for three sub-cases: (i) from $\mathcal{E}_1$ to $\mathcal{E}_1$, (ii) from $\mathcal{E}_2$ to $\mathcal{E}_2$, and (iii) from $\mathcal{E}_2$ to $\mathcal{E}_1$, and investigate whether the source is successively refinable in each case. Since $\mathbf{X} = \hat{\mathbf{X}}_*^1 \oplus \mathbf{Z}^1 = \hat{\mathbf{X}}_*^2 \oplus \mathbf{Z}^2$, the Markovity condition reduces to

$$\mathbf{Z}^1 = \mathbf{Z}^2 \oplus \mathbf{N}$$

where $\mathbf{N}$ is independent of $\mathbf{Z}^2$. To check this condition, we employ a powerful technique well-known in 2-D signal processing, namely, 2-D discrete Fourier transform (DFT). The main observation here is that $\mathbf{Z}^1 = \mathbf{Z}^2 \oplus \mathbf{N}$ with independent $(\mathbf{Z}^2, \mathbf{N})$ implies that the p.m.f.'s of these random variables satisfy

$$\mathbf{P}_{\mathbf{Z}^1} = \mathbf{P}_{\mathbf{Z}^2} \circ \mathbf{P}_{\mathbf{N}}$$

where $\circ$ denotes the 2-D circular convolution operation. This, in turn, implies

$$\mathcal{F}\left(\mathbf{P}_{\mathbf{Z}^1}\right) = \mathcal{F}\left(\mathbf{P}_{\mathbf{Z}^2}\right) \cdot \mathcal{F}\left(\mathbf{P}_{\mathbf{N}}\right)$$

where $\mathcal{F}$ and $\cdot$ denote 2-D DFT and element-by-element product, respectively.

<u>*Case i:*</u> $\mathbf{D}^1, \mathbf{D}^2 \in \mathcal{E}_1$. It can be shown using (9) that

$$\mathcal{F}\left(\mathbf{P}_{\mathbf{Z}^i}\right) = \begin{bmatrix} 1 & \delta_2^i \\ \delta_1^i & \delta_1^i \delta_2^i \end{bmatrix} \tag{12}$$

for $i = 1, 2$. Thus, we need

$$\begin{aligned} \mathbf{P}_{\mathbf{N}} &= \mathcal{F}^{-1}\left(\begin{bmatrix} 1 & \frac{\delta_2^1}{\delta_2^2} \\ \frac{\delta_1^1}{\delta_1^2} & \frac{\delta_1^1 \delta_2^1}{\delta_1^2 \delta_2^2} \end{bmatrix}\right) \\ &= \frac{1}{4} \begin{bmatrix} 1 + \frac{\delta_1^1}{\delta_1^2} \\ 1 - \frac{\delta_1^1}{\delta_1^2} \end{bmatrix} \begin{bmatrix} 1 + \frac{\delta_2^1}{\delta_2^2} & 1 - \frac{\delta_2^1}{\delta_2^2} \end{bmatrix} \end{aligned}$$

to be a valid p.m.f. But this is always granted since we only focus on $\mathbf{D}_2 \leq \mathbf{D}_1$.

Case ii: $\mathbf{D}^1, \mathbf{D}^2 \in \mathcal{E}_2$. We have from (11) that

$$\mathcal{F}\left(\mathbf{P}_{\mathbf{Z}^i}\right) = \left[\begin{array}{cc} 1 & \delta_2^i \\ \delta_1^i & 4p-1 \end{array} \right] \tag{13}$$

for $i = 1, 2$. This, in turn, implies that we need

$$\begin{aligned} \mathbf{P}_\mathbf{N} &= \mathcal{F}^{-1}\left(\left[\begin{array}{cc} 1 & \frac{\delta_2^1}{\delta_2^2} \\ \frac{\delta_1^1}{\delta_1^2} & 1 \end{array} \right]\right) \\ &= \frac{1}{4}\left[\begin{array}{cc} 2 + \frac{\delta_1^1}{\delta_1^2} + \frac{\delta_2^1}{\delta_2^2} & \frac{\delta_1^1}{\delta_1^2} - \frac{\delta_2^1}{\delta_2^2} \\ \frac{\delta_2^1}{\delta_2^2} - \frac{\delta_1^1}{\delta_1^2} & 2 - \frac{\delta_1^1}{\delta_1^2} - \frac{\delta_2^1}{\delta_2^2} \end{array} \right] \end{aligned}$$

to be a valid p.m.f. It can easily be seen that this requires

$$\frac{\delta_1^1}{\delta_1^2} = \frac{\delta_2^1}{\delta_2^2} \leq 1$$

which is granted only when the points (δ_1^1, δ_2^1) and (δ_1^2, δ_2^2) lie on the same line passing through the origin.

Case iii: $\mathbf{D}^1 \in \mathcal{E}_2$ and $\mathbf{D}^2 \in \mathcal{E}_1$. We assume without loss of generality that $\delta_2^1 = \nu\delta_1^1$ with $4p - 1 \leq \nu \leq \frac{1}{4p-1}$ and $\delta_1^1 \leq \sqrt{\frac{4p-1}{\nu}}$. It follows from (12) and (13) that we need

$$\mathbf{P}_\mathbf{N} = \mathcal{F}^{-1}\left(\left[\begin{array}{cc} 1 & \frac{\nu\delta_1^1}{\delta_2^2} \\ \frac{\delta_1^1}{\delta_1^2} & \frac{4p-1}{\delta_1^2\delta_2^2} \end{array} \right]\right) \triangleq \frac{1}{4\delta_1^2\delta_2^2}\left[\begin{array}{cc} r_{11} & r_{12} \\ r_{21} & r_{22} \end{array} \right]$$

to be valid, where

$$\begin{aligned} r_{11} &= \delta_1^2\delta_2^2 + \nu\delta_1^2\delta_1^1 + \delta_2^2\delta_1^1 + 4p - 1, \\ r_{12} &= \delta_1^2\delta_2^2 - \nu\delta_1^2\delta_1^1 + \delta_2^2\delta_1^1 - 4p + 1, \\ r_{21} &= \delta_1^2\delta_2^2 + \nu\delta_1^2\delta_1^1 - \delta_2^2\delta_1^1 - 4p + 1, \\ r_{22} &= \delta_1^2\delta_2^2 - \nu\delta_1^2\delta_1^1 - \delta_2^2\delta_1^1 + 4p - 1\ . \end{aligned}$$

Observe that the entries of $\mathbf{P}_\mathbf{N}$ always sum up to 1 and $r_{11} \geq 0$ is always granted. Thus, it suffices to check $r_{12} \geq 0$, $r_{21} \geq 0$, and $r_{22} \geq 0$, which can be re-written as

$$(\delta_1^2 + \delta_1^1)(\delta_2^2 - \nu\delta_1^1) \geq 4p - 1 - \nu\left(\delta_1^1\right)^2, \tag{14}$$

$$(\delta_1^2 - \delta_1^1)(\delta_2^2 + \nu\delta_1^1) \geq 4p - 1 - \nu\left(\delta_1^1\right)^2, \tag{15}$$

$$(\delta_1^2 - \delta_1^1)(\delta_2^2 - \nu\delta_1^1) \geq -\left[4p - 1 - \nu\left(\delta_1^1\right)^2\right]\ . \tag{16}$$

Since $\mathbf{D}^2 \leq \mathbf{D}^1$ translates into $\delta_1^2 \geq \delta_1^1$ and $\delta_2^2 \geq \nu\delta_1^1$, and $4p - 1 \geq \nu\left(\delta_1^1\right)^2$, (16) becomes vacuous.

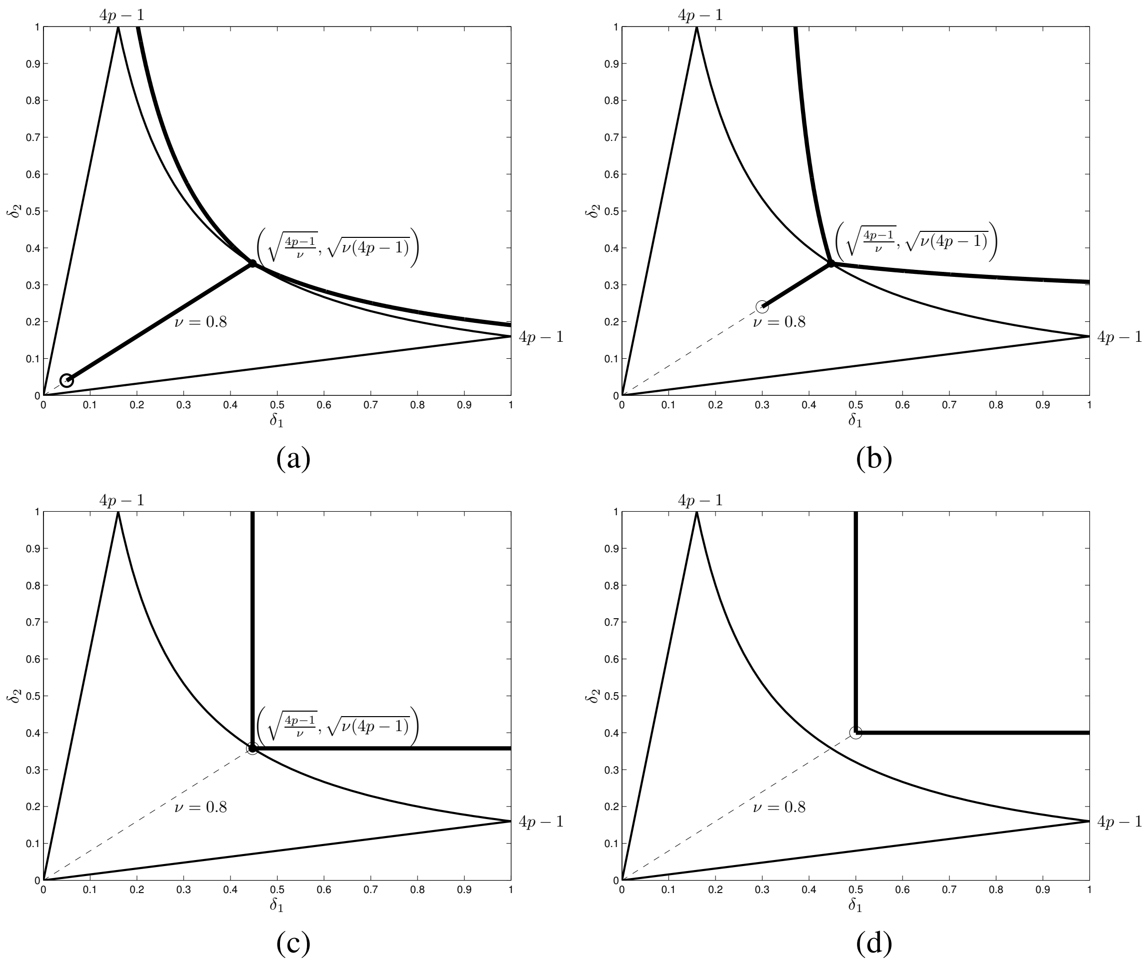

Fig. 7. The successive refinability region in the (δ_1^2, δ_2^2)-plane for several (δ_1^1, δ_2^1) pairs (indicated using $\circ$) satisfying $\delta_2^1 = \nu\delta_1^1$ and $\nu = 0.8$. To emphasize the similarity to the Gaussian case, we set $p = 0.29$ so that $4p - 1 = \rho^2$ with $\rho = 0.4$. The particular choices of δ_1^1 for (a)-(d) are 0.05, 0.3, $\sqrt{\frac{4p-1}{\nu}} = 0.4472$, and 0.5, respectively. When $\mathbf{D}^1 \in \mathcal{E}_2$ as in (a) and (b), the region of $\mathbf{D}^2$ for which successive refinability holds constitutes a line in $\mathcal{E}_2$ (indicated in bold) and everywhere above the bold curve that lies in $\mathcal{E}_1$. When $\mathbf{D}^1 \in \mathcal{E}_1$ as in (c) and (d), successive refinability holds for all (δ_1^2, δ_2^2) in the rectangular region above the bold lines.

Similar to the sanity check we had in Case (iii) for 2-D Gaussian vectors, we observe from (14) and (15) that this region includes all (δ_1^2, δ_2^2) such that $\delta_1^2 \geq \sqrt{\frac{4p-1}{\nu}}$ and $\delta_2^2 \geq \sqrt{\nu(4p-1)}$. Inclusion of this rectangular region intuitively follows from analyses of the previous cases and the observation that the point $\left(\sqrt{\frac{4p-1}{\nu}}, \sqrt{\nu(4p-1)}\right)$ is simultaneously on the line $\delta_2^2 = \nu\delta_1^2$ and on the common boundary of $\mathcal{E}_1$ and $\mathcal{E}_2$.

Figure 7 shows the successive refinability region for several choices of (δ_1^1, δ_2^1).

V. The Rate Loss for 2-D Gaussian Sources

We now analyze the rate loss for 2-D Gaussian sources when the first stage coding is performed optimally, i.e., by expending rate $R(\mathbf{D}^1)$ with $\mathbf{D}^1 \in \mathcal{D}_2$. From Lemma 1, it follows that the minimum second stage rate R^2 in this case is given by

$$\begin{aligned} R^2 &= \min_{E\{\mathbf{d}(\mathbf{X},\hat{\mathbf{X}}^2)\}\leq \mathbf{D}^2} I(\mathbf{X};\hat{\mathbf{X}}^2|\hat{\mathbf{X}}^1_*) \\ &= \min_{E\{\mathbf{d}(\mathbf{Z}^1,\hat{\mathbf{X}}^2-\hat{\mathbf{X}}^1_*)\}\leq \mathbf{D}^2} I(\mathbf{Z}^1;\hat{\mathbf{X}}^2-\hat{\mathbf{X}}^1_*|\hat{\mathbf{X}}^1_*) \ . \end{aligned}$$

Letting $\mathbf{X}^{\Delta}=\hat{\mathbf{X}}^2-\hat{\mathbf{X}}^1_*$ and recalling the independence of $\hat{\mathbf{X}}^1_*$ and $\mathbf{Z}^1$, we further have

$$\begin{aligned} R^2 &= \min_{E\{\mathbf{d}(\mathbf{Z}^1,\mathbf{X}^{\Delta})\}\leq \mathbf{D}^2} I(\mathbf{Z}^1;\mathbf{X}^{\Delta}|\hat{\mathbf{X}}^1_*)+I(\mathbf{Z}^1;\hat{\mathbf{X}}^1_*) \\ &= \min_{E\{\mathbf{d}(\mathbf{Z}^1,\mathbf{X}^{\Delta})\}\leq \mathbf{D}^2} I(\mathbf{Z}^1;\mathbf{X}^{\Delta})+I(\mathbf{Z}^1;\hat{\mathbf{X}}^1_*|\mathbf{X}^{\Delta}) \\ &= \min_{E\{\mathbf{d}(\mathbf{Z}^1,\mathbf{X}^{\Delta})\}\leq \mathbf{D}^2} I(\mathbf{Z}^1;\mathbf{X}^{\Delta}) \end{aligned}$$

where the last step follows from the fact that one can always choose $\mathbf{X}^{\Delta}$ so that it minimizes $I(\mathbf{Z}^1;\mathbf{X}^{\Delta})$, and at the same time, $\hat{\mathbf{X}}^1_*$ is independent from $(\mathbf{Z}^1,\mathbf{X}^{\Delta})$, and hence $\mathbf{Z}^1-\mathbf{X}^{\Delta}-\hat{\mathbf{X}}^1_*$ forms a Markov chain. Since $\mathbf{Z}^1$ is Gaussian (as are all variables) and the distortion measure is square error, we can utilize the solution (1), or equivalently, (2). More specifically, we can write

$$R^2 = R(\mathbf{D}^{\Delta}|\rho_{\mathbf{Z}^1}) \tag{17}$$

where $\mathbf{D}^{\Delta}=[D_1^{\Delta} \quad D_2^{\Delta}]^T \triangleq [D_1^2/D_1^1 \quad D_2^2/D_2^1]^T$. The normalization by D_1^1 and D_2^1 is due to the fact that those correspond to the variances of Z_1^1 and Z_2^1, respectively. Naturally, the behavior of $R(\mathbf{D}^{\Delta}|\rho_{\mathbf{Z}^1})$ depends on which region $\mathbf{D}^{\Delta}$ falls into. Define the sets $\mathcal{D}_j^{\Delta}$ for $j=1,2,3$ as

$$\mathcal{D}_j^{\Delta}=\{(\delta_1^2,\delta_2^2):(\delta_1^{\Delta},\delta_2^{\Delta})\in\mathcal{D}_j(\rho_{\mathbf{Z}^1})\}$$

with $\delta_i^{\Delta}\triangleq 1-D_i^{\Delta}$ for $i=1,2$. Figure 8 shows the three regions for a particular value of $\mathbf{D}^1$. It can be easily shown that $\mathcal{D}_1^{\Delta}\subset\mathcal{D}^1(\rho)$ and $\mathcal{D}_2^{\Delta}\subset\mathcal{D}^1(\rho)\cup\mathcal{D}^2(\rho)$.

We are ready to analyze the maximum possible rate loss. We tackle two possible scenarios separately: (i) non-degenerate refinement, corresponding to $\mathbf{D}^2\in\mathcal{D}_1^{\Delta}\cup\mathcal{D}_2^{\Delta}$, and (ii) degenerate refinement, corresponding to $\mathbf{D}^2\in\mathcal{D}_3^{\Delta}$. When $\mathbf{D}^2\in\mathcal{D}_3^{\Delta}$, at first glance, a direct comparison between two stage and single stage coding may be considered unfair since the rate loss will be partially due to the fact that the second stage is forced to operate in the degenerate region. More specifically, when $\mathbf{D}^2\in\mathcal{D}_3^{\Delta}$ one of the two distortion components can be further reduced without increasing the expended rate in the refinement stage. On the other hand, degenerate refinement is still interesting in those cases where $\mathbf{D}^1$ lies on the boundary between

$\mathcal{D}_2$ and $\mathcal{D}_3$ (depending on which of the two boundaries, only one of the sources is optimally encoded and the other is optimally estimated at the first stage), since these correspond to the sequential coding scenario in Figure 5. We shall analyze degenerate refinement only for such cases.

A. Maximum rate loss in non-degenerate refinement

When $\mathbf{D}^\Delta \in \mathcal{D}_1 \cup \mathcal{D}_2$, the backward channel is given by

$$\mathbf{Z}^1 = \mathbf{X}^\Delta + \mathbf{Z}^\Delta$$

where $\mathbf{Z}^\Delta$ is independent of $\mathbf{X}^\Delta$. Thus, (17) can be further expanded to

$$R^2 = \frac{1}{2}\log\frac{1-\rho^2_{\mathbf{Z}^1}}{D_1^\Delta D_2^\Delta(1-\rho^2_{\mathbf{Z}^\Delta})} \tag{18}$$

where

$$\rho_{\mathbf{Z}^\Delta} = \frac{\max\left\{0, \rho_{\mathbf{Z}^1} - \sqrt{\delta_1^\Delta \delta_2^\Delta}\right\}}{\sqrt{D_1^\Delta D_2^\Delta}} .$$

Now, using (18) with (2), we can write the rate loss at the second stage as

$$\begin{aligned} S^2 &= R(\mathbf{D}^\Delta|\rho_{\mathbf{Z}^1}) + R(\mathbf{D}^1|\rho) - R(\mathbf{D}^2|\rho) \\ &= \frac{1}{2}\log\frac{1-\rho^2_{\mathbf{Z}^1}}{D_1^\Delta D_2^\Delta(1-\rho^2_{\mathbf{Z}^\Delta})} + \frac{1}{2}\log\frac{1-\rho^2}{D_1^1 D_2^1(1-\rho^2_{\mathbf{Z}^1})} - \frac{1}{2}\log\frac{1-\rho^2}{D_1^2 D_2^2(1-\rho^2_{\mathbf{Z}^2})} \\ &= \frac{1}{2}\log\frac{1-\rho^2_{\mathbf{Z}^2}}{1-\rho^2_{\mathbf{Z}^\Delta}} . \end{aligned} \tag{19}$$

As a sanity check, one can re-derive the results in Section III by setting $\rho_{\mathbf{Z}^2} = \rho_{\mathbf{Z}^\Delta}$. For example, successive refinability from $\mathcal{D}_2$ to $\mathcal{D}_1$ corresponds to $\rho_{\mathbf{Z}^2} = \rho_{\mathbf{Z}^\Delta} = 0$, and thus to $\mathbf{D}^2 \in \mathcal{D}_1^\Delta$.

Its simplicity notwithstanding, the expression (19) resisted our efforts for a complete analysis of the maximum rate loss. Resorting to numerical trial and error, we observed that the rate loss can be as high as 0.161 bits. We next show that with proper choice of parameters, one can observe the same rate loss analytically.

For a given ρ, choose $\delta_2^1 = \rho^2\delta_1^1$ and $\delta_1^\Delta = \rho^2_{\mathbf{Z}^1}\delta_2^\Delta$. The intuition behind this choice is that (δ_1^1, δ_2^1) and (δ_1^2, δ_2^2) digress the furthest from being on a line that passes through the origin[2]. Since in this case

[2]With the same token, we could have chosen $\delta_1^1 = \rho^2\delta_2^1$ and $\delta_2^\Delta = \rho^2_{\mathbf{Z}^1}\delta_1^\Delta$. Without loss of generality, we only analyze the former case.

$\mathbf{D}^1, \mathbf{D}^\Delta \in \mathcal{D}_2$, we have

$$\begin{aligned}
\rho_{\mathbf{Z}^\Delta} &= \frac{\rho_{\mathbf{Z}^1} - \sqrt{\delta_1^\Delta \delta_2^\Delta}}{\sqrt{D_1^\Delta D_2^\Delta}} \\
&= \rho_{\mathbf{Z}^1} \sqrt{\frac{D_2^\Delta}{D_1^\Delta}} \\
&= \frac{\rho - \sqrt{\delta_1^1 \delta_2^1}}{\sqrt{D_1^1 D_2^1}} \sqrt{\frac{D_2^\Delta}{D_1^\Delta}} \\
&= \rho \sqrt{\frac{D_1^1 D_2^\Delta}{D_2^1 D_1^\Delta}} \\
&= \rho \sqrt{\frac{(1-\delta_1^1)(1-\delta_2^\Delta)}{1-\rho^2(\delta_1^1 + \delta_2^\Delta - \delta_1^1 \delta_2^\Delta)}},
\end{aligned}$$

and thus

$$1 - \rho_{\mathbf{Z}^\Delta}^2 = \frac{1-\rho^2}{1-\rho^2(\delta_1^1 + \delta_2^\Delta - \delta_1^1 \delta_2^\Delta)} .$$

Similarly, it can be verified after some algebra that

$$\rho_{\mathbf{Z}^2} = \frac{\max\left\{0, \rho - \sqrt{[\delta_2^\Delta(1-\rho^2\delta_1^1) + \rho^2\delta_1^1]\left[\frac{\rho^2(1-\delta_1^1)^2\delta_2^\Delta}{1-\rho^2\delta_1^1} + \delta_1^1\right]}\right\}}{\sqrt{(1-\delta_1^1)(1-\delta_2^\Delta)\left(1-\rho^2(\delta_1^1 + \delta_2^\Delta - \delta_1^1\delta_2^\Delta)\right)}},$$

and hence that

$$S^2 = \frac{1}{2}\log\frac{1}{1-\rho^2}\left(1-\rho^2(\delta_1^1 + \delta_2^\Delta - \delta_1^1\delta_2^\Delta) - \frac{\max\left\{0, \rho - \sqrt{[\delta_2^\Delta(1-\rho^2\delta_1^1) + \rho^2\delta_1^1]\left[\frac{\rho^2(1-\delta_1^1)^2\delta_2^\Delta}{1-\rho^2\delta_1^1} + \delta_1^1\right]}\right\}^2}{(1-\delta_1^1)(1-\delta_2^\Delta)}\right)$$

for any ρ, δ_1^1, and δ_2^Δ.

In all our numerical evaluations, S^2 is maximized as $\rho \to 1$, and up to some possible experimental inaccuracy, the values that attain the maximum are $\delta_1^1 = \rho^2$ and $\delta_2^\Delta = \frac{2}{3}$. Substituting those values into S^2 yields that

$$S^2 = \frac{1}{2}\log\frac{1}{1-\rho^2}\left(1 - \frac{\rho^2(2+\rho^2)}{3} - \frac{3\rho^2}{1-\rho^2}\max\left\{0, 1 - \sqrt{\frac{(5+\rho^2)(2+\rho^4)}{9(1+\rho^2)}}\right\}^2\right) .$$

It can be shown that the second argument of the maximum above is non-negative for $\rho \geq \sqrt{10} - 3 \approx 0.1627$. Thus, for high enough values of ρ, we have

$$S^2 = \frac{1}{2}\log\frac{1}{1-\rho^2}\left(1 - \frac{\rho^2(2+\rho^2)}{3} - \frac{3\rho^2}{1-\rho^2}\left[1 - \sqrt{\frac{(5+\rho^2)(2+\rho^4)}{9(1+\rho^2)}}\right]^2\right) . \tag{20}$$

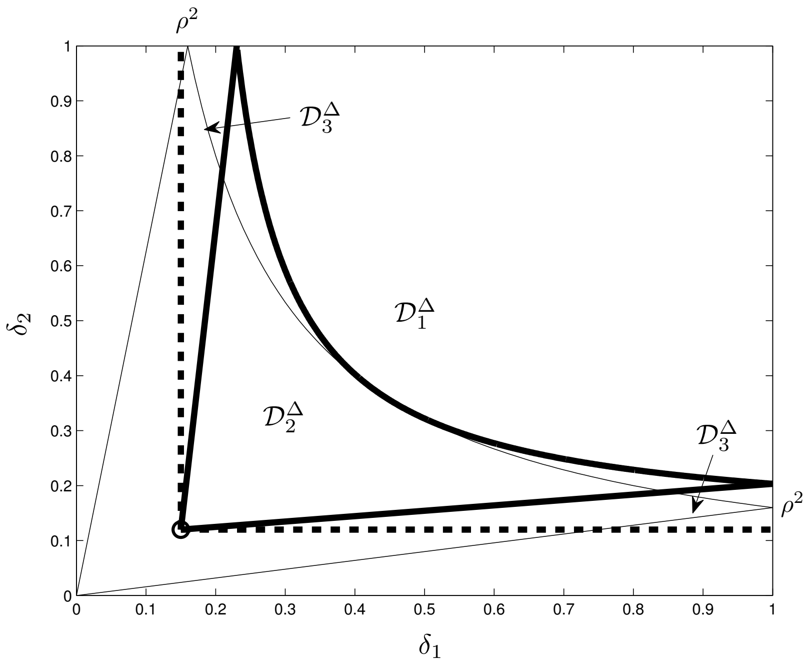

Fig. 8. The refinement regions $\mathcal{D}_j^{\Delta}$, $j = 1, 2, 3$, when $(\delta_1^1, \delta_2^1) = (0.15, 0.12)$ and $\rho = 0.4$.

Finally, taking the limit $\rho \to 1$ in (20), we have

$$S^2 \to \frac{1}{2} \log \frac{5}{4} \approx 0.161 \ .$$

Note that the rate loss can *potentially* be higher than 0.161 with some other choice of δ_1^1, δ_2^1, δ_1^{Δ}, δ_2^{Δ}, and ρ. What we have shown amounts to

$$\sup_{\delta_1^1, \delta_2^1, \delta_1^{\Delta}, \delta_2^{\Delta}, \rho} S^2 \geq 0.161 \ .$$

B. Maximum rate loss in sequential coding of correlated sources

Without loss of generality, we only consider the case where X_1 is optimally encoded and X_2 is optimally estimated, corresponding to $\delta_2^1 = \rho^2 \delta_1^1$. That is, in the first stage we operate on the boundary between $\mathcal{D}_2$ and $\mathcal{D}_3$. Also, since no refinement is required for X_1, we have $\delta_1^2 = \delta_1^1$. This implies that $\mathbf{D}^2 \in \mathcal{D}_1 \cup \mathcal{D}_2$ when $\delta_1^1 \geq \rho^2$ and $\mathbf{D}^2 \in \mathcal{D}_2 \cup \mathcal{D}_3$ otherwise (see Figure 8). The total rate loss for sequential coding of correlated sources is given by

$$\begin{aligned} S^{tot} &= R(\mathbf{D}^{\Delta} | \rho_{\mathbf{Z}^1}) + R(\mathbf{D}^1 | \rho) - R(\mathbf{D}^2 | \rho) \\ &= \frac{1}{2} \log \frac{1}{D_2^{\Delta} D_1^1} - R(\mathbf{D}^2 | \rho) \ . \end{aligned}$$

Now, if $\delta_1^1 < \rho^2$ and $\mathbf{D}^2 \in \mathcal{D}_3$, we have

$$\begin{aligned} S^{tot} &= \frac{1}{2}\log\frac{D_2^2}{D_2^{\Delta}D_1^1} \\ &= \frac{1}{2}\log\frac{D_2^1}{D_1^1} \\ &= \frac{1}{2}\log\frac{1-\rho^2\delta_1^1}{1-\delta_1^1} \end{aligned} \tag{21}$$

which is maximized for a fixed ρ when $\delta_1^1 \to \rho^2$, resulting in

$$S^{tot} = \frac{1}{2}\log(1+\rho^2),$$

which, in turn, is maximized as $\rho \to 1$, yielding $S^{tot} \to 0.5$ bits.

In all other cases,

$$\begin{aligned} S^{tot} &= \frac{1}{2}\log\frac{1}{D_2^{\Delta}D_1^1} - \frac{1}{2}\log\frac{1-\rho^2}{D_1^2 D_2^2(1-\rho^2_{\mathbf{Z}^2})} \\ &= \frac{1}{2}\log\frac{(1-\rho^2\delta_1^1)(1-\rho^2_{\mathbf{Z}^2})}{1-\rho^2}\ . \end{aligned} \tag{22}$$

If $\delta_1^1 \geq \rho^2$, the maximum S^{tot} is achieved when $\rho^2_{\mathbf{Z}^2} = 0$, i.e., when $\mathbf{D}^2 \in \mathcal{D}_1$, yielding

$$S^{tot} = \frac{1}{2}\log\frac{1-\rho^2\delta_1^1}{1-\rho^2},$$

which, as in the previous case, is maximized at $\delta_1^1 = \rho^2$ (this time equality is allowed) and $\rho \to 1$, resulting in $S^{tot} \to 0.5$ bits as well.

Finally, when $\delta_1^1 < \rho^2$ and $\mathbf{D}^2 \in \mathcal{D}_2$, S^{tot} is maximized by the choice $\delta_2^2 = \frac{\delta_1^1}{\rho^2}$, which corresponds to the boundary of $\mathcal{D}_2$ and $\mathcal{D}_3$, and hence the same rate loss is achieved as in the case $\delta_1^1 < \rho^2$ and $\mathbf{D}^2 \in \mathcal{D}_3$. To see that $\delta_2^2 = \frac{\delta_1^1}{\rho^2}$ indeed maximizes (22), it suffices to prove that $\rho^2_{\mathbf{Z}^2}$ is a non-increasing function of δ_2^2 in the interval $\rho^2\delta_1^1 \leq \delta_2^2 \leq \frac{\delta_1^1}{\rho^2}$, which easily follows from the fact that the derivative

$$\frac{d\rho^2_{\mathbf{Z}^2}}{d\delta_2^2} = \frac{\rho - \sqrt{\delta_1^1\delta_2^2}}{(1-\delta_1^1)(1-\delta_2^2)^2}\left(\rho - \sqrt{\frac{\delta_1^1}{\delta_2^2}}\right)$$

is non-positive in the same interval.

In [6], it is shown that the rate loss in the sequential coding problem can be universally bounded by 1 bit for square error distortion. It is somewhat surprising to see that the rate loss can be as high as 0.5 bits even for a jointly Gaussian source. Our result also implies that the universal bound of 1 bit cannot be decreased by more than 0.5 bits.

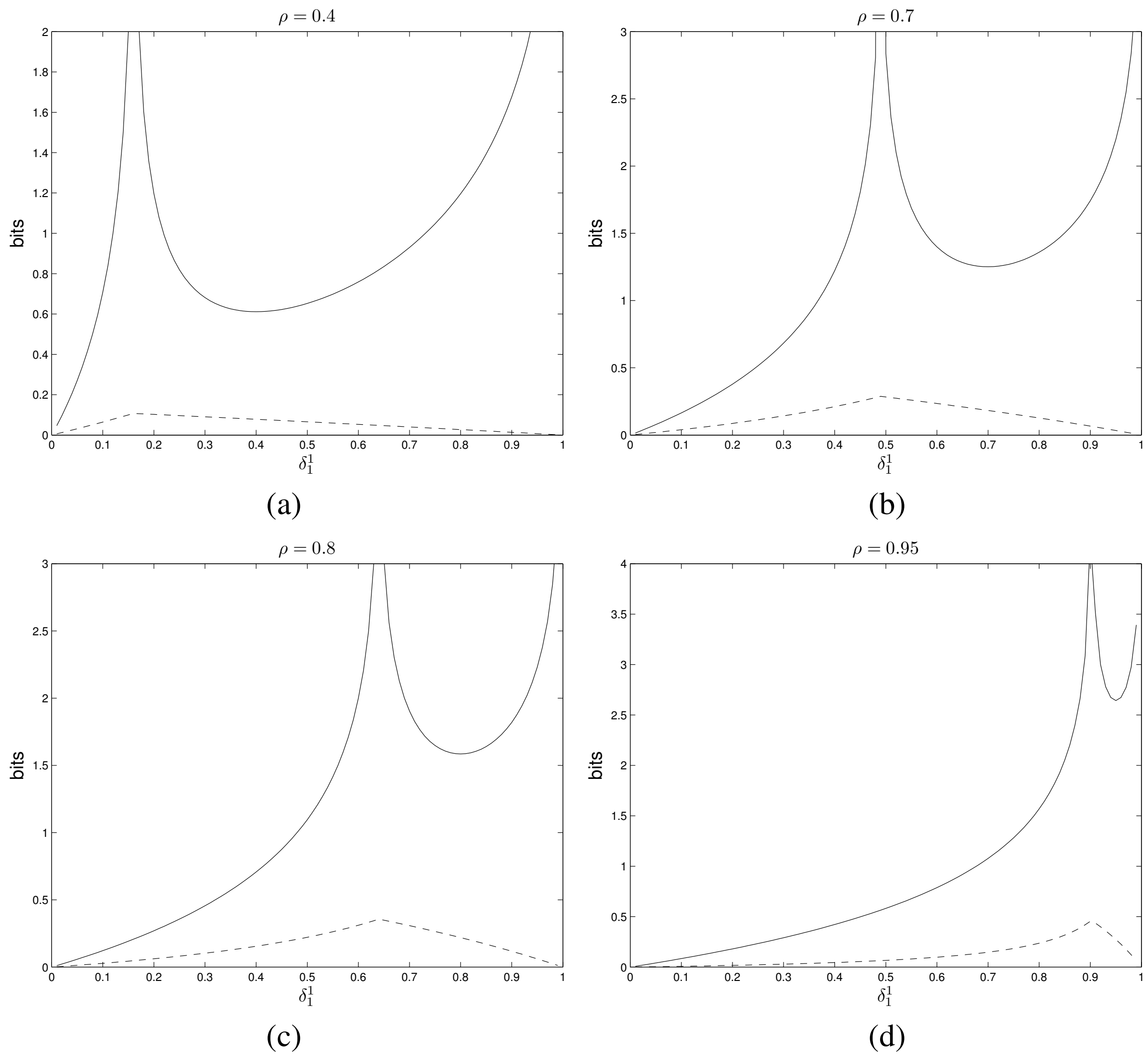

Fig. 9. Comparison of the single-stage rate $R(\mathbf{D}^2|\rho)$ (solid curve) and the rate loss S^{tot} (dashed curve) as a function of δ_1^1. Here, $\mathbf{D}^2 = [\delta_1^1 \;\; \delta_2^2]^T$ where δ_2^2 is chosen so as to maximize S^{tot}.

One may argue that a rate loss of 0.5 bits is insignificant because it is achieved when $\delta_1^1 \to \rho^2 \to 1$ and $\delta_2^2 \to 1$, in which case the single-stage rate-distortion function $R(\mathbf{D}^2|\rho)$ diverges. However, as shown in Figure 9, the maximum rate loss is considerably high when compared with the single-stage rate, except in the vicinity of the points where the latter diverges.

VI. Summary and Conclusions

We have investigated successive refinability of 2-D Gaussian and binary sources under square error and Hamming distortion measures, respectively. We showed that these sources are not successively refinable everywhere, while they exhibit extremely similar behavior. Turning then to rate loss, we showed that for Gaussian sources and with optimal first stage coding, the rate loss in the second stage can be as large as 0.161 bits in an appropriately defined fair scenario. We then argued that the unfair scenario corresponds to sequential coding of two sources and the rate loss in that case, which is universally upper bounded by 1 bit, can be as high as 0.5 bits.

Though the analysis in this paper is limited to $N = 2$ dimensions, the point is to show how vector sources deviate from scalar ones with respect to successive refinability under individual distortion criteria. Based on the results of [15], one can also rigorously analyze the case $N > 2$.

Appendix: Proof of Theorem 2

One can tackle the computation problem by solving the Lagrangian minimization

$$L(\beta_1, \beta_2) = \min_{p_{\hat{\mathbf{X}}}} \left[I(\mathbf{X}; \hat{\mathbf{X}}) + \beta_1 E\{X_1 \oplus \hat{X}_1\} + \beta_2 E\{X_2 \oplus \hat{X}_2\} \right]$$

for all $\beta_1, \beta_2 \geq 0$. We first observe that coding of vectors with individual distortion criteria corresponds to a special case of the successive refinement problem where the objective is to minimize the total rate only. Thus, we can specialize the Kuhn-Tucker conditions derived in [10][3] to

$$\sum_{\mathbf{x}} \frac{p_{\mathbf{X}}(\mathbf{x}) e^{-\beta_1 (x_1 \oplus \hat{x}_1)} e^{-\beta_2 (x_2 \oplus \hat{x}_2)}}{\sum_{\hat{\mathbf{x}}'} p_{\hat{\mathbf{X}}}(\hat{\mathbf{x}}') e^{-\beta_1 (x_1 \oplus \hat{x}'_1)} e^{-\beta_2 (x_2 \oplus \hat{x}'_2)}} \leq 1 \tag{23}$$

for all $\hat{\mathbf{x}}$. The corresponding backward channel is characterized by

$$p_{\mathbf{X}|\hat{\mathbf{X}}}(\mathbf{x}|\hat{\mathbf{x}}) = \frac{p_{\mathbf{X}}(\mathbf{x}) e^{-\beta_1 (x_1 \oplus \hat{x}_1)} e^{-\beta_2 (x_2 \oplus \hat{x}_2)}}{\sum_{\hat{\mathbf{x}}'} p_{\hat{\mathbf{X}}}(\hat{\mathbf{x}}') e^{-\beta_1 (x_1 \oplus \hat{x}'_1)} e^{-\beta_2 (x_2 \oplus \hat{x}'_2)}} \ . \tag{24}$$

for all $\hat{\mathbf{x}}$ with $p_{\hat{\mathbf{X}}}(\hat{\mathbf{x}}) > 0$. We henceforth use the simplified notation $s = e^{-\beta_1}$ and $t = e^{-\beta_2}$.

Guess 1: Our first guess for $p_{\hat{\mathbf{X}}}$ is given in matrix form as

$$\mathbf{P}_{\hat{\mathbf{X}}} = \begin{bmatrix} q & \frac{1}{2} - q \\ \frac{1}{2} - q & q \end{bmatrix} \tag{25}$$

for some $0 \leq q \leq \frac{1}{2}$. It can be shown that the choice

$$q = \frac{p(1+s)(1+t) - \frac{1}{2}(s+t)}{(1-s)(1-t)} \tag{26}$$

satisfies (23) with equality for all $\hat{\mathbf{x}}$. Translating $0 \leq q \leq \frac{1}{2}$ then yields

$$\frac{s+t}{(1+s)(1+t)} \leq 1 - 2p \ . \tag{27}$$

Also, (24) becomes

$$p_{\mathbf{X}|\hat{\mathbf{X}}}(\mathbf{x}|\hat{\mathbf{x}}) = \frac{1}{(1+s)(1+t)} s^{x_1 \oplus \hat{x}_1} t^{x_2 \oplus \hat{x}_2}$$

[3]Specifically, we use $\alpha = 0$ in the formulation of [10].

implying that the optimal backward channel is of the form $\mathbf{X} = \hat{\mathbf{X}} \oplus \mathbf{Z}$ where $\mathbf{Z}$ is independent of $\hat{\mathbf{X}}$ and $p_{\mathbf{Z}}$ is given in matrix form as

$$\mathbf{P}_{\mathbf{Z}} = \begin{bmatrix} \frac{1}{1+s} \\ \frac{s}{1+s} \end{bmatrix} \begin{bmatrix} \frac{1}{1+t} & \frac{t}{1+t} \end{bmatrix} . \tag{28}$$

Thus

$$s = \frac{D_1}{1 - D_1} \tag{29}$$

$$t = \frac{D_2}{1 - D_2} . \tag{30}$$

Using (29) and (30) in (26)-(28) yields $\mathbf{D} \in \mathcal{E}_1$ and (7)-(9). Finally, the value of $R(\mathbf{D})$ for $\mathbf{D} \in \mathcal{E}_1$ can be computed as

$$\begin{aligned} R(\mathbf{D}) &= I(\mathbf{X}; \hat{\mathbf{X}}) \\ &= H(\mathbf{X}) - H(\mathbf{X}|\hat{\mathbf{X}}) \\ &= H(\mathbf{X}) - H(\mathbf{X} \oplus \hat{\mathbf{X}}|\hat{\mathbf{X}}) \\ &= H(\mathbf{X}) - H(\mathbf{Z}) \\ &= H(\mathbf{X}) - \mathcal{H}(D_1) - \mathcal{H}(D_2) . \end{aligned}$$

Guess 2: Since the value of q in (25) becomes $\frac{1}{2}$ at the boundary of $\mathcal{E}_1$, our second guess is of the form

$$\mathbf{P}_{\hat{\mathbf{X}}} = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix} . \tag{31}$$

It can be shown that this guess satisfies (23) if and only if

$$\frac{s+t}{(1+s)(1+t)} \geq 1 - 2p \tag{32}$$

and (24) becomes

$$p_{\mathbf{X}|\hat{\mathbf{X}}}(\mathbf{x}|\hat{\mathbf{x}}) = \left(\frac{1-2p}{s+t}\right)^{x_1 \oplus x_2} \left(\frac{2p}{1+st}\right)^{1 \oplus x_1 \oplus x_2} s^{x_1 \oplus \hat{x}} t^{x_2 \oplus \hat{x}} \tag{33}$$

for $\hat{x}_1 = \hat{x}_2 = \hat{x}$. Observing

$$\begin{aligned} x_1 \oplus x_2 &= x_1 \oplus x_2 \oplus \hat{x} \oplus \hat{x} \\ &= (x_1 \oplus \hat{x}) \oplus (x_2 \oplus \hat{x}) \end{aligned}$$

we conclude from (33) that the optimal backward channel satisfies $\mathbf{X} = \hat{\mathbf{X}} \oplus \mathbf{Z}$ where $\mathbf{Z}$ is independent of $\hat{\mathbf{X}}$ also in this case. However, Z_1 and Z_2 are not independent as in the previous case since

$$\mathbf{P_Z} = \begin{bmatrix} \frac{2p}{1+st} & \frac{(1-2p)t}{s+t} \\ \frac{(1-2p)s}{s+t} & \frac{2pst}{1+st} \end{bmatrix} . \tag{34}$$

It then follows form (34) that

$$\frac{st}{1+st} = \frac{D_1 + D_2 + 2p - 1}{4p} \tag{35}$$

$$\frac{s}{s+t} = \frac{D_1 - D_2 + 1 - 2p}{2(1-2p)} \tag{36}$$

yielding (11). Finally, using $R(\mathbf{D}) = H(\mathbf{X}) - H(\mathbf{Z})$ as above yields

$$\begin{aligned} R(\mathbf{D}) &= H(\mathbf{X}) - \mathcal{H}(2p) - 2p\mathcal{H}\left(\frac{D_1 + D_2 + 2p - 1}{4p}\right) \\ &\quad -(1-2p)\mathcal{H}\left(\frac{D_1 - D_2 + 1 - 2p}{2(1-2p)}\right) \end{aligned}$$

It follows from (27) and (32) that we need not make any other guesses. However, not all $\mathbf{D} \in \mathcal{E}_1^c$ can be spanned using some $s, t \leq 1$. In fact, by careful inspection, we observe that only $\mathbf{D} \in \mathcal{E}_2$ can be attained using the current solution. The boundaries of $\mathcal{E}_2$ correspond to the extreme cases $s = 1$ and $t = 1$ for which $R(\mathbf{D})$ becomes $1 - \mathcal{H}(D_2)$ and $1 - \mathcal{H}(D_1)$, respectively. Thus, the expression for $R(\mathbf{D})$ for $\mathbf{D} \in \mathcal{E}_3$ can be compactly written as

$$R(\mathbf{D}) = 1 - \mathcal{H}(\min\{D_1, D_2\}) .$$

References

[1] M. Effros, "Distortion-rate bounds for fixed- and variable-rate multiresolution source coding," *IEEE Transactions on Information Theory*, 45(6):1887–1910, September 1999.

[2] M. Effros, "Universal multiresolution source codes," *IEEE Transactions on Information Theory*, 47(6):2113–2129, September 2001.

[3] A. El Gamal and T. M. Cover, "Achievable rates for multiple descriptions," *IEEE Transactions on Information Theory*, 28(6):851–857, November 1982.

[4] W. H. R. Equitz and T. M. Cover, "Successive refinement of information," *IEEE Transactions on Information Theory*, 37(2):269–275, March 1991.

[5] V. N. Koshelev, "Hierarchical coding of discrete sources," *Probl. Pered. Inform.*, 16(3):31–49, 1980.

[6] L. Lastras and T. Berger, "All sources are nearly successively refinable," *IEEE Transactions on Information Theory*, 47(3):918–926, March 2001.

[7] J. Nayak and E. Tuncel, "Successive coding of correlated sources," *IEEE International Symposium on Information Theory*, Nice, France, June 2007, pp. 1471–1475.

[8] J. Nayak and E. Tuncel, "Successive coding of correlated sources," submitted to *IEEE Transactions on Information Theory*. Available at http://www.ee.ucr.edu/∼ertem/publications.html

[9] B. Rimoldi, "Successive refinement of information: Characterization of the achievable rates," *IEEE Transactions on Information Theory*, 40(1):253–259, January 1994.

[10] E. Tuncel and K. Rose, "Computation and analysis of the N-layer scalable rate-distortion function," *IEEE Transactions on Information Theory*, 49(5):1218–1230, May 2003.

[11] E. Tuncel and K. Rose, "Error exponents in scalable source coding," *IEEE Transactions on Information Theory*, 49(1):289–296, January 2003.

[12] E. Tuncel and K. Rose, "Additive successive refinement," *IEEE Transactions on Information Theory*, 49(8):1983–1991, August 2003.

[13] H. Wang and P. Viswanath, “Vector Gaussian multiple description with individual and central receivers,” *IEEE Transactions on Information Theory*, 53(6):2133–2153, June 2007.

[14] H. Viswanathan and T. Berger, “Sequential coding of correlated sources,” *IEEE Transactions on Information Theory*, 46(1):236–246, January 2000.

[15] J. Xiao and Q. Luo, “Compression of correlated Gaussian sources under individual distortion criteria,” *43rd Allerton Conference on Communication, Control, and Computing*, September 2005, pp. 438–447.