Successive Refinement of Vector Sources under Individual Distortion Criteria

Jayanth Nayak*, Ertem Tuncel*, Deniz Gunduz†, Elza Erkip†
*University of California, Riverside, CA
Email: {jnayak,ertem}@ee.ucr.edu
†Polytechnic University, Brooklyn, NY
E-mail: dgundu01@utopia.poly.edu, elza@poly.edu

***Abstract*— The well-known successive refinement scenario 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 turns out that successive refinability is not granted everywhere, unlike in the 1-D case for the same sources. Moreover, the behavior of these sources with respect to successive refinability exhibit remarkable similarity.**

I. Introduction

We extend the well-known successive refinement scenario to vector sources where individual distortion constraints are posed on every vector component. The single-layer counterpart of this problem was addressed in [7], and it was later argued in [4] that the single-layer version is in fact a special case of what is referred to as *robust descriptions* [1].

In this work, the achievability region of the scalar successive refinement problem, which was derived independently by Koshelev [3] and Rimoldi [5], is extended to cover the vector sources under individual distortion criteria in a straightforward manner. This extension, in turn, is utilized for the derivation of a necessary and sufficient condition for "vector successive refinability." Not surprisingly, this condition is also a straightforward extension of the Markovity condition derived in [2]. We then use the Markovity condition to investigate whether vector successive refinability holds for two interesting cases: (i) 2-D Gaussian vectors and square-error criterion on each vector component, and (ii) 2-D binary symmetric vectors and Hamming distortion on each component. 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 these sources with respect to successive refinability exhibit remarkable similarity.

II. Extension of the Achievability Region for Vector Sources to Multi-Stage Coding

We first repeat here the single-letter characterization in [7] for single-stage coding. Denoting the memoryless source and the auxiliary reconstruction variable by $\mathbf{X} = [X_1\, X_2\, \ldots\, X_N]^T$ and $\hat{\mathbf{X}} = [\hat{X}_1\, \hat{X}_2\, \ldots\, \hat{X}_N]^T$, respectively, and using the shorthand notation $\mathbf{d}(\mathbf{X}, \hat{\mathbf{X}})$ for $d_i(X_i, \hat{X}_i)$ with $i = 1, 2, \ldots, N$, we have

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

where we use the convention that $\mathbf{a} \leq \mathbf{b}$ means $a_i \leq b_i$ for $i = 1, 2, \ldots, N$. Observe that this result is a special case of the robust descriptions result in [1]. More specifically, in the robust descriptions scenario, individual distortion criteria are allowed to be of the form $E\{d_i(\mathbf{X}, \hat{X}_i)\} \leq D_i$.

The following lemma extends this result to L stages.

Lemma 1: A distortion vector sequence $\mathbf{D}^1 \geq \mathbf{D}^2 \geq \cdots \geq \mathbf{D}^L$ and a cumulative rate sequence $R^1 \leq R^2 \leq \cdots \leq R^L$ are achievable[1] if and only if there exist auxiliary vectors $\hat{\mathbf{X}}^1, \hat{\mathbf{X}}^2, \ldots, \hat{\mathbf{X}}^L$ satisfying

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

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

We omit the proof, since it is a straightforward extension of the proofs in [3], [5].

Corollary 1: The 4-tuple $\{\mathbf{D}^1, \mathbf{D}^2, R(\mathbf{D}^1), R(\mathbf{D}^2)\}$ is achievable 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 $\{\mathbf{D}^1, \mathbf{D}^2, R(\mathbf{D}^1), R(\mathbf{D}^2)\}$, 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) from the chain rule, and (b) and (e) 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 $\mathbf{X} - \hat{\mathbf{X}}^2_* - \hat{\mathbf{X}}^1_*$. ■

III. Analysis of Vector Successive Refinability for Two Example Sources

We now investigate whether conditions in Corollary 1 are satisfied for (i) 2-D Gaussian sources under individual square-error distortion, and (ii) 2-D binary symmetric vectors and

[1] Achievability can be defined either in a weak sense as in [3] or in a strong sense as in [5].

individual Hamming distortion. It turns out that for certain values of $\mathbf{D}^1$ and $\mathbf{D}^2$, these sources are not successively refinable.

A. 2-D Gaussian Sources

Let the covariance matrix of the source $\mathbf{X}$ be given by

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

where $0 < \rho < 1$. For simplicity, we will use the notation $\delta_i = 1 - D_i$ and $\delta_i^j = 1 - D_i^j$ with proper i, j. Define three regions in the unit square on the $\mathbf{D}$-plane as

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

Figure 1 depicts the three regions. It was established in [4], [7] that in the non-degenerate region $\mathcal{D}_1 \cup \mathcal{D}_2$, the optimal *backward* test channel for the single-stage problem 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

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

Region $\mathcal{D}_3$ is degenerate in the sense that, for example, if $\rho^2 > \frac{\delta_2}{\delta_1}$, or equivalently $D_2 > 1 - \rho^2(1 - D_1)$, D_2 can be further reduced to $1 - \rho^2(1 - D_1)$ without increasing $R(\mathbf{D})$. Similarly for the case $D_1 > 1 - \rho^2(1 - D_2)$.

We now consider successive refinement for 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 case. In all three cases, letting $\mathbf{X} = \hat{\mathbf{X}}_*^1 + \mathbf{Z}^1 = \hat{\mathbf{X}}_*^2 + \mathbf{Z}^2$, the Markovity question boils down to whether or not

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

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

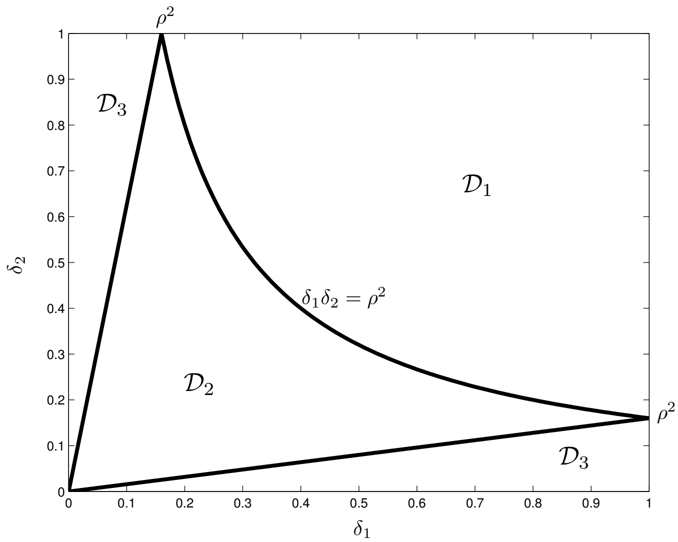

Fig. 1. The three regions $\mathcal{D}_1$, $\mathcal{D}_2$, and $\mathcal{D}_3$ for $\rho = 0.4$.

<u>*Case i:*</u> $\mathbf{D}^1, \mathbf{D}^2 \in \mathcal{D}_1$. Since

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

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

<u>*Case ii:*</u> $\mathbf{D}^1, \mathbf{D}^2 \in \mathcal{D}_2$. 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 $\mathbf{D}^2 \le \mathbf{D}^1$, it suffices to check

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

Re-writing this as

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

reveals that Markovity 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* the 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, this implies $\delta_1^1\delta_2^2 = \delta_1^2\delta_2^1$, or equivalently, 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, this implies $\delta_1^1\delta_2^2 = \delta_1^2\delta_2^1$, or equivalently,

$$\frac{\delta_2^1}{\delta_1^1} = \frac{\delta_2^2}{\delta_1^2} \ . \quad (2)$$

That is, (δ_1^1, δ_2^1) and (δ_1^2, δ_2^2) must lie on a line passing through the origin.

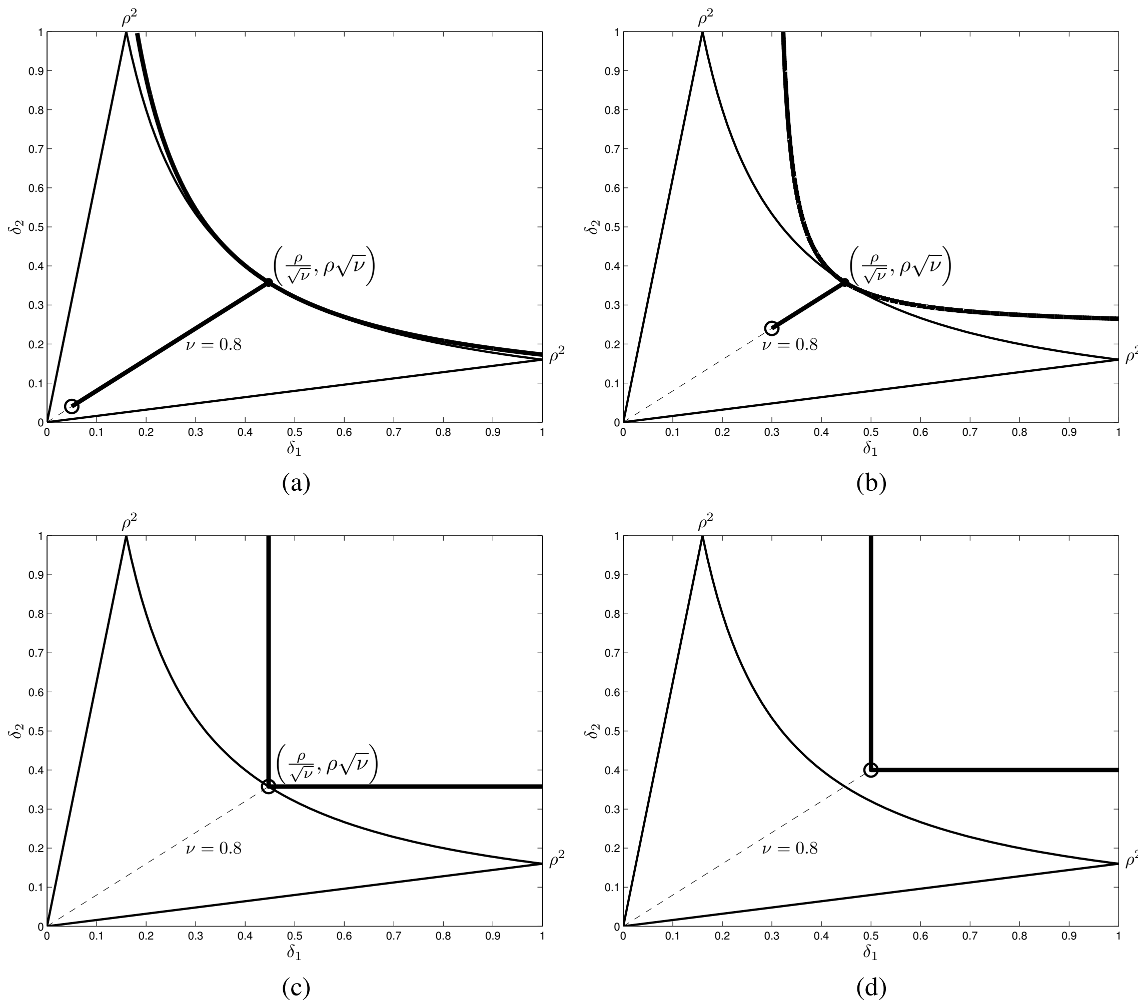

Fig. 2. The successive refinability region in the (δ_1^2, δ_2^2)-plane for several (δ_1^1, δ_2^1) pairs (indicated using $\circ$) satisfying $\delta_2^1 = \nu\delta_1^1$ and $\nu = 0.8$. We set $\rho = 0.4$. The particular choices of δ_1^1 for (a)-(d) are 0.05, 0.3, $\frac{\rho}{\sqrt{\nu}} = 0.4472$, 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 $\mathcal{D}_1$.

Case iii: $\mathbf{D}^1 \in \mathcal{D}_2$ 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 $\rho^2 \leq \nu \leq \frac{1}{\rho^2}$ and $\rho \geq \sqrt{\nu}\delta_1^1$. 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

$$(\delta_1^2 - \delta_1^1)(\delta_2^2 - \nu\delta_1^1) \geq \left(\rho - \sqrt{\nu}\delta_1^1\right)^2 . \tag{3}$$

As a sanity check, this region should include all (δ_1^2, δ_2^2) such that $\delta_1^2 \geq \frac{\rho}{\sqrt{\nu}}$ and $\delta_2^2 \geq \rho\sqrt{\nu}$. This follows from analyses 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 (2) 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 (1) that one can do the same from $\mathbf{D}^0$ to any $\mathbf{D}^2 \leq \mathbf{D}^0$. Indeed, the point $\delta_1^2 = \frac{\rho}{\sqrt{\nu}}$ and $\delta_2^2 = \rho\sqrt{\nu}$ satisfies (3) with equality, and the above claim is corroborated. However, it is also clear from (3) that the successive refinability region is not limited to that rectangular region.

Figure 2 shows the successive refinability region for several choices of (δ_1^1, δ_2^1).

B. 2-D Binary Symmetric Sources

Let the pmf of the source be given by

$$\mathbf{P}_{\mathbf{X}} = \begin{bmatrix} p & \frac{1}{2} - p \\ \frac{1}{2} - p & p \end{bmatrix}$$

where $\frac{1}{4} \leq p \leq \frac{1}{2}$. 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 $\delta_i = 1 - 2D_i$ and $\delta_i^j = 1 - 2D_i^j$ with 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 $\{(\delta_1, \delta_2) : 0 < \delta_1 \leq 1, 0 < \delta_2 \leq 1, \}$.

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

Theorem 1: The rate-distortion function for 2-D binary symmetric sources is given by

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

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

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

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

$$\mathbf{P}_{\hat{\mathbf{X}}_*} = \begin{bmatrix} q & \frac{1}{2} - q \\ \frac{1}{2} - q & q \end{bmatrix} \tag{4}$$

with

$$q = \frac{1}{4}\left[1 + \frac{4p-1}{\delta_1\delta_2}\right] \tag{5}$$

and

$$\mathbf{P}_{\mathbf{Z}} = \begin{bmatrix} (1-D_1)(1-D_2) & (1-D_1)D_2 \\ D_1(1-D_2) & D_1D_2 \end{bmatrix} \tag{6}$$

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

$$\mathbf{P}_{\hat{\mathbf{X}}_*} = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix} \tag{7}$$

and

$$\mathbf{P}_{\mathbf{Z}} = \frac{1}{2}\begin{bmatrix} 2 - D_1 - D_2 - (1-2p) & D_2 - D_1 + (1-2p) \\ D_1 - D_2 + (1-2p) & D_1 + D_2 - (1-2p) \end{bmatrix} \tag{8}$$

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 1 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 pmfs of these random variables satisfy

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

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

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

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 (6) that

$$\mathcal{F}(\mathbf{P}_{\mathbf{Z}^i}) = \begin{bmatrix} 1 & \delta_2^i \\ \delta_1^i & \delta_1^i\delta_2^i \end{bmatrix} \tag{9}$$

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 pmf. But this is always granted since we only focus on $\mathbf{D}_2 \le \mathbf{D}_1$.

<u>*Case ii:*</u> $\mathbf{D}^1, \mathbf{D}^2 \in \mathcal{E}_2$. We have from (8) that

$$\mathcal{F}(\mathbf{P}_{\mathbf{Z}^i}) = \begin{bmatrix} 1 & \delta_2^i \\ \delta_1^i & 4p-1 \end{bmatrix} \tag{10}$$

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

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

to be valid. It can easily be seen that this requires

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

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

<u>*Case iii:*</u> $\mathbf{D}^1 \in \mathcal{E}_2$ and $\mathbf{D}^2 \in \mathcal{E}_1$. We assume without loss of generality that $\delta_2^1 = \nu\delta_1^1$ with $4p - 1 \le \nu \le \frac{1}{4p-1}$ and $\delta_1^1 \le \sqrt{\frac{4p-1}{\nu}}$. It follows from (9) and (10) that we need

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

to be valid, where

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

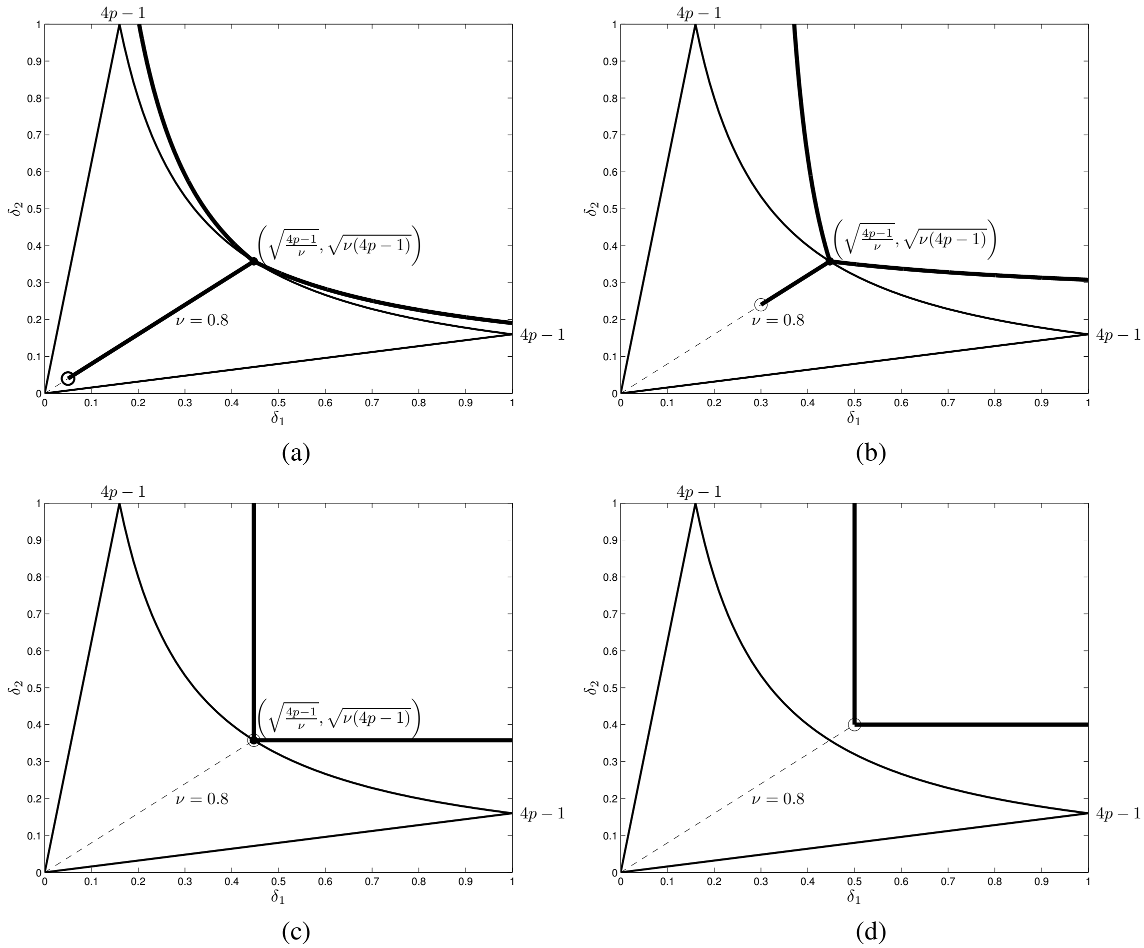

Fig. 3. The successive refinability region in the (δ_1^2, δ_2^2)-plane for several (δ_1^1, δ_2^1) pairs (indicated using $\circ$) satisfying $\delta_2^1 = \nu\delta_1^1$ and $\nu = 0.8$. To emphasize the similarity to the Guassian case, we set $p = 0.29$ so that $4p-1 = \rho^2$ with $\rho = 0.4$. The 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 $\mathcal{E}_1$.

Observe that the entries of $\mathbf{P_N}$ 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

$$\begin{aligned}
(\delta_1^2 + \delta_1^1)(\delta_2^2 - \nu\delta_1^1) &\geq 4p - 1 - \nu\left(\delta_1^1\right)^2 && (11)\\
(\delta_1^2 - \delta_1^1)(\delta_2^2 + \nu\delta_1^1) &\geq 4p - 1 - \nu\left(\delta_1^1\right)^2 && (12)\\
(\delta_1^2 - \delta_1^1)(\delta_2^2 - \nu\delta_1^1) &\geq -\left[4p - 1 - \nu\left(\delta_1^1\right)^2\right]\,. && (13)
\end{aligned}$$

Since $\mathbf{D}^2 \leq \mathbf{D}^1$ translates to $\delta_1^2 \geq \delta_1^1$ and $\delta_2^2 \geq \nu\delta_1^1$, and because $4p - 1 \geq \nu\left(\delta_1^1\right)^2$, (13) becomes vacuous.

Similar to the sanity check we had in Case iii for 2-D Gaussian vectors, we observe from (11) and (12) 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 3 shows the successive refinability region for several choices of (δ_1^1, δ_2^1).

Appendix: Proof of Theorem 1

One can tackle the computation problem by solving the Lagrangian minimization

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

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 [6][2] to

$$\sum_{\mathbf{x}} \frac{p_{\mathbf{X}}(\mathbf{x}) e^{-\beta_1(x_1 \oplus \hat{x}_1)} e^{-\beta_2(x_2 \oplus \hat{x}_2)}}{\sum_{\hat{\mathbf{x}}'} p_{\hat{\mathbf{X}}}(\hat{\mathbf{x}}') e^{-\beta_1(x_1 \oplus \hat{x}'_1)} e^{-\beta_2(x_2 \oplus \hat{x}'_2)}} \leq 1 \qquad (14)$$

[2]Specifically, we use $\alpha = 0$ in the formulation of [6].

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

$$p_{\mathbf{X}|\hat{\mathbf{X}}}(\mathbf{x}|\hat{\mathbf{x}}) = \frac{p_{\mathbf{X}}(\mathbf{x}) e^{-\beta_1(x_1 \oplus \hat{x}_1)} e^{-\beta_2(x_2 \oplus \hat{x}_2)}}{\sum_{\hat{\mathbf{x}}'} p_{\hat{\mathbf{X}}}(\hat{\mathbf{x}}') e^{-\beta_1(x_1 \oplus \hat{x}'_1)} e^{-\beta_2(x_2 \oplus \hat{x}'_2)}} . \quad (15)$$

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

$$\mathbf{P}_{\hat{\mathbf{X}}} = \begin{bmatrix} q & \frac{1}{2} - q \\ \frac{1}{2} - q & q \end{bmatrix} \quad (16)$$

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

$$q = \frac{p(1+s)(1+t) - \frac{1}{2}(s+t)}{(1-s)(1-t)} \quad (17)$$

satisfies (14) with equality for all $\hat{\mathbf{x}}$. Translating $0 \leq q \leq \frac{1}{2}$ then yields

$$\frac{s+t}{(1+s)(1+t)} \leq 1 - 2p . \quad (18)$$

Also, (15) becomes

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

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

$$\mathbf{P}_{\mathbf{Z}} = \begin{bmatrix} \frac{1}{1+s} \\ \frac{s}{1+s} \end{bmatrix} \begin{bmatrix} \frac{1}{1+t} & \frac{t}{1+t} \end{bmatrix} . \quad (19)$$

Thus

$$s = \frac{D_1}{1 - D_1} \quad (20)$$

$$t = \frac{D_2}{1 - D_2} . \quad (21)$$

Using (20) and (21) in (17)-(19) yields $\mathbf{D} \in \mathcal{E}_1$ and (4)-(6). 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 (16) becomes $\frac{1}{2}$ at the boundary of $\mathcal{E}_1$, our second guess is of the form

$$\mathbf{P}_{\hat{\mathbf{X}}} = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix} . \quad (22)$$

It can be shown that this guess satisfies (14) if and only if

$$\frac{s+t}{(1+s)(1+t)} \geq 1 - 2p \quad (23)$$

and (15) becomes

$$p_{\mathbf{X}|\hat{\mathbf{X}}}(\mathbf{x}|\hat{\mathbf{x}}) = \left(\frac{1-2p}{s+t}\right)^{x_1 \oplus x_2} \left(\frac{2p}{1+st}\right)^{1 \oplus x_1 \oplus x_2} s^{x_1 \oplus \hat{x}} t^{x_2 \oplus \hat{x}} \quad (24)$$

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 (24) 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

$$\mathbf{P}_{\mathbf{Z}} = \begin{bmatrix} \frac{2p}{1+st} & \frac{(1-2p)t}{s+t} \\ \frac{(1-2p)s}{s+t} & \frac{2pst}{1+st} \end{bmatrix} . \quad (25)$$

It then follows form (25) that

$$\frac{st}{1+st} = \frac{D_1 + D_2 + 2p - 1}{4p} \quad (26)$$

$$\frac{s}{s+t} = \frac{D_1 - D_2 + 1 - 2p}{2(1-2p)} \quad (27)$$

yielding (8). Finally, using $R(\mathbf{D}) = H(\mathbf{X}) - H(\mathbf{Z})$ as above yields

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

It follows from (18) and (23) 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] A. El Gamal and T. M. Cover, "Achievable rates for multiple descriptions," *IEEE Transactions on Information Theory*, 28(6):851–857, November 1982.

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

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

[4] J. Nayak and E. Tuncel, "Successive coding of correlated sources," submitted to *IEEE Transactions on Information Theory*.

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

[6] 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.

[7] 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.